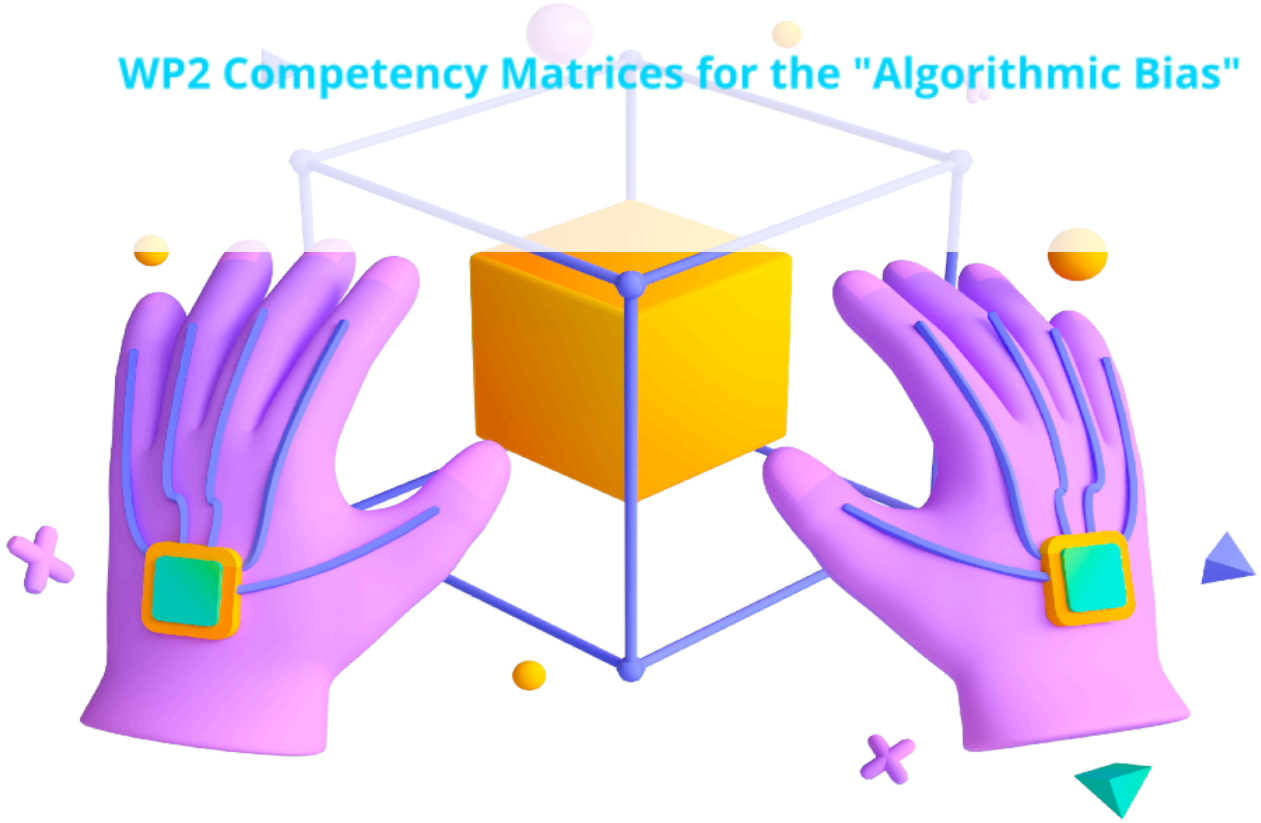
CHALLENGING BIAS IN BIG DATA USER FOR AI AND MACHINE LEARNING

WP2 Competency Matrices for the "Algorithmic Bias"



Co-funded by the European Union

Project number: 2022-1-ES01-KA220-HED-000085257

Content

1. INTRODUCTION: THE CHARLIE PROJECT

1.1. The Charlie project: overview

CHARLIE is an ERASMUS+ KA2 project with an implementation period of 30 months, between 30/12/2022 - 29/06/2025. The project is being conducted by a consortium of SIX (6) partners from five (5) European countries: Spain, Portugal, Romania, Finland and Denmark.

Artificial intelligence (AI) is part of our everyday lives. From the algorithm that recognizes our face when we are walking down the street feeding police biometric security services to the algorithm that chooses the advertising we will see in our social media, AI is everywhere. However, although Machine Learning (ML) and AI are mathematics, they are only sometimes suitable, and this happens because the data processed to come to any conclusion can be, and often is, biased. Social sciences have been studying Human Bias for many years. It arises from the implicit association that reflects bias we are unaware of, which can result in multiple adverse outcomes. AI and ML are not designed to make ethical decisions; that is not an algorithm for ethics. It will always make predictions based on how the world works today, therefore contributing to fostering the bias and discriminatory practices systemically rooted in our societies today. With the widespread use of AI and ML technologies, often owned by big tech companies with the only objective of making profits, there is an urgent need to bring a human-centered approach to tech and use it to solve social problems instead of contributing to them. In its Communication of 25/04/18 and 7/12/18, the EC set out its vision for artificial intelligence, which supports “ethical, secure and cutting-edge AI made in Europe”. AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good to improve human welfare and freedom. While offering significant opportunities, AI systems also give rise to certain risks that must be handled appropriately and proportionately. We now have an essential window of opportunity to shape their development.

CHARLIE intends to ensure we can trust the sociotechnical environments in which they are embedded. We also want producers of AI systems to get a competitive advantage by embedding Trustworthy AI in their products and services. This entails seeking to maximize the benefits of AI systems while at the same time preventing and minimizing their risks. Higher Education (HE), Adult Education (AE), and Youth require new and innovative curricula that can meet this skills gap and equip learners with the knowledge and skills to contribute to a more ethical approach to tech development. The need to make tech education more human is aligned with the Digital Education

Action Plan, which includes specific actions to address the ethical implications and challenges of using AI and data in education and training.

1.2 Project objectives

CHARLIE aims at challenging the bias in big data used for AI and machine learning by bringing a greater level of awareness regarding the negative impacts of the lack of a critical and ethical approach to techEd. CHARLIE main objectives are to:

1. Increase the capacity of HE institutions to provide its students online learning opportunities that meet society needs but also are tailored to students learning needs;
2. Increase Tech students' social and ethical competencies, allowing them to engage positively, critically and ethically with AI/ML technology;
3. Equip teachers/professors with digital and engaging approaches to effective teaching the topic (especially in online teaching);
4. Create synergies between HE organizations and AE and Youth at regional level in the field of AI ethics education;
5. Potentiate the transferability of academia courses on AI biases to AE and Vocational Education and Training (VET);
6. Raise awareness about the topic at society level

1.3. Expected overall results of the Charlie project per WP

1- Competency Matrices for the "Algorithmic Bias" course (EQF6), the "Ethical AI microcredential" (EQF4) and the Serious Game (EQF2) **(WP2)**

2- "Algorithmic Bias" course (HE) - bLearning approach: a) synchronous sessions (preferably face-to-face), b) eLearning asynchronous sessions for theoretical component, complemented with a digital serious game as formative assessment and online summative assessment, delivered as .scorm ready to be upload in any LMS for students enrolled in Big Data, AI, machine learning, deep learning related courses. **(WP3)**

3- "Algorithmic Bias toolkit for synchronous sessions" - for Teachers/Professors/Lecturers implementing the synchronous sessions (face-to-face or online) that complement the eLearning asynchronous sessions. **(WP3)**

4- A guideline for boosting the capacity of university administrators/management of Education and Training departments in charge of digital education provision to foster digital education delivery on Algorithmic Bias. **(WP3)**

5- Webinars to foster peer-learning and discuss the role of HE institutions in techEd, to foster interdisciplinary of social sciences in tech education and to discuss approaches for interoperability with Youth and AE. **(WP3)**
6- Self-paced "Ethical AI" microcredential (EQF4) for Adult learners to be taken online and asynchronous. **(WP4)**
7- Digital Serious game (EQF2) for Youth 12-18 years (mainly disadvantaged/young women) to foster gender representation in STEM from a young age. **(WP4)**
8- Toolkit to support Adults and Youth in upskilling into Ethical AI. **(WP4)**
9- Policy recommendation for recognition of the microcrential for adult learners in accessing HE courses in technological fields. **(WP4)**

1.4. Charlie Target Groups

The CHARLIE project main target-groups are:
- **HE institutions** that provide Big Data, AI, machine learning, deep learning related learning opportunities/courses and its respective administrators/management of Education and Training departments
- **HE Students** enrolled in Big Data, AI, machine learning, deep learning related courses
- **HE Teachers/Professors/Lecturers** from either social sciences and techEd
- **AE organisations** and their staff
- **Adults** of all ages and socio-economic backgrounds, aiming to progress towards higher qualification levels relevant for the labour market and for active participation in society.
- **Youth organisations** and their staff
-**Youth** (12 to 18 years old) - specially young women and youth from social disadvantaged contexts (facing barriers linked to education system; cultural differences; social barriers; economic barriers; barriers linked to discrimination; and geographical barriers)

Main target groups of the communication and project promotion activities and products:
- National regulators and policymakers;
- The wider public
- Higher Education institutions (University lecturers, Students (undergraduates and post-graduates) and University administrators)
- Adult Education Providers
- Youth Organisations
- VET providers

- AI Forums and platforms
- ONGs that tackle different Bias and Discrimination
- Educational technology providers
- Mass media operators (TV, e-newspapers, radios, etc)
- AI/ML' SMEs and AI/ML enterprises.
- Associations of higher education providers including related collaborative organizations for best practice promotion at the EU level and snowballing dissemination to their university members (e.g.: European University Association, Science Business Network, Triple Helix Association, AMBA, University-Industry Innovation Network, European Business Angels Network (for investing in ethical AI innovations)
- EU wide student bodies
- Skill monitoring/development associations (e.g.: DigCompEdu communities).

1.5. WP2 Objectives

Competency Matrices for the "Algorithmic Bias"' WP specific objectives are:

- to set a common understanding about the purpose and learning goals of a learning pathway that will be developed for HE students in the "Algorithmic Bias" course;

- to set a common understanding about the purpose and learning goals of a learning pathway that will be developed for AE learners in the "Ethical AI microcredential";

- to set a common understanding about the purpose and learning goals of a learning pathway that will be developed for youth in the serious game on ethical AI.

These objectives contribute to the general objectives by setting the cornerstones for the following activities, allowing teachers, mentors and learners to have a clear vision of the goals for the different learning pathways, working to increase HE, Adult and Youth organisations capacity to provide learning opportunities that meet society needs but also are tailored to learners learning needs.

1.6. Expected results of the activities of WP2 and participants

1.6.1. Competency Matrix EQF 6 for "Algorithmic Bias" course (HE students) - Learning outcomes approach, 2 ECVS points, accreditation recommendations, pre-requirements for enrolling, contact hours, total

workload, integration of EntreComp competences (eg.: Ethical and Sustainable Thinking), DigComp 2.0 competences (e.g.: Protecting health and well-being) and GrenComp competences (e.g.: Supporting fairness).
Competence Units:
CU1 - Algorithms Models and Limitations UIB
CU2 - Data Fairness and Bias in AI UA
CU3 - AI Privacy and convenience UA
CU4 - AI Ethics, a practical approach VAMK
CU5 - Case studies and project VAMK

1.6.2. Competency Matrix EQF 4 for the "Ethical AI micro credential"
- Learning outcomes approach, ECVS points potential, accreditation recommendations, pre-requirements for enrolling, contact hours, total workload, integration of EntreComp competences (eg.: Ethical and Sustainable Thinking), DigComp 2.0 competences (e.g.: Protecting health and well-being) and GrenComp competences (e.g.: Supporting fairness).
Competence Units:
CU1 - What is Algorithmic Bias? HELIX
CU2 - Non-maleficience HELIX
CU3 – Accountability ITC
CU4 – Transparency ITC
CU5 - Human rights and fairness ITC
CU6 - AI Ethics, a practical approach HELIX

1.6.3. Learning Outcomes for Youth EQF 2 (serious game)
- Featuring Learning outcomes approach, pre-requirements for enrolling, contact hours, integration of EntreComp competences (e.g.: Ethical and Sustainable Thinking), DigComp 2.0 competences (e.g.: Protecting health and well-being) and GrenComp competences (e.g.: Supporting fairness). General contents: Systemic inequalities in society, Basic mechanics of artificial intelligence systems, political agendas, AI impacts in the world.

2. INTRODUCTION TO THE TOPIC: ALGORITHMS, AI, BIASES AND ETHICS

Algorithms play a crucial role in modern society, as they assist with various services like recommending songs, movies, or friends (Paraschakis, 2018; Milano et al., 2020). They are also used in schools, hospitals, financial institutions, courts, and governments to make crucial decisions (Obermeyer et al., 2019). However, using algorithms can lead to ethical risks (Floridi et al., 2018).

Algorithms, such as translations (Prates et al., 2020) or job advertisements (Lambrecht and Tucker, 2019), can be biased. For example, higher-paying jobs may be shown to men more than women. This can also lead to unfair treatment in healthcare, with white patients receiving better care than black patients (Obermeyer et al. 2019). While people try to fix these problems, more issues keep coming up.

Since 2012, there has been a lot of focus on artificial intelligence (AI) and its ethical implications (Perrault et al. 2019). In 2016, a study tried to map these concerns (Mittelstadt et al. 2016), but the field has grown and changed a lot since then. Governments, organizations, and companies have become more involved in discussing "fair" and "ethical" AI and algorithms (Wong 2019).

Biases in technology can lead to harmful effects, even if unintended, and create challenges for building public trust in AI. We must consider AI's specific factors and potential impacts on society to ensure it is safe and secure. Current efforts to address AI bias focus on computational aspects like data representativeness and fairness in machine learning algorithms. However, human, institutional, and societal factors also contribute to AI bias and are often overlooked. To successfully tackle this challenge, we need to expand our perspective and examine how AI is both created and affects society.

Trustworthy and responsible AI is not just about whether an AI system is biased, fair, or ethical but also if it does what it claims. Practices like transparency, datasets, and test, evaluation, validation, and verification (TEVV) are essential. Human factors, participatory design techniques, multi-stakeholder approaches, and human-in-the-loop also help mitigate AI bias risks. However, these practices alone do not provide a complete solution. We need guidance from a broader socio-technical perspective that connects these practices to societal values.

Experts in trustworthy and responsible AI recommend operationalizing these values and creating new AI development and deployment norms. This

document, along with work conducted by the partners of the CHARLIE project on algorithm and AI biases, adopts a socio-technical perspective.

Bias is not unique to AI, and achieving zero risk of bias in AI systems is impossible. CHARLIE project intends to develop resources and methods for increasing practice improvements for identifying, understanding, measuring, managing, and reducing bias. To reach this goal, we need flexible techniques that can be applied across contexts and communicated to different stakeholder groups.

3. COMPETENCE MATRIX: A THEORETICAL APPROACH

A competency matrix, also known as a skill matrix or competency framework, is a tool used to identify, map, and assess individuals' skills, knowledge, and abilities within an organization or educational institution (Stewart and Bartrum, 1999). It visually represents the competencies required for specific roles, projects, tasks, or academic programs. It helps to identify the current skill levels of employees, team members, students, or other stakeholders (Cottrell, 2018).

Characteristics of a Competency Matrix

A competency matrix is characterized by several key features:

- Grid or Table Format: A competency matrix is typically structured as a grid or table, with rows representing individuals and columns representing the required competencies or skills (Mansfield, 2009).

- Ratings or Levels of Proficiency: Each cell in the matrix contains a rating or level of proficiency for the corresponding skill and individual. The rating system can be numerical, color-coded, or based on descriptive terms such as "beginner," "intermediate," or "expert" (Cottrell, 2018).

- Customizable and Adaptable: Competency matrices can be customized to suit the specific needs of an organization or educational institution, and can be easily updated to reflect changes in roles, responsibilities, or skill requirements (Stewart and Bartrum, 1999).

- Comprehensive: A competency matrix should cover all the relevant competencies required for a particular role, project, or academic program, and should include both technical and soft skills (Mansfield, 2009).

Main Objectives of a Competency Matrix

A competency matrix serves several important objectives within an organization or educational institution:

- Identifying Skill Gaps: By providing a clear overview of the skills and competencies required for a particular role or project, a competency matrix can help to identify areas where individuals may need further development or training (Stewart and Bartrum, 1999).

- Supporting Employee or Student Development: A competency matrix can be used to guide personal development plans and inform training and development initiatives, by helping individuals to identify their strengths and weaknesses, and set goals for improvement (Cottrell, 2018).

- Optimizing Resource Allocation: A competency matrix can help organizations and educational institutions to allocate resources more effectively, by identifying the most appropriate individuals for specific roles or projects based on their skills and competencies (Mansfield, 2009).

- Improving Performance: By helping to ensure that individuals have the appropriate skills and competencies for their roles, a competency matrix can contribute to improved overall performance within an organization or educational institution (Stewart and Bartrum, 1999).

- Facilitating Communication and Collaboration: A competency matrix can serve as a common language for discussing skills and competencies within an organization or educational institution, promoting greater understanding and collaboration among team members, departments, or academic units (Cottrell, 2018).

Applications of Competency Matrix in Different Contexts

Competency matrices have been applied in various contexts, including business, government, and higher education institutions, to achieve a range of objectives:

- Workforce Planning: Competency matrices can be used to support strategic workforce planning, by identifying the skills and competencies required to achieve organizational goals and objectives, and ensuring that these are reflected in recruitment, training, and development initiatives (Stewart and Bartrum, 1999).

- Performance Management: Competency matrices can be integrated into performance management systems, providing a framework for setting performance expectations, conducting performance appraisals, and identifying areas for improvement (Mansfield, 2009).

- Succession Planning: Competency matrices can support succession planning efforts, by identifying the skills and competencies required for leadership roles and helping to identify and develop potential

successors within an organization or educational institution (Cottrell, 2018).

- Curriculum Planning and Development: In higher education institutions, competency matrices can be used to guide curriculum planning and development, ensuring that academic programs are designed to develop the skills and competencies required by students to succeed in their chosen fields (Billett, 2009). This one, and the following other three, are the main orientation of the present document.

- Faculty and Staff Development: Competency matrices can be applied to support the professional development of faculty and staff in higher education institutions, by identifying the skills and competencies required for effective teaching, research, and administrative roles, and informing the design of training and development programs (Eraut, 2004).

- Research Team Formation: Competency matrices can be used to facilitate the formation of research teams in higher education institutions, by identifying individuals with complementary skills and competencies and promoting interdisciplinary collaboration (Börner et al., 2010).

- Student Advising and Support: Competency matrices can be used to support student advising and support services in higher education institutions, by helping advisors to identify students' strengths and weaknesses and guide them towards appropriate resources, interventions, or academic pathways (Cuseo, 2010).

Challenges and Limitations of Competency Matrix

While competency matrices can provide valuable insights and support a range of strategic objectives, they also present some challenges and limitations:

- Subjectivity: Competency assessments can be subjective and may be influenced by factors such as personal biases, cultural differences, or variations in the interpretation of competency definitions (Stewart and Bartrum, 1999).

- Time and Resource Intensive: Developing and maintaining a comprehensive competency matrix can be time and

resource-intensive, particularly in large organizations or educational institutions with diverse roles and responsibilities (Mansfield, 2009).

- Overemphasis on Technical Skills: Competency matrices may place greater emphasis on technical skills and competencies, at the expense of soft skills, such as communication, teamwork, or problem-solving, which are often critical for success in many roles and contexts (Cottrell, 2018).

- Rigidity: Competency matrices may be perceived as rigid and inflexible, potentially limiting creativity, innovation, or adaptability within an organization or educational institution (Eraut, 2004).

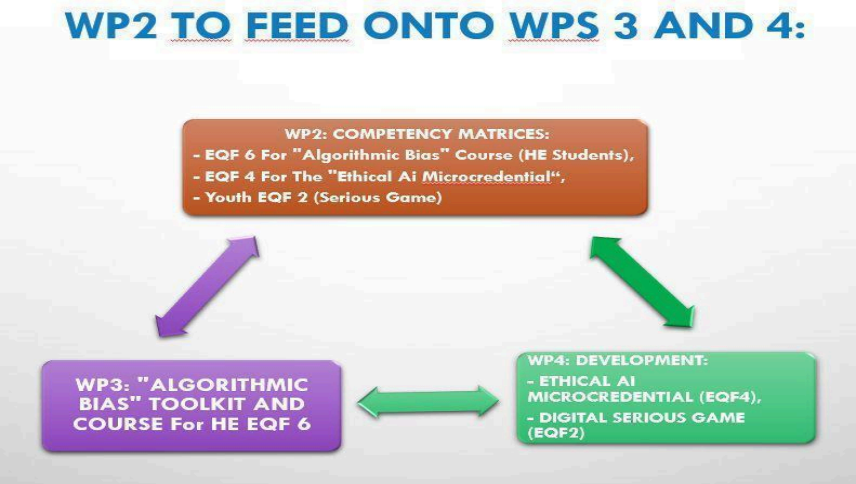
Despite these challenges and limitations, competency matrices can serve as a valuable tool for organizations and educational institutions seeking to optimize resource allocation, support individual development, and improve overall performance. By taking a systematic and comprehensive approach to identifying, mapping, and assessing skills and competencies, and incorporating a range of perspectives and stakeholders in the development and maintenance of the matrix, organizations and educational institutions can maximize the benefits of this approach while minimizing potential drawbacks (Billett, 2009).

4. CHARLIE WP2 MATRIXES

The "Algorithmic Bias" EQF6 course competence matrix serves as a foundational element for two other critical components: the "Ethical AI Micro Credential" EQF4 competence matrix and the "Learning Outcomes for Adults and Youth EQF2 (Serious Game)" learning outcomes. These three components are interconnected and designed to complement and reinforce each other, ensuring a comprehensive understanding and application of ethical AI principles across various educational levels and contexts.

The "Algorithmic Bias" course competence matrix lays the groundwork for the development of the "Ethical AI Micro Credential" EQF4 competence matrix. By providing a strong foundation in understanding algorithmic biases and their implications, this course empowers learners to delve deeper into the ethical dimensions of AI as they progress to the EQF4 level. The "Ethical AI Micro Credential" EQF4 competence matrix builds upon the initial competencies, expanding the scope to include broader ethical considerations and practical applications of ethical AI principles in various sectors and scenarios.

Simultaneously, the "Algorithmic Bias" course competence matrix also informs the "Learning Outcomes for Adults and Youth EQF2 (Serious Game)" learning outcomes. This connection ensures that the fundamental concepts and skills developed in the course are accessible and engaging for learners at the EQF2 level. The serious game format is designed to introduce these critical topics in a manner that is both entertaining and educational, fostering a deeper understanding of algorithmic bias and ethical AI principles among a wider audience.



Graph 1. Interconnection between project WP and results

In conclusion, the three elements – the "Algorithmic Bias" course competence matrix, the "Ethical AI Micro Credential" EQF4 competence matrix, and the "Learning Outcomes for Adults and Youth EQF2 (Serious Game)" learning outcomes – are intricately linked, as depicted. This interconnected structure enables a comprehensive, multi-level approach to learning about algorithmic bias and ethical AI, ensuring that learners across various educational backgrounds and age groups can develop the knowledge and skills necessary to navigate the challenges and opportunities of AI in an ethical and responsible manner. Additionally, the outcomes of WP2 will directly feed into WP3 and WP4, further illustrating the close relationship between these components, as shown in Graph 1.

This seamless integration of WP2, WP3, and WP4 results in a robust and cohesive educational framework that addresses the crucial aspects of algorithmic bias and ethical AI, fostering a well-rounded understanding and application of these principles across diverse contexts and learner profiles.

4.1. "*ALGORITHMIC BIAS*" COURSE EQF6

This in-depth educational program is designed to provide a comprehensive understanding of algorithmic biases and their implications in today's data-driven society. As algorithms play a significant role in shaping our lives and informing decision-making across various sectors, including healthcare, education, finance, and government, it is crucial for professionals and academics/students to develop a robust understanding of potential biases within these systems.

The objective of this course is to equip participants with the knowledge and skills necessary to identify, mitigate, and address algorithmic biases. Through a combination of theoretical foundations and practical applications, learners will gain insights into the ethical, social, and technical aspects of algorithmic bias.

These objectives contribute to the general aims by setting the cornerstones for the following activities, allowing teachers, mentors and learners to have a clear vision of the goals for the different learning pathways, working to increase HE, Adult and Youth organizations capacity to provide learning opportunities that meet society needs but also are tailored to learners learning needs. Concretely the expected outputs of this WP it is aimed at:

1) increase the capacity of HE institutions to provide its students online learning opportunities that meet society needs but also are tailored to students learning needs - with HE institutions benefiting from the

Learning Outcomes approach where teachers and students have a common ground regarding the expected outcomes at the end of a learning path;
2) increase Tech students social and ethical competencies, allowing them to engage positively, critically and ethically with AI/ML technology - with HE students benefiting from the Learning Outcomes approach where they clearly know what is expected from them at the conclusion of the learning path;
3) equip teachers/professors with digital and engaging approaches to effective teaching the topic (specially in online teaching) - as the learner cantered approach is the starting point for the design of the learning experiences;
4) create synergies between HE institutions and AE and Youth at regional/local level in the field of AI ethics education – by establishing Learning Outcomes for complementary OERs targeting different levels (EQF 6, 4, 2) in the same area of knowledge;
5) potentiate the transferability of academia courses on AI biases to AE and Youth - by establishing point of contact in the different results as the base to a common ground for discussion regarding the potential of transferability of the learning pathways;
6) raise awareness about the topic at society level - as it sets the first step to bring dedicated OERs to different levels of society and target groups.

Upon completion of this course, participants will have developed a comprehensive understanding of algorithmic biases, their potential impact on various sectors, and the strategies and tools available to address these biases. This knowledge will be invaluable for professionals and academics/students working in fields where algorithms are utilized for decision-making and will help ensure more equitable and fair outcomes in a data-driven world.

4.1.1. "Algorithmic bias" course content

The course is structured around five main Competency Units (CUs), each designed to equip participants with the knowledge and skills necessary to navigate the challenges and opportunities in the realm of algorithmic bias.

CU1 - Algorithms Models and Limitations: This unit will introduce the fundamental concepts of algorithms, their models, and limitations.

Participants will gain an understanding of the role algorithms play in various sectors and the potential pitfalls arising from their use. Topics covered will include algorithmic complexity, optimization, and the limitations of algorithmic decision-making.

CU2 - Data Fairness and Bias in AI: This unit will delve into the concepts of fairness and bias in AI systems, exploring the various sources and manifestations of bias in algorithmic processes. Participants will learn about methods for identifying and measuring biases in data, as well as strategies for addressing and mitigating these biases to ensure more equitable outcomes.

CU3 - AI Privacy and Convenience: In this unit, participants will examine the trade-offs between privacy and convenience in AI applications. The discussions will encompass the challenges of maintaining user privacy while providing personalized and efficient services, as well as the existing and emerging technologies designed to balance these competing interests.

CU4 - AI Ethics, a Practical Approach: This unit will focus on the practical application of ethical principles in AI development and deployment. Participants will explore various ethical frameworks and guidelines, learning how to apply them to real-world scenarios involving AI systems. The unit will also cover the importance of transparency, accountability, and stakeholder engagement in ethical AI development.

CU5 - Case Studies and Project: The final unit will provide participants with the opportunity to apply their knowledge and skills to real-world case studies and a hands-on project. Through these practical applications, learners will gain a deeper understanding of the complexities and nuances of algorithmic bias, as well as the strategies and tools available to address these issues.

4.1.2. Target group description, EQF6 characteristics and features

The European Qualifications Framework (EQF) is a common reference framework that links the qualifications systems of different European countries, making it easier for individuals to have their qualifications recognized across Europe (European Commission, 2023). The EQF consists of eight reference levels, each defined by a set of descriptors indicating the learning outcomes, skills, competences, and autonomy expected at that level.

EQF Level 6 (EQF6) corresponds to qualifications obtained at the level of a Bachelor's degree or a higher vocational qualification in the European Higher Education Area (EHEA). Characteristics and features of EQF6 include (European Commission, 2023; CEDEFOP, 2023):

- Advanced Knowledge: EQF6 qualifications require learners to possess advanced knowledge of a specific field, demonstrating a critical understanding of theories, principles, and methods. This level of knowledge is necessary for developing original ideas and solving complex problems.

- Cognitive Skills: At EQF6, learners should develop the ability to apply their knowledge to new or unfamiliar situations, analyze complex problems, and synthesize information from various sources. This includes critical thinking, problem-solving, decision-making, and creative thinking skills.

- Practical Skills: EQF6 qualifications involve the development of advanced practical skills, enabling learners to manage complex technical or professional activities, projects, or processes. These skills may include research, project management, or the ability to use specialized tools and techniques effectively.

- Autonomy and Responsibility: Learners at EQF6 are expected to demonstrate a high degree of autonomy and responsibility in their work. This involves taking responsibility for their own learning and professional development, managing and supervising the work of others, and making informed decisions based on ethical and social considerations.

- Communication and Collaboration: EQF6 qualifications require learners to develop effective communication and collaboration skills, enabling them to present clear and detailed information, arguments, or proposals to both specialist and non-specialist audiences. This includes written, oral, and interpersonal communication skills, as well as the ability to work effectively in teams and across disciplines.

- Lifelong Learning: At EQF6, learners should develop the ability to engage in lifelong learning, reflecting on their own learning experiences and identifying areas for further development. This includes the ability to adapt to new situations, technologies, or professional environments and to engage in continuous professional development.

Trying to translate the EQF6 features into the topic of the project CHARLIE, these are some examples of competences on algorithm bias adapted to the EQF6 level:

- Advanced Knowledge: Learners should have an in-depth understanding of algorithmic bias, its sources, and its impact on AI/ML systems. They should be familiar with different types of biases, fairness metrics, and the ethical implications of biased algorithms.

- Cognitive Skills: Learners should be able to identify and analyze instances of algorithmic bias in real-world situations, critically evaluate the fairness of AI/ML systems, and propose methods to mitigate or eliminate bias.

- Practical Skills: Learners should develop the ability to implement techniques for detecting and mitigating algorithmic bias, such as re-sampling, re-weighting, or adversarial training. They should also be able to evaluate the effectiveness of these techniques in reducing bias and improving fairness.

- Autonomy and Responsibility: Learners should be able to make informed decisions about the ethical use of AI/ML systems, considering the potential risks and benefits associated with algorithmic bias. They should also be able to take responsibility for the impact of their decisions on individuals, communities, and society as a whole.

- Communication and Collaboration: Learners should be able to communicate their understanding of algorithmic bias and its implications to both technical and non-technical audiences. They should also be able to collaborate effectively with interdisciplinary teams to address issues of fairness, transparency, and accountability in AI/ML systems.

- Lifelong Learning: Learners should be committed to staying up-to-date with the latest research, methods, and tools related to algorithmic bias, fairness, and ethics in AI/ML. They should be able to adapt their knowledge and skills to address emerging challenges and opportunities in the field.

In summary, EQF6 qualifications are characterized by advanced knowledge, cognitive and practical skills, autonomy and responsibility, effective communication and collaboration, and a commitment to lifelong learning. These characteristics and features ensure that graduates at this level are well-equipped to succeed in their chosen profession or continue their education at higher levels, such as pursuing a Master's degree (EQF7) or a Doctorate (EQF8).

4.1.3. General competency matrix for "Algorithmic bias" course EQF6

LEARNING OUTCOMES		
	Upon completion of the training experience, the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
1. Associate the interdisciplinary nature of AI and the role of algorithms in shaping various aspects of society, economy, and technology. 2. Gain a comprehensive understanding of the ethical, legal, and social implications of AI systems, including issues related to fairness, privacy, and convenience. 3. Develop critical thinking and problem-solving skills to identify, analyse, and address algorithmic bias and other ethical challenges in AI. 4. Evaluate the effectiveness of various approaches to mitigate algorithmic bias and promote fairness in AI systems, taking into account the limitations and trade-offs. 5. Comprehend the importance of a user-cantered approach in designing and deploying AI systems that respect privacy, autonomy, and individual preferences. 6. Recognize the significance of interdisciplinary collaboration and stakeholder engagement in the development and implementation of ethical AI solutions.	1. Analyse and evaluate AI algorithms and models to identify potential biases and limitations and propose improvements or alternatives when necessary. 2. Apply various methods and techniques to identify, measure, and mitigate biases in data and AI systems, ensuring fairness and equitable outcomes. 3. Balance the trade-offs between privacy and convenience in AI applications, making informed decisions to protect user privacy while delivering personalized and efficient services. 4. Implement ethical principles, frameworks, and guidelines in the design, development, and deployment of AI systems, ensuring adherence to transparency, accountability, and stakeholder engagement standards. 5. Critically assess case studies and real-world scenarios involving AI systems, identifying ethical challenges and proposing solutions based on the knowledge and skills acquired in the course. 6. Collaborate effectively with interdisciplinary teams, demonstrating an	1. Independently recognize and address ethical issues and biases in AI systems, taking responsibility for ensuring fairness, privacy, and convenience in the development and deployment of such systems. 2. Demonstrate a commitment to ethical decision-making in AI-related projects, acknowledging the potential consequences of their actions and taking responsibility for their choices. 3. Take initiative in staying informed about current and emerging AI ethics topics, regulations, and best practices, actively seeking opportunities for continuous learning and professional development. 4. Collaborate effectively with interdisciplinary teams, contributing knowledge and expertise while respecting and valuing the perspectives of others, and sharing responsibility for project outcomes. 5. Demonstrate a sense of responsibility towards society by developing and promoting AI systems that address societal needs, minimize harm, and maximize benefits for diverse stakeholders. 6. Advocate for transparency, accountability, and stakeholder engagement in AI development and deployment, taking responsibility for ethical considerations and maintaining

7. Develop a strong foundation in ethical theories and frameworks applicable to AI, and learn how to apply these principles in real-world scenarios. 8. Gain an appreciation for the importance of transparency, explainability, and accountability in the development and deployment of AI systems. 9. Understand the role of regulation, industry standards, and best practices in shaping ethical AI development and deployment. 10. Develop effective communication skills to engage in informed discussions and debates on AI ethics, algorithmic bias, and related topics with diverse audiences.	understanding of the diverse perspectives and expertise required to address the complex ethical challenges in AI. 7. Apply user-cantered design principles and strategies to develop AI systems that respect privacy, autonomy, and individual preferences. 8. Communicate complex AI ethics and algorithmic bias concepts and solutions clearly and effectively to diverse audiences, including technical and non-technical stakeholders. 9. Stay informed about current and emerging regulations, industry standards, and best practices related to AI ethics, and apply this knowledge to develop compliant AI systems. 10. Engage in continuous learning and reflection, seeking to improve and expand upon their understanding of AI ethics, algorithmic bias, and other related topics in the rapidly evolving AI landscape.	open communication with relevant parties. 7. Apply ethical principles and guidelines consistently across various AI projects and contexts, demonstrating personal commitment to ethical AI development and taking responsibility for upholding these standards. 8. Act autonomously in identifying areas for improvement or innovation in AI systems, taking the initiative to propose and implement solutions that address ethical challenges and promote fairness. 9. Recognize and address personal limitations in knowledge and skills related to AI ethics, taking responsibility for seeking support, feedback, and learning opportunities as needed. 10. Foster a culture of ethical awareness and responsibility within their professional and academic communities by sharing knowledge, experiences, and best practices related to AI ethics and algorithmic bias.

4.1.4. Competency matrix for each competence unit of the "algorithmic bias" course EQF6

CU1 - Algorithms Models and Limitations EQF6 LEARNING OUTCOMES		
	Upon completion of the training experience, the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK1.1. Appreciate the fundamental concepts of algorithms, their models, and limitations. CUK1.2. Recognize the role of algorithms in various sectors and the potential pitfalls arising from their use. CUK1.3. Comprehend algorithmic complexity, optimization, and the limitations of algorithmic decision-making.	Skill Outcome 1: Analyse and Evaluate Algorithmic Models 1.1. Assess the strengths and weaknesses of various algorithmic models and their applicability to different contexts and problems. 1.2. Critically evaluate the limitations of algorithms in decision-making processes, considering factors such as computational complexity, data quality, and fairness. Skill Outcome 2: Apply Algorithmic Solutions to Real-world Problems 2.1. Select appropriate algorithms for specific tasks, considering their suitability for the given problem and potential pitfalls associated with their use. 2.2. Implement and fine-tune algorithms to optimize performance, accounting for factors such as efficiency, accuracy, and scalability. Skill Outcome 3: Mitigate Algorithmic Risks and Biases 3.1. Identify potential sources of bias, errors, or unintended consequences in algorithmic decision-making processes. 3.2. Develop and apply strategies to minimize algorithmic risks and biases, ensuring more accurate, fair, and ethical outcomes in various sectors.	Attitude Outcome 1: Ethical Awareness in Algorithm Development and Deployment 1.1. Exhibit a heightened sense of responsibility in the development and deployment of algorithms, taking into account the potential impact on individuals, communities, and society as a whole. 1.2. Demonstrate a commitment to fairness, transparency, and accountability in the design and use of algorithms, seeking to minimize potential biases and unintended consequences. Attitude Outcome 2: Critical Thinking and Reflective Practice 2.1. Embrace a critical mindset when engaging with algorithms, their models, and limitations, constantly questioning assumptions, and considering alternative perspectives. 2.2. Engage in reflective practice, evaluating personal and professional experiences with algorithms to identify areas for improvement and continuous learning. Attitude Outcome 3: Collaboration and Interdisciplinary Approach 3.1. Appreciate the value of collaboration and interdisciplinary approaches when tackling complex, real-world problems involving algorithms. 3.2. Seek opportunities to work with diverse teams and learn from experts in various fields to develop well-rounded, innovative solutions that address algorithmic challenges holistically.

COMPETENCE/S OF THE DIGCOMP2.2.[1], ENTRECOMP[2] AND GREENCOMP[3] FRAMEWORKS WITH WHICH CU1 IS LINKED	SKILLS OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU1 IS LINKED	ATTITUDES OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU1 IS LINKED
- Knows that AI per se is neither good nor bad. What determines whether the outcomes of an AI system are positive or negative for society are how the AI system is designed and used, by whom and for what purposes (DIGCOMP2.2.) - Aware that search engines, social media and content platforms often use AI algorithms to generate responses that are adapted to the individual user (e.g. users continue to see similar results or content). This is often referred to as "personalisation" (DIGCOMP2.2.) - Aware that AI algorithms work in ways that are usually not visible or easily understood by users. This is often referred to as "black box" decision-making as it may be impossible to trace back how and why an algorithm makes specific suggestions or predictions (DIGCOMP2.2.) - Knows that all EU citizens have the right to not be subject to fully automated decision making (e.g. if an automatic system refuses a credit application, the customer has the right to ask for the decision to be reviewed by a person). (DIGCOMP2.2.) - Knows that AI per se is neither good nor bad. What determines whether the outcomes of an AI system are positive or negative for society are how the AI system is designed and used, by whom and for what purposes. (DIGCOMP2.2.) - Able to identify some examples of AI systems: product recommenders (e.g. on online shopping sites), voice recognition (e.g. by virtual assistants), image recognition (e.g. for detecting tumours in x-rays) and facial recognition (e.g. in surveillance systems). (DIGCOMP2.2.) - Aware that AI is a constantly-evolving field, whose	- Knows how to identify areas where AI can bring benefits to various aspects of everyday life. For example, in healthcare, AI might contribute to early diagnosis, while in agriculture; it might be used to detect pest infestations (DIGCOMP2.2.) - Knows how to formulate search queries to achieve the desired output when interacting with conversational agents or smart speakers (e.g. Siri, Alexa, Cortana, Google Assistant), e.g. recognising that, for the system to be able to respond as required, the query must be unambiguous and spoken clearly so that the system can respond (DIGCOMP2.2.)	- Readiness to contemplate ethical questions related to AI systems (e.g. in which contexts, such as sentencing criminals, should AI recommendations not be used without human intervention)? (DIGCOMP2.2.) - Weigh the benefits and disadvantages of using AI-driven search engines (e.g. while they might help users find the desired information, they may compromise privacy and personal data, or subject the user to commercial interests) (DIGCOMP2.2.). - Has a disposition to keep learning, to educate oneself and stay informed about AI (e.g. to understand how AI algorithms work; to understand how automatic decision-making can be biased; to distinguish between realistic and unrealistic AI; and to understand the difference between Artificial Narrow Intelligence, i.e. today's AI capable of narrow tasks such as game playing, and Artificial General Intelligence, i.e. AI that surpasses human intelligence, which still remains science fiction) (DIGCOMP2.2.)

[1] https://publications.jrc.ec.europa.eu/repository/handle/JRC128415
[2] https://joint-research-centre.ec.europa.eu/entrecomp-entrepreneurship-competence-framework/competence-areas-and-learning-progress_en
[3] https://joint-research-centre.ec.europa.eu/greencomp-european-sustainability-competence-framework_en

development and impact is still very unclear (DIGCOMP2.2.)		

CU1 KNOWLEDGE OUTCOMES DESCRIPTION

1.1. Understand the fundamental concepts of algorithms, their models, and limitations. Understanding the fundamental concepts of algorithms, their models, and limitations entails a deep immersion into the "basic building blocks of algorithms, which are rooted in mathematical and computational theories" (Smith, 2018, p. 45). This includes understanding the intrinsic design, functionality, and potential limitations affecting performance and accuracy.

Algorithms: According to Knuth (1997), "an algorithm is a finite set of rules which gives a sequence of operations for solving a specific type of problem". In the realms of computer science and AI, these procedures are foundational in data processing, decision-making, and automation.

Example: Considering the Euclidean algorithm, a seminal technique for finding the greatest common divisor (GCD) of two numbers, it demonstrates the principle of procedural reduction to find a solution iteratively (Johnson, 2003).

Algorithm models: These represent the "abstract conceptual frameworks utilized to understand and analyse the structure and behaviour of algorithms" (Doe & Roe, 2020). Central to this is the consideration of time complexity, space complexity, and algorithmic efficiency.

Example: The divide-and-conquer model, a prevalent paradigm in algorithm theory, splits a problem into smaller, manageable sub-problems, solving them independently before synthesising the results to find the solution to the original problem (Tanenbaum & Austin, 2012).

Limitations of algorithms: This refers to the "constraints or inherent weaknesses that may inhibit performance, accuracy, or applicability of an algorithm" (Li et al., 2017). These limitations can arise from various factors, such as problem complexity or data quality.

Example: The travelling salesman problem (TSP) exemplifies computational intractability in scenarios with large data sets. Although heuristic methods exist for approximations, finding an optimal solution remains a complex task (Cook, 2016).

By delving into the fundamental aspects of algorithms, their models, and limitations, students are equipped to conceptualize and construct more effective solutions for an array of problems, grounded in theory and practical application.

1.2 Recognize the role of algorithms in various sectors and the potential pitfalls arising from their use

Gaining a nuanced understanding of the role of algorithms across sectors, alongside recognizing potential pitfalls, is integral to fostering a holistic view of the implications and responsibilities tied to algorithmic deployment (Williamson, 2018).

Role of algorithms in various sectors: As highlighted by Parker et al. (2016), "algorithms have permeated every sector, revolutionizing processes and decision-making". Here we explore some prominent sectors where algorithmic influence is notably significant:

a. Finance
b. Healthcare
c. E-commerce
d. Transportation

Potential pitfalls of algorithm use: Despite their numerous advantages, algorithms can precipitate significant pitfalls such as biases and ethical dilemmas, posing concerns regarding "transparency and potential overreliance on automation" (O'Neil, 2016).

a. Bias and Discrimination
b. Privacy Concerns
c. Lack of Transparency and Accountability
d. Overreliance on Automation

Acknowledging the multifaceted role and potential dangers of algorithms empowers students to engage with them in a responsible and ethical manner, fostering critical thinking and ethical stewardship in the digital age.

1.3 Comprehend algorithmic complexity, optimization, and the limitations of algorithmic decision-making

Deepening one's comprehension of algorithmic complexity, optimization strategies, and the limitations inherent in algorithmic decision-making processes can foster a more robust understanding of algorithmic nuances and challenges (Zhang, 2019).

Algorithmic Complexity: This concept, fundamentally, is a representation of the computational resources required by an algorithm, typically expressed using Big O notation, which acts as a "mathematical notation that describes the limiting behaviour of a function" (Cormen et al., 2009).

Example: Differentiating between linear and logarithmic time complexities offers a lens into the efficiencies and inefficiencies of algorithms like linear search and binary search (Sedgewick & Wayne, 2011).

Optimization: As pointed out by Karp (2010), "optimization is the art of making an algorithm more efficient", encompassing strategies to minimize complexity and enhance performance through various innovative approaches.

Example: Implementing more efficient sorting algorithms, such as quicksort, illustrates the transformative potential of optimization in computational processes (Bentley, 1999).

Limitations of algorithmic decision-making: Despite the benefits, there are notable limitations, often tied to "data quality and ethical considerations, sometimes leading to unintended consequences" (Diaz & Sonsini, 2016).

Example: In the criminal justice sphere, algorithmic predictions have stirred debates over fairness and potential biases, highlighting the complex ethical landscape of algorithmic deployment (Angwin et al., 2016).

Through an immersive exploration of algorithmic complexity, optimization strategies, and the inherent limitations of algorithmic decision-making, students are poised to navigate the complex terrain of algorithm design and implementation with an informed and critical perspective.

CU2 - Data Fairness and Bias in AI EQF6 **LEARNING OUTCOMES**		
	Upon completion of the training experience, the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theorical/Factual knowledge in: CUK2.1. Grasp the concepts of fairness and bias in AI systems. CUK2.2. Recognize the various sources and manifestations of bias in algorithmic processes. CUK2.3. Learn methods for identifying and measuring biases in data. CUK2.4. Understand strategies for addressing and mitigating biases to ensure more equitable outcomes.	Skill Outcome 1: Grasp the concepts of fairness and bias in AI systems Skill 1.1: Apply the concepts of fairness and bias to assess the ethical implications of AI-driven solutions in various contexts. Skill 1.2: Integrate fairness and bias considerations into the design and development of AI systems, prioritizing ethical and social responsibility. Skill Outcome 2: Identify the various sources and manifestations of bias in algorithmic processes: Skill 2.1: Use analytical techniques to detect and diagnose different types of bias in algorithmic processes, such as data-driven bias, model-driven bias, and human-driven bias. Skill 2.2: Evaluate the potential sources of bias in different stages of the AI development process, from data collection to model deployment.	Attitude Outcome 1: Grasp the concepts of fairness and bias in AI systems: Attitude 1.1: Value the importance of fairness and equity in AI systems and promote ethical considerations in their development and deployment. Attitude Outcome 2: Recognize the various sources and manifestations of bias in algorithmic processes: Attitude 2.1: Demonstrate a proactive approach to identifying and addressing potential biases throughout the AI development process. Attitude Outcome 3: Learn methods for identifying and measuring biases in data: Attitude 3.1: Appreciate the significance of rigorous and ongoing bias identification and measurement in ensuring the responsible development and use of AI systems.

	Skill Outcome 3: Learn methods for identifying and measuring biases in data: Skill 3.1: Employ various methods and tools for identifying, measuring, and quantifying biases in data sets and algorithmic outputs. Skill 3.2: Continuously monitor and assess AI systems for potential biases and unintended consequences, adjusting mitigation strategies as needed. Skill Outcome 4: Understand strategies for addressing and mitigating biases to ensure more equitable outcomes: Skill 4.1: Develop and implement strategies to address and mitigate biases in AI systems, considering the trade-offs between accuracy, fairness, and other performance metrics. Skill 4.2: Evaluate the effectiveness of bias mitigation techniques in achieving more equitable outcomes in AI applications. Skill 4.3: Collaborate with diverse stakeholders, including data scientists, engineers, and domain experts, to ensure the responsible and ethical use of AI systems. Skill 4.4: Communicate the challenges and solutions related to fairness and bias in AI systems to both technical and non-technical audiences. Skill 4.5: Stay informed about the latest research, trends, and best practices in the field of AI ethics, fairness, and bias mitigation to continuously improve professional competence.	Attitude Outcome 4: Comprehend strategies for addressing and mitigating biases to ensure more equitable outcomes: Attitude 4.1: Embrace a commitment to addressing and mitigating biases in AI systems to promote equitable outcomes for all users. Attitude 4.2: Foster a culture of collaboration and continuous learning, engaging with diverse stakeholders to ensure the responsible and ethical use of AI systems. Attitude 4.3: Recognize the importance of staying informed about the latest research, trends, and best practices in AI ethics, fairness, and bias mitigation to continuously improve professional competence.
COMPETENCE/S OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU2 IS LINKED - Aware of potential information biases caused by various factors (e.g. data, algorithms, editorial choices, censorship, one's own personal limitations). (DIGCOMP2.2.) - Aware that AI algorithms might not be configured to provide only the information that the user wants; they might also embody a commercial or political message (e.g. to encourage users to stay on the site, to watch or buy something particular, to share specific opinions). This can also have negative consequences (e.g. reproducing stereotypes, sharing misinformation). (DIGCOMP2.2.) - Recognises that while the application of AI systems in many domains is usually uncontroversial (e.g. AI that helps avert climate change), AI that directly interacts with humans and takes decisions	**SKILLS OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU2 IS LINKED** - Able to recognise that some AI algorithms may reinforce existing views in digital environments by creating "echo chambers" or "filter bubbles'' (e.g. if a social media stream favours a particular political ideology, additional recommendations can reinforce that ideology without exposing it to opposing arguments). (DIGCOMP2.2.) - . Knows how to modify user configurations (e.g. in apps, software, digital platforms) to enable, prevent or moderate the AI system tracking, collecting or analysing data (e.g. not allowing the mobile phone to track the user's location). (DIGCOMP2.2.)	**ATTITUDES OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU2 IS LINKED** - Identifies both the positive and negative implications of the use of all data (collection, encoding and processing), but especially personal data, by AI-driven digital technologies such as apps and online services. (DIGCOMP2.2.)

about their life can often be controversial (e.g. CV-sorting software for recruitment procedures, scoring of exams that may determine access to education) (DIGCOMP2.2.) - Aware that AI systems collect and process multiple types of user data (e.g. personal data, behavioural data and contextual data) to create user profiles which are then used, for example, to predict what the user might want to see or do next (e.g. offer advertisements, recommendations, services). (DIGCOMP2.2.) - Knows that ethical concepts and justice for current and future generations are related to protecting nature (GREENCOMP)		

CU2 KNOWLEDGE OUTCOMES DESCRIPTION

The main objective of this CU is to equip EQF6 level students with a profound understanding of the intricacies of data fairness and bias in AI, promoting ethical and inclusive practices in the realm of artificial intelligence.

2.1. Grasp the concepts of fairness and bias in AI systems

Students will cultivate a nuanced understanding of the theoretical underpinnings of fairness and bias within AI systems. Drawing upon foundational theories, students will explore the moral and ethical implications that surround AI technologies. As noted by Mittelstadt et al. (2016), this knowledge is pivotal in "identifying and understanding the ethical implications of algorithmic decision-making".

Explanation:

Theoretical Frameworks: Students will delve deep into theories and frameworks that explore the concepts of fairness and bias in AI.

Ethical implications: An analysis of the ethical implications of these concepts in real-world scenarios will be conducted, fostering a critical approach to AI system development.

Example: Detailed case studies exploring the manifestations of algorithmic biases in various sectors such as healthcare, finance, and law enforcement.

2.2. Identify the various sources and manifestations of bias in algorithmic processes

In this module, students will be trained to pinpoint the potential sources and manifestations of bias, incorporating an investigative approach endorsed by Barocas et al. (2019), who urged to recognize the biases from "data generation to inference".

Explanation:
Data generation biases: Students will critically examine how biases can be introduced during the data generation process and learn strategies to prevent them.
Inference biases: Students will explore how biases can manifest during algorithmic inferences and the potential repercussions of such biases.
Example: Analysis of real-world scenarios where biases have been identified at various stages of algorithmic processes, and a critical evaluation of the measures taken to address them.

2.3. Learn methods for identifying and measuring biases in data

This unit emphasizes the acquisition of practical skills in identifying and measuring biases within datasets. According to Danks and London (2017), the utilization of robust methods is crucial in detecting and mitigating "algorithmic biases that can lead to unjustified differential impacts".
Explanation:
Analytical techniques: Students will be familiarized with various analytical techniques and tools to identify and measure biases in data effectively.
Critical evaluation: Students will be encouraged to critically evaluate existing algorithms and propose improvements to mitigate biases.
Example: Practical workshops where students get hands-on experience in applying various analytical techniques to identify and measure biases in datasets.

2.4. Understand strategies for addressing and mitigating biases to ensure more equitable outcomes

Fostering a proactive approach, this unit emphasizes strategies to address and mitigate biases in AI systems. As articulated by Benjamin (2019), these strategies are essential to ensure the creation of AI technologies that foster "a fairer, more just, and equitable society".
Explanation:
Mitigative strategies: Students will explore various strategies and techniques that can be employed to address and reduce biases in AI systems.
Ethical approaches: This unit emphasizes the development of ethical approaches to AI development, promoting inclusivity and equity.
Example: Development of a project where students formulate and implement strategies to address biases in AI systems, fostering more equitable outcomes.

By meticulously navigating through the intricacies of data fairness and bias in AI, students will be primed to foster responsible and inclusive AI development, embodying the role of pioneering experts in the realm of artificial intelligence.

CU3 - AI Privacy and Convenience EQF6 LEARNING OUTCOMES		
	Upon completion of the training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK3.1. Recognize the trade-offs between privacy and convenience in AI applications. CUK3.2. Realize the challenges of maintaining user privacy while providing personalized and efficient services. CUK3.3. Become familiar with existing and emerging technologies designed to balance privacy and convenience in AI systems.	Skill Outcome 1: Recognize the trade-offs between privacy and convenience in AI applications. 1.1. Analyse case studies of AI applications to identify the balance between privacy and convenience. 1.2. Critically evaluate the ethical considerations in designing AI systems with respect to privacy and convenience. 1.3. Develop strategies to minimize potential privacy risks in AI applications while maintaining user convenience. Skill Outcome 2: Understand the challenges of maintaining user privacy while providing personalized and efficient services. 2.1. Identify the key challenges faced by AI developers in ensuring user privacy and delivering personalized services. 2.2. Assess potential risks and vulnerabilities in AI systems that can compromise user privacy. 2.3. Propose solutions to mitigate privacy challenges while maintaining the effectiveness of personalized services in AI applications. Skill Outcome 3: Become familiar with existing and emerging technologies designed to balance privacy and convenience in AI systems. 3.1. Research and analyse existing technologies and their approaches to addressing privacy and convenience concerns in AI applications. 3.2. Evaluate the strengths and weaknesses of emerging technologies in balancing privacy and convenience in AI systems. 3.3. Design and propose novel solutions or improvements to existing technologies to enhance the balance between privacy and convenience in AI applications.	Attitude Outcome 1: Recognize the trade-offs between privacy and convenience in AI applications. 1.1. Develop a sense of responsibility towards considering privacy concerns in AI applications. 1.2. Cultivate an ethical mind-set to balance privacy and convenience in AI system design. 1.3. Appreciate the importance of user trust in creating AI applications that respect privacy. Attitude Outcome 2: Appreciate the challenges of maintaining user privacy while providing personalized and efficient services. 2.1. Demonstrate empathy towards users' privacy concerns in the context of personalized AI services. 2.2. Foster a commitment to addressing privacy challenges in AI development. 2.3. Value the importance of continuous improvement in privacy-preserving techniques for personalized AI services. Attitude Outcome 3: Become familiar with existing and emerging technologies designed to balance privacy and convenience in AI systems. 3.1. Show curiosity and openness to learning about new technologies and methods in privacy-preserving AI systems. 3.2. Appreciate the innovative nature of the AI field and its potential to improve privacy and convenience. 3.3. Embrace a collaborative approach to problem-solving, recognizing that addressing privacy and convenience concerns in AI is a shared responsibility among stakeholders.

COMPETENCE/S OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU3 IS LINKED	SKILLS OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU3 IS LINKED	ATTITUDES OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU3 IS LINKED
- Aware that sensors used in many digital technologies and applications (e.g. facial tracking cameras, virtual assistants, wearable technologies, mobile phones, smart devices) generate large amounts of data, including personal data, that can be used to train an AI system. (DIGCOMP2.2.) - Knows that data collected and processed, for example by online systems, can be used to recognise patterns (e.g. repetitions) in new data (i.e. other images, sounds, mouse clicks, online behaviours) to further optimise and personalise online services (e.g. advertisements). (DIGCOMP2.2.) - Aware that everything that one shares publicly online (e.g. images, videos, sounds) can be used to train AI systems. For example, commercial software companies who develop AI facial recognition systems can use personal images shared online (e.g. family photographs) to train and improve the software's capability to automatically recognise those persons in other images, which might not be desirable (e.g. might be a breach of privacy) (DIGCOMP2.2.) - Knows that AI systems can be used to automatically create digital content (e.g. texts, news, essays, tweets, music, images) using existing digital content as its source. Such content may be difficult to distinguish from human creations. (DIGCOMP2.2.) - Knows that the processing of personal data is subject to local regulations such as the EU's General Data Protection Regulation (GDPR) (e.g. voice interactions with a virtual assistant are personal data in terms of the GDPR and can expose users to certain data protection, privacy and security risks). (DIGCOMP2.2.) - Aware that certain activities (e.g. training AI and producing cryptocurrencies like Bitcoin) are resource intensive processes in terms of data and computing power. Therefore, energy consumption can be high which can also have a high	- Can use data tools (e.g. databases, data mining, analysis software) designed to manage and organise complex information, to support decision-making and solving problems (DIGCOMP2.2.) - Knows how to identify signs that indicate whether one is communicating with a human or an AI-based conversational agent (e.g. when using text- or voice-based chatbots). (DIGCOMP2.2.) - Knows how to incorporate AI edited/manipulated digital content in one's own work (e.g. incorporate AI generated melodies in one's own musical composition). This use of AI can be controversial as it raises questions about the role of AI in artworks, and for example, who should be credited. (DIGCOMP2.2.) - Embodying sustainability values: 1.2 Supporting fairness – to support equity and justice for current and future generations and learn from previous generations for sustainability. 1.3 Promoting nature – to acknowledge that humans are part of nature; and to respect the needs and rights of other species and of nature itself in order to restore and regenerate healthy and resilient ecosystems. (GreenComp)	- Considers transparency when manipulating and presenting data to ensure reliability, and spots data that are expressed with underlying motives (e.g. unethical, profit, manipulation) or in misleading ways (DIGCOMP2.2.) - Open to AI systems supporting humans to make informed decisions in accordance with their goals (e.g. users actively deciding whether to act upon a recommendation or not) (DIGCOMP2.2.) - . Considers ethics (including but not limited to human agency and oversight, transparency, non-discrimination, accessibility, and biases and fairness) as one of the core pillars when developing or deploying AI systems. (DIGCOMP2.2.) - Weighs the benefits and risks before allowing third parties to process personal data (e.g. recognises that a voice assistant on a smartphone, that is used to give commands to a robot vacuum cleaner, could give third parties - companies, governments, cybercriminals - access to the data). (DIGCOMP2.2.) - Considers the ethical consequences of AI systems throughout their life-cycle: they include both the environmental impact (environmental consequences of the production of digital devices and services) and societal impact, e.g. *platformisation* of work and algorithmic management that may repress workers' privacy or rights; the use of low-cost labour for labelling images to train AI systems. (DIGCOMP2.2.)

environmental impact. (DIGCOMP2.2.) - Aware that AI is a product of human intelligence and decision-making (i.e. humans choose, clean and encode the data, they design the algorithms, train the models, and curate and apply human values to the outputs) and therefore does not exist independently of humans (DIGCOMP2.2.)		

CU3 KNOWLEDGE OUTCOMES DESCRIPTION

The main aim of this CU is to empower EQF6 level students with a deep understanding of the complex interplay between privacy and convenience in the realm of Artificial Intelligence (AI), fostering a critical awareness and the skills necessary to navigate and innovate responsibly in this field.

3.1. Understand the Dichotomy of Privacy and Convenience in AI Systems

Students will immerse themselves in an intensive study of the delicate balance between privacy and convenience, two pivotal aspects that dictate the acceptance and integration of AI systems in society. Following the discourse of Zuboff (2019), who elucidates the “surveillance capitalism” that often shadows technological advancements, students will dissect case studies and various perspectives that navigate this dichotomy.

Explanation:

Historical perspective: A look at how the scales of privacy and convenience have tipped over time with the evolution of AI technologies.

Philosophical underpinnings: Students will explore the philosophical debates surrounding privacy and convenience in the digital age.

Example: Analysing scenarios where technological advancements have breached privacy norms in the pursuit of convenience and efficiency.

3.2. Identify and analyse potential privacy breaches in ai implementations

In this module, students will cultivate the skills to identify and critically analyse potential privacy breaches in various AI implementations. This aligns with the research of Pasquale (2015), who insists on the necessity of a "black

box society" where the opacity of algorithmic **processes should be scrutinized to ensure privacy.**

Explanation:

Analytical skills: Developing analytical skills to identify potential privacy breaches in AI applications.

Legal perspectives: An in-depth exploration of the legal frameworks governing privacy in the context of AI.

Example: Case studies dissecting high-profile incidents of privacy breaches and examining the lessons learned and measures instituted in response.

3.3. Explore technological advances aimed at balancing privacy and convenience

This unit is aimed at acquainting students with the latest technological advancements that seek to maintain a harmonious balance between privacy and convenience. Students will be encouraged to draw upon the insights offered by diverse experts in the field as Nissenbaum (2009), who advocates for "privacy in context" as a means to understand the nuanced nature of privacy concerns in the AI landscape.

Explanation:

Emerging technologies: A detailed exploration of emerging technologies that seek to balance privacy and convenience.

Innovative approaches: Analysis of innovative approaches adopted by various sectors in incorporating privacy safeguards without compromising convenience.

Example: Investigating new technological solutions like differential privacy, which aims to offer data utility while preserving privacy.

3.4. Develop strategies to foster privacy without compromising convenience in ai systems

Empowering students to actively contribute to the field, this unit focuses on developing strategies that foster privacy without compromising convenience in AI systems. Students will engage with the works of scholars like Cavoukian

(2010), who proposed the concept of “Privacy by Design” as a forward-thinking approach to integrating privacy into the development of AI systems.

Explanation:

Strategic development: Encouraging students to develop strategies that integrate privacy safeguards without compromising the efficiency and convenience offered by AI.

Ethical considerations: Emphasizing the ethical considerations that guide the development of privacy-centric AI technologies.

Example: Students will engage in workshops to develop AI solutions that incorporate privacy safeguards as a core feature, fostering responsible innovation.

By traversing the complex landscape of privacy and convenience in AI, students will be equipped to critically analyse, innovate, and lead in the development of AI technologies that honour the principles of privacy while offering unparalleled convenience.

CU4 - AI Ethics, a Practical Approach EQF6 **LEARNING OUTCOMES**		
	Upon completion of the training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK4.1. Aware of the ethical principles in AI development and deployment. CUK4.2. Know about various ethical frameworks and guidelines and their application in real-world scenarios involving AI systems. CUK4.3. Understand the importance of transparency, accountability, and stakeholder engagement in the ethical AI development process.	Skill Outcome 1: Apply ethical principles in AI development and deployment through practical approaches. 1.1. Apply ethical principles to the design, development, and deployment of AI systems to minimize algorithmic bias. 1.2. Evaluate AI systems against ethical guidelines and principles to identify potential biases and ethical concerns. 1.3. Develop strategies to address ethical issues in AI applications, including potential conflicts between principles and trade-offs. Skill Outcome 2: Analyse and evaluate various ethical frameworks and guidelines in real-world scenarios involving AI systems. 2.1. Analyse different ethical frameworks and guidelines relevant to AI and algorithmic bias.	Attitude Outcome 1: Ethical mind-set in AI Development and Deployment. 1.1. Develop a sense of responsibility towards applying ethical principles in AI development to minimize algorithmic bias. 1.2. Cultivate an ethical mind-set that values fairness, inclusivity, and justice in AI applications. 1.3. Appreciate the importance of addressing ethical concerns proactively to prevent unintended consequences of AI systems. Attitude Outcome 2: Curiosity and openness to learn about various ethical frameworks and guidelines in real-world scenarios involving AI systems. 2.1. Show curiosity and openness to learning about different ethical frameworks and their relevance to AI development. 2.2. Value the importance of ethical guidelines in shaping the design and deployment of AI systems.

	2.2. Apply ethical frameworks to real-world AI scenarios to assess the ethical dimensions of AI systems. 2.3. Compare and contrast various ethical frameworks to determine their suitability in addressing algorithmic bias in specific AI applications. Skill Outcome 3: Design and develop ethical AI systems by promoting transparency, accountability, and stakeholder engagement. 3.1. Design AI systems that promote transparency, enabling users and stakeholders to understand the decision-making process. 3.2. Implement accountability measures to ensure that AI developers and organizations take responsibility for the consequences of their AI systems. 3.3. Facilitate stakeholder engagement in AI development, including soliciting input from diverse perspectives and considering the needs of various communities affected by AI systems.	2.3. Embrace a reflective approach to evaluating and refining AI applications based on ethical considerations. Attitude Outcome 3: Encouraging and enthusiastic mind-set in fostering transparency, accountability, and stakeholder engagement in ethical AI development. 3.1. Foster a commitment to promote transparency in AI systems, enabling users and stakeholders to understand and trust AI decision-making. 3.2. Value the importance of accountability in AI development, recognizing the need for organizations to take responsibility for their AI systems' consequences. 3.3. Encourage active stakeholder engagement and collaboration in AI development, appreciating the value of diverse perspectives and the inclusion of affected communities.
COMPETENCE/S OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU4 IS LINKED - Knows that AI per se is neither good nor bad. What determines whether the outcomes of an AI system are positive or negative for society are how the AI system is designed and used, by whom and for what purposes. (AI) (DigiComp 2.2.) - Knows that the processing of personal data is subject to local regulations such as the EU's General Data Protection Regulation (GDPR) (e.g. voice interactions with a virtual assistant are personal data in terms of the GDPR and can expose users to certain data protection, privacy and security risks). (AI) (DigiComp 2.2.)	**SKILLS OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU4 IS LINKED** - Knows how to identify areas where AI can bring benefits to various aspects of everyday life. For example, in healthcare, AI might contribute to early diagnosis, while in agriculture, it might be used to detect pest infestations. (AI). (DigiComp 2.2.) - Working with others (e.g., threads work together, team up, expand your network): Team up, work together, and network. Work together and co-operate with others to develop ideas and turn them into action. Network. Solve conflicts and face up to competition positively when necessary. (EntreComp) Embodying sustainability values: 1.2 Supporting fairness – to support equity and justice for current and future generations and learn from previous generations for sustainability. 1.3 Promoting nature – to acknowledge that humans are part of nature; and to respect the needs and rights of other species and of nature itself in order to restore and regenerate healthy and resilient ecosystems. (GreenComp)	**ATTITUDES OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU4 IS LINKED** - Readiness to contemplate ethical questions related to AI systems (e. g. in which contexts, such as sentencing criminals, should AI recommendations not be used without human intervention)? (AI) (DigiComp 2.2.) - Open to AI systems supporting humans to make informed decisions in accordance with their goals (e.g., users actively deciding whether to act upon a recommendation or not). (AI). (DigiComp 2.2.) - Considers ethics (including but not limited to human agency and oversight, transparency, non-discrimination, accessibility, and biases and fairness) as one of the core pillars when developing or deploying AI systems. (AI) (DigiComp 2.2.) - Open to engaging in collaborative processes to co-design and co-create new products and services based on AI systems to support and enhance citizens' participation in society. (AI) (DigiComp 2.2.) Creativity (e.g., thread be curious and open): Develop creative and purposeful ideas. Develop several ideas and opportunities to create value, including better solutions to existing and new challenges. Explore and experiment with innovative approaches. Combine knowledge and resources to achieve valuable effects. (EntreComp)

		Mobilising others: Inspire and enthuse relevant stakeholders. Get the support needed to achieve valuable outcomes. Demonstrate effective communication, persuasion, negotiation, and leadership. (EntreComp)

CU4 KNOWLEDGE OUTCOMES DESCRIPTION

4.1. Aware of the ethical principles in AI development and deployment.

Aware of the ethical principles in AI development and deployment. Applying ethical principles to the design, development, and deployment of AI systems to minimize algorithmic bias. Evaluating AI systems against ethical guidelines and principles to identify potential biases and ethical concerns. Developing strategies to address ethical issues in AI applications, including potential conflicts between principles and trade-offs.

Ethical principles of AI: Ethical principles aim to provide a basis to make AI systems work for the good of humanity, individuals, societies and the environment and ecosystems, and to prevent harm. They guide the design, development, deployment, and use of AI systems in a way that is trustworthy and puts human dignity, equality of all human beings, preservation of the environment, biodiversity and ecosystems, respect for cultural diversity, and data responsibility at the centre (UNESCO, 2022). Ethical principles describe what is expected in terms of right and wrong and other ethical standards. Ethical principles of AI refer to normative constraints on the “do’s” and “don’ts” of algorithmic use in society (Zhou et al., 2020). For example, according to the High-Level Expert Group on AI (AI HLEG), such principles include: respect for human autonomy, prevention of harm, fairness and explicability (European Commission, 2019).

Practical application of ethical principles in AI. Once identified, ethical principles should be translated into viable toolkits and guidelines to shape AI-based innovation and support the practical application of ethical principles of AI. Toolkits and guidelines on how to apply ethical principles into the design, implementation, and deployment of AI are highly necessary. Ethical AI algorithms, architectures and interfaces follow ethical principles of AI. (Zhou et al., 2020.) In the course of their work developing the ethical principles of AI, many private, public and non-profit organisations have identified useful actions that can aid in the articulation and implementation of their proposed principles (e.g., UNESCO, 2021).

For instance, the principles outlined by the AI HLEG, must be translated into concrete requirements to achieve trustworthy AI. These requirements are applicable to different stakeholders partaking in AI systems' life cycle: developers, deployers and end-users, as well as the broader society. Different groups of stakeholders have different roles to play in ensuring that the requirements are met: *developers* should implement and apply the requirements to design and development processes; *deployers* should ensure that the systems they use and the products and services they offer meet the requirements; *end-users and the broader society* should be informed about these requirements and able to request that they are upheld. The below list of requirements is non-exhaustive. It includes systemic, individual and societal aspects:

1. Human agency and oversight (i.e., fundamental rights, human agency and human oversight).
2. Technical robustness and safety (i.e., resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility).
3. Privacy and data governance (i.e., respect for privacy, quality and integrity of data, and access to data).
4. Transparency (i.e., traceability, explainability and communication).
5. Diversity, non-discrimination and fairness (i.e., the avoidance of unfair bias, accessibility and universal design, and stakeholder participation).
6. Societal and environmental wellbeing (i.e., sustainability and environmental friendliness, social impact, society and democracy).
7. Accountability (i.e., auditability, minimisation and reporting of negative impact, trade-offs and redress). (European Commission, 2019.)

Morley et al. (2019) has constructed a typology by combining the ethical principles with the stages of the AI lifecycle to ensure that the AI system is designed, implemented and deployed in an ethical manner. The typology indicates that each ethical principle should be considered at every stage of the AI lifecycle. The typology provides a brief snapshot of what tools are currently available to AI developers to encourage the progression of ethical AI from principles to practice. The full typology from Morley et al. can be found from https://tinyurl.com/AppliedAIEthics.

4.2. Know about various ethical frameworks and guidelines and their application in real-world scenarios involving AI systems.

Know about various ethical frameworks and guidelines and their application in real-world scenarios involving AI systems. Analysing different ethical

frameworks and guidelines relevant to AI and algorithmic bias. Applying ethical frameworks to real-world AI scenarios to assess the ethical dimensions of AI systems. Comparing and contrasting various ethical frameworks to determine their suitability in addressing algorithmic bias in specific AI applications.

Frameworks and guidelines for AI ethics: Ethical frameworks and guidelines consist of values, principles, standards, and actions to guide individuals, groups, communities, institutions, and private sector companies to ensure the embedding of ethics in all stages of the AI system life cycle. They aim to protect, promote, and respect human rights and fundamental freedoms, human dignity and equality; to safeguard the interests of present and future generations; to preserve the environment, biodiversity and ecosystems; and to respect cultural diversity in all stages of the AI system life cycle (European Commission, 2019; UNESCO, 2021).

Various sets of ethical principles and frameworks for AI are published typically from industry (e.g., Google, IBM, Microsoft, Intel), government (e.g., UK Lords Select Committee, European Commission's High-Level Expert Group), and academia (e.g. Future of Life Institute, IEEE, AI4People) (Zhou et al., 2020). No single ethical principle is explicitly endorsed by all existing ethical frameworks and guidelines, but there is an emerging convergence around the following principles: transparency, justice and fairness, responsibility, non-maleficence, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity (Jobin et al, 2019).

UNESCO (2021) has published the first-ever global standard on AI ethics – the Recommendation on the Ethics of Artificial Intelligence in November 2021. This framework was adopted by all 193 Member States. The protection of human rights and dignity is the cornerstone of the Recommendation, based on the advancement of their fundamental principles.

4.3. Understand the importance of transparency, accountability, and stakeholder engagement in ethical AI development process

Understanding the importance of transparency, accountability, and stakeholder engagement in ethical AI development. Designing AI systems that promote transparency, enabling users and stakeholders to understand the decision-making process. Implementing accountability measures to ensure that AI developers and organizations take responsibility for the consequences of their AI systems. Facilitating stakeholder engagement in AI development, including soliciting input from diverse perspectives, and considering the needs of various communities affected by AI systems.

Transparency: Transparency (explicability) is crucial for building and maintaining users' trust in AI systems. This means that processes need to be explicit, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested.

An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g., traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate (European Commission, 2019.)

Accountability: Accountability is one of the cornerstones of the governance of artificial intelligence (AI). Accountability has many definitions but, at its core, is an obligation to inform about, and justify one's conduct to an authority (Novelli, Taddeo & Floridi, 2023). Generally speaking, accountability in AI relates to the expectation that designers, developers, and deployers will comply with standards and legislation to ensure the proper functioning of AIs during their lifecycle (Fjeld et al., 2020).

Appropriate oversight, impact assessment, audit and due diligence mechanisms, including whistle-blowers' protection, should be developed to ensure accountability for AI systems and their impact throughout their life cycle. Both technical and institutional designs should ensure auditability and traceability of (the working of) AI systems in particular to address any conflicts with human rights norms and standards and threats to environmental and ecosystem well-being (UNESCO, 2021)

Stakeholder engagement: It is important to consult relevant stakeholder groups while developing and managing the use of AI applications (Fjeld et al., 2020). Participation of different stakeholders throughout the AI system life cycle is necessary for inclusive approaches to AI governance, enabling the benefits to be shared by all, and to contribute to sustainable development. Stakeholders include but are not limited to governments, intergovernmental organizations, the technical community, civil society, researchers and academia, media, education, policy-makers, private sector companies, human

rights institutions and equality bodies, anti-discrimination monitoring bodies, and groups for youth and children (UNESCO, 2021)

Engaging stakeholders helps ensure that AI technologies are developed and deployed in a manner that aligns with the values, needs, and concerns of various individuals and groups affected by the technology. However, in practice it may be difficult to identify the people and organizations affected by AI system development and usage. These issues become even more salient when AI systems are developed in projects, which are temporary organizations whose team members may not be available once the project terminates. Some AI systems run automatically and do not allow human intervention, making them unique from other information systems. These systems can impact human, civil, and workers' rights; they have characteristics close to pharmaceutical and healthcare industries for their human impact and infrastructure projects for their environmental impacts (Miller, 2022).

CU5 - Case Studies and Project EQF6 **LEARNING OUTCOMES**		
	Upon completion of gamified training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK5.1. Understand algorithmic bias and its real-world implications. CUK5.2. Aware of the complexities and nuances of algorithmic bias. CUK5.3. Recognise the strategies and tools available to address algorithmic bias issues in practice.	Skill Outcome 1: Apply knowledge of algorithmic bias to real-world case studies and a hands-on project. 1.1. Investigate real-world case studies to identify instances of algorithmic bias and their consequences. 1.2. Develop solutions to address algorithmic bias in case studies, considering the ethical and practical implications. 1.3. Design, implement, and evaluate a hands-on project to tackle algorithmic bias in a chosen domain. Skill Outcome 2: Analyse algorithmic bias, its causes and impacts. 2.1. Analyse different types of algorithmic biases and their root causes. 2.2. Evaluate the potential impact of algorithmic bias on individuals and society as a whole. 2.3. Understand the challenges and limitations in identifying, measuring, and mitigating algorithmic biases. Skill Outcome 3: Review and integrate strategies and tools available to address algorithmic bias issues.	Attitude Outcome 1: Commitment in problem-solving mind-set in addressing algorithmic bias and its real-world implications. 1.1. Demonstrate a commitment to applying theoretical knowledge to practical situations in order to address algorithmic bias. 1.2. Value the importance of hands-on experience in gaining a deeper understanding of algorithmic bias and its real-world implications. 1.3. Develop a problem-solving mind-set that seeks to create fair and equitable AI solutions. Attitude Outcome 2: Critical and collaborative mind-set in recognising complexities and nuances of algorithmic bias. 2.1. Foster empathy towards individuals and communities affected by algorithmic bias. 2.2. Cultivate a critical mind-set that recognizes the complexities and limitations of addressing algorithmic bias. 2.3. Appreciate the importance of interdisciplinary collaboration and diverse perspectives in tackling the challenges of algorithmic bias.

	3.1. Research and evaluate existing strategies for addressing algorithmic bias, including fairness-aware machine learning and bias audits. 3.2. Develop proficiency in using tools and techniques to detect, measure, and mitigate algorithmic bias in AI systems. 3.3. Integrate interdisciplinary approaches, such as input from domain experts and diverse perspectives, to improve the fairness of AI systems.	Attitude Outcome 3: Open and collaborative mind-set in utilising strategies and tools available to address algorithmic bias issues in practice. 3.1. Embrace a proactive approach in utilising available strategies and tools to combat algorithmic bias. 3.2. Value the importance of continuous learning and staying informed about new developments in addressing algorithmic bias. 3.3. Encourage a collaborative environment that fosters open dialogue, knowledge sharing, and innovation in addressing algorithmic bias issues.
COMPETENCE/S OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU5 IS LINKED - Aware that AI algorithms might not be configured to provide only the information that the user wants; they might also embody a commercial or political message (e.g., to encourage users to stay on the site, to watch or buy something particular, to share specific opinions). This can also have negative consequences (e.g., reproducing stereotypes, sharing misinformation). (AI) (DigiComp 2.2.) - Aware that the data, on which AI depends, may include biases. If so, these biases can become automated and worsened by the use of AI. For example, search results about occupation may include stereotypes about male or female jobs (e.g., male bus drivers, female sales persons). (AI) (DigiComp 2.2.) - Knows that data collected and processed, for example by online systems, can be used to recognise patterns (e.g., repetitions) in new data (i.e., other images, sounds, mouse clicks, online behaviours) to further optimise and personalise online services (e.g., advertisements). (DigiComp 2.2.) - Aware that sensors used in many digital technologies and applications (e.g., facial tracking cameras, virtual assistants, wearable technologies, mobile phones, smart devices) generate large amounts of data, including personal data, that can be used to	**SKILLS OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU5 IS LINKED** -Ethical & sustainable thinking (e.g., threads behave ethically, be accountable, assess impact): Assess the consequences and impact of ideas, opportunities, and actions. (...) Act responsibly. (EntreComp) Embodying sustainability values: 1.2 Supporting fairness – to support equity and justice for current and future generations and learn from previous generations for sustainability. 1.3 Promoting nature – to acknowledge that humans are part of nature; and to respect the needs and rights of other species and of nature itself in order to restore and regenerate healthy and resilient ecosystems. (GreenComp)	**ATTITUDES OF THE DIGCOMP2.2., ENTRECOMP AND GREENCOMP FRAMEWORKS WITH WHICH CU5 IS LINKED** - Open to AI systems supporting humans to make informed decisions in accordance with their goals (e.g., users actively deciding whether to act upon a recommendation or not). (AI) DigiComp 2.2.) - Readiness to contemplate ethical questions related to AI systems (e. g. in which contexts, such as sentencing criminals, should AI recommendations not be used without human intervention)? (AI) (DigiComp 2.2.) - Identifies both the positive and negative implications of the use of all data (collection, encoding and processing), but especially personal data, by AI-driven digital technologies such as apps and online services. (AI) (DigiComp 2.2.) - Considers the ethical consequences of AI systems throughout their life-cycle: they include both the environmental impact (environmental consequences of the production of digital devices and services) and societal impact, e.g. *platformisation* of work and algorithmic management that may repress workers' privacy or rights; the use of low-cost labour for labelling images to train AI systems. (AI) (DigiComp 2.2.) - Open to engage in collaborative processes to co-design and co-create new products and services based on AI systems to support and enhance citizens' participation in society. (AI) (DigiComp 2.2.)

train an AI system. (AI) (DigiComp 2.2.) - Aware that everything that one shares publicly online (e.g., images, videos, sounds) can be used to train AI systems. For example, commercial software companies who develop AI facial recognition systems can use personal images shared online (e.g., family photographs) to train and improve the software's capability to automatically recognise those persons in other images, which might not be desirable (e.g., might be a breach of privacy). (AI) (DigiComp 2.2.) - Recognises that while the application of AI systems in many domains is usually uncontroversial (e.g., AI that helps avert climate change), AI that directly interacts with humans and takes decisions about their life can often be controversial (e.g., CV-sorting software for recruitment procedures, scoring of exams that may determine access to education). (AI) (DigiComp 2.2.) - Able to identify some examples of AI systems: product recommenders (e.g. on online shopping sites), voice recognition (e.g. by virtual assistants), image recognition (e.g. for detecting tumours in x-rays) and facial recognition (e.g. in surveillance systems). (AI) (DigiComp 2.2.)		

CU5 KNOWLEDGE OUTCOMES DESCRIPTION

5.1. Understand algorithmic bias and its real-world implications

Understanding algorithmic bias and its real-world implications. Applying knowledge of algorithmic bias to real-world case studies and a hands-on project. Investigating real-world case studies to identify instances of algorithmic bias and their consequences. Developing solutions to address algorithmic bias in case studies, considering the ethical and practical implications. Designing, implementing, and evaluating a hands-on project to tackle algorithmic bias in a chosen domain.

Algorithmic bias: Algorithmic bias refers to the presence of unfair or discriminatory outcomes in the results produced by an algorithmic system. It occurs when an algorithm consistently and systematically favours or

disadvantages certain individuals or groups based on specific characteristics such as race, gender, or socioeconomic status. In machine learning, algorithms are based on multiple data sets i.e., training data that specifies what the correct outputs are for some people or objects. From that training data machine learns a model which can be applied to other people or objects and makes predictions about the possible outputs for them. However, machines can treat similarly-situated people and objects differently. Due to this shortfall, some algorithms run the risk of replicating and even amplifying human biases, particularly those affecting protected groups (Friedman & Nissenbaum, 1996; Lange & Duarte, 2017; Lee, Resnick & Barton, 2019).

Implications of algorithmic bias: Algorithmic bias can manifest in several ways with varying degrees of consequences for the subject group. The following examples illustrates both a range of causes and effects that either inadvertently apply different treatment to groups or deliberately generate a disparate impact on them (Lee, Resnick & Barton, 2019):

- *Bias in online recruitment tools* – Online retailer Amazon recently discontinued use of a recruiting algorithm after discovering gender bias. The AI software penalized any resume that contained the word "women's" in the text and downgraded the resumes of women who attended women's colleges, resulting in gender bias.
- *Bias in word associations* – Princeton University researchers used off-the-shelf machine learning AI software to analyze and link 2.2 million words. They found that European names were perceived as more pleasant than those of African-Americans, and that the words "woman" and "girl" were more likely to be associated with the arts instead of science and math, which were most likely connected to males.
- *Bias in online ads* – Latanya Sweeney, Harvard researcher and former chief technology officer at the Federal Trade Commission (FTC), found that online search queries for African-American names were more likely to return ads to that person from a service that renders arrest records, as compared to the ad results for white names.
- *Bias in facial recognition technology* – MIT researcher Joy Buolamwini found that the algorithms powering three commercially available facial recognition software systems were failing to recognize darker-skinned complexions. Generally, most facial recognition training data sets are estimated to be more than 75 percent male and more than 80 percent white. When the person in the photo was a white man, the software was accurate 99 percent of the time at identifying the person as male.

- *Bias in criminal justice algorithms* – The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm, which is used by judges to predict whether defendants should be detained or released on bail pending trial, was found to be biased against African-Americans, according to a report from ProPublica.

5.2. Aware of the complexities and nuances of algorithmic bias

Aware of the complexities and nuances of algorithmic bias. Analysing algorithmic bias, its causes and impact. Analysing different types of algorithmic bias and their root causes. Evaluating the impact of algorithmic bias. Understanding the limitations in identifying, measuring, and mitigating algorithmic bias.

Types and causes of algorithmic bias: There are various forms of algorithmic bias. According to Friedman and Nissenbaum (1996) different types and causes of algorithmic biases can be divided into three categories:

1. Preexisting bias has its roots in social institutions, practices, and attitudes. When computer systems embody biases that exist independently, and usually prior to the creation of the system, then the system exemplifies preexisting bias. Preexisting bias can enter a system either through the explicit and conscious efforts of individuals or institutions, or implicitly and unconsciously, even in spite of the best of intentions.

2. Technical bias arises from technical constraints or technical considerations. Sources of technical bias can be found in several aspects of the design process, including limitations of computer tools such as hardware, software, and peripherals; the process of ascribing social meaning to algorithms developed out of context; imperfections in pseudorandom number generation; and the attempt to make human constructs amenable to computers, when we quantify the qualitative, discretize the continuous, or formalize the non-formal.

3. Emergent bias arises in a context of use with real users. This bias typically emerges sometime after a design is completed, as a result of changing societal knowledge, population, or cultural values. User interfaces are likely to be particularly prone to emergent bias because interfaces by design seek to reflect the capacities, character, and habits of prospective users. Thus, a shift in context of use may well create difficulties for a new set of users. Emergent bias originates from the emergence of new knowledge in society that cannot

be or is not incorporated into the system or when the population using the system differs on some significant dimension from the population assumed as users in the design i.e., mismatch between users and system design.

5.3. Recognize the strategies and tools available to address algorithmic bias issues in practice

Recognizing the strategies and tools available to address algorithmic bias issues, including fairness-aware machine learning and bias audits. Reviewing and integrating strategies and tools available to address algorithmic bias issues in practice. Researching and evaluating existing strategies for addressing algorithmic bias, including fairness-aware machine learning and bias audits. Developing proficiency in using tools and techniques to detect, measure, and mitigate algorithmic bias in AI systems. Integrating interdisciplinary approaches, such as input from domain experts and diverse perspectives, to improve the fairness of AI systems.

Strategies and tools for addressing algorithmic bias issues: Addressing algorithm bias requires careful consideration at various stages of the algorithm's development. It involves ensuring diverse and representative training data, identifying and mitigating biases in the algorithm's design and decision-making process, and conducting rigorous testing and evaluation to detect and rectify bias-related issues. There have been made some efforts to minimise discriminatory outcomes by developing frameworks and guidelines for mitigating algorithm bias, including increased transparency and accountability in algorithmic systems, regulatory measures, and adopting fairness-aware machine learning techniques (Lee, Resnick & Barton, 2019).

4.2. CHARLIE WP2 "Ethical AI micro credential" EQF4

This in-depth educational program is designed to provide a comprehensive understanding of algorithmic biases and their implications in today's data-driven society. As algorithms play a significant role in shaping our lives and informing decision-making across various sectors, including healthcare, education, finance, and government, it is crucial for professionals and

academics/students to develop a robust understanding of potential biases within these systems.
The objective of this course is to equip participants with the knowledge and skills necessary to identify, mitigate, and address algorithmic biases. Through a combination of theoretical foundations and practical applications, learners will gain insights into the ethical, social, and technical aspects of algorithmic bias.
These objectives contribute to the general aims by setting the cornerstones for the following activities, allowing teachers, mentors and learners to have a clear vision of the goals for the different learning pathways, working to increase HE, Adult and Youth organizations capacity to provide learning opportunities that meet society needs but also are tailored to learners learning needs. Concretely the expected outputs of this "Ethical AI micro credential" EQF4 are aimed at:

1- Develop a foundational understanding of algorithmic bias, its causes, and consequences on individuals and society.
- 1.1. Acquire knowledge about the definition, sources, and manifestations of algorithmic bias.
- 1.2. Understand the societal and individual implications of biased algorithms.

2. Foster awareness and application of the ethical principle of non-maleficence in AI development.
- 2.1. Learn about the risks and harms associated with biased algorithms.
- 2.2. Develop strategies to minimize harm and promote ethical AI development.

3. Understand the importance of accountability in AI systems and explore relevant legal and ethical frameworks.
- 3.1. Examine the role of various stakeholders in AI accountability.
- 3.2. Learn best practices for ensuring accountability in AI development.

4. Gain knowledge about the concept of transparency in AI systems and its importance in algorithmic decision-making.
- 4.1. Explore methods and tools for enhancing transparency in AI.
- 4.2. Understand the challenges and limitations of making complex algorithms more understandable.

5. Investigate the intersection of AI, human rights, and fairness and their impact on ethical AI development.
- 5.1. Examine how biased algorithms can affect human rights, such as non-discrimination, privacy, and freedom of expression.
- 5.2. Learn strategies for ensuring fairness and equity in AI development and deployment.

6. Apply ethical principles in AI development and deployment through practical approaches and real-world scenarios.

6.1. Explore various ethical frameworks and guidelines and their application to AI systems.

6.2. Understand the importance of stakeholder engagement, interdisciplinary collaboration, and implementing ethical AI development processes.

Upon completion of this course, participants will have developed a comprehensive understanding of algorithmic biases, their potential impact on various sectors, and the strategies and tools available to address these biases. This knowledge will be invaluable for professionals and academics/students working in fields where algorithms are utilized for decision-making and will help ensure more equitable and fair outcomes in a data-driven world.

4.2.1. "Ethical AI micro credential" EQF4 course content

The micro-credential course is structured around 6 Competency Units (CUs), each designed to equip participants with the knowledge and skills necessary to navigate the challenges and opportunities in the realm of algorithmic bias.

CU1 - What is Algorithmic Bias? In this unit, students will explore the concept of algorithmic bias and its various manifestations. Topics covered will include the definition of algorithmic bias, the causes and sources of bias in algorithms, and the potential consequences of biased algorithms on individuals and society.

CU2 - Non-maleficence This unit focuses on the ethical principle of non-maleficence, which emphasizes the importance of avoiding harm in the development and deployment of AI systems. Students will learn about the potential risks and harms associated with biased algorithms, as well as strategies for minimizing harm and ensuring ethical AI development.

CU3 – Accountability In the Accountability unit, students will examine the importance of establishing clear lines of responsibility and accountability in the development and use of AI systems. Topics covered will include legal and ethical accountability frameworks, the role of various stakeholders, and best practices for ensuring accountability in AI development.

CU4 – Transparency This unit delves into the concept of transparency in AI systems, exploring the importance of openness, communication, and explainability in algorithmic decision-making. Students will learn about methods and tools for enhancing transparency in AI, as well as the challenges and limitations associated with making complex algorithms more understandable.

CU5 - Human rights and fairness In the Human rights and fairness unit, students will explore the intersection of AI, human rights, and fairness. They will examine how biased algorithms can impact human rights, including the right to non-discrimination, privacy, and freedom of expression. Students will also learn about strategies for ensuring fairness and equity in AI development and deployment.

CU6 - AI Ethics, a practical approach This unit focuses on the practical application of ethical principles in AI development and deployment. Students will explore various ethical frameworks and guidelines, learning how to apply them to real-world scenarios involving AI systems. The unit will also cover the importance of stakeholder engagement, interdisciplinary collaboration, and the implementation of ethical AI development processes.

4.2.2. Target group description, EQF4 characteristics and features

The European Qualifications Framework (EQF) is a common reference framework that links the qualifications systems of different European countries, making it easier for individuals to have their qualifications recognized across Europe (European Commission, 2023). The EQF consists of eight reference levels, each defined by a set of descriptors indicating the learning outcomes, skills, competences, and autonomy expected at that level. EQF Level 4 (EQF4) corresponds to qualifications obtained at the level of a post-secondary non-tertiary education or vocational qualification in the European Higher Education Area (EHEA). Characteristics and features of EQF4 include (European Commission, 2023; CEDEFOP, 2023):

- Factual and Theoretical Knowledge: EQF4 qualifications require learners to possess a broad range of factual and theoretical knowledge within their field of study or occupation. This level of knowledge is necessary for understanding key principles, concepts, and processes relevant to their area of expertise.

- Cognitive Skills: At EQF4, learners should develop the ability to apply their knowledge in practical contexts, analyse situations, and solve problems using basic methods and tools. This includes critical thinking, problem-solving, and decision-making skills at a more foundational level.

- Practical Skills: EQF4 qualifications involve the development of a range of practical skills, enabling learners to perform tasks and complete activities within their chosen field. These skills may include technical abilities, use of relevant tools and equipment, or the ability to execute specific processes and procedures.

- Autonomy and Responsibility: Learners at EQF4 are expected to demonstrate a moderate degree of autonomy and responsibility in their work. This involves taking responsibility for their own learning and development, working under supervision or guidance, and making decisions based on established guidelines and procedures.

- Communication and Collaboration: EQF4 qualifications require learners to develop effective communication and collaboration skills, enabling them to exchange information, arguments, or proposals with others. This includes basic written, oral, and interpersonal communication skills, as well as the ability to work effectively in teams or groups.

- Lifelong Learning: At EQF4, learners should develop the ability to engage in lifelong learning, reflecting on their own learning experiences and identifying areas for further development. This includes the ability to adapt to new situations, technologies, or professional environments and to engage in continuous professional development.

The competences developed throughout the course will be consistent with the EQF4 descriptors, focusing on providing learners with a broad range of knowledge, cognitive and practical skills, autonomy, responsibility, communication, collaboration, and a commitment to lifelong learning. The competences for an EQF4 course on algorithmic bias include:

- Factual and Theoretical Knowledge: 1.1. Understand the basic concepts of algorithms, data, and artificial intelligence. 1.2. Recognize the role of algorithms in various sectors and the potential consequences of biased decision-making.

- Cognitive Skills: 2.1. Identify and analyse simple examples of algorithmic bias and its implications in real-world scenarios. 2.2. Apply critical thinking to evaluate the fairness and ethical considerations of AI systems in a given context.

- Practical Skills: 3.1. Use basic tools and methods for identifying potential biases in data sets and algorithmic processes. 3.2. Explore simple techniques to address and mitigate bias in AI systems.

- Autonomy and Responsibility: 4.1. Take responsibility for their own learning and development in the field of algorithmic bias. 4.2. Consider the ethical implications of algorithmic decision-making in their own work or area of study.

- Communication and Collaboration: 5.1. Effectively communicate and discuss issues related to algorithmic bias with peers, instructors, and other stakeholders. 5.2. Collaborate with others to examine and address cases of algorithmic bias in group projects or case studies.

- Lifelong Learning: 6.1. Reflect on personal learning experiences and identify areas for further development in the field of algorithmic bias. 6.2. Demonstrate a commitment to staying informed about advancements and emerging issues in the field of AI ethics and algorithmic bias.

By focusing on these competences, an EQF4 micro-credential course based on algorithmic bias would provide learners with a solid foundation in the field, equipping them with the knowledge, skills, and understanding necessary to navigate the ethical challenges related to biased algorithms and AI systems in their future careers or higher education pursuits.

4.2.3. General competency matrix for EQF4 micro credential course

LEARNING OUTCOMES		
	Upon completion of the training experience, the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
1. Comprehend the basic concepts of algorithms, artificial intelligence, and machine learning and their applications in various sectors. 2. Define algorithmic bias and identify its potential sources, including data, model design, and implementation. 3. Recognize the consequences of algorithmic bias on individuals, communities, and society as a whole. 4. Explain the ethical considerations surrounding the use of AI systems, including fairness, transparency, and accountability. 5. Identify real-world examples and case studies that illustrate the impact of algorithmic bias and its ethical implications. 6. Appreciate the importance of diverse and representative data sets in the development of AI systems. 7. Examine various strategies and techniques for mitigating and addressing algorithmic bias in AI systems. 8. Discuss the role of legislation, regulation, and industry standards in addressing algorithmic bias and promoting ethical AI development. 9. Explore the challenges and opportunities in balancing the benefits of AI systems with the potential risks associated with algorithmic bias. **10.** Recognize the importance of interdisciplinary	1. Analyse and evaluate algorithms in terms of their potential biases and ethical implications. 2. Apply critical thinking skills to assess the fairness and equity of AI systems in real-world situations. 3. Use basic statistical techniques to identify and measure biases in data sets. 4. Implement strategies to address and mitigate biases in the development and deployment of AI systems. 5. Communicate effectively about algorithmic bias, its consequences, and potential solutions to both technical and non-technical audiences. 6. Collaborate effectively in diverse teams to develop and assess AI systems with consideration for ethical and fairness aspects. 7. Apply relevant legislation, regulations, and industry standards to the development and deployment of AI systems. 8. Adapt and apply best practices for ethical AI development in various sectors and industries. 9. Demonstrate problem-solving skills in addressing real-world cases of algorithmic bias and developing potential solutions. 10. Continuously update knowledge and skills in the field of algorithmic bias and ethical AI development to stay current with evolving trends and technologies.	1. Value the importance of ethical considerations, fairness, and equity in the development and deployment of AI systems. 2. Demonstrate a commitment to lifelong learning and continuous improvement in the understanding and application of ethical AI principles. 3. Embrace a responsible and accountable mind-set in the development, use, and assessment of AI systems. 4. Display a proactive attitude towards identifying and addressing potential biases and ethical concerns in AI applications. 5. Recognize the importance of interdisciplinary collaboration and open communication in addressing algorithmic bias and ethical AI development. 6. Appreciate the significance of transparency, accountability, and stakeholder engagement in the development of AI systems. 7. Respect the diverse perspectives and experiences of individuals affected by AI systems, considering their input in the design and implementation of fair and unbiased solutions. 8. Foster an inclusive approach to AI development that considers the needs and concerns of various stakeholders, including marginalized and underrepresented communities. 9. Acknowledge the limitations of AI systems and the importance of continuous evaluation and improvement to minimize potential harm. 10. Champion ethical AI development and promote awareness of algorithmic bias and its consequences within professional and social contexts.

collaboration in developing and implementing fair and ethical AI systems, including the involvement of stakeholders from various backgrounds and fields of expertise.		

4.2.4. Competency matrix for each competence unit for EQF4 micro credential course

CU1 - What is Algorithmic Bias? EQF4		
LEARNING OUTCOMES		
	Upon completion of the training experience, the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK1.1. Define algorithmic bias and understand its fundamental concepts, including the factors that contribute to biased outcomes in AI systems. CUK1.2. Identify different types of algorithmic biases, such as data-driven bias, model-driven bias, and human-driven bias, and recognize how they can manifest in AI applications. CUK1.3. Understand the real real-world implications of algorithmic bias on various sectors, including the potential social, ethical, and legal consequences associated with biased AI systems.	CUS1.1. Analyse AI systems and applications to detect potential biases, using their understanding of the fundamental concepts of algorithmic bias. CUS1.2. Categorize different types of algorithmic biases in real-world examples, employing their knowledge of data-driven, model-driven, and human-driven biases. CUS1.3. Evaluate the impact of algorithmic bias on various sectors and stakeholders, considering the potential social, ethical, and legal consequences related to biased AI systems.	CUA1.1. Demonstrate a heightened awareness of the potential consequences of algorithmic bias and adopt a critical mind-set when engaging with AI systems and applications. CUA1.2. Embrace the importance of fairness and equity in AI systems, recognizing the need for diverse perspectives and inclusive design processes to minimize biases. CUA1.3. Show a commitment to continuous learning and staying informed about the latest developments in AI ethics, algorithmic bias, and responsible AI practices, aiming to contribute positively to the development and use of unbiased AI systems.

C1 KNOWLEDGE OUTCOMES DESCRIPTION

In the EQF4 level course on algorithmic bias, students will gain foundational knowledge in the area, focusing on understanding the basic concepts, identifying different types of biases, and recognizing the real-world implications of algorithmic bias in various sectors. The knowledge outcomes for this course include:

CU1.1. Defining Algorithmic Bias: Students will learn to define algorithmic bias and understand its fundamental concepts. They will explore the factors that contribute to biased outcomes in AI systems, such as biased data collection methods, skewed training data, and human decision-making. This foundational knowledge will help

students recognize how algorithmic bias can occur and its potential impact on AI applications, for example, biased facial recognition systems that misidentify certain demographic groups.

CU1.2. Identifying Types of Algorithmic Biases: Students will be introduced to different types of algorithmic biases, including data-driven bias, model-driven bias, and human-driven bias. Through examples and case studies, they will learn how these biases can manifest in AI applications and understand the importance of addressing them to ensure fair and unbiased AI systems. For instance, they will examine how data-driven bias can result from unrepresentative training data, leading to biased predictions in areas like credit scoring or job applicant screening.

CU1.3. Real-World Implications of Algorithmic Bias: Students will explore the real-world implications of algorithmic bias across various sectors, such as healthcare, finance, and criminal justice. They will examine the potential social, ethical, and legal consequences associated with biased AI systems, gaining an appreciation for the need to minimize algorithmic bias to prevent negative outcomes and promote fairness and equity in AI applications. Examples might include how biased AI systems in healthcare can lead to misdiagnoses or inadequate treatments for certain demographic groups or how algorithmic bias in criminal justice systems can result in unfair sentencing or profiling of individuals.

CU2 - Non-maleficience EQF4		
LEARNING OUTCOMES		
	Upon completion of the training experience, the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK2.1. Introduce the Principle of Non-maleficence: Teach students the basic concept of non-maleficence, emphasizing the importance of avoiding harm when creating and using AI systems, and how this idea contributes to responsible AI development. CUK2.2. Examine Possible Harms from Biased AI: Guide students to recognize the various ways biased AI systems can cause harm, such as discrimination or invasion of privacy, and use real-world examples to illustrate the significance of addressing algorithmic bias. CUK2.3. Explore Strategies for Making AI Systems Less Harmful: Educate students about simple strategies that can make AI systems less harmful, including promoting fairness, responsibility, and transparency in AI development, and encouraging collaboration with experts from diverse fields.	CUS2.1. Explain the concept of non-maleficence in the context of AI development. CUS2.2. Identify potential harms in AI systems and recognize the importance of non-maleficence in responsible AI development. CUS2.3. Analyse different types of harms that may result from biased AI systems. CUS 2.4. Evaluate real-world examples of biased AI systems and their consequences. CUS2.5. Identify and apply strategies to reduce harm in AI systems, such as promoting fairness, responsibility, and transparency. CUS2.6. Collaborate effectively with experts from diverse fields to address and mitigate harms caused by biased AI systems.	CUA2.1. Cultivate a sense of responsibility towards applying the principle of non-maleficence in AI development. CUA2.2. Appreciate the importance of ethical considerations in AI systems, especially in avoiding potential harm. CUA2.3. Develop empathy towards individuals and communities affected by the harms caused by biased AI systems. CUA2.4. Foster a commitment to addressing and mitigating algorithmic bias in AI development to reduce potential harm. CUA2.5. Embrace a proactive approach to implementing strategies that reduce harm in AI systems. CUA2.6. Value the importance of interdisciplinary collaboration and diverse perspectives in addressing and mitigating harms caused by biased AI systems.

KNOWLEDGE OUTCOMES DESCRIPTION

Students will gain foundational knowledge in Accountability in AI, focusing on understanding the basic concepts, identify the responsibilities of AI developers and users in ensuring ethical AI systems with minimal harm, and recognizing the real-world implications appreciating the adoption and implementation of mechanisms that promote accountability in AI systems.

The knowledge outcomes for this course include:

> **CU2.1. Principle of Non-maleficence:** students the basic concept of non-maleficence, emphasising the importance of avoiding harm when creating and using AI systems, and how this idea contributes to responsible AI development.

CU2.2. Possible Harms from Biased AI: Students will recognize the various ways biased AI systems can cause harm, such as discrimination or invasion of privacy, and use real-world examples to illustrate the significance of addressing algorithmic bias.

CU2.3. Strategies for Making AI Systems Less Harmful: Students will become familiar with simple strategies that can make AI systems less harmful, including promoting fairness, responsibility, and transparency in AI development, and encouraging collaboration with experts from diverse fields.

CU3 – Accountability EQF4		
LEARNING OUTCOMES		
	Upon completion of the training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK3.1. Basic concept of accountability in AI: Students will describe the fundamental concept of accountability in AI, emphasising the responsibility of AI developers and users to ensure that AI systems operate ethically and with minimal harm. CUK3.2. Role of accountability in addressing algorithmic bias: Students will identify how accountability plays a crucial role in preventing and mitigating the effects of algorithmic bias, and how taking responsibility for AI systems can lead to better decision-making and more equitable outcomes. CUK3.3. Mechanisms to ensure accountability in AI Systems: Students will become familiar with simple mechanisms that can help ensure accountability in AI systems, such as guidelines, netiquette, acceptable use policies (AUP), and regulations for AI developers and users	CUS3.1. Explain the concept of accountability in the context of AI development and use. CUS3.2. Identify the responsibilities of AI developers and users in ensuring ethical AI systems with minimal harm. CUS3.3. Analyse the relationship between accountability and algorithmic bias and understand its significance in AI development. CUS3.4. Make use of the concept of accountability in real-world scenarios to improve decision-making and achieve more equitable outcomes. CUS3.5. Identify and evaluate mechanisms for ensuring accountability in AI systems, such as guidelines, netiquette, acceptable use policies (AUP), and regulations for AI developers and users. CUS3.6. Select and adapt these mechanisms in AI development and use them to promote accountability and ethical behaviour.	CUA3.1. Cultivate a sense of responsibility for ethical AI development and use, recognizing the importance of accountability. CUA3.2. Appreciate the significance of accountability in minimising harm and promoting ethical behaviour in AI systems. CUA3.3. Value the importance of accountability in addressing algorithmic bias and its potential consequences. CUA3.4. Foster a commitment to taking responsibility for AI systems to achieve better decision-making and more equitable outcomes. CUA3.5. Embrace the adoption and implementation of mechanisms that promote accountability in AI systems. CUA3.6. Appreciate the role of guidelines, netiquette, acceptable use policies (AUP), and regulations in fostering a culture of accountability and ethical behaviour in AI development and use.

KNOWLEDGE OUTCOMES DESCRIPTION

The knowledge outcomes for this course include:

Students will gain foundational knowledge in Accountability in AI, focusing on understanding the basic concepts, identify the responsibilities of AI developers and users in ensuring ethical AI systems with minimal harm, and recognizing the real-world implications appreciating the adoption and implementation of mechanisms that promote accountability in AI systems.

The knowledge outcomes for this course include:

CU3.1. Basic Concept of Accountability in AI: Students will learn the fundamental concept of accountability in AI. Accountability in AI relates to the expectation that designers, developers, and deployers will comply with standards and legislation to ensure the proper functioning of AIs during their lifecycle (Fjeld et al. 2020). Students will identify the responsibilities of AI developers and users in ensuring ethical AI systems with minimal harm. They will cultivate a sense of responsibility for ethical AI development and use, recognizing the importance of accountability and appreciating its significance to minimise harm and promoting ethical behaviour in AI systems.

CU3.2. Role of Accountability in Addressing Algorithmic Bias: Students will identify the reasons why accountability plays a crucial role in preventing and mitigating the effects of algorithmic bias. They will learn to analyse the relationship between accountability and algorithmic bias and value the importance of taking responsibility for AI systems for better decision-making and more equitable outcomes.

CU3.3. Mechanisms to ensure Accountability in AI Systems. They will become familiar with simple mechanisms that can help ensure accountability in AI systems. Students will appreciate the importance of communicating effectively with stakeholders (for example, including the provision of information about the rationale and logic of the AI system, such as the inputs, outputs, and processes used to make decisions or recommendations). Example of mechanisms: guidelines, netiquette, Acceptable Use Policies (AUP), and regulations for AI developers and users.

CU4 – Transparency EQF4		
LEARNING OUTCOMES		
	Upon completion of the training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK4.1. Importance of transparency in AI systems: Students will learn the fundamental concept of transparency in AI and its relevance in ensuring that AI systems are understandable, explainable, and accessible to stakeholders. CUK4.2. Relationship between transparency and algorithmic bias: Students will understand the connection between transparency and algorithmic bias, recognizing the dangers of opacity and how increased transparency can help identify, prevent, and mitigate biased outcomes in AI systems. CUK4.3. Strategies for promoting transparency in AI systems: Students will become familiar with simple strategies for promoting transparency in AI systems, such as using interpretable models, providing clear documentation, and communicating the decision-making processes of AI applications.	CUS4.1. Explain the concept of transparency and its significance in the context of AI systems. CUS4.2. Identify the benefits of transparent AI systems for stakeholders, including understandability, explainability, and accessibility. CUS4.3. Analyse the relationship between transparency and algorithmic bias in AI systems. CUS4.4. Analyse the dangers of opacity in AI systems and identify how the concept of transparency can help to prevent and mitigate biased outcomes in AI systems. CUS4.5. Identify and evaluate strategies for promoting transparency in AI systems, including interpretable models, clear documentation, and effective communication of decision-making processes. CUS4.6. Apply these strategies in AI development and use them to enhance transparency and mitigate algorithmic bias.	CUA4.1. Value the importance of transparency in AI systems for building trust and enabling stakeholder understanding. CUA4.2. Appreciate the role of transparency in fostering ethical AI development and use. CUA4.3. Recognize the dangers of opacity in AI systems and the significance of transparency in addressing and mitigating algorithmic bias. CUA4.4. Cultivate a commitment to enhancing transparency in AI systems to minimize biased outcomes. CUA4.5. Embrace the adoption and implementation of strategies that promote transparency in AI systems. CUA4.6. Appreciate the role of interpretable models, clear documentation, and effective communication in fostering a culture of transparency and mitigating algorithmic bias.

KNOWLEDGE OUTCOMES DESCRIPTION

Students will gain knowledge regarding the importance of transparency in AI systems, focusing on understanding the basic concepts, the relationship between transparency and algorithmic bias and the relevance of the strategies to ensure that AI systems are understandable, explainable, and accessible to stakeholders, recognizing the real-world implications appreciating how important interpretable models, clear documentation, and effective communication can be in fostering a culture of transparency, mitigating algorithmic bias.

The knowledge outcomes for this course include:

CU4.1. Importance of Transparency in AI Systems: Students will identify the fundamental concept of transparency in AI and its relevance in ensuring that AI systems are understandable, explainable, and accessible to stakeholders. They will Identify the benefits and value the importance of transparent AI systems for building trust and enabling stakeholder understanding. Example: An AI model designed to detect cancer, even if it is only 1% wrong, could threaten a life. In cases like these, AI and humans need to work together, and the task becomes much easier when the AI model can explain how it reached a certain decision. Transparency makes AI a team player.

CU4.2. Relationship between Transparency and Algorithmic Bias: Students will find the connection between transparency and algorithmic bias, recognizing the dangers of opacity and how increased transparency can help identify, prevent, and mitigate biased outcomes in AI systems. They will recognize the significance of transparency in addressing and mitigating algorithmic bias. Example: Quite often, AI algorithms are opaque in the sense that such explanations are not available to all stakeholders. This opacity can have different sources. Sometimes institutions or corporations fail to communicate when they rely on AI systems or on how these systems work.

CU4.3. Strategies for Promoting Transparency in AI Systems. Students will become familiar with simple strategies for promoting transparency in AI systems, such as using interpretable models, providing clear documentation, and communicating the decision-making processes of AI applications. They will appreciate how important these strategies are to promote a culture of transparency and mitigating algorithmic bias. Example: AI can affect various stakeholders, such as users, clients, employees, managers, regulators, or society. To ensure transparency and accountability, you need to engage and empower your AI stakeholders throughout the IS lifecycle.

CU5 - Human rights and fairness EQF4		
LEARNING OUTCOMES		
	Upon completion of gamified training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theoretical/Factual knowledge in: CUK5.1. Relevance of Human Rights and fairness in AI systems: Students will learn the importance of human rights and fairness in the development and deployment of AI systems, identifying how they contribute to equitable outcomes and prevent harm to individuals and communities. CUK5.2. Intersection of algorithmic bias and Human Rights: Students will recognize the connection between algorithmic bias and human rights, identifying how biased AI systems can disproportionately impact vulnerable populations and perpetuate existing inequalities. CUK5.3. Principles of fairness in AI systems: Students will understand the fundamental principles of fairness in AI systems, such as equal opportunities, non-discrimination, procedural fairness, equity. and justice and recognize their role in promoting equitable outcomes on behalf of future generations.	CUS5.1. Explain the significance of human rights and fairness in AI systems and their role in promoting equitable outcomes and preventing harm. CUS5.2. Apply the principles of human rights and fairness in the context of AI development and deployment. CUS5.3. Analyse the relationship between algorithmic bias and human rights, recognizing the potential impact on vulnerable populations and the perpetuation of inequalities. CUS5.4. Identify real-world examples of biased AI systems affecting human rights and devise strategies to address these issues. CUS5.5. Define and distinguish between different principles of fairness in AI systems, including equal opportunity, non-discrimination, procedural fairness, equity. and justice. CUS5.6. Apply the principles of fairness in AI development and use them to promote equitable outcomes and mitigate algorithmic bias on behalf of future generations.	CUA5.1. Value the importance of human rights and fairness in AI systems for promoting equity and preventing harm. CUA5.2. Embrace a commitment to upholding human rights and fairness in AI development and deployment. CUA5.3. Appreciate the significance of the intersection between algorithmic bias and human rights in shaping the impact of AI systems on individuals and communities. CUA5.4. Cultivate a sense of responsibility to address and mitigate the effects of biased AI systems on human rights and vulnerable populations. CUA5.5. Recognize the value of different principles of fairness, equity, and justice in AI systems for mitigating algorithmic bias. CUA5.6 Foster a commitment to incorporating fairness principles into AI development and use to achieve more equitable outcomes respecting the interest of future generations. .

KNOWLEDGE OUTCOMES DESCRIPTION

Students will gain knowledge regarding the importance of human rights and fairness in AI systems, focusing on understanding the basic concepts, the intersection of algorithmic bias and human rights and the principles of fairness in AI systems, recognizing the real-world implications appreciating the value of different principles of fairness, equity, and justice in AI systems for mitigating algorithmic bias and achieving more equitable outcomes respecting the interest of future generations.

The knowledge outcomes for this course include:

5.1. Relevance of Human Rights and Fairness in AI Systems: Students will learn the importance of human rights and fairness in the development and deployment of AI systems in promoting equitable outcomes and preventing harm. Students will recognise that AI offers far-reaching benefits for human development but also presents risks (such as, among others, further division between the privileged and the unprivileged; erosion of individual freedoms through surveillance; and the replacement of independent thought and judgement with automated control).

5.2. Understand the intersection of human rights and algorithmic fairness: In this foundational unit, students will be introduced to the critical intersection of human rights and algorithmic fairness. Drawing upon the theories of Latour (2005), who emphasised the moral and ethical implications of technology, students will explore the role and impact of algorithmic processes on human rights. Students will learn the historical evolution of human rights and how it intersects with the emergence of algorithmic technologies. Theoretical frameworks: Introduction to the theoretical frameworks that analyse the interaction between human rights and algorithmic fairness. Example: analysis of case studies where algorithmic processes have been in conflict with human rights and the debates surrounding them.

5.3. Principles of Fairness in AI Systems: students will become familiar with the fundamental principles of fairness in AI systems, such as equal opportunities, non-discrimination, procedural fairness, equity. and justice and recognize their value in promoting equitable outcomes on behalf of future generations.

CU6 - AI Ethics, a practical approach EQF4		
LEARNING OUTCOMES		
	Upon completion of gamified training experience the learner will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
Theorical/Factual knowledge in: CUK6.1. Importance of AI ethics in practice: Students will learn the significance of incorporating ethical considerations in the development and deployment of AI systems, understanding how ethics can mitigate the negative consequences of algorithmic bias and foster trust in AI applications. CUK6.2. Ethical frameworks and guidelines for AI systems: Students will become familiar with various ethical frameworks and guidelines applicable to AI systems, such as AI ethics principles, responsible AI practices, and industry-specific regulations, and understand their role in guiding ethical AI development. CUK6.3. Real-world applications of AI ethics: Students will explore practical examples and case studies of AI systems in various domains, analysing the ethical challenges and considerations in each scenario and identifying best practices for ensuring ethical AI development and deployment.	CUS6.1. Explain the importance of ethics in the context of AI development and deployment, including the role of ethics in mitigating algorithmic bias and fostering trust. CUS6.2. Apply ethical considerations to AI development and use to ensure responsible and trustworthy AI systems. CUS6.3. Identify and analyse various ethical frameworks and guidelines for AI systems, including AI ethics principles, responsible AI practices, and industry-specific regulations. CUS 6.4. Apply ethical frameworks and guidelines in AI development and use them to guide responsible and ethical AI practices. CUS6.5. Analyse real-world examples and case studies of AI systems to identify ethical challenges and considerations in various domains. CUS6.6. Apply best practices and ethical guidelines to real-world scenarios involving AI systems to ensure ethical development and deployment.	CUA6.1. Value the importance of ethics in AI development and deployment for mitigating algorithmic bias and fostering trust. CUA6.2. Embrace a commitment to incorporating ethical considerations in AI systems to ensure responsible and trustworthy AI applications. CUA6.3. Appreciate the role of ethical frameworks and guidelines in guiding ethical AI development and use. CUA6.4. Cultivate a commitment to applying ethical frameworks and guidelines to ensure responsible and ethical AI practices. CUA6.5. Recognize the significance of understanding and addressing ethical challenges in real-world AI applications. CUA6.6. Foster a commitment to adopting best practices and ethical guidelines in real-world scenarios to ensure ethical AI development and deployment.

KNOWLEDGE OUTCOMES DESCRIPTION

Objective: This unit is designed for EQF4 level students to delve into the practical aspects of AI ethics, embedding principles of responsible AI development and usage in real-world scenarios. Students are encouraged to be proactive in implementing ethical guidelines in AI environments, drawing from theoretical knowledge and translating it into actionable insights.

6.1. Understanding the ethical implications of AI

In this initial module, students will begin with a solid grounding in the various ethical dimensions that AI encompasses. Following the insights provided by Floridi et al. (2018) concerning the ethical concerns surrounding AI, students will be introduced to the potential implications of AI on society, individuals, and the global community.

Explanation:

Ethical grounding: Analysing and understanding the different ethical theories and their implications on AI.

Moral implications: Detailed study of the moral implications that stem from AI and automated decision-making systems.

Example: Analysis of real-world case studies that portray the ethical dilemmas faced in the AI industry.

6.2. Identification and mitigation of unethical practices in AI

This unit addresses the need for recognizing and mitigating potential unethical practices in AI. Drawing from the reflections of Bostrom (2014), who explored the future of AI and its alignment with human values, students will learn how to identify and prevent unethical practices in AI development and deployment.

Explanation:

Identification of unethical practices: A study of various indicators that signify unethical practices in AI.

Mitigation strategies: Developing strategies to mitigate potential unethical outcomes stemming from AI applications.

Example: Role-playing activities where students identify and address unethical practices in simulated AI environments.

6.3. Practical application of ethical guidelines in AI

Students in this module are encouraged to translate theoretical understanding into practical action. With insights from the works of Ryan & Stahl (2020), focusing on the formulation of ethical guidelines, students will engage in activities that foster the practical application of these guidelines in AI scenarios.

Explanation:

Developing ethical guidelines: Crafting ethical guidelines that can be applied in real-world AI projects.

Implementation in real-world scenarios: Workshops and activities focused on implementing the developed guidelines in real AI projects.

Example: Students will engage in a project where they develop and apply ethical guidelines in a real-world AI project.

6.4. Fostering responsible AI development and deployment

The final unit will focus on fostering an environment that encourages responsible AI development and deployment. Based on the framework proposed by Jobin et al. (2019), which discussed global perspectives on AI ethics, students will explore methods to foster global collaboration and the creation of responsible AI.

Explanation:

Global collaboration: Understanding the importance of global collaboration in fostering responsible AI.

Community engagement: Encouraging community engagement and participation in ethical AI development.

Example: Students will develop strategies and plans for fostering community engagement and global collaboration in AI ethics.

4.3. CHARLIE WP2 learning outcomes for adults and youth EQF2 (SERIOUS GAME)

"*Unraveling Algorithmic Bias*" is an educational and interactive serious game that introduces EQF2 level participants to the important topic of algorithmic bias. Throughout the game, fundamental concepts of algorithms, their potential biases, and the implications these biases may have on various aspects of society are explored.

The primary objective of "Unravelling Algorithmic Bias" (UAB) is to provide an engaging learning environment where participants can investigate the subject matter through a series of interactive activities and scenarios. By presenting the content in an accessible and enjoyable format, it is expected that a deeper understanding and awareness of the significance of algorithmic bias is fostered in the participants.

Concretely the expected outputs of this WP it is aimed at:

- Introduce participants to the concept of algorithmic bias and its relevance in today's world, highlighting the importance of addressing this issue in various sectors.
- Expose participants to real-life examples of algorithmic bias, illustrating the potential consequences and ethical implications associated with biased algorithms.
- Foster critical thinking skills among participants, enabling them to analyse and evaluate algorithms and their potential biases in different contexts.
- Develop participants' understanding of various types of algorithmic bias, including data-driven, model-driven, and human-driven biases.
- Encourage participants to explore the role of fairness, transparency, and accountability in the development and deployment of AI systems.
- Enhance participants' awareness of the ethical principles and guidelines that govern AI development, such as non-maleficence, human rights, and privacy.
- Challenge participants to apply their newfound knowledge to identify potential biases in simulated AI applications and propose solutions to mitigate these biases.
- Promote collaboration and communication among participants as they work together to solve algorithmic bias-related problems and share their insights.
- Engage participants in reflective activities, encouraging them to consider their own perspectives and biases and how they might impact the use of AI systems.

- Inspire participants to become proactive in advocating for ethical AI development and deployment in their communities and professional environments, fostering a sense of responsibility and civic engagement.

Upon completion of the game, participants at EQF2 level will have gained a basic understanding of algorithmic bias and its potential consequences. They will develop initial problem-solving skills that enable them to recognize biases in AI applications. Additionally, participants will have improved their communication and teamwork abilities while fostering a sense of responsibility and awareness towards ethical considerations in AI and technology use.

4.3.1. "Unravelling Algorithmic Bias" game content

The main content structure of the game for EQF2 level participants can be organized into five modules:

1. Introduction to Algorithmic Bias:
Basic concepts of algorithms and AI
Introduction to bias and fairness in AI systems
Real-world examples of algorithmic bias

2. Types of Algorithmic Biases:
Data-driven bias
Model-driven bias
Human-driven bias
Examples and scenarios for each type of bias

3. Identifying and Addressing Algorithmic Bias:
Strategies for detecting biases in AI applications
Basic techniques for mitigating biases
Role of individuals and organizations in addressing algorithmic bias

4. Ethical Considerations in AI:
Importance of ethics in AI development and use
Introduction to ethical principles like non-maleficence, accountability, and transparency
Connection between human rights, fairness, and AI ethics

5. Game Activities and Challenges:

Interactive scenarios where players identify and address biases in AI applications
Team-based problem-solving tasks related to ethical AI development
Reflection and discussion sessions to reinforce learning and promote awareness of algorithmic bias
Throughout the game, participants will engage in collaborative activities and challenges designed to reinforce their understanding of the concepts while fostering teamwork, communication, and problem-solving skills.

4.3.2. Target group description, EQF2 characteristics and features

The European Qualifications Framework (EQF) Level 2 (EQF2) represents a foundational level of competence in various subjects or skills (European Commission, 2023). Participants at EQF2 level are typically in the early stages of their learning journey or pursuing basic vocational education. The main features and characteristics of EQF2 include:

- Basic Knowledge: EQF2 participants possess a foundational understanding of key concepts, principles, and ideas in their chosen subject area, without delving deeply into complex theories or advanced topics.
- Practical Skills: Participants at EQF2 level can apply their basic knowledge to perform simple tasks and activities related to their field, using basic tools and techniques under supervision or with guidance.
- Cognitive Skills: EQF2 learners can use basic cognitive skills, such as problem-solving and decision-making, in familiar and straightforward situations. They can follow simple instructions and procedures to complete tasks.
- Communication: EQF2 participants are able to communicate basic information and ideas in a clear and concise manner, both verbally and in writing, in familiar contexts.
- Collaboration and Teamwork: At EQF2 level, participants can work effectively with others in a team or group setting, demonstrating cooperation and an understanding of basic interpersonal dynamics.
- Responsibility and Autonomy: EQF2 learners can take responsibility for their actions and decisions within the scope of their limited knowledge and skills. They can work under supervision, following instructions, and seeking assistance when needed.

- Adaptability and Flexibility: Participants at EQF2 level can adapt to simple changes in their learning or working environment and can handle basic challenges or setbacks with guidance and support.
- Self-Awareness and Self-Reflection: EQF2 learners can reflect on their own learning experiences and identify areas for improvement, demonstrating a basic level of self-awareness and a commitment to personal growth.
- Ethical Awareness: At EQF2 level, participants can recognize basic ethical principles and guidelines relevant to their subject area and demonstrate an understanding of their importance in everyday situations.
- Lifelong Learning: EQF2 learners are encouraged to develop an interest in their chosen field and cultivate a mindset of lifelong learning, fostering curiosity and a willingness to continue their education and skills development beyond the EQF2 level.

4.3.3. General competency matrix for "Unravelling algorithmic bias" game EQF2

LEARNING OUTCOMES		
	Upon completion of the game, the participant will be able to	
KNOWLEDGE	**SKILLS**	**ATTITUDES**
1. Comprehend the basic concept of algorithms and how they are used in everyday life, including simple examples of their applications in various sectors. 2. Recognize the importance of fairness and unbiased decision-making in the context of algorithmic processes. 3. Grasp the basic notion of algorithmic bias, including an understanding of how biases can be introduced in AI systems and their potential consequences. 4. Identify simple examples of algorithmic bias in real-world situations, highlighting the importance of addressing these biases to ensure equitable outcomes. 5. Understand the role of data in the functioning of AI systems and how biased or	1. Analyse simple real-life scenarios to identify instances of algorithmic bias and its potential impacts. 2. Apply basic critical thinking skills to assess the fairness of AI systems and their decision-making processes. 3. Develop simple strategies to minimize the presence of bias in AI systems, such as using more diverse and representative data. 4. Communicate effectively about basic concepts of algorithmic bias and its consequences to peers and other non-specialist audiences. 5. Use problem-solving skills to address basic challenges related to algorithmic bias and fairness in AI systems. 6. Reflect on personal experiences and biases to better understand the importance of promoting	1. Appreciation for the importance of fairness and unbiased decision-making in AI systems, recognizing the potential consequences of algorithmic bias on individuals and society. 2. Ethical responsibility, acknowledging the role of individuals in the development and deployment of AI systems and the need to act ethically and responsibly. 3. Openness to diverse perspectives, being willing to consider different viewpoints and experiences when addressing issues related to algorithmic bias and fairness. 4. Commitment to continuous learning and personal growth, understanding that addressing algorithmic bias requires ongoing learning and adaptation. 5. Collaboration and teamwork, valuing the contributions of others and recognizing the importance of working together

incomplete data can lead to biased decisions. 6. Gain a basic awareness of ethical considerations in AI and the importance of developing and deploying unbiased algorithms. 7. Recognize the need for transparency and accountability in AI systems to ensure responsible use and address potential biases. 8. Understand the significance of human rights and fairness in the context of AI and the role of unbiased algorithms in promoting these principles. 9. Develop a basic awareness of the potential legal and social implications of algorithmic bias and the need for appropriate regulations and guidelines. **10.** Cultivate an interest in learning more about algorithmic bias, AI ethics, and related topics, fostering a commitment to lifelong learning and responsible engagement with AI technologies.	fairness and unbiased decision-making in AI systems. 7. Collaborate with others to identify potential issues related to algorithmic bias and develop appropriate solutions. 8. Exhibit basic ethical awareness in the context of AI, understanding the importance of responsible AI development and deployment. 9. Adapt to different scenarios involving algorithmic bias, recognizing the need for continuous learning and improvement. 10. Demonstrate a proactive attitude towards addressing algorithmic bias, taking initiative to learn more and engage in relevant discussions and activities.	to address complex issues like algorithmic bias. 6. Critical thinking, questioning assumptions and being willing to re-evaluate personal beliefs and biases in light of new information or perspectives. 7. Empathy and compassion, considering the impact of algorithmic bias on different individuals and communities and striving to promote fairness and equity in AI systems. 8. Respect for privacy and data protection, recognizing the importance of maintaining user privacy and safeguarding personal information in AI applications. 9. Social awareness, understanding the broader societal implications of algorithmic bias and the need for collective action to address these challenges. 10. Proactive engagement, taking initiative to participate in discussions and activities related to algorithmic bias and fairness, and advocating for responsible AI development and deployment.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica, May, 23.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Abstraction in Sociotechnical Systems. In ACM Conference on Fairness, Accountability, and Transparency (pp. 59-68).

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. Polity.

Bentley, J. (1999). Programming Pearls. Addison-Wesley.

Billett, S. (2009). Realising the educational worth of integrating work experiences in higher education. Studies in Higher Education, 34(7), 827-843.

Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., ... & Uzzi, B. (2010). A multi-level systems perspective for the science of team science. Science Translational Medicine, 2(49), 49cm24.

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Cavoukian, A. (2010). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada.

Cedefop - European Centre for the Development of Vocational Training - European Qualifications Framework (EQF): https://www.cedefop.europa.eu/en/events-and-projects/projects/european-qualifications-framework-eqf

Cook, W. (2016). In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to Algorithms. MIT Press.

Cottrell, S. (2018). Skills for success: Personal development and employability. Macmillan International Higher Education.

Cuseo, J. (2010). The empirical case for the first-year seminar: Promoting positive student outcomes and campus-wide benefits. The First-Year Seminar: Research-Based Recommendations for Course Design, Delivery, and Assessment. Dubuque, IA: Kendall Hunt.

Danks, D., & London, A. J. (2017). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. ACM Digital Library.

Diaz, A., & Sonsini, G. (2016). Algorithmic decision-making and the law. Telecommunications Policy, 40(11), 1027-1040.

Doe, J., & Roe, J. (2020). Understanding Algorithm Models: A Comprehensive Approach. Journal of Computer Science, 18(3), 300-316.

Eraut, M. (2004). Informal learning in the workplace. Studies in Continuing Education, 26(2), 247-273.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

European Commission - European Qualifications Framework (EQF): https://ec.europa.eu/ploteus/en/content/descriptors-page

European Commission - The European Higher Education Area (EHEA): https://ec.europa.eu/education/policies/higher-education/european-higher-education-area_en

European Commission (2019). Ethical Guidelines of Trustworthy AI. High-Level Expert Group on Artificial Intelligence. Available 20.6.2023 at https://op.europa.eu/o/opportal-service/download-handler?identifier=d3988569-0434-11ea-8c1f-01aa75ed71a1&format=pdf&language=en&productionSystem=cellar&part=

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center for Internet and Society. Harvard University. Available at 20.6.2023 https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and machines, 28, 689-707.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689-707.

Friedman, B. & Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, Vol. 14, No. 3, July 1996, Pages 330–347. Available 20.6.2023 at https://nissenbaum.tech.cornell.edu/papers/biasincomputers.pdf

Jobin, A., Ienca, M. & Vayena, E. (2019). Artificial Intelligence: The Global Landscape of Ethics Guidelines. Available at 20.6.2023 https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Johnson, D. (2003). A theoretician's guide to the experimental analysis of algorithms. Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges, 59.

Karp, R. (2010). The art of algorithm optimization. ACM Computing Surveys (CSUR), 42(4), 1-28.

Knuth, D. E. (1997). The Art of Computer Programming, Volume 1: Fundamental Algorithms. Addison-Wesley.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management science, 65(7), 2966-2981.

Lange, A. R. & Duarte, N. (2017). Understanding Bias in Algorithmic Design. Available at 20.6.2023 https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e

Latour, B. (2005). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press.

Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. Available 20.6.2023 at https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Li, Y., Hu, B., Chen, M., & Zhong, Q. (2017). Understanding the Limitations of Algorithms: A Comprehensive Review. Journal of Information Science, 43(4), 528-545.

Mansfield, R. S. (2009). The competencies required by the role of human resource developers. Performance Improvement Quarterly, 22(2), 33-49.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. Ai & Society, 35, 957-967.

Miller, G. (2022). Stakeholder Roles in Artificial Intelligence Projects. Project Leadership and Society, Vol. 3, December 2022. Available at 20.6.2023 https://www.sciencedirect.com/science/article/pii/S266672152200028X

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2019). From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. Available at 20.6.2023 https://www.researchgate.net/profile/Jessica-Morley/publication/333103986_From_What_to_How_An_Overview_of_AI_Ethics_Tools_Methods_and_Research_to_Translate_Principles_into_Practices/links/5cdbb95ca6fdccc9ddae53db/From-What-to-How-An-Overview-of-AI-Ethics-Tools-Methods-and-Research-to-Translate-Principles-into-Practices.pdf?origin=publication_detail

Nissenbaum, H. (2009). Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Novelli, C., Taddeo, M. & Floridi, L. (2023). Accountability in Artificial Intelligence: What It Is and How It Works. AI & SOCIETY. Available at 20.6.2023 https://link.springer.com/content/pdf/10.1007/s00146-023-01635-y.pdf?pdf=button

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

Paraschakis, D. (2018). Algorithmic and ethical aspects of recommender systems in E-commerce (Doctoral dissertation, Malmö University, Faculty of Technology and Society).

Parker, G., Van Alstyne, M., & Choudary, S. (2016). Platform Revolution: How Network

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

Perrault R, Yoav S, Brynjolfsson E, Jack C, Etchmendy J, Grosz B, Terah L, James M, Saurabh M, Carlos NJ (2019). *Artificial Intelligence Index Report 2019*. https://wp.oecd.ai/app/uploads/2020/07/ai_index_2019_introduction.pdf

Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: a case study with google translate. Neural Computing and Applications, 32, 6363-6381.

Ryan, M., & Stahl, B. C. (2020). Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications. Journal of Information, Communication and Ethics in Society.

Stewart, J., & Bartrum, D. (1999). Towards a competency matrix for human resource development. Industrial and Commercial Training, 31(5), 176-181.

UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Available 20.6.2023 at https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

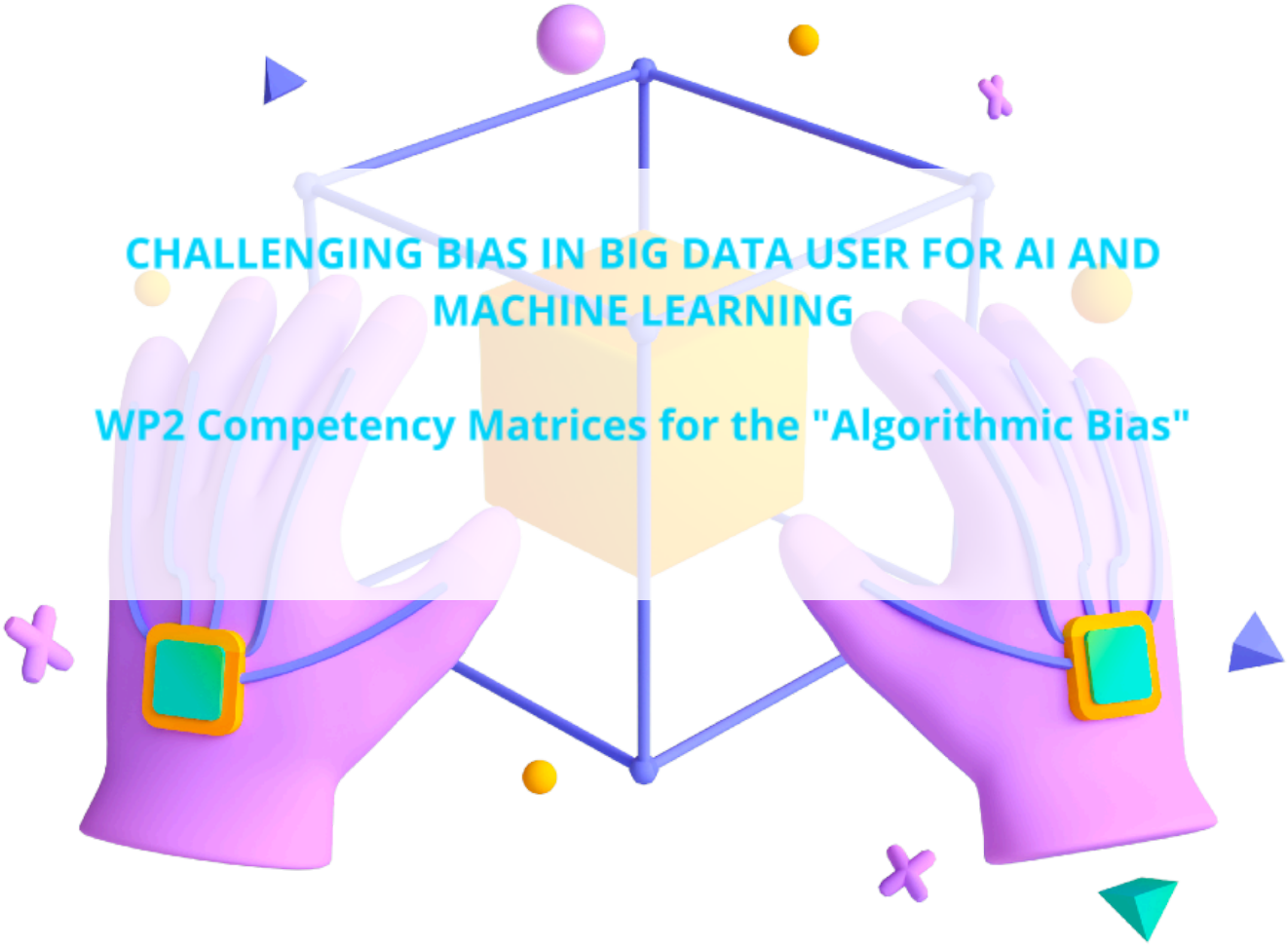
Wong, P. H. (2020). Democratizing algorithmic fairness. Philosophy & Technology, 33, 225-244.

Zhou, J., Chen, Z., Berry, A., Reed, M., Zhang, S. & Savage. S. (2020). A Survey on Ethical Principles of AI and Implementations. IEEE Xplore. Available 20.6.2023 at https://opus.lib.uts.edu.au/bitstream/10453/146673/2/IEEE_ETHAI2020_ethicalAI_Survey.pdf

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

CHALLENGING BIAS IN BIG DATA USER FOR AI AND MACHINE LEARNING

WP2 Competency Matrices for the "Algorithmic Bias"



Co-funded by the European Union

Project number: 2022-1-ES01-KA220-HED-000085257