AI4Gov

Trusted AI for Transparent Public Governance
fostering Democratic Values

Deliverable 5.3

Assessment tools, training activities, best practice guide V1

30-06-2024

Version 1.1

Funded by the European Union	*Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Agency. Neither the European Union nor the granting authority can be held responsible for them.*

PROPERTIES	
Dissemination level	Public
Version	1.1
Status	Final
Beneficiary	JSI
License	

AUTHORS		
	Name	**Organisation**
Document leader	Tanja Zdolšek Draksler	JSI
Participants	Andreas Miltiadous	AUTH
	Christos Gogos	AUTH
	Lila Iosif	MT
	Vicky Stroumpou	MT
	Silvina Pezzetta	WLC
Reviewers	Danai Kyrkou	VIL
	Dimitris Apostolopoulos and Maria Tsoukala	VVV

VERSION HISTORY				
Version	**Date**	**Author**	**Organisation**	**Description**
0.1	16/4/2024	Tanja Zdolšek Draksler	JSI	Initial ToC
0.2	26/4/2024	Tanja Zdolšek Draksler	JSI	ToC update with comments from partners
0.3	13/5/2024	Tanja Zdolšek Draksler	JSI	Initial draft
0.4	5/6/2024	Tanja Zdolšek Draksler	JSI	Draft
0.5	10/6/2024	Tanja Zdolšek Draksler, Lila Iosif, Vicky Stroumpou	JSI, MT	Update from contributions of partner MT
0.6	17/6/2024	Silvina Pezzeta, Andreas Miltiadous, Tanja Zdolšek Draksler	JSI, AUTH, WLC	Update from contributions of partners AUTH and WLC
0.7	18/6/2024	Tanja Zdolšek Draksler	JSI	Bibliography matching. Internal review version preparation.
0.8	21/6/2024	Dimitris Apostolopoulos, Maria Tsoukala	VVV	Internal review
0.9	21/6/2024	Danai Kyrkou	VIL	Internal review
1.0	25/6/2024	Tanja Zdolšek Draksler, Silvina Pezzeta, Christos Gogos	JSI, WLC, AUTH	Integrated corrections
1.1	27/6/2024	Tanja Zdolšek Draksler	JSI	Final version

Table of Contents

Table of Figures

Table of Tables

Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CC	Creative Commons
HRIAAI	Human Rights Impact assessment for AI
LMS	Learning Management System
MOOC	Massive Open Online Course
NGOs	Non-Governmental Organizations
OER	Open Educational Resources
XAI	Explainable Artificial Intelligence

Abstract

This document (D5.3 Assessment tools, training activities, best practice guide) has been developed within WP5 “Citizen-centric training for creating awareness & change” and is presenting the first version of the document, released on M18, while the second version will be delivered on M24 as D5.4. The deliverable describes the learning and training activities with the presentation of used materials (T5.2 and T5.3) and introduces a set of (self) assessment tools and checklists, a presentation of a best practice guide and blueprint covering ethical and technical aspects of AI development processes (T5.4 and partly T5.5).

1 Introduction

1.1 Purpose and scope

This deliverable is the result of the work in WP5 "Citizen-centric training for creating awareness & change" in tasks 5.2, 5.3 and 5.4 in the time period in M3-M18. As task 5.5 is also related to task 5.4, we incorporated some of the 5.5 results also. The present document (D5.3 Assessment tools, training activities, best practice guide v1) is presenting the first version of the deliverable, released on M18. The date for releasing the second version (D5.4 Assessment tools, training activities, best practice guide v2) is M24.

The deliverable describes the learning and training activities with the presentation of used materials (T5.2 and T5.3) and introduces a set of (self) assessment tools and checklists, a presentation of a best practice guide and a blueprint covering ethical and technical aspects of AI development processes (T5.4).

The WP5 is strongly interlinked with all other project tasks, especially with the technical tasks (WP2, WP3, WP6), given the fact that WP5 is transferring knowledge gained to the stakeholders and the general public. Specifically, WP5 will comprise information and documentation about all the different developed components of the AI4Gov platform to enable stakeholders to learn and train how to make use of it. However, the first period of WP5 or first deliverable (D5.3) is mostly focused on general knowledge and technology, to introduce the stakeholders and general public with the basic concepts about bias, artificial intelligence (AI), governance and ethics. On the other side, this deliverable presents a core element of the AI4Gov community, which is the reason why it is also closely linked with the dissemination and community building activities of the project (WP7).

WP5 main aims are to develop and implement training materials designed to enhance awareness and foster citizen engagement, participation, and inclusiveness in democratic processes. The specific objectives are to: (1) create the training materials and establish continuous learning capabilities, (2) promote user awareness and education through workshops and courses, (3) and enhance collaboration and linking with external policy frameworks and stakeholders.

Within the specific needs identification (impact canvas of AI4Gov), the following two were identified in relation to WP5 and D5.3/D5.4: [J] *Increase citizens' awareness and create change based on Inclusive AI* and [L] *User-centric guidelines/training materials produced to adapt and innovate AI and Big Data systems*. Expected results in relation to the stated needs were identified as R14 and R15: *Massive open online courses (MOOCs)* and *User-centric training materials and continuous learning solutions.*

1.2 Document structure

The deliverable is structured in 6 Chapters, starting with Chapter 1 that introduces the document, including the purpose and scope, and document structure. The following chapters (Chapters 2-5)

are dedicated to individual tasks from WP5, more in detail to T5.2, T5.3, T5.4 and T5.5. While Chapters 2 and 3 are focusing on learning and training, Chapters 4 and 5 are focused on assessment tools, guidelines and blueprints for ethical, transparent and trustworthy AI. Chapter 6 concludes the deliverable and Chapter 7 includes the references.

2 Learning and training

This section of the deliverable introduces the learning and training activities with the presentation of used materials building on T5.2 *Stakeholders' Training Activities* and T5.3 *AI4Gov Massive Open Online Courses (MOOCs)*. We will first describe the targeted stakeholders and then move to specific planned and already executed activities, explaining the trainings, educational resources, curriculum and MOOCs.

2.1 Targeted stakeholders

The WP5 content and services target the following stakeholders that were identified from the AI4Gov consortium as the project target groups:

- policy makers and public authorities,
- legal organizations,
- AI researchers and AI solutions integrators,
- citizens.

In relation to digital competencies, we are grouping the users to:

- Tech-savvy[1] people (tech users),
- Non-tech-savvy people (or non-tech users).

Moreover, we are additional grouping them to:

- internal (with special focus on AI4Gov pilot partners),
- external (outside the AI4Gov consortium).

2.2 Stakeholders' training activities

The main aim of the T5.2 *Stakeholders' training activities* is to deliver the AI4Gov training activities with addressing different audiences. The goal is to develop educational resources to address occupational needs and mismatches with the main focus on bias in AI. Training activities were divided into part one (year one of project timeline) and part two (year two and year three of project timeline), as an internal assessment at the start of the 5.2 task showed, that we need to focus on fundamentals first, to build a strong knowledge basis, otherwise the more complex topics will not be understood properly. In the second part of the T5.2 activities, we will include how to use the AI4Gov platform and technologies, allowing different audiences to gain insights into technical developments of the project. In that way, the aim is to close the AI talent gap within the consortium organizations and within other participants of the project's community. Table 1 shows the training activities and educational resources development overview – reporting the

[1] Knowing a lot about modern technology, especially computers. Definition of tech-savvy from the Cambridge Business English Dictionary © Cambridge University Press. https://dictionary.cambridge.org/dictionary/english/tech-savvy

work done in relation to T5.2 from M3 to M18 with a plan for M19 to M27, also highlighting the achieved KPIs (three or more training workshops foreseen: KPI=3).

Table 1: Trainings and educational resources development overview

T5.2	Project timeline		Activity
Part one	**Year 1**	M3 (March 2023)	Assessment of technical vocabulary and tasks understanding from non-tech partners (internal activities)
		M4 (April 2023)	Preparation of first training with focus on tech fundamentals
		M5 (May 2023)	1st AI4Gov training, hybrid, Ljubljana, Slovenia (KPI 1)
		M6 (June 2023)	Educational resources development (video lectures, transcripts, translations)
		M10 (October 2023)	Preparation of second training
		M11 (November 2023)	2nd AI4Gov training, hybrid, Thessaloniki, Greece (KPI 2)
		M12 (December 2023)	Educational resources development (video material)
Part two	**Year 2**	M15 (March 2024) / M16 (April 2024)	On March 13, 2024, the European Parliament formally adopted the EU AI Act. Preparation of video material: "Introduction to AI Act".
		M21-M23 (September – October - November 2024) - PLAN	3rd AI4Gov training (venue: Spain or Romania; hybrid) (KPI 3)
	Year 3	M25-M27 (January – March 2025) - PLAN	Possible 4th AI4Gov training, exploiting all developed AI4Gov technologies and the AI4Gov platform (venue: Spain or Romania; hybrid) (KPI 4)
*T5.2 duration is M3-M27.			

Next, we describe the carried-out activities from M3 to M18 in more detail.

2.2.1 1st AI4Gov training

The 1st AI4Gov training workshop "Bias in AI (focus on fundamentals)" was held on May 16th, 2023, in Ljubljana, Slovenia as a hybrid event. The first session focused on the fundamentals of AI and bias, as well as the impact of bias on human rights, especially for underrepresented groups. In the second session, participants listened to case study presentations and participated in practical

exercises divided into groups (hands on session). The program of the training is presented in Table 2.

Table 2: Program of the 1st AI4Gov training

Time	Title	Presenter
12.00-12.10	Welcome and a short introduction to the AI4Gov project	Jožef Stefan Institute, Slovenia, training organizer (Tanja Zdolšek Draksler) & Maggioli SPA, Greece, AI4Gov project leader (Spiros Borotis)
12.10-12.30	What is bias?	Georgia Panagiotidou, AUTH, Greece [online]
12.30-12.50	Fundamentals of AI	Erik Novak, JSI, Slovenia
12.50-13.10	AI and bias (bias in algorithms, bias in data)	George Manias, UPRC, Greece
13.10-13.30	Impact of bias on human rights (underrepresented groups)	Mpampis Chatzimallis, ViLabs, Greece
13.30-13.40	Coffee break and networking	
13.40-14.30	Hands on session: case studies on detecting bias. Followed by a panel discussion.	Group work (and breakout rooms in Zoom) lead by AI4Gov partners. Moderated by Alenka Guček, JSI.
14.30-14.40	Break and networking	
14.40-15.00	Policy/public administration example from Slovenia: "Slovenian national AI programme"	Samo Zorc, Ministry of Digital Transformation, Slovenia
15.00-15.20	Citizen's example: "Empowering civil society for oversight and policy-making"	Jasmina Ploštajner, Today is a new day, Institute for other studies, Slovenia (NGO)
15.20-15.40	How is AI regulated?	Federico Marengo, White Label Consultancy, Denmark [online]
15.40-16.30	Final discussion and closure with networking	

The training was very well received from the audience, which was mixed, representing policy makers and public organisations (also European Commission representatives), academia, IT

industry. All together there were 48 attendees - 18 persons participated in-person and 30 online (see Table 3 for more information).

Table 3: Online participants country of residence and organization/sector

Organization of participant or sector	Country
Maggioli	Greece
research	North Macedonia
Faculty of Law, University of Ljubljana	Slovenia
European Commission REA	Belgium
KT4D	Italy
European Commission REA	Belgium
Private sector	Slovenia
University of Piraeus	Greece
Ministry of Tourism	Greece
University of Ioannina	Greece
Leiden University	Netherlands
CSI Piemonte	Italy
PPC	Greece
European Commission	Belgium
Open Data Services	United Kingdom
University of Ljubljana	Slovenia
WLC	Italy
Municipality of Vari-Voula-Vouliagmeni	Greece
University of Thessaly	Greece
University of Piraeus	Greece
University of Piraeus	Greece
Aristotel University of Thessaloniki	Greece
Ministry of Tourism	Greece
Ministry of Tourism	Greece
European Commission	Belgium
Centre for European Perspective / Peace Operations Training Centre	Slovenia
Ministry of Defence/Secretary general/Legal service	Slovenia
Ministry of Labour, Family, Social Affairs and Equal Opportunities	Slovenia
Jožef Stefan Institute	Slovenia
University of Maribor	Slovenia

It should be further noted that two Slovenian ministries actively participated, namely the Ministry of Digital Transformation, and the Ministry of Foreign and European Affairs (members of the

Coordination Group on International Humanitarian Law). In particular, the latter gave us very positive feedback after the event, showing a strong interest in strengthening their knowledge on AI and also in future possible learning and training activities participation from their side. Figures 1 and 2 show one of the lectures and one group during the hands-on session (work in groups).

Figure 1: Lecture at the 1st AI4Gov training

Figure 2: Group work during the hands-on session (listening to the reported results from other groups)

The training was filmed by VideoLectures.NET and the filmed material forms the basis for the AI4Gov educational resources, that will be presented more in detail in Chapter 2.2.4.

2.2.2 2nd AI4Gov training

After an internal discussion, it was decided to organize a special training with a focus on Greek public authorities and in Greek language (personalised to their needs). This was the 2nd AI4Gov training workshop titled "Trusted AI for Transparent Public Governance" (Figure 3 shows a photo from the event). It was held on November 21st, 2023, in Thessaloniki at the Aristotle University of Thessaloniki, Greece as a hybrid event and in connection to the AI4Gov panel discussions, that was held in the morning at the same day. As the morning discussions had the audience of 170 attendees, the attendance of the training was lower: 50 people, 25 people in person and 25 people participating online.

Figure 3: Lecture at the 2nd AI4Gov training

The program was focused on technical aspects mostly, but involving also social sciences with the topic about bias (see Table 4 for the detailed program).

Table 4: Program of the 2nd AI4Gov training

Time	Title	Presenter
13.30- 13.50	AI-based Decision Making in the era of Big Data	George Manias, UPRC, Greece
13.50- 14.10	Bias – Are we aware of it?	Despina Natsi, AUTH, Greece
14.10- 14.30	Responsible and Bias-free AI-Based Decision-Making	Kostis Mavrogiorgos, UPRC, Greece
14.30- 14.50	Blockchain and e-government	Dimitris Ntalaperas, UBITECH, Greece
14.50- 15.00	Break	
15.00-15.30	Hands on session (examples from practise)	ViLabs, AUTH, UPRC
15.30	Summarizing and closing session work	

2.2.3 Future plans for trainings

Training activities in Y2/Y3 will be more oriented into how to use the AI4Gov platform and developed technologies, allowing different audiences to gain insights into technical developments of the project. For Autumn 2024 we are planning to organize a training in Romania with sister project Ithaca and an additional possible training in Madrid with the pilot partners (with special focus on pilot partner from Spain). The training content will shift from fundamentals to specific AI4Gov developed technologies and services.

2.2.4 Educational resources

The organised trainings present the main source for AI4Gov educational resources. Especially the first training was very valuable for this purpose, as it was filmed by the VideoLectures.NET[2] team and valuable video lectures were produced in the duration of 3 hours. The AI4Gov educational resources that derived from this event, consists of video materials, presentation slides, transcripts to English and subtitles to Greek, Spanish and Slovenian language (all languages of AI4Gov pilots). Video lectures from the 1st AI4Gov training are openly available online with presentation slides and subtitles on the VideoLectures.NET platform[3] (9 videos) and on YouTube via the AI4Gov project channel (playlist of the 1st training was created)[4] (8 videos). Moreover, the 2nd AI4Gov training delivered recordings of the presentations, that are also available online as videos on YouTube AI4Gov project channel (playlist of the 2nd training was created[5] (8 videos)). These materials, alongside with AI4Gov deliverables, form also a basis for AI4Gov MOOCs, that will be presented in Chapter 3.

Moreover, as the AI4Gov consortium is also up to date with the developments in the domain of AI, we found it crucial to produce a short video on the AI Act, after the European Parliament formally adopted this regulation in Spring. Prepared video "Introduction to AI Act" is an AI generated video included in the educational resources and also available at the AI4Gov YouTube project channel.

As there is a demand for online educational resources for teaching and learning, we are operating towards generating the educational resources in the type of the open educational resources (OER). UNESCO (n.d.) is defining OERs as "learning, teaching and research materials in any format and medium that reside in the public domain or are under copyright that have been released under an open license, that permit no-cost access, re-use, re-purpose, adaptation and redistribution by others". JSI being the UNESCO Chair on Open Technologies for Open Educational Resources and Open Learning[6], it is very important to disseminate the OER principles. Moreover, with the active partnerships with the UNESCO Chair in AI[7] and UNESCO Chair in teacher training

[2] https://videolectures.net/
[3] Online training course: https://videolectures.net/AI4GOVtraining2023_ljubljana/
[4] https://www.youtube.com/playlist?list=PLX1Lv490hMrMizdl5HDa-MOAqwy3ItOPM
[5] https://www.youtube.com/playlist?list=PLX1Lv490hMrN_bBcZSl8o58mmrorztBRX
[6] https://unesco.ijs.si/
[7] https://unesco.org.uk/unesco-chair-on-artificial-intelligence-at-university-college-london/

technologies with OER[8], we foreseen to build on the existing alliances also in the scope of the AI4Gov activities.

2.2.4.1 Existing materials, external resources

In addition to the specifically AI4Gov prepared educational resources, we are also using existing materials (external resources), like specific video lectures available on the VideoLectures.NET platform (managed from partner JSI), an award-winning free and open access educational video lectures repository. Since its establishment in 2007, the repository is providing high quality didactic contents not only to the scientific community but also to the general public. VideoLectures.NET contains over 20.000 lectures by scholars and scientists.

[8] https://chaireunescorelia.univ-nantes.fr/

3 AI4Gov Massive Open Online Courses (MOOCs)

This segment of the document presents the AI4Gov Massive Open Online Courses (MOOCs) that are being developed in the scope of T5.3, which started in M4 and will end in M27.

3.1 Methodology

The goal is to develop MOOCs, in which results produced from other AI4Gov WPs will be included. The entire MOOCs will comprise learning materials of 40-60 hours, while the methodology includes: a) theory, b) use case examples and c) assessment. In relation to the types of materials used, we used video materials, text content, presentation slides, images, quizzes and also comics. Within the scope of T5.3 KPIs, three MOOCs are foreseen (KPI=3).

Building on the educational theory, we are well aware about the advantages and limitations of MOOCs. On one side, MOOCs accommodate the worldwide demand for educational materials, making a significant difference by offering online accessibility without any geographical restrictions (Shah et al., 2022). On the other, low completion rates are being observed, despite a lot of investment in course development from the course provider side (Reich & Ruipérez-Valiente, 2019). Nevertheless, MOOCs are being considered to be an appropriate tool to reach the audience and raise awareness on the topic of bias in AI.

In the following section of this document, we present the curriculum for the AI4Gov MOOCs, which presents the basis for the carried-out work.

3.2 Curriculum “Trustworthy and democratic AI”

A curriculum is a pivotal element in education as it serves multiple critical functions, for example it provides structure and aligns with goals, but it also shapes learning experiences, cultivates critical thinking, reflects societal needs, encourages lifelong learning, reflects culture and is the basis for progress measurement. In summary, the aim of a curriculum is to create a dynamic and responsive educational environment that not only imparts knowledge but also develops the skills and values necessary for learners to thrive in a complex and rapidly evolving world. In essence, a curriculum is the backbone of the educational process, guiding both teaching and learning towards achieving desired outcomes (Darling-Hammond et al., 2020; Eilks & Hofstein, 2017).

In the realm of AI4Gov, a curriculum was developed titled “Trustworthy and democratic AI”. The main objective of the curriculum is to equip learners with the knowledge, skills, and ethical principles necessary to design, develop, and deploy AI systems that are both trustworthy and free from bias (or low bias, as unbiased systems are practically impossible to achieve). This framework aims to address the growing need for responsible AI development by fostering a deep understanding of the complexities and challenges related to AI bias, fairness, transparency, and accountability.

It is important to mention, that the curriculum is not a final defined document, but rather an evolving framework, that is constantly updated throughout the duration of the project. This model is driven by a participatory approach that defines a series of iterations in the development of learning courses and in the production of learning materials, with multiple revisions from internal and external stakeholders, in order to ensure high quality outcomes that are also tailored to end-users needs. The curriculum broken down into modules and lessons is presented in Table 5.

Table 5: Curriculum "Trustworthy and democratic AI

Module 1	Introduction to Trustworthy and Democratic AI (Fundamentals)
Lesson 1.1	What is Artificial Intelligence?
Lesson 1.2	The Role of AI in Society
Lesson 1.3	The Scope of Trustworthy AI
Lesson 1.4	The Importance of Ethical AI
Module 2	Bias in AI - Understanding Bias
Lesson 2.1	Defining Bias in AI
Lesson 2.2	Types of Bias and Sources of Bias
Lesson 2.3	Impact of Bias (Real-World Consequences, Bias in Decision-Making, Bias and Discrimination)
Module 3	Data and Bias
Lesson 3.1	Bias in Data Collection
Lesson 3.2	Data Sampling Methods
Lesson 3.3	Ethical Data Sourcing
Lesson 3.4	Data Pre-processing and Bias Reduction (Data Cleaning for Bias; Fair Data Pre-processing Techniques)
Lesson 3.5	Real-world Data Bias Case Studies
Module 4	Ethical AI Governance

Lesson 4.1	Legal and Regulatory Frameworks (Overview of AI Regulations, GDPR and Data Privacy, National and International AI Policies)
Module 5	Bias Mitigation
Lesson 5.1	Debiasing Techniques (Introduction to Debiasing, Post-processing Methods, In-processing Methods)
Lesson 5.2	Fairness Constraints and Reweighting (Fairness Constraints in Model Training, Reweighting and Fairness-aware Learning, Implementing Bias Mitigation in AI Models)
Module 6	Explainable AI (XAI)
Module 7	AI Trustworthiness Evaluation
Lesson 7.1	Evaluation Metrics for Trustworthy AI (Metrics for Fairness and Bias, Trustworthiness Assessment, Evaluation Tools and Frameworks)
Lesson 7.2	Ethical Auditing and Reporting (Ethical Auditing Processes, Reporting and Transparency, Corrective Actions and Re-evaluation)
Module 8	Responsible AI Development
Lesson 8.1	Applying Ethical Guidelines (Incorporating Ethical AI Principles, Case Studies in Ethical AI Development, Ensuring Ethical Compliance)
Lesson 8.2	Ethical AI Project Lifecycle (Integrating Ethical Considerations, Continuous Ethical Review, Ethical AI Development Best Practices)
Module 9	Real-World Applications and Case Studies
Lesson 9.1	Case Studies in Trustworthy AI (Analysis of Real-World Examples, Success Stories in Ethical AI, Addressing Bias and Trustworthiness Challenges)

This framework provides a comprehensive education in creating, maintaining and also using trustworthy and democratic AI systems. Each lesson builds on the previous one to ensure a deep understanding of the topic, and case studies and real-world applications are used to reinforce practical knowledge. In addition, the curriculum is linked to specific learning objectives that are presented in the following chapter.

3.2.1 Specific learning objectives

Participants/learners will be equipped with the knowledge and skills required to identify bias and to contribute to the responsible development and deployment of AI systems, ensuring that AI technologies are both effective and ethically sound. Specific learning objectives in the realm of this curriculum are:

1. Understanding AI Fundamentals:
   a) Understand the concepts in artificial intelligence, machine learning, and deep learning.
   b) Recognize the transformative potential and ethical implications of AI in various domains.
   c) Explore ethical frameworks and principles guiding AI development.
   d) Recognize the importance of fairness, transparency, and accountability in AI systems.
2. Bias Awareness and Recognition:
   a) Identify the types and sources of bias that can manifest in AI systems.
   b) Understand the real-world consequences of biased AI, including social and ethical implications.
3. Data Collection and Pre-processing:
   a) Learn how data collection and pre-processing can introduce bias into AI models.
   b) Implement best practices for collecting, cleaning, and preparing data to reduce bias.
4. Ethical AI Governance:
   a) Study legal and regulatory frameworks governing AI, such as GDPR and AI Act.
   b) Explore the role of responsible AI governance in organizations.
5. Algorithmic Bias Mitigation:
   a) Understand various techniques to mitigate algorithmic bias in AI models.
   b) Apply debiasing methods, reweighting strategies, and fairness constraints to model development.
6. Explainable/Interpretable AI:
   a) Examine methods for making AI models more explainable.
   b) Appreciate the importance of transparency in AI decision-making processes.
7. AI Trustworthiness Evaluation:
   a) Assess AI models and systems for fairness, bias, and trustworthiness.
   b) Implement evaluation techniques to ensure AI systems meet ethical and operational standards.
8. Responsible AI Development:
   a) Apply ethical guidelines and responsible AI principles to the end-to-end AI development lifecycle.
   b) Develop AI systems that prioritize fairness, accountability, and transparency.
9. Real-World Applications:
   a) Analyze case studies and real-world examples of trustworthy AI.
   b) Gain practical experience in addressing bias and trustworthiness challenges in AI projects.

The framework empowers learners to become advocates for trustworthy and democratic AI in their respective roles and organizations. Based on the developed curriculum, learners will gain different competencies.

3.2.2 General competencies gained

The "Trustworthy and Democratic AI" learning framework is designed to enhance a range of competencies and skills in individuals who complete the learning program. These competencies are essential for those who aim to develop, deploy, or manage AI systems while ensuring they are ethical, democratic, and trustworthy. However, these competencies are also crucial for those who are users of AI systems. Here are some of the key competencies that participants can expect to gain or improve through this learning framework:

- Ethical Competence:
  - Ability to recognize and navigate ethical dilemmas in AI development and deployment.
  - Proficiency in applying ethical principles to AI projects.
- Bias Awareness and Mitigation:
  - Skill in identifying sources of bias in AI systems.
  - Proficiency in implementing techniques to reduce bias in AI models and data.
- Transparency and explainability:
  - Capability to understand what makes AI models more transparent and understandable.
  - Understanding what does it mean to generate explanations for AI model outputs (XAI).
- Data Management and Privacy:
  - Competence in handling data ethically and ensuring privacy compliance.
  - Understand data pre-processing and cleaning to reduce data bias.
- Legal and Regulatory Knowledge:
  - Understanding of relevant AI regulations and laws.
  - Proficiency in navigating legal and regulatory requirements.
- Responsible AI Development:
  - Ability to incorporate ethical considerations throughout the AI development lifecycle.
  - Skill in designing AI systems that prioritize fairness, accountability, and transparency.
- Critical Thinking and Problem Solving:
  - Competence in analyzing complex issues related to bias and ethics in AI.
  - Skill in devising solutions to address ethical challenges in AI projects.
- Communication and Stakeholder Engagement:
  - Proficiency in communicating ethical and bias-related concerns to stakeholders.
  - Skill in engaging with diverse stakeholders, including ethicists, policymakers, and end-users.
- Auditing and Evaluation:

- Ability to evaluate AI systems for fairness and trustworthiness.
- Skill in conducting ethical audits and impact assessments.

- Project Management:
  - Proficiency in managing AI projects with ethical considerations.
  - Capability to balance project objectives with ethical priorities.
- Case Study Analysis:
  - Competence in learning from real-world case studies of ethical lapses.
  - Skill in applying lessons from case studies to project decisions.
- Collaboration and Teamwork:
  - Understanding the importance of multidisciplinary teams and collaborate work to address ethical and bias-related challenges.
- Leadership in Ethical AI:
  - Capability to lead and advocate for ethical AI within organizations and industries.

By gaining or improving these competencies, participants will be well-prepared to contribute to the development of trustworthy and democratic AI systems and to address the ethical challenges that arise in the rapidly evolving field of artificial intelligence. These competencies are highly valuable in roles related to AI development, governance, ethics, compliance, and more.

3.2.3 Digital competencies

The "Trustworthy and Democratic AI" curriculum includes various digital competencies, which are crucial in today's digital and AI-driven world. These competencies encompass both technical and non-technical digital skills. Here are the digital competencies covered:

- AI and Machine Learning Competence:
  - Understanding the fundamentals of AI, machine learning, and deep learning.
  - Proficiency in working with AI development tools and libraries.
- Data Management and Analysis Skills:
  - Competence in data collection, preparation, and analysis.
  - Skill in working with data analytics and visualization tools.
- Algorithmic Design and Implementation:
  - Understanding the process of designing and implementing algorithms for AI systems.
- Digital Privacy and Security:
  - Understanding of data privacy and security best practices.
  - Proficiency in securing AI systems and protecting data.
- Coding and Programming Skills:
  - Skill in optimizing AI code.
- Ethical AI Development with Technology:
  - Ability to implement ethical considerations in the AI development process.
  - Proficiency in using technology to reduce bias and ensure fairness.

Regulatory and Compliance Knowledge:

- Understanding of AI-related regulations and compliance requirements.
- Skill in adhering to legal and regulatory standards in AI development.

Digital Project Management:

- Skills in managing AI projects, including planning, execution, and evaluation.

Auditing and Evaluation Tools:

- Ability to use digital tools for auditing and evaluating AI models for fairness and trustworthiness.

Information Literacy:

- Competence in evaluating sources and staying up to date with AI developments.

Digital Communication Skills:

- Capability to effectively communicate and collaborate digitally.
- Skill in using digital communication tools and platforms.

These digital competencies are essential for individuals and organizations involved in AI development, ethics, and governance. They ensure that participants can navigate the digital landscape, understand and utilize the tools and technologies involved in AI, and implement ethical considerations in the digital development process.

After having the theory as the basis, we needed to decide on the learning platform, which is presented in the following chapter.

3.2.4 Open learning platform

As we wanted to ensure a great learning experience, we decided not to use possible existing solutions from partners, as not all needs would be addressed. The process of the platform selection started with preparing a list of criteria, namely:

- ✓ platform that supports open learning practices,
- ✓ free of charge/no fee for learning course developers and for learners/users,
- ✓ easy to use,
- ✓ includes a Learning Management System (LMS),
- ✓ if possible, Europe based,
- ✓ open licences support and usage (Creative Commons[9] (CC) licences),
- ✓ ownership,
- ✓ digital badge upon completion,
- ✓ number of learners not limited,
- ✓ learning materials not limited.

[9] "Creative Commons licenses give everyone from individual creators to large institutions a standardized way to grant the public permission to use their creative work under copyright law". https://creativecommons.org/share-your-work/cclicenses/

Based on the criteria list, we made an assessment of available learning platforms and selected the platform OpenLearn Create[10] as it meets all requirements. OpenLearn Create is "an innovative leading open educational platform where individuals and organisations can publish their open content, open courses and resources. It is Moodle based and has tools for collaboration, reuse and remixing" (see Figure 4).

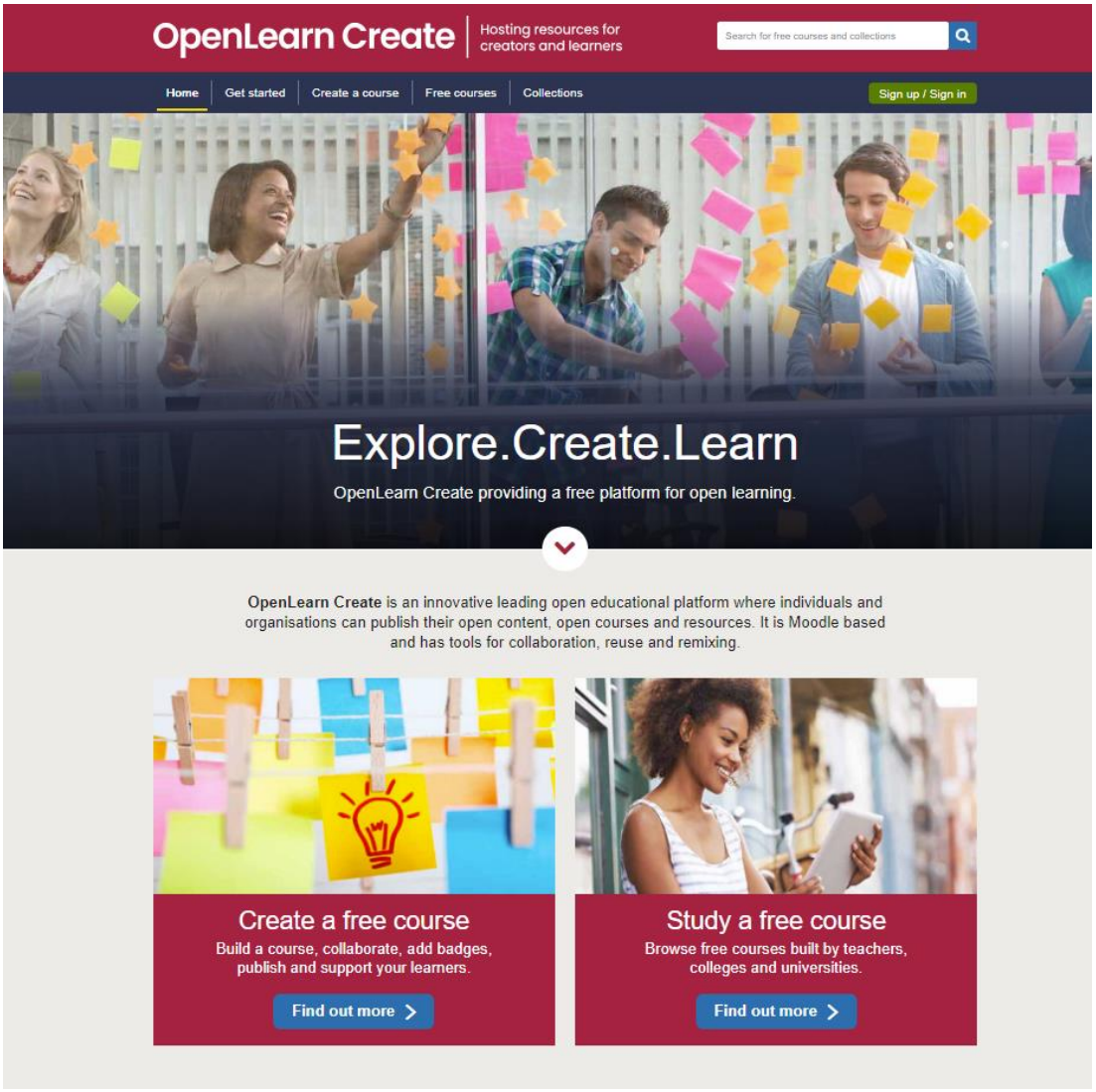

Figure 4: OpenLearn Create platform capture

In the next chapters, the MOOCs are being presented in detail.

[10] https://www.open.edu/openlearncreate/

3.2.5 MOOCs plan and focus group

The presented theory forms the basis for the AI4Gov MOOCs and was used for detailed planning of MOOCs development (Table 6).

Table 6: AI4Gov MOOCs plan/overview

No.	Title	Level	Duration	Available (public)
MOOC1	"Trustworthy and democratic AI – Fundamentals"	Beginner (basics)	10 hours	Spring 2024 (first half of 2024) → end of April
MOOC2	"Trustworthy and democratic AI - creating awareness and change"	Beginner to intermediate (AI4GOV developed results included)	15 hours	Expected second half of 2024
MOOC3	"Trustworthy and Democratic AI experts"	Intermediate to expert level (AI4GOV developed results included)	15 hours	Expected end of 2024/start of 2025

During the first MOOC development, we set up a focus group formed by five members from the Greek Ministry of Tourism. The main idea was to evaluate the developed MOOC on the project level, namely, to collect focus group feedback for possible corrections, updates etc. Each member of the focus group has registered (created account) in the OpenLearn Create platform with goal to assess all modules and give feedback. All four modules and lessons with quizzes were revised by the focus group and each member completed the assessment list in the form of a prepared online questionnaire, with feedback and suggestions for possible corrections. Overall, the focus group found the first MOOC very enlightening for a beginner in AI fundamentals with enriched learning and training material (video lectures, slides, comics, quizzes). Additionally, the focus group shared that the first MOOC gives a good perspective in understanding bias and ethics in AI.

The focus group will be active also in the future, for assessment of MOOC2 and MOOC3 development.

The first MOOC was developed and made publicly available last week in April 2024 and is in detail presented in next chapter.

3.2.6 MOOC 1: Trustworthy and democratic AI – Fundamentals

The first AI4Gov MOOC titled Trustworthy and democratic AI – Fundamentals is designed to provide learners around the globe with the foundational knowledge necessary to understand bias in AI. It aims to democratize the knowledge of AI ethics, governance, and technical standards to

ensure that future AI systems are developed with fairness, accountability, and transparency at their core. Figure 5 shows the visual graphic prepared for the first MOOC.

Figure 5: Graphics for the MOOC 1

The curriculum behind the MOOC 1 is presented in Table 7, showing the modules, lessons, specific learning objectives and also the AI4Gov WP content that was specifically included.

Table 7: Curriculum specification for MOOC 1

Module	Lessons	Specific learning objectives	AI4Gov
Module 1: Introduction to Trustworthy and Democratic AI (Fundamentals)	Lesson 1.1: What is Artificial Intelligence? Lesson 1.2: The Role of AI in Society Lesson 1.3: The Scope of Trustworthy AI Lesson 1.4: The Importance of Ethical AI	Understanding AI Fundamentals: (1) Understand the concepts in artificial intelligence, machine learning, and deep learning. (2) Recognize the transformative potential and ethical implications of AI in various domains.	WP4 WP5 WP2
Module 2: Bias in AI - Understanding Bias	Lesson 2.1: Defining Bias in AI Lesson 2.2: Types of Bias Lesson 2.3: Sources of Bias Lesson 2.4: Impact of Bias (Real-World Consequences, Bias in Decision-Making, Bias and Discrimination)	Bias Awareness and Recognition: (1) Identify the types and sources of bias that can manifest in AI systems. (2) Understand the real-world consequences of biased AI, including social and ethical implications.	WP2 WP5 WP4

Module	Lessons	Specific learning objectives	AI4Gov
Module 3: Data and Bias	Lesson 3.1: Bias in Data Collection Lesson 3.2: Data Sampling Methods Lesson 3.3: Ethical Data Sourcing Lesson 3.4: Data Pre-processing and Bias Reduction Lesson 3.5: Real-world Data Bias Case Studies	Data Collection and Pre-processing: (1) Learn how data collection and pre-processing can introduce bias into AI models. (2) Implement best practices for collecting, cleaning, and preparing data to reduce bias.	WP2 WP6
Module 4: Ethical AI Governance	Lesson 4.1: Legal and Regulatory Frameworks	Ethical AI Governance: (1) Study legal and regulatory frameworks governing AI, such as GDPR and AI Act. (2) Explore the role of responsible AI governance in organizations.	WP5 WP2

The learning course is openly available to all interested stakeholders and is completely free of charge. Figure 6 shows the QR code for accessing the MOOC 1.

Figure 6: QR code MOOC1

As we wanted to adapt to different stakeholders, we decided for two options of usage. First option, the user can go through the learning course without being registered (no access to assessment of knowledge, like quizzes). Second option, the user can register and exploit the full potential of the course (namely access also the assessment part, receive a digital badge and a certificate of completion, contact the course teachers and administrators etc.). In the process of the learning course development, we relied heavily on the focus group. Course management is done from partner JSI.

The learning analytics for the first month of active MOOC1 shows for now a low number of registered users being 30, but on the other side a high number of guests visits being 586. We plan to focus more on the MOOC1 dissemination in the following months in close collaboration with WP7 and with the aim to increase the numbers of registered and unregistered users (guest visitors). In the following chapter, we present the snapshots from MOOC1.

3.2.7 Snapshots

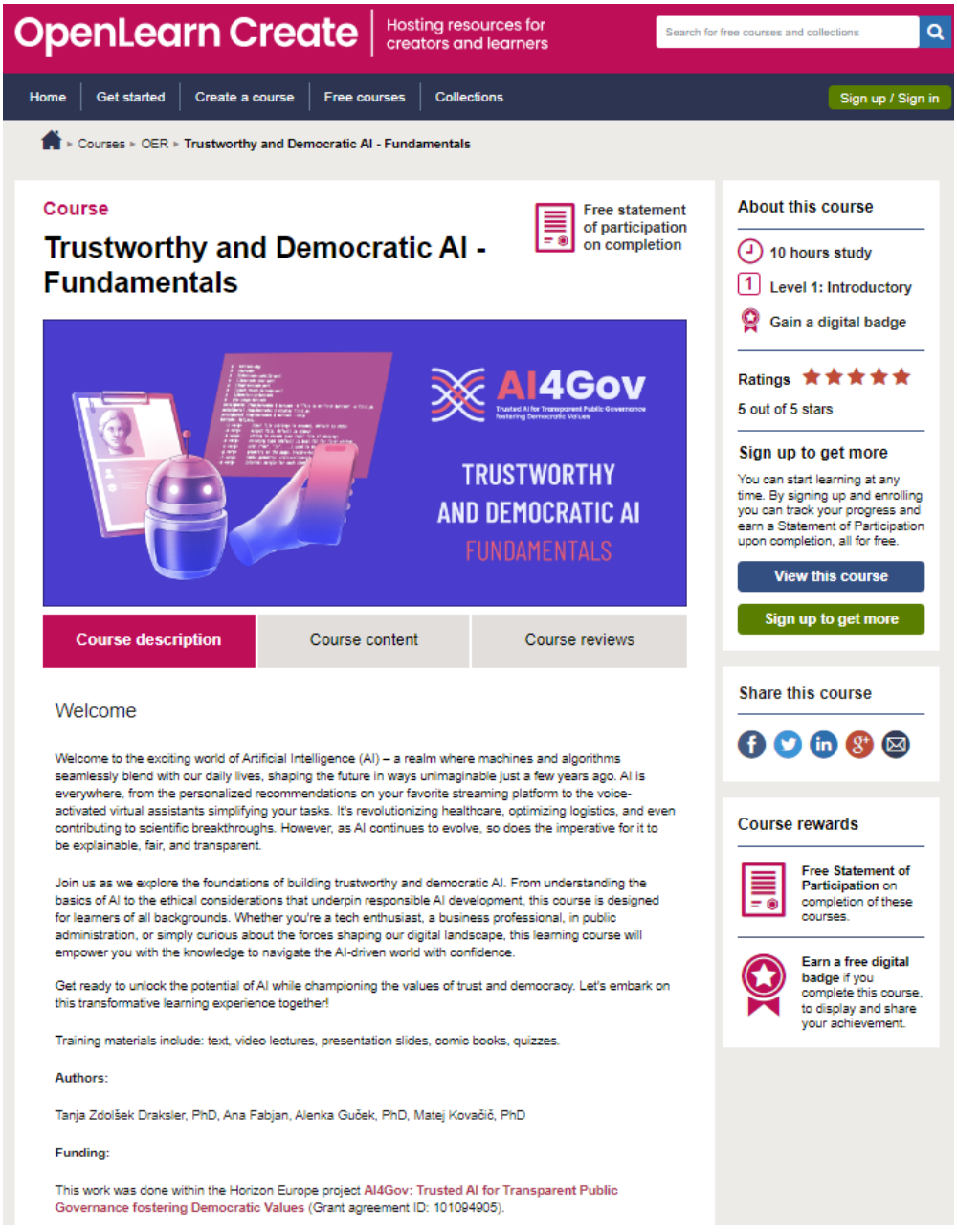

Figure 7: Snapshot from the MOOC1 number one

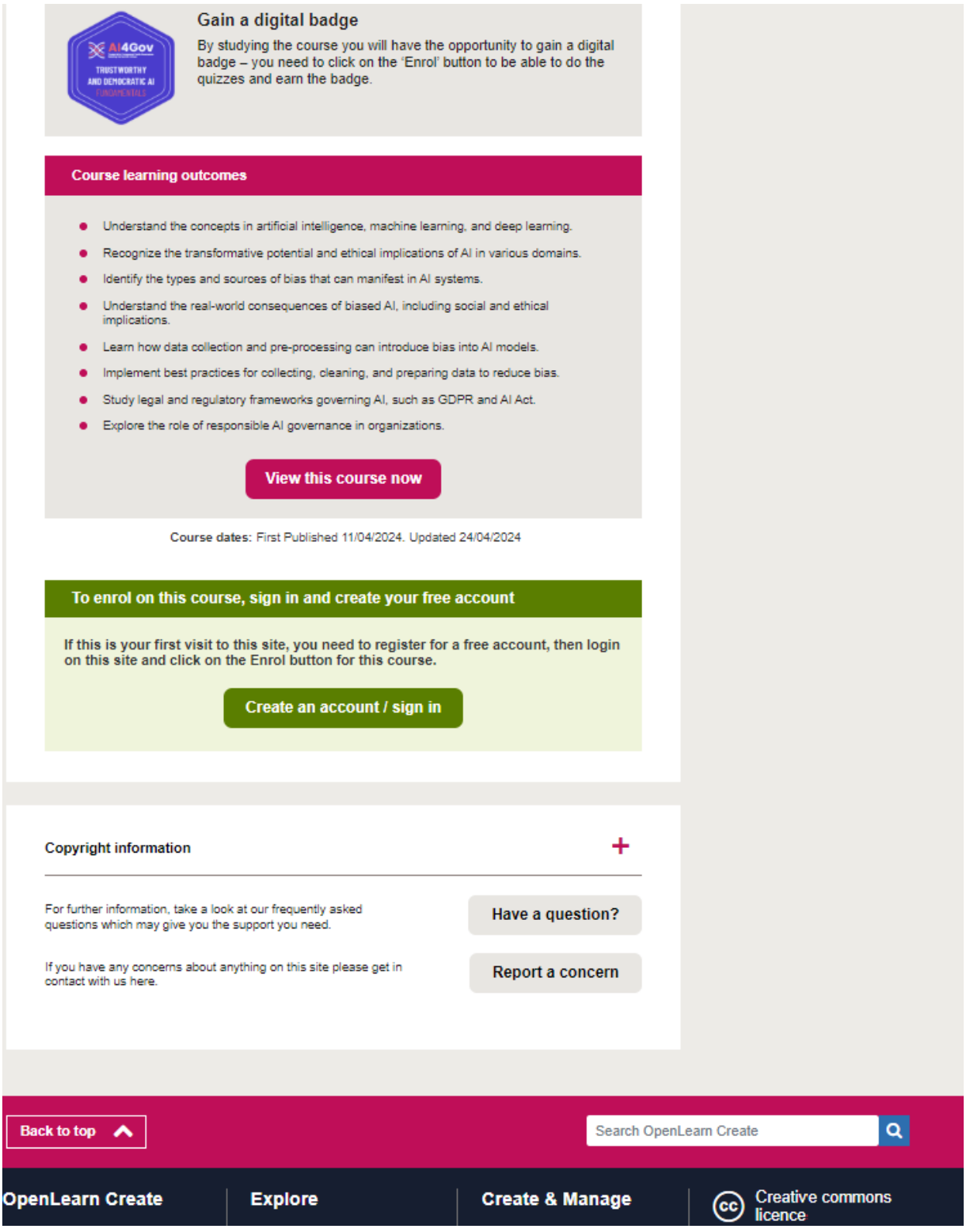

Figure 8: Snapshot from the MOOC1 number two

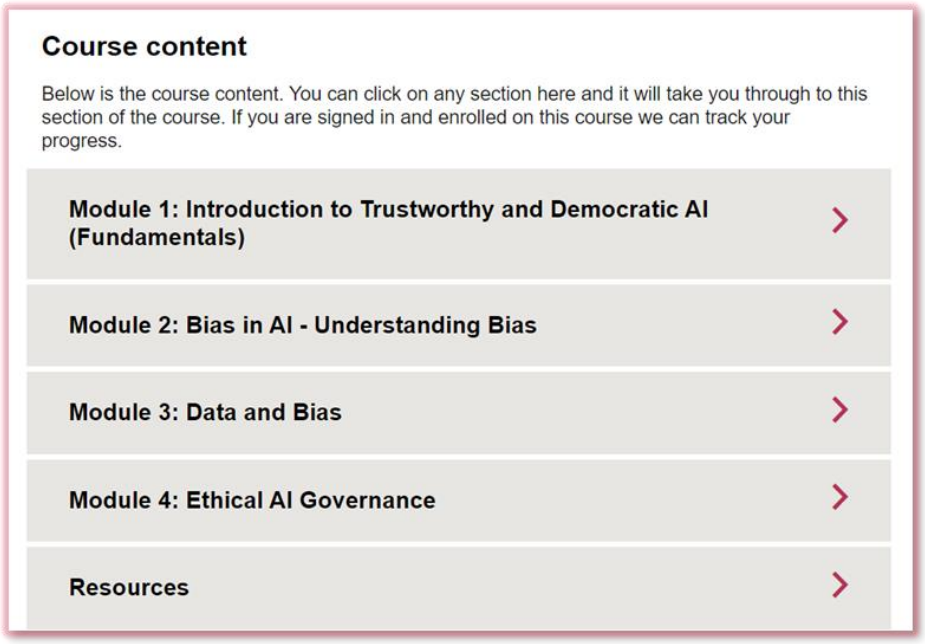

Figure 9: Snapshot from the MOOC1 number three

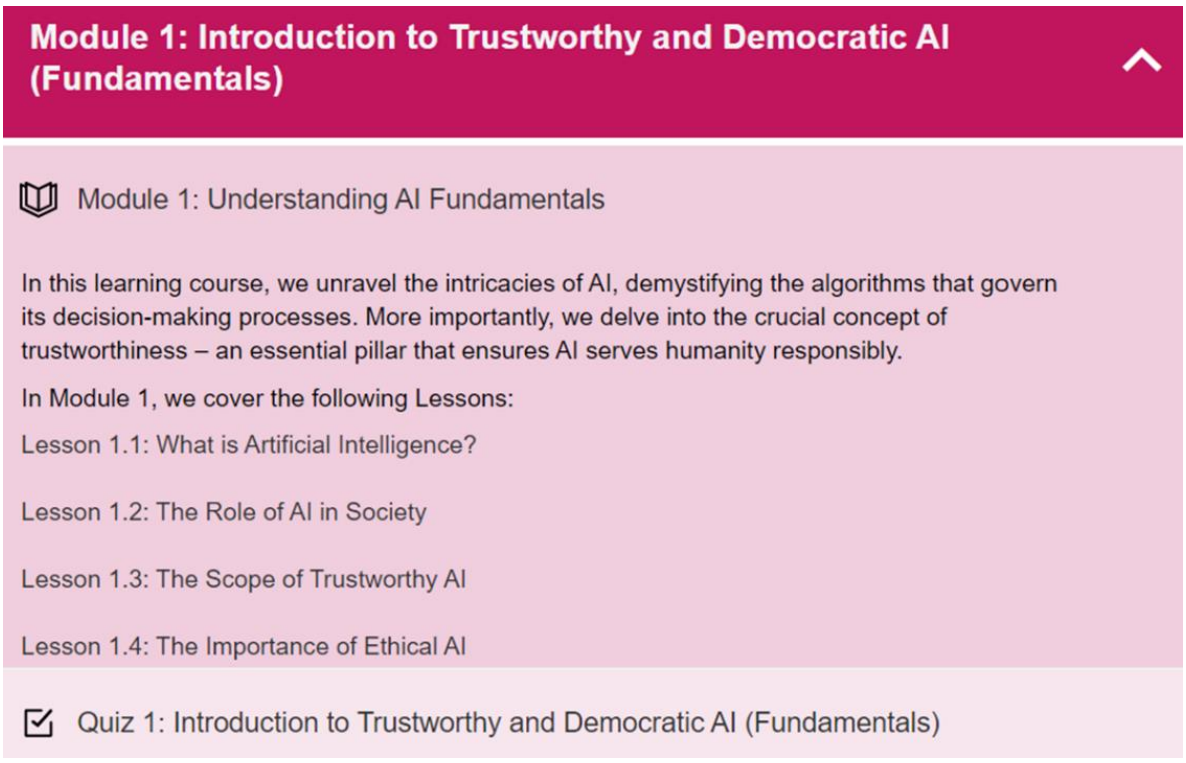

Figure 10: Snapshot from the MOOC1 number four

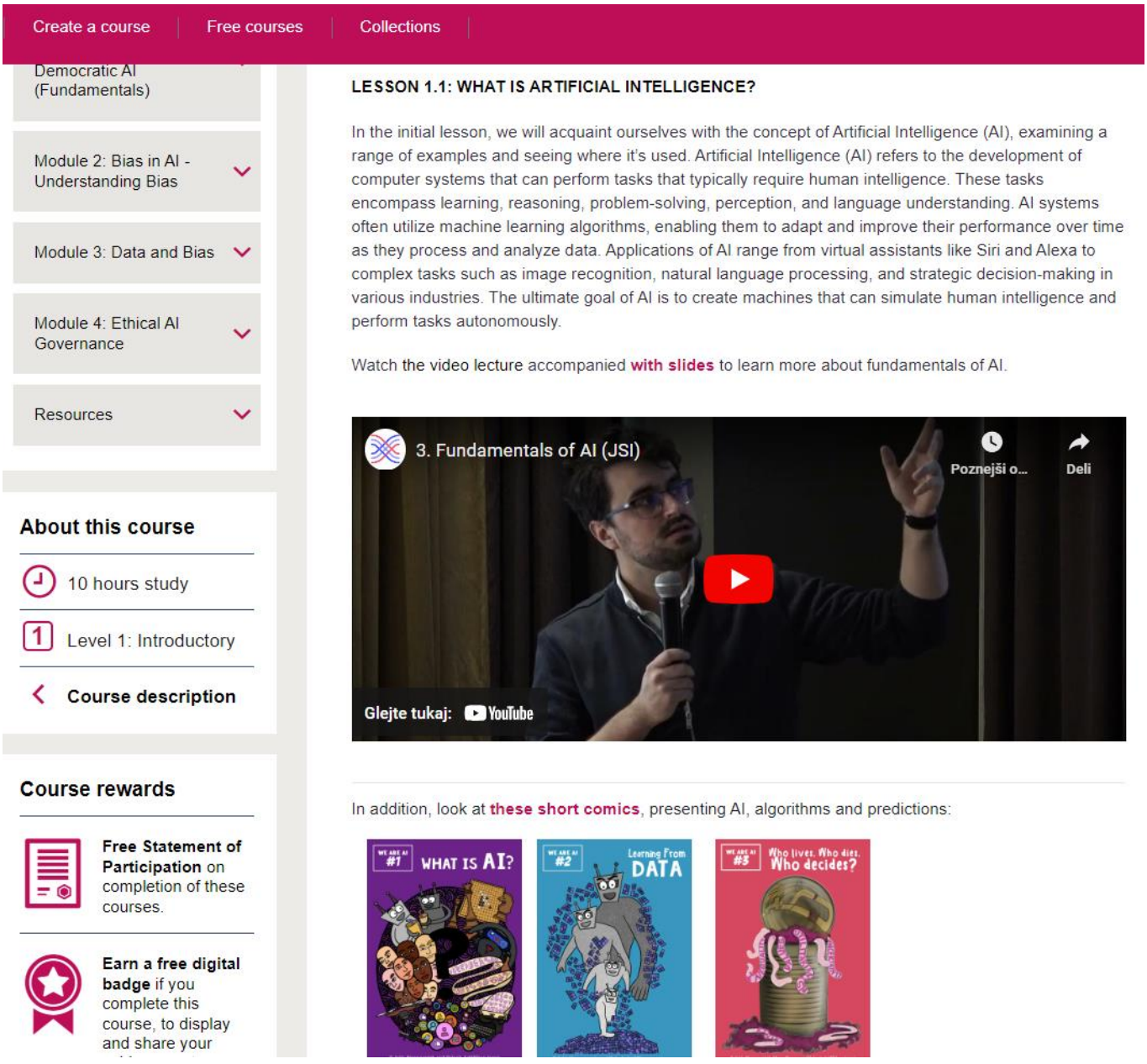

Figure 11: Snapshot from the MOOC1 number five

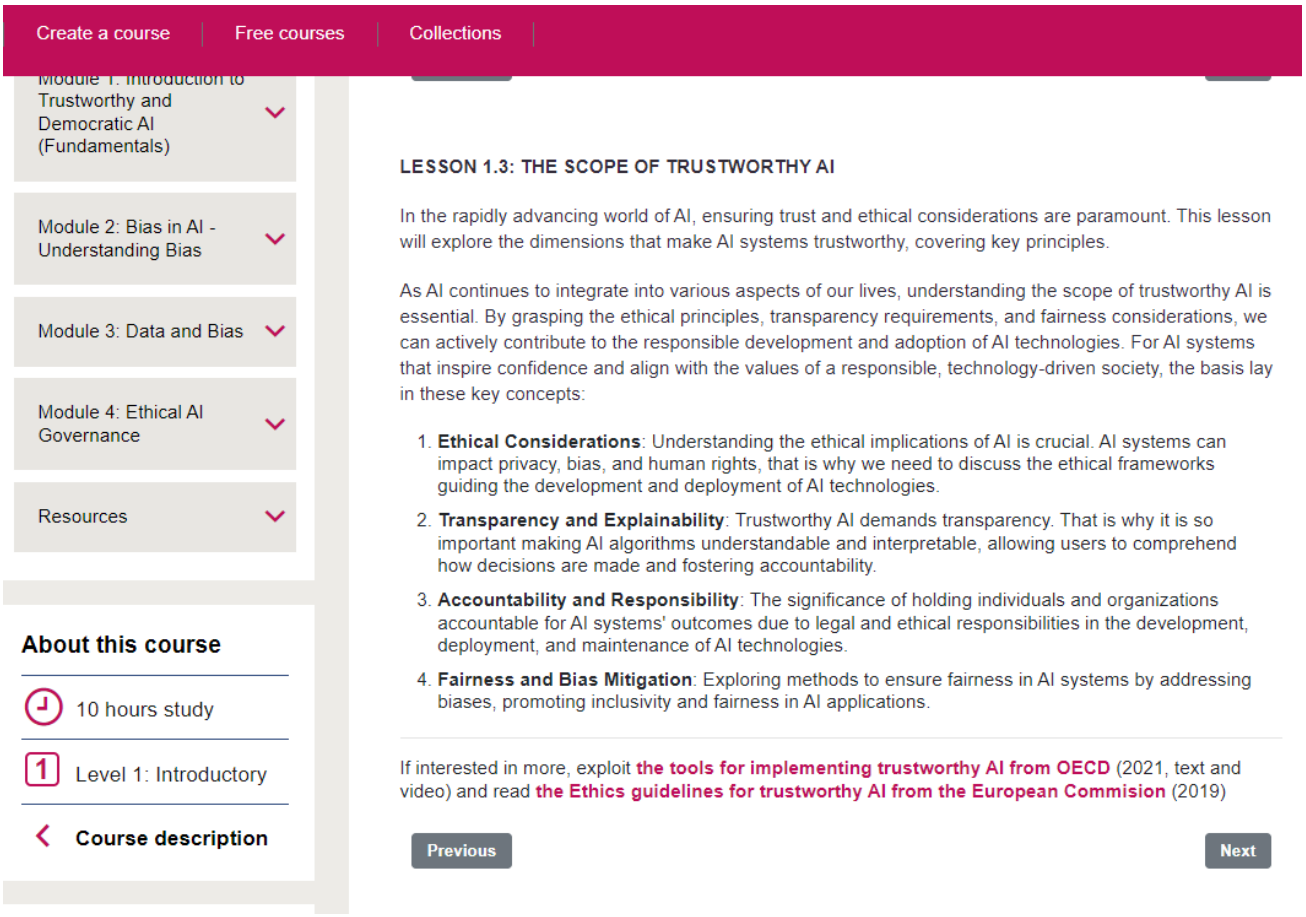

Figure 12: Snapshot from the MOOC1 number six

3.2.8 Further exploitation of MOOCs

Apart from the general specification of stakeholders listed in Chapter 2.1, the AI4Gov partners are constantly searching for new learning audience and collaboration opportunities in relation to WP5. Partner JSI is actively talking with several organisations to exploit the AI4Gov learning and training materials and learning services (MOOCs). This exploitation could be twofold. First, exploitation of existing materials and services as they are. Second, adaptation of the materials and services to make them serve the educational needs of specific groups, for example pupils, students, teachers, regulators, etc.

The exploitation of AI4Gov training and learning services are being discussed with:

1. The Ministry of public administration of Slovenia within their learning and training platform for public administration and public sector employees titled Administration academy ("Upravna akademija" in Slovene language)[11].

[11] Available online https://ua.gov.si/.

2. SAFE.SI[12] - Awareness-raising point on safe use of the internet and mobile devices for children, teenagers, parents and teachers. Project: Safer Internet Centre.
3. Court of Audit of Republic of Slovenia[13] - via international cooperation activities: International Organization of Supreme Audit Institutions (INTOSAI) / European Organisation of Supreme Audit Institutions (EUROSAI) / European Organisation of Regional Audit Institutions (EURORAI).
4. The Academic and Research Network of Slovenia[14] (ARNES), a public institute that provides network services to research, educational and cultural organizations, and enables them to establish connections and cooperation with each other and with related organizations abroad.

Within the active discussions, MOOC1 was already translated into Slovenian language from partner JSI and to Spanish language from partner MAG Spain.

The consortium is being open to further collaboration and exploitation opportunities in relation to learning and training. The activities and achievements in this relation will be reported in the next deliverable (version 2) and also within WP7 deliverables.

[12] https://safe.si/
[13] https://www.rs-rs.si/en/
[14] https://www.arnes.si/en/

4 Development of (self)assessment tools on ethical and transparent AI

The overarching aim of task 5.4 is to devise, refine, and disseminate tools and protocols that underpin the ethical and transparent development and utilization of AI within governance systems. The impetus for these tools is to provide a robust framework that ensures adherence to ethical standards and technical integrity, thereby circumventing potential deficits in transparency, inclusivity, and fairness.

4.1 Framework and set of tools for the adoption of practices that enable responsible algorithm development

The development of a robust framework and a comprehensive set of tools for the adoption of ethical and transparent practices in AI is crucial. This framework aims to ensure responsible algorithm development, fostering trust and accountability in AI systems. The following subsections detail the components of this framework, including self-assessment tools, best practice guides, and various protocols, guidelines, and checklists.

4.1.1 Set of (self)assessment tools

A set of assessment tools to be used, in alignment with widely adopted tools such as various Human Rights Impact Assessments (HRIAs) for AI and ALTAI, by public authorities and use case partners for ensuring that standard checks and balances are in place and are utilized in a more effective, transparent, and accountable way.

The self-assessment tools are:

- **Questionnaires**: Structured questionnaires that prompt detailed responses on ethical, legal, and technical aspects.
- **Evaluation Checklists**: Comprehensive checklists to ensure all critical areas are covered during the assessment process.
- **Scoring and Reporting Dashboards**: Interactive dashboards that provide real-time scoring and generate detailed reports based on the assessment outcomes.

The above tools should encompass the following features:

- **Multi-Dimensional Assessment Criteria**: Incorporating a wide range of ethical, legal, and technical criteria to ensure a holistic evaluation. These criteria align with international best practices and standards, making them applicable across various jurisdictions.
- **Granular Scoring System**: Utilizing a 7-point Likert scale for responses to allow a detailed assessment of compliance and ethical alignment, from complete non-compliance to full compliance.
- **Dynamic Questionnaire Framework**: The assessment is structured around a dynamic questionnaire that adapts based on previous responses, ensuring relevance and focus on specific areas of concern.

- **Comprehensive Requirement Coverage**: Covering a broad spectrum of requirements, including transparency, accountability, fairness, non-discrimination, privacy protection, and security.
- **Automated Score Generation**: Providing a cumulative score that reflects the AI system's overall ethical alignment, facilitating easy evaluation.
- **Actionable Feedback and Recommendations**: Offering detailed feedback and actionable recommendations tailored to the assessment results, providing specific steps for improvement.
- **Continuous Improvement and Update Mechanism**: Ensuring the assessment criteria and recommendations remain current and relevant through regular updates.
- **Guidance Documents**: Detailed documents that provide step-by-step guidance on using the tools effectively.
- **Training Modules**: Online and offline training modules to educate stakeholders on the ethical considerations and proper use of assessment tools.

4.1.2 Best practice guide, covering ethical and technical aspects of AI development processes

The best practice guide is designed to assist developers and policymakers (within and beyond the UC partners) in the ethical and technical development of AI systems. This guide covers:

- **Ethical Considerations:** Detailed guidelines on incorporating ethical principles into AI development, including fairness, accountability, transparency, and privacy.
- **Technical Standards:** Providing technical standards and best practices to ensure the robustness and safety of AI systems, including data management, model development, and system integration.
- **Regulatory Compliance:** Offering guidance on complying with relevant regulations and standards, such as GDPR and AI-specific laws.
- **Stakeholder Engagement:** Emphasizing the importance of involving stakeholders, including end-users and affected communities, in the AI development process to ensure diverse perspectives and needs are addressed.
- **Lifecycle Approach:** Covering the entire AI development lifecycle, from planning and design to deployment and monitoring, ensuring continuous ethical compliance and improvement.

This guide serves as a comprehensive resource for ensuring that AI development processes are ethically sound and technically robust, promoting responsible AI usage in governance.

4.1.3 Components of the Best Practice Guide

A set of tools, including protocols, guidelines and checklists facilitating the identification, reporting, and mitigation of possible shortcomings and risks of the AI model has been designed. These tools are crucial for maintaining the ethical and technical integrity of AI systems throughout their lifecycle.

4.1.3.1 Protocols

- **Risk Identification Protocols**: These protocols provide a structured approach for identifying potential risks at different stages of AI development and deployment. They include steps for conducting risk assessments, identifying biases, and evaluating ethical concerns.
- **Incident Reporting Protocols**: Protocols for reporting any ethical or technical incidents that occur during the AI system's operation. These protocols ensure that incidents are documented, reported, and addressed in a timely manner.
- **Mitigation Protocols**: Detailed procedures for mitigating identified risks, including strategies for bias reduction, enhancing transparency, and improving data security.

4.1.3.2 Guidelines

- **Ethical Development Guidelines**: Comprehensive guidelines on incorporating ethical principles into every phase of AI development. These guidelines cover data sourcing, model design, and system integration, ensuring that ethical considerations are integral to the development process.
- **Technical Development Guidelines**: Guidelines that focus on the technical aspects of AI development, including best practices for data management, algorithm design, and system architecture to ensure robustness and security.
- **User Interaction Guidelines**: Guidelines on designing user interfaces that promote transparency and allow users to understand how AI decisions are made. These guidelines ensure that AI systems are user-friendly and accessible.

4.1.3.3 Checklists

- **Development Phase Checklists:** Detailed checklists for each phase of AI development, from initial planning and design to final deployment. These checklists ensure that all ethical and technical considerations are addressed at each stage.
- **Compliance Checklists**: Checklists to ensure compliance with relevant regulations and ethical standards. These checklists help organizations verify that their AI systems meet all necessary legal and ethical requirements.
- **Continuous Monitoring Checklists**: Checklists for the continuous monitoring of AI systems to ensure ongoing compliance with ethical standards and to identify any emerging issues.

4.1.3.4 Risk Mitigation Strategies

- **Bias Detection and Reduction**: Strategies for detecting and reducing bias in AI systems, including techniques for fair data sampling, algorithm adjustments, and post-processing methods to ensure fairness in AI outcomes.
- **Transparency Enhancement**: Strategies to enhance the transparency of AI systems, including documentation practices, explainability features, and communication protocols that make AI decision-making processes more understandable to users and stakeholders.

- **Security Measures**: Comprehensive security measures to protect data and AI models from unauthorized access and potential breaches. These measures include data encryption, secure data storage, and access control mechanisms.

4.1.3.5 Monitoring and Reporting Mechanisms

- **Continuous Monitoring Tools**: Tools for the continuous monitoring of AI systems to ensure they remain compliant with ethical standards and perform as intended. These tools include automated monitoring systems that track performance and flag potential issues.
- **Reporting Mechanisms**: Mechanisms for reporting any ethical or technical issues that arise during the operation of AI systems. These mechanisms ensure that issues are promptly addressed and resolved, maintaining the integrity of the AI system.
- **Feedback Loops:** Structured feedback loops that incorporate user and stakeholder feedback into the continuous improvement of AI systems. These feedback mechanisms ensure that the AI system evolves based on real-world use and ethical considerations.

By providing these comprehensive tools, organizations can effectively manage the ethical and technical risks associated with AI systems. These tools ensure that AI systems are developed and deployed responsibly, fostering trust and accountability in AI governance.

4.1.3.6 Proposed Implementation Plan

In order to carry out the above-mentioned activities, a proposed implementation plan has been designed including five steps from the initial assessment to the evaluation:

- **Initial Assessment:** Conduct an initial assessment using the provided checklists and protocols to identify potential ethical and technical risks in the AI system.
- **Risk Mitigation Planning:** Develop a risk mitigation plan based on the identified risks, incorporating the provided strategies and guidelines.
- **Implementation of Monitoring Tools:** Implement continuous monitoring tools to track the AI system's performance and ensure ongoing compliance with ethical standards.
- **Regular Reviews and Updates:** Conduct regular reviews of the AI system and update the protocols, guidelines, and checklists as necessary to address any new risks or ethical concerns.
- **Stakeholder Involvement:** Engage stakeholders throughout the AI system's lifecycle to gather feedback and ensure that the system meets their needs and ethical expectations.

By following this implementation plan, organizations can ensure that their AI systems are ethically sound and technically robust, promoting responsible AI usage in governance contexts.

4.1.4 Integration of Framework and Tools

Task 5.4 aims to integrate the developed framework and tools into a cohesive system that can be used by public authorities and other stakeholders to ensure ethical and transparent AI development. The framework includes the following components:

1. **Integration of Assessment Tools:** Aligning the self-assessment tools with widely adopted frameworks such as HRIAAI and ALTAI to provide a comprehensive evaluation system.

2. **Implementation of Best Practices**: Utilizing the best practice guide to inform the development and deployment of AI systems, ensuring adherence to ethical and technical standards.
3. **Application of Protocols and Guidelines**: Implementing the protocols, guidelines, and checklists to facilitate effective risk management and ethical compliance throughout the AI lifecycle.
4. **Continuous Improvement and Feedback**: Establishing mechanisms for continuous feedback and improvement, ensuring that the framework and tools remain current and effective in addressing emerging ethical and technical challenges.

4.1.4.1 Specific Steps for Integration

To achieve the above-mentioned plan, a specific path of four steps has been designed:

- **Collaborative Workshops:** Conduct workshops with stakeholders to ensure the tools and frameworks are effectively integrated and understood.
- **Pilot Testing**: Implement pilot testing of the tools in real-world scenarios to gather feedback and make necessary adjustments.
- **Continuous Training**: Provide ongoing training for stakeholders to ensure they are proficient in using the tools and adhering to best practices.
- **Regular Updates:** Regularly update the tools and frameworks to reflect the latest developments in AI ethics and governance.

In conclusion, the development and integration of self-assessment tools, best practice guides, and comprehensive protocols and checklists are essential for ensuring the ethical and transparent deployment of AI systems in governance. These tools provide a structured approach to evaluating and managing the ethical and technical risks associated with AI, fostering trust and accountability.

5 Ethical and organizational guidelines and blueprints for trustworthy AI

Instances of AI systems being developed, funded, and implemented that are not fit for purpose, unethical, unjust, dangerous, or that further institutionalize inequality in society are alarmingly common. The lack of accountability in these cases is a glaring issue, with the question of ultimate accountability and who might be held accountable often left unanswered. As the number of AI systems facing difficulties continues to grow, it becomes increasingly likely that those backing questionable initiatives will face significant problems. This is a particularly pressing concern in the context of funding and adherence to government regulations. Given the potential risks associated with funding AI systems, it is imperative that non-governmental organizations (NGOs), public offices and the private sector take a proactive role in assessing the systems they choose to finance. This is the first and most crucial step in preventing potential failures in compliance with the legal and ethical framework. Task 5.5 aims to deal with this challenge.

A key approach to tackling this issue is to reform how funding organizations manage the application process to ensure compliance with the legal and ethical framework for trustworthy AI. Recently, soft laws, guidelines, and recommendations have gained legal force with the enactment of the AI Act. The AI Act, a pivotal piece of legislation, incorporates the latest advancements in generative AI and serves as a crucial reference point when assessing AI systems and, consequently, financing them. This legal regulation provides a robust and obligatory framework. As is well known and clearly stated in D1.4, the EU aims to position itself as a frontrunner in AI regulation development, which should be human-centric, without impeding innovation and investment. Specifically addressing the latter concern. T5.5 is explicitly designed to support public or private organizations interested in financing AI. T5.5 will offer tools for these organizations to effectively guide applicants in completing their trustworthy AI statement.

At this stage, partner WLC, the leader of T5.5, has assessed the ethical tools available, incorporating the newly approved AI Act and reviewing available guidelines, including those suggested in the AI4Gov first review meeting. The list of reviewed frameworks and guidelines can be seen in Table 8. Also, considering that D5.5, the direct output connected to T5.5, is to be delivered in M27, we include the drafts of one tool, a Stop-and-Think instrument, and a blueprint to assist those funding the development of AI systems based on the AI Act. However, it should be clarified that the final tools, blueprint, and D5.5 will include the frameworks and guidelines reviewed; thus, it will go beyond the legal framework provided in the AI Act.

Table 8: Relevant ethical AI frameworks and guidelines

Relevant Ethical AI Frameworks
AI Act[15]
EC Guidelines for Trustworthy AI[16]
Blueprint for an AI Bill of Rights[17]
Australia's AI Ethics Principles[18] [19]
CA Responsible use of AI[20]
IEEE ethically aligned design[21]
Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal[22] [23]
Relevant guidelines
Center for Inclusive Change, Essential Considerations in AI Contracting[24]
WEF, Guidelines for AI procurement[25]
WEF, AI Procurement in a Box: AI Government Procurement Guidelines[26]
WEF, AI Procurement in a Box: Workbook
UK Office for AI, Guidelines for AI procurement[27] [28]

[15] https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
[16] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[17] https://www.whitehouse.gov/ostp/ai-bill-of-rights/
[18] https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles
[19] https://www.csiro.au/en/research/technology-space/ai/ai-ethics-framework
[20] https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html
[21] https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
[22] https://ai4si.gzs.si/uploads/XbVUzDbn/Artificialintelligencehumanrigthsdemocracyandtheruleoflaw.pdf
[23] https://arxiv.org/ftp/arxiv/papers/2202/2202.02776.pdf
[24] https://www.inclusivechange.org/ai-governance-solutions/ai-contract-clauses
[25] https://www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf
[26] https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf
[27] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/990469/Guidelines_for_AI_procurement.pdf
[28] https://www.linkedin.com/pulse/hitchhikers-guide-through-ai-procurement-guidance-mr-ashley-moore/

5.1 Stop-and-Think: A Guide for Applicants Developing AI Systems

The European Union's AI Act aims to ensure that AI systems developed and deployed within the EU are safe, transparent, and respect fundamental rights. Before applying for funds to build AI systems, it is crucial to understand and comply with these regulations. This guide will help you evaluate your AI project against the critical requirements of the EU AI Act and, consequently, prepare your Trustworthy AI Application Statement.

5.1.1 Key Areas to Consider when preparing your Trustworthy AI Application Statement

Stage 1: Understanding AI Act Risk Classification

Stop: Before preparing your statement, ensure all your group members understand the AI Act Risk Classification.

Think: Read carefully the summary of the AI Act classification and decide according to it.

1. Risk Classification

- ✓ **Understand Risk Levels**: The EU AI Act classifies AI systems into four risk categories: Unacceptable Risk, High Risk, Limited Risk, and Minimal Risk.
- ✓ **Unacceptable Risk**: AI systems that deploy harmful manipulative “subliminal techniques”; AI systems that exploit specific vulnerable groups (physical or mental disability); AI systems used by public authorities or on their behalf, for social scoring purposes, “Real-time” remote biometric identification systems in publicly accessible spaces for law enforcement purposes, except in a limited number of cases.
- ✓ **High Risk**: AI systems that adversely impact people’s safety or fundamental rights. The AI Act differentiates between two categories of high-risk systems. Systems used as a safety component of a product falling under EU health and safety harmonization legislation; systems deployed in eight specific areas detailed in Annex III.
- ✓ **Limited Risk**: AI systems that interacts with humans (i.e. chatbots), emotion recognition systems, biometric categorization systems, and AI systems that generate or manipulate image, audio or video content (i.e. deepfakes), would be subject to a limited set of transparency obligations.
- ✓ **Minimal Risk**: these systems could be developed and used without conforming to any additional requirements.

Stage 2: Identify Your System's Risk Level

Stop: Which classifications mentioned above does your AI system project fall under?

Think: Determine which category your AI system falls into. Remember that high-risk systems include those used in critical infrastructure, education, employment, essential public services, law enforcement, and migration, among others.

Stage 3: Application that includes a High-Risk AI system

Stop: Revise the High-Risk AI System Requirements of the AI Act

Think: If your AI system is classified as high-risk, ensure it complies with the following requirements:

- ✓ **Risk Management System**: Implement a risk management system to identify, assess, and mitigate risks.
- ✓ **Data Governance**: Ensure the quality and integrity of the data used. This includes proper data collection, annotation, and handling procedures.
- ✓ **Technical Documentation**: Maintain comprehensive technical documentation detailing the system's purpose, design, development, testing, and deployment.
- ✓ **Transparency and Information Provision**: Provide clear information to users about the system's capabilities and limitations.
- ✓ **Human Oversight**: Design mechanisms that allow human oversight and intervention when necessary.
- ✓ **Robustness, Accuracy, and Security**: Ensure your system is resilient, accurate, and secure against potential threats.

Stage 4: Ethical Considerations

Stop: Adhering to ethical principles is key to complying with guidelines and the AI Act. Ensure your team discusses the potential ethical challenges.

Think: Revise the following key principles when developing your project and AI Trustworthy Statement.

- ✓ **Fairness and Non-Discrimination**: Design your AI system to avoid bias and discrimination. Implement measures to detect and mitigate any potential bias in data and algorithms.
- ✓ **Privacy and Data Protection**: Comply with GDPR and other relevant privacy regulations. Ensure the AI system does not infringe on individuals' privacy rights.
- ✓ **Accountability**: Establish clear accountability for the AI system's decisions and actions. Ensure there are processes in place for redress and remedy in case of harm or misuse.

Stage 5: Transparency and User Awareness

Stop: Before diving into full development of the AI System, make sure transparency and user awareness will be considered and AI Act standards will be followed, regardless of the risk classification.

Think: Does the project include the following critical points?

- ✓ **Clear Communication**: Inform users when they are interacting with an AI system. Provide understandable information about how the AI system makes decisions.
- ✓ **Documentation for Users**: Offer comprehensive documentation and user guides that explain the AI system's functionality, limitations, and correct usage.

Stage 6: Sustainability and Societal Impact

Stop: Before delving deeper into the technical aspects of the AI system development, stop and reflect on the sustainability and societal impact of the proposed project.

Think: Consider the following environmental challenges and discrimination problems that may arise from your project.

- ✓ **Environmental Impact**: Consider the environmental impact of developing and deploying your AI system. Aim for energy-efficient algorithms and sustainable practices.
- ✓ **Social Impact**: Evaluate the broader societal implications of your AI system. Ensure it contributes positively to society and does not reinforce existing inequalities or create new ones.

Final Consideration: Application Checklist

Before submitting your funding application, ensure you have addressed the following:

1. **Risk Assessment:** Have you classified your AI system's risk level?
2. **Compliance with High-Risk Requirements**: If applicable, have you met all the high-risk AI system requirements?
3. **Ethical Considerations:** Have you implemented measures to ensure fairness, privacy, and accountability?
4. **Transparency Measures**: Are you providing clear information and documentation to users?
5. **Impact Assessment:** Have you evaluated and mitigated the environmental and societal impacts of your AI system?

5.2 Blueprint for Developing AI Systems According to the EU AI Act

This blueprint is a comprehensive guide for applicants seeking funding to develop AI systems complying with the EU AI Act. It covers the essential steps and considerations to ensure conformity with regulations, helping you prepare a robust and compelling funding application.

Step 1: Preliminary Assessment

1.1 Understand the EU AI Act

Familiarize yourself with the EU AI Act: Understand its objectives, scope, and requirements.

Identify critical provisions relevant to your project: Focus on risk classification, ethical guidelines, and specific requirements for high-risk AI systems.

1.2 Define Your AI System

Describe the AI system: Clearly outline the purpose, functionality, and intended use of your AI system.

Determine the risk classification: Assess whether your AI system falls under unacceptable risk, high risk, limited risk, or minimal risk.

Step 2: Compliance with High-Risk AI System Requirements

2.1 Risk Management

Implement a risk management system: Develop processes to identify, assess, and mitigate potential risks associated with your AI system.

2.2 Data Governance

Ensure data quality and integrity: Establish protocols for data collection, annotation, storage, and processing to maintain high standards.

Comply with data protection regulations: Adhere to GDPR and other relevant data privacy laws.

2.3 Technical Documentation

Create comprehensive documentation: Include details on the system's design, development, testing, deployment, and maintenance.

Document data sources and methodologies: Ensure transparency in data handling and algorithm development.

2.4 Transparency and Information Provision

Provide user information: Clearly explain the capabilities, limitations, and intended use of the AI system.

Ensure transparency: Make information about the AI system's decision-making processes available to users.

2.5 Human Oversight

Design for human oversight: Incorporate human intervention and oversight mechanisms in the AI system's operations.

2.6 Robustness, Accuracy, and Security

Ensure system robustness: Design the AI system to be reliable and resilient against failures and attacks.

Focus on accuracy: Implement measures to maintain high accuracy and performance.

Prioritize security: Protect the AI system from unauthorized access and cyber threats.

Step 3: Ethical Considerations

3.1 Fairness and Non-Discrimination

Implement bias mitigation strategies: Use techniques to identify and reduce bias in data and algorithms.

Promote fairness: Ensure the AI system does not discriminate against individuals or groups.

3.2 Privacy and Data Protection

Safeguard personal data: Implement strong privacy measures to protect users' data.

Follow GDPR guidelines: Ensure compliance with the General Data Protection Regulation.

3.3 Accountability

Establish clear accountability: Define who is responsible for the AI system's decisions and actions.

Create redress mechanisms: Provide ways for users to report issues and seek remedies.

Step 4: Transparency and User Awareness

4.1 Clear Communication

Inform users about AI usage: Make it clear when users are interacting with an AI system.

Explain decision-making: Provide understandable information about how the AI system makes decisions.

4.2 User Documentation

Offer comprehensive guides: Provide user manuals and documentation that explain the system's functionality and limitations.

Step 5: Sustainability and Societal Impact

5.1 Environmental Impact

Minimize environmental footprint: Develop energy-efficient algorithms and use sustainable practices.

5.2 Social Impact

Assess societal implications: Evaluate your AI system's potential social impact and ensure it contributes positively to society.

Avoid reinforcing inequalities: Design the AI system to promote social good and avoid exacerbating existing inequalities.

Step 6: Funding Application Preparation

6.1 Gather Documentation

Compile all necessary documents: Ensure you have all the required technical, ethical, and compliance documentation ready.

Prepare a detailed project plan: Include timelines, milestones, and deliverables.

6.2 Write the Application

Clearly articulate your project's objectives: Explain how your AI system aligns with the goals of the EU AI Act.

Highlight compliance efforts: Demonstrate how you have addressed the Act's requirements and ethical considerations.

Detail the expected impact: Outline the potential benefits and positive societal impact of your AI system.

6.3 Review and Submit

Review for completeness and accuracy: Ensure your application is thorough and error-free.

Submit to the relevant funding body: Follow the specific submission guidelines the funding organization provides.

6 Conclusions

This deliverable provides an overview of the work done in WP5, in more detail for T5.2, T 5.3, T5.4 and also partially for T5.5 in the period of first 18 months. It covers the topics of AI4Gov trainings, educational resources, MOOCs, (self)assessment tools on ethical and transparent AI and also ethical and organizational guidelines and blueprints for trustworthy AI.

In the context of the AI4Gov trainings and educational resources, organized trainings were presented along with the future plans. Moreover, educational resources were explained and presented. For the MOOCs, first the methodology was introduced and afterwards the curriculum was presented in detail. Along with the introduced plan of the AI4Gov MOOCs, also the first MOOC that is already available for learners was presented.

In relation to the (self)assessment tools on ethical and transparent AI, the framework and set of tools for the adoption of practices that enable responsible algorithm development were presented. As we see some correlations of this topic to the topic of ethical and organizational guidelines and blueprints for trustworthy AI, the last chapter is presenting a list of relevant reviewed frameworks and guidelines. Also, a draft of one tool is added (Stop-and-Think instrument) and a blueprint to assist those funding the development of AI systems based on the AI Act.

The progress of the targeted tasks is clearly shown and the work is being continued as planned.

7 References

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Eilks, I., & Hofstein, A. (2017). Curriculum Development in Science Education. In K. S. Taber & B. Akpan (Eds.), *Science Education* (pp. 169–181). SensePublishers. https://doi.org/10.1007/978-94-6300-749-8_13

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, *363*(6423), 130–131. https://doi.org/10.1126/science.aav7958

Shah, V., Murthy, S., Warriem, J., Sahasrabudhe, S., Banerjee, G., & Iyer, S. (2022). Learner-centric MOOC model: A pedagogical design model towards active learner participation and higher completion rates. *Educational Technology Research and Development*, *70*(1), 263–288. https://doi.org/10.1007/s11423-022-10081-4

UNESCO. (n.d.). *Open Educational Resources*. Retrieved June 10, 2024, from https://www.unesco.org/en/open-educational-resources