Learning framework/curriculum for: **TRUSTWORTHY AND DEMOCRATIC AI**

Objective: The objective of the "Trustworthy and Democratic AI" learning framework is to equip learners with the knowledge, skills, and ethical principles necessary to design, develop, and deploy artificial intelligence systems that are both trustworthy and free from bias. 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.

Specific Learning Objectives:

Understanding AI Fundamentals:

- Understand the concepts in artificial intelligence, machine learning, and deep learning.
- Recognize the transformative potential and ethical implications of AI in various domains.

Bias Awareness and Recognition:

- Identify the types and sources of bias that can manifest in AI systems.
- Understand the real-world consequences of biased AI, including social and ethical implications.

Ethical Considerations in AI:

- Explore ethical frameworks and principles guiding AI development.
- Recognize the importance of fairness, transparency, and accountability in AI systems.

Data Collection and Preprocessing:

- Learn how data collection and preprocessing can introduce bias into AI models.
- Implement best practices for collecting, cleaning, and preparing data to reduce bias.

Algorithmic Bias Mitigation:

- Understand various techniques to mitigate algorithmic bias in AI models.
- Apply debiasing methods, reweighting strategies, and fairness constraints to model development.

Explainable/Interpretable AI:

- Examine methods for making AI models more explainable.
- Appreciate the importance of transparency in AI decision-making processes.

Ethical AI Governance:

- Study legal and regulatory frameworks governing AI, such as GDPR and AI Act.
- Explore the role of responsible AI governance in organizations.

Responsible AI Development:

- Apply ethical guidelines and responsible AI principles to the end-to-end AI development lifecycle.
- Develop AI systems that prioritize fairness, accountability, and transparency.

AI Trustworthiness Evaluation:

- Assess AI models and systems for fairness, bias, and trustworthiness.
- Implement evaluation techniques to ensure AI systems meet ethical and operational standards.

Real-World Applications:

- Analyze case studies and real-world examples of trustworthy AI.
- Gain practical experience in addressing bias and trustworthiness challenges in AI projects.

By the end of this learning framework, participants 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. The framework empowers learners to become advocates for trustworthy and democratic AI in their respective roles and organizations.

learning framework broken down into modules and lessons:

MODULES, LESSONS:

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 (The Ethical Imperative; AI Ethics Frameworks; Case Studies in Ethical Lapses)

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)

Module 3: Ethical Considerations
Lesson 3.1: Ethical Principles in AI
Lesson 3.2: Fairness, Transparency and Accountability - Ethical Decision Frameworks

Module 4: Data and Bias
Lesson 4.1: Bias in Data Collection
Lesson 4.2: Data Sampling Methods
Lesson 4.3: Ethical Data Sourcing
Lesson 4.4: Data Preprocessing and Bias Reduction (Data Cleaning for Bias; Fair Data Preprocessing Techniques)
Lesson 4.5: Real-world Data Bias Case Studies

Module 5: Algorithmic 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: Interpretable AI
Lesson 6.1: Explainable AI Approaches (XAI)

Module 7: Ethical AI Governance
Lesson 7.1: Legal and Regulatory Frameworks
- Overview of AI Regulations
- GDPR and Data Privacy

- National and International AI Policies

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: AI Trustworthiness Evaluation
Lesson 9.1: Evaluation Metrics for Trustworthy AI
- Metrics for Fairness and Bias
- Trustworthiness Assessment
- Evaluation Tools and Frameworks

Lesson 9.2: Ethical Auditing and Reporting
- Ethical Auditing Processes
- Reporting and Transparency
- Corrective Actions and Reevaluation

Module 10: Real-World Applications and Case Studies
Lesson 10.1: Case Studies in Trustworthy AI
- Analysis of Real-World Examples
- Success Stories in Ethical AI
- Addressing Bias and Trustworthiness Challenges

This structured learning framework provides a comprehensive education in creating and maintaining 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.

Competencies gained with the Trustworthy and democratic AI learning framework

COMPETENCIES, general: The "Trustworthy and Democratic AI" learning framework is designed to enhance a range of competencies and skills in individuals who complete the program. These competencies are essential for those who aim to develop, deploy, or manage AI systems while ensuring they are ethical, democratic, and trustworthy. 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 make AI models more transparent and understandable.

- Skill in generating explanations for AI model outputs (XAI).

Data Management and Privacy:

- Competence in handling data ethically and ensuring privacy compliance.
- Skill in data preprocessing 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:

- Ability to work collaboratively in multidisciplinary teams 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.

Digital competencies: The "Trustworthy and Democratic AI" learning framework 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 by the framework:

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:

- Proficiency in designing and implementing algorithms for AI systems.
- Ability to create and optimize AI models.

Interpretable AI and Explainability:

- Capability to develop AI models that provide human-understandable explanations.
- Skill in using tools and methods for model interpretability (XAI).

Digital Privacy and Security:

- Understanding of data privacy and security best practices.
- Proficiency in securing AI systems and protecting data.

Coding and Programming Skills:

- Competence in programming languages commonly used in AI, such as Python.
- Skill in writing, debugging, and 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:

- Proficiency in managing AI projects, including planning, execution, and evaluation.
- Skill in using digital project management tools.

Auditing and Evaluation Tools:

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

Information Literacy and Research Skills:

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

Digital Communication Skills:

- Capability to effectively communicate and collaborate digitally, especially in multidisciplinary teams.
- 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.

This will be achieved via WP5. Building blocks: training workshops (T5.2), MOOCs (T5.3), panel discussions (T.51), (self)assessment tools on ethical and transparent AI (T5.4), ethical and organizational guidelines and blueprints for trustworthy AI (T5.5).