



**Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization**  
**TAILOR**

**Grant Agreement Number 952215**  
**PhD Curriculum Report**

<b>Document type (nature)</b>	Report
<b>Deliverable No</b>	D9.6
<b>Work package number(s)</b>	WP9
<b>Date</b>	Due
<b>Responsible Beneficiary</b>	UNIVBRIS, ID #16
<b>Author(s)</b>	Peter Flach and Miquel Perello Nieto
<b>Publicity level</b>	Public
<b>Short description</b>	This deliverable proposes a PhD curriculum in Trustworthy AI: a specification of the structure and content of a PhD programme that could be delivered by (consortia of) academic institutions.

<b>History</b>			
<b>Revision</b>	<b>Date</b>	<b>Modification</b>	<b>Author</b>
1.0	11/04/2023	First version	PF, MPN

<b>Document Review</b>		
<b>Reviewer</b>	<b>Partner ID / Acronym</b>	<b>Date of report approval</b>
Fredrik Heintz	#1 / LIU	April 12, 2023
Barry O'Sullivan	#4 / UCC	April 13, 2023

*This document is a public report. However, the information herein is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.*

---

## Table of Contents

<b>Summary of the report</b>	<b>2</b>
<b>Organisation</b>	<b>3</b>
<b>1. Introduction to the deliverable</b>	<b>3</b>
<b>2. PhD Curriculum in AI for the 21st-century</b>	<b>4</b>
<b>3. Methodology</b>	<b>5</b>
<b>4. TAILOR PhD curriculum in Trustworthy AI</b>	<b>6</b>
4.0. Foundations of Artificial Intelligence	7
4.1. Foundations of Trustworthy AI	7
4.2. AI Paradigms and Representations	8
4.3. Deciding and Learning How to Act	9
4.4. Reasoning and Learning in Social Contexts	10
4.5. Automated AI	11
<b>5. Relationship with the AIDA PhD curriculum</b>	<b>13</b>
<b>6. Concluding remarks</b>	<b>14</b>
<b>References</b>	<b>15</b>
<b>Appendix A. Glossary of terms</b>	<b>16</b>
<b>Appendix B. Badges and consortium-based programmes</b>	<b>18</b>
References	21

## Summary of the report

This deliverable proposes a PhD curriculum in Trustworthy AI: a specification of the structure and content of a PhD programme that could be delivered by (consortia of) academic institutions. The topics of the curriculum are closely aligned with the TAILOR scientific work packages. Work on the TAILOR PhD curriculum was coordinated with a collaboration with the four ICT-48 AI networks within the VISION Coordination and Support Action, taking place under the auspices of the International AI Doctoral Academy (AIDA) to define a more general PhD curriculum on Artificial Intelligence.

This deliverable report is structured as follows. [Section 1](#) gives a brief introduction, and [Section 2](#) discusses a modern perspective on the shape and role of PhD curricula in general. [Section 3](#) details the methodology followed in developing the TAILOR PhD curriculum, which is stated in detail in [Section 4](#). [Section 5](#) describes how the TAILOR PhD curriculum in Trustworthy AI aligns with and contributes to the AIDA PhD curriculum in AI, and [Section 6](#) concludes. Two appendices give a glossary of terms ([Appendix A](#)) and a discussion how badges could facilitate delivery and certification in practice ([Appendix B](#)).

## Organisation

The following people have been involved in the Deliverable:

Partner ID / Acronym	Name	Role
ID #16, UNIBRIS	Miquel Perello Nieto	Researcher
ID #16, UNIBRIS	Peter Flach	WP Lead
ID #4, UCC	Barry O'Sullivan	VISION and AIDA Curriculum Lead
ID #2, CNR	Umberto Straccia	WP3 Lead
ID #5, KUL	Luc De Raedt	WP4 Lead
ID #6, UOR	Giuseppe De Giacomo	WP5 Lead
ID #8, IST-UL	Ana Paiva	WP6 Lead
ID #7, LEU	Holger Hoos	WP7 Lead

## 1. Introduction to the deliverable

The TAILOR network comprises most of the research centres of excellence in Artificial Intelligence in Europe. Of the 54 partner laboratories of TAILOR, 44 are related to higher education institutions while 10 of them are industrial partners that build important and necessary synergies with real-world applications. This means that we have the best programs of higher education about Artificial Intelligence in Europe and beyond. This includes our involvement in master's and doctoral programs, workshops, conferences and other notable events. However, the rapid growth of the field of AI requires quick adaptation of existing curricula. In TAILOR, we have world-leading expertise on Trustworthy AI, which we have built on to define a **model curriculum** for a PhD in this topic. Here, we understand *curriculum* as the specification of the structure and content of a PhD programme, which is distinct from actual delivery.

This deliverable complements Deliverable 9.5 Mapping of AI-oriented PhD programmes at TAILOR partners by proposing a TAILOR PhD curriculum on Trustworthy AI. Deliverable 9.5 was developed as a bottom-up mapping of all the available material and courses from BSc, MSc and PhD programmes currently being delivered by TAILOR partners. On the other hand, this Deliverable 9.6 follows a top-down approach by defining a model curriculum on Trustworthy AI based on essentials. The proposed curriculum can later materialise into a PhD program by an accredited institution of higher education or a consortium. The previous mapping will potentially be valuable to match the proposed curriculum with the available resources.

---

Based on the expertise within the TAILOR consortium we have designed a PhD curriculum in Trustworthy AI which is described in [Section 4](#). This has been possible thanks to the collaboration of the work package leaders in the TAILOR network, as well as the joint forces of the four ICT-48 networks and the VISION consortium which led to the creation of the Artificial Intelligence Doctoral Academy (AIDA). This collaboration and the methodology followed for the elaboration of this curriculum are explained in [Section 3](#). Our contribution also helped to define the AIDA PhD curriculum in AI, as it is complementary to the curriculum presented in this deliverable. [Section 5](#) contains details about the synergies between the two curricula. We conclude the delivery with the opportunities and difficulties that a cross-network curriculum entails and the future directions opened in [Section 6](#). Additionally, [Appendix A](#) contains a glossary of terms that have been crucial to communicate about the curriculum without misunderstandings. Finally, we provide an initial exploration of how the curriculum could be reused in multiple PhD programs by the use of graphical representations of necessary competencies and the acquisition of badges (see [Appendix B](#)).

## 2. PhD Curriculum in AI for the 21st-century

Doctorate education dates back to medieval Europe in the 13th century (Park, 2005 and 2007). Originally a doctorate was awarded as a license in order to teach in Universities, although it did not require an innovative research contribution (Park, 2005). Since then, it has evolved over the years with the first doctorate programme in research in Germany when Humboldt founded the University of Berlin in 1810 (Park, 2005; Wyatt, 1998). It required innovative research, attendance to seminars, production of an acceptable thesis, and an oral examination (Goodchild and Miller, 1997). This doctorate was followed by the USA in 1861 at Yale, followed by Harvard, Michigan and Pennsylvania. The UK introduced higher doctorates (DSc and DLitt) in the 1870s at the University of London, Edinburgh, Oxford and Cambridge. The lower doctorate (PhD) was spread to Britain in 1917 and later to other English-speaking countries (Simpson, 1983). Nowadays, there are multiple doctorate education degrees: Traditional PhD, PhD by publication, Professional Doctorate (EdD), Practice-based doctorate, higher doctorates (DSc and DLitt; Barnes, 2013), New Route PhD and industrial focus doctorates (DEng) (Park, 2005 & 2007; Gould, 2015). All of these have similarities but also different requirements, which have evolved over the centuries adapting to the purpose and needs of the work placement expected after the award. And even if the PhD requirements have been standardised in Europe, European countries have different implementation requirements (e.g. the viva has a different role) (Gould, 2016).

Nowadays, the number of awarded PhDs is growing faster than the required academic and R&D work positions, which means that more than half of the awardees end up in government and non-government organisations, businesses and industry (Gould, 2015; Sharmini & Spronken-Smith, 2020). For that reason, it is being questioned if the PhD curricula should be adapted to the current work expectations (Park, 2005 & 2007; Gould, 2015 & 2016; Sharmini & Spronken-Smith, 2020; Coates et al. 2020; Sarrico 2022).

---

Because Trustworthy AI in Europe is a field inherently connected to multiple stakeholders, we propose to design a curriculum that integrates transferable skills that would be required for communication and synthesis of ideas in written and oral form, collaboration between multidisciplinary and multicultural teams, and manage individual projects. We have been inspired by the concept of a *doctoral architecture* to help the development of the student in all the phases of the PhD, to ensure that their expectations are aligned with the outcome of the award. Following Coates et al. (2020) doctoral architecture which is organised in the next steps: (1) preparations: awareness, foundation, onboarding, and application; (2) experiences: research, development; (3) successes: personal, academic, professional. Finally, any program based on this curriculum should provide answers to the following questions (Sarrico, 2022): (a) how do the graduates contribute to society once they obtain the degree? (b) what do the candidates know about their job prospects before starting? (c) how does the job in which they end up working relate to the skills obtained during the degree? (d) how to promote diversity? Every student that wishes to enrol in such a doctorate should have access to those answers which should reduce the dropout rate.

We have designed a European curriculum in Trustworthy AI that is aligned with the requirements expected for a doctorate in information and communications technologies in the 21st century. We encourage the transparency of the curriculum by embedding preparation, experience and success as an essential part of it. And a required set of transversal skills that should be accredited as a portfolio in order to be awarded the doctorate.

### 3. Methodology

The design of this curriculum has been aligned with a collaboration with the four ICT-48 AI networks (AI4Media, ELISE, HumanE-AI NET, and TAILOR) within the VISION Coordination and Support Action. This collaboration took place under the auspices of the International AI Doctoral Academy (AIDA) to define a world-level PhD curriculum on Artificial Intelligence. It is hence important to distinguish two different curricula:

- The TAILOR PhD curriculum on Trustworthy AI as described in this deliverable;
- The AIDA PhD curriculum on AI which includes elements of the TAILOR curriculum as described below.

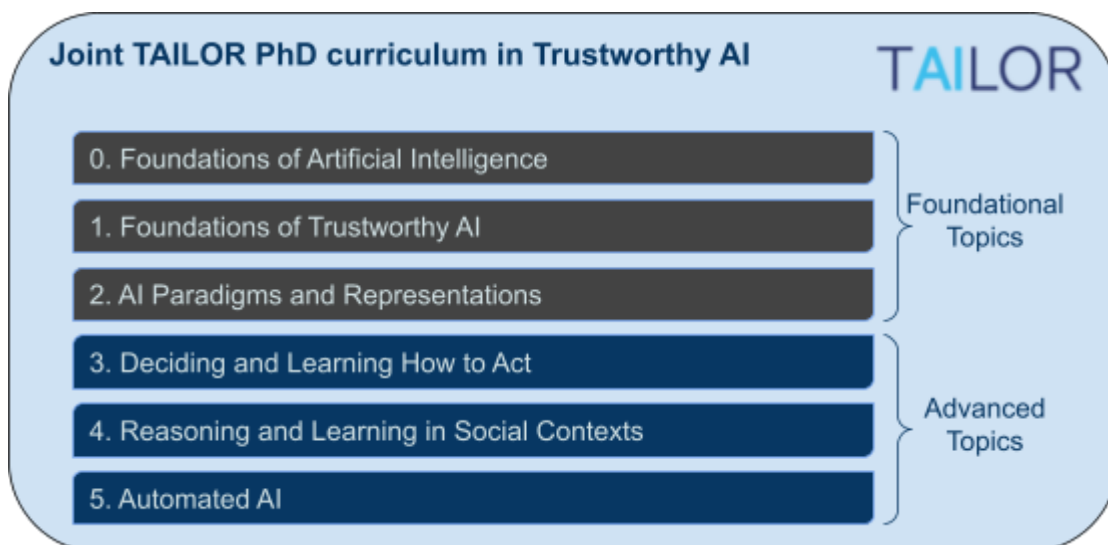
The structure of the AIDA curriculum is based on a set of interdisciplinary core topics and special topics which have been defined by the different networks. In TAILOR WP9 we first defined a **glossary of terms** which has been crucial for the proper communication between the network representatives (see [Appendix A](#)). Then, we designed a TAILOR PhD curriculum on Trustworthy AI based on the main topics covered by the work packages Trustworthy AI (WP3), Paradigms and Representations (WP4), Acting (WP5), Social AI (WP6), and Auto AI (WP7). Then, representatives of each work package contributed to defining the learning outcomes, content, knowledge, methodological skills, transferable skills, applications and courses available. The resulting TAILOR PhD curriculum is presented in [Section 4](#). Once the four ICT-48 networks defined their specific PhD curricula, we determined the interdisciplinary

core topics that could be repurposed in order to define the core topics of the AIDA PhD curriculum. The TAILOR PhD curriculum has been adapted to be a specialisation in Trustworthy AI in the AIDA curriculum (see [Section 5](#)).

Each of the presented topics includes a detailed description of various dimensions which were developed by the AIDA PhD Curriculum committee chaired by Barry O'Sullivan. The **academic level of the intended audience** has three levels: foundational, intermediate and advanced. The **specificity of the topic** can be broad, niche or specialised. And the **type of content** varies between theoretic, algorithmic or methodological. Then the **learning outcomes** are divided into (i) **content/knowledge** which specifies the expected knowledge that a student should obtain, and what types of questions should be able to reason about; (ii) the **methodological/skills** that demonstrate that a student has acquired the theoretical knowledge and can understand how it maps into a methodological pipeline to build and evaluate an AI system; and (iii) **transferable/application** defines cross-disciplinary skills that can help the students in group work or real application scenarios. Additionally, some topics also provide a list of available online courses that could be provided by one of the TAILOR partners.

## 4. TAILOR PhD curriculum in Trustworthy AI

This TAILOR PhD curriculum equips students with essential knowledge and skills for developing trustworthy Artificial Intelligence systems. With a general foundation on Artificial Intelligence, students learn how an integrated approach to learning, optimisation and reasoning can achieve trustworthiness. The first three topics are introductory, with (0) a general overview of Artificial Intelligence that ensures a common understanding across the curriculum; (1) a special focus on the trustworthy aspects of AI; and (2) how to leverage complementary AI paradigms and representations. The last three advanced topics include (3) AI agents deciding and learning how to act; (4) AI agents acting and learning in society; and (5) ensuring that AI tools and systems are performant, robust and trustworthy.



## 4.0. Foundations of Artificial Intelligence

This topic presents the foundations, scope, history and methodologies of AI.

**Level: Foundational, Broad, Theoretical.**

### Content/Knowledge

Students should be able to:

- Comprehend and compare the various **definitions of AI**.
- Understand/describe the **history of AI** and the eras into which it can be periodized.
- Properly **position AI within computer science** and analyse its links with other fields of science or philosophy (neuroscience, philosophy of mind, electrical/electronic engineering, mathematics, cognitive science).
- Understand and historically order the most important **propositions in the philosophy of AI** (e.g., Turing test, physical symbol system hypothesis, etc.).
- Comprehend the specific **relationship of AI** with logic, applied maths, game theory and other areas of mathematics.
- Compare and discriminate between different **AI methodological paradigms** (symbolic, computational, etc.).
- Understand/describe the concept of the **intelligent agent**.

### Methodological/Skills

Students should be able to:

- Apply their critical and analytical faculties, in order to argue about the comparative advantages/disadvantages of different methodological paradigms from the rich history of AI.
- Clearly argue about similarities and differences between natural/human intelligence and artificial intelligence, given the current level of technological progress and potential near-future advances.

### Transferrable/Application

Students should be able to:

- Work effectively with others in an interdisciplinary and/or international team.
- Clearly and succinctly communicate their ideas to technical and non-technical audiences.

## 4.1. Foundations of Trustworthy AI

This topic covers the dimensions of Trustworthy AI: (i) Explainability, (ii) Safety, (iii) Fairness, (iv) Accountability and Reproducibility, (v) Privacy, and (vi) Sustainability.

**Level: Foundation, Broad, Theory, Methodological.**

## Content/Knowledge

Students should be able to understand/describe current discourse on the following questions:

- How can we guarantee user trust in AI systems through explanation? How to formulate **explanations as Machine-Human conversation** depending on context and user expertise?
- How to bridge the gap from safety engineering, formal methods, verification as well as validation to the way AI systems are built, used, and reinforced?
- How can we build algorithms that respect **fairness constraints** by design through understanding causal influences among variables for dealing with bias-related issues?
- How to uncover **accountability** gaps w.r.t. the attribution of AI-related harming of humans?
- Can we guarantee **privacy** while preserving the desired utility functions?
- Is there any chance to reduce energy consumption for a more **sustainable AI** and how can AI contribute to solving some of the big sustainability challenges that face humanity today (e.g. climate change)?
- How to deal with properties and tradeoffs among multiple dimensions? For instance, accuracy vs. fairness, privacy vs. transparency, convenience vs. dignity, personalization vs. solidarity, efficiency vs. safety and sustainability.

## Methodological/Skills

Students should be able to:

- apply their critical and analytical faculties on specific case studies, in order to argue about the need and content of AI trustworthiness issues.

## Transferrable/Application

Students should be able to:

- Work effectively with others in an interdisciplinary and/or international team.
- Clearly and succinctly communicate their ideas to technical and non-technical audiences.

## 4.2. AI Paradigms and Representations

This topic covers the challenge of integrating different representations and paradigms for AI in order to enable both learning, reasoning and optimisation. The integrated representations are intended to engender trustworthiness.

**Level: Intermediate, Broad, Algorithmic, Methodological.**

## Content/Knowledge

Students should be able to:

- Understand the motivations for the need to integrate learning, reasoning and optimisation, and the role of prior knowledge and knowledge representation.
- Understand integrated representations for Trustworthy AI.



- 
- Understand different paradigms that integrate different representations. In particular:
    - Statistical relational AI: the integration of logic and probability/fuzziness for both reasoning and learning
    - Neurosymbolic AI: integrating logic with neural networks to enable perception and reasoning
    - Knowledge graphs, ontologies, graph neural networks and embeddings.
    - Constraint satisfaction and optimisation techniques: integrating solvers and learners for better performance and for learning CSP models.
  - Apply the above methods in perception, spatial reasoning, natural language processing, vision, and other societal/industrial domains

## Methodological/Skills

Students should be able to:

- Use a wide variety of representations (graphical models, logic, neural networks, knowledge graphs) for both learning and reasoning
- Discern the power and limitations of different types of representations.
- Combine different representations for a particular AI task.
- Use both knowledge and data for a particular AI problem.
- Understand and use the above-mentioned categories of techniques (StarAI, NeSy, CSP, Knowledge graphs, Ontologies (OWL/RDFS)).
- Understand the limitations and challenges of the integrated representations and paradigms.
- Understand the trustworthiness of these techniques.

## Transferrable/Application

Students should be able to:

- Work effectively with experts in different learning, reasoning and optimisation paradigms.
- Collaborate with domain experts to identify suitable integrated learning, reasoning and optimisation techniques for Trustworthy AI.

## 4.3. Deciding and Learning How to Act

This topic covers ways in which AI agents can be empowered with the ability to deliberate autonomously how to act in the world.

**Level: Advanced, Broad, Theory, Algorithmic.**

## Content/Knowledge

Students should be able to:

- Understand the different approaches in the fields of Artificial Intelligence and Formal Methods that can be applied in synergy to develop autonomous agents.
- Recognize the mathematical and algorithmic techniques as well as the key challenges to solving sequential decision-making problems.
- Integrate data-driven learning methods with model-based reasoning methods for deciding and learning how to act.

- Identify the limitations of current machine learning and reasoning methods to act in the real world.

## Methodological/Skills

Students should be able to:

- Program advanced agents using learning and planning techniques for solving sequential decision-making tasks that involve other agents.
- Analyse autonomy in dynamic, partially observable settings involving a single agent or multiple agents.
- Develop methods for optimising control policies in complex sequential decision-making problems.
- Implement techniques to balance exploration and exploitation in decision-making tasks that require learning from the environment while acting on it.
- Use linear time logic as a specification language for formulating complex tasks as well as environment properties.
- Apply synthesis from formal specifications to solve planning problems in nondeterministic environments.

## Transferrable/Application

Students should be able to:

- Work effectively with others in an interdisciplinary and/or international team to reach a collective objective by sharing knowledge, learning and building consensus.
- Present materials coherently and concisely in written or oral form, with clear use of language to a technical audience.

## 4.4. Reasoning and Learning in Social Contexts

This topic covers the foundations, techniques, algorithms and tools for allowing autonomous AI agents to be social and act within societies. It will offer a breadth of understanding in technologies that allow building Social AI systems and a multidisciplinary take on this topic that will impact every aspect of our daily life in the future.

*Level:* **Advanced, Broad, Theory, Algorithmic.**

## Content/Knowledge

Students should be able to:

- Comprehend that capturing the social aspects of human behaviour is essential in understanding **how people think and how people react to each other**, which is a fundamental step to developing reasoning algorithms that can operate effectively in social contexts.
- Demonstrate a good **understanding of computational models of social reality**. That is, how social contexts determine human behaviour through norms, practices, conventions, rituals and other rules of human social nature.
- Understand current methodologies to model social cognition, collaboration and teamwork.
- Understand /describe theoretical models for cooperation between agents.

- Understand the process of creating systems equipped **with perception and social capabilities** that allows them to adapt to different social contexts and **learn from other agents** in such environments.
- Understand how models of social reality generate emergent behaviour and the impact of such models in agent societies and social networks of multi-agent systems.

## Methodological/Skills

Students should be able to:

- Correctly identify different ways to sense the environment and understand how to use off-the-shelf solutions and how to make sense of the captured data.
- Explore the creation of a simple Social AI System, using a perception technology whose data feeds into a reasoning mechanism that outputs social (and intelligent) acts in a context of choice.
- Evaluate social reasoning and learning algorithms in the form of simulations or with a human.
- Analyse the solutions to a problem and critically think about the societal impact.

## Transferrable/Application

Students should be able to:

- Work effectively with others in an interdisciplinary and/or international team.
- Design and manage individual projects.
- Clearly and succinctly communicate their ideas to technical audiences.

## 4.5. Automated AI

This topic covers meta-level methods to ensure that AI tools and systems are performant, robust and trustworthy.

**Level: Advanced, Broad, Theory, Algorithmic.**

## Content/Knowledge

Students should be able to:

- Explain the basic problems solved by AutoAI methods, including (but not limited to) automated algorithm configuration, automated algorithm selection, automated performance prediction, model selection, hyperparameter optimisation and neural architecture search.
- Explain, in general and using specific examples, the significance of AutoAI problems and methods for the broader field of AI, including (but not limited to) the areas of machine learning, automated reasoning and optimisation.

In addition, students should be able to achieve a selection of the following, more specific learning outcomes:

- Demonstrate a working knowledge of Neural Architecture Search, notably how to define search spaces and optimise over these spaces, with both differential and black box methods.

- 
- Identify and define a Hyperparameter Optimization Problem (HPO), specifically in the domain of Algorithm Configuration and Neural Architecture Search. They should also be familiar with hyperparameter importance techniques to interpret different solutions to these problems.
  - Be familiar with Gaussian Processes and their modelling capabilities, specifically in the domain of algorithm configuration.
  - Assess the strengths and weaknesses of various HPO methods, notably bayesian and evolutionary strategies for doing such.
  - Demonstrate knowledge on various speedup techniques to HPO, including leveraging previous information through meta-learning, learning curve prediction and bandit based scheduling techniques.
  - Define multiple objectives for an optimization problem and various evolutionary and bayesian techniques for solving such problems.
  - Explain Dynamic Algorithm Configuration (DAC) and its difference to Static Algorithm Configuration. They should also be able to demonstrate how to use Reinforcement Learning to solve such optimization problems in DAC.
  - Demonstrate knowledge of AutoAI methods for tasks that go beyond supervised learning. This includes knowledge of the underlying theoretical principles and algorithms as well as knowledge of specific tools and systems, including their correct and effective use, strengths and limitations.
  - Demonstrate knowledge of AutoAI methods, tools and systems for problems in areas outside of machine learning (i.e., knowledge beyond automated machine learning).
  - Understand the way how AI systems interact with their environment, and what are possible pitfalls of that (e.g., badly calibrated confidence statements, adversarial examples, (un)explainable decisions)
  - Explain the importance for AI tools and systems to be able to detect situations in which their use becomes problematic (e.g., ineffective or unsafe).
  - Demonstrate knowledge of techniques and approaches for achieving self-monitoring in at least one major area of AI.
  - Evaluate AI systems for safety problems in interacting with their environments
  - Demonstrate awareness of meta-learning, transfer learning, and continual learning techniques that can be leveraged to transfer information from earlier tasks.
  - Explain how this transfer of knowledge can be used to make AutoML techniques and systems more efficient.

## Methodological/Skills

Students should be able to:

- Correctly use a range of AutoAI techniques in at least one major area of AI.
- Critically assess (in technical and general ways) and explain the limitations of AutoAI methods.
- Recognise and explain technical problems that may arise in the use of AutoAI methods.

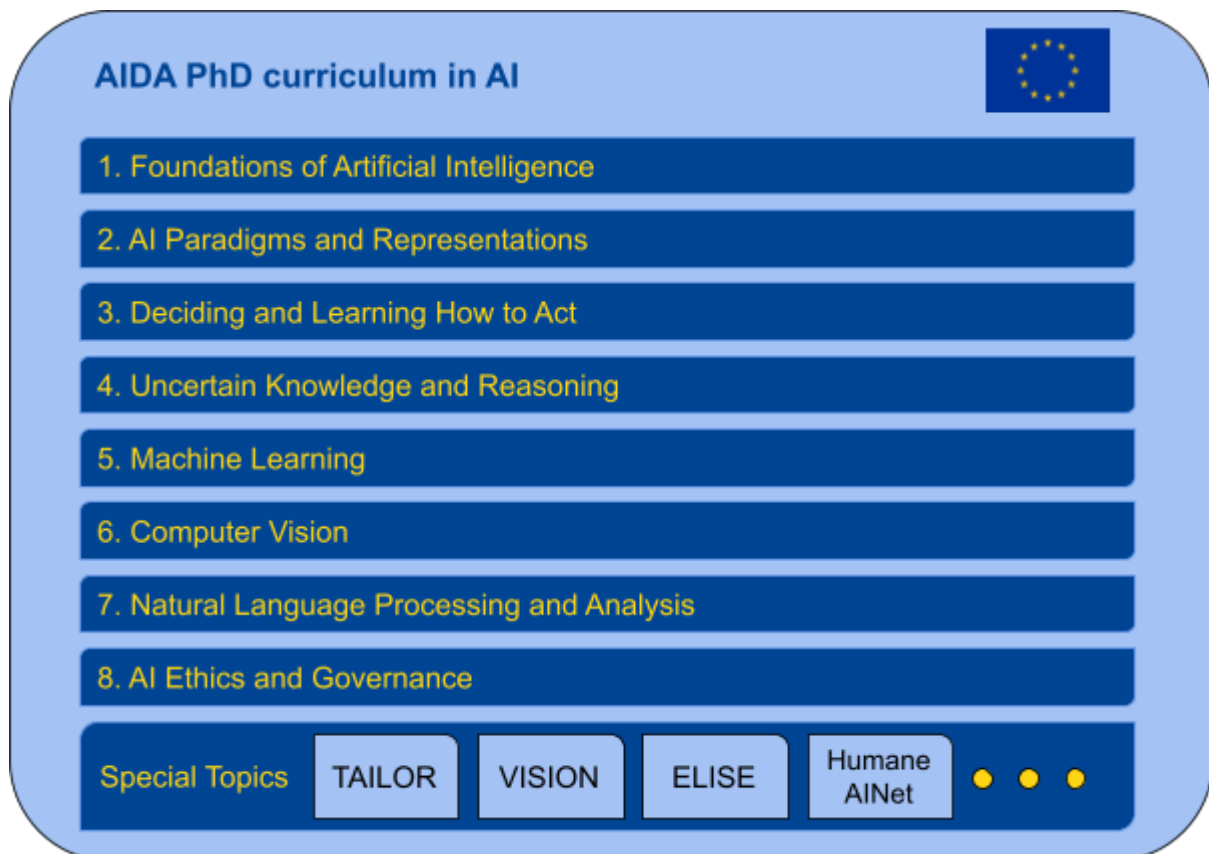
## Transferrable/Application

Students should be able to:

- Work effectively with others in an interdisciplinary and/or international team.
- Design and manage individual projects.
- Clearly and succinctly communicate their ideas to technical audiences.

## 5. Relationship with the AIDA PhD curriculum

Europe has multiple research networks around the topic of Artificial Intelligence (e.g. TAILOR, AI4Media, HumaneAI, Elise). It is then natural to see how the previous TAILOR PhD curriculum could be complementary to potential joint curricula. With this in mind, an AI PhD curriculum could be used to define a program that could be delivered by multiple partner institutions. Certain core topics should be covered in any PhD curriculum in AI, which could be used for other specialised curricula. We have closely worked with AIDA's PhD Curriculum committee to design a cross-network curriculum with a core component and special topics proposed by four ICT-48 networks. Based on the topics covered by every network, AIDA proposed to define the following eight core modules.



The specialisation could be done through a set of pillars led by a network of partner universities. Thus, the aforementioned TAILOR PhD curriculum could be split into two parts: the general core and the special topics. In particular, part of (1) Foundations of Trustworthy AI could be used in the general Foundations of Artificial Intelligence (core), (2) AI Paradigms and Representations and (3) Deciding and Learning How to Act could be completely moved to the core, (4) Reasoning and Learning in Social Contexts and (5) Automated AI could be left as special topics.

<b>TAILOR module</b>	<b>Relationship</b>	<b>AIDA module</b>
0. Foundations of Artificial Intelligence	partially contributes to	1. Foundations of Artificial Intelligence
1. Foundations of Trustworthy AI	partially contributes to	1. Foundations of Artificial Intelligence
2. AI Paradigms and Representations	is the same as	2. AI Paradigms and Representations
3. Deciding and Learning How to Act	is the same as	3. Deciding and Learning How to Act
4. Reasoning and Learning in Social Contexts	is a special topic	
5. Automated AI	is a special topic	

## 6. Concluding remarks

In this deliverable we have provided an overview of the process to design the TAILOR PhD curriculum with the collaboration of core consortium partners, and presented the outcome in detail. The TAILOR project has created a network of excellence in Trustworthy AI, and we exploited the expertise of each technical work package to define a curriculum that covers all aspects of Trustworthy AI. We have described how the proposed TAILOR PhD curriculum relates and contributes to a more comprehensive PhD curriculum in AI that is currently being developed by AIDA.

Possible future work involves a feasibility study on how to materialise the curricula, and to credit the required skills for a doctorate award. We are currently working on topic ontologies and badges that could be earned in different institutions, and be certified by a consortium to obtain the award.

## References

- ❖ Barnes, T. (2013). *Higher doctorates in the UK 2013*. UK Council for Graduate Education.
- ❖ Coates, H., Croucher, G., Weerakkody, U., Moore, K., Dollinger, M., Kelly, P., Bexley, E., & Grosemans, I. (2020). An education design architecture for the future Australian doctorate. *Higher Education*, 79(1), 79–94.  
<https://doi.org/10.1007/s10734-019-00397-1>
- ❖ Goodchild, L. F., & Miller, M. M. (1997). The American Doctorate and Dissertation: Six Developmental Stages. *New Directions for Higher Education*, 1997(99), 17–32.  
<https://doi.org/https://doi.org/10.1002/he.9902>
- ❖ Gould, J. (2015). How to build a better PhD. *Nature*, 528(7580), 22–25.  
<https://doi.org/10.1038/528022a>
- ❖ Gould, J. (2016). What's the point of the PhD thesis? *Nature*, 535(7610), 26–28.  
<https://doi.org/10.1038/535026a>
- ❖ Park, C. (2005). *New Variant PhD: The changing nature of the doctorate in the UK*. *Journal of Higher Education Policy and Management*, 27(2), 189–207.  
<https://doi.org/10.1080/13600800500120068>
- ❖ Park, C. (2007). *Redefining the doctorate*.
- ❖ Sarrico, C.S. The expansion of doctoral education and the changing nature and purpose of the doctorate. *High Educ* 84, 1299–1315 (2022).  
<https://doi.org/10.1007/s10734-022-00946-1>
- ❖ Simpson, R. (1983). *How the PhD Came to Britain. A Century of Struggle for Postgraduate Education*. SRHE Monograph 54. ERIC.
- ❖ Sharmini, S., & Spronken-Smith, R. (2020). The PhD – is it out of alignment? *Higher Education Research & Development*, 39(4), 821–833.  
<https://doi.org/10.1080/07294360.2019.1693514>
- ❖ Wyatt, J. (1998). “The lengthened shadow of one man”: the public intellectual and the founding of universities. *Higher Education Review*, 30(2), 29.  
<https://bris.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/lengthened-shadow-one-man-public-intellectual/docview/1297984863/se-2?accountid=97>

---

## Appendix A. Glossary of terms

This Glossary was originally written by Peter Flach to facilitate inter-network discussion and adopted by the AIDA Curriculum Committee chaired by Barry O'Sullivan.

**Cohort-based doctoral training** - PhD programmes are increasingly offered in a cohort-based way, where a group of students work on thematically related PhD topics. Such programmes can be organised within a single University (e.g., Centres for Doctoral Training in the UK) or by a consortium of universities (e.g., the WASP Graduate School in Sweden).

**Learning outcome** - the measurable skills, abilities, knowledge and values that a **PhD student** will be able to demonstrate as a result of successfully completing a given **training module**.

**PhD curriculum** - a specification of the structure and content of a **PhD programme**. This typically concentrates on the doctoral training part, which may include **training modules**, self-study, practical assignments, internships and research visits, the acquisition of **transferable skills**, etc. Elements of the doctoral training programme are typically delivered to cohorts of students on the same **PhD programme** (see **cohort-based doctoral training**), or to groups of students across different **PhD programmes** but within one University. Completion of parts or all of the doctoral training programme may be mandatory in order to progress to (or proceed with) the research part of the **PhD programme**, in which case the corresponding modules will be examined through written or oral exams and/or coursework.

**PhD degree** - a doctoral degree conferred by an accredited institution for higher education (typically a University).

**PhD examination** - the process by which PhD examiners determine whether a **PhD thesis** is worthy of awarding a **PhD degree**. This can involve one or more of the following:

- written reports from the examiners after reading the thesis;
- a *viva voce* examination behind closed doors;
- a public PhD defence.

The outcome of the PhD examination can be binary (whether the PhD is awarded or not) or may stipulate minor or major corrections of the written thesis.

**PhD programme** - a combination of research training and supervised research practice, offered by an accredited institution for higher education, with the aim of achieving a **PhD degree** after submitting and successfully defending a **PhD thesis**. **PhD students** need to formally enrol on a PhD programme, typically 3-4 years in duration, for which they may need to pay a tuition fee.

**PhD student** - a person enrolled on a **PhD programme** with the aim of obtaining a **PhD degree** after several years of research training and practice.

**PhD supervisor** - an academic with responsibility for guiding a **PhD student** during their PhD studies, and monitoring their progress.

**PhD thesis** - a written treatise on a chosen research topic, typically 150-200 pages long, submitted at the end of a **PhD programme** to satisfy the requirements of being awarded a **PhD degree**, subject to **PhD examination**.



**Training module** - part of a **PhD programme** typically consisting of a series of lectures or seminars, and (if mandatory) assessed through written or oral exams and/or practical assignments.

**Transferable skill** - a skill that is relevant for successful PhD study but more generally applicable, such as writing and presentation skills.

## Appendix B. Badges and consortium-based programmes

Badges have been used in online platforms to incentivize the accomplishment of certain tasks by platform users. The concept of “gamification” was used already in 2008 for marketing and user engagement (Currier, 2008), and can be defined “as the use of game design elements in non-game contexts” (Detering et al., 2011). Other forms of gamification are the use of points and leaderboards to compare the progress of students. The use of badges has been shown to motivate users and influence their behaviour (Anderson et al. 2013). However, some of the competitive aspects may be counterproductive as the students may focus on acquiring higher scores rather than knowledge and skills.



For that reason, caution needs to be exercised if badges are used as a means of “gamification” and motivation. But they could be used as a means of certification of the gained skills, necessary to be awarded the doctorate in Trustworthy AI. The badges could facilitate the creation of PhD programs involving multiple institutions.









The idea of certified digital badges is not new. On March 14th, 2013, Mozilla started an initiative of Open Badges and its "application-programming interface" to allow issuers to create badges that users could earn (Gibson et al., 2015). However, there was no widespread adoption which led to Mozilla announcing the transfer of Open Badges to the IMS Global Learning Consortium in 2016<sup>1</sup>. In 2018, Mozilla retired its tool to store badges (known as Backpack)<sup>2</sup> and its users migrated to Concentric Sky's Badgr platform. During the migration, some of the originally issued badges lost their authentication, which made their migration impossible and the loss of badges by some users. Some of the possible reasons for the lack of adoption may be (1) lack of relevant badges, (2) lack of information about the existence of the accreditation digital system, (3) companies and organisations currently perform manual checks of accreditations with institutions during the hiring or acceptance process successfully, (4) lack of trust among users regarding maintenance of the obtained badges, (5) a chicken and egg problem of organisations not checking them because people do not use them, and people not using them because the organisations do not require them.

Worldwide digital certified badges are still in the early stages and lack adoption. For that reason, a TAILOR PhD program should agree beforehand on the participating institutions and the accreditation badges. A (non-)centralised committee should ensure that the

























<sup>1</sup> “Mozilla Foundation and Collective Shift/LRNG to Transition Management of the Popular Open Badges Project to 1EdTech to Ensure Long-term Support and Sustainability”  
<https://www.imsglobal.org/article/1edtech-mozilla-foundation-and-lrng-announce-next-steps-accelerated-evolution-open-badges>

<sup>2</sup> “An Update on Badges and Backpack”  
<https://medium.com/read-write-participate/an-update-on-badges-and-backpack-5a06fab252ea>





























certifications are obtained by the student and that the PhD can be awarded. In the future, if digital certificates are adopted worldwide, the revision of awards could be facilitated. Based on the PhD curriculum defined in this deliverable, the following is an example of a set of badges that could be used to specify its requirements.

Badge	Description
 ... 	Topic-specific badges. The examples on the left are “Artificial Intelligence” and “AI Paradigms and Representations”.
 ... 	The level of difficulty of the course can be foundation, intermediate or advanced.
	The topic has a strong theoretical component.
	The topic assesses methodological skills.
	The topic provides multiple algorithms.
	The topic has a vital practical component.

The previous badges can be combined to create a set of requirements for each TAILOR module. The following table shows each module and the corresponding set of badges.

TAILOR module	Badges
0. Foundations of Artificial Intelligence	  
1. Foundations of Trustworthy AI	   
2. AI Paradigms and Representations	    
3. Deciding and Learning How to Act	   
4. Reasoning and Learning in Social Contexts	   
5. Automated AI	   

With the help of the specified badges, it is easy for a TAILOR institution to assess its potential to provide part of the PhD curriculum. We take as an example the foundation year of the Center for Doctoral Training (CDT) in Interactive Artificial Intelligence at the University of Bristol<sup>3</sup>. The following table lists the modules offered by the CDT and the corresponding set of badges that could be obtained after completion.

University of Bristol courses for the IAI CDT	Badges
Computational Logic for Artificial Intelligence	  
Dialogue and Narrative	  
Machine Learning Paradigms	   
Responsible AI	   
Applied Data Science	 
Uncertainty Modelling for Intelligent Systems	   
Interactive AI Team Project	  
Research Methods in Interactive Artificial Intelligence	 
AI Summer Project	  

By mapping the badges from the TAILOR PhD curriculum to the ones offered by the Interactive AI CDT we see that the CDT covers most of the necessary syllabus but would require additional courses (Deciding and Learning How To Act, Automated AI) which could be provided by other TAILOR partners.

---

<sup>3</sup> See <https://www.bristol.ac.uk/cdt/interactive-ai/programme-details/foundation-year/>

## References

- ❖ Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013). Steering User Behavior with Badges. *Proceedings of the 22nd International Conference on World Wide Web*, 95–106. <https://doi.org/10.1145/2488388.2488398>
- ❖ Currier, J. (2008, November 5). *Gamification: Game Mechanics is the New Marketing*. Ooga Labs Blog. <https://blog.oogalabs.com/2008/11/05/gamification-game-mechanics-is-the-new-marketing/>
- ❖ Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining “Gamification.” *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. <https://doi.org/10.1145/2181037.2181040>
- ❖ Gibson, D., Ostashewski, N., Flintoff, K., Grant, S., & Knight, E. (2015). Digital badges in education. *Education and Information Technologies*, 20(2), 403–410. <https://doi.org/10.1007/s10639-013-9291-7>