



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

**TAILOR**

**Grant Agreement Number 952215**

**Report for Second TAILOR Summer School, D9.8**

<b>Document type (nature)</b>	Report
<b>Deliverable No</b>	9.8
<b>Work package number(s)</b>	9
<b>Date</b>	Due M20
<b>Responsible Beneficiary</b>	ID #31, CSIC
<b>Author(s)</b>	Sierra Carles, Cester Laura
<b>Publicity level</b>	Public
<b>Short description</b>	Report from the second TAILOR Summer School, held in June 2022 in Barcelona.

<b>History</b>			
<b>Revision</b>	<b>Date</b>	<b>Modification</b>	<b>Author</b>
1.0	2022-10-04	-	Laura Cester

<b>Document Review</b>		
<b>Reviewer</b>	<b>Partner ID / Acronym</b>	<b>Date of report approval</b>
Peter Flach	# 16, UNIBRIS	September 12, 2022
Umberto Straccia	# 2, CNR	October 4, 2022

## Table of Contents

<b>Summary of the report</b>	<b>3</b>
<b>Introduction</b>	<b>3</b>
<b>Organisation</b>	<b>4</b>
Summer school objective	4
Organisation details	4
Summer School Dissemination Actions	4
Flyer	5
Program	5
Tutorials	7
<b>Data on participants</b>	<b>10</b>
Lecturers	12
<b>Certificate of attendance</b>	<b>13</b>
<b>Pictures posted on social media</b>	<b>14</b>
<b>Appendices</b>	<b>16</b>
1: Grants and fees	16
2: Social programme	16
3: SubSift Summary	17

## Summary of the report

The present deliverable reports on the second TAILOR Summer School, held in Barcelona in June, 2022.

The summer school was co-arranged with ACAI and was held on 13 to 17 June. A total of 124 people participated, of which 67 were from TAILOR partners. The summer school had two parallel tracks and one of them was on trustworthy AI. Of the 16 tutorials, 8 were on different aspects of trustworthy AI and many of them given by esteemed researchers from TAILOR partners.

AI-based tools for the grouping of people according to professional interests and personalities were tested for the purpose of encouraging social connections between students. One of the objectives was to promote the TAILOR Connectivity Fund.

## Introduction

The second TAILOR summer school took place in Barcelona, on June 13-17, 2022. The school shared organisation with the Intelligent Data Science and Artificial Intelligence Research Center at Universitat Politècnica de Catalunya (IDEAI-UPC), which celebrated its 19th Advanced Course on AI (ACAI) specialised in explainable AI.

The main topic of the TAILOR Summer School was Trustworthy AI where different international researchers participated as lecturers offering updated training on the requirements to achieve a trustworthy AI, including transparency, diversity, robustness, and privacy.

The school had diverse formats, such as keynote presentations, lectures, labs/hands-on sessions, short tutorials on cutting-edge topics, and longer in-depth tutorials. Besides, a Doctoral Consortium was arranged where Ph.D. students could discuss their Ph.D. theses.

The target group of the Summer School was mainly Ph.D. students and postdocs in the research areas related to the TAILOR project. In addition, a few professors, scientific researchers, and master students also participated.

In total, there were 124 participants, of which 85% were doctoral students. As for the gender distribution, 71% were men. The countries with the most participants were Spain, Italy, Netherlands, Germany, Ireland, and France. Regarding the research fields, 53% of the participants worked in Machine Learning, followed by Computer Vision, Multiagent systems, and AI ethics. More details about this is provided in the section Data for more information.

From TAILOR, 67 people associated with TAILOR participated in the summer school, most of them students.

As a novelty, SubSift has been considered, which is a tool to match the participants for their topics of interest, creating groups that have similar motivations to meet and discuss future collaborations. The objective was to facilitate the interaction and networking between participants in the 2nd TAILOR Summer School. It had a positive impact with 70 participants and feedback from 23 of them.

## Organisation

The following people have been involved in the Deliverable:

Partner ID / Acronym	Name	Role
# 31, CSIC	Carles Sierra	Eurai President
# 31, CSIC	Laura Cester	Content Manager

## Summer school objective

The WP9 (Network Collaboration) commits TAILOR to hold summer schools dedicated to the best young talent and scientist networking. This report focuses on the second summer school in Barcelona during the M20, June 2022.

## Organisation details

The Summer School took place in Barcelona, specifically at the Universitat Politècnica de Catalunya. The registry had different fees depending on the situation of the participant. For TAILOR members there was a special price that can be checked at the appendix 1. The fee for up to three students per TAILOR partners was covered by the CSIC TAILOR funds for the conference.

The local organisation partner was the Intelligent Data Science and Artificial Intelligence Research Centre at the UPC. They carried out the registration process and organisation of rooms and schedules. On the part of TAILOR, a small committee had a few online meetings in preparation of the summer school.

In addition, the school offered a social program (included in the registration fee) to promote networking and to know the UPC and the city of Barcelona and its culture and social life. It can check the social events in the appendix 2.

## Summer School Dissemination Actions

The summer school has been advertised through different channels:

- Emails:
  - from TAILOR in emails to the consortium (TAILOR Newsletter; emails to remind the TAILOR Summer School)
- Post on the UPC [Website](#) and TAILOR [website](#)
- Reminder during TAILOR events and meetings
- Internal communications within the partner institutions.
- TAILOR and partner institution's social networks

## Flyer

A flyer was produced and used to advertise the event through multiple channels (see section above Summer School dissemination actions)



## Program

The program consisted of five days of classes and was prepared jointly by ACAI and TAILOR with parallel tutorials where students could choose which one to attend.

Below, the complete program:

	TRACK - ACAI-TAILOR School 2022, Barcelona 13-17 June									
	Mon 13 June		Tue 14 June		Wed 15 June		Thu 16 June		Fri 17 June	
Morning	Monday		Tuesday		Wednesday		Thursday		Friday	
9:00-9:30	Registration		Jose M Juarez <i>(chair: E Almirall)</i>  Farah Benamara Zitoune <i>(chair: A. Nebot)</i>		Anna Monreale <i>(chair: A. Valls)</i>  Clara Neppel <i>(chair: R. Morros)</i>		Thomas Villmann <i>(chair A. Vellido)</i>  André Meyer-Vitali <i>(chair: R. Sangüesa)</i>		Pilar Dellunde/Vicent Costa <i>(chair: C. Barrué)</i>  Andreas Herzig <i>(chair: J. Delgado)</i>	
9:30-10	Opening ceremony									
10:00-10:30	Opening keynote Paulo Cortez <i>(chair: K. Gibert)</i>									
10:30-11:00										
11:00-11:30	Break									
11:30-12:00	Elvira Oliva <i>(chair: J.Vázquez)</i>	Amos Yehuda Azaria <i>(chair: A. Cortés)</i>	Break		Break		Break		Break	
12:00-12:30										
12:30-13:00			Riccardo Guidotti <i>(chair: E. Sayrol)</i>		Pompeu Casanovas <i>(chair: C Sierra)</i>		Giorgio Visani <i>(chair: R. Sangüesa)</i>		Nava Tintarev <i>(chair: L. Padró)</i>	
13:00-13:30										
13:30-14:00	Lunch		Lunch		Lunch		Lunch		Lunch	
14:00-14:30										
Afternoon										
15:00-15:30	Jose M Alonso <i>(chair: J. Hernando)</i>	Neil Yorke-Smith <i>(chair: D. Garcia)</i>	Paulo Lisboa <i>(chair: F. Marquès)</i>	Loizos Michael <i>(chair: X. Angerri)</i>	Aïda Valls/ Núria Agell <i>(chair: A. Cabellos)</i>	Carme Torra <i>(chair: J Aranda)</i>	Doctoral Consortium		Closing ceremony and Doctoral Consortium Awards Fredrik Heintz <i>(chair:C. Sierra)</i>	
15:30-16:00										
16:00-16:30										
16:30-17:00	Break		Break		Break		Break			
17:00-17:30	Jose M Alonso	Neil Yorke-Smith	Paulo Lisboa	Loizon Michael	Aïda Valls/Núria Agell	Carme Torra	Doctoral Consortium			
17:30-18:00										
18:00-18:30										
Social Program	Visit to the suppercomputer MareNostrum <b>Care: LUNCH TIME</b>		Glass of Cava in the Torra Girona gardens (UPC)		Concert of Choir Llevant at Sant Pau del Camp church		Gala Dinner at Balmesiana Palace		Get Together Cocktail at the Indoor garden of Hotel Catedral (paid on its own)	



## Tutorials

In particular, these were the tutorials of the TAILOR track:

### Monday 13th

#### **Irrational, but Adaptive and Goal Oriented: Humans Interacting with Autonomous Agents**

*Amos Yehuda Azaria (Ariel university, Israel)*

Autonomous agents that interact with humans are becoming more and more prominent. There are several interaction modes. The humans and agents may compete with each-other, as in zero-sum games. The humans and agents may be fully cooperative, and the human may assist the agent in its task. However, while interaction between two humans is usually neither zero-sum nor fully cooperative, not much work on such interaction appears in the literature. We will explore the challenges of such interaction, such as the development of an environment that is intuitive for humans and data collection. In addition, we will observe common approaches such as composing a model of human behaviour based on observing human actions and then optimising the agent's actions based on this model, and model free reinforcement learning, which learn which actions the agent should take at which states. Finally, we will discuss several shortcomings of common approaches and explore solution concepts that attempt to overcome these problems.

#### **Reliable Autonomous Systems: Certification, Trust and Hybrid Intelligence**

*Neil Yorke-Smith (Delft University of Technology (TU Delft), The Netherlands)*

Should a self-driving car have a driver's licence? The multiple manifestos for 'trustworthy AI' have hardly, as yet, translated into practice in the development and deployment of automated and autonomous systems. This tutorial considers the operation of autonomous systems in human society, in the light of current certification and licencing of humans and automated systems. Often, autonomous systems must interact with humans -- for example, food delivery robots -- and this means that trust and concepts of hybrid man-machine 'intelligence' further play into questions of trustworthiness and reliability. The tutorial considers systems on the axes of the scope of autonomy, the safety-critical nature of the tasks, and the amount of available regulation. We will take examples from household, industrial, vehicle and aerospace domains. The tutorial combines interactive discussion with break-out case studies.

### Tuesday 14<sup>th</sup>

#### **Please Do Not Moderate Me: The Challenges of Hate Speech Detection in Social Media**

*Farah Benamara Zitoune (Dep. Computer Science. U. Paul Sabatier, Toulouse. France)*

Hate Speech (HS) and harassment are particularly widespread in online communication, especially due to users' freedom and anonymity and the lack of regulation provided by social



media platforms. This phenomenon has determined a growing interest in using artificial intelligence and Natural Language Processing techniques to address social and ethical issues. An extensive body of work has been proposed to automatically detect HS relying on a variety of deep learning methods (Founta and Nunes, 2018; Schmidt and Wiegand, 2017). In this talk, I'll survey main existing approaches to HS in social media from corpus collection, annotation to automatic detection, highlighting the main challenges and the solutions proposed to address them. The second part of the talk will focus on hateful content towards women, with a special attention to the detection of denunciation of sexist content (Chiril et al, 2020).

#### **AI 4.0: Learn. Argue. Explain.**

*Loizos Michael* (Open University of Cyprus, Cyprus)

With the oncoming Fourth AI Revolution and the demand for enhanced collaboration between humans and machines, it is imperative to re-evaluate the role of machine learning in the design of human-centric AI systems. Instead of focusing on achieving predictive accuracy, learning should be viewed as a means to the eventual end of supporting the meaningful consumption of the learned knowledge by the machine's human collaborators. Explanations in this context act as proxy translations of the learned knowledge into a form that is cognitively-compatible with the reasoning abilities and prior knowledge of those collaborators. With psychological evidence pointing to argumentation as the lingua franca of human reasoning, it is natural to investigate how explanations can be represented and communicated in an argumentative language, and how our perspective of machine learning can be recast to support argumentative dialogues between humans and machines. After discussing these issues, the tutorial will introduce the autodidactic and coaching-based learning paradigms for the acquisition of commonsense and collaborator-specific knowledge, and it will demonstrate concrete algorithms under these two paradigms that acquire knowledge in an ante-hoc explainable argumentative language.

Wednesday 15<sup>th</sup>

#### **The need for new measures of success and risk for AI**

*Clara Neppel* (IEEE European Business Operations)

The proposed EU AI ACT is a risk based approach. Thereby, risk itself is redefined, since besides the traditional dimensions familiar to technologists, such as safety and security, also the risks to human rights are taken into account. In order to maximize human and environmental wellbeing, we consequently also need to redefine the measures of success. How to achieve these aims and what are the possibilities to consider them during the whole AI lifecycle, from design to deployment?

#### **(Keynote organized by TAILOR) The European Artificial Intelligence Act: An Overview and Some Critical Points**

*Pompeu Casanovas* (Advanced Research in Philosophy and Sociology, Autonomous University of Barcelona UAB, Spain)

Tuta quia diffidens, 'Safe for Distrusting', was the motto of the old literary Catalan Academy of Suspicious (or Distrustful)—Acadèmia dels Desconfiats (1700-1714). It could be the motto of the new European proposal of Regulation of the EU Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (May 2021). It is called the Artificial



Intelligence Act, and it can be figured out as a follow-up of the General Data Protection Regulation, in force since 2018. What are the main traits of this proposal? What type of regulation is it? How is the field of Artificial Intelligence defined? How does it deal with the risks that AI applications may entail? Which are the pros and cons of the proposal and how can it affect the practical work of engineering design? I will introduce in my presentation some ethical and legal issues to start answering these questions.

### **Intelligent Robotic Assistants: Research challenges and ethics teaching inspired by science fiction**

*Carme Torras* (IRI-CSIC. Barcelona, Spain)

What would a perfect robotic assistant look like? What abilities should it be endowed with? Which are the challenges for both AI research and ethics education? The need to function in dynamic human-centered environments poses new defies to the robotics and AI research communities. Differing from industrial manipulators and vehicles, robotic assistants need to have friendly interfaces that allow easy teaching by non-experts, be highly adaptable and personalizable, trustworthy and intrinsically safe to people, and able to handle deformable materials, among other demanding skills. Besides these technical defies, robotic assistants raise also ethical challenges, which have led to devoting increasing attention to AI ethics and robot ethics. Several institutions are developing regulations and standards, and many ethics education initiatives are being pursued, where science fiction often plays a prominent role by highlighting the pros and cons of possible future scenarios. In this tutorial, the state of the art in intelligent assistive robotics will be briefly reviewed, and educational materials from a university course on "Ethics in Social Robotics and AI" will be presented, which are based on the science fiction novel *The Vestigial Heart* and can be downloaded free of charge from MIT Press website.

Thursday 16<sup>th</sup>

### **Trustworthy Hybrid Team Decision-Support**

*André Meyer-Vitali* (DFKI Saarbrücken, Germany)

The aim to empower human users of artificially intelligent systems becomes paramount when considering the coordination and collaboration in hybrid teams of humans and autonomous agents. Hereby, we consider not only one-to-one interactions, but also many-to-many situations (multiple humans and multiple agents), where we strive to make use of their complementary capabilities. Therefore, mutual awareness of each other's' strengths and weaknesses is crucial for beneficial coordination. Each person and agent has individual knowledge, facilities, roles, capabilities, expectations, and intentions. It should be clear for each of them what to expect from each other, in order to avoid misleading anthropomorphism, and how to delegate which tasks to whom. In order to address these goals, and in accordance with a hybrid theory of mind, we propose the use of trustworthy interaction patterns and epistemic orchestration with intentions and causal models.

Friday 17<sup>th</sup>

### **Artificial Intelligence and Social Intelligence**

*Andreas Herzig* (Institut de Recherche en Informatique de Toulouse (IRIT) of Université Paul Sabatier, France)

Social intelligence is the ability to put oneself into somebody else's shoes: an agent possessing a so-called theory mind can understand other agents' beliefs, goals, and

emotions and take them into account in her own reasoning and acting. When people tried to define human intelligence they often neglected this social aspect, and artificial intelligence by and large reproduced that omission. It is however well-known in psychology that theory of mind is of fundamental importance in human-human interaction, indicating that robots and other intelligent devices should also be equipped with such a theory in order to interact with humans in a flexible and meaningful way. A formal framework for theory of mind is offered by epistemic logics, which provide a tool to reason about other agents' knowledge and belief. The course overviews these logics and their role in an account of social intelligence.

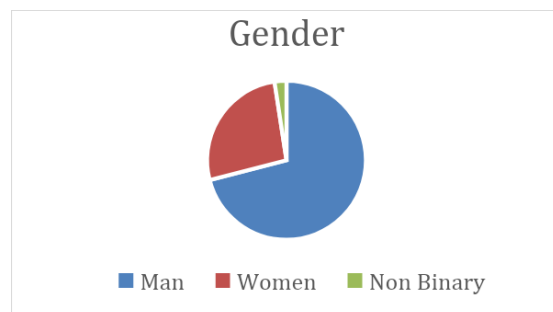
### (Closing keynote) **Learning and Reasoning with Trajectory Data**

*Fredrik Heintz* (Linköping University, Sweden)

Accurately analysing and predicting movements of objects through time and space is central to many applications. This talk presents research on learning generative models based on trajectory data, probabilistic logical reasoning over observed and predicted trajectories (and other time series data), and privacy-preserving synthetic data generation. Methods include Gaussian Processes and temporal GANs (Generative Adversarial Networks) for learning generative models, probabilistic signal temporal logic for reasoning, and Bayesian Optimization for synthetic data generation. The research is applied to both autonomous systems, such as unmanned ground and aerial vehicles, and traffic monitoring and analysis together with for example the Swedish Transport Administration and the local transportation authority.

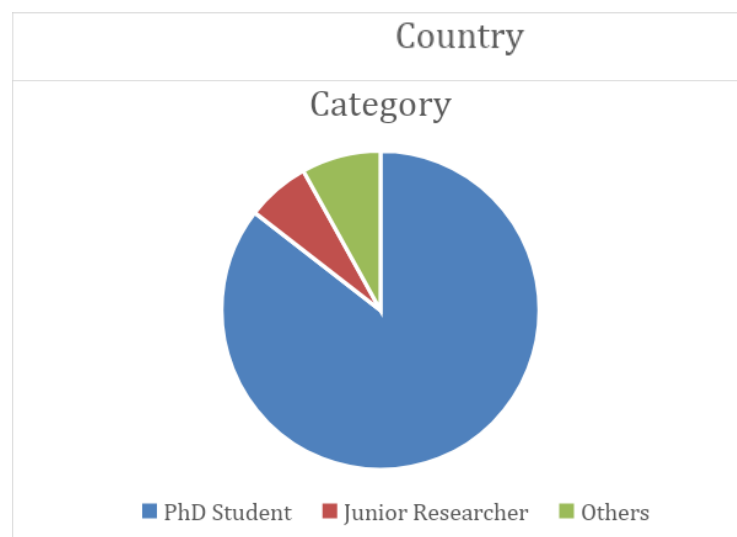
## Data on participants

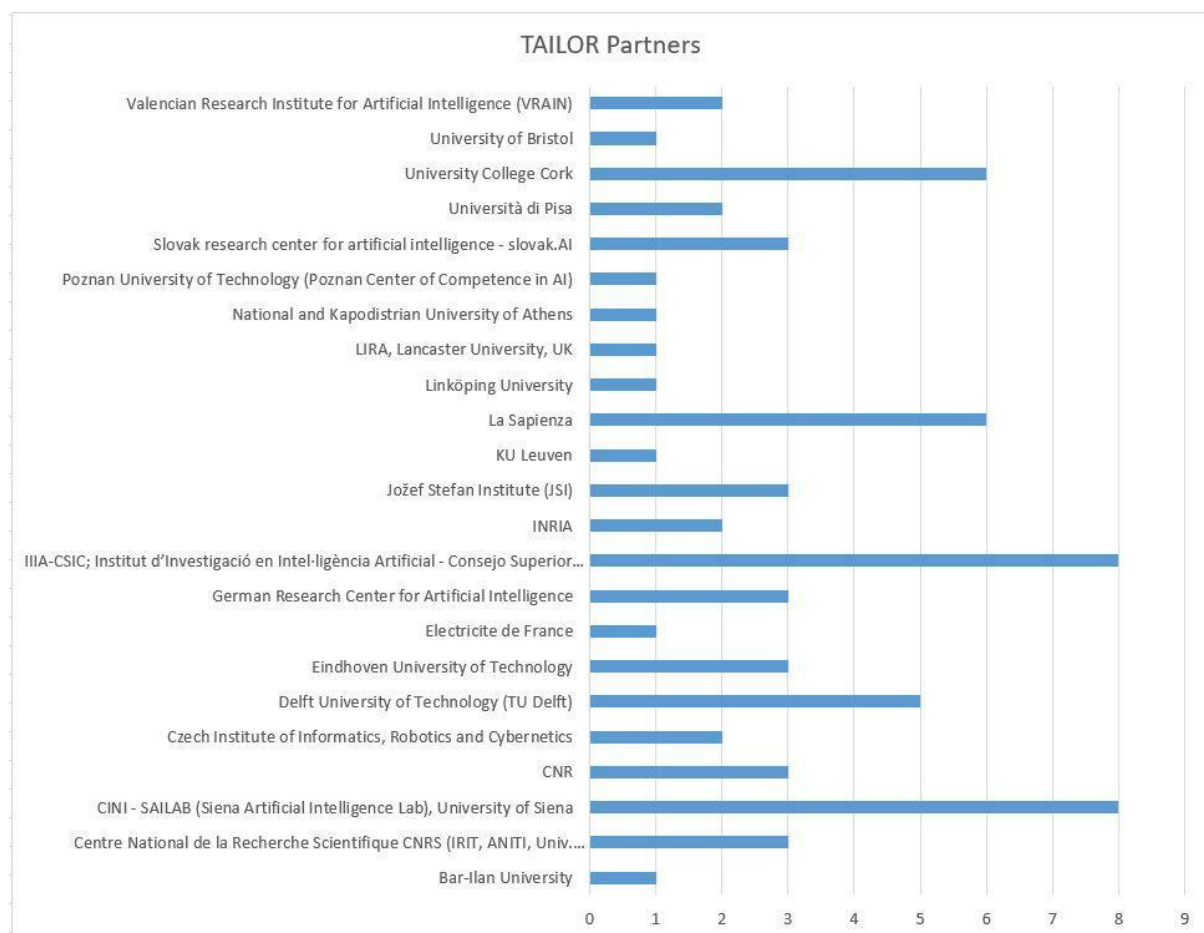
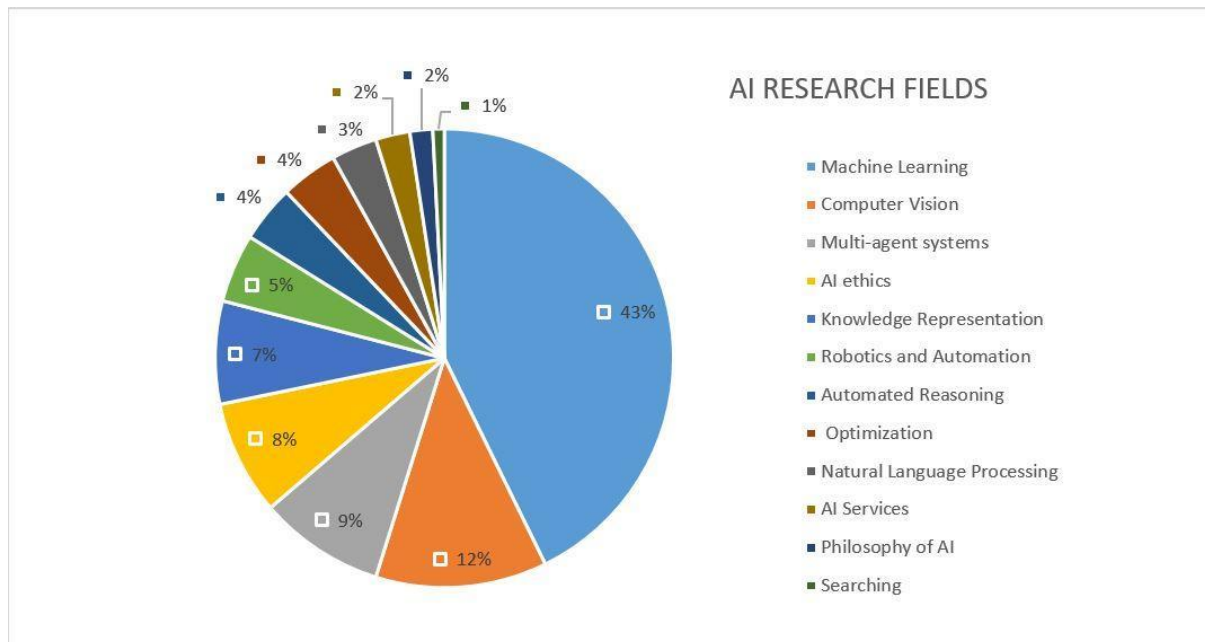
Here below some general data about the lecturers and participation.



	n	p
Spain	40	0,32
Italy	30	0,24
Netherlands	12	0,10
German	9	0,07
Ireland	7	0,06
France	7	0,06
Slovakia	3	0,02
Sweden	3	0,02
UK	3	0,02
Czech_R	2	0,02
Israel	2	0,02
Poland	2	0,02

Austria, Belgium, Greece, Portugal





## Lecturers

Here below some general data about the lecturers.

Name and Affiliation of the lecturers:

- [Fredrik Heintz](#). Associate Professor of Computer Science at Linköping University, SE.
- [Amos Yehuda Azaria](#). Associate Professor. Ariel University, Israel.
- [Neil Yorke-Smith](#). Associate Professor. Delft University of Technology, NE.
- [Farah Benamara Zitoune](#). Associate Professor of Computer Science at Toulouse University Paul Sabatier, France.
- [Loizos Michael](#). Associate Professor at Open University of Cyprus, Cyprus.
- [Clara Neppel](#). Responsible for the growth of IEEE's operations and presence in Europe, IEEE European Business Operations.
- [Carme Torras](#). Research Professor at the Robotics Institute (CSIC-UPC), Spain
- [André Meyer-Vitali](#). Senior researcher at DFKI, Germany.
- [Andreas Herzig](#). CNRS researcher at the Institut de Recherche en Informatique de Toulouse (IRIT) of Université Paul Sabatier, France.
- [Pompeu Casanovas](#). Director of Advanced Research, Professor of Law and Legal Studies at Autonomous University of Barcelona (Spain), and Adjunct Professor at La Trobe University Law School (Melbourne, Australia)

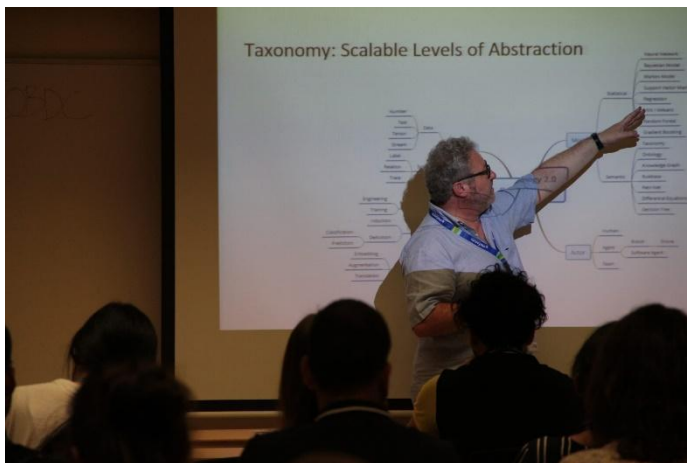
## Certificate of attendance





## Pictures posted on social media

These pictures are from the Twitter profile of Karina Gibert, the chair of the local organizing committee.









## Appendices

### 1: Grants and fees

TAILOR members had a special price as is seen in the table below. The fee was covered by the CSIC TAILOR funds for the summer school for up to three students per TAILOR partner.

#### Grants and fees

Find advantages on **registration fee** and **travel grants** to attend ACAI-Tailor 2022.

Registration fees	Early	Late	SuperLate
<b>ECAI or Tailor members</b>			
Student	250 €	300 €	415 €
Regular	350 €	400 €	515 €
<b>Other members</b>			
Student	290 €	340 €	455 €
Regular	400 €	450 €	565 €

### 2: Social programme

This is the resume of the social program that offered the school. It can be checked the social events in detail in this [link](#).

Social Program	Visit to the supercomputer MareNostrum <b>Care: LUNCH TIME</b>	Glass of Cava in the Torra Girona gardens (UPC)	Concert of Choir Llevant at Sant Pau del Camp church	Gala Dinner at Balmesiana Palace	Get Together Cocktail at the Indoor garden of Hotel Catedral (paid on its own)
----------------	--	---	--	----------------------------------	--

### 3: SubSift Summary

## **A pilot study to facilitate networking among attendees of a Summer School**

*Miquel Perello-Nieto, Peter Flach*

In preparation for the second TAILOR Summer School, TAILOR's Work Package 9 considered a set of AI-tools to encourage social connections between students. One of the objectives was to promote the TAILOR Connectivity Fund, which provides a budget for PhD students to visit research groups belonging to the TAILOR network of excellence, or for TAILOR members to visit labs out of the network. With only one week's notice for the call for voluntary participation in this study we got a 41% participation rate (70 out of 170 students).

We proposed to use two different AI-tools from TAILOR partners to match students with similar interests (SubSift [1]) and diverse personalities (SynTeam [2]).

SubSift [1] is a tool to match profiles represented as bags of words. It was originally used to match submitted publications for a conference (or journal) to a set of candidate reviewers based on their topics of expertise. In this study we requested from the students some free text indicating their topics of interest to generate their profiles. We performed term frequency-inverse document frequency (TF-IDF) including 1 to 3-grams, removing stop words from English and l2 norm. We then used k-means to divide the students into 4 clusters. The 5 most representative terms of each cluster were: (1) learning, deep learning, machine, deep, machine learning, (2) ai, language, natural, natural language, processing, (3) systems, mining, data, fuzzy, recommender systems, and (4) vision, computer, computer vision, music, neural networks.

SynTeam [2] is based on the Post-Jungian Personality theory for team composition [3], which has been shown to improve the performance of teams in various studies [2-9]. A set of 20 questions was completed by all participants which evaluates two sets of psychological functions: Sensing vs iNtuition (SN), Thinking vs Feeling (TF); and two sets of psychological attitudes: Perception vs Judgement (PJ), Extroversion vs Introversion (EI). Each of the 16 possible combinations forms a personality. Each of the previously defined clusters of students is then divided into smaller teams of 4 or 5 students with SynTeams, ensuring that there is at least one leading personality, and a mixed set of personalities and genders. A total of 16 teams were generated.

After the event we sent a form to the 70 participants to ask for their feedback, from which we obtained 23 answers (32%). The overall impressions appeared positive, but 6 of them had concerns about the transparency of the full process and the lack of communication channels, or common time and space slots to meet their peers. The lack of a specific space and time slot meant that 13 of them (56%) couldn't meet each individual member of their team, while 15 of them (65%) didn't manage to meet all the team together. This is a clear indication that setting a specific task, location and time may be necessary to motivate the team members to meet. Finally, a clear indicator of the success of this activity was that from 23 students that

provided feedback, 16 of them (69%) considered future collaborations using the TAILOR Connectivity Fund; 12 of those (75%) with students in their assigned teams.

The overall success of this pilot study has been really valuable to prepare a similar setting to perform a joint activity during the 2nd TAILOR Conference held in Prague, and will be improved and extended in additional academic events.

## References

- [1] Simon Price, Peter A Flach, and Sebastian Spiegler. SubSift: a novel application of the vector space model to support the academic research process. In Tom Diethe, Nello Cristianini, and John Shawe-Taylor, editors, *Proceedings of the First Workshop on Applications of Pattern Analysis*, volume 11 of *Proceedings of Machine Learning Research*, pages 20–27, Cumberland Lodge, Windsor, UK, 2010. PMLR.
- [2] Ewa Andrejczuk, Juan A Rodríguez-Aguilar, Carles Sierra, Carme Roig, and Yolanda Parejo-Romero. Don't Leave Anyone Behind: Achieving Team Performance Through Diversity. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, 2018.
- [3] John H Bradley and Frederic J Hebert. The effect of personality type on team performance. *Journal of Management Development*, 16(5):337–353, jan 1997.
- [4] J. Arnold and R. Randall. *Work psychology*. Pearson Education Limited., Harlow, England, 2010.
- [5] Larry K Michaelsen, Dean X Parmelee, Kathryn K McMahon, Ruth E Levine, and Diane M Billings. *Team-based learning for health professions education: A guide to using small groups for improving learning*. 2008.
- [6] M. K. Mount, M. R. Barrick, and G. L. Stewart. Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, (11):145–65, 1998.
- [7] M. A. West. *Effective Teamwork: Practical Lessons Learned from Organizational Research*. Wiley-Blackwell, West Sussex, 2012.
- [8] K. B. White. Mis project teams: An investigation of cognitive style implications. *MIS Quarterly*, (8(2)):95–101, 1984.
- [9] D.J. Wilde. *Post-Jungian Personality Theory for Individuals and Teams*. SYDROSE LP, 2013.