Matchmaking work at Bristol

Miquel Perello Nieto

12 July 2023

This project is funded by the EC
under H2020 ICT-48, GA 952215

Summary

- SubSift
- Peer review process for the ACM SIGKDD'09 data mining conference
- TAILOR partners ranking website
- Synergetic groups of students with similar interests
- Matching synergetic groups of researchers and tasks
- Conference networking suggestions from previous publications

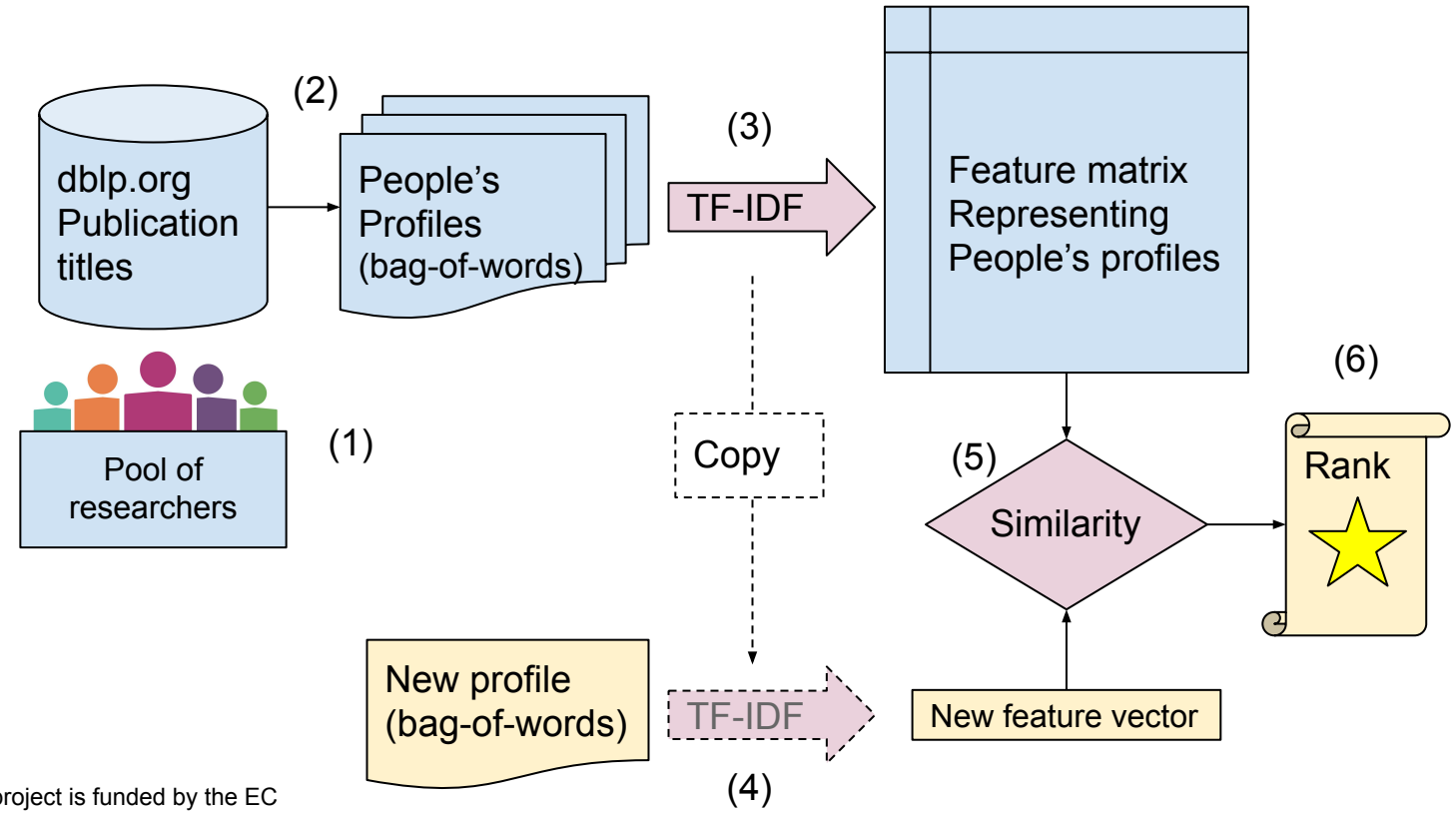
This project is funded by the EC
under H2020 ICT-48, GA 952215

SubSift

[Price et.al. 2010]

This project is funded by the EC
under H2020 ICT-48, GA 952215

SubSift [Price et.al. 2010]



This project is funded by the EC
under H2020 ICT-48, GA 952215

SubSift [Price et.al. 2010]

- Automatic matching of articles and reviewers
  - Articles/Topics are characterised by TF-IDF text representation
  - Reviewers by TF-IDF titles of their articles in the DBLP database
  - A similarity score is computed between topics/articles and reviewers (e.g. cosine similarity)
- Other applications:
  - Matching of people with similar interest
  - Matching of articles of similar topics
  - Grouping people in teams by topic, or similarities

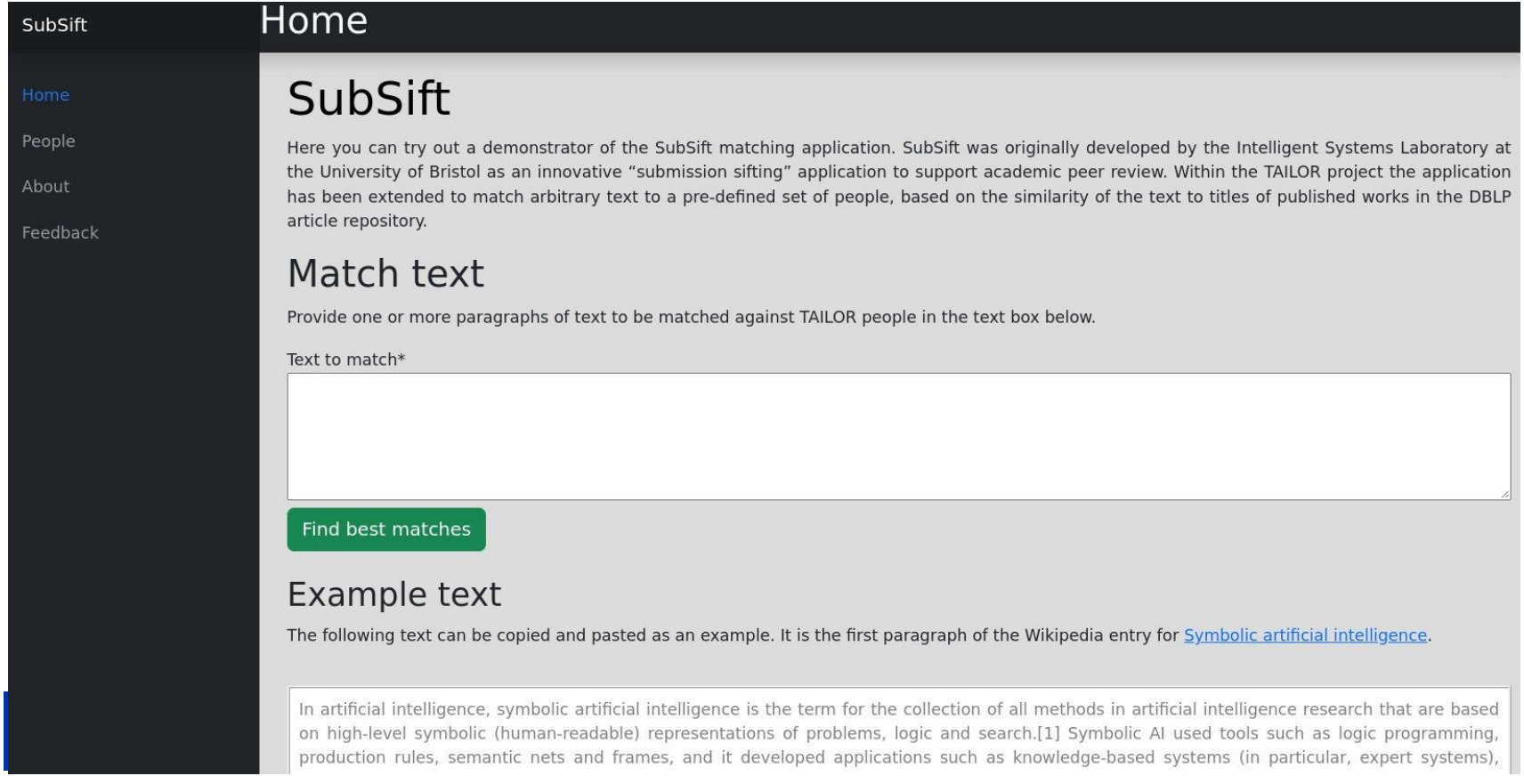
This project is funded by the EC
under H2020 ICT-48, GA 952215

TAILOR partners ranking website

This project is funded by the EC
under H2020 ICT-48, GA 952215

TAILOR
https://tailor-network.eu/

https://subsift-webapp.onrender.com/subsift/



Synergetic groups of students with similar interests

2nd ACAI-TAILOR Summer School, June 13-17, 2022

This project is funded by the EC under H2020 ICT-48, GA 952215

SynTeams

[Andrejczuk et. al. 2018]

This project is funded by the EC
under H2020 ICT-48, GA 952215

SynTeam [Andrejczuk et. al. 2018]

- Partition of students into subgroups of similar size and competence
- Individuals are evaluated by competence, personality and gender
- Basic student's definition
  - Unique ID:
  - Gender:
  - A personality: four personality traits {-1, 1} (See Jung, 1921)
    - (SN) Sensing / Intuition
    - (TF) Thinking / Feeling
    - (PJ) Perception / Judgment
    - (EI) Extroversion / Introversion
  - A set of competences: task related (e.g. skill, knowledge, attitude)

This project is funded by the EC
under H2020 ICT-48, GA 952215

TAILOR
https://tailor-network.eu/

Post-Jungian personality test

2 Psychological functions:
- **S**ensing vs i**N**tuition
- **T**hinking vs **F**eeling

2 Psychological attitudes:
- **P**erception vs **J**udgment
- **E**xtroversion vs **I**ntroversion

5 ternary questions per dimension (20 in total)

16 personality combinations

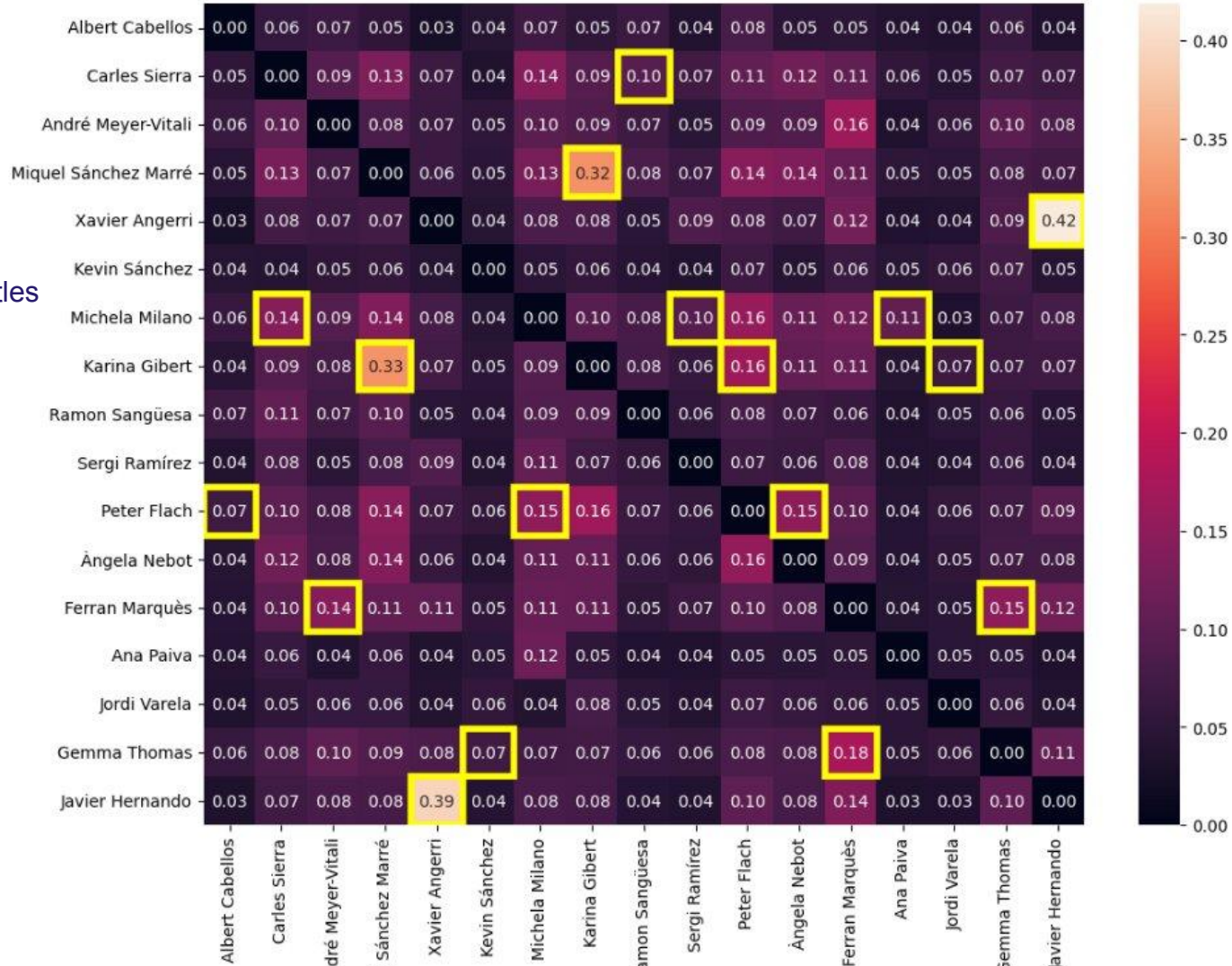
Facilitators: E_TJ

	.ST.	.SF.	.NF.	.NT.
I..J	ISTJ The Inspector	ISFJ The Protector	INFJ The Advocate	INTJ The Arcthitect
I..P	ISTP The Crafter	ISFP The Artist	INFP The Mediator	INTP The Thinker
E..P	ESTP The Persuader	ESFP The Performer	ENFP The Champion	ENTP The Debater
E..J	ESTJ The Director	ESFJ The Caregiver	ENFJ The Giver	ENTJ The Commander

This project is funded by the EC under H2020 ICT-48, GA 952215

Match people to people

- TF-IDF from publication titles
- Cosine similarity



This project is funded by the EC under H2020 ICT-48, GA 952215

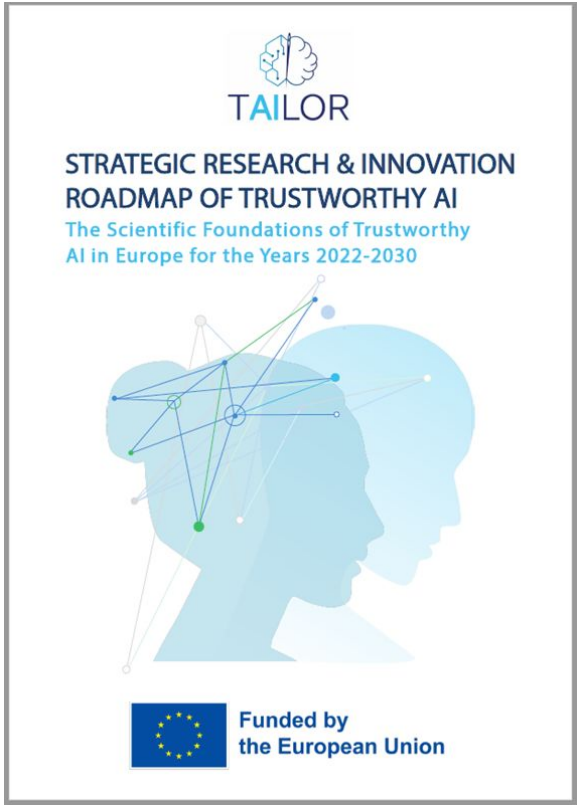
Matching synergetic groups of researchers and tasks

2nd TAILOR Conference
September 13-14, 2022
Prague

This project is funded by the EC
under H2020 ICT-48, GA 952215

The Strategic Roadmap, SRIR

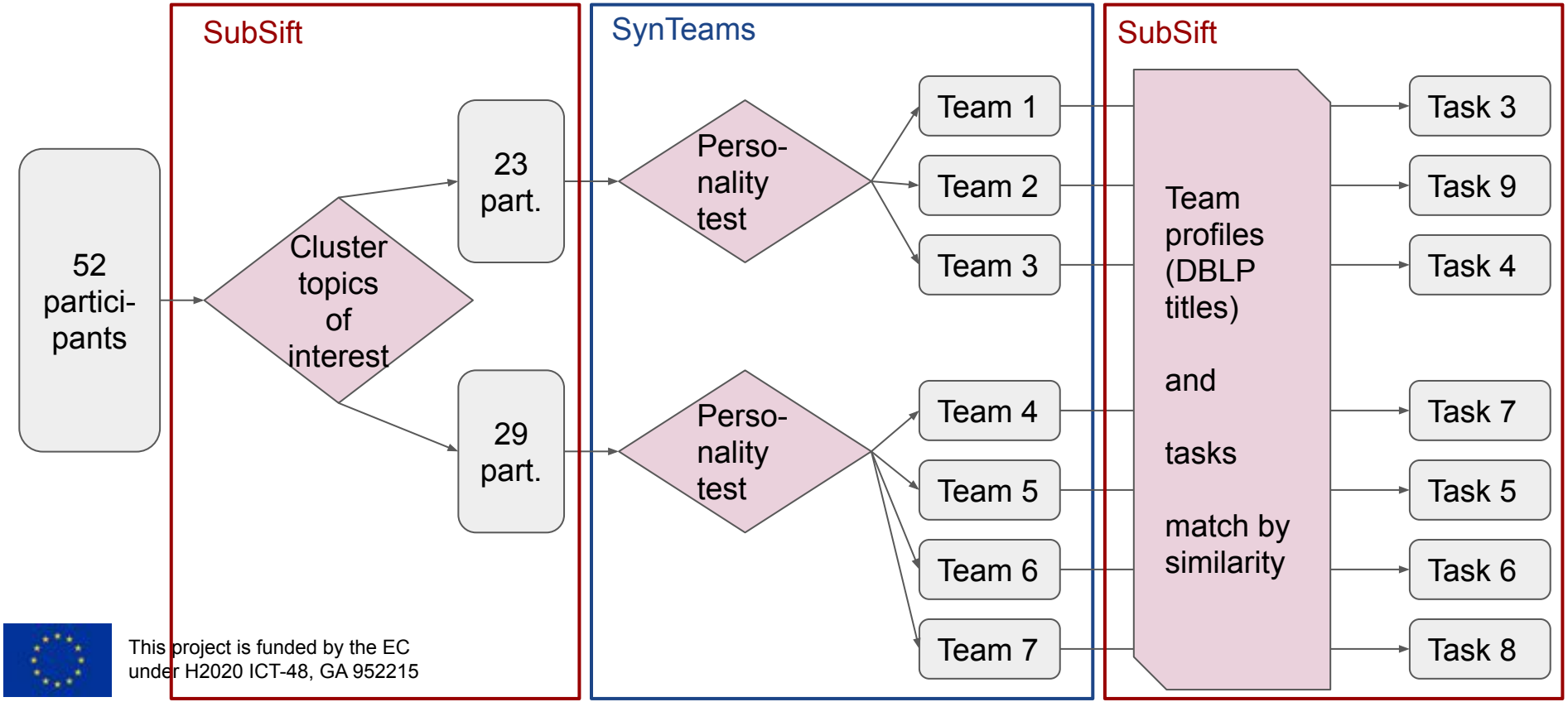
- Definition of the foundations of Trustworthy AI for the years 2022-2030
- Clear recommendations for future research on Trustworthy AI
- For researchers, policy makers and funding agencies
- The first version filed M23
- **<u>The Group work will contribute to version 2</u>**



This project is funded by the EC under H2020 ICT-48, GA 952215

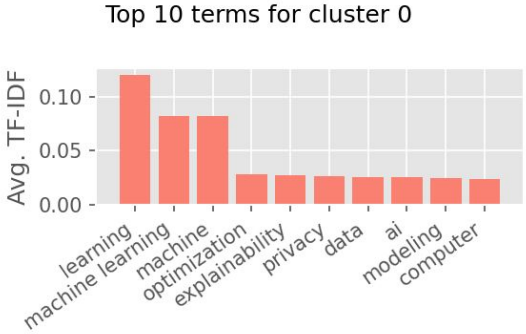
TAILOR
https://tailor-network.eu/

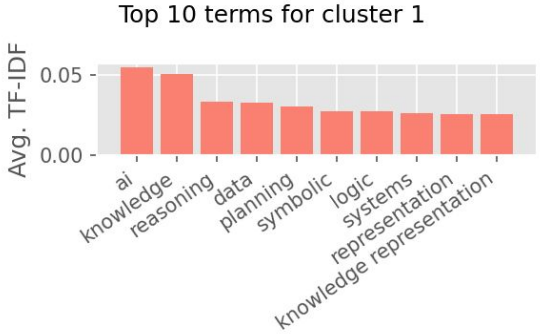
Task assignation overall pipeline

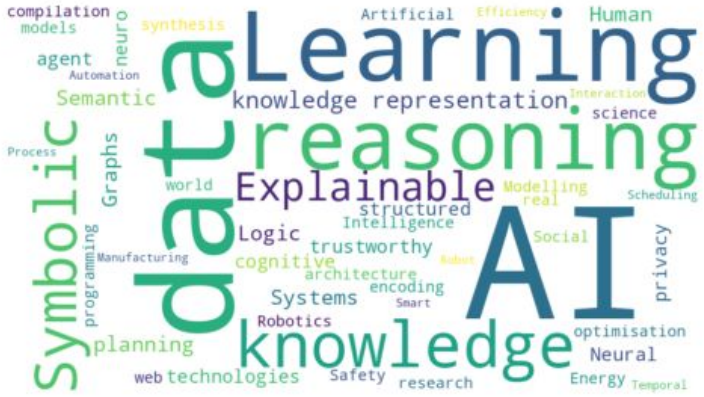


This project is funded by the EC under H2020 ICT-48, GA 952215

2 Clusters based on topics of interest







This project is funded by the EC
under H2020 ICT-48, GA 952215

23 participants

29 participants

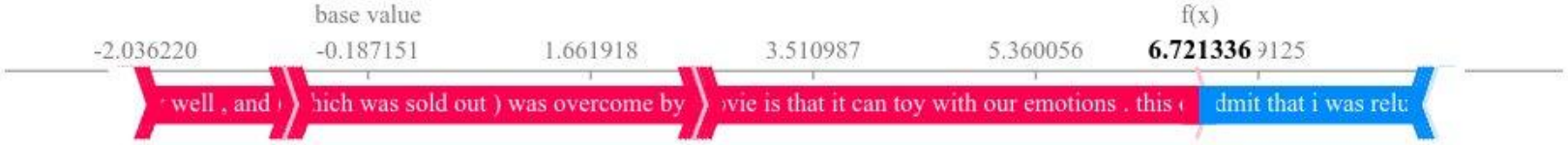
Conference networking suggestions from previous publications

3rd TAILOR Conference,
June 5-6, 2023, Siena

This project is funded by the EC
under H2020 ICT-48, GA 952215

SHapley Additive exPlanations

Example of a positive (red) vs negative (blue) sentiment classifier trained on movie reviews of IMDB.

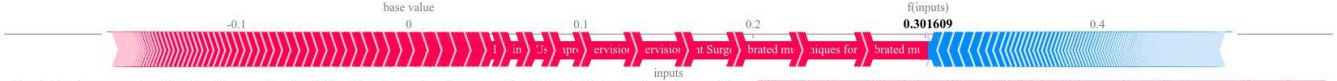


i went and saw this movie last night after being coaxed to by a few friends of mine . i ' ll admit that i was reluctant to see it because from what i knew of ashton kutcher he was only able to do comedy . i was wrong . kutcher played the character of jake fischer very well , and kevin costner played ben randall with such professionalism . the sign of a good movie is that it can toy with our emotions . this one did exactly that . the entire theater (which was sold out) was overcome by laughter during the

This project is funded by the EC
under H2020 ICT-48, GA 952215

Example of ranking and explanation

- Score 1.00 - Miquel Perello-Nieto
- Score 0.30 - Peter Flach
- Score 0.09 - Saso Dzeroski
- Score 0.08 - JERZY STEFANOWSKI
- Score 0.07 - Francesca Pratesi
- Score 0.06 - Stefano Teso
- Score 0.05 - Maria Bielikova
- Score 0.05 - Luciano Cavalcante Siebert
- Score 0.05 - Beniamino Di Martino
- Score 0.05 - Francisco Chicano
- Score 0.05 - Andrea Visentin
- Score 0.05 - Neil Yorke-Smith
- Score 0.04 - Joaquin Vanschoren
- Score 0.04 - Giuseppe De Giacomo
- Score 0.04 - Holger Hoos
- Score 0.04 - Flavio Lombardi
- Score 0.04 - Ana Paiva
- Score 0.04 - Andrea Orlandini
- Score 0.04 - Fredrik Heintz



Co-designing opportunities for Human-Centred Machine Learning in supporting Type 1 diabetes decision-making. When the Ground Truth is not True: Modelling Human Biases in Temporal Annotations. MIDI-Draw: Sketching to Control Melody Generation. Empirical Evaluation of Predictive Models: A keynote at ECIR 2022. FAT Forensics: A Python toolbox for algorithmic fairness, accountability and transparency. LIMESegment: Meaningful, Realistic Time Series Explanations. Understanding Reinforcement Learning Based Localisation as a Probabilistic Inference Algorithm. Self-Enhancer: A Self-supervised Framework for Low-Supervision, Drifted Data with Significant Missing Values. The Weak Supervision Landscape. The Weak Supervision Landscape. Simply Logical - Intelligent Reasoning by Example (Fully Interactive Online Edition). FAT Forensics: A Python Toolbox for Implementing and Deploying Fairness, Accountability and Transparency Algorithms in Predictive Systems. What and How of Machine Learning Transparency: Building Bespoke Explainability Tools with Interoperable Algorithmic Components. Efficient and Robust Model Benchmarks with Item Response Theory and Adaptive Testing. Multi-label thresholding for cost-sensitive classification. Co-Designing Personal Health? Multidisciplinary Benefits and Challenges in Informing Diabetes Self-Care Technologies. Human Activity Recognition Based on Dynamic Active Learning. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. Machine Learning Explanations as Boundary Objects: How AI Researchers Explain and Non-Experts Perceive Machine Learning. Continual Density Ratio Estimation in an Online Setting. You Only Write Thrice: Creating Documents, Computational Notebooks and Presentations From a Single Source. Risk Sensitive Model-Based Reinforcement Learning using Uncertainty Guided Planning. Classifier Calibration: How to assess and improve predicted class probabilities: a survey. Explainability Is in the Mind of the Beholder: Establishing the Foundations of Explainable Artificial Intelligence. Uni- and multivariate probability density models for numeric subgroup discovery. Modelling Patient Behaviour Using IoT Sensor Data: a Case Study to Evaluate Techniques for Modelling Domestic Behaviour in Recovery from Total Hip Replacement Surgery. FAT Forensics: A Python Toolbox for Implementing and Deploying Fairness, Accountability and Transparency Algorithms in Predictive Systems. One Explanation Does Not Fit All. Reflections on reciprocity in research. FACE: Feasible and Actionable Counterfactual Explanations. Model-Based Reinforcement Learning for Type 1 Diabetes Blood Glucose Control. Explainability fact sheets: a framework for systematic assessment of explainable approaches. Polsar Image Classification via Robust Low-Rank Feature Extraction and Markov Random Field. One Explanation Does Not Fit All: The Promise of Interactive Explanations for Machine Learning Transparency. LIMEtree: Interactively Customisable Explanations Based on Local Surrogate Multi-output Regression Trees. Bypassing Gradients Re-Projection with Episodic Memories in Online Continual Learning. Towards Faithful and Meaningful Interpretable Representations. Model-Based Reinforcement Learning for Type 1Diabetes Blood Glucose Control. A Big Data platform for smart meter data analytics. Setting decision thresholds when operating conditions are uncertain. An application of hierarchical Gaussian processes to the detection of anomalies in star light curves. Performance Evaluation in Machine Learning: The Good, the Bad, the Ugly, and the Way Forward. Counterfactual Explanations of Machine Learning Predictions: Opportunities and Challenges for AI Safety. Desiderata for Interpretability: Explaining Decision Tree Predictions with Counterfactuals. Sβ^3$-IRT: A New Item Response Model and its Applications. Distribution calibration for regression. Beyond temperature scaling: Obtaining well-calibrated multi-class probabilities with Dirichlet calibration. β Distribution Calibration for Regression. HyperStream: a Workflow Engine for Streaming Data. FAT Forensics: A Python Toolbox for Algorithmic Fairness, Accountability and Transparency. FACE: Feasible and Actionable Counterfactual Explanations. Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with Dirichlet calibration. bLIMEy: Surrogate Prediction Explanations Beyond LIME. Explainability Fact Sheets: A Framework for Systematic Assessment of Explainable Approaches. Introduction to the special issue on Data Science in Europe. A Comprehensive Study of Activity Recognition Using Accelerometers. Activities of Daily Living Ontology for Ubiquitous Systems: Development and Evaluation. Anomaly detection in star light curves using hierarchical Gaussian processes. Analysis of Patient Domestic Activity in Recovery From Hip or Knee RePlacement Surgery: Modelling Wrist-worn Wearable RSSI and Accelerometer Data in The Wild. The Facets of Artificial Intelligence: A Framework to Track the Evolution of AI. Conversational Explanations of Machine Learning Predictions Through Class-contrastive Counterfactual Statements. Glass-Box: Explaining AI Decisions With Counterfactual Statements Through Conversation With a Voice-enabled Virtual Assistant. Releasing eHealth Analytics into the Wild: Lessons Learnt from the SPHERE Project. Non-Parametric Calibration of Probabilistic Regression. Computational support for academic peer review: a perspective from artificial intelligence. Unsupervised learning of sensor topologies for improving activity recognition in smart environments. Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers. The Role of Textualisation and Argumentation in Understanding the Machine Learning Process. Classifier Calibration. First-Order Logic. ROC Analysis. Probabilistic Sensor Fusion for Ambient Assisted Living. CASP-DM: Context Aware Standard Process for Data Mining. Reframing in context: A systematic approach for model reuse in machine learning. On the need for structure modelling in sequence prediction. Cost-sensitive boosting algorithms: Do we really need them? Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data. Declaratively Capturing Local Label Correlations with Multi-Label Trees. Active transfer learning for activity

This project is funded by the EC under H2020 ICT-48, GA 952215