



TAILOR

Foundation of Trustworthy AI:
Integrating Learning, Optimisation and Reasoning

IMPACT STORIES

from the TAILOR Network of Excellence
Centres in AI



**Funded by
the European Union**

The TAILOR Project Partners

Linköping University, Consiglio Nazionale delle Ricerche, Institut national de recherche en informatique et en automatique, University College Cork, KU Leuven, La Sapienza, Leiden University, University of Lisbon, AI and Machine Learning Group, Universitat Pompeu Fabra, Alma Mater Studiorum – Università di Bologna (UNIBO), Bar-Ilan University, Eindhoven University of Technology, Centre National de la Recherche Scientifique, Jožef Stefan Institute, TU Darmstadt, University of Bristol, University of Freiburg, University of Oxford, University of Trento, Vrije Universiteit Brussel, Charles University, Commissariat à l’Energie Atomique et aux Energies Alternatives, Université d’Artois, Czech Institute of Informatics, Robotics and Cybernetics, Delft University of Technology, German Research Center for Artificial Intelligence, Ecole Polytechnique Fédérale de Lausanne, Fondazione Bruno Kessler, Fraunhofer IAIS, Graz University of Technology, Institut d’Investigació en Intel·ligència Artificial – Consejo Superior de Investigaciones Científicas, Lancaster University, National and Kapodistrian University of Athens, NEO: Networking and Emerging Optimisation, Poznan University of Technology, RWTH Aachen Center for Artificial Intelligence, Siena Artificial Intelligence Lab, University of Siena, Slovak research center for artificial intelligence, Netherlands Organisation for applied scientific research, Università di Pisa, Université Grenoble Alpes, University of Basel, AI Research Group, Valencian Research Institute for Artificial Intelligence, Volkswagen Group Machine Research Lab, Engineering – Ingegneria Informatica SPA, Tieto Finland Oy, Philips Electronic Nederland B.V., Electricite de France, Imperial College of Science Technology and Medicine, ZF Friedrichshafen AG, Luxembourg Institute of Health, Centraal Bureau voor de Statistiek, Bosch, ABB AB.

...and more than 120 network members!

Image credits

Pages 6 (top) and 19: Lena Tasse, Linköping University
Pages 8-9: Charlotte Perhammar, Linköping University
Page 12: Giovanni Casini, CNR-ISTI
Page 14-15: Kinit
Pages 16 and 22 (bottom): Anna Nielsen, Linköping University
Page 17: Website Governo da República Portuguesa
Page 20: GHGSAT
Page 22 (top): Per Wistbo Nibell, Linköping University
Cover: CK - stock.adobe.com

Funding

TAILOR is an ICT-48 Network of AI Research Excellence Centres funded by EU Horizon 2020 research and innovation programme under grant agreement No 952215.

Contents

Using AI competitions to advance science and attract talent.....	4
Trustworthy AI in Education	6
Standardisation to Allow AI Access to Critical Services	10
The TAILOR Handbook of Trustworthy AI.....	12
Regional AI research and innovation for talent Attraction and democracy.....	14
Factual and Trustworthy Contributions to European Policy making	16
AI-Powered Collaboration Tools	18
Safe and powerful AutoAI solutions amplify future research and innovation in many domains.....	20
Shaping the Future of AI within the EU	22

Using AI competitions to advance science and attract talent

Humans like to compete, not only in sports but also in computer science. This makes data competitions an excellent way of reaching very diverse goals and interest groups. The universal language of competitions allows to showcase what is possible and to create incentives for applicable and experimental research. It is also used to attract talent.

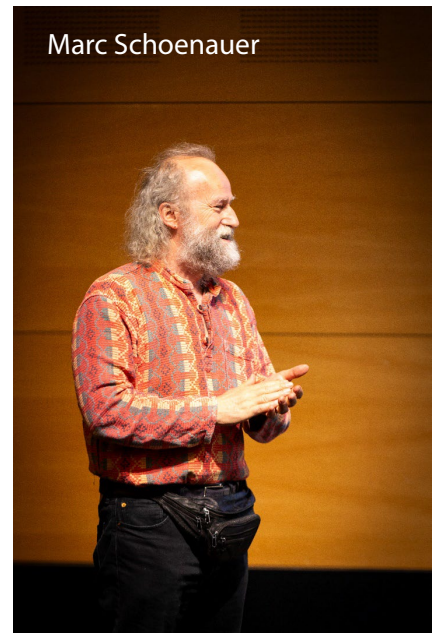
The competitions put forward by the TAILOR network cover various fields from charging station infrastructure to neuroscience, controlling the power grid, developing numerical twins or creating stable fusion reactors. An example of such a TAILOR challenge was the Smarter Mobility Data Challenge, developed in partnership with the industrial groups behind the Manifeste IA* together with INRIA Saclay scientists Marc Schoenauer and Sebastien Treguer.

– We believe that this most human of instincts can help further the development of AI. Since the dawn of time, we have been driven to compete. From the Olympic Games to the original space race and on to hot dog eating contests and now machine learning, we have always enjoyed competing and it has made us all better, says Marc Schoenauer.

The Smarter Mobility Data Challenge was open to all European students and had them develop machine learning models that could forecast the states of car charging stations around Paris, at several levels of geographical resolutions, based on historical data. The competition sat at the intersection of public interest, industry and academia and was meant to

showcase the possibilities of AI when it comes to making any complex system more efficient. The participants were asked not only to deliver access to their code but also explain their approach and explainability of the final model was taken into account by the jury. The best teams received job propositions by some of the industrial partners of the “Manifeste IA”.

AI competitions are used by participants to learn more applied knowledge by practice, gain experience, develop their skills and sharpen their tooling, build a team for future projects, and in some cases make a name for oneself within the machine learning and AI space. Industry uses them as equal parts to evaluate the benefits they can derive from AI applied to their real-world problem, develop proofs of concept, new talents into the field. Scientists use competitions to federate participants from various backgrounds to collaboratively explore new ways to advance science and solve various problems whether on fundamental research subjects or more applied topics and share new common building blocks, such as scientific papers and open-source code.



Marc Schoenauer

– It allows us to crash test, optimise, and mitigate any threats by finding any biases, weaknesses, hallucinations, or inconsistencies that may arise. Being able to do that before we apply these algorithms on actual services or critical applications adds a level of scrutiny and security, explains Marc, noting the value of competitions beyond the added recognition and excitement.

The possibilities of AI are truly endless, as all things are part of complex systems and if we have the capability to find new patterns that make these systems more efficient, we can better utilise our common resources.

– I can see this technology having a profound impact on society, and as long as we can make sure that the system is trustworthy and that it is aligned with our common goals, I think it will add enormous value. Getting there and finding the best algorithms will take some trial and error though, and I think competitions are the perfect place to do so, concludes Marc. ■

CONTEST PLATFORMS

Kaggle and CodaLab are machine learning competition platforms and communities. While Google-owned Kaggle is the most famous, CodaLab is a much-appreciated open-source alternative. University Paris-Saclay and the joint LISN/Inria team TAU are community lead of Codalab platform. Most TAILOR competitions are hosted on CodaLab. According to the independent organisation MLcontest, CodaLab was the leading contest platform in 2023.

<https://mlcontests.com/state-of-competitive-machine-learning-2023/#platforms>

About the researchers:

Marc Schoenauer is Principal Senior Researcher at INRIA, the French National Institute for Research in Computer Science and Control, and Deputy Research Director of AI.

Sebastien Treguer is a Research Engineer at INRIA Saclay.



** The "Manifeste IA" is a strategic document signed by the French Ministry of Economy and Finance and eight French large corporations (Air Liquide, Dassault Aviation, EDF, Renault, Safran, Thales, Total, and Valeo). The text is part of the AI for Humanity strategy initiated by the French Government.*

Trustworthy AI in Education

University teachers are shaping the hearts and minds of the next generation of AI researchers. It is important that PhD students in computer science and AI learn about the results of TAILOR researchers in the domain of trustworthy AI. Researchers Peter Flach and Miquel Perello-Nieto have been leading a team of researchers collecting a set of courses to cover the topic. The resulting curriculum will be free to use by any academic institution that needs guidance when developing coursework and classes for trustworthy AI.



The TAILOR curriculum promotes the teaching of trustworthy AI. It is a program that covers all the necessary topics for future professionals of AI.

– The number of AI models, methods and research has grown very rapidly in the latest decade, which makes it necessary to map AI solutions. Furthermore, AI tools will be used in critical applications that require auditability, explainability and interpretability, and need to be covered in PhD programs, says Flach, elaborating on why there is a need to coordinate these efforts.

The different courses and tools being developed have been tested at the Tailor Network summer school on three occasions. Over 800 participants have tested and provided valuable feedback which has been used to update and finetune the curriculum.

– Standardizing the educational material will facilitate the movement of students across Europe, fueling interdisciplinary collaboration of European institutions and advancing the research efforts in AI. With the rapid advancements in the field, this also needs to be a continuous cooperative process, as new courses get added to the curriculum and complementary ones get developed by different schools and universities. Ideally, it can become a toolkit and a creative commons for trustworthy AI education, continues Flach looking towards future implementations.

– Given enough adoption of the curriculum and this framework for trustworthy AI development, there is also the possibility to create specific certifications for those that have completed enough of the coursework. But in order to build this out and



Peter Flach

implement it properly, we need more resources, explains Perello-Nieto, sharing his hope for a continuation.

In truly collaborative spirit, the work that was done within the TAILOR Network was also extended and adopted as part of the AIDA project (the International AI Doctoral Academy). Two different curricula were developed, a PhD

curriculum in Trustworthy AI, which focuses on foundations of AI and advanced topics, and a second one, to show how these courses could also be used as a part of the more general curriculum in AI.

– A driving force behind this collaboration was Barry O’Sullivan, part of the steering committee of the VISION

consortium which is building a network to support world-level AI education, that has developed AIDA. These efforts encompass the networks TAILOR, Elise, AI4media, and AI NET, explains Flach and Perello-Nieto, emphasizing how much further we can get when working together. ■



About the researchers:

Peter Flach has been a Professor of Artificial Intelligence at the University of Bristol since 2003, and is the leader of Work Package 9 Network Collaboration in the TAILOR project.

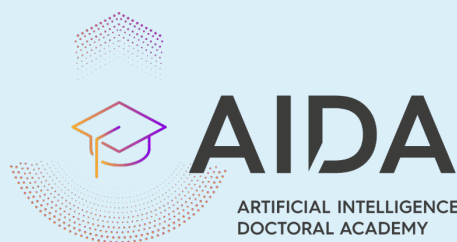
Miquel Perello-Nieto has been a Research Associate at the University of Bristol since 2015, and is a researcher of Work Package 9 Network Collaboration in the TAILOR projec

AIDA AND THE NETWORKS OF EXCELLENCE CENTRES IN AI AND ROBOTICS

The Networks of Excellence Centres in AI and Robotics includes nine projects: AI4Media, ELISE, HumanE-AI-Net, TAILOR (since 2020), ELSA, euROBIN (since 2022) and ELIAS, dAIEDGE, ENFIELD (since 2023).

Barry O’Sullivan is part of the steering committee of the VISION consortium which is building a network to support world-level AI education and has developed AIDA.

The International AI Doctoral Academy (AIDA) was created by the four ICT-48 networks (AI4Media, ELISE, HumanE-AI NET, TAILOR) and the VISION project, to offer access to knowledge and expertise and attracting PhD talents in Europe. AIDA offers courses, lectures and summer schools. Read more on www.i-aida.org

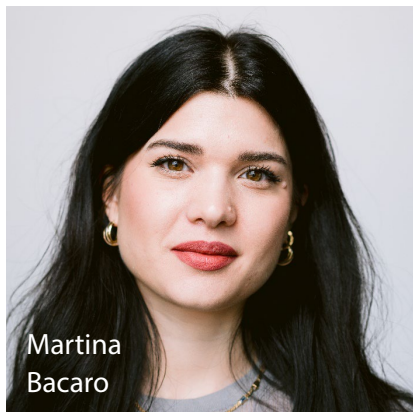


Creating links between European AI researchers of tomorrow

The concept of robotic workers in our home and workplace isn't new. We've all wished we had a robot vacuum cleaner, seen grass cut automatically, interacted with some level of autonomy in vehicles, or used voice recognition to play the next song. All with varying degrees of success, and varying levels of anthropomorphizing.

The next step forward in AI and robotics will merge all previous iterations of robots in our surroundings and bring about a giant leap in their capabilities. But as the technology is getting there, we also need to figure out how to navigate this new reality where robot and human interactions will be a natural occurrence in every part of society.

To avoid unintended consequences in human robot interactions, researchers strive to figure out what is known as the "attribution of intentionality". Which means determining with what certainty that we can predict what an autonomous agent (robot or human) will do as their next action. Spearheading these efforts within the TAILOR Network are Martina Bacaro and Tom Ziemke, whose research sits at the intersection between autonomy, intentionality, interaction, agency, and ascribed and perceived trust.



Martina Bacaro

– Trustworthy AI is a big thing in the TAILOR network, and we believe it will become just as important in the robotics space. AI and robotics are of course intrinsically linked, as many robots will be fitted with the latest AI software. However, when adding a mechanical layer, some of the trust sits with the appearance as well as with the fluidity and humanness of movement, says Martina Bacaro, highlighting the

additional challenges of autonomous agents in a real-world environment.

Bacaro was one of many young European AI researchers who were awarded funding for research mobility through one of TAILOR's funding instruments. Scientific mobility, especially early career mobility, has been shown to improve academic performance. Mobile scholars tend to publish in higher impact publications, take more risk and establish new scientific trajectories.

– I think Martina's line of study is crucial for the future. We need to create a better understanding of humans and robot interactions as they will only become more frequent. And if that means formulating a few laws of robotics, we'll have a pretty good place to start, says Tom Ziemke.

Currently there is a divide between those who ascribe too much capability to AI systems and robots, and those who don't believe at all. But once the presence of robots becomes ubiquitous, there needs to be a way to bridge this divide. Clear declarations of what level of autonomy each robot or chatbot provides can help manage expectations, make outcomes more predictable and ultimately build trust over time.

– Going forward, no one will be able to avoid interacting with robots. What we are trying to determine is how to build trust, how people ascribe intentionality

and how they choose to interpret them, says Ziemke, clarifying that it's a matter of when, not if.



Martina's line of study is crucial for the future. We need to create a better understanding of humans and robot interactions as they will only become more frequent. And if that means formulating a few laws of robotics, we'll have a pretty good place to start.

TOM ZIEMKE

A well-known example is the uncanny valley. It's the unsettling feeling towards something that can almost pass as human, except for a lingering notion that something is slightly off. This can happen with robots, 3D computer animations and even when just chatting. Here there are two ways to go, to make robots perfectly human, or to make them obviously robotic.

TAILOR FUNDING FOR SCIENTIFIC MOBILITY

2 funds: Connectivity Fund & Collaboration Exchange Fund
 1,7 M€ total budget
 84 research visits (~20 % female)
 17 workshops





– I would be very excited to continue my research in a post-doc, and my hope would be to work with an android and look at whether “to err is human”, and if this implies a requirement for us to interact in a completely “human-to-human way” with robots. And to find out

if it truly is our flaws that makes us human, says Bacaro.

Looking to the future, automated autonomous agents will do our taxes, work in our factories, take care of our dirty laundry, drive our vehicles, do deliveries, educate us and be our social companions.

– When that future arrives, we want to understand when, where and why we trust these agents, and whether we always should, say Bacaro and Ziemke, explaining the profound implications the technology has. ■

About the researchers:

Martina Bacaro, PhD student in Philosophy, Science, Cognition, and Semiotics at the University of Bologna.

Tom Ziemke, Professor of Cognitive Systems in the Department of Computer & Information Science at Linköping University.

THE UNCANNY VALLEY

An unsettling feeling people experience when robots closely resemble humans in many respects but aren't quite realistic.

Standardisation to Allow AI Access to Critical Services

Recent developments in the field of AI show great potential for the future of our society. However, to fulfil this great promise, we must be able to apply AI to our most constrained and critical services. But to do so, we must also make sure that we have true alignment with our goals and values, and that we can standardise technical requirements for safety, robustness and resilience against attacks.

When applying new technology to critical applications such as healthcare, the power grid or climate prediction models, we need to understand how the AI agent is generating its output. We also need to know which dataset it is being trained on – as to not put our trust in black box model. Championing these efforts within the TAILOR Network are André Meyer-Vitali and Chokri Mraidha, who are creating a standardised framework for evaluating models and output for AI systems. The standardisation efforts are orchestrated by CEN & CENELEC, two European Standardisation Organizations mandated by the European Commission to develop standards for AI trustworthiness to support the AI Act.

The EU AI Act was introduced to make sure that models that are being deployed in the EU are aligned with our goals and values through legislation. However, how to apply the legislation, how to determine which requirements need to be met, and how to apply them to different types of models is still up for debate. There are a few different approaches to consider:

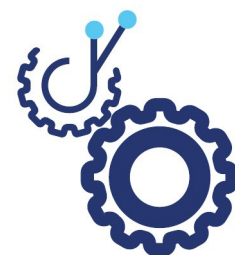
– Using neurosymbolic models would provide easier validation, better transparency, and accountability whereas causal models give us causality and allow us to determine why the model reached a certain output. Just relying on bigger datasets and compute scaling seems too uncertain and unreliable, says André Meyer-Vitali, detailing some of the differences and trade-off between different models.

The recent discussions and progress have created both a lot of hype, but also awareness of the issues with trust in AI systems. However, it has resulted in two camps where there is usually either exaggerated enthusiasm or exaggerated fear. And neither is good.

– We need objective measures, and standardisation is vital to build a collective understanding of AI systems for certification. And then we need to sort out what this means in a legal sense, as terms such as transparency, privacy, fairness, robustness, accountability, etc. need to be specified and measured, explains Chokri Mraidha, outlining the multi-step process of getting to trustworthy AI.

The potential gains are enormous. Getting greater efficiency in every form of critical service would mean huge savings financial, environmental, and in terms of human resources. Europe will also be able to stay competitive and it would help alleviate much of the current concerns when it comes to providing public services.

– We hope that the framework for evaluation that we are developing will lead to a wide adoption of trustworthy AI systems in industry and the public sector. And that the gains from doing so will be spread evenly amongst the population, providing better services and a better society for all, concludes André Meyer-Vitali and Chokri Mraidha, reiterating what is at stake. ■



About the researchers

Dr. André Meyer-Vitali, Senior Researcher at DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz) and Lead Scientist / Principal Investigator at Centre for European Research in Trusted AI.

Dr. Chokri Mraidha, Director of Research at CEA LIST institute and Head of the Embedded and Autonomous Systems Design Laboratory at CEA LIST.



André Meyer-Vitali

Chokri Mraidha

EU standardisation bodies

European standards are developed through one of the three European Standards Organisations: the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC), and the European Telecommunications Standards Institute (ETSI).



The TAILOR Handbook of Trustworthy AI

The possibilities and promises of AI technology might seem endless, and it's easy to get blinded by the pace of development and visions of a utopian future. We will need massive technical breakthroughs, ethical considerations as well as the development of rules, guidelines, definition and legal requirements before we allow algorithms to access all corners of our life.



The main question, as identified by the EU, is that of trustworthiness. In pursuing trustworthiness for AI systems, it gives us an opportunity to create universal definitions for privacy, safety, fairness, transparency, sustainability, accountability, reproducibility, justice, and equity within the context of AI. With an agreement as to what these terms mean, there is also the possibility to create a standard to which AI systems can be held. This would allow us to know that the output can be trusted, and that the model will let us know when it doesn't know the answer, as opposed to just trying to make up a reasonable sounding one.

At the forefront of these considerations, we find Francesca Pratesi and Umberto Straccia. Together they are leading the team of editors behind the TAILOR Handbook of Trustworthy AI.

The handbook provides an overview of ethical dimensions in the definition and deployment of AI systems, and while this is specifically geared towards the scientific community, it is equally important for the general public. It will allow those that are interested can have a common ground and a common understanding of what is being discussed. Having a defined terminology also makes possible to make relevant comparisons. When adding requirements and criteria for trustworthiness, this

provides an opportunity to hold developers accountable, and to rank different models.

Research in this domain is taking huge steps forward at the moment, and in addition to that, new legal requirements coming from the EU AI Act are now being finalized. These will need to be accommodated and incorporated in the Handbook as well. Given the pace of AI innovation, the Handbook is a living document that will extend and redefine itself to accommodate the development of new models and new technical approaches.

– Aside from the technical and legal requirements we also think that there is

more ethical consideration that we want to add, and we would like to extend the work that has been done on the environmental impact of AI, says Francesca Pratesi, emphasizing the need to take a truly holistic approach to AI.

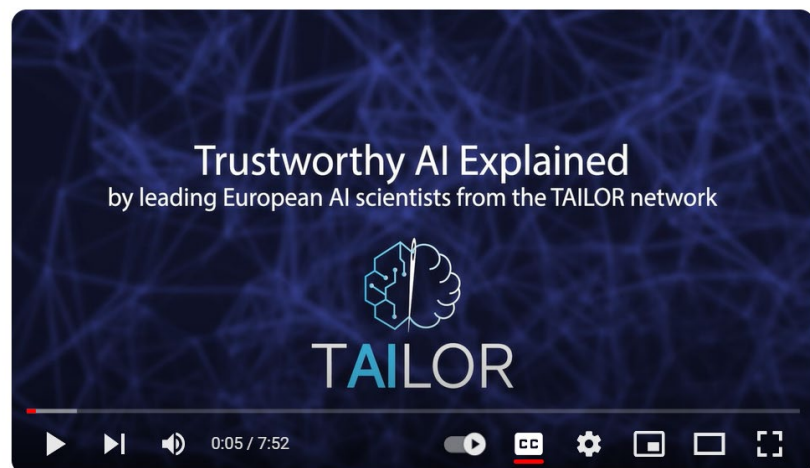
The successful implementation of the handbook would mean the adoption of it as a de facto standard with the AI space. It is a work that was inspired by, and implemented, because of the EU AI Act. The handbook has the potential to hold great significance in the implementation of the AI act on the European and national level. In addition, the handbook provides value for academia, as a guideline for evaluating models. For policymakers and the general public, the handbook can be used as a tool for understanding how models work and what they need to be tested on.

– The handbook has the potential to create better understanding, promote discussions and stimulate critical thinking about AI in all parts of society, says Pratesi.

The content of the handbook also matters for those that are currently

developing new AI models. If they know what the rules and requirements are, then they can adjust accordingly. And it is important that there is competition within this space, so not everything ends up being run by a single model developed by a single company.

– If we can build this framework, it would create a blueprint for building models that can be trusted to benefit all of society, says Umberto Straccia, underlining exactly what is at stake here.



The main concepts discussed in the Handbook are also available as short videos in the “Trustworthy AI Explained” series on YouTube.

About the researchers:

Francesca Pratesi and Umberto Straccia are researchers working at the Institute of Information Science and Technologies of Pisa. The institute is part of CNR (the Italian National Research Council), which is the largest public research organization in Italy.



THE TAILOR HANDBOOK OF TRUSTWORTHY AI

The TAILOR Handbook of Trustworthy AI is freely available online. Part of the material that was developed for the handbook has also been used to enrich Wikipedia.

Regional AI research and innovation for talent Attraction and democracy

To ensure that all members and citizens of the European Union can share the profits generated by deploying AI systems, there is also a need to democratize and decentralize the access to education, research and AI resources. Not only will this make sure that we can catch any talents no matter where they are from, but also that the algorithms being developed are being done so with local considerations in mind, making sure that we get a diversity of models and reasoning going in to them.

One such initiative that has proven greatly successful is the newly founded first independent research institute in Slovakia, the Kempelen Institute of Intelligent Technologies (KInIT), with founder and Director General Maria Bielikova at the helm. Bielikova conducts research focusing on human-computer interaction analysis, user modelling and personalization. Before founding KInIT, she was employed at the Slovak University of Technology for more than 30 years.

Being part of the TAILOR Network has a number of benefits, including collaborators and stakeholder within the EU, as part of the scope of the initiative is to make sure to gather diverse perspective from all over Europe. But for people to have both an informed view on how to shape the future of AI and the technical skills and understanding to participate in doing so, educational institutions and resources must also be locally available no matter where someone grows up or goes to school.



Our vision for the future is one where we feed and promote innovation with Slovakia, build unicorns and create exciting opportunities.

MARIA BIELIKOVA



To make sure that Slovakia and Eastern Europe are no exceptions, Bielikova used the know-how from the TAILOR Network as a catalyst. By making sure that the application-driven basic research being done is conducted in close collaborations with the private sector within the region, Bielikova secured the funding that allowed her to start KInIT. The research institute makes sure that local talent has an opportunity to stay and still get to conduct cutting edge research with an international outlook, whilst also tackling local problems in its application.

– Having spent most of my career in academia, I knew the power that comes from being able to bring excellent science to business in order to leverage their resources in a way that ultimately benefits both worlds. So, when I decided to start an independent research institute I wanted to do so in close collaboration with the private sector from the beginning”, says Bielikova, explaining the

benefits and opportunities that basic research in industry demanding challenges get from having a business partner.

Extending collaboration to the whole research and innovation ecosystem is key. One area that already brings different business partners together is language. KInIT has launched an initiative to develop large language models for the low resource Slovak language. Even though the companies are competitors, they all see the benefit of having locally developed and fine-tuned AI models.

– Making sure that we secure funding for basic research, which then get practical implications locally, both add value back to the private sector but also gives us use cases that get attention and can attract new talent to the field. This talent may otherwise have left the region to pursue a career elsewhere, says Bielikova, explaining the importance of both providing interesting opportunities and

advancing the regions research capabilities.

The fundamental idea of bringing excellent science to industry has proven very successful. From its inception in 2020, aided in part by the TAILOR Network that provided credibility, networking opportunities and part of the funding, KInIT has grown from 20 to 60 people, including 10 PhD students, in the span of just 3,5 years. It has grown to

encompass almost 30 collaborators within industry, two industry PhDs, and getting 20 % of its funding from there, and has also emerged as the preferred Slovak AI partner in its domain of expertise.

– Our vision for the future is one where we feed and promote innovation with Slovakia, build unicorns and create exciting opportunities within tech and research through collaborative efforts

between academia, industry and governing bodies. Extending that vision to Eastern Europe means having regional collaborations with neighbouring countries combining resources and competencies for stronger AI capabilities. All of which will in turn help the chance of developments of competitive AI models that are true to an open, liberal, free and western society – and our way of life, concludes Bielikova, sharing her hopes and dreams for the future. ■



About the researcher:

Maria Bielikova is the founder and Director General of KInIT. She is also the chair of the Permanent Committee for Ethics and Regulation of AI of the Slovak Republic, and a former member of the High Level Expert Group on Artificial Intelligence established by the European Commission.

Factual and Trustworthy Contributions to European Policy making

At the heart of the TAILOR Network's mission is the desire to see the recognition of the need for strategic AI investments within the EU, and a policy framework adapted to the needs of European businesses and citizens. Fredrik Heintz, the Coordinator of the TAILOR network, works intensively towards this target.

– The ultimate goal would be that everyone realises that we need to make strategic investments and to develop our own models that share our values, goals, and interests, says Fredrik Heintz, professor at Linköping University and the leader of the network.

Despite his extensive commitments as a researcher and a university teacher, Heintz is devoting a lot of time to activities contributing to public awareness and the broader societal dialogue surrounding AI. Mainly focused on meeting high-level policy makers and industry representatives, this also means making time in his busy schedule for journalists and local politicians.

– It's not just the tech industry that stands to benefit from AI. The real game-changer lies in AI's capacity to revolutionise the public sector, making it more efficient, responsive, and human-centric, Heintz explains. It's crucial that we, as scientists, share not only our concerns and fears but also our vision for a huge societal benefit.

Large parts of the EU public sector are currently facing a serious predicament. Financial constraints, coupled with a significant workforce shortage, pose a



AI IN THE EU PUBLIC SECTOR

AI Watch is the European Commission's knowledge service to monitor the development, uptake and impact of Artificial Intelligence in Europe. The report* "AI Watch: European landscape on the use of artificial intelligence by the public sector", from 2022, found 686 AI use cases in 30 European countries. The use cases were in diverse domains: General Public Services (30 %), Economic Affairs (18 %), Health (15 %), Public Order and Safety (14 %), Social Protection (9 %) and Environmental protection (4 %).

*Tangi L., van Noordt C., Combetto M., Gattwinkel D., Pignatelli F., AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector, EUR 31088 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53058-9, doi:10.2760/39336, JRC129301.

severe threat to the availability and quality of public services. The Swedish Association of Local Authorities and Regions has released a report, that in the future, 95 % of the working population would have to work within the public sector to keep up the current level of service. This highlights the urgency of finding a solution and is one of the major motivations for the policy work done by the TAILOR scientists.

Their idea is that AI can be used to alleviate the pressure on the public sector. By automating administrative tasks and facilitating seamless integration across various government departments, AI can free human workers to engage in

the core activities of caring, nursing, and educating.

– The promise of AI is that the people can focus on the social and humane aspects, and AI can do the admin, the diagnosis,

the diagnostics and optimize the workflow, says Heintz, sharing a vision of humans and AI working together.

The economic benefits of integrating AI into the public sector are staggering.



It's crucial that we, as scientists, share not only our concerns and fears but also our vision for how AI can have a huge societal benefit.

FREDRIK HEINTZ

Heintz refers to a study by DIGG, the Swedish agency for digital government, which projects annual savings of up to 140 billion SEK. The cost savings aside though, as it's more about how public services are delivered, and how workers are deployed in more fulfilling, impactful roles.

The scientists of the TAILOR network are relentlessly calling for governments, policymakers, and technologists to embrace AI as a tool for societal improvement. Large strategic investments and a focus on developing AI that aligns with our common values, goals, and interests is needed to allow a future where AI and humans work side by side. ■

About the researcher:

Fredrik Heintz is professor of computer science at Linköping University and Coordinator of the TAILOR project. Fredrik Heintz is also the director of the Wallenberg AI and Transformative Technologies Education Development Program (WASP-ED), the vice-president for AI Research Adra, and a member of the Swedish national AI Commission.

TAILOR scientist Ana Paiva, Secretary of State for Science in Portugal



Ana Paiva, one of the founding members of the TAILOR network, was appointed Secretary of State for Science in Portugal in April 2024. Her scientific work in TAILOR was focused on investigating how AI agents act and learn in a society.

AI-Powered Collaboration Tools

AI-powered collaboration tools can be used to find new and exciting collaborations and collaborators within the research community. As the TAILOR Network is tasked with developing trustworthy AI, making use of AI tools to further AI research leads to both new exciting opportunities and allows the researcher to put theory into practice. Leading the charge and responsible for the development and deployment of these tools within the Tailor Network are Miquel Perello-Nieto and Peter Flach at the University of Bristol.

One of the first AI-powered collaboration tools is called SubSift and was originally designed to assign reviewers to articles sent for publication in several major data mining conferences. As part of the development process, SubSift was also used to match a text query to the different expertise of the TAILOR partner leads, to show the search capabilities using internal data. A live demonstration of the tool can be found online.

Another unique feature of the SubSift AI tool is that it offers explainability by giving insight as to why it makes its recommendation. This, coupled with the fact that it is not allowed to hallucinate, makes it compliant to the guidelines of trustworthy AI and provides transparency.



Pushing these capabilities even further, the team was using SubSift together with SynTeam (a tool developed by a group of researchers led by Carles Sierra at IIIA-CSIC, Barcelona) to find synergies and teams to work with on similar topics at the TAILOR Network summer school and at two TAILOR Conferences.

– SubSift was also used at the 3rd TAILOR conference to discover potential connections among attendees with favourable results, explains Perello-Nieto,

so we think there is great potential for applying them to the research community at large.

The use of SubSift to assign reviewers to scientific articles and conference contributions reduces time spent matching papers and reviewers, and greatly increases the quality of matching. And when using both tools for collaborations and suggesting networking activities, the feedback from participants was overwhelmingly good and positive, with 66 % of the researchers that discovered each other through SubSift saying that they would be open to collaborating, and 96 % agreed that it had added value to their professional network. Providing further encouragement is the fact that SubSift was highlighted by the European Commission's Innovation Radar as an excellent innovation.

The positive outcome of the pilot testing lends credence to the utility of the tool, and going forward the hope is that more AI researchers integrate these AI tools – accelerating the reviewing process and enabling better and more relevant networking and collaborations.



– We think this holds great promise, and anything that can help us as a community to spend more time on research and exchanging ideas with our peers the better, says Flach, hoping to facilitate more research and less admin.

SubSift

SubSift is a collection of tools developed in Bristol for matching academic researchers and text. It works by converting text documents to a bag-of-words (BoW) representation using term frequency-inverse document frequency (TF-IDF) and calculating pairwise cosine similarities. For researchers, publication titles as listed on their DBLP computer science bibliography page are converted to BoW. SubSift was highlighted by the European Commission's Innovation Radar as an excellent innovation.



To continue building on the successful trials of SubSift and SynTeam, there would need to be deployment on a larger scale, and integration into more widely adopted review management tools, as well as seeing adoption of these systems and the go to networking and

collaboration tool for conferences and large events.

– There are somewhat similar tools on the market, and this may be more of a step up than a complete reinvention, summarizes Perello-Nieto and Flach, but

we do think that it speaks to the incredibly broad range of fields and instances where AI can be used to create greater efficiency, and how important it is that we continue to explore all of those venues. ■

About the researchers:

Miquel Perello-Nieto has been a Research Associate at the University of Bristol since 2015, and is a researcher of Work Package 9 Network Collaboration in the TAILOR project.

Peter Flach has been a Professor of Artificial Intelligence at the University of Bristol since 2003, and is the leader of Work Package 9 Network Collaboration in the TAILOR project.

Carles Sierra is the Director of the Artificial Intelligence Research Institute (IIIA) of the Spanish National Research Council (CSIC) in Barcelona, and the President of EurAI, the European Association of Artificial Intelligence.



SynTeam

Subsift and SynTeam was used at several TAILOR conferences and summer schools to create diverse and creative teams. SynTeam* is based on the Post-Jungian Personality theory for team composition, which has been shown to improve the performance of teams in various studies. A set of questions was completed by all participants, to evaluate psychological functions, psychological attitudes. The result is 16 possible combinations forming a personality used for creating teams.

*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.

Safe and powerful AutoAI solutions amplify future research and innovation in many domains

Machine learning and other AI technologies have the potential to greatly increase scientific output in many domains, while even increasing quality and lowering costs. However, due to the current exponential growth and influx of investment and interest, access to computer scientists and machine learning specialists is often a limiting factor. Scientists working in the group of Holger Hoos work on AutoAI, solutions that will democratize AI by allowing experts from other domains to develop their own AI tools.

Scientists from all domains working with large amounts of data face similar challenges. Curating, reading and interpreting data is time-consuming and tedious. Humans also tend to make errors and introduce biases.

Computational solutions to this problem were introduced in the early seventies. Typically, domain experts have developed task-specific models that can successfully handle a specific data analysis. However, these models are not easily transferable to other tasks or domains, and any

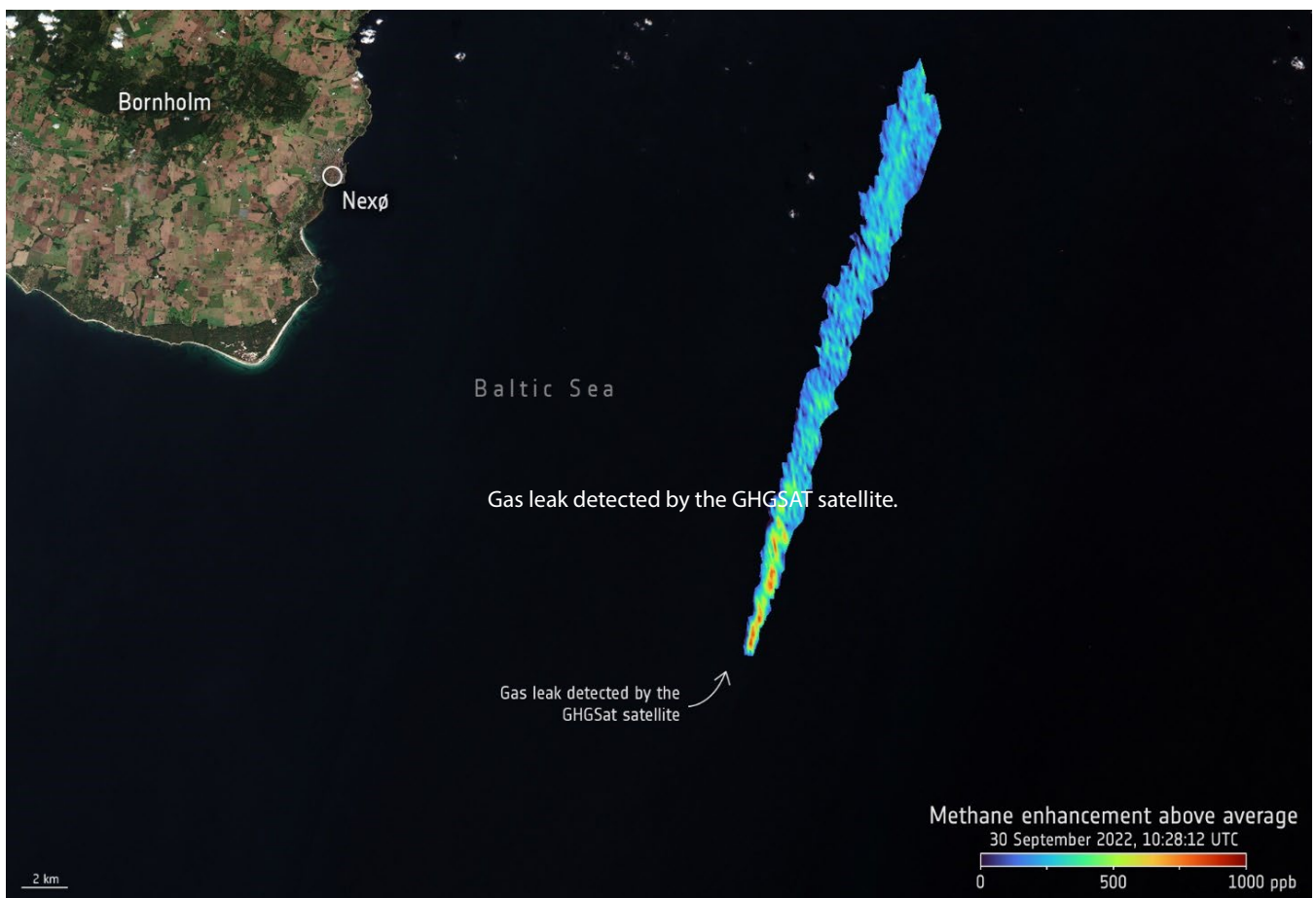
change in the task or equipment often implies the re-design of the whole model.

– For example, switching from one satellite to another might make your model unusable, forcing you to design a new one, explains Julia Wąsala. With AutoAI methods, you don't have to redesign a whole new system each time.

In earth observation studies, domain experts use data from sophisticated instruments in space to tackle today's greatest environmental challenges, such as climate change and air pollution. One

such phenomenon is the release of greenhouse gases. Typically, scientists interpret and organise satellite data to combine it with data from maps and meteorological information. With the help of machine learning, this process can be made faster, more accurate and less biased.

– The problem is that you need a lot of machine learning experience to be able to design machine learning models for these applications, says Wąsala. With AutoAI algorithms, earth observation scientists will have access to smart



algorithms that design the models for them. The scientist provides the data and the AutoAI assembles modules, like building blocks, and then assembles the best possible model for the data. The AutoAI actually takes away a lot of choices you otherwise have to make and are tricky to make for non-experts.

Using AutoAI methods to create new models not only increases the possibilities for scientists outside the machine learning sector to build their own models, but it also has the potential to increase the trustworthiness of the models.

– A huge problem in machine learning, in particular neural networks, is that they

NEURAL NETWORKS

Artificial neural networks are models inspired by the biological neural network in the animal brain. It contains different nodes, connected by edges and assembled into layers. The term deep learning, or deep neural networks, refers to the use of multiple layers.



can be easily fooled. To make sure that doesn't happen, the field of verification has emerged. Verification is very expensive, so what we're working on is making AutoAI do the verification to

make sure the neural network is safe, explains Annelot Bosman.

Eventually the researchers would like to incorporate verification features in a framework that would create a neural network that is already safe from the beginning and ready for use without additional verification.

AutoAI technology also allows for using less computational power and energy. Usually, AI models need extensive training and testing but in AutoAI, smart algorithms do the testing. This reduces the time spent using supercomputers, freeing valuable computational power for other applications and reducing energy use.

– With AutoAI frameworks, users don't need to have in-depth programming knowledge, but they need to understand machine learning in theory. And that can be learnt in class, concludes Bosman. ■



Julia Waşala (left) and Annelot Bosman

About the researchers:

Julia Waşala, PhD student in Automated Machine Learning for Earth Observation, supervised by dr. Mitra Baratchi and **Annelot Bosman**, PhD student in Neural Network Verification, supervised by dr. Jan van Rijn. They all work with Holger Hoos at Leiden University, the Netherlands.

Holger Hoos, is an Alexander von Humboldt Professor of Artificial Intelligence, RWTH Aachen University, Germany, and Professor of Machine Learning, Leiden University, The Netherlands. Hoos is also an Adjunct Professor of Computer Science at the University of British Columbia, Canada and the Chair of the Board for the Confederation of Laboratories for AI Research in Europe (CLAIRE).

Best paper award at AAI 2023 SafeAI Workshop

König, M., Bosman, A. W., Hoos, H. H., & van Rijn, J. N. (2023). Critically Assessing the State of the Art in CPU-based Local Robustness Verification. In Proceedings of the Workshop on Artificial Intelligence Safety.



Shaping the Future of AI within the EU

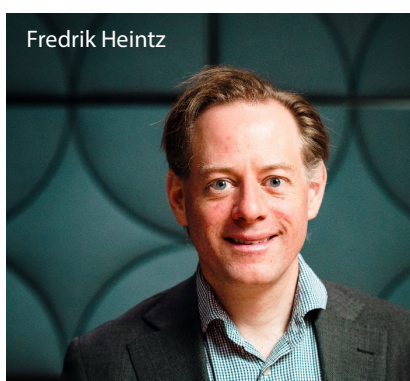
In an era marked by rapid technological advancements, the European Union is taking decisive steps to harness the power of Artificial Intelligence responsibly. Recognizing AI's vast potential alongside the imperative for safety and trust, the EU has spearheaded the TAILOR Network, a pioneering research initiative aimed at shaping a future where AI technologies are developed within a framework of ethical guidelines and human-centric values.



Funded by the European Union, the TAILOR network is an ambitious project that seeks to navigate the complexities of AI development while ensuring that the technology remains a force for good. At the helm of this transformative endeavour is Fredrik Heintz, whose passion and expertise are guiding the network towards a future where AI enhances every aspect of our lives.

– The EU initiative was what got us started, Heintz recalls, highlighting the collaborative effort to establish scientific foundations for trustworthy AI. I was excited to help develop the scientific foundations for determining what constitutes a trustworthy AI – and to help companies and organizations to implement the technology, he adds, emphasizing his multifaceted role in steering the project towards its ambitious goals.

Getting the TAILOR Network off the ground required both a big investment and a collaborative spirit.



– A big part of this work was to actually build out the TAILOR Network, and as with any endeavour, there is an initial startup cost associated with that, Heintz explains. Surmounting these challenges,

the network has set the stage for continuous innovation in the AI field.

Success for the TAILOR Network, according to Heintz, involves crafting a technical solution that adheres to eight ethical principles, deemed essential for AI's trustworthiness.

– We are hoping to develop a technical solution that can make sure that any model being evaluated meets the requirement for the 8 ethical principles that was deemed to constitute trustworthiness in AI, Heintz explains, underlining the human-centric focus.

Beyond scholarly output, the network aims to influence policymaking, public awareness, and the broader societal dialogue surrounding AI. The network's aspiration is to foster an environment where healthy scepticism is balanced with an openness to innovation. Through

its steadfast commitment to collaboration, innovation, and ethics, the network is shaping the future of AI.

Looking ahead, Heintz envisions a future where the TAILOR Network remains at the forefront of AI research and development.

– My hope is that we can now continue this collaboration, keep iterating, and

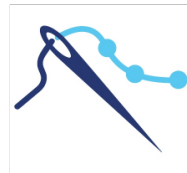
keep developing new tools as AI technology is a very rapidly moving target, he states. Keeping pace with the fast advancements in AI technology is vital for ensuring that these developments gain society as a whole. ■



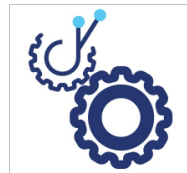
DEVELOPING THE SCIENTIFIC FOUNDATIONS FOR **TRUSTWORTHY AI** THROUGH THE INTEGRATION OF LEARNING, OPTIMISATION AND REASONING



TRUSTWORTHY AI



PARADIGMS AND REPRESENTATIONS



ACTING



SOCIAL AI



AUTO AI

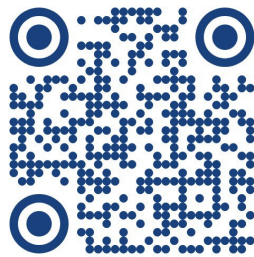
About the network

TAILOR is an EU project with the aim to build the capacity to provide the scientific foundations for Trustworthy AI in Europe. TAILOR develops a network of research excellence centres, leveraging and combining learning, optimisation, and reasoning (LOR) with the key concepts of Trustworthy AI (TAI). These systems are meant to provide descriptive, predictive, and prescriptive systems integrating data-driven and knowledge-based approaches.

TAILOR has 54 partners and is coordinated by Fredrik Heintz, Linköping university. It is an ICT-48 Network of AI Research Excellence Centres funded by EU Horizon 2020 research and innovation programme under grant agreement No 952215

READ MORE

TAILOR Network
of Excellence Centres in AI



Funded by the
European Union