Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization
TAILOR
Grant Agreement Number 952215

NEW D8.2 Industry transfer congress and microprojects report

Document type (nature)	Report
Deliverable No	NEW D8.2
Work package number(s)	WP8
Date	October 2024
Responsible Beneficiary	DFKI, P26
Author(s)	Janina Hoppstädter, Kyra Kiefer
Publicity level	Public
Short description	Summary of the Organisation and the Results from the 1st Industry Collaboration and Transfer Exchange on AI “From Research to Market”

History			
Revision	**Date**	**Modification**	**Authors**
v1	07.10.2014	First version	DFKI - Janina Hoppstädter, Kyra Kiefer

Document Review		
Reviewer	**Partner ID / Acronym**	**Date of report approval**
Fredrik Heintz	#1, LiU	06.10.2024
Annelot Bosman	#7, LEU	01.10.2024

Table of Contents

Figures

Introduction to the Deliverable

This document serves as a deliverable under Work Package 8 (WP8), titled "Industry and Innovation Programme." The key objective of WP8 is to foster synergies between the industry and the TAILOR network of excellence centres, laying the foundation for Trustworthy AI in Europe. This includes organising and conducting Theme Development Workshops (TDWs) that focus on specific industry sectors within the AI ecosystem.

This particular deliverable offers a comprehensive summary of the outcomes from the first TAILOR Transfer Congress "Industry and Collaboration Exchange – From Research to Market". This event was designed specifically to enhance collaboration between research and industry in the field of Trustworthy AI, aiming to bridge the gap between academic advancements and industry needs. The goal was to ensure that AI technologies are developed and applied in ways that are both innovative and reliable.

The document is structured as follows:

- **Chapter 1:** Provides an in-depth look at the event's organisation, highlighting key participants and the strategic significance of the venue.
- **Chapter 2:** Covers the congress's main objectives and outlines key outcomes, such as the development of new collaborative frameworks and targeted micro-projects.
- **Chapter 3:** Summarises insights and strategies from workshops and working groups, with a focus on addressing practical challenges faced by both industry and research.
- **Chapter 4:** Reflects on the congress's impact and presents the next steps and future directions for the TAILOR network.

The following people have been involved in the Deliverable:

Partner Acronym	Name	Role
DFKI	Janina Hoppstädter	Senior Project manager
DFKI	Kyra Kiefer	Project manager

Executive summary

The first TAILOR Industry Collaboration and Transfer Congress, held on May 6-7, 2024, at the Innovation Center in Saarbrücken, was an important step in connecting foundational research on Trustworthy AI with the practical needs of industry. Organised by the TAILOR Industry Collaboration and Transfer Group, the congress focused on addressing two major challenges facing Europe: the loss of valuable AI talent and the long delay in transferring AI research breakthroughs into industrial applications.

The event brought together 60 participants from academia, industry, and policymaking in a lively environment supportive to collaboration. The congress featured keynote speeches and interactive workshops, all aimed at fostering the development of innovative, human-centred AI solutions that align top-down research with bottom-up market needs.

Keynote speakers, including Thomas Neubert from Transatlantic AI eXchange and Laure Poirson from EIT ICT Labs Germany, emphasised the importance of enhancing research commercialization and strengthening the connection between academia and industry. These efforts led to the formation of the Industry Collaboration and Transfer Alliance, a new initiative designed to support greater collaboration among researchers and industry partners. This alliance is planning several flagship programmes, including joint workshops, hackathons and a dedicated platform to facilitate interaction between stakeholders, offering brokerage services for collaboration on specific AI solutions.

The congress also highlighted the critical role of trustworthy AI in societal applications, with Kevin Baum from the German Research Center for Artificial Intelligence (DFKI) discussing AI's potential to drive beneficial societal innovations. The event showcased best practices in academic-industry collaboration that will inform future projects.

In summary, the 1st TAILOR Transfer Congress successfully established a strong foundation for the collaboration between academia and industry required to develop trustworthy AI. The results of this event are expected to lead to intensified large-scale AI research and the widespread adoption of this new technology, positioning Europe at the forefront of trustworthy AI applications guided by advanced ethical principles.

About the Congress

The following sections provide an overview of the congress, outlining its objectives, the context in which it was organised, and its significance within the broader scope of the TAILOR project. Additionally, the facts and figures highlight key aspects of the event, including its collaborative nature, the organisations involved, and the impact it had on fostering innovation and practical applications in Trustworthy AI.

Objectives and Purpose of the Event

The issue of brain drain, particularly in the AI field, is a major concern in Europe. Many researchers, industrial scientists, and other highly skilled workers, including junior staff, are emigrating, often to the United States, where opportunities for career growth, flexibility, and higher compensation are more readily available. This trend is especially problematic because the U.S. leads the world in recruiting AI specialists, offering them the chance to "change the world" within a supportive and rewarding environment. Unfortunately, such opportunities are far less common in Europe.
Besides the brain drain, Europe faces another significant challenge: the need to accelerate the transfer of AI technologies from research to practical industry applications.
To address these challenges, the 1st TAILOR Industry Collaboration and Transfer Exchange on AI,“From Research to Market” was initiated. The main goal of this congress was to foster collaboration and streamline the transfer of knowledge and expertise between industry and academia in the rapidly evolving field of Trustworthy Artificial Intelligence. The event aimed to bridge the gap between industry needs and research capabilities through inspiring keynote speeches and engaging workshops designed to identify industry requirements and develop targeted micro-projects in partnership with research institutions.
In addition to promoting direct engagement between industry professionals and researchers, the congress provided a unique opportunity for participants to brainstorm ideas for innovative micro-projects. These micro projects were designed to be quickly realisable, addressing specific challenges faced by industry partners. The event encouraged the formation of interdisciplinary teams and collaborations with various partners, facilitating the implementation and financing of these micro-projects later on. This initiative was intended not only to meet immediate industry needs but also to lay the groundwork for long-term partnerships and funding opportunities.
The congress served as a pivotal platform where industry professionals could directly communicate their insights, challenges, and needs with researchers. This direct engagement helped align research efforts with real-world problems and market demands. At the same time, researchers had the chance to showcase their latest advancements and collaborate with industry partners, ensuring that their work translated into practical and impactful solutions.

Significance of the Event in the Context of the Project

To tackle the challenges and risks highlighted during the 1st Industry Collaboration and Transfer Exchange on AI – “From Research to Market,” it is crucial for research, business, and politics to collaborate closely and improve education, research, and innovation policies. The potential for cooperation needs to be fully explored and continually expanded, as these challenges cannot be addressed by any single government or entity alone. The advancement of Europe requires a collective effort, pulling in the same direction.
Therefore, the Industry Collaboration and Transfer Alliance was established and will bring together research and industry partners to jointly develop new AI applications and foster cooperation. Initially led by some of Europe’s top research institutes involved in TAILOR, the alliance will soon be joined by additional members from both academia and industry. The initiative will grow to include more

research institutes across Europe as it evolves. Current members of the Industry Collaboration and Transfer Group include ABB, CBS, Tietoevry, Inria, EDF, FBK, DFKI, ENG, and TNO.
The Group is now planning several key initiatives to promote greater collaboration and knowledge sharing. These include analysing and sharing requirements, both internally and externally, to find efficient ways of working together. These efforts are to begin with an 'information event' where existing work and solutions will be presented, accessible not only to participants but also to a wider range of external stakeholders. The aim is to identify specific requirements for new solutions and ideas for closer cooperation between research institutions and industry, including through joint workshops. In addition, the group is planning new collaboration formats, such as hackathons and challenges, to encourage creative, solution-oriented approaches.
Other plans include knowledge creation, content development and the completion of the joint strategy and solution development programme. This includes building knowledge databases and developing AI workshop formats to engage stakeholders from research, industry and politics. The group also aims to further develop these solutions and integrate them into new and existing programmes, building on the lessons learned from previous implementation efforts.
A key element in the group's future plans is the development of a dedicated platform. This platform will provide information on calls for proposals and collaboration opportunities, broker enquiries between research and industry and facilitate communication between different scientific disciplines and industry sectors. It will also serve as a platform for bringing together stakeholders and as a sandbox for trialling and validating AI solutions.
Through these and related initiatives, the TAILOR Industry Collaboration and Transfer Group aims to create a solid framework for future pan-European collaborative AI development that supports innovation and societal applications across different economic sectors.

Facts & figures

The first Transfer Congress, titled "Industry and Collaboration Exchange on AI – From Research to Market," was organised by the TAILOR Industry Collaboration and Transfer Group and held on the 6th and 7th of May 2024. This event took place at the newly established Innovation Center in Saarbrücken, chosen for its symbolic significance as a hub for innovative ideas in the Greater Region. The Innovation Center serves as a beacon for future visions, fostering an entrepreneurial spirit and promoting productive dialogue between science and industry—making it an ideal venue for the congress. The workshop was co-organized by prominent organisations such as CERTAIN, CLAIRE, DFKI, TNO, and the Transatlantic AI eXchange and the VISION project to further enhance the collaborative nature of the event.

Figure 1: Congress Flyer

The congress brought together 60 participants from academia and various industry sectors, creating an exciting platform for collaboration and innovation. The gathering provided an environment conducive to exchanging ideas, sharing expertise, and charting clear paths for the practical application of trustworthy AI technologies. The diversity of participants and co-organizing bodies highlighted the interdisciplinary approach needed to address both the challenges and opportunities in AI, fostering lively discussions. The connections made and the knowledge shared during the congress are expected to inspire continued collaborative work and innovation in the development and deployment of trustworthy AI, laying a strong foundation for future efforts in this field.

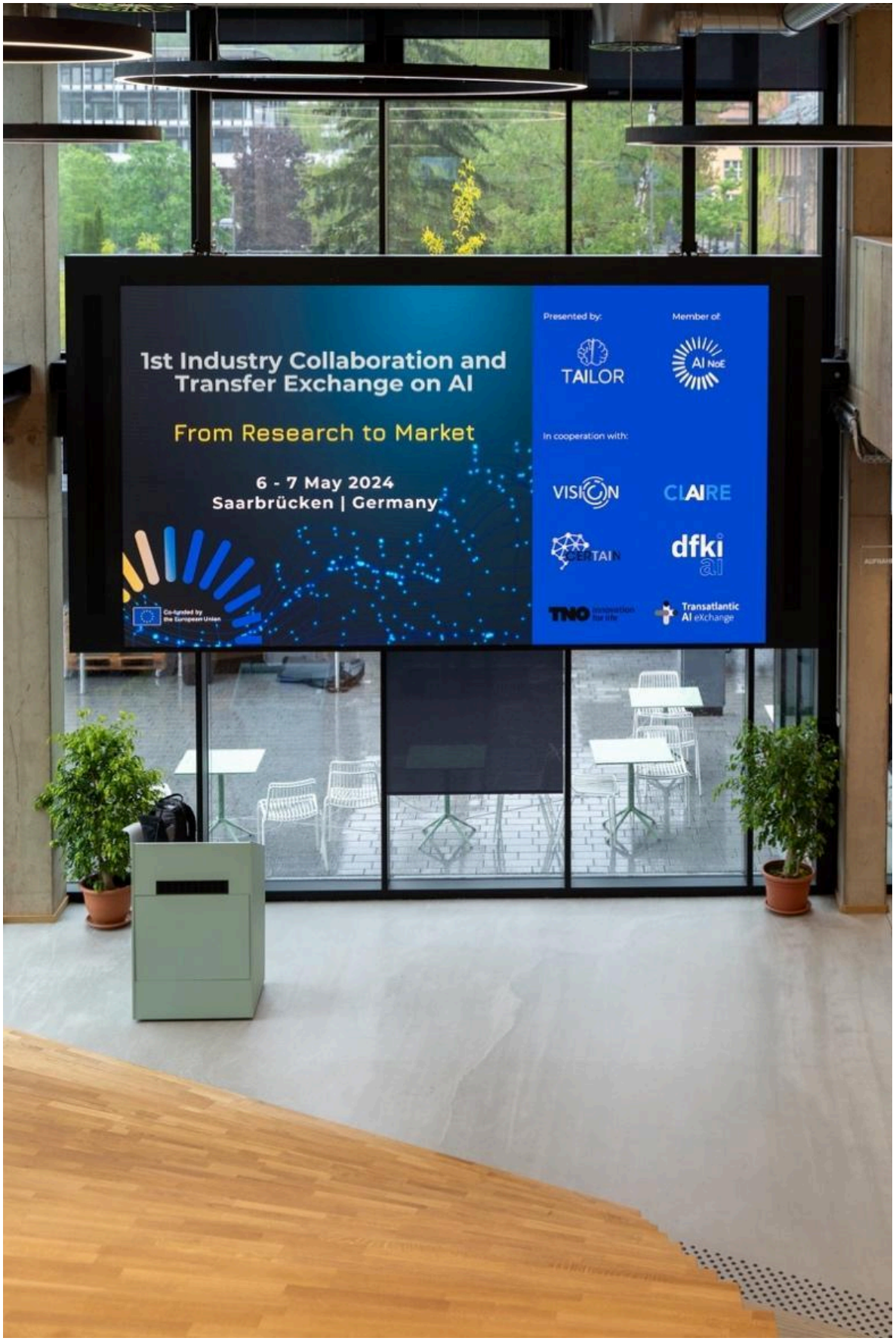


Figure 2: Congress setting at the Innovation Center

Figure 3: Networking opportunities

Program Overview and Schedule

The congress, moderated by Freek Bomhof (TNO) and André Meyer-Vitali (DFKI), kicked off with enlightening keynote speeches. Over the course of two days, participants engaged in parallel working groups, generating valuable insights and strategies aimed at concrete changes in AI implementation. These findings were later shared in plenary sessions.
Below is a detailed outline of the program for both days of the congress:

Day 1

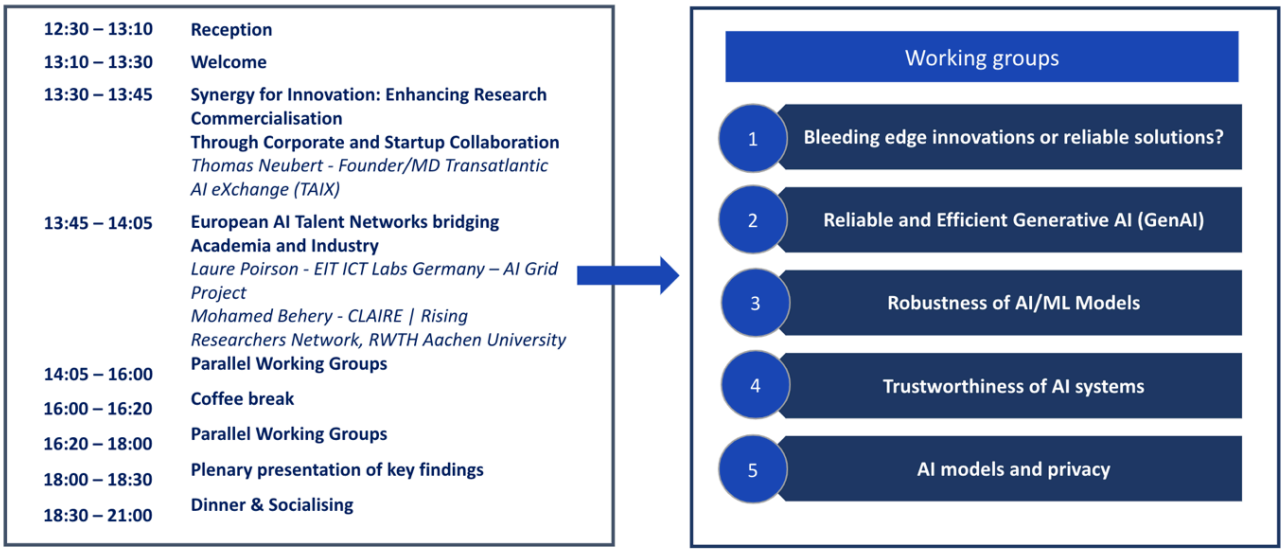


Figure 4: Schedule Day 1

Day 2

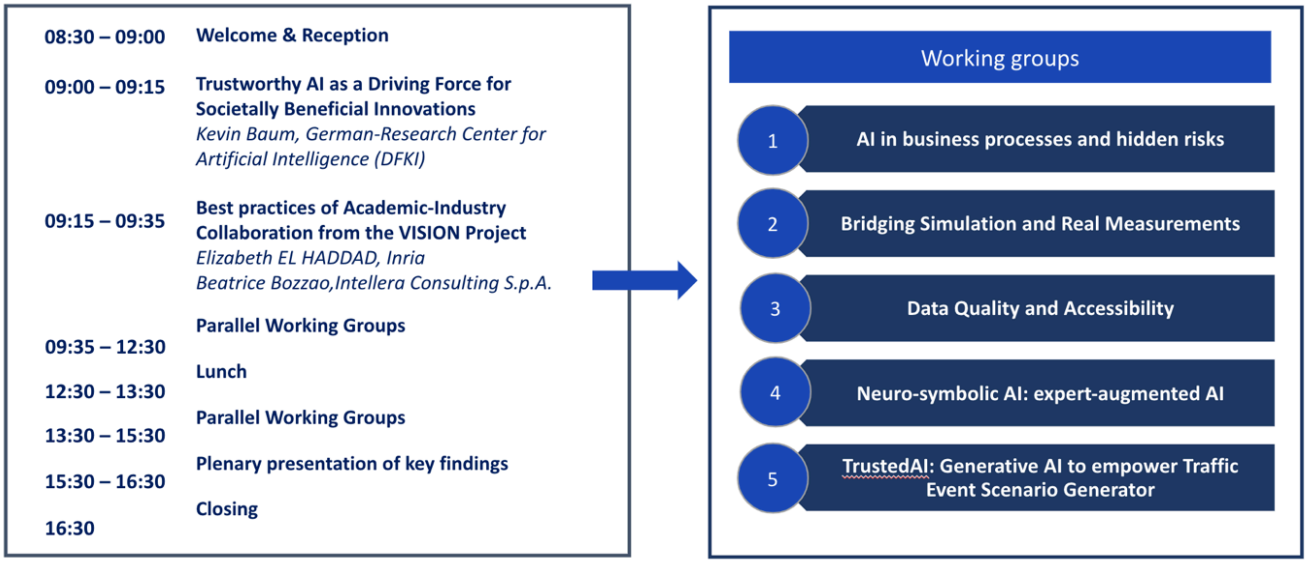


Figure 5: Schedule Day 2

Welcome & Introduction

The congress was moderated by Freek Bomhof (TNO) and André Meyer-Vitali (DFKI). They outlined the objectives of the TDW as well as the agenda and programme and introduced the invited keynote speakers to the participants.

Figure 6: Welcome & Introduction by André Meyer-Vitali (DFKI)

Figure 7: Welcome & Introduction by Freek Bomhof (TNO)

On the first day of the congress, Prof. Dr. Antonio Krüger and Prof. Dr. Philipp Slusallek held welcome messages that emphasised the critical importance of Trustworthy AI and the effective transfer of knowledge from academia to industry. They stressed that advancing AI is not only about technological innovation but also about ensuring that these innovations are reliable, ethical, and aligned with societal values. Their speeches set the tone for the event as a collaborative space focused on developing AI technologies that are both advanced and trustworthy, while also challenging participants to think deeply about how their research can be practically applied.

Figure 8: Welcome note by Antonio Krüger (DFKI)

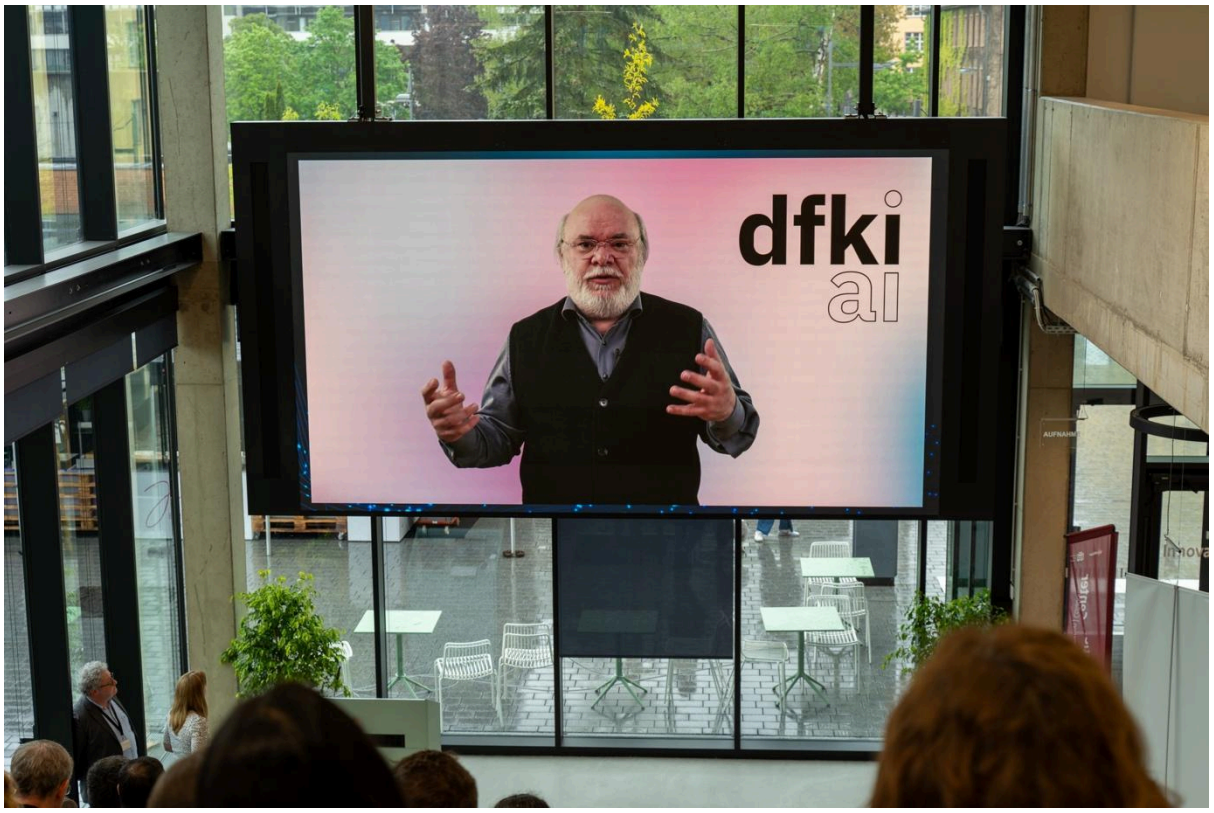


Figure 9: Welcome note by Philipp Slusallek (DFKI)

Keynotes

The inspiring keynotes set the stage for the congress, featuring high-level experts from both academia and industry. Thomas Neubert (Transatlantic AI eXchange), Laure Poirson (EIT ICT Labs Germany), and Mohamed Behery (CLAIRE | RWTH Aachen University), shared their perspectives on bridging the gap between academic research and industry application. Additionally, the day featured discussions on the role of Trustworthy AI in societal applications, led by Kevin Baum (DFKI), who highlighted AI's potential to drive beneficial societal innovations. The congress concluded with presentations from Elizabeth EL HADDAD (Inria) and Beatrice Bozzao (Intellera Consulting S.p.A.), who showcased best practices from the VISION project on academic-industry collaboration.

These foundational presentations were not just an in-depth look at the most important issues, but they also served as a springboard for later discussions on the impact of bias and combating misinformation in the AI field. They were accompanied by compelling real-world use cases, which further fueled expert dialogue in the subsequent Working Groups.

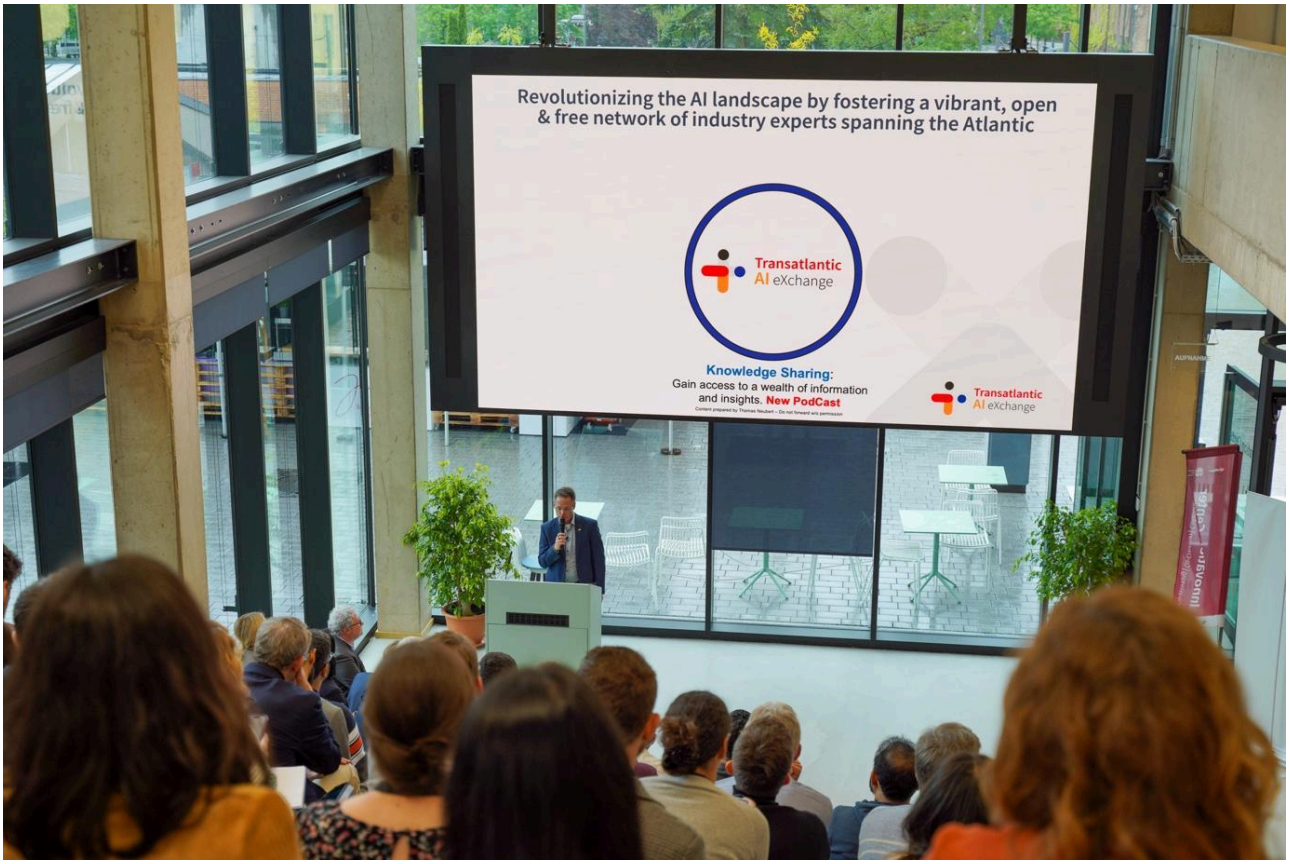


Figure 10: Keynote speech by Thomas Neubert (TAIX)

Figure 11: Keynote speech by Mohamed Behery (CLAIRE R2N Nwtwork)

Figure 12: Keynote speech by Laure Poirson (EIT ICT Labs Germany – AI Grid Project)

Working Groups

Following the keynote speeches, participants engaged in active parallel working group sessions over two days to produce a series of insights and strategies for AI implementation. The topics discussed ranged from cutting-edge innovations to AI in business processes. These breakout sessions concluded with plenary discussions where groups reported back on their discussions and proposed solutions.

This overview provides a comprehensive account of the working group sessions that took place over both days of the congress.

Working Group topics - Day 1

1. **Bleeding edge innovations or reliable solutions? (TNO)**
   The suitability of research results for industry transfer depends on their Technology Readiness Level (TRL), ranging from cutting-edge (TRL8) to moderately advanced (TRL5). While TRL8 findings are readily adoptable, TRL5 results can be tailored to align closely with industry requirements. Hence, the key considerations are identifying the TRL of research outcomes and understanding the industry partners' specific needs. Let's discuss these aspects and the existing achievements together.

2. **Reliable and Efficient Generative AI (GenAI) (Tietoevry)**
   Join us in discussing the future of GenAI and its on-device & in-cloud applications. Efficient fine-tuning proves its effectiveness against long and extensive training. Consequently, forthcoming hybrid AI models will have the ability to operate both locally on devices and in the cloud. Decisions about which model to employ will be influenced by factors such as 5G/6G network latencies, the availability of computational resources, and legal considerations.

3. **Robustness of AI/ML Models (DFKI)**
   Robust AI/ML models are designed to maintain validity, performance, and resilience throughout their lifecycle, even in the face of changes or attacks. This involves implementing strategies to continuously monitor and adapt models to evolving conditions and potential threats. By prioritising robustness, organisations can enhance the reliability and effectiveness of their AI/ML solutions.

4. **Trustworthiness of AI systems (ABB)**
   Join us to explore how user needs influence the effectiveness of explanations versus analysability in promoting trust in AI/ML models, particularly in specialised contexts, such as neural network-based control systems. In some cases, such as those involving process engineers, analysability (e.g., sensitivity analysis), may be more beneficial than explanations. Tailoring approaches to provide the most relevant insights for users' specific needs can ultimately increase confidence in AI/ML models.

5. **AI models and privacy (EDF)**
   AI models applied in upstream activities rely on sensitive data, such as individual load curves and contract details, necessitating privacy safeguards to align with GDPR and national regulations. This involves implementing privacy-preserving practices like Differential Privacy in Stochastic Gradient Descent and defending against privacy attacks, enabling scalability

across diverse applications and data types, including consumer or IoT data and structured or unstructured data. Discuss with us the necessary measures and their effective implementation to enable scalability across different data types and applications.

Working Group topics - Day 2

1. **AI in business processes and hidden risks (TNO)**
   AI can streamline existing business processes, delivering immediate efficiency gains and simple return on investment calculations. Conversely, AI can also revolutionise non-existent or fundamentally changing processes, which is a riskier, but potentially more impactful route. Ultimately, the industry's appetite for risk will determine its approach to adopting AI solutions. But how high is it?

2. **Bridging Simulation and Real Measurements (ABB)**
   When developing AI/ML/statistical models, the gap between simulation (surrogate models) and real measurements must be closed – both sources of data are necessary, as typically either one alone lacks sufficient quantity or accuracy. How can these be closed?

3. **Data Quality and Accessibility (CBS)**
   Discover with us the growing acknowledgement of the importance of prioritising data quality, accessibility, and regulatory measures alongside algorithms to foster trustworthy AI. Initiatives like the European Commission's Data Governance Act and Common European Data Spaces aim to enhance data access. Furthermore, explore the potential of synthetic data as a promising alternative for AI applications, particularly when real data access is limited, aligning with efforts to improve trustworthiness and facilitate research and industry transfer in AI/ML.

4. **Neuro-symbolic AI: expert-augmented AI (EDF)**
   Multiple industry organisations use expert systems that include rules, ontologies and knowledge graphs for downstream activities, alongside neural models for learning from data. Combining symbolic reasoning with neural approaches improves performance, provides explainability and increases confidence in the model results, especially when symbolic knowledge cannot be directly converted into physical equations for neural networks. The aim of the workshop is to share methods and experiences in integrating symbolic and neural thinking, especially in scenarios where symbolic knowledge cannot be easily translated into neural networks.

5. **TrustedAI: Generative AI to empower Traffic Event Scenario Generator (DFKI)**
   Join us in discussions while tackling the issue of out-of-distribution (OOD) data by leveraging generative AI and constraint satisfaction problem (ML) techniques, enhanced by a collection of explicit (symbolic) attributes. Our aim is to generate in-distribution traffic scenarios, including corner/critical traffic events, to achieve automation levels L3-L5 (SAE). Through a transfer of knowledge and know-how, delve into the significance of these challenges and the proposed solutions for the industry.

Outcomes from the working groups - Micro projects funded by TAILOR

One of the key outcomes from the working groups during the 1st TAILOR Industry Collaboration and Transfer Exchange was the development of 3 micro-projects. These initiatives were not just small-scale projects; they were a testament to the collaborative spirit fostered during the congress. They demonstrate how collective innovation can take root in real-world development, especially in the realm of Trustworthy AI.

These micro-projects were designed as collaborative efforts among a diverse community of multidisciplinary researchers and industry professionals, aimed at addressing specific challenges identified during the working groups. This approach allowed developers to consolidate knowledge and resources, ensuring that the projects were relevant not only for academic advancement but also reflective of industry realities.

Moreover, these micro-projects extend beyond the TAILOR network, encouraging collaboration with new partners, whether within or outside the TAILOR consortium. By broadening the range of expertise and perspectives involved, the projects were able to explore more cross-sectoral, innovative solutions, ultimately enhancing the overall quality and impact of the initiatives.

To facilitate the quick and easy initiation of these micro-projects, a streamlined process was established. Workshops were expected to generate ideas for transfer projects, and participants could apply for funding through a simple 1-2 page application immediately following the event. The review process was designed to be swift, taking only two weeks, allowing projects to start almost immediately and be completed within a 2-3 month timeframe. At the 4th TAILOR Conference in Lisbon, there will be an opportunity to present projects and secure specific funding for third-party partners, with each micro-project allocated a budget of €50,000.

These micro-projects needed to be part of an approved funding submission, aligning with the broader goals of the TAILOR initiative. At the end of each project, a short progress report is to be submitted to WP8 leader Philipp Slusallek and another TAILOR member for approval, with funding disbursed upon approval.

These micro-projects listed below serve as proofs of concept, demonstrating quick-to-implement, short-term solutions for critical industry needs. They are designed to generate new interactions and collaborations, thereby enhancing the reach and impact of the TAILOR network.

TrustedAI – Trustworthy Healthcare Professional (Doctor) Assistant for as Text Summarization, Assisting and Validating the Prescription for Patients

Organization: University of Jyväskylä, Jyväskylä, Finland **External Collaborator:** Tommi Lehtinen (Data Engineer) **External Collaborator Organization:** Lakeframe Oy, Finland **Host Organization**: Tietoevry Oy, Finland

Public summary

This research aims to improve healthcare services in the Nordic countries (with plans to later extend to Europe) by utilising Large Language Models (LLMs) to assist healthcare professionals and protect personal data. It focuses on developing tools to quickly generate extensive synthetic healthcare records, data anonymization, de -anonymization, model finetuning, and text summarization tailored to doctors' needs. Additionally, it involves translating into the Finnish language for more efficient consultations. The research also evaluates prescriptions based on patient-specific data, such as allergies and medication interactions, to ensure personalised and safe treatment. By initially using existing LLM tools, the project emphasises enhancing trustworthiness in clinical applications, streamlining healthcare processes, and alleviating the workload on medical staff, while ensuring the ethical and effective use of AI in medicine.

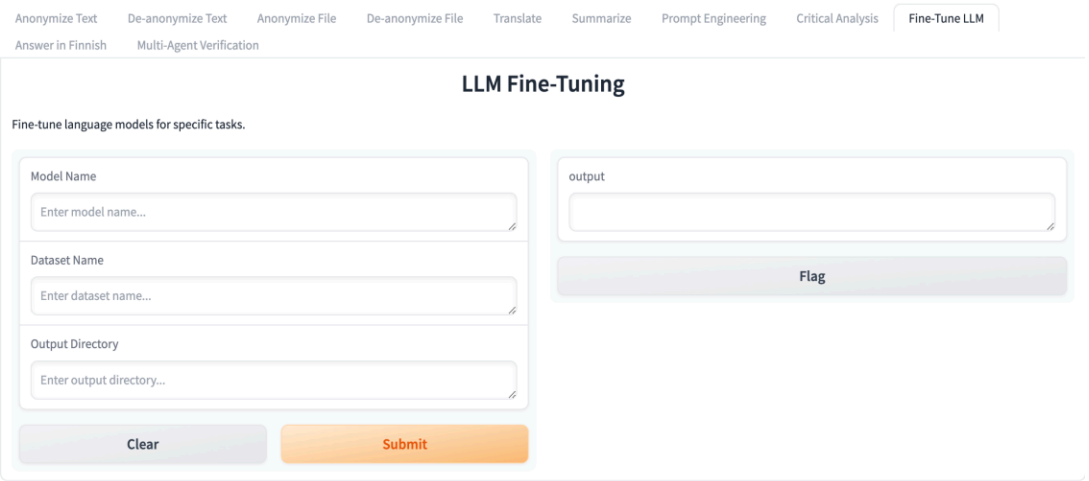
Research Objectives

The objective of this research was to advance the trustworthiness of healthcare AI systems and to ensure robust protection of personal data within these systems. This project addressed two main challenges: protecting personal healthcare data under GDPR and providing high- quality LLM models to assist healthcare professionals. They used English-based pre-trained models, incorporating Finnish, Finnish-Swedish, and Swedish data, to apply advanced techniques for text summarization and sentiment analysis.

Detailed description

The technical approach of this research is designed to advance the trustworthiness and efficiency of healthcare Large Language Models (LLMs), protect the personal healthcare data, employing a comprehensive, multi-faceted strategy:

- **GUI**: They provide the user-friendly interface through Gradio. This GUI platform allows easy medical text processing—essential in most medical applications—with interfaces for text anonymization, language translation, text summarization, and even fine-tuning of large language models to further improve special medical applications. These include backend scripts to support advanced NLP tools, among others: SentenceTransformer for semantic analysis and cosine similarity from Sklearn for text comparison for functionalities such as anonymization and summarization. This setting gives the user the ability to interactively anonymize texts, get translations, get summaries that can be tweaked to allow specific details such as medications and allergies, or even fine-tune language models right from the GUI. GUI with extra options for further processing enables critical analysis and, through that, comprehensive overall support for improving the handling of medical text in multiple languages, including Finnish. Compleate script of the GUI is available here https://github.com/wasimsse/MediTrustAI/tree/master/gu



- **Synthetic Medical data:** To achieve the objectives, they utilised synthetic data. Initially, they generated data for 250,000 individuals, encompassing various age groups (from newborns to those over 65 years) and genders. They also created records for multiple visits by these demographic groups. This dataset primarily included synthetic patient IDs, names, addresses, disease history, symptoms, basic health conditions such as blood pressure and heart rate, and doctor's notes, complete with confirmation of diagnoses. They chose to use synthetic medical data for three main reasons: to protect real patients' privacy and security; to enable large-scale and controlled research studies that are not possible with real data due to ethical concerns; and to ensure that researchers can get consistent results when repeating studies. They generated advanced probabilistic modelling techniques that simulate real-world occurrences of diseases based on age and gender, which are defined in the scripting logic https://github.com/wasimsse/MediTrustAI/blob/dataset/data/medical_conditions.txt). This includes condition-specific data generation to reflect realistic medical scenarios that healthcare data scientists and providers frequently encounter. Additionally, dynamic data generation was employed through the Faker library, allowing them to create realistic but fictional patient data dynamically, enhancing the richness and variability of the dataset. The generated dataset is available here https://drive.google.com/drive/folders/1N3Q-ftvaLTiiAOsdghlH7Sp-Jd5K7wD3, while the script for generating the data can be found on their public GitHub repository at https://github.com/wasimsse/MediTrustAI/tree/master.
- **Data Anonymization and Deanonymization:** Data Anonymization and De- anonymization: Before fine-tuning the selected data, they developed another component which was data anonymization and de-anonymization. They implemented this component with two approaches; the purpose of anonymization was to not provide any type of personal information to machine learning models for training in order to align with GDPR. The first approach utilised the Presidio Analyzer and Anonymizer Engines for automated data protection, employing token generation for reversible anonymization and dynamic handling based on the detected data type. Additionally, de-anonymization was also facilitated, allowing for the secure retrieval of original data values from their anonymized forms using a mapped token system. The second approach combined regex rules with a Named Entity Recognition (NER) pipeline to quickly substitute sensitive data such as SSNs, phone numbers, and email addresses with placeholders, while also detecting and anonymizing personal and location entities in texts. Both methods provided robust frameworks for ensuring data privacy, with Gradio interfaces enabling interactive data processing, verification of the anonymization effectiveness, and de-anonymization capabilities. The script for this anonymization and de-anonymization process can be found on GitHub at https://github.com/wasimsse/MediTrustAI/tree/master/data_anonymization

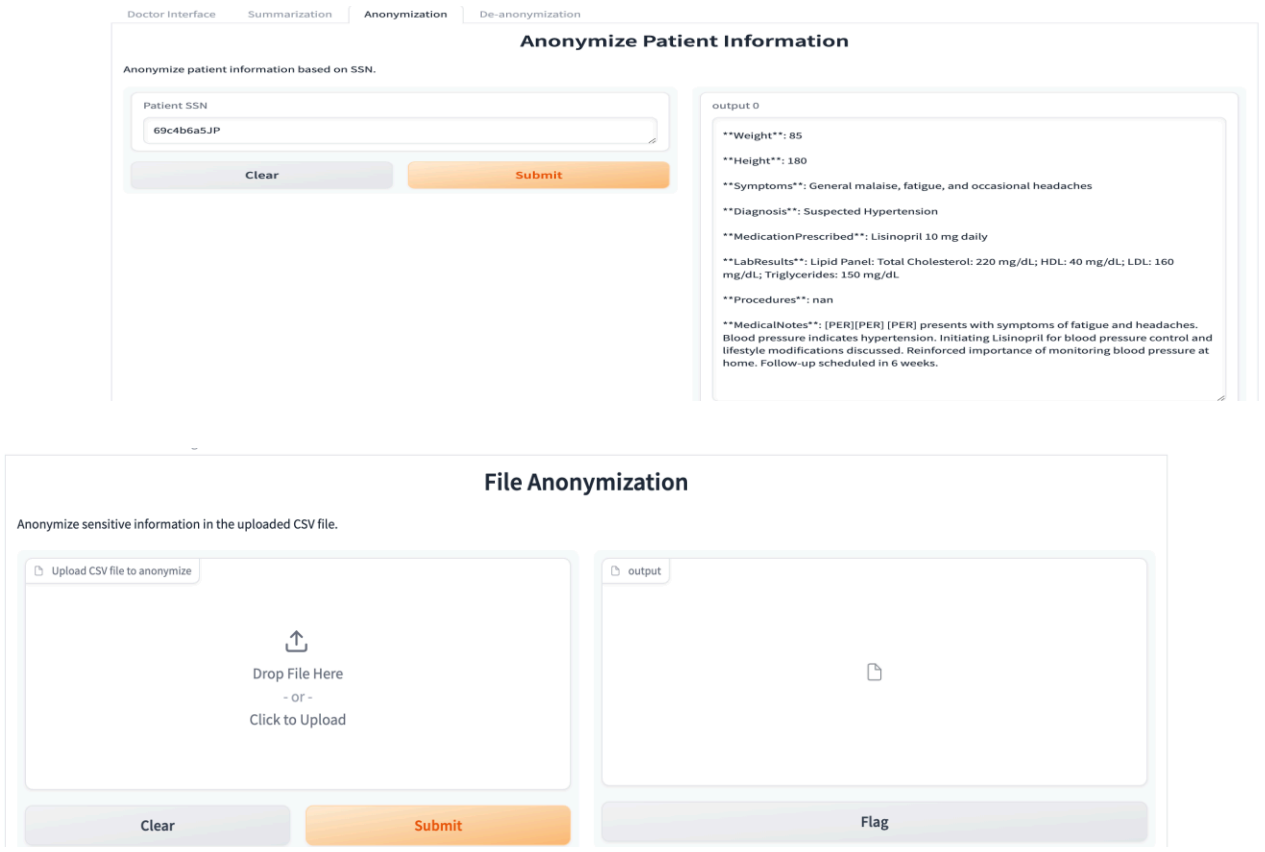
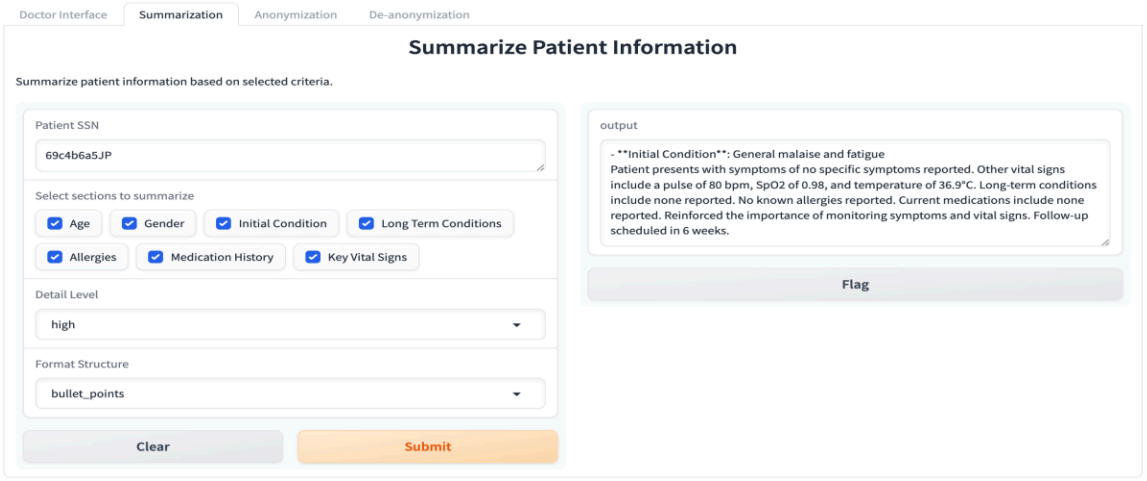


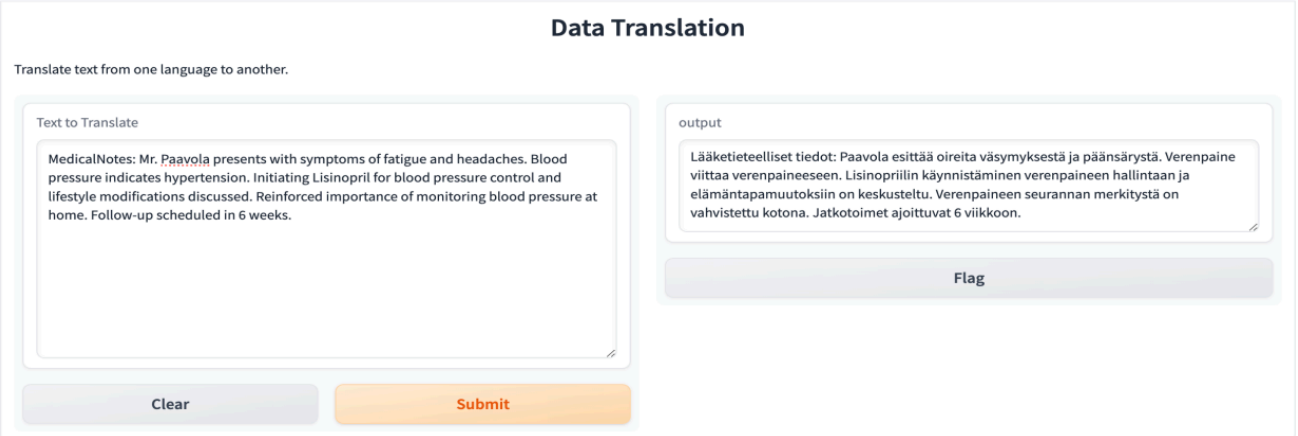
Image 3: File Anonymization Interface

- **Fine Tuning:** After cleaning the synthetic data, they proceeded to fine-tune the models using the CSC computer facilities, specifically targeting the "lama3-openbiollm-8" and "Open BioLLM" models. They utilised the transformers and datasets libraries to streamline this process. The models and tokenizers were initialised from pre-existing settings using AutoModelForCausalLM and AutoTokenizer. They have loaded the dataset using a load_dataset function, which was processed to fit the models. All text entries were tokenized to maintain a constant length and format; it is required for better model training. They also created a TrainingArguments class with several parameters such as learning rate, batch size, and the number of training epochs. The training was conducted with the Trainer class, fine-tuning the models to have an advanced understanding and generation of biomedical text. Lastly, the trained models and tokenizers were saved into the output directory specified, packaging them for deployment and further use. The detailed scripts for the fine-tune module are available on GitHub at https://github.com/wasimsse/MediTrustAI/blob/master/llm_fine_tuning.py .
- **Text Summarization:** They enhanced the frontend of their system so that a doctor could view the important history details of the patients including chronic diseases, medications, allergies, etc. They also made a text summarization module with the transformers and sentence transformers libraries. The module uses the 't5-small' model for generating short, informative summaries about patients, and the Sentence Transformer model 'paraphrase-MiniLM-L6-v2' for refining it. This redundancy reduction function makes sure that there are no redundant sentences. The algorithm achieves this by computing the cosine similarity between the sentence embeddings. All sentences with a similarity higher than 0.8 are discarded, ensuring that each sentence contributes something different to the summary. This step is very important in the necessity to have a short and relevant summary. It checks the accuracy of the summaries by cosine similarity between the summarised content and the original text to ensure that summaries are reliable and consistent with the original data. Detailed scripts for

text summarization module can be found on GitHub: https://github.com/wasimsse/MediTrustAI/tree/master/text_summarization



- **Translation**: They have integrated a translation feature in our system using the transformers library with MarianMTModel and MarianTokenizer, both pre-trained on the model of Helsinki, for English to Finnish translations. This will allow translation of medical or general text from English to Finnish, while at the same time, making it easy for medical practitioners to communicate with and access medical records or information prepared in a language other than English. The process is, at a high level, tokenization of the English text; translation to Finnish using the model; and regeneration of the decoded tokens back to string format for readability. So, this would be very useful in many healthcare scenarios that are highly diverse due to its value in communicating across languages. The other function that has also been optimised and considered on par or better than every other optimization here, is the translation function, being able to process a batch of texts. This is meant to be optimised for processing time and to improve the user experience, where the resulting translations are both fast and accurate for multiple entries at a time. The script for fine-tuning is at GotHub: https://github.com/wasimsse/MediTrustAI/tree/master/data_translation



- **Chatbots**: They also demonstrate the use of Gradio and Databricks technologies to build interactive, intelligent systems that enhance healthcare operations. The first script leverages

Gradio to create a user-friendly graphical interface, enabling medical professionals to perform tasks such as anonymization, translation, and summarization of medical texts directly through the web. This interface is connected to backend services hosted on Databricks, ensuring secure and efficient data processing. The second script expands on this by incorporating Databricks' vector search and LangChain capabilities to power a health chatbot capable of handling complex medical inquiries. It uses advanced NLP models to retrieve relevant medical information and generate precise answers, integrating embeddings and vector search to access and utilise data from extensive medical databases. Together, these implementations highlight a robust integration of state-of-the-art technologies, streamlining medical data processing and enhancing user interaction within healthcare environments. The script for these chatbots are available on this link https://github.com/wasimsse/MediTrustAI/tree/master/databricks_demo

In less than two months, they've made significant efforts to demonstrate proof of concept for anonymization, summarization, and synthetic data generation, collaborating closely with experts in technology and the medical field. While they've made promising progress, we acknowledge that further development is required to evolve these initial models into a fully- fledged prototype. Summary of each contribution is listed below.

- **Synthetic Medical Data Creation:** They generated a synthetic dataset that mimics real patient information without compromising individual privacy. This dataset helps in research and is safe and repeatable for scientists**.**
- **Data Anonymization Techniques:** They used advanced tools to anonymize and de-anonymize healthcare data, ensuring it stays private and secure while complying with legal standards like the GDPR**.**
- **Fine-Tuning Models:** They fine-tuned healthcare-specific models to better understand medical terminology and contexts, enhancing their ability to work with real-life medical data**.**
- **Text Summarization and Translation:** They developed tools to summarise medical documents and translate between English and Finnish, making it easier for healthcare professionals to access and understand medical records quickly.
- **Healthcare Chatbots:** By integrating Gradio and Databricks, they set up smart chatbots that handle complex medical questions efficiently, enhancing how healthcare professionals retrieve information and interact with patients.
- **Interactive GUI Development:** They created a user-friendly graphical interface using Gradio that allows medical staff to easily interact with AI for tasks like summarising and translating texts without needing technical skills.

Future plans

Building on this early achievement, they will take steps to constitute even more efficient and more personalised health delivery by enhancing patient history access. They will work to ensure
trust of the public in the use of AI in healthcare through the following steps of stringent data security:

- **Data Security**: They will use the latest data-encoding methods to ensure better patient data security.

- **Bias Mitigation**: To do this, they will use AI Fairness 360 to detect and reduce any biases for equality of healthcare services
- **Transparency in AI**: XAI tools, such as LIME and SHAP, will be integrated to make the AI decisions transparent and explainable.
- **Evaluation and Development**: They will do proof of value assessments that test whether our innovations actually work in real-life settings.
- **Intellectual Contributions**: They will continue to publish our findings while filing potential patents for the greater AI and healthcare community and commercialising them.

Progress against planned goals

Despite our well-laid plans, less than two-month deadline provided by TAILOR made it challenging to fully achieve our objectives:

- **Data Security Enhancements**: The basic encoding techniques were started off first for data security improvements, but due to time constraints, we could not establish the thorough standards of data protection as desired.
- **Bias Mitigation Initiatives**: They started using AI Fairness 360 for the identification of biases, and systematisation is ongoing.
- **Transparency in AI**: Preliminary attempts have been made with tools like LIME and SHAP to open AI-based decisions for reasons of transparency. However, much more remains to be done
- **Practical Evaluations**: They set up a demo for testing our applications but haven't managed to carry out these tests yet.
- **Academic and Intellectual Contributions**: Academic papers for publication are under preparation

Self-assessment

- AI Excellence: The TrustedAI project substantially contributed to the integration of trust in AI applications for healthcare. We focused on enhancing data privacy and personalising patient care according to GDPR standards. These efforts enabled the value addition of AI to healthcare and greatly furthered the ability to obtain data security in a way that allows more personalised treatment options.
- Scientific step-up: This project further developed my professional and scientific experience. Assuming the role of AI integrator in practical healthcare enabled insight into unique challenges and opportunities at that junction. The partnership with the University of Jyvaskyla and Tietoevry Oy has expanded my experience base and positively influenced my scientific reputation.
- Suitability of the host: The setting for this research was provided by Tietoevry Oyj. Their commitment to innovation and robust technological infrastructure positioned us well to tackle many complex challenges. The collaborative environment was crucial for the interdisciplinary nature of our project, facilitating productive exchanges of ideas. Special thanks to Contact Person: Iftikhar Ahmad (iftikhar.ahmad@tietoevry.com), who provided significant guidance at every step, ensuring the project stayed on track and aligned closely with industrial and societal needs.
- Suitability of the visit length: The allotted time for this project was quite restrictive. A longer duration would have allowed for a more thorough exploration of our research objectives and further refinement of our AI tools. Extended time could have enhanced the depth of our testing and ensured a smoother integration of AI technologies into clinical practices.

TEC4CPC - Towards a **T**rustworthy and **E**fficient **C**ompanion for **C**ar **P**art **Catalogues**

Organisation: N4

Public summary

N4 publishes procurement platforms for the automotive sector. Especially car parts catalogues (N4Parts), that are used by customers to purchase parts and to obtain information (installation instructions, maintenance intervals, etc.), require a deep understanding of vehicles to obtain the necessary information for procurement or repair jobs. Besides automotive specific know-how, users with different skill levels in the usage of digital media may face challenges when using a browser-based application. All product information is based on data provided by master data suppliers in a given format, making it difficult to adjust the information (N4 would be liable for the truthfulness of the data) or the presentation of the core data itself. Based on these circumstances, it is necessary to provide a solution to find and display all necessary data without a deep knowledge of the master data or N4Parts itself. A chatbot could provide a solution, because users could simply ask for the location of an information (path) or request the information itself, without any navigation necessary. However, a chatbot might also have risks: the quality of the answers may vary or - based on the user's experience - for some users using the chatbot might actually be less efficient compared to classical interaction with the system. This leads to the questions for which users a chatbot might work and what advantages users and companies gain by using the chatbot.

Research objectives

The aim of the research project is to implement a chatbot companion and evaluate it concerning its ability to provide quality- and efficiency gains in the user interaction with N4Parts.
Integration of LLM chatbot with retrieval augmented generation into the parts shop software.
The integration of a Large Language Model (LLM) chatbot with Retrieval Augmented Generation (RAG) into the parts store software is done for several reasons, with increased trustworthiness at the forefront. RAG combines the power of an LLM with database-driven information retrieval, ensuring that the chatbot's responses are both accurate and up-to-date. This hybrid approach allows the chatbot to access a wide range of stored information and process it in real time to provide users with reliable and contextually relevant answers. Integration into the store software takes place via an API interface that connects the chatbot with the existing database and the store's backend. This allows the chatbot to efficiently process customer inquiries, make product recommendations and provide detailed information on spare parts.

Quality: Can users rely on the information provided by the companion?
Language model (LLM) hallucinations can result in incorrect answers that lead to poor decisions, such as ordering the wrong parts or following incorrect installation instructions. This issue is critical, especially in contexts where the LLM's responses guide the purchase of items. Therefore, the reliability of the information provided by the chatbot is paramount. Users must be able to trust the answers, and sources should be transparently communicated for cross-checking. Additionally, the information presented should be reasonably prioritised to ensure relevance and accuracy. Our project will evaluate the LLM-based companion with respect to these aspects in an in-house user study with participants having different amounts of prior experience with N4Parts.
Efficiency: Can users save time in information retrieval by using the companion?
Given a wide range of different user-groups (Personas) that will use the catalogue on a daily basis, it is necessary to provide the best user experience for each of these groups. Since the affinity for the use of digital media varies within the user groups, the solution should provide additional user journeys and meet different needs. If a user already has a lot of experience with N4Parts, they might be able to

more efficiently reach their goals via manual navigation, without the use of a chatbot companion. For a novice user however, the additional effort needed to formulate their request to the companion might pay off in reduced times needed to reach the desired goal. Furthermore, users who do not have any experience with generative AI might also struggle in efficiently using the chatbot. The goal of TEC4CPC is to evaluate the different efficiency gains or -losses from the usage of the chatbot companion in different user groups.

Detailed Description

The initial solution consists of three services that allow the retrieval of information from car manuals and question-answering using this information: A Llama3 language model that utilises retrieval-augmented generation (RAG), a Chroma vector database to store embeddings of car manuals and an API backend that N4's products can later connect to.
The system can retrieve real-world HTML car manuals via SOAP endpoints provided by the company TecAlliance GmbH. It parses and splits the content of these documents into smaller chunks and computes embeddings via the open-source GPT4All language model. The resulting embeddings as well as relevant metadata of the source documents are then stored in the Chroma Database for more efficient retrieval. The API manages users, chats, and document metadata in an SQLite Database to allow the continuation of previous chats and automatization for the retrieval of manuals. It provides endpoints for queries to the Llama3 language model using the most relevant embeddings found in the manual documents as context providing backlinks to the source documents. An Ollama service was used to provide the language model for easier experimentation with various models and settings while Langchain was utilised for fast prototyping of the RAG pipeline and experimenting with various text-splitters and embeddings.

Results

As stated above, initial research questions mainly addressed the following questions:

- Answer validity: Are the answers relevant and helpful?
- Time efficiency: Do users save time?

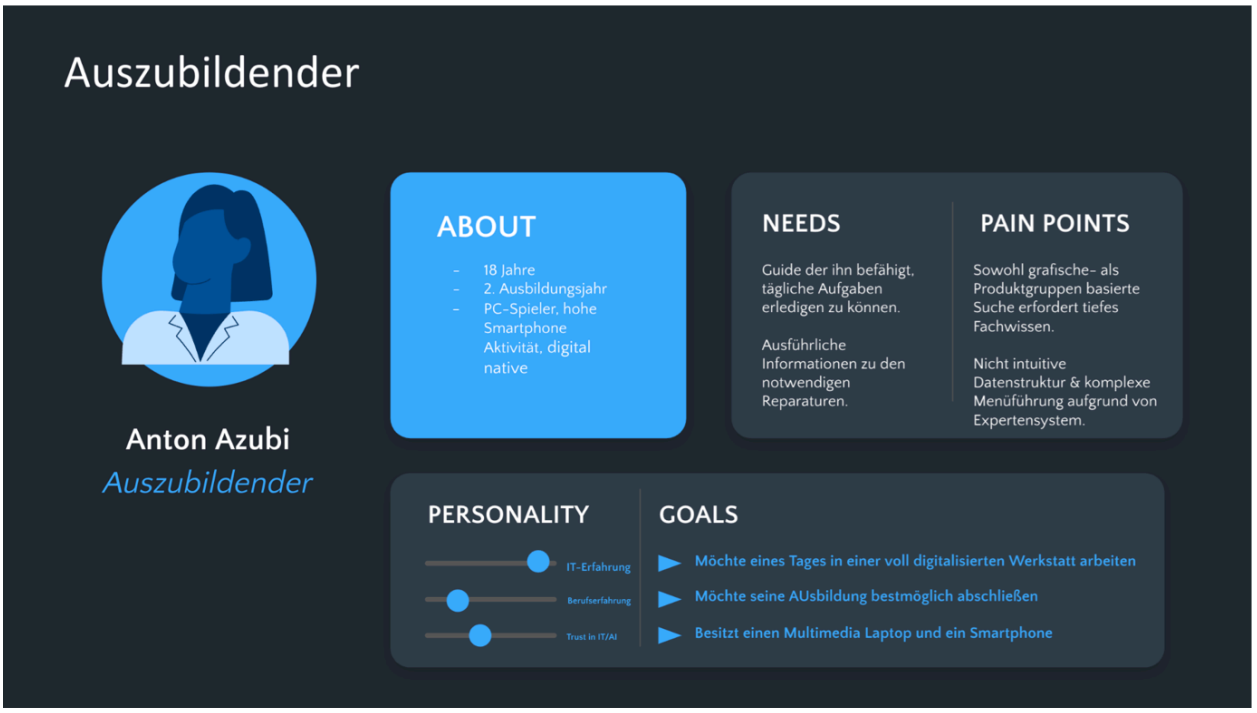
With regards to the evaluation, N4 draw on standard conceptualizations that seek to capture the relevancy, correctness and completeness of a RAG system. Additionally, they tried to compare times obtained in the chatbot-supported settings to those coming from human candidates.
For the sake of clarification, standard metrics for the evaluation of a minimal test set can be seen in the table below: For a given set of queries and responses, they manually annotate whether relevant documents (i.e. documents that contain the info) are identified during the retrieval stage and whether the answer provided uses these documents from the retrieval step as context in the generation step.
As N4 developed the solution in an iterative manner, they quickly came to realise that they need more experiments with regards to different chat templates (used for different personas) and with potential enhancements to the data ingestion pipeline and retrieval stage. Especially due to the amount of imprecise answers, they enhanced the current system. As can be seen above, even if the correct document has been retrieved, this does not mean that it will be part of the generated answer. To mitigate this behaviour, N4 added a richer metadata structure to the documents and complemented the vanilla retriever with a self-querying retriever. The self-querying retriever transforms natural language into a structured query that can be applied to the underlying vector store.

Item	Item Text	Correct Anser in Context	Context Info in Answer	Notes
1	Wie tausche den Innenraumfilter?	No	No	Provides generic answer.
2	Was ist der Feststellwert für die Ölablassschraube?	Yes	Yes	Uses exact figures.
3	Wie programmiere ich einen Schlüssel?	Yes	Partially	Level of granularity is too rough; no info on distinct tools.
4	Welches Schaltgetriebeöl muss ich für das Fahrzeug verwenden?	Yes	Partially	Provides generic answer. Shows need to include more (relational) data.

As can be seen in this table, these enhancements increased the correctness and completeness of the answers. Yet it is important to note that these modifications also involve higher investments in terms

of development effort which is needed to test and fine-tune the respective constructor and provide it with custom examples that teach it how natural language queries are translated into adequate filters. Taken together, the observations stress the importance of data quality, data understanding and data engineering. Without a solid basis neither the vectorsearch nor the extraction of powerful filters can work reliably.

During the initial phase of the project, N4 developed user personas to guide the testing and evaluation processes. These personas were categorised into three distinct levels: beginner, intermediate, and expert. Each category was carefully defined to reflect varying degrees of experience and familiarity with both automotive content and digital tools. These definitions serve as a foundational framework for future, more extensive testing that will take place in subsequent phases of the project. Due to the limited timeframe of this phase, conducting a representative survey was not within the scope of our research. Instead, they focused on laying the groundwork for more comprehensive studies to follow.



Werkstattmitarbeiter

Tina Toolkit

Werkstatt Mitarbeiterin

ABOUT

- 34 Jahre
- Seit 16 Jahren im Peugeot Autohaus
- Bouldern, PC Gaming

NEEDS

Companion der ihr hilft, tägliche Aufgaben schneller zu erledigen.

Kurzübersicht mit allen kritischen Infos mit der Möglichkeit Details abzurufen.

PAIN POINTS

Fülle an Informationen ist teilweise hinderlich bei dem schnellen Informationsbezug.

Navigation durch Stammdaten nicht immer einheitlich, dadurch Schwierigkeiten vor allem bei fremden Fabrikaten (!=Peugeot)

PERSONALITY

- IT-Erfahrung
- Berufserfahrung
- Trust in IT/AI

GOALS

- Möchte effizienter arbeiten
- Möchte befördert werden zur Werkstattmeisterin
- Möchte langfristig AI Tools in den Alltag einbinden

Werkstatt Meister

Manfred Meister

Werkstatt Meister

ABOUT

- 58 Jahre
- Seit 40 Jahren im Peugeot Autohaus
- eingeschränktes Seh- und Hörvermögen
- Lesen, spazieren, Märklin, Oldtimer

NEEDS

Companion der ihm hilft, tägliche Aufgaben schneller zu erledigen.

Kurzübersicht mit allen kritischen Infos mit der Möglichkeit Details abzurufen.

PAIN POINTS

Informationen sind nicht leicht zugänglich.

Besonders große Hürde bei ihm unbekannten Fahrzeugen.

Hohe Komplexität bei Vielzahl an neuen Hard- und Softwareprodukten.

PERSONALITY

- IT-Erfahrung
- Berufserfahrung
- Trust in IT/AI

GOALS

- Möchte ein simples System einsetzen
- Goal 2
- Goal 3

In the initial testing phase, they focused on evaluating time savings for users. Importantly, times for the non-assisted scenario are needed to measure how long users without chatbot support interact with the part finding interface to find answers. To do so, N4 conducted trials with a non-representative sample across all personas. The results already suggest that inexperienced users with limited automotive knowledge but high IT affinity struggle significantly in locating information, often fail to find the desired result or, if they find answers, this takes an extended amount of time. While these observations are insightful, they are preliminary. This means that the basis of the test data, they cannot report meaningful results on the effect of different personas. This also holds true for the data collected in the assisted scenario. As can be seen in the second table, different personas that have been modelled

by initialising the chatbot with distinct templates mainly point towards the fact that it is harder to meet the distinct requirements of expert users. Especially for them, N4 expects incoming queries to be more specific and on point. With the given data structure, efforts are needed to resolve ambiguities and push the relevant features in the data set that satisfy the user needs.
To gain a more precise understanding of the potential time savings and efficiency improvements, it is imperative to conduct further testing with a larger, more representative sample. This should involve employing appropriate methodologies tailored to capture the expected concrete time savings across different user profiles.
As a final remark, N4 tested cases where the user input is out-of-topic (see item 11 above as a reference). Overall, the chatbot recognizes these and, in a friendly manner, refuses to provide answers.

Future Plans

Building upon the foundational work accomplished in the TEC4CPC microproject, several avenues have been identified to enhance the system's capabilities, broaden its applicability, and ensure its robustness. The following outlines the strategic directions envisioned for the project's evolution:
First, to refine the chatbot's efficacy, extensive user studies encompassing a broader spectrum of personas are planned. These studies will delve deeper into user interactions, capturing nuances across varying expertise levels and familiarity with digital tools. Feedback mechanisms will be established to continuously gather user insights, enabling iterative improvements and fostering a user-centric development approach.
Second, recognizing the pivotal role of data quality, efforts will be intensified to enrich the metadata associated with car manuals and other relevant documents. This includes standardising data formats, enhancing the granularity of information, and ensuring timely updates. Implementing more sophisticated data validation and cleansing processes will further bolster the reliability of the chatbot's responses.
Third, with regards to the production environment and as the user base and data volumes grow, scaling the system to maintain performance becomes imperative. Future work will focus on optimising the retrieval and generation pipelines, leveraging cloud-based solutions, and employing efficient indexing strategies to ensure rapid and reliable responses. In production at the latest, ensuring the privacy and security of user data remains a top priority. Future developments will incorporate regular security audits and vulnerability assessments that will be conducted to safeguard the system against potential threats.
By pursuing these strategic initiatives, the TEC4CPC project aims to evolve into a comprehensive, trustworthy, and efficient companion for users navigating the complex landscape of car part catalogs and beyond. The envisioned enhancements will not only elevate user experience but also position N4 at the forefront of innovative solutions in the automotive procurement sector.

Advancing Alzheimer's Diagnosis: Multimodal Explainable AI for Early Detection and Personalized Care

Organisation: University of East London

Public summary

Alzheimer's disease (AD) is becoming more common, emphasising the need for early detection and prediction to improve patient outcomes. Current diagnostic methods are often too late, missing the opportunity for early intervention. This research seeks to develop advanced explainable AI models that combine various types of data, such as brain scans, genetic information, and health records, to identify early signs and risk factors of AD. By doing so, the project aims to enhance early diagnosis, reduce healthcare costs, and enable personalised treatments. This innovative approach could reveal new biomarkers and improve our understanding of AD, potentially transforming how we predict and manage the disease. By making the model's predictions transparent and explainable, clinicians can understand the driving factors behind assessments, enhancing trust and patient communication. Additionally, explainable AI addresses ethical concerns related to bias, fairness, and accountability in automated decision-making.

Research objectives

This research addresses the scientific challenge of improving early detection and personalised treatment strategies for Alzheimer's disease (AD) within the healthcare industry. The main objective is to develop a multimodal explainable deep learning model capable of integrating diverse data sources, such as neuroimaging, genetic markers, and health records, to accurately predict AD onset in genetically predisposed individuals. This challenge is critical due to the increasing prevalence of AD and the limitations of current diagnostic methods, which often rely on late-stage symptomatic manifestations. By addressing this challenge, the research aims to enhance early detection, reduce healthcare costs, and improve patient outcomes, ultimately advancing the field of neurodegenerative disease management. One of the additional objectives is to scrutinise the contributions of diverse data types (e.g., neuroimaging, genetic markers, psychometric responses, demographics, health records, and blood test results) to the prognostication of AD markers and risk factors.

Detailed description

This project will develop a multimodal deep learning model for predicting Alzheimer's disease (AD). Data from various sources, including neuroimaging scans, genetic markers, and clinical records, will be collected and preprocessed. Features will be extracted from each modality, and a deep learning architecture will be used to integrate the data. The model will be trained and evaluated using standard metrics, with a focus on interpretability. Techniques such as feature importance analysis and visualisation will be employed to enhance explainability. Compared to existing approaches, this project will emphasise a wider range of data modalities and enhanced model interpretability.

Comparison with State of the Art:

This project's approach will involve integrating diverse data modalities and emphasising model interpretability, setting it apart from existing methods. While the model may be more complex, it is

expected to achieve superior performance in AD prediction. This comprehensive and interpretable approach will represent a significant advancement in the field of AD research.

Results

This work is expected to yield novel insights into the early detection and prediction of Alzheimer's disease (AD) through the development of a multimodal machine learning model. By integrating diverse data modalities such as neuroimaging, genetic markers, and clinical records, the model aims to uncover new biomarkers and pathological mechanisms underlying AD development.
The application of advanced machine learning techniques to multimodal data integration may lead to the discovery of previously unrecognised patterns or relationships between different AD risk factors. These discoveries could advance our understanding of the disease and inform future research directions.
The developed multimodal machine learning model has the potential for technology transfer, with applications in both clinical practice and research settings. By providing early and accurate predictions of AD onset, the model could assist healthcare professionals in making informed decisions regarding patient care and treatment strategies.

Future plans

In the future, the plan is to further refine the multimodal machine learning model by incorporating additional data sources and optimising the model architecture for even greater accuracy and efficiency. There is also an intention to explore opportunities for commercialization, with the potential to partner with healthcare organisations or technology companies to integrate the model into clinical practice. Collaboration with experts in neurology, genetics, and machine learning will be essential for advancing the research and translating it into real-world applications. This collective effort aims to improve early detection and personalised treatment of Alzheimer's disease, ultimately making a positive impact on patient care and healthcare systems worldwide.

Lessons learned from the Congress

The congress provided many important insights that will lay a solid foundation for future events. Perhaps the most significant revelation was how post-pandemic travel habits are changing. Participants are increasingly reluctant to travel, making it more challenging to host on-site workshops. This shift led to a noticeable number of no-shows, some due to genuine issues, but others without any prior notice, which affected the overall attendance and atmosphere.

Another critical takeaway is the importance of selecting the right topics for discussion. The relevance and appeal of the workshop topics are crucial for attracting and retaining participants. In this case, the selection of working groups and discussion topics was essential in driving engagement. However, the rapid organisation of the event posed some challenges, highlighting that high-level workshops are best prepared with ample lead time.

Despite these challenges, the event underscored that in-person networking remains highly valued, even in the face of difficulties. Many attendees expressed gratitude for the opportunity to network face-to-face after years of virtual events. Approximately 60 participants were unanimous in their appreciation of the event's quality, particularly the well-designed individual sessions, catering, and the enjoyable evening program on the first day. The overall atmosphere of the event was well-suited to its goals, which were broadly welcomed by participants.

In summary, the post-pandemic reality adds a new layer of complexity to organising in-person events, bringing issues like travel reluctance and the need for longer preparation times to the forefront, while also confirming the enduring importance of on-site participation. Considering these factors, it is recommended that future events carefully select topics and provide ample time for planning. This will ensure alignment with participant interests and help continue to create meaningful experiences that advance professional development and build strong networks.

Conclusion and further Outlook

This is where the success stories of working groups, especially their trials around micro-projects, come to the fore. To the extent that they have built bridges between research institutions and industry, some of these small-scale programs have begun to foster a two-way flow of expertise. This approach helps in successful delivery by addressing industry needs immediately with short, focused projects. This predilection to contribute with such quickly initiated, smaller projects instead of larger longitudinal studies is a sign of the industry's wish for agility and applied relevance. These micro-projects have the added benefit of advancing applied research while encouraging sustainable practices in corporate settings. Instead, with its emphasis on concrete, well-defined issues, even fairly modest investments have been able to show striking progress in finding real-world solutions. The imperative to maintain a flexible and agile research framework is set up exactly in response to the needs of industry partners.

Vast potential exists for ongoing and new applications of the micro-projects to further enhance the overall impact as an adjunct to the Industry Collaboration & Transfer Group. Achieving long-term sustainability and expansion would require a sturdy funding mechanism, for example, in the form of an industry-supported annual membership fee. Such a model would provide a reliable source of funding and incentivize continuous collaboration between academic researchers and industry partners.

Micro-projects may eventually grow to include other industries and research areas. This would further increase the number of scopes and people attending to the localised project, implying a positive impact for the Industry Collaboration & Transfer Group. This can also entail the development of new instruments and frameworks to organise and evaluate projects more efficiently, permitting cross-organization or even inter-company cooperation with accurate project outcome metrics.

On the whole, the micro-projects set a successful first step toward industry progress in knowledge transfer of research. Supported by an ongoing iterative process of format optimization and sustained funding, the Industry Collaboration & Transfer Group remains an important driver in connecting research to practical application across academia and industry, with a focus on addressing real-world challenges.