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

D8.1 Report on Theme Development Workshops

Document type (nature)	Report
Deliverable No	D8.1
Work package number(s)	WP8
Date	25.08.2023
Responsible Beneficiary	DFKI, P26
Author(s)	Janina Hoppstädter, Kyra Kiefer, Philipp Slusallek, Marlies Thönnissen (in alphabetical order)
Publicity level	Public
Short description (Please insert the text in the Description of Deliverables in the Appendix 1.)	Summary of the Organisation and the Results from the Joint series of Theme Development Workshops during the report period.

History			
Revision	**Date**	**Modification**	**Authors**
v1.0	5.08.2023	First version	DFKI - Janina Hoppstädter, Kyra Kiefer, Marlies Thönnissen

Document Review		
Reviewer	**Partner ID / Acronym**	**Date of report approval**
Fredrik Heintz	#1, LiU	05.09.2023
Holger Hoos	#7, LEU	14.09.2023

Table of Contents

Introduction to the Deliverable

This document is a deliverable of Work Package 8 “Industry and Innovation Programme”. The objective of WP8 is to develop synergies and cross-fertilization between industry and the TAILOR network of excellence centres to provide the basis for Trustworthy AI in Europe and to run and organise so-called Theme Development Workshops in their areas. The objective of such a workshop is to bring together key players from specific industry sectors with key AI researchers and other stakeholders to jointly identify the key research topics and challenges, outline priorities and develop and provide input to strategic roadmaps for AI in a certain area or for a specific industry sector. These roadmaps can then be further refined, e.g., over the course of the projects, thereby forming the basis for ongoing and future AI research, industry collaboration, and corresponding funding.

This deliverable serves to highlight and outline the results of the joint series of the sector-specific Theme Development Workshops led by TAILOR that emerged as a result of the work of WP8.

The document is structured as follows:

- Chapter 1 outlines the concept and idea of a sector-specific Theme Development Workshop and describes how this innovative format can be embedded in the broader context of strategic AI research and innovation agendas and roadmaps, and the template/blueprint as a process with five phases with the corresponding activities (process steps). It also provides a checklist derived from the comprehensive process description, which can be used by the Networks of AI Excellence (NoE) to support their own planning process of a TDW as well as lessons learned from the previous TDWs.
- Chapter 2 provides an overview of the past sector-specific Theme Developments and a comprehensive description of the results, in particular their contribution to the Strategic Research and Innovation Roadmap[1].
- Chapter 3 closes this deliverable with some further insights and conclusions as well as an outlook on planned activities in the future.

[1] See https://tailor-network.eu/research-overview/strategic-research-and-innovation-roadmap/

Organisation

The following people have been involved in the Deliverable:

Partner Acronym	Name	Role
DFKI	Janina Hoppstädter	Senior Project manager
DFKI	Kyra Kiefer	Project manager
DFKI	Philipp Slusallek	Project lead
DFKI	Marlies Thönnissen	Project manager

Executive summary

In 2021, the concept of the Theme Development Workshop was initially tested and successfully implemented within TAILOR and the respective industry sectors. Due to Covid-19 and the corresponding travel restrictions, the organisers were forced to reschedule the initially planned on-site workshops to an online format. Therefore, the concept was adapted and supported virtually. Mechanisms were selected for implementation and tested thoroughly before the workshops were held. The high participation of research and industry participants in all workshops reflects the success of the online events and first collaborations and transfer actions have already been started.

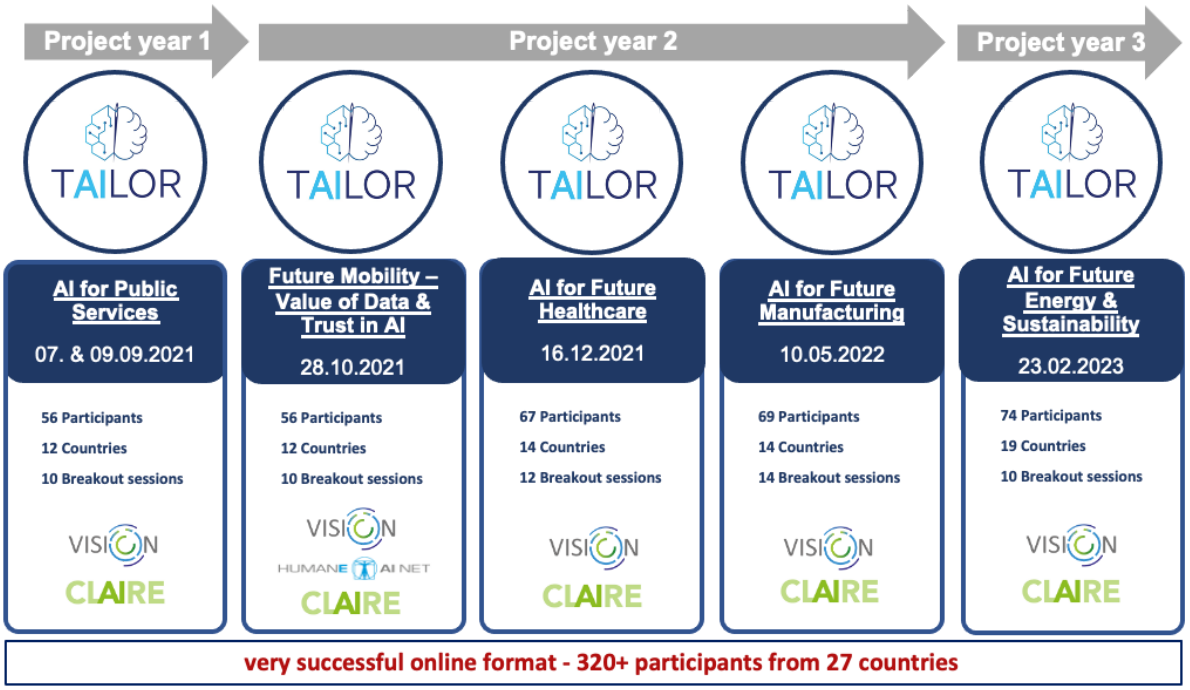

Figure 1: Overview of the past Joint series of Theme Development Workshops

In the workshops of the individual industry sectors, current topics and questions were elaborated that have also been included in the creation of the European Research and Innovation Agenda.

The first "AI in the Public Sector" workshop, organised by CLAIRE, TAILOR, and VISION, culminated in pivotal outcomes shaping AI's role in government functions. Experts from diverse sectors converged to discuss AI applications in urban mobility, safety, security, statistics, and innovative public services. Crucially, the workshop emphasised transparency, ethics, accountability, and trust as key components for AI's success in governance. The breakout sessions spotlighted specific challenges and opportunities in areas like public safety, urban mobility planning, reliable statistics, next-generation public services, AI expertise development, accountability mechanisms, and more. These discussions underscored the complex interplay between AI and the public sector, leading to the identification of themes that will significantly influence the European AI research and innovation roadmap. These themes encompass education, algorithm registers, AI ecosystem performance measurement, and data availability, quality, and accessibility. The workshop's

outcomes lay the groundwork for AI's responsible integration into public services, fostering a commitment to ethical considerations and societal benefit.

The second Theme Development Workshop "Future Mobility - Value of Data and Trust in AI", organised by CLAIRE, HumaneAI Net, TAILOR, and VISION, yielded significant outcomes for advancing AI in the mobility and transportation sector. With 56 participants from diverse backgrounds across 14 countries, including major industry representatives and European Commission members, the workshop achieved robust engagement. Notable keynotes addressed the importance of trustworthy AI, emphasising data value and model transparency. Breakout sessions focused on critical themes, such as Trustworthy AI, Explainable AI and Verification, Data Value Estimation, AI Expertise, AI Sensitivity Analysis, Energy Autonomous Assets, and AI and GDPR. These discussions culminated in actionable outcomes for the European AI research and innovation roadmap. Calls were made for a holistic approach to Trustworthy AI, standardised, accessible data to drive AI advancements, transparent communication, fostering AI expertise development, addressing AI sensitivity and safety, and navigating the intersection of AI and GDPR. These outcomes reinforce AI's growing impact on mobility and transportation and provide valuable direction for responsible AI integration in Europe.

The third Theme Development Workshop "AI for Future Healthcare", organised by CLAIRE, TAILOR, and VISION, generated pivotal outcomes for advancing AI in healthcare. Attended by a diverse audience from 14 countries, including industry leaders and European Commission members, the workshop achieved robust engagement. Keynotes stressed trustworthy AI and data privacy, while breakout sessions delved into themes like Data Governance, Trustworthiness, and AI Literacy. These discussions converged into actionable insights for the European AI research and innovation roadmap, advocating data privacy, emphasising trust in AI, and promoting AI education. These outcomes bolster AI's role in healthcare and provide vital guidance for responsible integration across Europe.

The collaborative effort of CLAIRE, TAILOR, and VISION also resulted in the fourth impactful "AI for Future Manufacturing" Theme Development Workshop. During this event, experts from diverse fields converged to deliberate on pivotal AI subjects within the manufacturing sector. The discussions encompassed various dimensions, such as trust, robustness, human-robot collaboration, and the challenges of data labelling. These insights have been instrumental in shaping the European AI roadmap. The workshop not only underlined the significance of explainable AI, worker inclusion, and dynamic risk management but also highlighted the potential synergy between advanced technologies like Edge Computing, Edge AI, and high-performance communication systems such as 5G, all of which hold the promise of substantial industrial enhancements.

During the collaborative fifth workshop on "AI for Future Energy & Sustainability," co-organised by CLAIRE, TAILOR, and VISION experts from academia, industry, and politics converged to explore emerging AI topics in the Energy sector. This one-day event laid initial groundwork for the European AI research and innovation roadmap, detailed in a public report on the TAILOR website. Following the workshop's findings, the Organising Committee pinpointed key topics. These include energy efficiency through AI, machine learning for energy optimization, sustainable business models, collaboration, and integrating renewable energy systems. These focal points, while central, will be further enriched by inputs from the

broader Theme Development Workshop. In the Energy sector, emphasis lies on energy efficiency, machine learning's role in optimization, collaboration, and integrating renewable energy systems. Beyond that, there's a need for transparent, trustworthy AI models, effective human-AI interaction, data quality, privacy, and optimization decision-making.

In summary, the TDW methodology has proven to be an appropriate and valuable tool to identify sector-specific requirements. As part of other actions in TAILOR, some of these requirements have been addressed and further progressed into challenges and even to co-innovation labs. A co-innovation lab is another TAILOR concept, in which two or more organisations work closely together to co-invent solutions to agreed problems, i.e. those identified in a TDW.

In addition to the sector specific needs and requirements, a few common needs came out of the five TDWs covered in this report. These are

- education (schools, academia, workforce and general public)
- data management and accessibility
- a common terminology (wrt trustworthiness etc)
- tools to evaluate trustworthiness

Finally, in the centre of all TAILOR workshops is the concept of enabling Trustworthy AI (TAI) through research integrating Learning, Optimisation and Reasoning (LOR). How to achieve this is explored in the TAILOR roadmap, SRIR, v2, due spring 2024.

The concept of sector-specific Theme Development Workshops

In the next subsections, the TDW concept will be placed in a broader context of developing strategic AI research and innovation roadmaps, also referring to insights from scientific literature. Based on this, as well as the authors' previous experiences and success stories in organising such workshops, the adaptations of the TDW format to the specific requirements in the context of ICT-48 will be further elaborated.

Objectives of TAILOR and the innovation activities

The identification of emerging trends, societal and industrial challenges is a difficult process especially in the area of AI which is generally seen as one of the most disruptive technologies of our time. Furthermore, deriving a coherent and prioritised scientific and technical picture that takes advantage of existing assets in such a disruptive field is a particular challenge. In general, developing a roadmap is seen as a useful strategic planning tool to forecast development needs and requirements as well as to define the most important steps and activities in order to achieve major advances in the field addressed.[2] Accordingly, the four ICT-48 Networks of Excellence (NoE) have all defined such activities in their work plan, aiming at the development of a strategic research and innovation agenda/roadmap for AI in their specific topic areas.[3]

Existing AI strategies, e.g., on the European level[4] or in national contexts are basically focusing on political areas of activities and high-level objectives, omitting a concrete definition of AI research and innovation roadmaps. In the German national AI strategy for example, it is explicitly stated that predictions about the development of AI research and its areas as well as about technologies and their use in applications and sectors are not included on purpose[5]. Accordingly, more focused roadmaps like the "AI, Data and Robotics SRIDA"[6] are trying to close this gap. However, also the SRIDA provides a more high-level perspective by identifying five so-called technology enablers defined as equally important to AI, Data, and Robotics, and an indication of some prospective short-, medium-, and longer-term outcomes in these areas.[7] Since these enablers are defined as cross-cutting, relations to specific industry sectors or application areas are not included in these aspired outcomes.

---

[2] See Camarinha-Matos, L.M. and Afsarmanesh, H. (2004): A roadmapping methodology for strategic research on VO. In: Collaborative networked organizations – A research agenda for emerging business models, Kluwer Academic Publishers, p. 275.
[3] See for example TAILOR: "Objective 2 Define and maintain a unified strategic research and innovation roadmap for the Foundations of Trustworthy AI": https://www.tailor-network.eu/about/tailor-objectives/
[4] See Artificial Intelligence Strategy (2018): https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf, p. 12
[5] See Artificial Intelligence Strategy (2018): https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf, p. 12
[6] See Strategic Research, Innovation and Deployment Agenda (Third Release, September 2020): https://bdva.eu/sites/default/files/AI-Data-Robotics-Partnership-SRIDA%20V3.1.pdf
[7] See ibid., p.36-37.

Given that nowadays the lines between pure academic research and development of innovations are more and more blurring, more integrated approaches bringing basic and applied research as well as innovation together in a focused research agenda seem to be needed and more promising. This appears to be of particular importance in the area of AI where tech giants like Google/Alphabet and Facebook are headhunting for academics and finance big research departments producing major breakthroughs in Deep Learning and other areas. Accordingly, new and innovative formats are required to support the co-creation of research and innovation roadmaps between academia and industry, also including other stakeholders to support interdisciplinary perspectives and to work on the corresponding challenges.

One of these innovative mechanisms that the VISION consortium identified and plans to further explore together with the EU ICT-48 NoE (and beyond) are so-called Theme Development Workshops (TDWs). The objective of these TDWs is to bring together key players from specific industry sectors with key AI researchers to jointly identify the key research topics and challenges, outline priorities, and develop and provide input to strategic roadmaps for AI in a certain area or for a specific industry sector. These roadmaps can then be further refined, e.g., over the course of the projects, thereby forming the basis for ongoing and future AI research and corresponding funding.

The Big Picture: Roadmapping activities

Roadmaps are used in academic, industrial and governmental settings for strategic planning as well as defining or mapping structures and relationships between science, technology, and/or applications. The overall objectives are to better align activities and resource allocations in complex and highly-dynamic environments by identifying gaps and opportunities. However, depending on the purpose and context, there is a wide variety of roadmaps, from science/research via industry to product and project/issue roadmaps for example.[8] There seems to be a kind of consensus in scientific literature that roadmaps reflect the knowledge and expertise of a group of visionaries and experts who should be mixed with different backgrounds in order to ensure a balance of perspectives and contributions.[9][10]

Given this broad range of roadmap purposes and objectives, it makes sense to focus the further analysis more on the specific context of developing a research and innovation agenda/program. In the EU research project BRAID (FP7-ICT-2009-4)[11] for example, the consortium developed an extensive Research Technology and Development roadmap for ICT and Ageing, based on existing roadmaps from previous EU research projects. In order

[8] See Kostoff, R.N. and Schaller, R.R. (2001): Science and Technology Roadmaps. In: IEEE Transactions on Engineering Management, vol. 48, no. 2, pp. 132, 134.
[9] See Camarinha-Matos, L.M. and Afsarmanesh, H. (2004): A roadmapping methodology for strategic research on VO. In: Collaborative networked organisations – A research agenda for emerging business models. Kluwer Academic Publishers, pp. 275-288.
[10] See Kostoff, R.N. and Schaller, R.R. (2001): Science and Technology Roadmaps. In: IEEE Transactions on Engineering Management, vol. 48, no. 2, p. 135.
[11] Project website does not seem to exist anymore, but a summary can be downloaded at: https://cordis.europa.eu/docs/projects/cnect/5/248485/080/reports/001-01BRAIDPublishableSummary.pdf

to accomplish this in a systematic way and based on proven scientific methods, a roadmapping method was developed that is depicted in Figure 1.

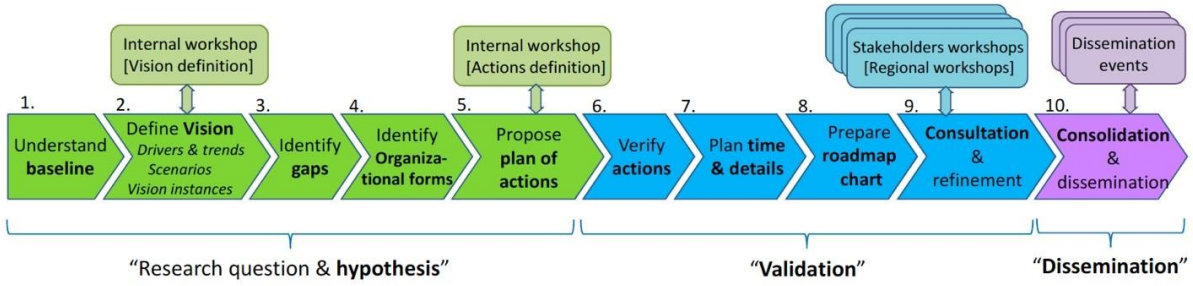

Figure 2: Roadmapping method and steps[12]

In the 10 step process, three main phases are proposed, starting with elaborating the research question(s) based on an understanding of the current state of the art in research and the identification of the roles of all major stakeholders. In the BRAID project, this was done by the consortium to gain an understanding of the baseline. In the next steps, the desired vision for the future is identified and the gaps between this vision and the current state are analysed in order to deduce a first plan of required actions as well as supportive organisational forms and structures. To achieve these results, the BRAID consortium conducted a series of internal workshops and brainstorming sessions until a first set of actions emerged. According to the process, the results from the first phase are then further refined and transitioned to a roadmap chart for validation with various stakeholders in dedicated (regional) workshops. In the BRAID project's "Consensus Building Events" in five European countries as well as two special conferences were organised to consult with more than 150 stakeholders for example. Last but not least, the roadmapping method closes with the third phase of preparing the final version of the roadmap with consolidated results.[13]

Taking this roadmapping method and process as a guideline, a TDW is an innovative format that allows to address several activities at once and thereby spans across different steps of the process. This seems prudent as there is already a lot of preparatory work done by the four NoE under EU ICT-48 regarding the "research question and hypothesis" phase (proposals with state of the art, identification of existing gaps and first ideas, and work plan of how to address them, including a competitive review and selection process by the Commission). Furthermore, the four NoE already comprise more than 150 Centres of Excellence and partners in AI all across Europe, so there is a huge pool of expertise and stakeholder perspectives available, which should be connected and exploited in an efficient manner. Accordingly, from our point of view, and based on our previous experience (see next section), a TDW will mainly contribute to the following process steps:

- (2)[14] Define vision: Input to and refinement of strategic and long-term topics for AI research and innovation in Europe, based on the work plans of the four NoE and their first versions of roadmaps respectively.

[12] Camarinha-Matos et al. (2013): A comprehensive research roadmap for ICT and Ageing. Studies in Informatics and Control, Vol. 22, No. 3, September 2013, p. 235.
[13] See Camarinha-Matos et al. (2013): A comprehensive research roadmap for ICT and Ageing. Studies in Informatics and Control, Vol. 22, No. 3, September 2013, p. 235, 247-248.
[14] Number is referring to the process step in Figure 1.

- (3) Identify gaps: Input to and refinement of barriers and challenges in relation to these strategic long-term topics.
- (5) Plan of actions: First ideas of how to address the identified challenges, including some more focused activities which can be addressed by a smaller group of partners within the NoEs respectively.
- (7) Plan time & details: First rough estimation of _timeline and next steps to address the identified challenges and more focused activities.
- (8) Prepare roadmap chart: Results of the TDW can directly feed into the roadmapping activities of the NoE.
- (9) Consultation and refinement: Given the large pool of expertise and stakeholders (from academia and industry) involved in the four NoE, the results of the TDWs can and will provide a high level of confidence in the adequacy[15] of the proposed research and innovation activities and the derived input to the roadmaps.

Last but not least, we would like to come back to the understanding of roadmaps as living documents. Although this perspective of continuous evolution due to new results and emerging trends is also stated in the paper outlining the roadmapping process[16], a corresponding process step (11) is missing in the proposed method. In the context of a highly dynamic field like AI in general, and given the objectives of the four NoE, specifically "to reinforce Europe's capacity and progress in critical technologies"[17], this continuous improvement and adaptation of the roadmaps is a key success factor to achieve progress in identifying and addressing the major scientific or technological challenges in AI over time. Therefore, we propose to exploit and further develop TDWs as an innovative format to support this highly dynamic and complex process within the NoE. Accordingly, we outline our previous experience with this format and develop a first template for organising and executing such an event in the following.

Previous experience and success factors for a TDW

In spring 2019, CLAIRE[18] and ESA[19] collaborated to organise the first "CLAIRE-ESA Theme Development Workshop on Artificial Intelligence and Earth Observation" which took place at ESRIN[20] in Frascati, Italy with over 70 participants. For this one-day workshop, AI researchers as well as Earth Observation researchers and engineers came together to work on the challenge of how to utilise AI to get the most out of the large amounts of data produced by earth-observing satellites every day. Accordingly, the experts discussed common goals and strategies as well as the next steps to develop new approaches and solutions at the intersection of AI and Earth and Space science.

[15] With regard to confidence in adequacy, see also: Camarinha-Matos et al. (2013): A comprehensive research roadmap for ICT and Ageing. Studies in Informatics and Control, Vol. 22, No. 3, September 2013, p. 251.
[16] See Camarinha-Matos et al. (2013): A comprehensive research roadmap for ICT and Ageing. Studies in Informatics and Control, Vol. 22, No. 3, September 2013, p. 251.
[17] https://cordis.europa.eu/programme/id/H2020_ICT-48-2020
[18] Confederation of Laboratories for Artificial Intelligence Research, website: https://claire-ai.org/
[19] European Space Agency, website: https://www.esa.int/
[20] See: https://www.esa.int/About_Us/ESRIN

This Theme Development Workshop was a huge success, and at the end of the one day event, the participants had collaboratively developed an initial research and innovation agenda, including:

- A strategy for further collaboration (next steps to establish a closer, long-term cooperation between CLAIRE and ESA),
- a list of corresponding challenges and projects to start with, and
- first ideas how to approach these challenges and projects, including some requirements as well as a first timeframe.

In the meantime, some of these challenges and projects have been started, and CLAIRE and ESA have created the World's first AI Special Interest Group (SIG) on Space. The SIG's goal is to accelerate the development of both AI and space-related research, and forms a key part of ESA's wider strategy to accelerate the use of AI in Europe's space and related industries and research communities. It has a very practical focus, organising events and workshops to bring European AI and space experts together, supporting exchanges between these communities and identifying the key themes for collaboration that could be funded by ESA's and other research programmes.[21]

This success story strongly encourages the use of this innovative mechanism of a TDW to provide input to the research and innovation roadmap of the four ICT-48 Networks of Excellence. Given the authors' experience in co-organising and participating in the first CLAIRE-ESA Theme Development Workshop, it is recommended to keep in mind the following aspects that we identified as key to the success of such a workshop:

- Start early with the preparation of the TDW. Depending on the number of participants and logistics required we recommend 3+ months in advance.
- Work with a small and agile team to plan and organise the TDW, involve experts from all areas/research disciplines addressed to create a sound schedule for the workshop day, and to follow-up with them about the results and outcomes of the workshop.
- Select the participants for the TDW according to their proven expertise and most importantly their strategic thinking in their research/working field, and make sure that interdisciplinary groups are assembled especially for the breakout sessions[22].
- The mindset of the participants is key to the success of the workshop: People should be visionary and open to discuss creatively in smaller groups about the specific topics and challenges, and contribute their expertise proactively. Participants must have the longer-term goals (e.g., 5-10 years) in mind that are needed for fundamental research, even though more short-term topics should be covered as well as they come up in the discussions. This includes the collaborative documentation of the group findings during the breakout sessions to ensure a valuable outcome of the TDW beyond networking.
- Motivate the discussions in the breakout sessions with some initial inspiring plenary presentations, focusing on specific and exciting challenges.
- Pre-structure the workshop day and prepare the documents for the breakout sessions according to the scope, objectives, and desired outcomes, but leave

[21] See: https://claire-ai.org/sig-space/
[22] Small group of 5-8 people working on a specific sub-topic for a predefined amount of time during the TDW.

also room for some creative elements and spontaneous ideas, because interesting topics might emerge during the TDW and the discussions in the breakout sessions.

Development of a blueprint to support the NoE

Based on the success and the experiences from the first "CLAIRE-ESA TDW on AI and Earth Observation" and the projects, activities and workshops initiated since then, this innovative approach was further refined. In the context of the AI strategy in Europe, and more specifically under the umbrella of EU ICT-48 and the VISION project as its Coordination and Support Action, the following challenges and corresponding requirements were further addressed:

1. How to best integrate such a workshop format in the context of a NoE, also given its specific nature as a European Research and Innovation Action (project character)?
2. Which adaptations are needed to support academia-industry collaboration in particular?
3. How to deal with the challenges of the Covid-19 pandemic that makes it impossible to physically meet and discuss in larger groups (at least during the first year of the ICT-48 projects)?
4. How can we facilitate the innovation and research transfer via TDWs beyond the European frame of ICT-48?

In order to do so, ideas, requirements and contributions from various sources were further addressed and generalised and integrated into a concept. In particular the following activities supported the concept development:

- First ideas for the adaption of the concept already started in the proposal phase of the ICT-48 call. Based on discussions with industry representatives from the CLAIRE network, specifically the requirements for academia-industry collaboration were discussed (à 2, 4)[23]. Furthermore, with partners involved in the proposals for VISION and TAILOR, ideas for integrating the TDW concept into the networks (à 1) were developed, specifically to fit into the planned activities around roadmapping.
- When the NoE and VISION started, kick-off meetings and workshops were organised as purely virtual events due to the Covid-19 pandemic. This supported the development of ideas how to best organise also the TDWs as virtual events. (à 3)
- Based on these results, an initial concept for TDWs in the context of ICT-48 was developed and discussed in-depth within WP8 "Industry, Innovation and Transfer program" of TAILOR[24]. These discussions included separate 1-hour calls with seven partners from different industry sectors[25]. (à 2)
- Based on the results of these in-depth discussions, the initial concept for the TDWs was refined. This improved concept was then discussed within WP4

[23] Reference in brackets means: Addressing challenges (2) and (4) as described above.
[24] For TAILOR WP-structure, see: https://www.tailor-network.eu/about/organization/.
[25] For TAILOR industry partners, see: https://www.tailor-network.eu/network/industry/.

"Academia-Industry Joint AI Forces" of VISION (à 2). In parallel, the concept was also presented and discussed with the coordinator and two partners from HumanE-AI-Net, representing WP6 "Applied Research with industrial and societal use cases" and WP7 "Innovation Ecosystem and Socio-Economic Impact"[26] of the NoE (à 1, 2).

- The idea and objectives of a TDW were discussed with representatives of AI4EU, who are coordinating the pilot experiments for the platform[27] and who expressed their interest to cooperate in future iterations of TDW (à 2, 4).
- The presented concept for the TDWs was supported and much appreciated by all discussion partners. Thus, an agreement between the respective coordinators and leadership teams of VISION, TAILOR, HumanE-AI-Net, and CLAIRE was made to plan and execute a series of Joint (co-organised) Theme Development Workshops, starting in spring 2021 (à 1). Accordingly, a corresponding announcement to the NoE and beyond (AI4EU, CLAIRE, etc. à 4) was published and distributed in mid-December.
- Reacting to this broad announcement, the coordinator of AI4Media[28] reached out to discuss the TDW concept. At the end of this call, AI4Media expressed interest to organise a joint TDW, specifically with a focus on the media sector (à 1, 2).

We would like to thank all partners involved in this process, and especially those people who supported us by sharing their ideas and valuable input.

Organisation of a Theme Development Workshop

In the following subsections, the template for planning, organising and executing a TDW will be shown and explained in detail. This includes a checklist and some recommendations for the specific steps and activities required. The following is taken in part from the VISION Deliverable "Template for Theme Development Workshops" which is publicly available on the VISION website[29].

TDW template and process

Based on previous experience and focused discussions with some representatives and partners from the NoE, a template for a five-phase process to plan and execute a TDW was developed. This overall process is visualised in Figure 1 and will be further elaborated on in the following subsections, including some recommendations and sharing of best practices.

[26] For a description of this NoE, see (access only for registered users of AI4EU-platform): https://www.ai4eu.eu/humane-ai-net-0.
[27] For AI4EU pilot experiments, see: https://www.ai4eu.eu/pilot-experiments.
[28] See project website: https://ai4media.eu/.
[29] See project website: https://www.vision4ai.eu/d41-tdw/

Figure 3: Process overview – steps to organise a Theme Development Workshop

Phase 1: Start planning the Theme Development Workshop

Phase 1 of the overall process consists of three main activities. Based on our previous experience, we recommend starting early with the planning, at least 3 months before the planned workshop date. However, this timeline depends on the scope and size of the TDW as well as some other circumstances (e.g., is there a longer holiday period in-between, are there any special time constraints given the planned location of the TDW, do people have to arrange their travel in advance, or is an online workshop planned).

Activity 1: Define the topic and scope/purpose of the TDW

The first step is to define the topic and scope of the TDW from a high-level perspective. The guiding strategic rationale behind this are the objective/s of the NoE as defined in their project plans, especially in relation to their planned roadmapping activities. If there is already a first roadmap available, we recommend to use it to plan and derive the TDW/s and their topics accordingly. If this is not the case, the development of the roadmap and the TDW planning should be intertwined as much as possible to align those activities, and thereby maximising the benefit for the network.
In general, there are two different perspectives to define the high-level topic of a TDW and provide input to the strategic AI research and innovation roadmap: (1) From an industry perspective: How will and can AI be used in a particular industry sector or to address a relevant societal topic. And (2) from a research perspective: Given a specific AI area or research field, how will and can this be used in different application areas and/or industry sectors. Accordingly, some examples for high-level topics of a TDW could be:

1. Trustworthy AI for autonomous driving (focused on an industry sector)
2. AI approaches and technologies to prevent and expose fakes news (focused on a societal topic)

3. Application scenarios of AutoML in various industry sectors (focused on a research topic)

Apart from the purpose to provide input to the strategic AI research and innovation roadmap, the requirements analysis also revealed some additional objectives of a TDW. Besides the longer-term strategic perspective and its corresponding objectives (roadmapping), it seems important to the NoEs to use the TDWs also to identify and define some mid-to-short term activities, which then can be further addressed by some partners within the networks collaboratively. Due to their short-term results they are particularly interesting to industry. But they should be avoided as the main focus of a TDW, as they can easily detract from the long-term focus and may come up anyway in the discussions, at which point they can be explicitly documented as a result. In particular, ideas for use-cases and challenges/benchmarks (TAILOR) and microprojects (HumanE-AI-NET) were mentioned in this context and should therefore be taken into account for defining the topic and scope of a TDW.

Activity 2: Define the target/participant groups for the TDW

The next step after the high-level definition of the topic and purpose is to decide which groups of stakeholders should participate in the TDW and why. This step is not about selecting the individual participants of the workshop (see Phase 3), but more to gain an understanding which backgrounds and expertise are needed to contribute towards achieving the objectives of the TDW. In the context of the NoE and based on the considerations outlined in Section 2, the following groups are considered as important stakeholders in general:

- **AI researchers from academia** working on specific AI topics and/or application areas.
  à Contribution: They should bring their specific expertise from academia as well as ground-breaking new ideas about their areas of expertise to the TDW but they need to be open to engage with the industry perspective.
- **AI researchers from industry** working on specific AI topics and/or application areas
  à Contribution: they should bring their specific research expertise and challenges from an industry perspective to the TDW but they need to bring a strategic, long-term perspective and be open to engage with the research perspective.
- **Strategic thinkers from industry/business units**
  à Contribution: they should bring requirements from their business units and specific market challenges to the TDW.
- **Strategic thinkers from non-profit/associations and/or politics**
  à Contribution: They should bring requirements and challenges from a broader/societal perspective in relation to the defined topic/s to the TDW.

Activity 3: Establish an Organising Committee for the TDW

In order to ensure a proper planning and organisation of the TDW, including its content and focus (long-term strategic and mid-to-short term activities), we recommend establishing a so-called **Organising Committee** (OC). Given our previous experience and also the somewhat complex nature of the TDWs as an innovative instrument, it makes sense to bring

together a group of people with different backgrounds and expertise. Depending on the scope and objectives of the TDW, we suggest an OC consisting of 5-8 persons[30], collaboratively agreeing to pursue the following tasks:

- Refine the high-level topic of the TDW into specific sub-topics/topic-lines that are interesting to the target groups of the workshop.
- Define the schedule for the workshop day(s), including presentations/speakers and breakout sessions according to sub-topics/topic-lines[30].
- Suggest experts and select participants for the defined sub-topics/topic-lines of the TDW and with all the needed background and perspectives.
- Attend the TDW as experts and/or facilitate a specific topic-line/breakout session
- Contribute to/proof-read and amend the final TDW report.

Accordingly, we advise to select the members of the OC purposefully and take the following prerequisites into consideration:

- Members should be experts in the topic area of the TDW.
- Members should have an extensive network of/know other experts related to the topic/s of the TDW.
- Mix members with academic and industrial background for the OC.
- OC should include at least one person from/in close contact with on-site/online workshop (technical) organisation team.
- OC members agree to meet on a regular basis/several times before the TDW takes place.

Phase 2: Promote the TDW to the target audience

As soon as the high-level definition of the topic, purpose and target groups is finished, we recommend starting activities to promote the TDW to the target audience. This is very important to select and invite the most appropriate candidates for the workshop (see Phase 3), which is a key success factor for a valuable outcome of the TDW. We recommend fixing the date and location of the TDW before promoting it to the target groups, so that the experts can check their availability before their registration. Also a first outline of the schedule for the workshop day/s might be helpful in this context, and therefore should be included in the information material if possible.

Activity 4: Create information/marketing material for the TDW

Based on our previous experience, we advise to create an appealing information package to promote and announce the TDW, and distribute it to the identified target group/s. This package should include:

- Some general information about the TDW, and its purpose and scope,
- The main topic/s of the TDW,
- Some information about the target group/s, and why people should participate,
- Some information why it is beneficial to be a participant, and

[30] Depends on their willingness to share the results with the VISION CSA. Based on our experiences with the NoE so far (see previous section), we expect a very good exchange of ideas and results though.

- how to participate/register as well as the planned date, time, and location of the TDW.

In December 2020, we released a broader announcement for a series of planned joint Theme Development Workshops, co-organised by TAILOR, HumanE-AI-Net, VISION, and CLAIRE, and distributed it within the four NoE and other EU projects/ initiatives (AI4EU, etc.). This announcement (see Annex 5a) can be used as an example how to promote a TDW and can be adapted by the NoE according to their specific objectives and requirements. As an optional step, we also recommend creating a survey to support the identification of appropriate participants for the TDW.

Activity 5: Create a survey to identify suitable participants for the TDW (optional)

This step is indicated as optional because the OC can also decide to create a list of participants based on their own networks, including (associated) partners from the NoEs. However, we recommend at least complementing this list with candidates from an open application process to guarantee a certain level of fairness and broader inclusion of participants. We advise to include all information into the survey that the OC needs to select the most appropriate participants for the TDW, specifically:

- Some personal data
- Affiliation and background of the applicant
- Additional data supporting the selection process and/or organisation of the TDW, e.g.,
    - areas and level of expertise in AI,
    - possible contributions to the TDW, and
    - suggestion of further (sub-)topics to be included in the schedule.

Furthermore, the survey should not be too time-consuming and easy to submit, because otherwise applicants might not finish the questionnaire.

Phase 3: Start organising the Theme Development Workshop

If this has not happened before promoting the TDW to the target audience, Phase 3 should start with a definition of the date and location of the workshop. For on-site/physical meetings, we also recommend defining and reaching out to the local organisers as early as possible in order to clarify organisational issues (needed rooms and equipment, catering, etc.). Especially during the Covid-19 pandemic, a TDW might also be organised as an online or hybrid event. In this case, we recommend to clarify who is in charge of the technical support for the TDW (procurement of software licences, technical support staff required during the workshop etc.) via the OC as early as possible.

Activity 6: Define the schedule for the TDW

In order to support the selection of participants and speakers, a schedule for the workshop day/s should be defined by the OC. Based on our previous experience, a well-prepared one-day workshop is able to produce an initial research and innovation roadmap, including some first ideas for next steps and estimations on the time frame. However some partners from the NoE also expressed interest to extend the TDW to a second day in order to allow for deeper discussions about specific (sub-)topics. This seems to be reasonable depending on the objectives, purpose, and scope of the workshop.

An exemplary schedule for a one-day, on-site workshop is outlined in the following table, addressing the requirements mentioned by the NoE to discuss roadmapping (long-term strategic perspective) as well as use-cases and challenges (mid-to-short-term activities). Furthermore, the emergence of new topics during the TDW is taken into consideration to support the creative nature of the workshop and address the special interests of participants.

Topic/Purpose	Timeline	Description/activities
Arrival and Intro	8:00	Arrival and welcoming (registration, first socialising,etc.)
	9:00	Official welcome by organisers and information about workshop and procedures (first outlook about the day)
Session 1- long-term strategic perspective	9:30	Short invited talks (~15 -20 minutes) on (sub-)topics relevant to the workshop (from academia and/or industry) à inspiring & challenging as preparation for breakout sessions
	10:30	Coffee break and preparation for roundtable work (breakout sessions)
	10:45	Round table group work on specific (sub-) topics
	12:00	Plenary discussion/presentation of key findings from round table work
Lunch and Socialising	13:00	Lunch break and informal discussions
Session 2: Mid-to-short-term activities	14:00	Plenary selection of emerging topics (from morning session) for further development
	14:15	Short invited talks (~ 5-8 minutes max.) on to introduce use cases and challenges
	14:45	Round table group work on specific use cases and challenges, as well as selected topics
	15:45	Coffee break
	16:00	Plenary discussion/presentation of key findings from round table work and next steps
Closing and Socialising	17:00	Closing and informal discussions

Table 1: Exemplary schedule for one-day on-site workshop

Activity 7: Select and invite the participants and speakers

This step is very important because the success of the TDW depends to a large degree on the participants and their contributions. As explained in Section 2, their expertise and visionary mindset is key, as is an interdisciplinary staffing of the smaller groups discussing in the breakout sessions. These considerations should guide the selection of participants by the OC accordingly. We recommend to develop a fair and transparent set of criteria for the selection process, keeping the following aspects in mind:

- Aim for a balanced distribution among the main target groups, and also include a variety of professional backgrounds (researchers, experts from business units, engineers, and developers, etc.),
- Aim for geographical balance, include participants from all over Europe,
- Aim for gender balance,
- Include some early-stage researchers/young talents and invite additional stakeholders to guarantee a broader perspective,
- Try to include as many organisations (academic and private sector) as possible, and
- Include participants from big companies, SMEs and startups.

We advise to create a participant list, including a waiting list in case some of the pre-selected experts are not available for the TDW. The candidates for this list can be generated via different sources, e.g., the network of the OC members, the partners in the NoE, and/or from a broader survey as described in Activity 5. As soon as a prioritised list is ready, the OC should organise the invitation of the selected participants and get their confirmation to further continue the planning and organisation of the TDW.
In parallel, the OC should discuss and define the presentations and speakers according to the defined TDW schedule. The presentations should be brief and include some real challenges in order to inspire and support the discussions in the following breakout sessions. Speakers can be either participants of the TDW or invited guests. Especially for on-site TDWs with longer travel distances for the experts, we advise to provide some interesting networking and collaboration opportunities to those speakers who are not participating in the workshop, or to focus on speakers who live close to the location of the TDW (e.g., a dinner on the evening before the event to already welcome the participants and get them to socialise).
For on-site TDWs and hybrid events, participants, and speakers might ask for support regarding their travel plans and accommodation. Therefore, we recommend to prepare an information package together with the local organisers, including some information on how to get to the location and about nearby hotels with special rates, and provide this to the speakers and participants as early as possible.

Activity 8: Prepare the event including the breakout sessions

Before the start of the workshop day/s, it makes sense to check once more if all needed infrastructure is available and ready, and if the staff is familiar with their supporting and facilitating roles. We recommend to check in particular:

- Is the location and local organisation team ready (on-site/hybrid TDW):
    - Are the needed rooms ready, including required equipment for presentation and breakout sessions/group work (separate rooms or roundtables in a bigger room)?

    - Is there a stable WIFI connection for all participants?
    - Is catering available and coordinated with the workshop schedule?
    - Are participants informed about how to get to the location?
    - Who can support in case there are questions/issues emerging during the workshop day/s, etc.?
- Are the virtual meeting solutions ready and tested, is the technical support team ready (hybrid-/online TDW)?
- Is the registration process organised (on-site/hybrid TDW):
    - Is there a list of confirmed speakers and participants for the registration team?
    - Are name tags and WIFI access info ready for hand-out?
- Are the plenary and breakout sessions organised:
    - Are speakers informed about the schedule and modalities for their presentations?
    - Are the documents for the collaborative work and note-taking during the breakout sessions prepared?
    - How is the distribution of participants to breakout sessions organised (pre-defined groups or random distribution, etc.)? Are there technical solutions to support this for online TDWs/hybrid events? (see also Activity 10)
    - Optional: Are facilitators assigned and briefed for the breakout sessions? Are any recordings planned for the plenary/result presentations, and how is this organised?

Phase 4: Execute the TDW – the day(s) of the workshop

During the workshop day(s), it is important to ensure a smooth pass through the event. This includes welcoming and supporting the participants and speakers, and being available to deal with emerging issues during the day. Some of the organisers might also be involved as speakers and/or participants of the TDW, so we recommend to staff the supporting teams accordingly to avoid distractions due to organisational issues.

Activity 9: Registration and support

As already described in Activity 8, the workshop day starts with welcoming the participants, including a registration process (on-site/hybrid TDW). In case of an online workshop, the organisers might use technical options (password protection, etc.) to make sure that only registered participants get access to the TDW. Depending on the number of participants and the peculiarities of the location (e.g., longer distance from registration to plenary room), we recommend to open registration at least 45 minutes before the official start of the workshop to avoid situations where people are rushing into the first session at the last minute. Furthermore, people can use the time before the official start of the TDW to talk to other participants informally and familiarise themselves with the location and programme. A welcome package for each participant (e.g., name tag, info about location and breakout rooms) may support this process.
We also advise planning for multiple participants arriving at the same time, so staffing for the registration process should be able to cope with such situations. Otherwise there may be

unpleasant long queues and waiting times, especially if there are further questions during the registration process. Last but not least, if you plan to take photos or record videos/streams during the TDW, participants need to be informed upfront and indicate their consent.

Activity 10: Plenary and breakout sessions

We recommend to start the workshop officially in plenary with a welcoming session, including the most relevant information about the objectives and procedures planned for the workshop (see Table 1). It is important to explain to the participants what is planned in the plenary and breakout sessions, what the expectations are, what tools they can use for support, which kind of results and outcomes are expected, etc. Especially before the start of the first breakout sessions, we advise to offer support in case there are any questions regarding the distribution and/or relocation to breakout rooms/tables, and schedule this transfer during a short break – otherwise breakout sessions might start late because participants are looking for their working groups (on-site/hybrid events). In case of an online workshop, the distribution of participants to the virtual breakout rooms needs to be organised via the technical solution. Furthermore there might be an opportunity to use a virtual conferencing tool for the TDW, which supports more activities similar to a physical event[31].

Regarding the distribution of participants to breakout sessions, we recommend to assemble interdisciplinary groups (e.g., 2 AI researchers, 2 experts from business units, 2 engineers/developers in one breakout session). This allows to include different perspectives in the discussion of a specific topic and supports creative ideas with regard to possible solution approaches. Accordingly, a corresponding distribution mechanism or even a list of predefined participants per breakout session is useful, based on the interest, expertise, and background of the participants. If a more random or agile[33] distribution of participants is desired, we recommend to ensure at least that participants from academia and industry mingle as much as possible in each of the breakout sessions. However, this also depends on the discussed topic and desired outcome of such a session. It is helpful to have someone in each group who takes notes, so that participants can focus on the discussion and are not distracted. This could be one of the participants or a designated person from the organisers. Sometimes, joint note taking using a shared document can also work quite well in either way as each participant can add to the notes or adapt them as needed.

It is very important for the outcome of a TDW (see Activity 12) that participants in the breakout session document their discussion results and insights. Thus, we recommend using a proven tool to support collaborative note-taking in these sessions, and to provide at least a rough structure and some guidelines for such a document. An example for such a structure can be found in the following table.

[31] One example is https://gather.town/ – a tool with a computer game-like user interface and spaces for presentations/plenary sessions, poster presentations, social rooms, meeting areas, etc., including options for private chats and conversations. We advise to develop a list of criteria to fulfil based on the scope and objectives of the TDW, and then check conditions and alternative solutions before selecting the technical environment and tools to support the workshop.

Headline/section of the document	Description/ à purpose
Headline: Topic Optional: Subtopic	State the topic to be discussed in this breakout session à Indicate as headline of the document, so that everybody knows directly what the breakout session is about
	State the subtopic to be discussed in this session à we recommend to formulate a question that is as concrete as possible
Names and affiliations of the participants	Participants can and should indicate their names and affiliations. à Supports the preparation of final report and follow-up activities à Give people credit
Objectives and/or guidelines for the breakout session (see also next section)	Describe the objective of the breakout session, and provide some guidelines if necessary à Participants should know what topics to discuss, including the expected outcome of their collaborative work in this session à Precise and as concrete as possible, we recommend to use bullet list or similar to not overwhelm participants with too much information à If the breakout session is moderated by someone, this description can be very brief: Moderator to introduce the objectives and thereby motivate the discussion
Optional: Guiding questions	We recommend to use 3-5 guiding questions/subheadings to structure the discussion à These can be specific subtopics to be discussed for example à Recommendation: Depending on the topic, ask for measures for achievement, next steps, and estimated timeline, challenges, and barriers of achievement, etc. à If the breakout session is moderated by someone, discussion can be guided with this predefined structure/questions instead

Table 2: Exemplary structure to support collaborative note-taking in breakout sessions

Phase 5: Summarise and distribute results of the TDW

The five-phase process to organise and execute a TDW closes with a review and summary of the main results and outcomes of the workshop day/s. As outlined in the following activities, these steps support the NoE in their project-specific objectives, and contribute to a further improvement of TDWs as an innovative format.

Activity 11: Review of the TDW results

After the TDW, the results need to be reviewed and edited for further use. Basically, these are recorded in the documents collaboratively produced during the breakout sessions, and can be complemented by further sources, e.g.,:

- Facilitators who took part in the breakout sessions and contribute their perspective and assessment,
- Notes taken during the plenary/result presentations by support staff,
- Recordings from the TDW if available (of presentations, etc.[34]),
- Selected experts, including the Organising Committee, who can contribute their experiences from the TDW, as well as their expertise to the results.

Based on this reviewing and editing process, we recommend creating a report summarising the main results of the TDW (see Activity 12), as well as some further documents supporting follow-up activities in the NoE (see Activity 13).

Activity 12: Create a report and distribute it

A report summarising the main results of the TDW might be interesting to all participants and further recipients. Especially to initiate some activities in the NoE based on the outcome of such a workshop, we recommend creating a report immediately after the event and distributing it for feedback to the participants. This helps in incorporating different perspectives and adds input that might otherwise would not be included.
If a summary of the results is needed for recipients outside the participant list of the TDW (e.g. for dissemination purposes), we recommend checking the report carefully for any participant-specific content and/or contents that should not be distributed to a broader audience (for confidentiality reasons, etc.). If such content is part of the report, there should be either a consent of the participants to the publication or the report should be revised accordingly by deleting/anonymising/generalising sensitive results. For transparency reasons, we advise to inform participants before the workshop if a dissemination/publication of workshop results is planned and to indicate what information might be critical to them. With participants from competing companies it might be necessary to remind them not to discuss any topics that might violate rules and laws regarding anti-competitive behaviour.

Activity 13: Take care of follow-up activities (optional)

Depending on the purpose and objectives of the TDW, there might be further activities to follow-up with after the event. In the context of the NoE, the TDWs are planned to contribute to the roadmapping activities as well as to use-cases and challenges to be addressed in the networks for example (see Activity 1). Accordingly, corresponding processes and collaborations are needed to ensure this, e.g., by incorporating the TDWs results and/or involve participants from the TDW into subsequent roadmapping workshops, supporting smaller groups of participants to follow-up with specific use-cases, etc.

We also recommend asking the participants for their general feedback about the TDW after the event. This can be done ideally immediately via an online survey for example but could also be sent to all participants together with the workshop report. Such a feedback process is very valuable to further improve the TDW concept in the context of a NoEs (and ICT-48 in general) with lessons learned, thereby supporting the development of this innovative format.

TDW checklist

The following checklist summarises the main activities to plan and execute a TDW. Accordingly, it is intended to support and guide the NoE in organising their own workshops.

Checklist: Main activities to plan, organise and execute a TDW	
☑	Define the topic and scope of the TDW ➢ What topic area/challenges should be addressed, and from which perspective (AI topic/area or specific industrial or societal perspective)? ➢ What is the purpose and objective of the workshop?
☑	Define the target group of the TDW ➢ Who (groups of stakeholders) should participate in the TDW and why? ➢ What can and should these groups contribute to the TDW?
☑	Establish an Organising Committee ➢ Who is in charge to plan and organise the TDW? ➢ Who can support regarding content and selection of participants and speakers?
☑	Define Date and location for the TDW ➢ What is an appropriate date and location for the TDW? ➢ Who is the local organiser/contact person, and how will they support the TDW?
☑	Promote the TDW to the target audience ➢ Is the marketing/information material ready? ➢ Who should get the information, and how is the delivery managed? ➢ Optional: Do you need a survey to generate a pool for candidates?
☑	Select participants and invite speakers to the TDW ➢ What should a schedule for the workshop day/s look like? ➢ Which candidates should be selected to discuss the defined topics? ➢ Which presentations and speakers can contribute and motivate the discussion?
☑	Prepare the TDW, including the breakout sessions ➢ Is the local team ready, are rooms booked (on-site/hybrid event), is the technical support team ready (online/hybrid event)? ➢ Is registration and catering organised (on-site)? ➢ Are all handouts and documents for the TDW and breakout sessions prepared?
☑	Execute the TDW (workshop day/s)

	➢ How do you manage registration and support to participants (on-site)? ➢ How do you execute and support plenary presentations and breakout sessions?
☑	Review results and create report about TDW ➢ Who is reviewing the results from the TDW, especially from breakout sessions? ➢ Who is generating the summary/final report, who can support?
☑	Follow-up activities and feedback ➢ Who should be provided with the final report, and who is in charge of this? ➢ Optional: Do you plan to ask participants for their feedback about the TDW, how? ➢ Who is taking care of follow-up activities resulting from the TDW, and how?

Table 3: Main activities to plan, organise and execute a TDW

Lessons learned from organising a Theme Development Workshop

Based on the previous experience of organising and planning the first sector-specific Theme Development Workshops, we were able to identify some valuable specific and general lessons learned. In the following, these will be explained in more detail and, in addition, included in the ongoing planning of future workshops and will serve for further adaptation of the template for the organisation of Theme Development Workshops.

- **Early Planning and Coordination:** Initiating the planning process well in advance, preferably around six months before the workshop, proved essential for overcoming challenges related to holiday periods and other scheduling conflicts. Moreover, involving the Organizing Committee right from the outset, along with aligning their availability, ensured a smoother planning process.
- **Monthly Meetings and Communication Strategies:** Holding monthly meetings with the Organizing Committee proved to be an effective balance between coordination and minimising time consumption. This approach acknowledged that a majority of Committee members were unable to attend regular meetings. However, to ensure ongoing communication, email exchanges were utilised between meetings, enhancing efficiency.
- **Mitigating Keynote Speaker Challenges:** Recognizing the potential for keynote speaker cancellations or technical issues, it was prudent to have backup arrangements in place. This proactive approach safeguarded against last-minute disruptions and upheld the quality of the workshop's content.
- **Dynamic Agenda:** Flexibility in the agenda is important to accommodate unforeseen discussion needs or participant interests. Dynamically adjusting the agenda can increase the workshop's relevance.

- **Participant Commitment and Engagement:** Despite efforts to secure confirmed participation from attendees for the entire workshop day, some individuals either failed to attend without prior notice or participated only partially. Implementing mechanisms to enforce full-day commitment could improve the overall workshop experience.
- **Diversity and Gender Balance:** Ensuring a diverse range of participants, both in terms of industry and academia backgrounds as well as gender balance, was identified as a critical factor for the workshop's success. Striving for a balanced representation allowed for richer discussions and perspectives during the event.
- **Sustainability and Follow-up Activities:** The workshop's success was not limited to its immediate outcomes. The collaboration's sustainability depended on follow-up activities, such as joint research initiatives, challenges, and hackathons, which require dedicated efforts to materialise the ideas generated during the workshop.
- **Involvement of Key Opinion Leaders:** Engaging renowned experts and key opinion leaders in the AI industry can enhance the workshop's interest and credibility.
- **Public Communication:** Targeted promotion of the workshop within the professional community and relevant industry circles can increase participant numbers and elevate the event's visibility. This requires
- **Collecting Feedback:** Participants' opinions about the workshop are valuable. Gathering feedback both during the workshop and in the post-event phase can help improve future events. However, this proved to be difficult, as only very few participants filled out the survey that was created for this purpose, despite multiple requests during the workshop.
- **Report generation**
  As a result of the follow-up of the workshop, the Organising Committee as well as the experts and moderators of the breakout sessions have to be involved in the preparation of the report. In the past, however, it has been shown that these are not always available due to scheduling restrictions, so it is important to send multiple reminders and therefore to plan enough time in advance for the preparation of the report.

In conclusion, the Theme Development Workshops demonstrated the significance of meticulous planning, diverse participation, effective communication, and continuous engagement for fruitful collaboration between industry and academia in the field of artificial intelligence.
Incorporating these valuable lessons learned into future Theme Development Workshops will undoubtedly enhance the effectiveness and efficiency of organising and conducting those workshops that unite researchers and industry representatives in the realm of AI research. By leveraging these insights, workshop organisers can navigate challenges with greater foresight, foster deeper engagement among participants, and ultimately contribute to the advancement of AI research and its practical applications.

Results of the Theme Development Workshops

As already outlined, the projects HumanE AI Net, VISION and CLAIRE AISBL are organising a series of joint so-called Theme Development Workshops (TDWs) under the lead of TAILOR to bring together key players from specific industry sectors with key AI researchers, as well as additional stakeholders. The objectives are to jointly identify the strategic AI research areas and challenges in the scope of the workshop, and to initiate further activities to address them beyond the event (e.g., joint working groups, research papers, challenges, hackathons, transfer labs).

Figure 4: TAILOR Industry sectors and partners/representatives

Concerning the concept of the Theme Development Workshops, we mainly focused on the TAILOR sectors shown in Figure 4 in cooperation with the corresponding TAILOR industry partners and representatives, which are mainly involved in the activities of TAILOR WP8 and beyond. Through their valuable participation and insights, they have contributed significantly to the results of the subsequent Theme Development Workshops, which are listed in detail below.

Theme Development Workshop “AI in the Public Sector”

About the workshop

The first Joint Theme Development Workshop (TDW) co-organised by CLAIRE, TAILOR and VISION[32] on “AI in the Public Sector” took place on the 7th and 9th of September 2021. The full report is publicly available on the TAILOR website[33].

Facts and figures

At this two half-day workshop, experts from public and governmental institutions, industry and academia jointly developed initial input for the European AI research and innovation roadmap. The high international interest that was expressed in response to the announcement of the AI in the Public Sector TDW, translated into excellent attendance of the event: In total, 56 participants from twelve countries joined the two workshop days, with around ⅓ of participants being female. The workshop also caught the attention of high-level political institutions, thus representatives of the European Commission, the United Nations Interregional Crime and Justice Research Institute (UNICRI), the UK Office for National Statistics, the State Chancellery of Saarbrücken, the Den Haag Ministry of Interior & Kingdom Relations, City of Amsterdam, and City of Frankfurt/Main joined the workshop.

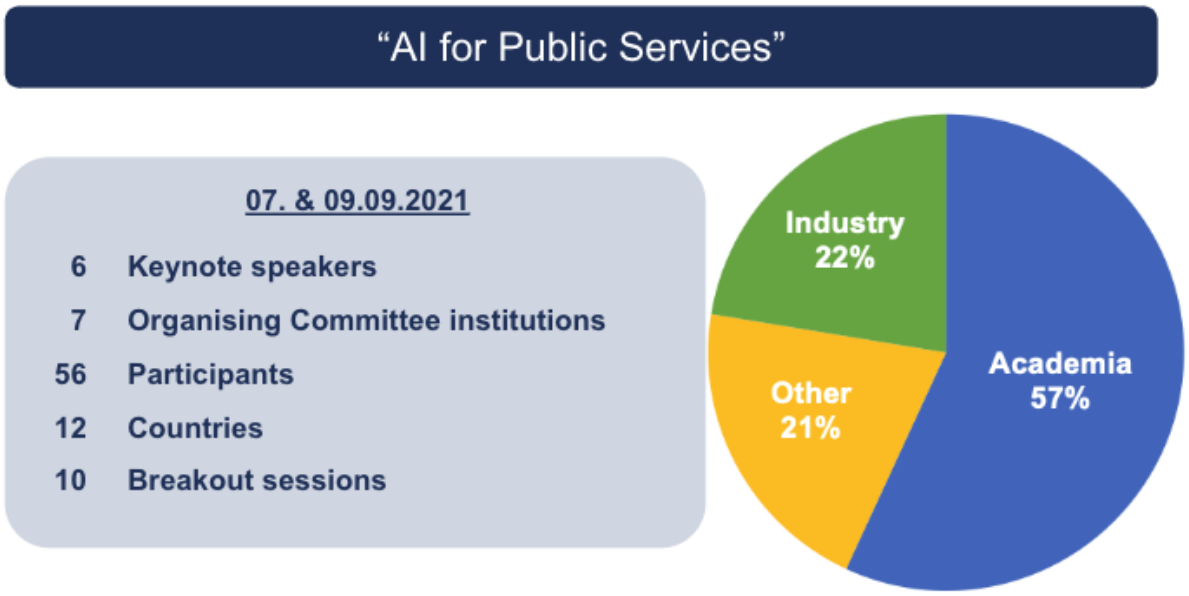

Figure 5: Facts and figures of the TDW “AI for Public Services”

Program

The TDW “AI in the Public Sector” was opened by the Co-Chairs Silke Balzert-Walter (DFKI) and Freek Bomhof (TNO) on behalf of the Organising Committee (OC), which included further representatives from CBS, CLAIRE, Engineering Ingegneria Informatica, FBK and

[32] In alphabetical order
[33] See project website:
https://tailor-network.eu/events/theme-development-workshop-ai-in-the-public-sector/

Intellera Consulting. They outlined the objectives of the TDW as well as the agenda and programme for the two days, and introduced the invited keynote speakers to the participants. The inspiring keynotes on both workshop days were provided by high-level experts from several European countries. Their introductory presentations served as a basis for the discussions about the opportunities and risks of AI usage in the public sector, and provided some interesting examples of application areas. Accordingly, these presentations stimulated the expert discussions in the following breakout sessions.

Figure 6: Flyer of the TDW "AI for Public Services"

The first day of the TDW was focused around application areas for AI in the public sector. The topics for the breakout sessions were prepared by the Organising Committee and introduced by selected experts, and the community had the opportunity to suggest further topics via an online form before the workshop.

Daniel Sarasa Funes in his role as the Director of Fundación Zaragoza Ciudad del Conocimiento (Spain) stressed the importance of so-called urban labs, which can serve as an important connection between different stakeholders and lower the barrier of technology adoption by citizens. In his keynote he also mentioned the example of 'Toronto tomorrow', a project by Google/Sidewalk Labs, and explained why the project has failed, which he attributed to various factors such as the resistance of public servants. He concluded his presentation by stressing the importance of accountability for AI in the public sector, because municipalities are the institutions closest to Citizens.

Rasmus Hauch and Christine Kvist, CTO and Public Sector Lead of 2021.AI (Denmark), gave an interesting keynote explaining the role of startups and scaleups in supporting the public sector. They summarised their focus on responsible leadership, empowering transparency while respecting privacy and compliance by design, outlining that these topics are very important in dealing with AI. They also emphasised further aspects like AI governance and AI frameworks in assuring compliance with external regulations, internal codes-of-conduct and other best practices and also mentioned some interesting activities and initiatives in Denmark, and provided an overview of potential use cases in the public sector like chatbot assistance, optimised traffic routing or AI assisted tax calculations.

The second day of the TDW changed the perspective towards horizontal topics of AI spanning many application areas. In particular, the discussions were focussed on different aspects of Trustworthy AI, including some topics for breakout sessions which emerged during the first workshop day.

Ammar Alkassar in his role as the State Commissioner for strategy as well as CDO & CIO at the Saarland State Government (Germany) delivered a welcoming message to the participants of the second workshop day and emphasised the importance of AI for governmental institutions. In particular, he referenced the ongoing digitisation of public services, and pointed out the necessity for AI to be trustworthy in order to generate acceptance of the technology among public administrations and citizens.

Andrea Renda, Senior Research Fellow and Head of Governance, Innovation & Digital at the Centre for European Policy Studies (CEPS, Belgium) introduced in his keynote “Trustworthy AI: an imperative for government” the work of the Independent High-Level Expert Group on Artificial Intelligence (HLEG), that has designed the Ethics Guidelines for Trustworthy AI. In his keynote, Andrea Renda also addressed the new AI Regulation with a critical look on its effects, for example with respect to broader societal risks emerging when humans lose control of the process. Furthermore, digitalisation demands governments to be active in a number of roles such as a regulator, auditor, service provider, and trustee. Last but not least, Andrea Renda outlined four urgent actions that are needed to trigger government uptake in the future: matching AI use with administrative and constitutional law, developing sandboxes on key use cases, guidance on AI procurement, and providing ad hoc guidance on specific AI use cases.

Marieke van Putten, Senior Innovation Manager at the Ministry of the Interior & Kingdom Relations Den Haag (The Netherlands), outlined in her introductory presentation the evolution of more complex AI solutions in the public sector, with a variety of different inputs, levels of complexity and types of applications. She also gave some insights into the policy and the work on Trustworthy AI in the Ministry of the Interior & Kingdom Relations in The Netherlands, which includes research on what trusted AI really means, its possible impacts and how to develop instruments to ensure and enhance it. Marieke van Putten concluded with the statement, that more public organisations than ever are now active in the area of AI. Accordingly, there are various application areas for AI in this area like tailored solutions for citizens, process optimisation, maintenance in cities, inspection and enforcement, crime investigations as well as forecasting and policy development.

Key results from the breakout sessions

The following subsections will provide a detailed explanation of the results of the breakout sessions of the Theme Development Workshop “AI for Public Services”.

AI for public safety and security

The breakout session started with a brief presentation on how AI is used in the area of law enforcement on an international level, including some use case examples. The guiding questions which were discussed during this breakout session can be summarised to: (1) “How can we protect AI tools from being used for the wrong purposes?” and (2) “How can future AI support decision making while protecting and respecting privacy?”

The discussion showed that to win the race against organised crime, it is highly recommended that the public sector should possess the knowledge, tools, and budget to fight against criminals using AI. It is of utmost importance that the public sector is aware of this, and that corresponding tools are developed. Especially regarding law enforcement, the topic is highly complex and requires understanding of a fundamental tension, namely

- Combining data across different public organisations and (potentially) private actors, and the building of AI-enabled tools, is increasingly necessary to combat new forms of (cyber)crime that rely on more advanced technologies and techniques.
- At the same time, this sharing and AI system development creates incredibly powerful tools that, if used without care/principle, for the wrong reasons or problems or in the hands of the wrong actors, may lead to serious forms of harm.
- As a result, there is both a serious political problem, as well as a need for insight and knowledge about how both of these issues can be addressed integrally.
- Building trust in AI systems is the key challenge. It must exist on both the public and private side in order to encourage its use. Here, an independent and objective institution is needed to guarantee the safety of AI systems. But should it be at local, regional or national European level? Situations are different but it is important to understand the local level before extending. Also the guidance in the field of AI on both sides is essential. The importance and opportunities of the technology must be demonstrated and explained so that both sides know exactly what they are dealing with. An exchange between all parties involved and their use of AI is also of great importance.

During the discussion it was noted that currently dominant topics (fairness, accountability, transparency, explainability) are all *contextual* and need to be understood, studied, developed and implemented across the above topics. As such the experts in this breakout session propose a shift in paradigm from *technology*-centred to *system and human*-centred, in which *human* refers to the protection of human rights and prevention of citizen harms by AI systems (both in terms of AI systems aiding crime prevention and the prevention of AI systems’ harmful uses), and *system* refers to the technology situated in complex sociotechnical practices, organisations and institutions. From a research perspective, the challenges and topics discussed in this breakout session could motivate the further study and development of the following topics: Development of ecosystems of trust for enabling and safeguarding AI systems in law enforcement; sociotechnical specification of AI systems

in law enforcement; institutional design for AI systems in law enforcement; meaningful human control of AI systems in law enforcement; boundary conditions for public-private collaboration in AI systems for law enforcement.

AI for urban mobility

In this session focused on AI for urban mobility, seven participants discussed the key future innovative AI technologies for the planning and management of urban mobility for citizens, civil servants and decision-makers. The discussion covered various concepts and topics, from the availability of data to the transparency and trust of solutions to be adopted by the public sector.

One of the key results was to focus not only on urban but also on rural mobility and interfaces between urban and rural areas. Also, the question how the public sector can build trust on AI solutions provided by private companies in the public sector played an important role during the session. This aspect, which is closely related to transparency, Intellectual Property Rights and monetisation addressed the question of the openness and the availability of data and the possibility and ability to define parameters that measure the level of trust in AI solutions. Further points of discussion were the replicability of AI solutions linked to the availability of data and representation as well as the interaction between developers and civil servants and on how to integrate AI solutions that impact the kind of solution that can be implemented in the public sector. During the session, several ideas for use cases, hackathons and other challenges were discussed, for example in regards to an AI applied to solving the surface parking problem at the centres of our cities or on how to make best use of satellite images in building AI solutions for mobility. One of the long-term project ideas was on how to combine car sharing/car pooling/Automatic Driving and optimisation algorithms.

AI for reliable statistics

This breakout session looked at the topic of AI for reliable statistics. The group consisted of representatives from the private sector, academia and official statistics institutes. One of the three identified key elements on this topic was on the community of official statistics. A non-competitive environment allows for excellent opportunities and great potential for collaboration at the European level. The group also looked actively into the use of AI for new data sources (satellite data, websites, sensors, etc.) that require new technologies since they produce an enormous amount of data sets that are too large to be processed manually. The use of reliable statistics also requires a guarantee of quality for the used data, which raises the question on how to define certain quality aspects that need to be taken into account for the reliability of statistics. The Covid crisis, for example, has shown that aspects like timeliness can be of particular importance. Therefore, relevant statistics, relevant policy and decision making should be used rather than historical data. Another key element addressed during the session was the role to produce trustworthy and transparent statistics. The questions that came up in this regard was if users should be educated and to what extent AI-based statistics can be explained with the definition of the target group in question. The participants of the breakout session also came up with ideas for Hackathons, i.e. on

explainability or to concretely work with data sources like satellite data and to use AI technologies to gain experience.

AI for next generation public services

This breakout session focused on the next generation of public services based on AI. After an introductory presentation of a so-called “no-stop-shop” concept, eight people intensively discussed to what extent AI is suitable to support citizens in a seamless interaction with public services provided by public administrations and authorities, and how this can be practically implemented. In this context the justification of AI in public services is crucial. It must be made clear to users which services use which data and for what purpose they are used (transparency). Only through transparency can the acceptance and thus also the use of the services by the citizens be guaranteed. But there are also barriers to overcome on the part of the authorities. Civil servants working with these new technologies also need to be educated to see the benefits and be encouraged against the idea that technology will take over their jobs. With the mass of data and sources available another important aspect is to increase the availability and reliability of data sets. A possible two-step process for verifying data and services was proposed and discussed. Moreover, it is necessary to think about how to deal with false positives & false negatives, what the consequences could look like and who is authorised to use which service. Accordingly, the aim for AI models that are reliable and explainable is crucial but they will also depend on which kind of service is being addressed and which criteria has to be fulfilled. Instead of a one-fits-all approach, a fully integrated research approach including several perspectives and aspects is needed. This requires different experts to work together: AI experts, the local authorities, lawyers, etc.

AI expertise in the public sector

The public sector faces several challenges in attracting talents and empowering their employees to provide AI-based solutions. In this breakout session, the experts discussed the specific needs for AI training and upskilling programmes, and how these needs can be aligned with existing academic activities. First of all, there was a clear need for education on AI identified, including the importance to understand the different ‘awareness levels’ which need to be addressed: ignorance, belief in AI fairy tales, not believing in AI at all, or even being afraid of AI. Also terms like ‘prediction’ for example can mean different things for an AI researcher and for a policymaker or public health authorities. Furthermore, politicians can sometimes perceive ‘AI as magic’ and thus expect way too much. So, it is important to train administrators on what AI is, and to manage expectations for civil servants on the one hand, and on the other hand also train AI experts on the specificities of public administrations’ work.
One of the main themes identified in this breakout session is to challenge the current mentality in order to make use of AI technologies. In the public sector, it is sometimes difficult to accept failure even in small scale projects, so a mindset change is required towards an openness for experimentation. To make the benefits obvious for all stakeholders, testlabs for citizens to play around with AI could be created for example. Also, AI-based support for citizen involvement could be beneficial, for example by promoting citizen science and citizen-based control mechanisms. Furthermore, municipalities sometimes face barriers based on a ‘not invented here syndrome’. This could be addressed with knowledge transfer

between private sector and public sector, and between governments, and be supported by a coordinated action plan on AI. It is important to share data and models, but mostly experience is fundamental. It would be useful to have a resource pool of experts and cases to draw from. Skills and capacity are also needed to make better procurement, including an open source strategy.

Statistics of AI usage and accountability

On the second day of the workshop, representatives from the private sector, academia and official statistics institutes continued their discussions from the first day (topic: Reliable statistics) in the breakout session on Statistics of AI usage and accountability. One of the five key elements identified during the session focused on a certification approach (including governance) as a suitable area to consider for AI statistics, also on a European level as it is crucial for the public acceptance of AI. The second key element that was discussed was an algorithm register linked to the importance of reflecting the usefulness and benefit from an innovator-like perspective for the use of AI. The participants of the breakout session also talked about the possibility of making software code available to the public (open source) in order to improve accountability and trust individually addressed per project and to give the right perception to the public that the use of AI is ethical and transparent. The last two key results of the discussion dealt with the measuring performance of AI ecosystems and a discussion on how diversity and bias can be measured. This also brought up the idea to start working on the question of how to measure ecosystems performance based on data driven approaches, e.g., by organising a hackathon around this topic.

Transparency

The discussions in this breakout session picked up different insights and findings from the first day regarding transparency aspects in different application areas. The session was mainly focused on but not limited to data transparency, and also included topics like having datasets available or providing a clear communication on what the data is used for. The outcomes on this topic relied also on the requirements published by the High Level Expert Group, namely on explainability, communication and traceability. Explainability needs to be tailored to a specific audience and on different levels to be explicit, e.g. via user panels. It was also considered important to communicate statistical principles of uncertainty, meaning the limits of the respective methods in documentation, education, interpretation and making conclusions. For communication, existing misleading narratives of AI might hinder the communication for transparency in AI. Therefore, the traceability of AI methods plays an important role to transparency by making information like the intended use of the data (who is involved) or explanations (who is responsible) accessible. All three components for these requirements have to go hand in hand to enable transparency. The participants also talked about the idea of looking into the EU AI Act, specifically into the methods and not only into the applications. Another topic discussed was transparency and having a human in the loop with stricter requirements on keeping human awareness and information to ensure transparency. One of the main takeaways of the session was that there are many facets and requirements to fully implement transparency which should be accomplished together with all involved stakeholders.

Systemic analysis of AI effects

In the breakout session, a variety of different topics were discussed. The participants talked about how the role of the public sector could be transformed in terms of AI, and how new elements like AI can be identified in the ongoing digitisation of the public sector. Linked to these new elements, new risks can appear that could cause harm (data leaks, language models, etc.), so one of the emerging questions was how the government should be equipped to deal with these issues. The participants also discussed the topic of inhouse development in comparison to existing dependencies to the private sector and how different stakeholders handle the rules i.e. on image search related content. It was considered important how systemic effects and government functioning could be measured, also taking into account social and technical issues like existing distrust of sharing data with the government versus sharing data with private stakeholders. In terms of funding, the group identified the need to rethink how investments in technology and change need to be organised from a governmental point of view.
In summary, the key points of the discussion were the understanding of unknown unknowns (risks and systemic risks) with the Covid crisis as a good example to learn from, as well as the question on how to make public services for all by also meeting the user expectations. One of the main insights in this context was, that considering AI technology alone is not sufficient, but a more systems-based approach is necessary, also in order to guarantee system safety for example. As one idea to address this, sandboxes to reproduce some algorithmic decisions were discussed. Another key element identified during the discussion in this breakout session was the challenge of oversight and public administrative reform and the political alignments and to approach trustworthy AI as interdisciplinary.

Public trust in AI systems

The fourth breakout session addressed the topic of **Public Trust in AI**, which was also discussed frequently during the first day. The experts identified four important topics/actions which have to be taken into account to achieve a beneficial use of AI:

1) Ensure **trustworthiness** at both the technical and process level

   'Technical' level of trustworthiness: Transparency, robustness, explainability, accuracy, data quality and bias

   Process-related trustworthiness: Organisational context, goal function, testing & certification, political decision-making; also not too much transparency but "just enough transparency"

2) **Better educate** the public, the professional users, and developers/designers in AI

For the public, AI needs to be 'demystified'. Many people today understand the basics of cars without being mechanics, or understand the internet without being computer scientists. So we need to convey the basics of AI in the same way, without expecting the public to become AI scientists. A new generation will take AI technologies and applications for granted that already existed when they were born, but they may take some aspects too

much for granted, like privacy/sensitive data for instance. This needs to be taken into account.

Public-private collaboration

Collaboration among the public and private sector can unlock the potential of AI in the provision of services for citizens and improve the effectiveness of the public administration. However, the capacity to process technical requirements at large and within complex organisations is challenging, especially in the context of the Public Sector. Public administration usually lacks the competences that will enable an effective acknowledgment and monitoring of the impact of AI adoption in their context. The provision of practical examples may make this process smoother and simpler, even to organisations that face this challenging set-up. On the procurement side, rules still have a steep road ahead due to complex formal procedures. There is a need to transform these rules and make them more flexible as technology evolves. Flexibility is particularly important on small tenders and bids, so that Public Administrations have the opportunity to test advancements in strategic fields before scaling up to widely extended programs.
During the discussions in this breakout session, the following four main topics were identified:

- Structuring innovation partnerships between the public and private sector for enhancing collaboration should have a high priority. Partnerships need to rely on co-creation and co-development of AI applications, single events and workshops are not enough
- An economic model for innovation regarding the public and private partnerships should be developed in order to implement AI.
- In the education context, new and innovative ways of transferring knowledge between the public and private sector should be elaborated. Both in terms of upskilling human resources and providing practical examples of the possibilities of technology in their fields.
- Regarding potential implementation, writing tender requirements is difficult when you do not know what to ask for, or what the questions could be. Accordingly, there is a need to develop and disseminate guidelines for civil servants to procure AI solutions.

Input to the AI research and innovation roadmap

Based on the results summarised in the previous section, the Organising Committee identified several topics which could be a valuable input to a European AI research and innovation roadmap. These topics will be presented to and further discussed with experts from TAILOR, VISION and CLAIRE in order to enrich the respective roadmap activities.

The below topics are the ones that stood out most prominently and will thus provide the ‘core’ of the input. However, when the roadmaps will be constructed, all inputs from the Theme Development Workshop will be considered.

Public sector specific

- **Education**
  Due to a lack of proper understanding, new technologies are marginalised in the processes of public administration, and the dialogue with technical developers is complicated, reducing the potential benefits and impact of AI technologies and solutions. Tailored education for civil servants and other public sector workers could be a possible approach to address this challenge, especially by focussing on
    - A better understanding of the general framework for the potential introduction of AI in the processes of public administration,
    - Balancing expectations, having a more concrete view of limits and capabilities of AI
    - Increasing acceptance of AI as part of the future working activities.
- **Measure performance of AI ecosystems**
  The deployment of AI in society by governments will have systemic effects. Citizens, companies and other organisations will change their behaviours. In order to detect potential harmful side-effects, it is necessary to be able to measure the effects of AI in society in a systematic way. One of the underlying aspects that could be measured is the performance of AI ecosystems: clusters of companies that develop AI, (local) governments that stimulate the update of AI, end users (both companies and citizens) that use AI.
- **Algorithm register**
  The development and deployment of algorithm registers addresses a number of concerns that are related to the usage of AI in the public sector. Such a register provides a way to implement transparency and could also be a basis for public accountability. The concept could be extended so that it fosters citizen engagement, for instance by supporting citizen science initiatives.
- **Procurement, and market creation**
  Governments can play a role in market creation and thus influence developments in a favourable way. A number of considerations are relevant here: is there a place for in-house development and how does this relate to procurement from private companies? How should investments be organised, what is the role of public-private partnerships? Also, guidelines for tender processes need to be updated as a result of changing requirements.

More general topics not limited to the public sector

- **Data**
  Data is vital to produce AI solutions, but the availability of a large amount of data is not a unique requirement; also the quality and accessibility of the data are key requirements to produce and replicate trustworthy AI solutions, for both public and private sectors. This includes:
    - the need to overcome information silos in different public organisations and (potentially) private actors,
    - the design of both governance models and technologies for data sharing infrastructure (as enabler for trustworthy AI solutions) ensuring availability, quality and accessibility of the data.

- **Requirements of AI**
  The requirements for AI systems as identified by the High-level Expert Group on AI are still very valid and relevant, and many aspects still need to be further detailed and made more concrete. Important is also a broad and integrated view on these aspects. An additional consideration is how certification could support the promotion of trust and adoption of AI systems that are used in the public sector. This could be organised alongside the previously mentioned algorithm register.
- **Systemic approach and life-cycle management**
  A more integrated approach towards procurement and deployment of AI is necessary. It is not only important to set out clear procurement guidelines, at the same time, it should be ensured that the necessary knowledge and resources to operate and maintain the system are in place. This is specifically important because AI systems can adapt themselves, so monitoring of the system performance is needed: is the system still operating within the scope for which it was designed and trained? How does the system react to new or adapted other systems? When will a system become end-of-life? There are also links to the previously mentioned topics of ‘measuring systemic effects’, ‘Requirements of AI’ and ‘Education’, which could be further investigated.

Theme Development Workshop “Future Mobility - Value of data and trust in AI”

About the workshop

The second Joint Theme Development Workshop (TDW) co-organised by CLAIRE, HumaneAI Net, TAILOR and VISION[34] on “Future Mobility - Value of Data & Trust in AI” took place on the 28th of October 2021 with the aim to develop and identify the most promising and emerging AI topics in the mobility and transportation sector. At this one-day workshop, experts from academia, industry and politics jointly developed initial input for the European Artificial Intelligence (AI) research and innovation roadmap. The full report is publicly available on the TAILOR website[35].

Facts and figures

The high international interest that was expressed in response to the announcement of the AI for Future Mobility Theme Development Workshop translated into excellent attendance of the event.

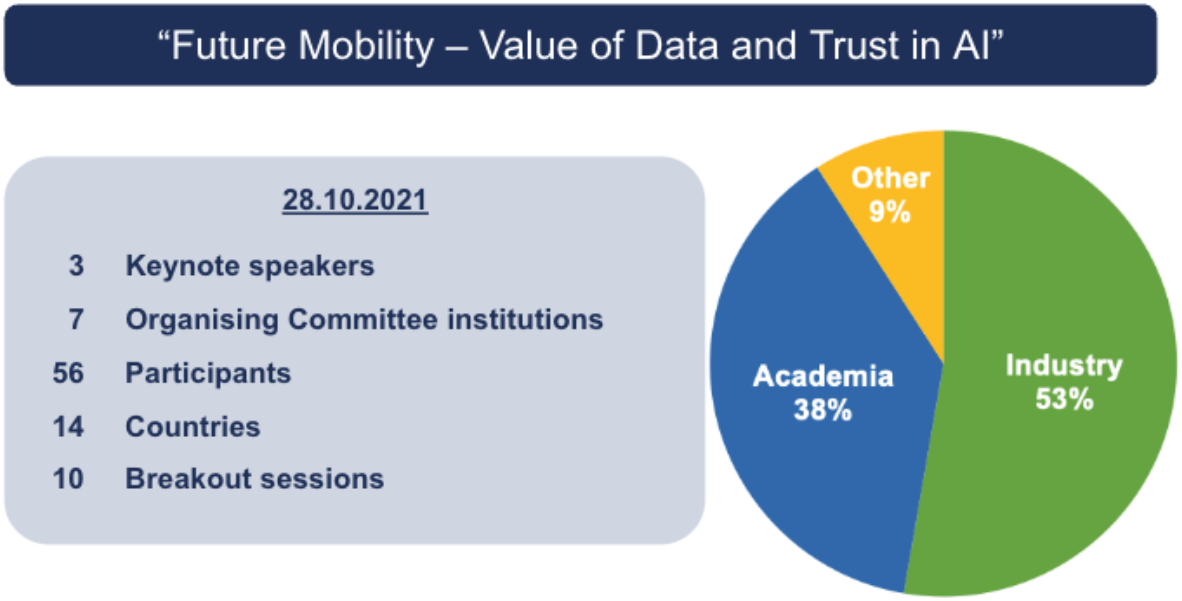

Figure 7: Facts and figures of the TDW “Future Mobility - Value of Data & Trust in AI”

Fifty-six participants joined the TDW, ranging from a diverse set of backgrounds. Fourteen (predominantly EU) countries were represented, with thirty participants indicating that they are affiliated with industry, whilst twenty-one participants indicated that they are affiliated with academia (five participants indicated “other”). The participation of major industry representatives, with companies like ZF Group and the Volkswagen Data Lab, is particularly noteworthy and testifies to great interest on the part of industry. Equally important being the participation of those affiliated with the European Commission. The TDW, therefore, caught the attention of some of the most important actors in the field of Future Mobility and brought

[34] In alphabetical order
[35] See project website: https://tailor-network.eu/events/future-mobility-value-of-data-and-trust/

together representatives from key companies, supra-national institutions, and academia. The workshop thus successfully provided a platform for discussions between representatives from academia, industry and politics: Discussions that are key in unlocking the full potential of AI in Europe.

Program

The TDW was opened by the Co-Chairs Philipp Slusallek (DFKI) and Jan Alpmann (German Entrepreneurship) on behalf of the Organising Committee (OC), which included further representatives from CLAIRE, DFKI, German Entrepreneurship, TNO, Volkswagen AG, LMU and ZF Group. The Co-Chairs outlined the objectives of the TDW as well as the agenda and programme, and introduced the invited keynote speakers to the participants.

Figure 8: Flyer of the TDW "Future Mobility - Value of Data & Trust in AI"

The inspiring keynotes were provided by high-level experts from academia and industry. These introductory presentations served as a basis for the discussions about the value of data and the aspect of Trustworthiness in the mobility and transportation sector, and provided some interesting examples of application areas. Accordingly, these presentations stimulated the expert discussions in the following breakout sessions. Introductory presentations by Dr.-Ing. Christian Müller, Dr. Manuel Götz and Dr.-Ing. Marc Hilbert

Dr.-Ing. Christian Müller, Head of Competence Center Autonomous Driving DFKI Saarbrücken, primarily focused on advancements in core AI and validation and verification in his keynote, but also featured examples from explainable AI and human-machine

cooperation. Specifically, Dr. Müller focused his attention on a very promising set of methodologies, most commonly known as Hybrid AI – the combination of Machine Learning and Deep Learning with symbolic AI methods.He also focused on patterns for explainable learning systems through rational reconstructions. Furthermore, his presentation covered informed learning with prior knowledge, where a learning system is guided by information and learning can be constrained through symbolic constraints, enforcing safety conditions. Finally, he focused on patterns for learning an intermediate abstraction which have a high potential of overcoming the key limitation of learning purely statistical correlations.

Dr. Manuel Götz, Head of AI & Cyber Security Technology Center at ZF Group, started his keynote with an introduction to Autonomous Driving – highlighting the huge step-up in complexity for highly automated driving. He then moved on to the main part of his presentation, explaining what trustworthy AI means in the context of automotive products, and which issues the industry is facing. Another topic covered was verification and validation, where realistic simulation models and improved statistical methods are particularly relevant. Last but not least, Dr. Götz mentioned AI governance and monitoring, addressing in particular the necessity of appropriate documentation, traceability and quality control of data, labels and AI models in product software.

Dr.-Ing. Marc Hilbert, Team Lead VW Data: Lab Munich at Volkswagen AG, started his keynote with a quick overview of Data:Lab Munich and how it has utilised Machine Learning to develop data driven products for Volkswagen (VW). The first major topic of the keynote was focused on the difference between innovation and invention and that innovation is key for companies. The second major topic that Dr. Hilbert focused on was the current state and development of how to achieve AI innovation. The last major topic of the presentation was focused on explainability and trust.

Key results from the breakout sessions

The following subsections will provide a detailed explanation of the results of the breakout sessions of the Theme Development Workshop “Future Mobility - Value of Data & Trust in AI”.

Trustworthy AI for Future Mobility - Group A

The Breakout Session ‘Trustworthy AI for Future Mobility ’ was very popular among the participants, so the discussion was split into two groups. Group A was well balanced with participants with a background in Autonomous Driving and other domains in Mobility. During the session, commonalities were identified, for example that many aspects are applicable outside of Autonomous Driving in terms of trustworthy AI for Mobility. To tackle trustworthiness, a fusion of the domains would not necessarily help to reach this objective due to known problems like compatibility with SOTIF standards in Autonomous Driving. However, some of the Key Trustworthiness Indicators (KTIs) like fairness, non-bias or robustness are also needed in public transportation. One of the key takeaways of this session was to correctly classify domains and to neither neglect their commonalities nor their differences. One of the ideas to expose data to the community was to use large-scale publicly governed data to establish trust as a long-term goal. As for challenges and

roadmapping, one idea was to use concise data bases that are easier to obtain. Another idea was to generate a roadmap showing how to approach Trustworthy AI in a holistic way.

Trustworthy AI for Future Mobility - Group B

The second group discussing 'Trustworthy AI for Future Mobility' consisted mainly of participants from academia, startups and SMEs, allowing for a view on this topic from different perspectives. During the session, the group agreed that Europe is currently handling data in the right way, and that this creates trust among the public. One of the main topics of the discussion was the question how black boxes can be transformed into a more transparent "glass box" to create more trust in the model. The key takeaways of this session were to explain how data is being used to increase trust, to have more experts and resources in Europe and also to invest in projects to educate people how to build good models and to remove biases as well as collecting data in the right way. The participants also identified wishes towards a EU research roadmap, namely the wish for more standards for data-driven programming, to improve the media coverage on the topic and to also have more investments on the hardware side to have secure supply chains for a fully trustworthy system.

Explainable AI for time series & verification approaches

The breakout session 'Explainable AI for time series & verification approaches' dealt with the question which possibilities exist to verify methods and how their suitability to certain problems could be measured. The group considered the topic as very complex, also in terms of a clear distinction of trustworthiness and explainability, namely asking which use cases require explainability or trustworthiness. One idea to follow-up on this was to organise a dedicated workshop, gathering stakeholders from various backgrounds to discuss these distinctions between trustworthiness and explainability in more detail, especially as many open questions of explainable methods exist in general. Therefore, a generalisation was 7 considered difficult. A question that came up in this regard was how generalisation could be established not only from a local but also from a global approach. It was also agreed that it is important to consider multiple methods in close cooperation with experts to ensure that the algorithms work properly and give reasonable results. The overall goal would be to create a xAI rulebook. To get there it could be helpful to first categorise AI application areas. As an outlook, the participants regarded it as necessary to generate ideas and find suitable partners for cooperation as the topic is relatively new. They shared the idea to use a big data pool for everyone to work and train on common datasets for an easier understanding and to identify potential errors. Estimating the value of data In the breakout session 'Estimating the value of data' the participants addressed the complex question of how the value of a given dataset could be estimated. During the discussion, a lot of ideas were exchanged, i.e. the question how to put numbers (or Euros) behind data to establish trust in customers who share their data. The key takeaways of this session for estimating data and generating a roadmap were that data has value to different customers, so estimating the value of data is considered as an important step concerning writing down the actual value of the data. Also, the group identified different types of values to the customers, i.e. on an individual, business or social level. An important element would be to generate the value on the customer and the vendor side to allow for a comfortable feeling when contributing data to certain

applications. The group also agreed that data always has a certain bias which needs to be taken into account when formalising and generating a roadmap for estimating value of data. Finally, the discussion concluded with the insight that there is no linear relationship between value and data and that data has to have a good quality to be rightly estimated in its value.

Towards standardisation & certification of AI

The breakout session “Towards standardisation & certification of AI" aimed to discuss the challenges of developing standardisation and validation processes for machine learning components in the context of camera sensor-based algorithms, i.e. in Autonomous Driving. After a quick warm-up session, the participants dived into the discussion about the aims of certification. One of the main aims identified is to bring AI into the market while safety, security and privacy (e.g., video recording by an autonomous car) as well as explainability, robustness and steadiness are further important aspects and aims of AI solutions. In addition, another goal of the certification of AI systems is the controllability by humans, so that humans have control over the AI system at any time and can intervene if necessary, for example when the driver still has the opportunity to disagree with the Advanced Driver Assistance Systems (ADAS). In this context, a new way of controllability might be the process of keeping the human in the whole system through the human-in-the-loop mechanism. The second topic discussed by the participants was the aspect and meaning of self-awareness of the systems, meaning that the system can assist itself in case it is unfamiliar with the situation. This aspect could also address the safety argument of certification. Some possible solutions were highlighted by the participants, like Bayesian 8 Inference Technique for handing over the control to humans or how to quantify the uncertainty or methods for quantification of uncertainty. Another item discussed in this breakout session was the idea of “Breaking the Rule” in Autonomous Driving, which means that Autonomous Driving systems should be allowed to break traffic rules in case of an emergency. However, this would make it more difficult to get a certification of the AI system, which has led to solutions for special cases like this by redefining the rules for Autonomous Driving where necessary.

AI expertise in Future Mobility - Group A

Also this topic was very popular among the participants, so the discussion was split into two groups. In group A, participants from industry and academia discussed the challenges in accessing expertise and attracting talents to AI research departments in academia and industry. While many unsolved problems and undefined needs in developing AI expertise remain, industry and academia in particular should work together to identify the kind of expertise that is needed in a specific field, and also to define a basic level of AI related knowledge needed for leadership as well as for the general public. In this context, the participants also discussed the question of what the specific needs for AI training and upskilling programmes are, and how these needs can be aligned with academic activities and doctoral programmes. They came to the conclusion that knowledge management, especially sharing AI knowledge in the AI communities and initiatives, is a key factor besides bridging the gap between the needs in industry and the training in academia through small courses and activities in universities (out of degree). It was also considered worth mentioning, that it is becoming increasingly difficult to attract interdisciplinary AI researchers

to industry and research across Europe's borders or to prevent them from leaving. One reason identified by the group was that visa applications and entry requirements are becoming more complex. They also added that AI researchers in academia often cannot gain enough practical experience, while many individuals move to industry too early and therefore do not deepen their basic research and application-driven research.

AI expertise in Future Mobility - Group B

The second group focusing on “AI expertise in Future Mobility” also discussed specific needs for AI training and upskilling programmes and how these needs can be aligned with academic activities and doctoral programmes. To structure the discussion within the group, the participants divided the topic into three pillars: (1) AI technology for AI experts who build AI systems; (2) Other users who use AI systems to build other non-AI-systems; (3) People who simply use AI systems. The discussion has shown that within the first pillar, scientists from various disciplines with problem solving skills, for example physicists and engineers, start to work with AI, but they might need more training in AI. This is why AI or Data Science in general should be taught more in the various scientific courses in academic studies. This issue also became relevant in the second pillar where people need more skills on statistical knowledge and data handling, i.e. comparable to the level in algebra. Therefore, Data Science and Statistics should play an important role in all curricula in a way that it complements but not replaces disciplines like algebra or such. To ease the use of AI for 9 non-AI researchers, the solution could be to build (modular) frameworks in order to reduce the complexity of AI models. On the other hand, a guarantee for these modules would be required. It seems that a variety of frameworks exist that are widely used but may be too complex for other disciplines (e.g., biology) and too tailored to Computer Science. Therefore, visual and/or graphical user interface tools and “AI as a Service with a web interface” are needed. During the discussion, it also became clear that technology needs to move towards the average user (pillar 3), but users still need to “understand” the general behaviour of an AI system, including the capabilities and limitations of those systems. Otherwise it might cause the risk of a so-called “Uncanny Valley” in the Mobility & Transportation sector, leading to mistrust and fear of using AI. Accordingly, it is necessary to allow people to have a realistic impression of the technology through public engagement and communication for example, and to build trust in the wider community.

Reliable Confidence Measure

Currently used deep learning methods do not produce sensible confidence. Therefore, the reliability of deep learning methods in safety critical use cases cannot be assessed and trusted. In this breakout session the participants discussed reliable confidence measures to handle these challenges. The participants jointly elaborated the term "Confidence Measure" as a basis for further discussion. They came to the conclusion that measurable metrics for this term could be safety, robustness, “out of distribution data”, “adversarial attack” as well as certain quantifiable measures from 3rd party validation, which seems to be more trustable to the user. Based on their common understanding, the participants also identified the challenges of Reliable Confidence Measure in the Mobility & Transportation sector, where solutions need to be developed. According to the group, the Deep Neural Network (DNN) models are nowadays over-confident, whilst, with respect to the Reliable Confidence

Measure, it might be better if a model is under-confident, enabling the user to take better action, e.g., in case of an emergency break. The Mobility sector faces another challenge, namely that in the research community people from industry as well as academia are often more interested in “high accuracy” in Machine Learning when it might be better to have another type of matrix to keep track of reliable confidence for cases like Expected Calibration Errors (ECE). These challenges led the group to to formulate some ideas for possible Hackathons in the Mobility and Transportation sector like “DNN Models Benchmarking keeping Reliable confidence measure in focus” to identify additional performance metrics like ECE or adversarial attacks to assess their suitability and reliability as performance measures and regularisation elements. In this context, the group also discussed potential assessment strategies, resulting in the finding that large test datasets are needed in a series of runs to make overfitting difficult.

Machine Learning in the context of personal data and GDPR

The driving question of the discussion in this breakout session was how data from individuals can be used for Machine Learning to create GDPR compliant human-centred AI applications. This also includes the issue of how ‘private’ data can be related to the certain exclusivity between privacy and data utility. The more private a data set gets, the less utilisable it is for 10 building up Machine Learning models. Under the umbrella of GDPR, the most important aspect for developing Machine Learning algorithms in research institutions is to have a trusted middleman that governs over the data together with an ethics committee (that only grants access to people for research purposes falling under Article 89 of the GDPR). This could also partly apply to industry applications. In industry, other options need to be considered, so that the personal data can still be used without any infliction on people’s privacy. One way that was discussed in this group could be Edge Computing for extracting data without touching individuals’ sensitive data, i.e. by using blurred images. The arising challenges for such a solution however are computational limitations for example. Another option using personal data would be to use federated learning with morphic description. An obvious problem is that these data sets are often biassed, which raises the question of whether conclusions can be drawn about the individual based on these biases.

AI sensitivity analysis for time series

In this breakout session, the application of AI in safety-critical situations and the importance of knowing the influence of input signals on an output due to safety reasons was widely discussed among the participants. It was also discussed whether pre-processed features can improve safety. The drawback of using pre-processed features is that the AI might miss some important input signals. Another item discussed in this group was the question whether there is some kind of benchmark for safety in AI. It was further elaborated whether safety categories with recommendations can further improve the safety development of AI algorithms. Nowadays, there are still some unsolved problems regarding the safety of AI. An example of this are the so-called ‘black swan events’ (tail of distribution). To handle such events, AI systems should be robust and resilient to these sorts of events. The participants agreed that it would be beneficial for the AI system to be able to tell the user that it does not know how to handle a specific situation and delivers a kind of certainty value. All participants

agreed that much more research on that field is necessary to overcome the lack of knowledge.

AI for energy autonomous assets

In the transport and logistics sector, there are many moving goods without power supply. More and more of these assets are connected via (battery-powered) IoT devices to collect their positions and data. However, data transmission is a major challenge here. Accordingly, this breakout session discussed the opportunities and challenges for the use of AI under such conditions. From an energy consumption point of view, it is certainly necessary to plan the device's computing and communication operations carefully to optimise energy consumption. AI algorithms (constraint planning) have shown in the past, also before IoT, that they are able to successfully manage devices in a high constraint environment, e.g., satellites and rovers for space exploration. So, it is necessary to reflect on whether the AI application should run on the device itself or on a cloud server. Decentralised, multi-agent oriented AI algorithms can offer solutions in multi-IoT systems where centralised solutions are difficult with regard to data transmission limitations. The centralised versus decentralised AI dilemma is discussed within the classical computing paradigm/hardware. The group also 11 touched on some future issues in this context, for example that quantum computing with its potential application to AI would make the dilemma even more interesting/harder, since quantum machines nowadays are all cloud-based and too size-overwhelming to be IoT devices.

Input for the roadmap

Based on the results summarised in the previous section, the Organising Committee identified several topics which could be a valuable input to a European AI research and innovation roadmap. These topics will be presented to and further discussed with experts from TAILOR, HumaneAI Net, VISION and CLAIRE in order to enrich the respective roadmap activities. The below topics are the ones that stood out most prominently and will thus provide the ‘core’ of the input. However, when the roadmaps will be constructed, all inputs from the Theme Development Workshop will be considered.

Mobility and transportation sector specific

- **Grasp Trustworthy AI holistically**
  Trustworthy AI is a difficult to grasp topic that should be perceived and approached holistically, including the areas of “Robustness & Security”, “Human-in-the Loop & Explainability”, “Ethics, Privacy & Liability”, “AI Governance & Monitoring”, “Verification & Validation”, “Data Availability/Quality”, “Reliability & Safety”. Accordingly, these topics should be addressed in a European AI research and innovation roadmap. On a political level, it would be beneficial to consider using the same terms in ongoing and future initiatives, especially in the area of AI Ethics (digital Ethics)
- **Data**
  Data enables new AI technologies to be developed, tested, and implemented. This is particularly true in the automotive industry which relies heavily on data to, among

other things, develop autonomous driving solutions. There are, however, significant issues related to the availability and usage of data that are holding back the industry. These challenges could in part be addressed by:

1. Creating and using easy to obtain data bases.
2. Creating better standards for data-driven programming and increasing investments on the hardware side to have secure supply chains for a fully trustworthy system.
3. Creating a big data pool for everyone in the automotive industry to work with and train on common datasets for easier understanding and error identification.
4. Redefining some rules regarding the certification of AI systems and specifically Autonomous Driving technologies, e.g., if an AI system does not know how to react/handle a situation, it should inform the user.
5. Ensuring that users are able to understand the general behaviour of AI systems, including their capabilities and limitations which requires increased data transparency. – Otherwise there is a large risk of mistrust and fear of using AI.

- **Communication**
  The wider public must be able to trust new AI technologies that the automotive industry develops; otherwise such technologies will fail. One of the best ways to increase such trust is by communicating information about new AI technology to the wider community and creating positive public engagement

More general topics not limited to the Mobility and Transportation sector

- **Trustworthy AI & Explainable** AI
  Trustworthy & explainable AI are very important for the success of AI in Europe. This TDW has uncovered some important insights in relation to these two important topics:
  - Improving media coverage on the topic of trustworthy AI should greatly improve trust in AI.
  - Differences between explainable and trustworthy AI must be addressed and clarified to avoid spreading confusion.
  - Knowledge management regarding AI and especially sharing knowledge in the AI community is very important to spread awareness of trustworthy & explainable AI.
- **Data**
  Some important conclusions regarding data from this TDW are:
  1. Data has different value to different customers – estimating the value of data is very important.
  2. Data is always biassed to a degree, which should be considered accordingly. Also there is no linear relationship between value and data.
- **Academia / Education**
  - Discussions in this TDW revealed important issues regarding education and AI across Europe. Specifically, it has become clear that too little AI training is offered at universities and that research roadmaps are often lacking this

important topic. Additional AI training through small courses is essential (outside of degree or secondary school) in academia/school.

- It is becoming increasingly difficult to attract interdisciplinary AI researchers to Europe (visa process/entry requirements are some of many limiting factors). This issue should be addressed.
- AI researchers in academia often cannot gain enough practical experience, but some also move to industry too early and fail to deepen their basic research skills. From the perspective of the industry, however, future talents gain practical experience too late. Industry needs both people that have a deep research background, but on the other side people who are willing to start a career in industry early with focus on applying AI knowledge. In addition, because of the lack of talent, the industry has to overcome this problem in upskilling people quickly. The resulting gaps in training and experience should be closed.

Theme Development Workshop “AI for Future Healthcare”

About the Workshop

The third Joint Theme Development Workshop (TDW) co-organised by CLAIRE, TAILOR and VISION[36] on “AI for Future Healthcare” took place on the 16th December 2021 with the aim to develop and identify the most promising and emerging AI topics in the healthcare sector. The full report is publicly available on the TAILOR website[37].

Facts and figures

The high international interest that was expressed in response to the announcement of the “AI for Future Healthcare” Theme Development Workshop translated into excellent attendance of the event.

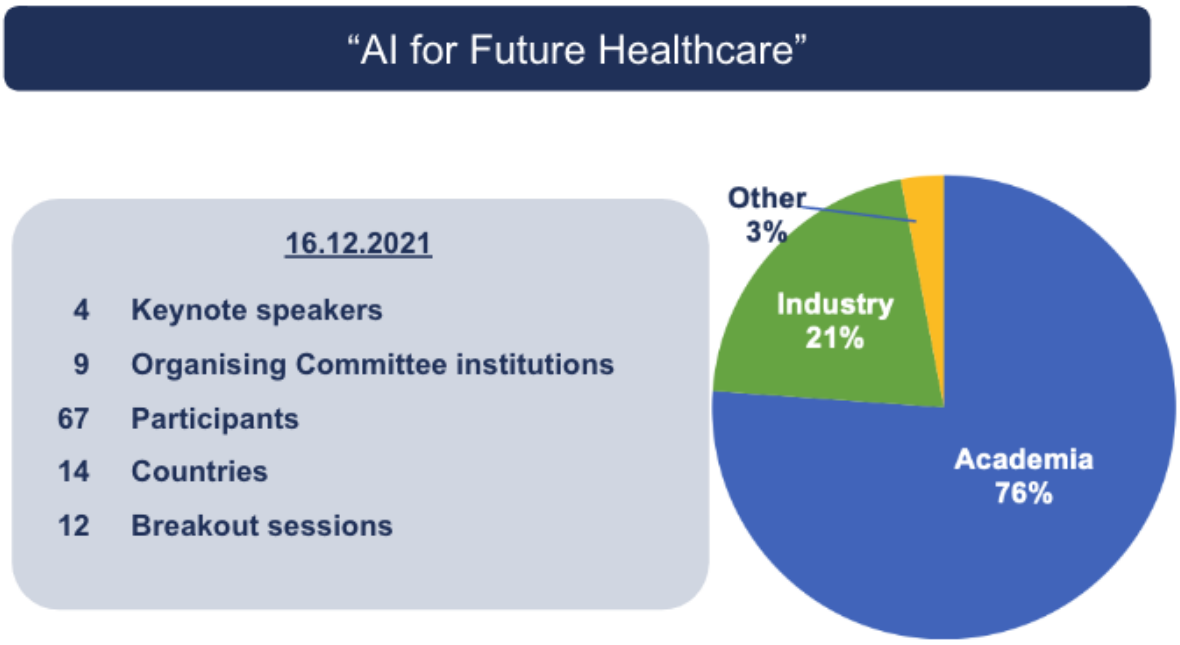

Figure 9: Flyer of the TDW “AI for Future Healthcare”

Sixty-seven participants joined the TDW, ranging from a diverse set of backgrounds. Fourteen (predominantly EU) countries were represented, with fourteen participants indicating that they are affiliated with industry, whilst fifty-one participants indicated that they are affiliated with academia (two participants indicated “other”). The participation of major industry representatives, with companies like Philips and NTT Data is particularly noteworthy and testifies to great interest on the part of industry. Equally important being the participation of those affiliated with the European Commission. The TDW, therefore, caught the attention of some of the most important actors in the field of Future Healthcare and brought together representatives from key companies, supra-national institutions, and academia. The

[36] In alphabetical order
[37] See project website: https://tailor-network.eu/events/ai-for-future-healthcare/

workshop thus successfully provided a platform for discussions between representatives from academia, industry and politics: Discussions that are key in unlocking the full potential of AI in Europe.

Program

The TDW was opened by the Co-Chairs Ricardo Chavarriaga and José M. Sempere on behalf of the Organising Committee (OC), which included further representatives from CLAIRE, DFKI, LIH, NTT Data, Philips, VRAIN and ZHAW. The Co-Chairs outlined the objectives of the TDW as well as the agenda and programme, and introduced the invited keynote speakers to the participants.

Figure 10: Flyer of the TDW "AI for Future Healthcare"

The inspiring keynotes were provided by high-level experts by Prof. Dr. Ulf Nehrbass, Dr. Alfonso Valencia, Dr. Nicola Pezzotti and Anna Forment.

In his keynote, **Prof. Dr. Ulf Nehrbass from the Luxembourg Institute of Health (LIH)** presented the Clinnova project. The Clinnova project consists of a large network with partners from Luxembourg, Metz/Nancy, Strasbourg, Basel, Mannheim, Saarbrücken and Freiburg with the strategic goal to unlock the potential of data science and AI in healthcare, i.e. with better data driven methods, increase in diagnostics efficiency, precision medicine, medico economic savings, opening prevention effort and better patient care. Clinnova considers AI innovation in healthcare an operational and organisational challenge where the

key innovation driver is not the AI algorithm but the data-enabling environment that produces standardised and quality-controlled data around relevant use cases.

The second keynote was presented by **Dr. Alfonso Valencia** on open questions in the topic of AI, genomics and personalised medicine. The path for precision or personalised medicine generally consists of a computational infrastructure, like databases, methods, systems and resources to perform the analysis of the data. The two following components are the research environment, like epigenomics, proteomics or metabolomics and the clinical environment, like epidemiology, medical devices or medical images. Both components are required for the interpretation of the data and consequently the multi-omic report that is helping the patient.

In his keynote, **Dr. Nicola Pezzotti, Senior Scientist Ai from Philips research,** showcased AI use cases in the Healthcare Industry that have already been deployed within Philips. It is Philips' overall purpose to improve people's health and well-being through meaningful innovation by improving the lives of 2.5 billion people per year by 2030. To achieve this goal Philips developed a concept called the "health continuum" which is basically the idea that health is not only about diagnosis and treatment but rather starts with the aspects of Healthy living (Personal Health) to the prevention (e.g., Precision Diagnosis and Image Guided Therapy), Diagnosis and Treatment and the After care at peoples home. Dr Nicola Pezzotti presented specific use cases for each of these areas in order to explain them in more detail. that monitoring performance and impact in the real world will be key to mature as an industry.

The fourth keynote, held by **Anna Forment from NTT Data,** was about how AI can be applied in the healthcare sector to improve management of healthcare systems. With regard to the global approach, she explains the fields where AI can support healthcare management and its improvement. The presentation was concluded with three steps to support healthcare professionals decision-making: supporting decisions with imaging data by starting to implement image data analysis through deep learning capabilities, personalised and precision medicine by using all the information related to the individual's experience and health outcomes and applying genomics across all healthcare pathways and levels of assistance and finally, improving clinical assistance by exploiting clinical record data by leveraging the treatment of clinical records using NLP to make predictions about prognostics and aggravation episodes.

These introductory presentations served as a basis for the discussions about Trustworthy AI for Future Healthcare, data sharing in the Healthcare sector and Explainable AI. Accordingly, these presentations stimulated the expert discussions in the following breakout sessions.

Key results from the breakout sessions

The following subsections will provide a detailed explanation of the results of the breakout sessions of the Theme Development Workshop "AI for Future Healthcare".

Trustworthy AI for Future Healthcare - Group A

The breakout session started by identifying key enabling factors of Trustworthy AI in the healthcare sector. Participants rapidly pointed out that Standards are one of the most important aspects for Trustworthy AI. Currently, there are several international initiatives aimed at developing governance frameworks that can support AI trustworthiness. These initiatives cover aspects like ethical guidelines, principles for accountability and regulation proposals, but they do not have yet a global dimension and differences can be observed e.g., between the US, China and the EU. Increased efforts in the standard development are necessary, including incentives that promote the involvement of all relevant stakeholders. The explainability of the AI models is a crucial element for gaining the trust of all stakeholders as well as a likely requirement for regulatory compliance. In this respect, significant support is required in the development of methods that allow opening of the black-box. Regarding this, it is important to keep in mind that not all AI systems for healthcare rely on deep-learning techniques and approaches involving hybrid AI may play an important role in the future.

Also important is the availability of adequate infrastructure for the conception, development, and validation of AI systems. AI performance is dependent on the processes for data collection, curation and management. Therefore, the trustworthiness of these systems will require suitable, well-resourced infrastructures to efficiently perform these processes in a way that satisfies the performance criteria of service providers, regulators and patients while respecting social and ethical norms.

Involvement of field experts and practitioners is a sine qua non condition for achieving trustworthy AI systems for health. It is worth noticing that the trust of an AI-supported healthcare system will depend on all the elements involved, including the AI systems and practitioners who use them. Hence, practitioners need to be involved as early as possible in the life cycle of the AI system and be able to spend enough time and resources to understand the capacities and limitations of the AI systems. As such, cultivating a culture where practitioners exert healthy scepticism will be important to better identify the areas where AI can have more significant impact in the health of society, while limiting the risk of feeding the hype cycle and triggering excessive expectations. Multi-Stakeholder collaborations, like the one established by CLAIRE and the Hippo AI Foundation are a viable approach to facilitate this involvement.

In the foreseeable future the decision-making process in healthcare will remain on the human-side. These decisions will be supported by information inferred by AI-systems. There is thus a need to better understand how decision making and practitioner’s behaviour are affected by the introduction of AI-systems.

As mentioned above, the involvement of practitioners, patients and other multiple stakeholders is required throughout the entire life cycle of the AI systems (covering their conception, development, validation, use and decommission), as well as in the development of standards and governance approaches. It is thus important to stimulate the formation and career development of professionals who can be a bridge between all stakeholders. At the immediate level, it is important to identify which are the individuals and organisations that are leading current initiatives in the global sphere and ensure that the EU has a significant contribution.

Trustworthy AI for Future Healthcare - Group B

The breakout session started by identifying that data quality is still an important topic that should receive sustained attention. It is still not possible to have widely agreed metrics for data quality. With plans for the European Health Data Space, topics like reusability of data, metadata (provenance), but also detection and mitigation of bias, should all be elaborated. Not specific for the healthcare sector, but probably the most intricate and tightly related concept is privacy. Medical data is very sensitive and exploiting that data to its maximum use will inevitably create tensions between values. Privacy Enhancing Technologies are a technological development that can alleviate some of the problems and these need to be developed further. Another important related concept is sovereignty, meaning that the involved persons should be given good opportunities for consent management. Explainability has similar tensions with other values. Not only because associated transparency creates challenges with confidentiality of data, but also because autonomy and agency of users can be challenged when a smart AI gives explanations in a situation where information asymmetry between patient and doctor exists. AI should be able to mitigate the knowledge gap between patients and professionals. The main problem with explainability itself is the lack of a commonly agreed understanding of what 'explainability' is, how to measure it, how it promotes calibrated trust, and how to assess the quality of an explanation for different user groups. Explainability is an interdisciplinary topic and should have a roadmap of its own.
Another relevant topic is certification. This should develop beyond risk minimization and also take performance more and more into account. Apps and software, especially when they contain AI, should be treated like drugs in terms of certification. There was a concern that 'the horse has already left the stable' here, given the large amount of health-related apps that are currently being put on the market already. Just like in the food industry, where ingredients and key data about the food inside the package have to be put on the label, some kind of 'model passport' that informs users could be considered.

Data sharing in the Healthcare Sector

More and more data is being produced from various sources (e.g., medical devices, smart devices, public records), in different geographies, and is often owned by different parties like academia, hospitals, industry as well as governments. Data sharing can improve AI analysis: The richer the data collection, the more robust and reliable the model will be and it is a reality that a considerable amount of data is required to grant that the training process will have success finding these patterns. Taking this challenging scenario as reference, the breakout session "Data Sharing in the Healthcare Sector" gathered the expertise of relevant stakeholders – Data owners, data scientists, researchers and lawyers mainly – to discuss and brainstorm on the challenges and benefits of data sharing as well as on the key technologies and enablers currently coming into scene. Most relevant conclusions derived from the discussion and questions triggered by the participants can be summarised as follows:

1) GDPR is blocking research unwillingly. In future there should be more discussion about how to adapt GDPR for research purposes.
2) Data Silos are partially motivated by the lack of trust in data privacy and security mechanisms. It is a fundamental right to know what others are doing with your data.

3) The AI environment claims a number of benefits while gaining access to Data to train models. Nevertheless, data owners are sceptical to some extent in the sense that they don't yet see the real benefits of adopting AI in healthcare settings as a counterpart for providing access to data.
4) With regards to the inclusion of Real world data and synthetic data as a means to gain access to larger datasets, these are great approaches that help to train AI, however, it is essential that we have access to the real data.

One thing is clear: „sharing is caring“; and all stakeholders should contribute to address aforementioned challenges and collaborate to make health data really actionable data to leverage on and benefit research, benefit healthcare professionals and carers and far above all patients, real owners of the health data.

Trustworthy AI aspects on time series data analysis

In personal, connected and in-hospital care, time series data is a common and important form of input data (e.g., toothbrush localization, sleep phase determination through headphones). This breakout session focussed on important trustworthy AI components, challenges, and solutions for this type of data and its analysis.
The session started with a presentation by Rien van Leeuwen about the development of the Philips Smart Toothbrush, a customer device that works with a mobile app. The challenges that Philips encounters lie at the levels of user adoption as well as on the level of model development. The overall idea is that the app provides feedback about the daily brushing process. At first sight the situation and challenges seem to be specific for the product itself. However, the data modelling process as well as the attempts for efficient feedback from the user to the system holds for other products, in other markets, as well. In the group Feedback loops were found to be key in the process of adoption of the product. For time series data also the moment of feedback is important. Not only to improve the process of (technical) data modelling but also for the adoption (and thereby the amount of input data) is important. Another aspect that was mentioned in the discussion was the combination of datasets. E.g. brush activity data, safely kept local and group behaviour data. During the session we discussed different time intervals, different contexts and the use of federated learning on multimodal/multisensor data and early vs late fusion. From a technical as well as GDPR point of view it is important to think about similarity; when is the input similar to samples in a user specific dataset, when is it (significantly) different. Can SSL + clustering help us here? It is challenging to find the ground truth with respect to the datasets, and to relate that to unusual/random behaviour. Thereby it keeps challenging to interpret the data, discover outliers, and provide useful feedback to the end user (the brusher). A different approach for increasing further adoption is to combine different AI apps for health. This goes in the direction of lifestyle and prevention, where it is often not only one isolated problem that needs to be addressed. The challenges then shift to finding ways to fuse data for better interpretability, thus not only focus on precision, but instead focus on combining data. It is here where the field touches social implications social interactions.
At the end of the session the group discussed activities they would like to follow/organise to address the challenges. One one the ideas were to organise a hackathon / work session in which experiments with other cases can be compared to identify similarities in the challenges, possibly combine them, and relate them to technical and societal research

activities.
Perhaps it is useful to test a different approach in the direction of persuasion and coaching. Focus more on the social aspects which might lead to better (interpretable) data, and thereby classifiers. Maybe some themed sessions can be held. Also some practical advice came from the group: Maybe low(est) hanging fruit: a robust uninterpretable detection that warns you when the input is “different” and therefore network output can not be trusted. Next step would be to make this detector interpretable/less black box. The group coined the term “brushworthy AI”.

Trustworthy AI aspects on image segmentation and reconstruction

Medical imaging is a powerful tool for the diagnostics of diseases. With the advent of AI, classical segmentation algorithms have been replaced by much better performing convolutional neural networks. This development has opened up avenues for automatic image segmentation and reconstruction. Industrial use cases comprise tumour detection in computed tomography or magnetic resonance imaging scans, as well as real-time X-ray image denoising to enable low doses in angiography or image guided surgery. Since the consequences of wrong results often are literally a matter of life and death, the trustworthiness of these AI algorithms is very crucial.
One major long term challenge the break out session identified is the availability of fully annotated public datasets for image segmentation. There is always the issue of patient data security which puts restrictions on the use of the data in AI research and applications. Often explicit patient consent has to be asked for to gain access to the data. For some use cases like X-ray image denoising, there simply is no ground truth available because that would mean irradiating patients with different doses simply to get the same picture with different noise levels. That is neither medically nor ethically acceptable.
For data sets with annotations, there is often the problem of quality. Experts do not always agree on a single segmentation of a tumour for example, so either several experts have to vote on a given segmentation or the annotated data set could suffer from bias depending on who did the annotation. A possible solution could be quality measure / curated data sets perhaps provided in a GAIA-X context.
An important discussion point was the general role of AI in medical applications. Should AI act as a stand-alone medical application or should it be restricted to the role of a decision support system for humans. So far, only systems which support the clinician are commercially available. This is also due to ethical and legal issues when using AI assistance. For low risk use cases like computing sports training advice from smart health watch data the hurdles are much lower than for robotic surgery systems actually operating on a patient. The final responsibility has to stay in the hands of the clinician or the user.
Another type of use case not typically considered when talking about medical image segmentation that came up during discussion is robot-assisted monitoring and interacting with a patient during recovery/rehabilitation or during daily routines. From an algorithmic point of view, these machine-assisted therapeutic procedures touch not only upon aspects of computer vision similar to autonomous driving but also on machine-human interaction where the intentions of the patient have to be predicted from their behaviour in order to prevent falls or other injuries.

AI and genomics: Building Precision Medicine using reliable AI

In the breakout session on AI and genomics, reliable AI techniques (especially machine learning and deep learning),and how they support bioinformatics in clinical diagnosis,were discussed and analysed. The integration of information on patients of a diverse heterogeneous nature (genomic, clinical, environmental, ...) was discussed quite extensively. The main conclusions reached regarding this topic was that new protocols and standards are needed for the collection of information that takes into account the polyhedral aspects of health and its diseases, and quality controls must be implemented to give a minimum reliability of the information collected. Importantly, in the current day and age, this information includes software as well, thus requiring AI and ML predictive approaches to adhere to some essential quality principles (e.g. DOME recommendations). Workshops on how FAIR principles introduced for scientific data can be extended to ML and AI are underway to fill this gap (e.g.RD alliance).

The importance of collecting information about negative results on clinical experimentation leading to avoid undesirable biases in AI systems was also highlighted. In addition, there exists an increasing need to extend the amount of information available in order to integrate the information effectively.

Regarding the methodologies and techniques to carry out the information processing, organised in multiple data repositories with a low information density, transfer learning technologies seem to be able to provide successful scenarios in multiple health application domains. These techniques together with a combination of data-driven and model-based approaches can be an adequate framework for the development of a successful paradigm of AI applied to health, especially in the field of genomic medicine.

Another fundamental aspect regarding information processing is the transfer of primary knowledge to secondary knowledge. In this case, the need to articulate and regulate transparent methods for sharing information was highlighted, as well as the interest in gradually implementing standards for the effective development of a uniform collaboration scenario.

The other main aspect that was discussed in the session was related to the characteristics that the various AI engines applied to the field of health genomics should have. In the first place, the simulation of biosystems related to disease and the processes affected by genomic information (from biomolecular processes to digital twins) require new data-driven methods and translation techniques. These methods and techniques should be able to connect various models at different scales (i.e. from specific biomolecular processes to complex disease pictures). In this case, methods known in other areas of AI, such as multi-agent systems, natural computing (specifically, membrane computing) and other correlation methods should be explored. Similarly, the combination of knowledge-based methods and data-driven models seems to be a promising approach.

Regarding prediction systems, and the application of machine learning techniques, the existence of black boxes can compromise the acceptance of these systems in a field of application as sensitive as that of health. A dichotomy is produced since, in the field of genomics and health research, systems with black boxes have proven their usefulness, while in the case of clinical practice, fully explainable AI systems are required for their application, as is the case with any other diagnostic tool used today. Ultimately, working in the field of genomics and health should tend to produce AI systems that explain the relationship between genotypes and phenotypes. This should be a leitmotif for AI systems to

be produced in this field of application. In any case, the balance between usefulness and explainability should be modulated depending on the task in which the AI engines are used. Last but not least, the existence

e of biases in AI systems applied to genomics and health should be explored. Some of these biases, due to the biological nature of genomic information, far from being undesirable are necessary. There is a need to study and control the introduction of biases in AI systems that occur in this application area.

AI in infodemics

The complex infodemic phenomenon concerns the overabundance of information, not necessarily reliable, circulating online and offline about an epidemic outbreak with a huge impact on public health. Therefore the participants of the breakout session “AI in Infodemics” discussed infodemics and the role of AI to assess the infodemic risk, with potential applications to public health. Within these discussions it has become apparent that there are 2 main challenges of AI in infodemics that need to be overcome, namely the questions how this phenomenon can be first detected and quantified and how AI analytics can be translated as an output to policy. To inform the policy and also to involve legislators to establish appropriate regulation a AI assisted risk monitoring and assessment is required. One of the most relevant results on this topic is the C19 Infodemic Observatory that heavily rely on AI techniques and science to make an assessment of the risk of infodemics in collaboration with the infodemic management team from World Health Organization (WHO). They have gathered a unique team of computer scientists, social scientists, and business leaders to gather massive datasets from the private sector in order to provide live and actionable insights for other researchers, policymakers, and the general public, while respecting privacy. The C19 Infodemic Observatory analysed more than 2.2 Billion Tweets and 200 Million URL which led them to the conclusion that only 58% of the content is produced by Humans while 42% is produced by Non-Humans like Social Bots. It has also become apparent that 71% of the content online comes from reliable sources while 29% of the content is not reliable. In addition to this it is also of utmost importance to recognise the relevance and interdependence with other phenomena, like spreading of the epidemics. Within the discussion of the breakout session the participants gathered some thoughts for follow-up discussions like the question wow to design a new way of communication to provide reliable information able to convey the same emotional payoff of an unreliable information/content? In this case AI plays an important role to scale the solutions. They have also worked out that it would be of great interest for the future research in infodemics to combine insights from interdisciplinary disciplines like Computational Social Science, Behavioural Neuroscience, Complexity Science to achieve better results and a broader insight. In addition, the topic of Trustworthy AI will play a major role in this area of research in the future by developing trustworthy AI, which will be the key to fight infodemics.

Trustworthy aspects for NLP

In patient intake and engagement, medical documentation, automatic report generation, EMR analysis and forecasting, NLP is widely used. This breakout session focussed on the most important trustworthy AI components, challenges, and solutions for these use cases.

The group started with a presentation of Aki Härmä from Philips Research, which highlighted some important aspects and questions regarding the use of Trustworthy AI/NLP in healthcare. In particular, they touched upon the topics of Explainable AI, Safety and Robustness, Respect for Privacy, Fairness, Equity, and Justice by Design, Accountability and Reproducibility by Design and Sustainability aspects of large scale utilisation of such systems. This kicked off a broad discussion amongst the workshop participants, in which they tried to define when AI/NLP systems are trustworthy, arriving at the following list of requirements. AI/NLP systems can be considered trustworthy if they are:

- Bias-free
- Explainable
- inclusive, NLP powered AI for healthcare should be available for all (not just
- particular language groups)
- securite/privacy-preserving (particularly regarding data sharing)
- support (user) sovereignty
- sustainable.

Furthermore, the following key observations were distilled from the breakout discussion:

1. NLP is an enabling technology, meaning that it is at the core of a wide range of interaction scenarios within a healthcare environment. It can support making AI solutions explainable, e.g. through dialogue.
2. Language carries a lot of (cultural) information, and as such there exist higher risks for bias. Bias is everywhere, and awareness is the first step towards mitigating this (using synthetic data may be one mitigation strategy). Additionally, due to its interactive nature, language/NLP offers an opportunity to remedy (bias) issues through interaction/conversations.
3. Ambiguity and context understanding are the most challenging aspects for NLP applications. Ultimately this may require the development of Artificial General Intelligence (AGI). Knowledge graphs can be a powerful tool in helping understanding the context, particularly if these are dynamic, updating knowledge could happen through interaction with users.
4. Proper data sharing solutions are needed in support for creating the next generation of Trustworthy NLP systems. These solutions should support the sharing of trained models while preserving privacy. Initiatives such as the European Data Spaces are examples of promising solutions for this. An alternative could be the generation of synthetic data to train models.
5. Sustainability aspects are currently an issue, e.g. in terms of energy required for training contemporary NLP models such as BERT. Particularly for scaling up this technology to wide-range application this needs to be solved. Precise definitions of tradeoffs between e.g. energy usage and utility of NLP/AI solutions will aid in this.

Federated learning approaches for the Healthcare sector

The participants of this breakout session discussed and analysed federated learning approaches to facilitate the analysis of health data stored across different stakeholders and/or borders. This should, for instance, avoid the transfer or exchange of data and ensure

increased security. An industrial use case, where federated learning and preprocessing is performed on devices/smartphones serves as a starting point to discuss the challenges of applying federated learning in real world applications.

One of the first points to keep in mind is that federated learning is a broad concept with algorithmic approaches heavily depending on the actual use case. During the discussion the participants identified two main types: use cases featuring a few big data silos (hospitals for example), which are always online and are endowed with sufficient computing power on the one hand and use cases where many small devices (like wearables or smart phones) are the data sources, but are only sometimes online and with very limited computing capacity. It is obvious that while both types are federated learning scenarios, their needs and requirements are vastly different and thus require different algorithmic solutions. A fundamental issue for all use cases is the legal question of data security and anonymity. How can these be guaranteed when using federated learning? The ensuing discussion found the need for a binding legal framework in which to conduct federated learning. This tied in with the problem, that a precise definition of “Federated learning” is necessary to make sure all stakeholders are aligned on objectives.

A high communication load is required for successful federated learning. This is difficult for low battery distributed devices with low connectivity, so there has to be a tradeoff between communication and computation. How to choose the best trade-off theoretically leads to a multi-objective optimization problem, which is hard to solve practically. Making use of sparsification and quantification strategies or a combination thereof could be a way to handle this challenge.

An additional point which has to be taken into account is the problem of devices dropping on and off due to individual user behaviour. This causes a data bias, because some devices deliver much more data than others, which in turn can skew the federated learning results. A solution could be the use of asynchronous training when building the federated learning models. Assuming that we have successfully trained such a model, then how to evaluate it, since the original data on which it was trained is no longer available? This is especially hard for use cases with many devices (wearables for example), where data is not stored for a longer period of time.

During the discussion the question arose, whether there are any centralised data platforms shared at EU level for federated learning and if not, is the EU planning to put in place such a platform. This illustrates another challenge: the availability of training data for federated learning research - quote: “I want to do federated learning, but I don’t know where to find data”. A solution to the data storage issue could be an European cloud with citizen empowerment for uploading personal health data. This is another aspect which could function as an actual, albeit challenging, use case for GAIA-X. It also shows the need for a central information site, where one can get the data needed, a measure of the data’s quality, as well as legal and economic structures like incentives around it.

A final challenge for every federated learning use case is the standardisation of the input data. Data acquired through different standards are usually not comparable. Therefore interoperability between data providers is a necessary prerequisite for the subsequent actual federated learning itself.

Explainable AI in Healthcare - Group A

Health data is particularly sensitive and solutions developed with the help of AI are often difficult to understand. The aim of this breakout session was therefore to dive into explainable AI to promote the acceptance of digital health solutions in society.
The discussion started with possible reasons that require explainability, with the concept of ethics being particularly emphasised by the participants, as explainability is an essential and important factor in strengthening human trust in non-human systems.
Furthermore, another important aspect related to the overarching topic of explainability is the generalisation of AI models to ensure the content transfer from one model to another. In the context of medical knowledge, it is also crucial to construct a common knowledge model between the technical specialist/engineers and the medical specialists like doctors or medical nurses as well as the right training in AI for these groups on how to improve reasoning via assistive diagnostic tools with the help of AI. Also, this presupposes that humans still have the possibility to intervene in the decision-making process of the AI tool at any time.
Another important aspect highlighted by the participants of the breakout sessions is the fact that when considering software as a medical device, the effectiveness and robustness of software as a medical device must be proven and ensured. Considering the current situation, it can be determined that the interpretability of data or processes is highly context-dependent. In addition, medical experts are accustomed to black-box systems (i.e. drugs whose underlying mechanism is unknown, clinicians' own solutions for certain measurements), which in turn is associated with risks. Therefore, there is also a great responsibility to ensure that these instruments are robust, explainable and trustworthy.
It was also found that explainability is often not that important for the end user (i.e. the patient, the clinician, the caregiver) as long as the end user benefits from it. In this context, the research has also shown that industry stakeholders are not necessarily interested in creating better (explainable) models and optimising the models.
In terms of the potential for standardising explainability, the relatively slow adoption and acceptance of AI models due to lack of trust can be addressed. There is also the problem that there are no definitions for many medical observations, so explainability can also be a problem in medicine (e.g. the form of aneurysms). For this reason, explainability should be documented, similar to a guideline, and workflows can be standardised (like unit tests). It is also important to establish and standardise tools for explainability (e.g. SHAP, LIME, saliency maps, graphical networks).

Explainable AI in Healthcare - Group B

In the beginning of the workshop, the participants acknowledged that there is no objective measure of interpretability. This also poses a difficulty within the evaluation of explanations of an AI-system.
The criteria which define the quality of explanation are highly context dependent. Relevant factors here are for instance the user-role and the purpose of the explanation. A medical professional might have very different information needs than an AI-Developer or a layman. Thus, a good explanation does not only depend on the AI-System but also takes the information needs of the user into account. Another point which should be taken into account is the domain of application. For many domains, an extensive amount of human knowledge

is already available. Ideally, an AI system should make usage of this knowledge and integrate it in its decisions but also explanations.
The participants also discussed why explanations are particularly important when it comes to AI systems in healthcare. Medical professionals are used to blackbox systems (i.e. drugs for which the underlying mechanism is not known. Explainability is often not very relevant for the end user. However, for AI-Systems in healthcare interpretability plays an important role in building trust of the user. Especially for decision support systems, trustworthiness is essential. If a user trusts the system entirely, they can be over dependent and run the risk not to detect mistakes and risks within the suggestions of the AI-system. If the user does not trust the system at all, it will not aid the decision at all. Therefore, the user should be enabled to understand the decisions of the system and decide when to trust a suggestion and when to decide differently.
An important point which was mentioned as well is that the explainable model should ideally fit the cognitive model of the user. It was argued that users would expect causal relationships. Causal relationships which fit the cognitive model of the user are generally not inherent to AI-systems, therefore extracting those casualties from abstract models becomes an additional challenge. Furthermore, attention was raised to the factor of interactivity. It was argued that the user's understanding of the AI-system would be improved and they would be able to question the system's decisions in different ways. Examples that were mentioned include contrastive explanations, but also the usage of natural language, and conversational systems.
Overall, the participants identified a gap between currently available explainability models and the human cognitive models. One of the future challenges lying ahead is to develop methods and ways to close that gap. This will be essential for the development of trustworthy systems and accountability.

AI expertise in the Healthcare Sector

The healthcare sector nowadays faces several challenges in attracting talents and empowering their employees to provide AI-based solutions. Therefore the breakout session “AI expertise in the Healthcare Sector” aimed to address and to answer what the specific needs for AI training and upskilling programmes are and how these needs can be aligned with academic activities and doctoral programmes.
The participants of the breakout session started the discussion from the need to include a mandatory course of studies of AI in medical schools, but they all agreed that this is not enough and they o proposed to include a computer science and AI course in elementary, middle and high schools. This would create a good and solid base of AI, will make it easier for a University student of medical school to choose a course on AI for healthcare, and will provide all citizens with the necessary skills set to actively participate in the new AI society.
The healthcare sector faces several challenges in attracting talents and empowering their employees to provide AI-based solutions. What are the specific needs for AI training and upskilling programmes, and how can these needs be aligned with academic activities and doctoral programmes?
The proposal of the participants was to create a multi-disciplinary group, where the medical doctors, nurses, data and AI specialists participate in workshops on a regular basis to know and understand each other’s problems (Multi-disciplinary Workshop).
However, not only doctors or nursing staff such as nurses but also AI researchers in the

healthcare domain need to be trained on domain expertise such as healthcare ones and new jobs should be created in the hospitals for AI experts (Train AI specialists in healthcare).
A key point to be considered in AI & health is “Trust”: trust is essential for physicians to use AI systems. Medical doctors are interested in understanding the reasoning behind the model. AI explicability becomes a key element in the adoption of AI by medical doctors.
In addition the group proposed the idea of a European AI centre that creates a network with University hospitals and is able to act quickly when there is a health crisis: during the pandemic, the local hospitals got in touch with local universities and research centres to develop solutions to tackle the COVID-19. Most of these solutions were developed locally using only the data collected locally at the hospital. If there had been an existing network of AI experts working on AI & healthcare, it would have probably been faster to give support to the medical doctors in tackling the pandemic.

HPC-AI convergence and the Healthcare Sector

The ability of artificial intelligence techniques to analyse data accurately is growing at a breakneck pace. Among these techniques, Deep Learning (DL) has benefited from crucial results in machine learning theory and the large availability of data to extract useful knowledge from them. The accuracy of this process is closely related to the quality and quantity of the data and the computing power needed to digest the data. Therefore High-Performance Computing (HPC) is an AI-enabling platform. On the other hand, supercomputers are shifting to GPUs due to their improved power efficiency and the need for increasingly GPU-enabled workloads, such as DL. Healthcare is one of the critical sectors of the global economy, as any improvement in health systems impacts the well-being of society. European public health systems are generating large datasets of biomedical data, especially images that create an extensive database of untapped knowledge, as most of its value comes from expert interpretations. Nowadays, this process is still done manually in most cases. In the field of healthcare processes combining traditionally separate environments and HPC, Big Data analytics and Artificial Intelligence (AI) can overcome current problems and promote innovative solutions, in a clear path towards more efficient healthcare, for the benefit of people and the public budgets. High-Performance Computing (HPC) can propel AI applications toward grand challenges in healthcare: genomics, drug design, and diagnostics. Despite their potential, supercomputers are rarely used for AI. They are not yet equipped to effectively support specific AI software tools or securely acquire large amounts of data as medical applications require. Furthermore, artificial intelligence researchers are not used to the batch execution model of supercomputers
Within this breakout session, there has been brainstorming on the open problems of HPC-AI convergence, such as programming and execution models, accuracy, reproducibility, portability. Based on these discussions, it was evident that High-Performance Computing is not yet perceived as an urgent need in the medical AI community, especially in Small and Mid-sized Enterprises (SMEs).
On the contrary, on-demand services offered by cloud services are increasingly used in SMEs of the Healthcare Sector. They are not having big/urgent enough problems, and cloud computing offers (from over-the-top providers) that come with practical (and easy to use) tools for exploiting AI solutions. HPC is, therefore, mainly used in the academic sphere. The discussions have also shown that multiple (modern) languages and interoperable data formats are needed for data analysis: Python toolboxes (TensorFlow, PyTorch ...), Julia, etc..

AI and bioinformatics: integrating learning and biomedical knowledge

Developing learning models that are aware of and consistent with biomedical concepts and knowledge is a key research challenge to enable a widespread adoption of AI-based solutions in the life-sciences. Two key advantages have been identified throughout the discussion in the breakout session. First, learning-reasoning integration would strengthen trust of the life science community towards the use of data-driven methods and enhance self-explicability of the models, e.g. by having the model provide interpretations rooted on well-understood biomedical concepts. Second, learning-reasoning integration seems to be fundamental to surpass limitations of purely data-driven methods, such as machine learning and deep learning models, in unfavourable conditions such as data scarcity. In this respect, the breakout session has identified rare diseases as a relevant challenge which can highly benefit from an integrated approach capable of fusing symbolic knowledge, available under the form of knowledge graphs and interactomes, with high-dimensional/small-sample-size data. Rare diseases also require an integrative approach at the level of expert collaboration: life science and AI communities need to unite efforts and a pan-European collaboration needs to be sought and promoted. Rare disease patient associations can effectively support advocacy initiatives in this sense. The workgroup warmly advises to pursue and establish a collaboration between the CLAIRE-TAILOR-VISION community and the ELIXIR initiative, in particular as pertains to the machine learning focus group.
On a methodological level, the breakout session identified the research field of learning from complex data structures as a key enabler to effectively pursue the integration. On the one hand, learning models for graphs allow us to parse relational information in knowledge graphs and bio-networks, transforming such symbolic knowledge into numerical embeddings that can be effectively processed and incorporated by data-driven methods. At the same time, these models need to be extended to better integrate the temporal dimension and temporal evolution of diseases, to make them actionable and effective on biomedical data. Additionally, it is also advised to carefully investigate and consider the role of bias in knowledge representation, and how this can affect black-box systems that integrate such knowledge.

Input for the roadmap

Based on the results summarised in the previous section, the Organising Committee identified several topics which could be a valuable input to a European AI research and innovation roadmap. These topics will be presented to and further discussed with experts from TAILOR, VISION and CLAIRE in order to enrich the respective roadmap activities.

The below topics are the ones that stood out most prominently and will thus provide the ‘core’ of the input. However, when the roadmaps will be constructed, all inputs from the Theme Development Workshop will be considered.

Healthcare sector specific

- Ownership of Health Data needs (global/EU) privacy by design governance guidance for all involved multi-stakeholders based on the new data economy principle.
- Availability of public data sets difficult due to patient data security

- Developing trustworthy AI tools is key to fight infodemics
- Legal question of data security and anonymity → how to guarantee these when using federated learning? → legal framework required
- Explainability is often not very relevant for the end user (i.e. patient, clinician, care professionals), as long as end user benefits from it
- Industrial stakeholders are not necessarily interested in creating better (more explainable) models and optimising the models
- Need advocacy on public healthcare systems and organisations to create funding opportunities and to support clinical data collection and sharing.

Not specific to the healthcare sector

- Trustworthy & Explainable AI are closely linked – more precise definitions needed for both
- AI competency courses in middle & high school to make it easier for students to take AI related courses at university – increasing the number of people with expert AI knowledge significantly
- Incentives for the participation in standards development needed (Especially for academics and SMEs)

Theme Development Workshop “AI for Future Manufacturing”

About the workshop

The Joint Theme Development Workshop (TDW) co-organised by CLAIRE, TAILOR and VISION1 on “AI for Future Manufacturing” took place on the 10th of May 2022 with the aim to develop and identify the most promising and emerging AI topics in the manufacturing sector. At this one-day workshop, experts from academia, industry and politics jointly developed initial input for the European Artificial Intelligence (AI) research and innovation roadmap. Inspired by introductory speeches and presentations from selected experts, the participants actively discussed a wide variety of topics during the breakout sessions and shared their main results in the subsequent plenary presentations. Furthermore, some initial ideas for follow-up activities and further collaborations have been identified. The full report is publicly available on the TAILOR website[38]

Facts and figures

The high international interest that was expressed in response to the announcement of the AI for Future Manufacturing Theme Development Workshop translated into excellent attendance of the event.

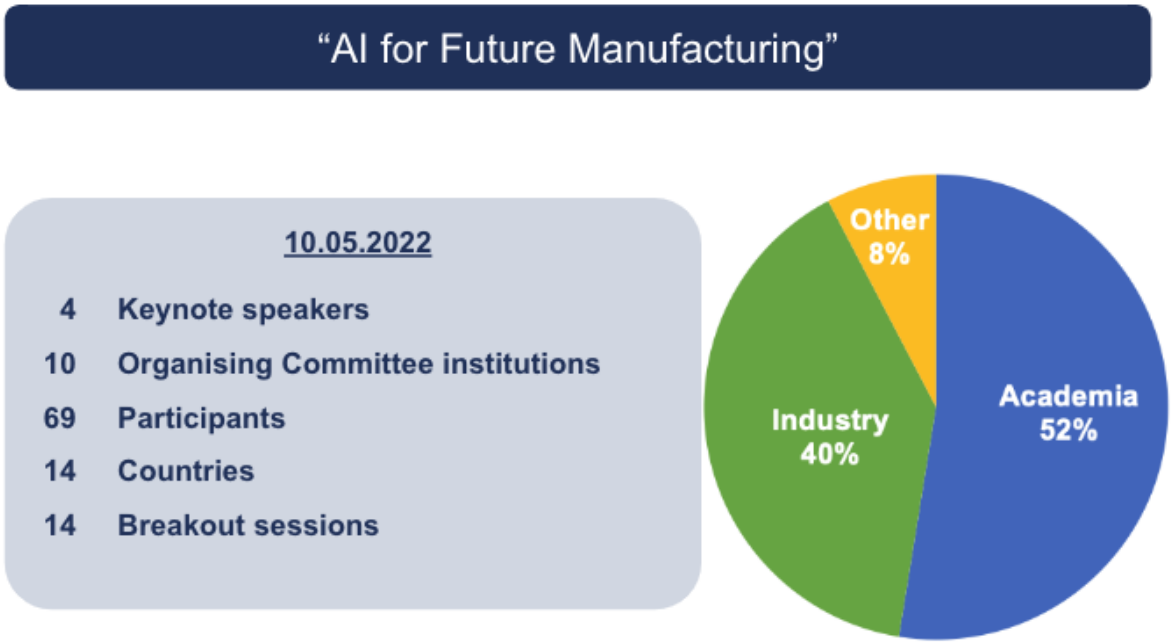

Figure 11: Flyer of the TDW “AI for Future Manufacturing”

Seventy-seven participants joined the TDW, ranging from a diverse set of backgrounds. nineteen (predominantly EU) countries were represented, with thirty-three participants indicating that they are affiliated with industry, whilst forty-one participants indicated that they

[38] See project website: https://tailor-network.eu/events/theme-development-workshop-ai-for-manufacturing/

are affiliated with academia (three participants indicated “other”). The participation of major industry representatives, with companies like ZF Group, ABB, Airbus and ArcelorMittal is particularly noteworthy and testifies to great interest on the part of industry. Equally important being the participation of those affiliated with the European Commission. The TDW, therefore, caught the attention of some of the most important actors in the field of Future Manufacturing and brought together representatives from key companies, supra-national institutions, and academia. The workshop thus successfully provided a platform for discussions between representatives from academia, industry and politics: Discussions that are key in unlocking the full potential of AI in Europe.

Program

The TDW was opened by the Co-Chairs Tilman Becker (CIIRC) and André Meyer-Vitali (DFKI) on behalf of the Organising Committee (OC), which included further representatives from ABB, ArcelorMittal, CLAIRE, CIIRC, DFKI, EIT Manufacturing, NTT DATA, RICAIP, TNO, University of Málaga and ZF Friedrichshafen AG. The Co-Chairs outlined the objectives of the TDW as well as the agenda and programme, and introduced the invited keynote speakers to the participants.

Figure 12: Flyer of the TDW “Ai for Future Manufacturing”

The inspiring keynotes were provided by the high-level experts Prof. Dr. rer. nat. Dr. h.c. mult. Wolfgang Wahlster, Dr. Georg Schneider, Prof. Dr. Vladimír Mařík and Sanjit Shewale.

Prof. Dr. Wolfgang Wahlster, German Research Center for Artificial Intelligence (DFKI), gave his introductory keynote on the topic of industrial AI across the next decade of industry 4.0. He started by giving an overview of the idea of industry 4.0 and how it encompasses the challenge of injecting AI into manufacturing processes to retain the competitiveness of the European economy. Professor Wahlster concluded his keynote by revealing that deep learning is not enough for the next generation of Industry 4.0 systems and that hybrid AI systems are needed.

Within the second keynote of the workshop day **Dr. Georg Schneider, AI Tech Center at ZF Group,** shared some insights on the benefits of Trustworthy AI in manufacturing. He started his keynote by highlighting the problem of explainability in relation to the decisions that AI systems may take, as experts may lack insight into how a given AI system has arrived at a certain conclusion. He then discussed possible solutions to this issue by discussing an example related to anomaly detection. Dr. Schneider concluded his keynote by highlighting four key benefits of explainability. The first being that it allows for the verification and validation of AI models, The second benefit being that it improves AI models and labelled data bases. The third being that it enables great insight into production systems and the root causes of problems. Lastly, explainability allows for an increase of trust in AI and the acceptance of it as a solution to various problems.

Prof. Dr. Vladimír Mařík, Czech Institute of Informatics, Robotics and Cybernetics (CIIRC CTU), started his keynote by giving an overview of the Czech Institute of Informatics, Robotics and Cybernetics CTU in Prague (CIIRC). Specifically, he discussed the RICAIP project, which is currently the largest EU project in the field of AI for applications in Industry 4.0. Next, Professor Mařík discussed the role of AI and machine 5 learning in the first decade of industry 4.0. Professor Mařík concluded that in the second phase of Industry 4.0, we will need global, flexible, and self-learning engineering solutions to be able to control and manage complex and intelligent manufacturing systems and their environments.

Sanjit Shewale, Global Head of Digital, Process Industries Division at ABB, started the fourth keynote by highlighting the importance of explainable AI in the context of AI research's applicability in industry. Standardisation and interoperability are a must to enable explainability. As the industry is still struggling to see value in AI due to failed pilot projects, explainability is key to increase trust. Furthermore, Sanjit Shewale argued that there are no AI-powered operations without quality data. Many projects fail or get extended due to a lack of quality data. Lastly, he gave an overview of the Explainable AI Consortium, whose goal it is to increase trust and the reliability of AI solutions.

These introductory presentations served as a basis for the discussions about the value of data and the aspect of Trustworthiness in the manufacturing sector, and provided some interesting examples of application areas. Accordingly, these presentations stimulated the expert discussions in the following breakout sessions.

Key results from the breakout sessions

The following subsections will provide a detailed explanation of the results of the breakout sessions of the Theme Development Workshop "AI for Future Manufacturing".

Trustworthy AI in Future Manufacturing

The aim of this session was to initially define the strategic challenges and derive relevant major topics and industrial applications where it is critical that the Artificial Intelligence (AI) is trustworthy. The session had the luxury of having two experts giving brief presentations with their perspectives to kick-start the discussions. First Juan Carlos Nieves gave an introduction to the technical and social aspects of trustworthiness and related this to the Ethics Guidelines on Trustworthy AI introduced by the High-Level Expert Group on AI appointed by the European Commission. Then Lena Weirauch spoke about her experience from the aviation industry prior to co-founding ai-omatic. She presented a scalable solution for automated condition monitoring where an AI finds anomalies in machine data without historical data needed and transfers the expert knowledge of an engineer into an AI to create a digital maintenance assistant to make the maintenance process more efficient.
The discussion during this breakout session was centred around 3 topics: 1. Key challenges with Trustworthy AI in Manufacturing; 2. Typical industrial applications that require trustworthy AI, and finally 3. Key recommendations for the future.
Perhaps the most important challenge discussed was to what extent an industry can give guarantees on AI trustworthiness of its products. A very related challenge is how to verify that a solution is trustworthy, and who takes responsibility during deployment: supplier(s) or customer?
The discussion around industrial conclusions was kept at a very general level without going into specific industries. It was concluded that Trustworthy AI is essential for anything related to safety, that can lead to unplanned downtimes, or if there is potential environmental impact. Regarding this point, Lena from ai-omatic mentioned that their solution of ai-omatic contributes to the topic of trustworthy AI and explainable AI in the sense that the user sees which sensor is responsible for an increasing or decreasing health value. Another example mentioned is that errors in AI for logistics value chain may lead to economic loss.
The breakout session resulted in a number of key recommendations: It is critical to require trustworthiness for analytics on all different data-types (images, time series, events, textual, etc.). We need to work on how to carry over trust within traditional industrial manufacturing ecosystems from the “analog” business world into the digital & AI business relationships. Fundamental research is needed between academy and industry into methods and their application to provide guarantees about AI systems. Furthermore there is a need for research into the different dimensions of trust, for example, robustness with respect to changing work conditions, interpretability as a means for true human-machine collaboration), and verification that the AI fits the intended purpose. Last but not least, we need to identify good practices using the Trustworthy AI principles and include this in our teaching.

Federated Learning for Future Manufacturing

This session addressed the challenge of training AI models without giving up data sovereignty and the question of what approaches to share models instead of data. Federated Learning enables Privacy by Design by training partial data sets separately and combining only the resulting models without sharing the data. However, privacy may be less relevant in manufacturing. Rather than privacy, the confidentiality of corporate data (e.g., manufacturing processes) may be more relevant. However, the privacy of employees and consumers can be affected. Therefore, regulation is important to guarantee the legal

certainty of people, processes, things, and to secure intellectual property. Data sharing of "near" original data needs to be encouraged.
By processing training sets in parallel Federated Learning could be more efficient and scalable for large-scale manufacturing ecosystems. Sequential processing is required due to interdependence of data in some cases, however. Small data sets can increase the level of bias and training on small data sets is more difficult. In addition, the costs for communication need to be considered, as well as the limited processing power of low-power edge devices. A challenge is to decide how to split up the data and which parts of the models to share with which parties. Agreements need to be reached about the unified structure of neural network models, because they cannot be (easily) merged otherwise. Compression of weights in the model (quantisation) can reduce the size of data to be communicated. Sharing of semantic models is another possibility to achieve common architectures.

Human-Robot and Human-Machine Interaction in Manufacturing Processes

The aim of this breakout session was to identify ways to realise partly automated processes or to stabilise highly automated processes through smart dialogue systems that incorporate and combine a plethora of different interaction methods in Human-Robot or Humane-Machine interaction.
Rather than fully automate manufacturing, Human-Robot Collaboration (HRC) bears the promise of including humans in the production process while easing burdensome tasks and emphasising human strengths in flexibility, experience and understanding.
Besides the physical aspects of safely coordinating humans and robots working in the same space, there are challenges in communication from robots to humans and vice versa and the integration of humans and robots and their collaboration into the larger manufacturing processes.
The physical aspects of HRC encompass safety issues and pose research challenges in a wide range of sensor technology needed to understand where the actors (humans and robots) are and what they are doing and what they intend to do next.
HRC is about communication with all applicable modalities, including spoken dialogue, gestures and classical IT communication channels like displays, touch-screens, and also wearables and mobile devices, like smart watches and tablets.
The tasks for teams of one or more human workers and one or more robots must be integrated into the overall manufacturing process. Tasks from production planning must be communicated to the human-robot teams, results from the team's work must be entered back into the planning and MES systems.
HRC can support many aspects of human involvement in production. It can enable workers to do more, including physical support for workers with disabilities or caused by age or (temporary) medical conditions. HRC can support demanding, stressful and repetitive tasks. HRC can also enable human workers to increase their expertise by using the communication aspects to provide assistance to workers about next steps in new or not well-trained production processes. HRC raises challenges in data privacy and ethical aspects as data about human workers is collected and its use bears potential for misuse.

Optimising assembly component

This session was about simulation approaches to find the best combination of groups of parts before the assembly. For the particular case of assembling, if the total number of possible decisions is limited, exhaustive search might be an affordable option. Otherwise, application of some optimization techniques can help. For this aim, a good model is needed. The model development needs a lot of data with good quality. Here, AI can help to train a model from data without the need for background information. In this respect reinforcement learning (RL) can help to solve an optimization problem in the absence of a model, i.e. via learning from experience.
Furthermore, within the breakout session, some keywords on ML/AI-based optimisation approaches were discussed, such as surrogate-assisted optimisation and variants of efficient global optimisation. Surrogates are (non-linear) regression models that are trained on the data points using cross-validation and hyperparameter optimization. The best model and the confidence measures of the model are then available for optimising the model, understanding the structure of the solutions, extracting knowledge, etc. However, it is worth mentioning that, if (as mentioned above) the dimensionality is low and the quality assessment is fast, the grid search may be quite sufficient.
Instead of building the model, since your objective function is cheap to evaluate, you could also optimise it directly (e.g. with an evolutionary strategy like CMA-ES or a derandomised strategy or differential evolution, to name a few).
The discussions have shown important trends in optimization using AI/ML: warm-start, meta-learning, switching between optimizers, surrogate-assisted, and explainable AI to understand characteristics of good solutions. The application domains in this context are simulation-based product engineering, product assembly and production process optimization.
In addition to this, some important key challenges in the field optimising assembly components were identified like Time-consuming objective functions (e.g., passenger safety in automotive, with 24h for a simulation being normal, or CFD) as well as handling visual simulation output (flow fields, for example) and extracting that information to make it usable by the optimizer. It also became apparent that optimisers need to be capable of learning from problem instances they have seen over time. In addition, they must be able to deal with very time-consuming function evaluations, as well as with multiple objectives and constraints, even for very time-consuming function evaluations. Another key challenge with regard to the overall topic of optimising assembly components is formulating the problem in the right way (what are objectives, constraints, user preferences, interactions between different disciplines, and even working processes within a company, i.e., which departments need to coordinate what they do in terms of design variables etc.).
During the breakout session, the participants identified key recommendations regarding the overarching topic of the session, such as research on the use of ML to pick the right optimizer and its configuration as well as integrating optimization and transfer learning, meta-learning, to make optimizers learn from problem instances. In addition to that, there is also a need for research on the issue of extracting problem knowledge using explainable AI from optimization runs, across both singular or multiple runs.

Retrofitting of AI

Currently, we are attempting to bring a number of AI based decision making into the production processes as well as machines. But how will AI react in the future when new sensors are added, new machinery is brought in, or if new software upgrades are installed on existing systems? During this session, these issues were addressed with the question of whether it will be necessary to upgrade or update the system.
The session started with a discussion of the definition of what it means to retrofit AI. The experts discussed the definition differentiating between (1) no (pre) AI to AI and/or (2) from existing to new AI and also (3) adapting existing AI to other changes in the manufacturing system. Furthermore, there are aspects of retrofitting AI deployment at macro-levels from an ecosystem to an organisation level and a micro-level from machines, production process to the plant level. The experts in the session emphasised that SMEs are lagging in digital transformation at the macro-level. The experts recommended a gradual transition to AI-driven manufacturing systems for SMEs, through collaboration with suppliers and clusters. The retrofitting of AI has more relevance for SMEs as they often cannot opt for embedding AI into newer generations of expensive machines. In this context, equipment with older design can be enhanced by applying AI methods to recognize the background reasons of unsolved design issues. The industry understanding of retrofitting and reuse of AI is nascent. The session was an example of continuous learning to get a practical perspective on the concept. It is important to continuously exchange industry-academia knowledge for effective implementation of the solution. An expert highlighted the importance of research perspective and technology advancement in transfer learning. The current academic research is looking at the end-to-end-supply chain (lifecycle).
Looking ahead, the session pointed out a few challenges concerning the implementation of and retrofitting of AI. The importance of privacy and cybersecurity are concerns that are valid for implementation and data engineering solutions were highlighted. The environmental impact of computing is highly relevant and could be an incentive for wide-scale usage of existing AI (2) and direct adoption (3) to new systems.
The experts agreed on the need for ecosystems and workforce training as the most relevant trends in the current times. The session concluded on a philosophical note on how retrofitting and ‘doing more with less’ will help to solve the societal challenge for business to be profitable yet sustainable especially for SMEs that need to compete in the next industry 5.0 decade.

AI in product development

This breakout session dealt with the use of data for development and construction of products by adjusting materials and measurements according to the need and best results from the manufacturing processes, to enhance the design of products. The use of data and AI for development or construction of products ist still an open field that has to be addressed. Adjusting materials, design or measurements according to the need of usability, duration and efficiency of products is very important. This can influence either the manufacturing processes directly or adjust the product use case to the customer's needs and expectations. Throughout the workshop, certain issues were considered in particular. There was a mutual understanding that the constraints set for a product pose the first difficulties, be it a final product, assembly or just a machine part itself. On the engineering part, it is a challenging

task to connect each constraint to the impact it has as well as to include an in-depth understanding about the interactions between each constraint defining the product. An AI linking all these factors together would help to develop the product in a more efficient way with less iterations in the process itself.
A possible use case for this is the exploration of the design space through AI, combining the space with simulation tools as well as constraints and physical limitations.
Nevertheless, understanding of the problem to solve is essential. Experts will still have to challenge themselves incrementally until an appropriate result is generated and enhanced through AI. Also taking the risk of a given solution into account, AI could help to deliver some form of risk certainty management, including knowledge about the uncertainty of certain constraints. This derives from the fact that certain constraints are not questioned due to a prior use case or missing data showing the invalidity of the constraint in the development process. Therefore, the question of the certainty of a constraint should be asked, however, often it is not. What do you know about what you don't know? Mainly this affects interdependence on constraints which can also pose a great risk in the development process. A systematic approach for the consideration of constraints is needed, solved by an AI algorithm bringing all the things together to form the whole picture.
An evolving Digital Twin - evolved during the use of the product - can also enhance the knowledge about constraint, not just on the engineer side but also on the user side. If this could be wrapped into a service the data generated would enhance every product.
With increasing time the maschine increases data as well as interdependencies between different digital twins which can be connected. On the other hand, when it comes to digital twins on a process level, tracking the operation of the machine/product by different personnel could also influence the data generated through the different levels of knowledge as well as maybe a different use of the product itself. Including humans into simulations is always a great effort, but turning the integration into a challenge with a defined goal or scoring could enhance the willingness to work with a Digital Twin and AI included.

Smart Manufacturing for Space Applications

This breakout session focused on the challenges and opportunities for AI that arise in manufacturing in the context of Space exploration and Earth observation, which is a challenging but also cutting-edge domain. It consists of a stream of data that goes from design to manufacturing then into testing and, finally, operation and maintenance. The ultimate goal is to achieve a continuously-improving single end-to-end algorithmic process that integrates these four areas.
Manufacturing in/for Space presents particular challenges. Delays in sensors, monitoring and control mean that fully-autonomous predictive models are desirable. Ideally, the models will operate next to the sensors to avoid such delays, however, energy consumption during operation is a concern, thus energy-efficient AI methods would be required. The design and testing phases seem to benefit from physics-informed methods, as the manufacturing process often does not produce enough high-quality data. However, current physics-informed deep-learning models are limited in terms of the dimensionality of the problem. Nowadays, the combination of data-driven and model-based (physics-informed) methods are the most promising.
Another important challenge is the semi-automatic generation of requirements for guiding automatic design processes. Currently, the design process can be automatized if the precise

requirements are provided by human experts. However, identifying these requirements is a difficult and complex task that requires a diverse engineering knowledge combined with experimental validation. AI methods that could generate and refine requirements would significantly improve the design phase.
Cybersecurity concerns also limit the capabilities of the AI systems, their sensors and the data collected. In addition, the trustworthiness of systems deployed in a spacecraft or space station must be extremely high and must be verified independently from the vendors who provide the various parts of the system.
The industry is moving towards distributed manufacturing thanks to the advances in AI. However, we must take into account the bigger picture of how Industry 4.0 may lead to the delocalization of highly-technical manufacturing industries due to automation. The Space industry should also consider the ethical implications of semi- or fully-autonomous manufacturing systems both in Space and in Europe.

AI in Digital Factories

Given the background of the participants, one of the first challenges identified in this session was to continue working on bridging the gap between academia and industry (State-of-Art vs. Reality Check) with flexible multidisciplinary teams. To this end, defining the processes of knowledge transfer from academia to industry is one of the first steps to be defined. Other challenges identified were how to ensure interoperability given the lack of greater acceptance of standards, the lack of data and intelligence for building specific profiles, including availability, quality and reliability to build better models or aspects related to cybersecurity to prevent risks even more with open connections (such as data spaces). Among the recommendations discussed were those related to working together in multidisciplinary teams namely data analytics, data science, AI / MLOps engineers, cloud architects, business analysts and robotics including the figure of Joint labs with the industry for a better understanding to create teams that can carry out the implementations required by the industry. Also relevant was the discussion on the difference between the terms "understandable AI" and "explainable AI" given the interest from the industry to better understand the inner workings of the models beyond that they maintain accountability and bias-free. On cybersecurity issues, self-protection models and secure environments both at the plant level and along the supply chain drove the discussion towards dynamic risk management should be considered when implementing AI. The topic of Human-Technology collaboration was not ignored, hybrid intelligence is the key to successful AI systems. Importance of augmented intelligence, human augmentation, and better decision-making capability, without removing the human factor.
Finally, emerging technologies such as Digital Twin and multi-agent systems based on reinforcement learning were identified as a driver for moving towards more autonomous processes. Also the combination of Edge Computing and Edge AI together with high-performance communication technologies such as 5G will act as an enabler for industrial process improvements.

Industrial applications of Explainable AI

The explainability of the AI algorithms is a necessary criterion to be fulfilled prior to deployment in real industrial application. Explainability is a requirement creating trust for

further usage of the solution. Thus, the aim of this breakout session was to dive into explainable AI to promote the further scaling and usage of advanced solutions in industrial applications. The session opened up with a very interesting keynote where Dr. Rafia Inam- Senior Research Manager, AI Research at Ericsson shared the experience and knowledge from Ericsson: "*The role of AI & Explainable AI in Telecom industry*". Describing different aspects of building a trustworthy AI, she emphasised on the importance of the explainability as one of the cornerstones towards trustworthy AI. Following the valuable input from the opening presentation we discussed among the others how and when the requirement of interpretability should and could be intrinsic or post-hoc in different industrial applications and examples. Furthermore we discussed different trade-offs surrounding the explainability such as; complexity and robustness all together in relation to the context in which the system is deployed. The complexity and associated uncertainty vs. the risk of errors in an application requiring even better methods for explainability will push this boundary as they reduce our uncertainty about the system. We noticed that although interpretability and explainability are in most cases interchangeable, explainability is at a higher level, dealing with human interaction. While the interpretability is of great value for the developer the explainability is most valuable for the operator. In other words, explainability/ interpretability are a kind of personalization/ contextualization for different stakeholders.
Furthermore, we emphasised on the importance of the so called *Explainable AI by design* as a key to increase trust and most probable more reliable systems. Explainable AI is dealing with different challenges and solutions on design level as well as at deployment level. One of the biggest challenges to be considered is the risk level of the proposed solution associated with the specific application. Generating a concise, meaningful, and tailored explanation to different users and stakeholders is very important.
Finally we conclude that static explanations are often not enough - users and experts should be able to interact with the AI. Robust and model agnostic explanation methods are required to leverage the full breadth of available AI methods and models, to create trust or interpret the model. Explainable AI is integral not just to enable trust but also being compliant with regulations.

Zero Defect Manufacturing

This breakout session dealt with AI technology methods like data-based modelling to detect out-of-control (OOC) states in classical Statistical Process Control (SPC) from in-process data enabling continuous monitoring and earlier detection of problems with a simultaneous reduction or near elimination of expensive post-process quality inspection.
The most important key challenges within the topic of Zero Defect Manufacturing, which was identified by the participants of this session, are people-related issues as well as the Acceptability and Explainability of new AI solutions. Explainable AI is necessary to move to the next level of acceptance of advanced diagnostic systems in industry. What is crucial is that the people involved in the process have a clear understanding of why the system is recommending what it is recommending. Besides that it is essential to ensure the necessary acceptance of the workers. This can be achieved by a common language to connect the managers with the workers and also to connect individual engineer roles and tools. However, this can be covered by a culture accepted by the company, which focuses on hybrid production as the highest priorities.
When it comes to Zero Defect Production, the second key challenge is the shift from a local

to a global perspective by having a holistic approach to cover the whole product life cycle (from product design, through supply chains, shared data, process monitoring to final consumer) which requires a process structure for formal representation of the information/data.
Another key challenge identified by the participants of the breakout session is the ageing of models caused by changing processes/concept drift through modified structures as well as a lack of training data after a change and fast transfer. To tackle this challenge a long term solution of constantly self-updating/self-learning models based on the observator is needed. There is also a great potential of using simulation models to have at least some training data immediately after a change. Furthermore, the Zero Defect Manufacturing system must know about the changes, people can also enter this, but correctly, e.g a tool replacement must be inserted to a system correctly. The fourth key challenge regarding Zero Defect manufacturing is false alarms. People often get overwhelmed by false alarms, e.g. in space apps thousands of independent sensors and problem detectors can have a high false positive rate, which can lead to an ignorance of the quality control system. This can be solved by techniques to double check the generated alarm by the system via multivariate methods, dimensionality reduction and correction of alarm thresholds.

Psychological approach for data labelling

During this breakout session the participants discussed sustainable approaches in combining psychology with elements of labelling technologies to ensure high data quality (standardisation) which is needed to apply AI in manufacturing.
These sustainable approaches aim to combine psychology with elements of labelling technologies to ensure high data quality (standardisation) which is needed to apply AI in manufacturing. However, data often appears messy, because they are not labelled accurately due to the lack of motivation of human annotators. These operators should be involved more to understand the reasons for labelling the data and to minimise their fear of losing their job through automation of their know-how. Human-centred AI should include the labelling operators. Machine Learning is only a part of AI and should be combined with other approaches, which also applies to labelling. Operators can be supported by using better tools and heuristics and by understanding the context and purpose of their work. Feedback should be used to improve the tools and methods. They would also be helped by collaborating amongst each other to increase social interaction and better results through standardisation of their methods.At the same time, the work can be customised to individual operators depending on their backgrounds.

Reality independent AI

This breakout session addressed the question of how the potential of digital twins with synthetic data can be uncovered in image data (e.g. for AI quality inspection or object recognition) for manufacturing.
Synthetic data are algorithmically produced (synthesised) to mimic the characteristics of real (sensor) data. The process of creating synthetic data can range from as simple as adding entirely random (but sensible) data to the dataset, to changing some features of real data, to as complex as creating new data entirely from scratch. E.g. pictures of road surfaces in good condition and digitally adding surface cracks to create thousands of synthetic images of poor

road surfaces to train algorithms, or generating images of faces using gan technologies. Main applications of synthetic data are data anonymization (privacy aspects) and training machine learning algorithms from sparse or biassed datasets. In this breakout session a structured approach to generate training data synthetically and a use case for optical inspection in production of microchips was presented and discussed.
Deep neural networks become an integral part of online inspection. They learn to interpret deviations and cracks while distinguishing them still from intended deviations in special product configurations.
The availability of training data is the main problem of using such deep learning methods. E.g. in a production environment, most data will show undamaged parts while actual defects are rare. A solution to this is to simulate measurements based on scenes that are generated by parametric models or the real world. By investigating the parameter space of such models, training data can be generated in a controlled way.The method involves the creation of partial models by learning individual aspects of the product, such as geometry, surface properties and lightning conditions generated in cad/cam environments and shallow models respresting e.g. cracks. The partial models are called parametric, because each model is controlled by a set of input parameters and describes a part of the scene in a simulation environment as a function of these parameters. The parameter space of the parametric scenario is the union of the input parameters of all the partial models and the scenario-specific parameters. Such an approach also allows for controlled sampling of the parameter space. E.g. using a strategy that generates additional samples close to misclassified samples. For more details and background information see ‘Digital reality: a model-based approach to supervised learning from synthetic data’.

Democratising AI

The aim of this breakout session was to uncover who is already able to use AI in manufacturing and whether only large companies or also smaller companies benefit from the use of AI and if both parties also benefit from each other and exchange knowledge.
Above all there is the risk included in taking decisions on ground data. How can anyone make a decision based on data without knowledge and training? A lot of decision makers are sceptical about AI as well and still new to the topic and possibilities. Therefore it is essential to start slowly with simple application, with slow upscaling roadmaps showing value through pilots, with a scalable solution, as measurable as possible.
Including every person influenced by the solution will also help with the cultural acceptance in a company. Acceptance (explainability) and inclusion (controllability) of workers to take away the fear, including an demographic worker structure is essential for a successful implementation and further usage. Always thinking about the Questions: Who should start with AI? - if you have no idea about AI it may be risky and lead to errors. How do workers interact with AI, i.e. to what extent are they responsible for the AI enabled processes? A clearly defined road of inclusion and responsibility is the goal you should aim for. Delivering proper rules, like a guidance system defining the transformation in a very structured way as well as elaborating on how to establish this in a company processes is challenging. Therefore a management championship is needed to really bring it through, to make it a strategic decision and make everyone adapt to the change. Also collaboration with existing experts (e.g. include the Process engineer into the solution development) will also bring acceptance as the inclusion of workers, onboarding them as early as possible.Keep in mind

that we have a rising need for data scientists. We cannot have as many data scientists as we need, as this will not be economical in the future. Therefore it is essential to democratise AI in any company. This counts for any dedicated topic involved in the data science pipeline from labelling of data to MLOps. If we now look at the transition from big companies to small ones the missing standardisation causes errors. Taking over bigger models into smaller companies is therefore a challenging task. For once island solutions can be the right thing for bigger companies on a broad scale with different divisions choosing different solutions. This is not possible for SME as it has to be standardised along each department. But to really enhance circular economy and sustainability there will have to be a democratisation between global players and SME. This will most likely happen along the value-chain and should be fostered as much as possible.

Optimization and Machine Learning for Better Manufacturing Processes

This breakout session discussed what the obstacles and challenges for the implementation of advanced methods combining Optimisation and Machine Learning in manufacturing companies are. Optimization problems in industrial manufacturing companies are characterised by large complexity, huge volume of data, mixed-integer and combinatorial decision variables, where small changes in these variables may affect a whole supply chain. In addition, there is often the desire to solve problems in a very short time despite the fact that evaluating the quality of a single candidate solution is often expensive and may require the simulation of a process. Typically there is not a single optimal solution but a number of feasible solutions that trade-off multiple (often many) conflicting criteria from which an expert will choose the solution to be deployed. Classical optimization methods such as mixed-integer programming and metaheuristics are too slow in many cases and cannot compete with human experts who are already highly-skilled at identifying good solutions based on factors and knowledge that are not immediately available as potential input data for AI models. The quality of the data is itself an open challenge, since it is often noisy, missing and/or erroneous. It is not unusual to have at the same time too much data and too little data because only a limited amount of information is useful within a very large dataset. The breakout group identified the need for more research in the context of large, non-sparse and combinatorial problems, in particular, better surrogate models for such types of data. Cutting-edge methods such as Alpha-Zero require a large investment in terms of engineering effort, data collection and training time and it is unclear at the moment how well they will work in the context of manufacturing processes. More theoretical/basic research work is needed to understand their adaptability and limitations.
The explainability of complex models was highlighted as an ongoing concern. We discussed examples where clients or suppliers have requested to scale back the complexity of a model because it was not sufficiently explainable. In the context of optimization, explainability includes identifying which changes in the data lead to a change in the optimised solutions, as well as how the hyper-parameters of the methods influence the solutions generated. Finally, the group also identified a lack of collaborations between different companies to release data and try to solve common problems, even though it was also agreed that the applicability of academic research would improve if industry would make real data and simulation models publicly available for research. One proposal put forward to foster such collaborations was to create consortiums of non-competing companies from diverse industries.

Input for the roadmap

Based on the results summarised in the previous section, the Organising Committee identified several topics which could be a valuable input to a European AI research and innovation roadmap. These topics will be presented to and further discussed with experts from TAILOR, AI4Media, VISION and CLAIRE in order to enrich the respective roadmap activities.

The below topics are the ones that stood out most prominently and will thus provide the 'core' of the input. However, when the roadmaps will be constructed, all inputs from the Theme Development Workshop will be considered.

Manufacturing sector specific

- Fundamental research is needed between academy and industry into methods and their application to provide guarantees about AI systems. Furthermore there is a need for research into the different dimensions of trust, for example, robustness with respect to changing work conditions, interpretability as a means for true human-machine collaboration), and verification that the AI fits the intended purpose.
- By processing training sets in parallel Federated Learning could be more efficient and scalable for large-scale manufacturing ecosystems. Federated Learning also enables Privacy by Design by training partial data sets separately and combining only the resulting models without sharing the data.
- Human-robot collaboration (HRC) can help integrate humans into the production process without replacing them by relieving them of burdensome tasks and highlighting their strengths such as flexibility, experience and understanding.
- Products are in many cases constrained by restrictions, which also pose difficulties. In this case an evolving Digital Twin - evolved during the use of the product - can also enhance the knowledge about constraint, not just on the engineer side but also on the user side.
- The trustworthiness of systems deployed in a spacecraft or space station must be extremely high and must be verified independently from the vendors who provide the various parts of the system.
- When it comes to data labelling, sustainable approaches aim to combine psychology with elements of labelling technologies to ensure high data quality (standardisation) which is needed to apply AI in manufacturing. However, data often appears messy, because they are not labelled accurately due to the lack of motivation of human annotators. These operators should be involved more to understand the reasons for labelling the data.
- The availability of training data is the main problem of using deep learning methods. E.g. in a production environment, most data will show undamaged parts while actual defects are rare. A solution to this is to simulate measurements based on scenes that

are generated by parametric models or the real world. By investigating the parameter space of such models, training data can be generated in a controlled way.

More general topics not limited to the Manufacturing sector

- Explainable AI by design as a key to increase trust and most probable more reliable systems. Therefore, a meaningful, and tailored explanation to different users and stakeholders is very important.
- Users and experts should be able to interact with the AI. Robust and model agnostic explanation methods are required to leverage the full breadth of available AI methods and models, to create trust or interpret the model.
  Reinforcement learning (RL) can help to solve an optimization problem in the absence of a model, i.e. via learning from experience.
- AI could help to deliver some form of risk certainty management, including knowledge about the uncertainty of certain constraints.
- Joint labs with the industry should be established for a better understanding and to create teams that can carry out the implementations required by the industry.
- Acceptance (explainability) and inclusion (controllability) of workers to take away the fear, including an demographic worker structure is essential for a successful implementation and further usage.
- Dynamic risk management should be considered when implementing AI.
- The combination of Edge Computing and Edge AI together with high-performance communication technologies such as 5G will act as an enabler for industrial process improvements.

Theme Development Workshop “AI for Future Energy & Sustainability”

About the Workshop

The Joint Theme Development Workshop (TDW) co-organised by CLAIRE, TAILOR and VISION1 on “AI for Future Energy & Sustainability” took place on the 23rd of February 2023 with the aim to develop and identify the most promising and emerging AI topics in the Energy sector. At this one-day workshop, experts from academia, industry and politics jointly developed initial input for the European Artificial Intelligence (AI) research and innovation roadmap. The full report is publicly available on the TAILOR website[39].

Facts and figures

The high international interest that was expressed in response to the announcement of the AI for Future Energy & Sustainability Theme Development Workshop translated into excellent attendance of the event.

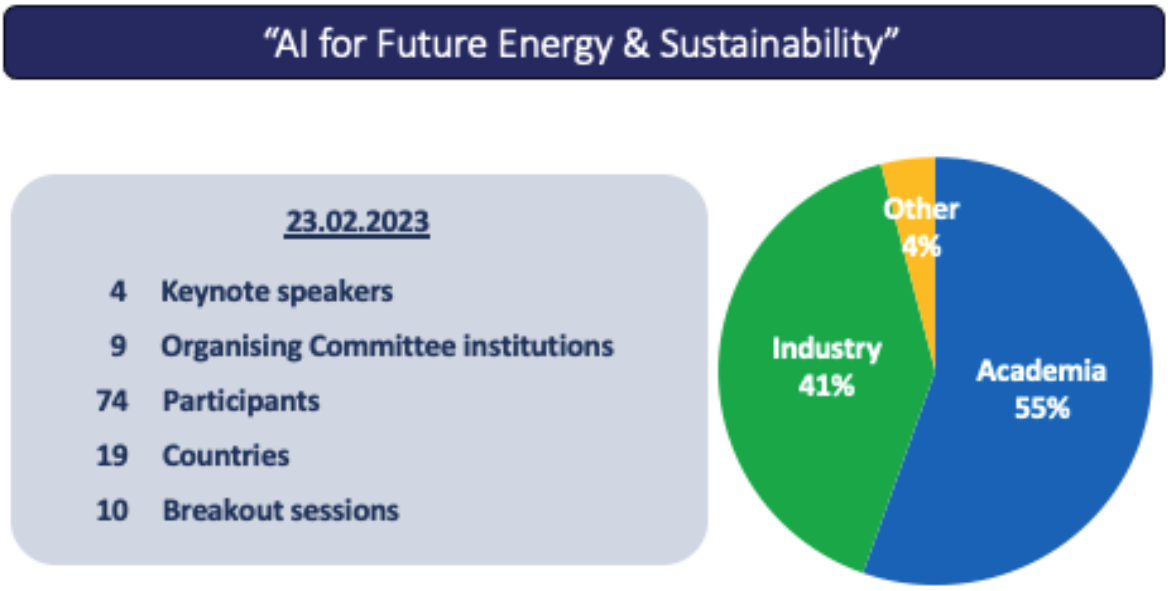

Figure 13: Flyer of the TDW “AI for Future Energy & Sustainability”

Forty-seven participants joined the TDW, ranging from a diverse set of backgrounds. Nineteen (predominantly EU) countries were represented, with thirty participants indicating that they are affiliated with industry, whilst forty-one participants indicated that they are affiliated with academia (three participants indicated “other”). The participation of major industry representatives, with companies like ABB, EDF, Huawei Deutschland Technologies GmbH, Stromnetz Hamburg GmbH and Tietoevry is particularly noteworthy and testifies to great interest on the part of industry. The TDW, therefore, caught the attention of some of the most important actors in the field of Future Energy and Sustainability and brought together

[39] See project website: https://tailor-network.eu/events/theme-development-workshop-ai-for-future-energy-sustainability/

representatives from key companies, supra-national institutions, and academia. The workshop thus successfully provided a platform for discussions between representatives from academia, industry and politics: Discussions that are key in unlocking the full potential of AI in Europe.

Program

The TDW was opened by the Co-Chairs Freek Bomhof (TNO) and Ghislain Agoua (EDF) on behalf of the Organising Committee (OC), which included further representatives from ABB, CLAIRE, German Research Centre for Artificial Intelligence (DFKI), EDF, NTT DATA, Tietoevry, Netherlands Organisation for Applied Scientific Research (TNO), University of Valencia (UPV) and Philips. The Co-Chairs outlined the objectives of the TDW as well as the agenda and programme, and introduced the invited keynote speakers to the participants.

The inspiring keynotes were provided by high-level experts from academia and industry, namely Sherif El-Meshad, Dr. Adamantios Marinakis, Aliene van der Veen and Claire Lambert**.**

Sherif El-Meshad, ABB Electrification, gave his introductory keynote on the role of Artificial Intelligence (AI) in achieving a sustainable future in Energy. Starting his keynote, he outlined the current state and future assessment of AI and emphasised its immense potential. Overall, Sherif El-Meshads keynote speech underscored the importance of balancing the potential benefits of AI with its environmental impact. By following sustainable development practices and using AI responsibly, we can create a more sustainable and equitable future.

Dr. Adamantios Marinakis, Research Center for Energy Networks (ETH Zurich), gave a keynote about the role of AI in the future energy systems. Dr. Marinakis started his keynote by examining our ongoing energy system highlighting the transition from fossil fuels towards more sustainable, renewable energy sources, which is no longer a matter of 'if' but 'when'. In the second part of the keynote, Dr. Marinakis highlighted the main challenges and opportunities that this energy transition presents. Dr. Adamantios Marinakis continued his presentation by addressing the significant changes in the energy sector and the respective challenges and opportunities that they present. Dr. Adamantios further outlined various applications of AI in the energy sector and identified four main challenges associated with the use of AI in the energy system: data privacy, data security, explainability or transparency, and accountability. Each of these challenges needs to be addressed to ensure that AI is used responsibly and effectively in the transition towards a sustainable and intelligent energy future.

In the keynote "Intelligent Agents in the Energy System and How to Make Them System Operators", **Aliene van der Veen, TNO,** set the stage with a classical definition of AI from the textbook "Artificial Intelligence, A Modern Approach" by Russels & Norvig (1995): *"AI is about designing and building intelligent agents that receive precepts from the environment and take actions that affect that environment."* On the basis of this definition, she posed the question, "Where can we expect such intelligent agents in the energy system?" Furthermore,

she emphasised the need for distributed system operation by AI, which was once viewed as a grand challenge in 2012 but is now considered a Social Grand Challenge in 2023.

Claire Lambert, EDF, gave her keynote on the challenges for reliable Artificial Intelligence (AI) in the energy industry. She presented EDF's goal to develop innovative solutions around the world to produce clean energy and guarantee access to electricity for all. She explained why AI is an important lever that will be activated to achieve this goal. Overall, Lambert's keynote emphasised that AI is a key tool that is helping and will help the energy sector build a net-zero energy future with electricity and innovative services. Its trustworthiness and ability to respect regulatory constraints will favour its widespread use.

Figure 14: Flyer of the TDW "AI for Future Energy & Sustainability"

Results from the breakout sessions

The following subsections will provide a detailed explanation of the results of the breakout sessions of the Theme Development Workshop "AI for Future Energy & Sustainability".

AI for Energy Efficiency

The focus of this session was on how to achieve improved energy efficiency using software solutions in general, and in particular the AI-enabled ones. The session began with an introductory presentation by the invited expert Juha Mäntysaari from ABB Process

Automation. Firstly, it was pointed out that there are standard indicators for energy performance, and then stated that it all starts with access to data. Furthermore, the importance of energy consumption anomaly detection and time-series forecasting was stressed. The capabilities of machine learning for time-series forecasting were illustrated by comparing multiple forecast methods for a real scenario.

It should be possible to obtain significant energy savings in many applications by applying machine learning solutions. For example, in any transportation system, be it on land, air or sea, there is room for energy optimization. In particular, with the current clear shift towards electrical vehicles and vessels. Similarly, in traditional heavy industries, such as mining & metals, pulp & paper, power utilities, and chemicals, there is also a shift towards electrification with large savings possibilities. Other applications mentioned were heating & ventilation, data centres and energy storage (battery as well as hydrogen-based).

In order to increase energy efficiency with machine learning, there are, however, several challenges, e.g. coming up with suitable business models giving the proper customer incentives since there are often conflicting objectives, such as, e.g., carbon neutrality and cost minimization. There is also a need to understand that industry deals with large and interconnected systems, which poses a challenge in itself. Other important challenges include data availability & quality, the need for explainable & accurate models and paying attention to the entire AI life cycle, including ML model quality monitoring.

The discussion led to the following key recommendations: It is important to combine machine learning and optimization, and to focus on business models and use co-innovation with multiple stakeholders involved (universities, automation suppliers and industry end-users), i.e., to have cross-domain collaboration. Moreover, it became clear from the initial presentation how important the energy consumption anomaly detection and time-series forecasting aspects are. Finally, active participation in the on-going legislation activities in Europe is important from both industry and academia.

AI for Collaborative Sustainable Buildings

This breakout session focussed on AI solutions for buildings, interactions and the formation of Human-AI-Ecosystems. In order to meet the climate sustainability goals, existing smart building systems need to be improved on the level of interaction and collaboration. Solutions go into deliberation between buildings and interactions with humans in order to reach collaborative goals, such as e.g. peak shaving and insights into various forms of flexibility.
In order to achieve a resilient grid infrastructure powered by renewable energy sources, the focus must shift from individual systems and isolated intelligence to fostering interaction and collaboration among them. This requires the integration of energy management and distributed smart buildings systems, as well as a mutually beneficial partnership between humans and smart energy systems.
Therefore, the challenges of energy transition and AI must be approached in a cross-disciplinary manner. While simple systems can sometimes suffice, the key lies in facilitating effective interaction between various systems. The overarching idea builds upon human-AI ecosystems where humans and machines work together in teams with a shared objective, characterised by collaboration and mutual support. The aim is to create an

environment where machines are not seen as mere tools, but rather as collaborators working towards a common goal, much like an ecosystem.
Within this breakout session the group identified several challenges. The first is the need for semantic dataspaces to facilitate data communication. On the technical AI level, there are self-optimization problems, and questions about decision-making, ranging from regression to supervised and reinforcement learning. In the field of human-machine interaction, researchers must consider human requirements and provide explainable solutions, including preference elicitation and aggregation. Privacy and conflicts of interest arise when private data is shared to enable AI-based decision-making, scheduling, and other data-driven solutions. Additionally, the agility of AI-based solutions must be considered.

The domain opens up for an interdisciplinary research discipline, containing topics that require contributions from individuals in various fields. There are similar initiatives happening in parallel, which should be tailored and discussed in follow-up activities, such as workshops, hackathons and show-case presentations. We should further encourage participation and refine the projects and participants.
The group discussion resulted in several recommendations. First, it's important to consider the objectives of various parties, including building occupants and building owners. Second, key domain challenges must be addressed, in addition to AI challenges. Third, rooms or buildings must be properly connected to the IoT ecosystem to enable trustworthy information sharing. Finally, explainable AI is crucial for increasing reliability and ensuring that the partnership between humans and algorithms remains on the right track.
Moreover, scalability is a crucial factor to consider. Currently, data streams still need to be analysed to link them to the AI model. There is a need to explore how AI can carry out this step without human intervention.

Energy at the Edge

The session about energy at the edge of the grid started with a short introduction by moderator André Meyer-Vitali (DFKI), followed by an inspiring presentation by expert Kim Bybjerg (Tata Communications). Bybjerg explained how the Internet of Things (IoT) lays the foundations for the buildings of the future with several important examples. The examples mentioned were two of the world’s most sustainable buildings: Edge in Amsterdam and Berlin’s 10-storey Cube. The buildings are equipped with multiple sensors that measure and collect vast amounts of behavioural and environmental data using the Internet of Things (IoT) and 5G wireless Internet. Using these data and anticipating future states allows for precisely controlling the buildings for optimal energy efficiency. It’s important to move from merely collecting data to understanding their value, e. g., data for climate management. Resources can be saved in parts of the buildings that are currently unused or unoccupied. This can be scaled from buildings to neighbourhoods and cities and also involves traffic management to avoid congestions.
An important concern is the privacy of collected data. Some data can be aggregated and anonymised and a lot of data can be processed locally. For example, image data can be processed for face recognition or motion detection and deleted afterwards. This is not always the current practice, such as for car driving behaviour data that is continuously sent to the manufacturer. Reducing the amount of transmitted data is also beneficial for the mobile networks, which are expensive to set up and maintain. Edge devices, however, such

as smart meters, are inexpensive and can be upgraded for up-to-date functions.
Regarding the value of data, it is essential to improve predictions, for example, for the charging schedule and duration of electric vehicles. Fair use of such prediction results should guarantee that all users are served sufficiently. The main issue is that efficient and fair scheduling cannot be performed on local data only. With only local optimisations the overall scheduling may become inefficient. Therefore, it is necessary to include global data - at the cost of increased complexity. The advantages of processing at the edge (and thereby saving communications and avoiding central control) need to be balanced by smart decentralised coordination mechanisms to exchange only the information that is necessary for all local nodes to achieve global optimisation. Smart AI systems can help dealing with uncertainties and negotiations among users. They need to be human-aware, i.e., understandable and controllable with personal preferences.

Explainable AI for Energy

This session dealt with the importance of Explainability in the use of AI in the energy sector. There are many critical industrial applications in this sector, both downstream (from production to delivery) and upstream (from delivery to services to customers). The acceptability of AI solutions for the management of these systems depends not only on their ability to be explained to decision-makers or industrial operators, but also on their ability to respect regulatory constraints (European AI law, audits by control authorities, privacy, etc.). For example, the black box effect of many tools/models (Deep Neural Networks) limits their acceptability and widespread use.
Explainable Artificial Intelligence (XAI) tools are crucial to convince decision makers of the developed AI technology or method. Currently available tools or packages for XAI models are unsatisfactory because they are too data-oriented, can require difficult mathematics and, for some, lack strong theoretical demonstrations/proofs. They also do not cover the whole chain: feature interpretability - model interpretability (either in-situ or post-hoc) and decision interpretability.
It is therefore important to be cautious about off-the-shelf tools for XAI, and to consider explainability aspects at the beginning of AI technology development and throughout the development process. The level of explainability required should be discussed between industry experts and algorithm developers to build a fit-for-purpose tool. This co-construction allows tools to benefit from domain-specific recommended practices. Explainability methods/frameworks that enable interoperability of data and features should be privileged. Verification and Validation of the algorithms/tools could follow a systems engineering approach: guarantee the behaviour of the models/tools even if a component is not fully explained/understood.

Data-driven Microgrids powered by renewable energy - Group A

This breakout session was centred around the concept of data-driven microgrids powered by renewable energy. The discussion highlighted the importance of balancing, storing, and flexibly utilising renewable energy within microgrid systems to optimise their performance.
Leveraging artificial intelligence (AI) and machine learning (ML) was deemed crucial in achieving the functional microgrid capabilities necessary for optimal system-level performance. In this regard, one aspect that needs to be considered is the Weather

Prediction and Energy Storage. The participants emphasised the need for more localised and accurate models to predict weather conditions specific to microgrid areas. Additionally, understanding energy storage mechanisms within the grid was considered essential for efficient utilisation of renewable energy resources, including inertia and battery systems.
Followed by this, the breakout session delved into the importance of robust data governance frameworks and consent handling protocols for data collection within microgrids. Addressing data privacy concerns while enabling effective steering of assets was a crucial aspect of data-driven microgrid operations.
Furthermore, there is also a need for appropriate data legislation, specifically discussing Automatic Meter Reading (AMR) information. Establishing legal frameworks for data sharing and utilisation within microgrids was a key consideration.
Beyond that, detecting abnormalities and identifying potential cyber-attacks were identified as areas where AI-powered solutions could enhance the security and resilience of microgrid systems. Simulations, including simulated attacks on communication lines, were also discussed as a part of this topic.
AI and Machine Learning were seen as valuable tools when it comes to smarter management of various energy resources, such as hydro, solar, and wind. Optimising energy production and consumption based on environmental conditions could significantly enhance the efficiency and sustainability of microgrid operations. Moreover, AI-driven solutions were highlighted as potential enablers of enhanced resilience against extreme weather events. By leveraging real-time data and predictive algorithms, microgrids can adapt and respond to adverse weather conditions, ensuring reliable energy supply during challenging circumstances.
The session also touched upon the legislative aspects of microgrid operations, emphasising the need for appropriate frameworks governing microgrids and enabling seamless integration with larger energy infrastructure. APIs (Application Programming Interfaces) for effective communication between microgrids and external organisations were also discussed.

Data-driven Microgrids powered by renewable energy - Group B

Microgrids with a primary energy source from their own renewable production require balancing, storing and flexibility of usage to work at an optimal level. The balancing needs an asset level forecasting and steering of all connected assets. AI and ML will play a crucial role in achieving a functional microgrid capable of providing optimal microgrid system level performance.
There are multiple challenges in this area that have been identified by the participants like the proper modelling of the aggregation level. In larger grids, the variability of behaviour is large (buffers, industries, shopping malls) and this is harder to regulate. However, the law of the large numbers also plays a role which can actually make regulating larger grids easier and more challenging for smaller grids. Additionally, taking into account border conditions like thermal loads of heat pumps is a challenge that is closely related.
With regard to exploiting storage capacity of EVs, the challenge is to train AI to take end user requirements into account, like minimum battery level and expected travel range. Excessive interaction with the user should be avoided, which in turn would need to be automatically optimised. This may involve personal data which is a challenge in itself. Other aspects discussed within the session include battery life and warranty.

Furthermore, business models for companies and end users are an aspect that is not yet elaborated in terms of the business ecosystem, including regulation. Specifically, there is a challenge to address the changing role of prosumers, which will be given a larger role. A better understanding of the economical and psychological incentives is also considered very important.
Social interactions in smaller communities play a role; besides pricing being an incentive, also behaviours and norms will be important. Agent-based modelling can help understand such dynamics in a better way.

Furthermore, there is a need for AI-supported system design for future energy systems, especially integrated systems that make use of flexibility of subcomponents. This also involves knowledge management and knowledge transfer from experts.

AI for Optimal Prosumer Management

The presence of intermittent renewable energy sources requires an increased flexibility of the generation in the energy system. One way to balance the stochastic availability of renewable energy is to foresee significant reserves and to cover for the uncertainty by generation. Another way is to optimally schedule both producers and consumers of energy while accounting for uncertainty. This breakout session focused on the use of AI to optimally schedule industrial generation and consumption of energy. In this session, challenges and relevant industrial applications were discussed, identified, and concluded with some recommendations and concluding remarks.
The session started with a talk by Dr. Marija Zima, who has many years of experience in the relevant field from ABB.. The presentation "*How is AI related to relativity?*" addressed the three keywords for a successful renewable proliferation: space, time and mass (energy). It was demonstrated how AI can help to firstly design the *space* optimally accounting for the grid topology, to provide possibilities to identify the potential sources of waste and, in real time, optimise the generation as well as the consumption of *energy*: Defining power and energy ratings of resources, forecasting the conditions, optimising planning and scheduling— including uncertainty—, considering the asset health, and maintenance and finally fulfilling the production schedule with an optimal energy consumption.
Following the valuable input from the opening presentation, several challenges were discussed. In this regard, a cluster of challenges around data quality, availability, and accessibility as one of the major challenges were identified. Another big challenge is the necessity for developed models to fulfil the sufficient level of trustworthiness, explainability, scalability and maintainability during their lifecycle. The complexity of the objectives in this context was raised for discussion. The distributed nature of the energy management problem, with limited certainty of forecasting the future load together with the intermittent nature of the renewable energy sources on one hand and increasing demand on energy resilience and security on the other hand, makes the problem even more challenging. It was emphasised that the need for expertise and staff with sufficient know-how at different levels will be an increasingly urgent matter. Considering the industrial applications where AI for optimal prosumer management would play a role, three clusters of applications, from industrial plants to grid operation and private sites were covered. In all these three areas, AI can provide different smart solutions and services, such as price prediction on hourly basis, sustainability support and guidance, design optimization, optimised usage of self-generated

energy, using industrial plants as intelligent nodes in a grid, etc.
Finally, it was concluded that there is a need to ensure access to high quality data through a structured data acquisition to get the most out of AI in service for an optimal prosumer management. It is necessary to have a tight integration between design tools and optimization configurations. Here, there is an emerging need for standardisation and legislation for both data as well as the developed models. Trustworthiness, ease-of-use, and Explainability are the key features for the models developed in this context. Last but not least, competence and know-how in this emerging area are necessary.

AI for the Integration of Renewable Energies

The power feed-in of distributed renewable energy sources and the demand for electrical power increase on the distribution level and confront low-voltage grids by challenges like unmonitored overloads or violations of the voltage range. Additionally, the volatility of loads and renewable energies makes it challenging to predict future grid states and plan and apply preventive measures. The breakout session focused on essential building blocks for integrating renewables, such as AI-based forecasting algorithms, and how to train them. Initially, some key challenges in the field of renewable energies like Thermal Load Integration and how they can be overcomed were discussed. Cold countries face the unique challenge of providing enough thermal load to a significant amount of end-user demand. The potential of converting excess renewable energy into heat and storing it in thermal storage systems, which provide a cost-effective solution, were discussed. However, determining the optimal share of energy conversion depends on factors such as generation capacity, load profile, and expected thermal demand. The session also explored the development of AI models to effectively address this challenge. The production of green hydrogen as a storage medium posed another challenge with regard to the technical and economic aspects of producing green hydrogen from renewable sources. Furthermore, the potential solutions were elaborated to ensure its usability in energy systems. Finally, the goal was to identify ways to use AI in optimising the production and use of green hydrogen.
In addition, another challenge defined by the participants related to the Energy Equilibrium in Hybrid Renewable Energy Systems. Integrating multiple renewable energy sources into a hybrid system requires maintaining energy equilibrium. To address this challenge, AI models for optimising energy distribution and achieving a balanced hybrid grid were explored. These models are able to intelligently allocate energy resources by analysing real-time data, including energy demand, generation capacity, and weather conditions.
For effective energy management, accurate forecasting is crucial. During the session, the importance of data availability and generation for both regular and extreme conditions was emphasised. AI models are capable of processing real and simulated scenarios, including rare events, which can significantly improve forecast accuracy supporting decision-making processes.

Automation and Autonomy

The topic was first introduced by the moderator with a mind map of themes related to automation and autonomy with a preliminary definition of these terms relating to the dilemma of AI, moving from the automated repetition of procedures to the delegation of tasks to autonomous agents, and the resulting reduction in (direct) control. This general theme can

be applied to urban scenarios, including the production, distribution and consumption of energy.
It was mentioned that trust in autonomous systems requires human-machine interaction. However, it may be the case that trust does not require deep understanding. Much of the discussion revolved around experimenting in simulated environments (including virtual reality) with synthetic data (digital twins) as a means to achieve trust in autonomous systems by experience. Among the various examples, the case of autonomous drilling for oil or gas on a full-scale test rig was discussed (a system capable of taking its own decisions by evaluating the current conditions and adapting to them while considering multiple horizon strategies to fulfil the drilling operation goal). While such test environments are of great help, it is important to stress that they also have their limitations, such as the quality of sensors (resolution, latency, accuracy, reliability, etc.). Interactions of humans with robots in a factory were also discussed as a case where experience can be gained in a relatively well-controlled environment.

AI for e-mobility integration onto the grid

In the context of climate change and stricter regulations on greenhouse gas emissions, electric mobility (e-mobility) offers a sustainable response to urban pollution. The technological performance of batteries makes Electric Vehicles (EVs) a key element in the development of e-mobility, with almost 9 million EVs expected on Europe's roads by 2030. This growing number of EVs is challenging for both grid operators and energy providers. AI can help facilitate the integration of EVs into the grid by balancing the needs of the grid/energy resources, providing flexibility services and grid stability, and accelerating smart charging.
More specifically, AI models could be used to better understand user habits in terms of driving and charging patterns. In fact, energy demand for vehicle charging overlaps with other electrical uses, making it important to have predictive models (load, charging point occupancy) to anticipate and meet demand. Moreover, AI optimisation or reinforcement tools could help develop smart charging solutions: These solutions help to manage EV charging power in an efficient, flexible and economical way. With smart charging solutions, EV batteries can be viewed as energy reservoirs that can be used to feed energy back into homes (V2H), buildings (V2B) or electricity grids (V2G). AI could also help optimise the sizing and location of charging infrastructure for different uses such as residential, commercial and public.
A specific focus was on the Smarter Mobility Data Challenge which aimed to predict the occupancy of charging points in Paris. This challenge highlighted issues such as the lack of publicly available datasets, the quality of the available datasets, the difficulty of modelling user behaviour, and evaluating the predictions. One important point is the lack of a complete dataset including state of charge information, mobility data, EV connections, traffic data, etc. Networking activities brought by TAILOR, challenge organisations could help to fill this gap..
In summary, this session highlighted the transformative potential of AI in addressing the challenges associated with integrating EVs into the grid. By leveraging AI capabilities such as predictive modelling, optimisation, and reinforcement learning, it will be possible to develop efficient and intelligent charging solutions that meet the demands of a growing EV market, while ensuring grid stability and maximising the benefits of EVs as an energy resource. However, it is critical to address data challenges and continuously improve model

accuracy to enable effective decision making and planning for EV deployment and grid integration.

Individual energy contracts based on AI

Within this breakout session it was discussed how AI can be used to tailor energy contract offers for each consumer or consumer group, e.g., based on energy consumption profiles, demand forecasts, demand-side management options, and energy price forecasts. Furthermore, risk-sharing models based on price guarantees and minimum charges could also be tailored based on consumer knowledge.
Concerning the idea of AI-based energy contracts, however, it is questionable whether and how these can be applied in the future and which challenges need to be overcome or reconsidered. For instance, one of these challenges in evaluating if micro-contracts (peer-to-peer or agent-based) or individual contracts are performing best.
In addition, another important aspect of the discussion was to determine who holds the authority in designing energy contracts - whether it is the retailer or the consumer. Addressing this potential conflict of interest requires thoughtful auditing to maintain fairness and avoid any exploitation of consumers.
Furthermore, the impact of EU regulations on energy contract tailoring was discussed, particularly regarding social aspects. The participants raised concerns about ensuring equitable access to favourable contracts, regardless of consumers' means. Reducing the energy loads of high-energy consumers with greater resources became a priority.
The breakout session also explored the potential of newly rising community-based models for tailored energy contracts. By fostering collaboration and shared benefits within local energy communities, these models aimed to enhance resilience, efficiency, and social cohesion.
Another discussed challenge related to data exchange when consumers switch energy retailers. Ensuring the seamless transfer of historical data while protecting consumer privacy was crucial in facilitating transitions between energy providers. Continuing the idea of transparency, namely making AI algorithms transparent and accessible, could enhance trust among stakeholders and promote understanding of how tailored energy contracts are created. To implement this, the applicability of an open-source foundation could be the key.

Input for the roadmap

Based on the results summarised in the previous section, the Organising Committee identified several topics which could be a valuable input to a European AI research and innovation roadmap. These topics will be presented to and further discussed with experts from TAILOR, VISION and CLAIRE in order to enrich the respective roadmap activities.

The below topics stood out most prominently and will thus provide the ‘core’ of the input. However, during the actual development of the roadmaps, all inputs from the Theme Development Workshop will be considered.

Energy sector specific

- **Energy efficiency using AI**: Focus on achieving improved energy efficiency through AI-enabled software solutions. This includes energy consumption anomaly detection and time-series forecasting.
- **Machine learning applications**: Explore the application of machine learning solutions for energy optimization in various sectors such as transportation (land, air, and sea), heavy industries, heating and ventilation, data centres, and energy storage.
- **Business models and incentives**: Develop suitable business models that align with conflicting objectives such as carbon neutrality and cost minimization to promote the adoption of AI solutions for energy efficiency.
- **Collaboration and cross-domain innovation**: 2BFoster collaboration among stakeholders, including universities, automation suppliers, and industry end-users, to combine machine learning and optimization approaches for energy efficiency.
- **Integration of energy systems**: Focus on the integration of energy management and distributed smart building systems to achieve a resilient grid infrastructure powered by renewable energy sources. Emphasise the partnership between humans and smart energy systems.
- **Scalability and decentralisation:** Consider scalability as a crucial factor for AI solutions in energy systems, especially when dealing with large and interconnected systems. Explore the potential of edge computing and decentralised coordination mechanisms.
- **Interdisciplinary research and knowledge transfer**: Encourage interdisciplinary collaboration and knowledge transfer among experts in various fields to address the complex challenges at the intersection of energy and AI.
- **Integration of electric vehicles (EVs) into the grid**: Leverage AI models to manage the integration of EVs into the grid effectively. Focus on predicting user behaviour, optimising charging infrastructure, and using EV batteries as energy reservoirs for grid stability and flexibility.

More general topics not limited to the Energy sector

- **Explainable AI and Trustworthiness**: There is a need for AI models and systems to be explainable, transparent, and trustworthy, which is of importance across all domains. It is crucial to address the black box effect of AI models and develop tools and methods for explainable artificial intelligence. This involves considering interpretability at different levels, including feature interpretability, model interpretability, and decision interpretability.
- **Human-AI Collaboration and Interaction**: The concept of human-AI ecosystems and collaborative sustainable buildings highlights the importance of effective interaction and collaboration between humans and machines. AI should be designed to work together with humans as collaborators in teams, aiming for shared objectives and mutual support. The field of human-machine interaction, including preference elicitation, aggregation, and understanding human requirements, plays a significant role in achieving successful human-AI collaboration.
- **Data Quality, Accessibility, and Privacy**: There is a need for high-quality data, structured data acquisition, and semantic dataspaces for facilitating data

communication. Additionally, privacy concerns regarding the collection, processing, and sharing of data as well the importance of data anonymization and local processing should be taken into account.

- **Optimization and Decision-Making**: AI's role in optimization and decision-making processes is important for AI in the energy sector particularly in the context of energy management, EV charging, and industrial generation scheduling. AI can provide smart solutions for optimising energy consumption, designing efficient systems, and achieving global optimization.

Conclusion and Outlook

This deliverable pointed out the organisation of the so-called Theme Development Workshops as an innovative engagement strategy between academia, industry and other stakeholders and also explained the most promising results of the past TDWs under the lead of TAILOR as well as their valuable input to the European Research and Innovation Agenda.

Given the great deal of interest we experienced with more than 300 participants from nineteen countries of the past five sector-specific Theme Development Workshops and the overwhelming results, these workshops will serve as a starting point for further TDWs in different sectors or current topics like Climate Change, Green Deal etc. Together with the most important players from industry and research as well as public administration, these topics need to be further developed in a kind of deep-dive workshop to move from broad topics with industry to specific topics with broad industry participation from all sectors. After various discussions with the industry partners from WP8 as well as the Commission, we see a great interest in this further development of the workshops and have already received requests for further TDW on specific topics. In addition to the further development of the Theme Development Workshops, the organisation of follow-up meetings is another major topic of interest. The participants expressed interest in further defining the identified topics and deepening the cooperation. However, in order to organise such follow-up meetings, further capacities would be necessary that are currently not covered in the TAILOR project.

Nevertheless, the exploitation of the results of the Theme Development Workshops and the collection of ideas from industry will open up new opportunities for applying for new proposals and research tenders in the future.

Following the planned series of Theme Development Workshops, initial discussions on further opportunities for cooperation were arranged and carried out. Within these discussions, the participants agreed that the knowledge gained must now be transferred into industrial solutions in order to generate impact. A planned approach for this is the organisation of industrial internships, research visits and PhD positions between the partners. During the Theme Development Workshops various collaborations have already laid the foundation for longer-term cooperation. For example, DFKI and the Luxembourg Institute of Health have established a joint TAILOR Transfer Lab which is now being revitalised with the first small projects in the TAILOR context. The exchange of young researchers between the two parties is also being promoted with the support of the TAILOR Exchange Fund.

In addition to the exchange between research and industry, the results and data of the TAILOR project are to be made more visible and processed and presented in a simple and visible way for outsiders. This should ensure that industry partners outside TAILOR can also access the results and contact TAILOR if necessary. For this, a central office must be set up to process these requests and forward them to the appropriate partners. To enable this communication and distribution between TAILOR partners and external parties, it is planned to establish a so-called alliance of research and industry partners as well as politics in order to combine forces and to create and operate a joint education, research and innovation policy. Together, potentials of cooperation have to be used and developed. With the planned

alliance a platform can be conceived and built up, with the help of which all partners can offer their services and expertise and cooperation can be facilitated.

This platform also holds great potential to promote new collaborations also with partners outside TAILOR in so-called micro-projects. Thus, with the help of the integrated partners, funding opportunities could be made available to enable joint small projects between research, industry and politics. The structures for this will be made available with the planned platform.