



# DATA QUALITY REQUIREMENTS FOR INCLUSIVE, NON-BIASED AND TRUSTWORTHY AI

Putting Science Into Standards



*JRC Conference and  
Workshop Report*

Balahur, A.; Jenet, A.; Hupont Torres, I.; Charisi, V.;  
Ganesh, A.; Griesinger, C.B.; Maurer, P.; Mian, L.;  
Salvi, M.; Scalzo, S.; Soler Garrido, J.; Taucer, F.;  
Tolan, S.

This publication is a Conference and Workshop report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

#### Contact information

Name: Andreas Jenet  
Address: Rue du Champ de Mars 21, 1050 Ixelles  
Email: JRC-PSIS-WORKSHOP@ec.europa.eu  
Tel: +32 229-82302

#### EU Science Hub

<https://joint-research-centre.ec.europa.eu>

JRC131097

PDF ISBN 978-92-76-59091-0 doi:10.2760/365479 KJ-03-22-173-EN-N

Luxembourg: Publications Office of the European Union, 2022

© European Union, 2022



The reuse policy of the European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of photos or other material that is not owned by the European Union/European Atomic Energy Community, permission must be sought directly from the copyright holders.

How to cite this report: Balahur, Alexandra; Jenet, Andreas; Hupont Torres, Isabelle; Charisi, Vasiliki; Ganesh, Ashok; Griesinger, Claudius B.; Maurer, Philip; Mian, Livia; Salvi, Maurizio; Scalzo, Salvatore; Soler Garrido, Josep; Taucer, Fabio; Tolan, Songül, *Data quality requirements for inclusive, non-biased and trustworthy AI. Putting-Science-Into-Standards*, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/365479, JRC131097.

# Contents

## Table of Contents

|  |    |
|--|----|
| Foreword.....  | 1  |
| Abstract.....  | 2  |
| 1 Introduction.....  | 3  |
| 2 State of play.....   | 4  |
| 2.1 International recommendations and guidelines.....  | 5  |
| 2.2 Data in the context of the European Union’s draft AI Act.....                                      | 6  |
| 2.3 Selected international standardisation activities and platforms.....                               | 7  |
| 2.4 Standardisation of AI data in the EU.....  | 7  |
| 3 Horizontal initiatives for data quality assessment and bias mitigation in research and industry..... | 8  |
| 3.1 Creating and documenting datasets for AI.....  | 8  |
| 3.1.1 State of the art, challenges and ongoing standardisation activities.....                         | 9  |
| 3.1.2 Pre-normative research gaps and standardisation opportunities.....                               | 10 |
| 3.1.3 Prioritisation and conclusions.....  | 12 |
| 3.2 Data quality and bias examination and mitigation in AI.....  | 12 |
| 3.2.1 Pre-normative research gaps and standardisation opportunities.....                               | 12 |
| 3.2.2 Prioritisation and conclusions.....  | 13 |
| 4 Data quality needs and practice in selected sectors.....   | 14 |
| 4.1 Education and Employment.....  | 14 |
| 4.1.1 State of the art, challenges and ongoing standardisation activities.....                         | 15 |
| 4.1.2 Pre-normative research gaps and standardisation opportunities.....                               | 16 |
| 4.1.3 Prioritisation and conclusions.....  | 17 |
| 4.2 Law enforcement and the public sector.....   | 17 |
| 4.2.1 State of the art, challenges and ongoing standardisation activities.....                         | 18 |
| 4.2.2 Pre-normative research gaps and standardisation opportunities.....                               | 19 |
| 4.2.3 Prioritisation and conclusions.....  | 21 |
| 4.3 Finance.....   | 21 |
| 4.3.1 State of the art, challenges and ongoing standardisation activities.....                         | 21 |
| 4.3.2 Pre-normative research gaps and standardisation opportunities.....                               | 22 |
| 4.3.3 Prioritisation and conclusions.....  | 22 |
| 4.4 AI for media, including social media, content moderation, recommender systems.....                 | 22 |
| 4.4.1 State of the art, challenges and ongoing standardisation activities.....                         | 22 |
| 4.4.2 Pre-normative research gaps and standardisation opportunities.....                               | 24 |
| 4.4.3 Prioritisation and conclusions.....  | 25 |
| 4.5 Medicine and Healthcare.....   | 25 |

|       |  |    |
|-------|--|----|
| 4.5.1 | State of the art, challenges and ongoing standardisation activities..... | 25 |
| 4.5.2 | Pre-normative research gaps and standardisation opportunities.....       | 29 |
| 4.5.3 | Prioritisation and conclusions.....                                      | 30 |
| 4.6   | AI for Industrial Automation and Robotics.....                           | 32 |
| 4.6.1 | State of the art, challenges and ongoing standardisation activities..... | 34 |
| 4.6.2 | Pre-normative research gaps and standardisation opportunities.....       | 35 |
| 4.6.3 | Prioritisation and conclusions.....                                      | 36 |
| 5     | Discussion on ways forward.....  | 37 |
| 6     | Conclusion.....  | 41 |
| 7     | References.....  | 43 |
|       | List of abbreviations and definitions.....                               | 48 |
|       | List of figures.....   | 49 |
|       | List of tables.....  | 49 |
|       | Annexes.....   | 50 |
|       | Annex 1. Workshop agenda.....  | 50 |
|       | Annex 2. Workshop Advisory board members.....                            | 51 |

## Foreword

Nearly ten years ago, the Joint Research Centre published the guiding document “Science for standards: a driver for innovation” about how science and standards interact along examples of JRC’s own involvement in pre-normative research and standardisation. The document recommended the translation of research and innovation into standardisation, which led in April 2013 to the organisation of the first Putting-Science-Into-Standards workshop.

The Putting-Science-Into-Standards workshops are co-organised by the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC) and the Joint Research Centre of the European Commission, bringing together the scientific, industrial, and standardisation communities. These workshops aim at facilitating the identification of emerging science and technology areas that could benefit from standardisation activities to enable innovation and promote industrial competitiveness. Seven workshops have been held since 2013 in different fields of science.

The Putting-Science-Into-Standards has played a crucial role in anticipating and foreseeing the transformation of emerging technologies into industrial valorisation, contributing to the establishment of at least 3 standardisation platforms: a Joint Technical Committee on hydrogen, a Focus group on quantum technologies, and a Focus group on organ on chip.

This year’s Putting-Science-Into-Standards workshop focussed on data quality requirements for inclusive, non-biased and trustworthy artificial intelligence, anticipating future standardisation needs resulting from various ongoing regulatory activities, such as the artificial intelligence act. The workshop kick-started a forum that discussed priorities, particular technologies and the drafting of a potential standardisation roadmap.

The Putting-Science-Into-Standards workshop demonstrated the value offered by the unique position of the Joint Research Centre on being on the one side integrated in the scientific community, and on the other side active in technical committees of European and International Standardisation Organisations and other standardisation bodies.

## Acknowledgements

The authors would like to thank the speakers Bernard Magenmann (Deputy Director General JRC), Stefano Calzolari (President CEN), Lucilla Sioli, Gabriela Ramos, Karine Perset, Ansgar Koene, Reva Schwartz, Gabriele Mazzini, Sebastian Hallensleben, Patrick Bezombes, the chairs and discussion panellists for their contributions: Felix Naumann, Emmanuel Kahembwe, Kasia Chmielinski, Flora Dellinger, Rasmus Adler, Francisco Herrera, David Reichel, Fred Morstatter, Dee Masters, Nikoleta Giannoutsou, Enrique Fernandez Macias, Patrick Grother, Javier Rodríguez Saeta, Robin Allen, Rosalía Machín Prieto, Karen Croxson, Andrea Caccia, Jörg Osterrieder, Symeon Papadopoulos, Maja Pantic, Manuel Gomez Rodriguez, Jochen Leidner, Sandra Coecke, Thorsten Prinz, Alpo Väri, Koen Cobbaert, Aurélie Clodic, Roland Behrens, Emmanuel Kahembwe, Adil Amjad, Matteo Sostero., As well as Emilia Gómez, Emilia Tantar, Salvatore Scalzo, Antonio Conte, David Reichel, Agnès Delaborde, Philippe Saint-Aubin, Elena Santiago Cid for their concluding and closing remarks. Finally, all the participants for their active involvement during the event and validation of the report.

The workshop would not have been a success without the contribution of JRC and CCMC colleagues Angel Alvarez Martinez, Els Somers, Giovanni Collot, and Wallis Raekelboom. We are grateful for the proofreading by Rikst van der Schoor.

## Authors

Balahur, Alexandra; Jenet, Andreas; Hupont Torres, Isabelle; Charisi, Vasiliki; Ganesh, Ashok; Griesinger, Claudius B.; Maurer, Philip; Mian, Livia; Salvi, Maurizio; Scalzo, Salvatore; Soler Garrido, Josep; Taucer, Fabio; Tolan, Songül.

## **Abstract**

A decade of rapid development of artificial intelligence (AI) has resulted in a large diversity of practical applications across different sectors. Data play a fundamental role in AI systems, which can be seen as adaptive data processing algorithms that adjust outputs to input training data. This fundamental role of data is reflected in the EU policy agenda where for example guidance on handling the data is specified in the AI Act. In response to the needs of the AI Act, the Joint Research Centre, in collaboration with the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation, organised the Putting Science Into Standards workshop on data quality requirements for inclusive, nonbiased, and trustworthy artificial intelligence. The workshop took place on 8 and 9 June 2022, with more than 178 participants from 36 countries gathering for the first time European standardisation experts, legislators, scientists, and societal stakeholders to map pre-normative research and standardisation needs. The workshop highlighted existing and the need of new standards from the creation and documentation of datasets all along to data quality requirements, bias examination and mitigation of AI systems. The workshop also identified the steps needed to start the process of drafting new standards and recognised that inclusiveness and full representation of all relevant stakeholders, including industry, SMEs representatives, civil society, and academia is crucial. Building a stronger engagement of experts in AI standardisation is essential to contribute to the development of standards not only to support the market deployment of AI systems in accordance with the AI act, but also to support this growing field of research.

# 1 Introduction

The European Commission's Joint Research Centre (JRC) and the European Standardisation Organisations, CEN and CENELEC, carry out annual 'foresight on standardisation' exercises under the Putting-Science-Into-Standards (PSIS) initiative.

The PSIS initiative aims at identifying emerging science and technology areas that would benefit from standardisation activities to enable innovation and enhance industrial competitiveness. Every year, CEN and CENELEC and JRC select a topic for a PSIS workshop from a variety of proposals made by JRC scientists.

PSIS workshops bring together regulators, scientific communities, industry partners and the standardisation community to map standardisation needs arising from European and international initiatives and to translate them into proposed actions for the technical committees in charge of drafting standards.

The field of artificial intelligence (AI) has experienced an unprecedented rate of development in the last decade, which is accompanied by a wide range of practical applications across sectors as varied as medicine, finance, manufacturing, employment, social media, law enforcement, robotics, energy and environmental politics, and agriculture. Data play a fundamental role for AI systems, which can be seen as adaptive ("learning") data processing algorithms that adjust outputs to input training data. This fundamental role of data in this context is reflected in the European Union (EU) policy agenda.

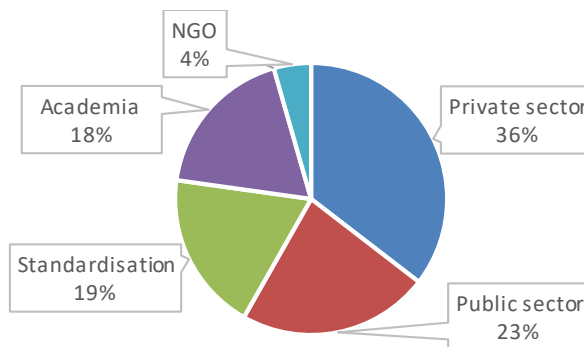
Examples of ongoing policy developments for AI and data include the:

- European Data Strategy [ref COM(2020) 66 final]
- General Data Protection Regulation [ref Regulation (EU) 2016/679]
- Digital Services Act [ref COM(2020) 825 final]
- Data Governance Act [ref COM(2020) 767 final]
- AI Act [ref COM/2021/206 final],
- Data Act [ref COM(2022) 68 final]

European and international standardisation organisations developing the standards to support these regulations are faced with the task of capturing in their specifications the existing landscape of best practices, existing state-of-the-art techniques, and tools and methods in the field, while taking into account all stakeholders' needs.

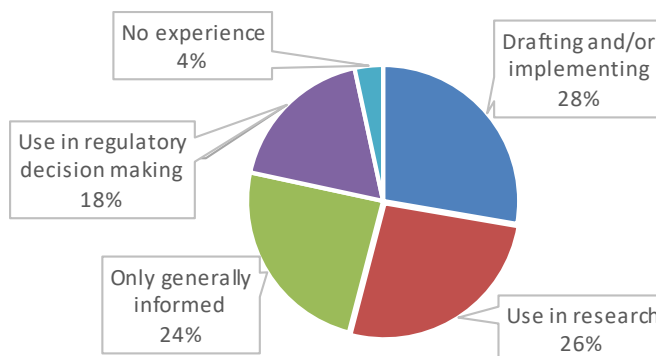
Given the crucial role of data for developing AI systems and the path towards the standardisation of AI and data for AI, the 2022 PSIS workshop focused on the topic of "Data quality requirements for inclusive, nonbiased, and trustworthy artificial intelligence". It took place online on 8 and 9 June 2022, with more than 178 participants from 36 countries, out of which 137 from 21 EU Member States. The main objectives of the workshop were defined as follows:

- Exploring current and future needs and recommendations to address data quality and related ethical concerns in the context of AI and the future AI Act;
- Mapping ongoing standardisation efforts and identification of potential standardisation gaps that might lead to road-blocks concerning the necessary clarity for developers of AI systems or which may affect the trustworthiness of emerging AI systems;
- Gathering key elements for ensuring data quality and associated standards for AI models; and
- Identifying priorities and potential timelines for the process of developing standards or other relevant documents, e.g. guidance documents for addressing specific sectorial needs or technical reports that summarise pertinent technical and scientific aspects in sufficient detail.



**Figure 1 Type of workshop participating organisations.**

The organisations participating in the workshop were well distributed over the different organisation types. With 36% the private sector filled the largest share, followed by the public sector, including the European Commission, at 24% (see Figure 1). Both the public sector, standardisation organisations and academia, including universities and research organisations, were represented each by around a fifth. Civil society organisations, such as NGOs, trade unions, and some individuals, made up 4% of the total registrations.



**Figure 2 Level of experience in standardisation among the participants**

Overall the knowledge of the participants was to a large extent at a professional expert level, and well distributed among the different types of organisation. Nearly a third of the registered participants are involved in drafting and/or implementing data standards (see Figure 2).

The first day of the workshop discussed the political and technical implications of the AI Act, as well as international initiatives, e.g. from UNESCO and OECD focusing on data and AI, and its link with trustworthiness, human rights and democratic values.

The second day of the workshop consisted of parallel sessions of technical character, involving prominent AI practitioners and researchers, who helped summarise the current state of the art on data-driven techniques and tools to ensure the trustworthiness of AI systems, with a focus on fairness and transparency.

## 2 State of play

Technological advances in digital transformation have created a situation where the volume of information generated and shared is outpacing the ability of humans to review and use such information. Novel artificial intelligence technologies, such as machine learning models and big data analytical tools are making sense of this information and providing insights.

Technological developments in computing infrastructures and algorithms have led to ever-more powerful AI algorithms and AI applications, capable of revolutionising virtually every aspect of our lives. In several areas, better processing of information has enabled big advantages, in sectors such as healthcare, law enforcement, finance, media and education.



These developments have led to a situation where AI models are becoming more complex, more accurate and more widely used than ever before. Developments in storage and processing capabilities provide the base for deep neural network models that are trained on vast data, reaching unprecedented accuracy on many automated tasks.

Applications of AI models today range from well-known fields, such as recommender systems and search result ranking, to sensitive use cases, such as medical diagnostics, bank loan approval or CV filtering. Concerns of potential biases in AI with important and immediate effects in our lives, are growing. Biases are prevalent in existing state-of-the-art models, raising concerns on societal consequences, in particular for AI-informed decisions in the fields of health, security, finance, recruitment and education. Reproducibility is also important to ensure accountability. Not less important is the transparency about data use, i.e. as which processes are based on data, or how data is used and what data is used.

All this leads to the conclusion that data quality is of fundamental importance for AI tools to function well. The High-Level Expert Group on AI, appointed by the European Commission in 2018, stated "Diversity, non-discrimination and fairness" as one of the seven key requirements for trustworthy AI systems (European Commission 2019). This includes the avoidance of unfair bias, accessibility and universal design, and stakeholder participation. "Privacy and data governance" is another key requirement stated by the expert group. However, data quality relates to most of the groups' seven requirements, such as "technical robustness and safety", "transparency", and to some extent also "accountability"

Although the complexity of AI deep learning models makes their inner workings challenging to explain, the scientific community and tech companies are responding to the criticisms and expanding efforts to alleviate the root causes of AI biases.

Proving that the data on which models are trained fulfil quality standards is of paramount importance to ensure inclusiveness, ethical use of AI and absence of bias.

## **2.1 International recommendations and guidelines**

Artificial Intelligence has the potential to immensely improve the lives of people around the world. Today, AI is already used to optimise food production, predict disasters, monitor pollution, map infection outbreaks, and many other uses. However, the benefits of AI come with the risk of encoding into the digital biases discriminations and inequality.

The need to institute data quality standards to ensure the development of inclusive and trustworthy AI is pressing. The lack of gender and ethnic diversity in AI research and development means that AI systems are in danger of recreating and perpetuating existing forms of structural inequality, even when working as intended. Studies have shown that voice and facial recognition systems work much better for white males than for all the other groups. This opens the risk to more serious discrimination, especially in situations where AI decision-making can affect the health or human rights of individuals. A solution to these risks can only result from a cross-sectoral, multidisciplinary and collaborative approach – one which engages all relevant stakeholders.

The Recommendation on the Ethics of AI of UNESCO, adopted in 2021, represents such an approach. Addressing the need for transparency, it calls upon Member States to implement policies to promote and provide for diversity and inclusiveness – reflecting their populations in AI development teams and training datasets. This potentially includes investing in the creation of gold standard datasets, which are open, trustworthy and diverse – as well as constructed on a valid legal basis, including the consent of data subjects when required. The recommendation also includes capacity-building and assessment mechanisms to ensure its implementation: in particular, the readiness and ethical impact assessment tools. These tools apply to all stages of the AI life cycle and engage all relevant actors involved in its development and deployment, thereby allowing for the credible and transparent monitoring and evaluation of policies relating to AI technologies. By continuing to centre ethics in the design, development and deployment of AI, we can – as states, institutions and as individuals – chart a collective path towards a more just and equitable future.

The OECD AI Principles (2019) include 10 principles covering two different areas: principles for responsible stewardship of trustworthy AI, and national policy and international cooperation for trustworthy AI. This is in line with work ongoing within the European Commission and UNESCO. Ongoing projects to support these principles include:

- The OECD framework for the classification of AI systems, which is a structure that helps classify AI systems depending on their impact on the public policy areas covered by the OECD principles.

- AI incident tracking to collect information about AI incidents and controversies from all over the world. The objective is to help create the evidence base for effective AI policy, and to avoid reproducing known incidents across jurisdictions.
- Catalogue of tools for trustworthy AI presents instruments and structured methods that people can use to make sure their AI systems respect OECD principles.
- Towards accountability in AI & assessing risks with the objective to contribute to the general understanding of key components of accountability in AI. This includes contributing to existing initiatives, (e.g. ISO, IEEE, NIST, CEN-CENELEC) and integrating OECD work on AI and well-recognised risk management frameworks.

## **2.2 Data in the context of the European Union's draft AI Act**

The EU considers it very important that artificial intelligence is developed, deployed and used in a way that benefits our citizens, societies and economies. These benefits of AI will only materialise through uptake, but uptake requires trust. The AI package adopted in April 2021 pursues the twin objective of supporting innovation and excellence and ensuring trust in AI.

The AI Act seeks to define the uses of AI technologies that are allowed and the conditions under which they can be deployed on the EU market. This goal will be achieved through a proportionate, risk-based approach, and that is why the AI Act mainly focuses on high-risk systems. At the same time, it is innovation-friendly because it addresses the real risks of the technology, while shielding systems that do not pose high risks to safety or fundamental rights from diverging national regulations.

In order to ensure trust and a consistent and high level of protection of safety and fundamental rights, the AI Act proposes mandatory requirements for all high-risk AI systems (Hupont, Micheli, et al. 2022). The compliance with such requirements will have to be demonstrated through a conformity assessment procedure, which may require under certain circumstances the intervention of a third-party certification body.

Data plays a prominent role in this context, as there is no AI without data and the quality of the data used is paramount to ensure that AI is trustworthy. For this reason data quality stands as one of the key requirements that high-risk AI systems are expected to fulfil.

As the AI Act is modelled along the lines of the product legislation regulatory scheme, only high-level technical objectives/requirements are contained in the main legal act, while the more granular and specific technical solutions to demonstrate the compliance with those requirements are expected to be set primarily through harmonised standards developed by European Standardisation Organisations. Hence, the role of harmonised standards will be key to ensure that the expectations set in the AI Act with regard to data quality translate into implementable and effective technical solutions, which reflect the latest state-of-the-art. In order to provide the European standardisation organisations with a solid mandate to start their work as soon as possible, the Commission has already launched the adoption process for an official and first standardisation request in the field.

Specifically on data standardisation, given the nature of the challenges associated with data, the Commission considers that ensuring appropriate involvement of concerned stakeholders in the standardisation process is key. The standardisation process should adequately gather and reflect the views of the stakeholders which are part of the AI value chain. For example: involvement of both big and small tech companies is essential for a technology such as AI where SMEs play a fundamental role; or, another example, the views of health professionals and patients are crucial for the development of standards related to data quality in healthcare AI.

In terms of the specific requirements in the legal text, the AI Act sets on the one hand requirements related to good data governance and management, and on the other hand substantial technical objectives related to training, validation and testing data (e.g. relevance, representativeness) that the provider of an AI system shall fulfil. These are also intended to prevent the risk of bias and discrimination, which is one of the most challenging threats posed by AI systems to fundamental rights.

Furthermore, in addition to the already mentioned pre-market expectations and requirements, there is an important post-market dimension in the AI Act (and it could not be otherwise given the nature of AI technologies), which substantiates notably into post-market monitoring requirements for providers. In order to make pre-market and post-market controls more effective in relation to data, the AI Act explicitly provides certification bodies and national authorities with the empowerment to access the datasets utilised by providers.

## **2.3 Selected international standardisation activities and platforms**

The EU is committed internationally at multilateral and bilateral level to promote and build a vision of trustworthy and human-centric AI, ensuring that all efforts at international level are appropriately reflected into standardisation supporting the future AI Act. Agreements between European Standardisation Organisations (ESO) and International Standardisation Organisations (ISO), as well as relevant ad-hoc initiatives, can ensure that international standards can be used at European level (also as harmonised standards).

In terms of international AI standardisation ISO/IEC JTC1 SC42 (Joint Technical Committee 1, Sub committee 42) is the main source, with a considerable history of AI work, and a substantial number of standards published or on development at different stages, ranging from foundational standards (covering e.g. terminology) to standards on AI management and trustworthiness. ISO/IEC will play a fundamental role as a source of standards to support the AI Act under the frame of existing cooperation agreements (Frankfurt and Vienna agreements) with European Standardisation Organisations.

In terms of data-related AI standardisation, current coverage in ISO/IEC is highly relevant. Specific standards include those in the 5259 series (ISO 2022), which provide a broad coverage of data quality including terminology, data quality measures, management requirements and guidelines, and a quality process framework. These are complemented by other standards as well as by technical reports, such as the draft TR 24368 covering ethical and societal concerns or the TR 24027:2021 on bias in AI systems (ISO/IEC 2022).

Other international standardisation organisations are also active in the AI field. This includes the Institute of Electrical and Electronics Engineers (IEEE) Standards Association, which is working on a wide range of AI standards, including an entire series (the 7000 series) with a strong focus on ethical aspects. Relevant content in this series ranges from methodologies to support AI providers to integrate ethical considerations in the design process to standards on topics such as transparency, bias and privacy, and also includes specific work on a certification programme.

A particularly relevant IEEE standard in terms of data quality is the 7003 on bias considerations aiming to minimise unintended, unjustified and unacceptable bias. It aims to provide guidance for developers and deployers across the AI lifecycle, ensuring that AI systems are designed in full consideration of the impact on relevant stakeholders, that they are used in their intended context, and that documentation and transparency is present throughout the process.

Across the Atlantic, in the United States, NIST activities also include relevant work on managing bias in AI and developing an AI risk management framework. Indeed, bias is a crucial risk in AI, and mitigating it is an important element of risk management. In this regard, bias considerations extend beyond computational bias. Other relevant categories include human cognitive bias, e.g. in the decision making processes in the design lifecycle or in use of AI when deployed, and also systemic biases, for example in terms of how institutions make decisions. These biases can be prevalent in data, models and organisations, and their impact must be understood and measured.

A risk management framework is being developed by the National Institute of Standards and Technology (NIST) that defines the processes to identify, assess and prevent harms, taking into account context while being aware of potential sources of bias, such as selecting data based purely on availability or limiting testing systems to optimised scenarios. In terms of key factors specific to data, scientific integrity, proper consideration of context and the adoption of methods based on social sciences can be an important part of an overall risk management strategy.

## **2.4 Standardisation of AI data in the EU**

The European Standardisation Organisations CEN and CENELEC recognised the urgent need for AI standardisation and launched last year the Joint Technical Committee 21 'Artificial Intelligence' (JTC 21), responsible for the development and adoption of standards for AI, as well as providing guidance to other technical committees concerned with AI.

The priority of JTC 21 is to produce standardisation deliverables that address European market and societal needs and support European Union legislation. Notably, support for the AI Act is a priority, and in the future this may extend to other regulations, e.g. on data or AI sustainability aspects. Whenever possible, supporting European standardisation needs is done through the adoption of suitable international standards, preventing duplication of efforts and supporting European and international convergence. Therefore, a primary activity in

JTC 21 is the identification and adoption of international standards available or under development from other organisations such as ISO/IEC.

While some concrete standardisation areas, such as terminology, can largely rely on international standards, others may require a combination of European and international standards. In this manner, specific gaps at the international stage are addressed at European level. Complementary European standardisation coverage is expected on topics such as AI trustworthiness characteristics (human oversight, accuracy, robustness and others), the provision of concrete metrics and controls for data quality, risk management, and on the definition of conformity assessment methodologies for high-risk AI systems. Future work may also cover the mitigation of ethical and societal concerns, as well as the environmental sustainability of AI.

This will result in a solid foundation of standards applicable to high-risk cases across the board, in line with the horizontal nature of the European AI Regulation. Such standards will ensure full coherence between the different specifications, whether European or international, and capture European values and principles. On the practical side, these standards aim to be actionable, with concrete technical specifications and metrics, as well as a range of good practices supporting AI providers to innovate in compliance with legal requirements.

### **3 Horizontal initiatives for data quality assessment and bias mitigation in research and industry**

The first part of the workshop focused on horizontal data quality and transparency approaches coming from industry and academia, covering different phases of an AI system's lifecycle (e.g. dataset building, model training, system deployment and post-market monitoring). These horizontal initiatives encompass data biases with respect to different groups considering gender, nationality, culture, language and disciplines. The two horizontal approaches addressed by this first part are about creating and documenting datasets for AI and aspects about data quality and bias examination and mitigation in AI.

#### **3.1 Creating and documenting datasets for AI**

Data is the raw material needed to train, test and validate AI systems. Furthermore, data is central throughout the AI development lifecycle, including, among others, steps devoted to data preparation, curation, annotation, cleaning, and sharing. The creation of high-quality datasets in terms of completeness, correctness, representativeness and preservation of privacy have a direct impact in the development of trustworthy AI systems. However, recent studies have shown that most popular and widely used datasets for AI are highly noisy (e.g. samples incorrectly labelled, missing data, bad quality samples) (Wang, et al. 2018)] and biased (ie imbalanced in terms of demographics, with little/no representation of certain minority groups) (Hupont and Chetouani, Region-based facial representation for real-time action units intensity detection across datasets 2019, Hupont, Gomez, et al. 2022). It has also been acknowledged that a clear degradation in AI systems' performance occurs as the noise level in the training dataset increases. For instance, Reale et al. (Reale, Nasrabadi and Chellapa 2016) demonstrate that a 10% manual correction of mislabelled samples yields a performance increase similar to that of doubling the size of the dataset.

For all the above reasons, an increasing number of initiatives and good practices for the creation and documentation of trustworthy datasets for AI have shed light in the last four years (Hupont, Micheli, et al. 2022). Starting by the pioneering work "Datasheets for Datasets" by Gebru et al. (Gebru, et al. 2018), different types of documentation approaches have been proposed by different stakeholders, from both industry and academia, to promote transparency regarding the creation process and the contents of datasets. Prominent examples take the form of questionnaires (Gebru, et al. 2018, OECD 2022), checklists (Madaio, et al. 2020, European Commission 2019), information sheets/cards (M. Mitchell, et al. 2019, Matthew, et al. 2019) and widgets (Chmielinski, et al. 2022, Bäuerle, et al. 2022, Holland, et al. 2018).

At present, with the imminent adoption of the AI Act, there is a need to bridge the gap between existing voluntary practices in terms of dataset creation and documentation and the requirements defined in the legal text of the AI Act. In this context, this parallel session on 'creating and documenting datasets for AI' aimed at presenting current state-of-the-art approaches for trustworthy dataset documentation and reflecting on whether they could be leveraged for future standards development.

In an expert panel discussion, Felix Naumann from the Hasso-Plattner-Institut, Emmanuel Kahembwe from the Verband der Elektrotechnik, Elektronik und Informationstechnik (VDE), Kasia Chmielinski from the Data Nutrition Project at Harvard Kennedy School and Flora Dellinger from Confiance.ai highlighted pre-normative

issues and standardisation needs in the field of creating and documenting datasets for AI by discussing practical examples from their work. In the following we present the main discussions, findings and conclusions in terms of standardisation needs that arose during this session.

### 3.1.1 State of the art, challenges and ongoing standardisation activities

Nowadays, the development of AI systems – especially when underlying AI models are based on Deep Learning – requires vast amounts of data in the form of very large datasets. This panel discussed the main challenges related to the collection, creation and documentation of such datasets, that should be tackled in order to achieve trustworthy datasets for AI. The panellists identified six key challenges that need urgent attention, namely:

- *Data collection and preparation is time consuming and error prone.* Data scientists spend about 60% of their time preparing, labelling and cleaning data for the training and validation of AI systems. Data collection (e.g. gathering data from the Internet, via crowdsourcing or from other sources) is the second most time-consuming task, taking around 19% of their time. Therefore, data collection and preparation accounts for around 80% of the AI system development pipeline (Hameed and Naumann 2020). Furthermore, these tasks require a lot of human supervision and manual work, are extremely tedious, exhausting and highly prone to errors (e.g. incorrectly labelled, bad quality samples) (Song, et al. 2020).
- *Webscraping is a common and potentially harmful practice.* Webscraping, i.e. the process of using bots to automatically extract data from websites, is one of the most popular methods to collect massive amounts of data for AI (Zhu, et al. 2021, Kumar, et al. 2021, Hupont, Tolan and Gunes 2022). This practice might result in copyright issues and the difficulty in tracing the provenance of datasets. Additionally, collecting data from uncontrolled sources (e.g. data generated by internet users) can perpetuate malignant stereotypes and search engine biases as well as lead to the collection of illicit material (Birhane and Prabhu 2021, Vincent and Hecht 2021).
- *Datasets are not sufficiently representative of real-world use cases.* Current datasets for AI are not representative enough of the real-world situations in which AI systems are to be deployed (Hupont, Tolan and Gunes 2022). The most common domain gaps that can be encountered include: demographic biases (e.g. facial datasets dominated by white young men while other demographic groups – particularly people of colour – are highly underrepresented (Hupont and Fernández 2019), not considering corner cases (e.g. no/little representation of minority groups or classes in the dataset), and quality gaps between dataset vs. operational data (Vincent and Hecht 2021). As a result, AI systems' performance is highly impacted when moving to real operational settings (Hupont and Chetouani 2019).
- *Evaluation metrics are generic and research-oriented.* Dataset quality, completeness and correctness is generally evaluated in very simple terms using classic and generic metrics such as the number of samples, the distribution of data classes, data correlation, the percent of missing data, and the percent of incorrectly labelled samples (Afzal, et al., 2021). Novel AI-specific quality dimensions need to be considered, covering factors such as diversity (e.g. data richness, evenness (Hupont, et al., 2022), representativity, biases (e.g. demographic gaps) or ethical transgressions in data content (Birhane & Prabhu, 2021). It is equally important to move from theoretical research-oriented evaluation contexts to novel evaluation benchmarks considering AI systems' real operational settings (e.g. use of different sensors to collect real-world data, ensure the ecological validity of the benchmark).
- *Lack of well-defined data governance, curation and documentation processes.* The quality and integrity of a dataset must not only be checked during its creation process but also be ensured throughout its lifecycle. Until 2018, dataset documentation processes and methodologies were generally lacking. The publication of "Datasheets for Datasets" (Gebru, et al. 2018) contributed to raise awareness among AI practitioners about the need for transparency in the dataset creation process and marked a turning point in the field. Other prominent data documentation methodologies followed, such as the "Data Nutrition Label" (Chmielinski, et al. 2022) or "Accountability for Machine Learning Datasets" (Hutchinson, et al. 2021). These initiatives promote transparent dataset documentation practices, clearly disclosing data collection, cleaning, labelling and curation steps, data liability, governance and lineage issues, data distribution, data gaps and quality measures, among others. However, the documentation of large – sometimes very unstructured and multimodal – datasets is extremely challenging and far from being solved. Current approaches have a rather horizontal focus and are limited when it comes to address all dataset lifecycle phases (Hupont, Micheli, et al. 2022). More vertical (i.e. domain-specific) approaches are yet under-explored.

- *Lack of incentives to foster documentation practices and data sharing.* The culture of data documentation, data sharing and data governance is not yet “in the DNA” of most AI practitioners, institutions and companies (Vincent and Hecht 2021). This is partly due to the lack of incentives to implement this type of practices, particularly in the private sector. Management support and adjustment of organisational structures is necessary, but transparency and data sharing are perceived as being against companies’ interests. The power that large companies have over large datasets favours an AI progress gap between tech giants and SMEs. The public sector also owns a large amount of data, which could potentially be shared with adequate privacy and security measures. Nevertheless, well-established and trusted private-private, private-public and public-public data sharing mechanisms are virtually non-existent.

The panel highlighted some ongoing activities and initiatives that are already starting to address some of these challenges, and that could eventually inspire future standards on the matter:

- *Dataset and AI documentation methodologies.* Existing documentation methodologies for datasets, particularly “Datasheets for Datasets” (Gebru, et al. 2018) and the “Data Nutrition Label” (Chmielinski, et al. 2022) which have a focus on data for AI, are well-structured approaches starting to have a wide adoption and support from AI practitioners worldwide. Other successful methods for the documentation of AI models and systems, such as Model Cards (M. Mitchell, et al. 2018), AI Factsheets (Matthew, et al. 2019) or the OECD framework (OECD 2022), also make considerations related to data quality on which future standards could rely.
- *Programming toolkits.* In recent years, several programming toolkits aiming at automatically generating documentation in the format proposed by the previous methodologies have reduced the potential workload for good documentation practices. Examples include Symphony<sup>1</sup>, Model Cards Toolkit<sup>2</sup> and the Data Nutrition Project<sup>3</sup>. Most existing toolkits allow for the automatic assessment of data quality issues, being able to automatically compute facts and metrics such as data duplicates, outliers, biases in data distributions, mislabelled data instances, and cross-correlations.
- *Checklists.* Checklists are deemed a useful format for auditors, as well as for AI developers and providers, to ensure that all relevant data-quality factors have been addressed throughout an AI system’s lifecycle. Examples include the “Assessment list on trustworthy artificial intelligence” (ALTAI<sup>4</sup>) by the High-Level Expert group on AI from the European Commission, and the checklist around fairness on AI proposed in (Madaio, et al. 2020).
- *Industry initiatives for trustworthy AI.* There are some activities supported by key industry players that aim at fostering the development and commercialisation of trustworthy AI systems, putting particular emphasis on challenges related to data. A prominent ongoing initiative, driven by the German VDE<sup>5</sup> organisation, is the “AI trust label”<sup>6</sup> whose objective is to define a specification for attaching a trust label to AI products. Another example is the “Ensuring trust by data” workforce by the French Con fiance.ai<sup>7</sup> industrial consortium, which focuses on the development of methods, tools and collaboration mechanisms to obtain trustworthy datasets for AI-based critical systems.

The previous activities, tools and methods are independent from ongoing European regulations such as the AI Act, the Data Act and the Digital Services Act, and might differ from them, e.g. in terms of taxonomy and requirements. Nevertheless, they have the potential to serve as a good basis for future related standards.

### 3.1.2 Pre-normative research gaps and standardisation opportunities

The previous activities and tools constitute a promising and clear step forward on the road to high-quality data for AI. However, the development in the future of standards in the field of AI still have the room to bridge some existing research and standardisation gaps as described in the following paragraphs.

The panellists first emphasised the need for specific standards on “data for AI”, as current general standards on data do not address all the concerns related to the datasets that are used to train and validate AI systems (e.g. very large-scale, annotated data, frequently web-scraped, captured by sensors and multimodal). These

<sup>1</sup> <https://www.whysymphony.com/>

<sup>2</sup> [https://www.tensorflow.org/responsible\\_ai/model\\_card\\_toolkit/guide](https://www.tensorflow.org/responsible_ai/model_card_toolkit/guide)

<sup>3</sup> <https://datanutrition.org/>

<sup>4</sup> <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-alta-i-self-assessment>

<sup>5</sup> <https://www.vde.com/en>

<sup>6</sup> <https://www.vde.com/de/presse/pressemitteilungen/ai-trust-label>

<sup>7</sup> <https://www.confiance.ai/>

standards should include representativity considerations, ensuring data that sufficiently represents the operational conditions and demographic setting on which AI systems are to be deployed.

Standards should address the lack of well-defined processes through the development of methodologies on how data has to be collected, stored, labelled, accessed, shared and used within AI. This includes the definition of standardised, accessible dataset documentation guidelines.

Guidelines are also needed when it comes to efficiently measure data quality. On the one hand, new AI-specific data quality metrics must be defined, in order to consider novel dimensions beyond traditional ones (e.g. social, ethical, diversity dimensions). On the other hand, standardised toolkits able to automatically compute metrics from raw data and to correlate model errors with data quality problems can be a valuable tool for AI practitioners to identify problems caused by data.






Algorithmic standardised tools should not be limited to the computation of quantitative metrics. They can also aid with more qualitative – though very tedious – tasks such as data curation, labelling, cleaning and documentation.

All the previous considerations can benefit by aligning with ongoing regulations such as the European AI Act, in terms of processes, metrics and taxonomy. This, together with related checklists, can strongly help companies –especially SMEs– to comply with legal requirements.

Finally, the panellists foresaw the need for different data standards for different AI domains, although this point could be addressed in a later stage, after finding more mature solutions for the gaps laid out in 3.1.1.

The following matrix compiles the standardisation needs identified during the session, which are ordered according to 5 dimensions: terminology, metrology, performance characterisation, compatibility and regulatory assessment.

**Table 1 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain for creating and documenting datasets**

|   |  Terminology   |  Metrology  |  Performance Characterisation   |  Compatibility                             |  Regulatory assessment  |
|---|---|--|--|---|--|
| <b>AI system deployment &amp; marketing</b> <ul style="list-style-type: none"> <li>– Regulatory assessment</li> <li>– Users</li> <li>– Transparency &amp; specification</li> <li>– Accountability &amp; responsibility</li> <li>– Maintenance, post-market follow-up &amp; bias monitoring</li> <li>– Supply network</li> </ul> |   |  |  |   |  |
| <b>AI system creation and production</b> <ul style="list-style-type: none"> <li>– Data sets &amp; algorithms (incl. bias)/models</li> <li>– Cybersecurity</li> <li>– System design &amp; integration</li> <li>– Upscaling &amp; evaluation</li> <li>– Quality control</li> </ul>  |   | <ul style="list-style-type: none"> <li>– Automated tools (data+model behaviour analysis)</li> <li>– New dimensions (e.g. ethical, social, diversity).</li> </ul> |  |   |  |
| <b>Data creation</b> <ul style="list-style-type: none"> <li>– Compilation, preparation, bias testing</li> <li>– Analysis, processing, labelling</li> <li>– Licensing &amp; restrictions</li> <li>– Sharing &amp; marketing</li> </ul>   | <ul style="list-style-type: none"> <li>– Unified taxonomy.</li> <li>– Documentation as integrated part of the data creation process</li> <li>– Standardised processes on data creation and collection</li> <li>– Automated tools for data documentation and creation</li> </ul> | <ul style="list-style-type: none"> <li>– Standards on AI-specific data quality metrics</li> <li>– Fairness and bias metrics</li> </ul>                           | <ul style="list-style-type: none"> <li>– Standards on data integrity through lifecycle</li> <li>– Ensure representativity (operational setting, real-world data, demographic setting)</li> </ul> | <ul style="list-style-type: none"> <li>– Standards on data format and labelling</li> <li>– Standards on data sharing</li> </ul> | <ul style="list-style-type: none"> <li>– Data quality checklists for auditors</li> <li>– Data lineage provenience mapping</li> <li>– Data maintenance</li> </ul> |

### **3.1.3 Prioritisation and conclusions**

Among the previous standardisation opportunities, the panel members identified four of them as high priority. Two could strongly leverage on existing pre-normative initiatives and can be implemented with relatively small effort, namely 1) the definition of a standard checklist for dataset quality assessment and 2) the implementation of a methodology for data lineage mapping. The other two are 3) the definition of a unified taxonomy and data labelling scheme and 4) the implementation of standardised toolkits for data documentation, labelling, cleaning and extraction of metrics. Both would require a much greater effort because of the lack of consensus in the field (3, 4), the large variety of data types depending on the application domain (3, 4), and the strong implementation effort needed to bring them to operation (4).

Nevertheless, while standards are of the utmost importance to ensure quality data for AI, they must come together with the conviction of the community on the need of such practices. Creating data quality and data sharing values and culture is equally important on the way towards trustworthy AI.

## **3.2 Data quality and bias examination and mitigation in AI**

This session presented the latest research on methodologies, tools and best practices in the area of trustworthy AI, with a focus in the area of data quality, as well as bias examination and mitigation. In scope are, for example, concrete techniques that concern data quality, such as data augmentation, weighting loss functions (e.g. based on demographics), blinding methods (e.g. from a protected variable), feedback loops, fine-tuning, federated learning, transfer learning, robustness, evaluation techniques and benchmarks, adversarial attacks, explainability, human oversight, and post-market monitoring. Even in the presence of strong data quality measures, unwanted biases can still be present at the training and deployment stages of the AI life-cycle, causing unintended and possibly unexpected and harmful outcomes. It is therefore important to remain vigilant about best practices for continuous bias examination and mitigation during algorithm training, deployment and operation.

At the training stage, biases may feed into the system through data manipulation and augmentation techniques, fine-tuning steps, the definition of objective functions or the use of specific algorithmic techniques. Even at the deployment stage, AI systems remain vulnerable to biases through potential feedback loops, improper evaluation techniques and benchmarks, or adversarial attacks. Detecting and addressing bias and other data quality issues throughout the entire AI system lifecycle requires the adoption of specific best practices, e.g. related to fairness, transparency, explainability, robustness, and the specification of sustainable monitoring and evaluation techniques.

According to the 2020 PricewaterhouseCoopers AI Predictions report (PricewaterhouseCoopers 2020), 68% of organisations still need to address fairness in the AI systems they develop and deploy.

The session had four speakers: Rasmus Adler from the Fraunhofer Institute for Experimental Software Engineering, Francisco Herrera from the University of Granada, David Reichel from the Fundamental Rights Agency and Fred Morstatter from the University of Southern California Information Sciences Institute.

### **3.2.1 Pre-normative research gaps and standardisation opportunities**

The session addressed the question on data quality to define critical applications of AI, stressing the relevance of fairness in legal frameworks relevant to AI. There is a need to assure data are robust to be trustworthy and to assess test data sets, for both data quality and the limit of the systems used to process them. A proper definition of AI is missing, together with the objectives of its uses in proposed regulations. Furthermore, assurance cases, and their identification, are a very important source for trustworthy data. It would be ideal to move from prescriptive to more goal-based regulations and standardisation based on assurance cases.

It was suggested to address as a key element the contrast between the data centric view point and the model centric view point, the latter focused on data and algorithm codes. It was pledged to avoid discrimination in AI to address fairness. Three main challenges were stressed: i) processing data to obtain smart data; ii) avoid bias in model through data fairness; and iii) protect data privacy. The following steps were identified to address such challenges: establish an AI supervisory agency; create methodologies to ensure data quality, and creating data silos for a data-centric approach by AI users.






The European Union Agency for Fundamental Rights (AFR) is active in linking AI with fundamental rights, including bias and algorithms, in particular on issues related to discrimination and data quality, and on the bias of data that may determine discriminatory outputs (socio-ethnic nature, e.g. or low quality data or



missing data). There are challenges to identify bias in relation to these characteristics. The lack of assessments, identified proxies and data on protected characteristics exacerbate these challenges. Nevertheless, there are limits to the mitigation of poor quality data and their subsequent use. There are also legal requirements and fundamental rights that need to be respected when it comes to documentation and dataset description. Social sciences and statistical offices may improve data and avoid bias. The European Commission, the Council of Europe, OECD, UNESCO and other relevant bodies advocate the need to further carry out research on AI case studies.

The session also discussed the topics of fairness, cultural modelling and conversational agents, where robust models and a removal of biased data is needed in order to mitigate bias. The difficulties that arise from the different meanings of fairness call for a definition of fairness metrics supported by a wide range of stakeholders. In this sense, open data, standardisation and auditing algorithms can be very important to avoid inconsistencies and contradictions.

**Table 2 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain for data quality and bias examination and mitigation**

|   |  Terminology |  Metrology   |  Performance Characterisation   |  Compatibility   |  Regulatory assessment   |
|---|---|---|--|---|---|
| <b>AI system deployment &amp; marketing</b> <ul style="list-style-type: none"> <li>– Regulatory assessment</li> <li>– Users</li> <li>– Transparency &amp; specification</li> <li>– Accountability/responsibility</li> <li>– Maintenance, post-market follow-up &amp; bias monitoring</li> <li>– Supply network</li> </ul> |   |   |  | <ul style="list-style-type: none"> <li>– Define standards for use of AI systems in different contexts, including those where they were deployed for cross-domain use</li> </ul> | <ul style="list-style-type: none"> <li>– Define goal-based regulatory assessment</li> <li>– documentation standards and transparency of system as condition for deployment</li> </ul> |
| <b>AI system creation and production</b> <ul style="list-style-type: none"> <li>– Data sets &amp; algorithms (incl. bias)/models</li> <li>– Cybersecurity</li> <li>– System design &amp; integration</li> <li>– Upscaling &amp; evaluation</li> <li>– Quality control</li> </ul>  | <ul style="list-style-type: none"> <li>– Definition of fairness in AI systems</li> </ul>      | <ul style="list-style-type: none"> <li>– Definition of measures to detect and measure bias in AI models</li> <li>– Measure suitability of data in context of AI system</li> <li>– Definition of fairness metrics in AI systems</li> </ul> | <ul style="list-style-type: none"> <li>– Assurance cases throughout the lifecycle of AI systems</li> <li>– perform research on case studies to properly assess ethics implications of AI used in complex systems</li> <li>– benchmarking datasets for system bias testing</li> <li>– test for hidden discrimination and protected variables</li> </ul> |   |   |
| <b>Data creation</b> <ul style="list-style-type: none"> <li>– Compilation, preparation, bias testing</li> <li>– Analysis, processing, labelling</li> <li>– Licensing &amp; restrictions</li> <li>– Sharing &amp; marketing</li> </ul>   |   | <ul style="list-style-type: none"> <li>– Data quality assessment metrics</li> </ul>   |  |   |   |

### 3.2.2 Prioritisation and conclusions

In order to proceed with data quality standards, identification and mitigation of bias along the AI value chain, there is a need to agree on adequate data quality assessment metrics. There is also a need to define fairness and its metrics in AI systems. Furthermore, definitions are lacking for detection and measure bias in AI models, so as to measure the suitability of data. To allow further system integration, common terms to allow interoperability of AI systems in complex contexts need to be agreed, particularly in the cross-domain use. A goal-based regulatory assessment and a standard to document transparency should be developed as a pre-condition for the deployment of AI systems.

Pre-normative research is needed to assess ethics implications in complex systems and test for hidden discrimination and protected variables.

## 4 Data quality needs and practice in selected sectors

The AI Act sets out a horizontal framework to avoid fragmentation of the Digital Single Market and ensure harmonisation of provisions on AI across different sectors. This choice is thoroughly analysed in the accompanying impact assessment (European Commission 2021). Typically the same technology is characterised by the same problems and risks to fundamental rights (e.g. autonomy, data dependency, opacity etc.), irrespective of whether the AI system is developed by a public or private entity and irrespective of the sector where the system is deployed.

Therefore, horizontal, cross-cutting standards are needed to ensure a level playing field and to avoid inconsistencies in how the same AI applications are regulated. Within such an approach, it is important that certain risks which are specific to certain sectors are properly considered in the context of the standardisation process supporting the future AI Act. This would ensure that the development of horizontal standards serves well the needs of the different sectors.

Hence, the second block of parallel sessions of the workshop focused on different sectors considered as being at high-risk under the AI Act:

- Medicine and healthcare
- Law enforcement
- Finance
- Education and employment
- Industrial automation and robotics

While not being classified as high-risk under the AI Act, the sector of media, including social media, content moderation and recommender systems need also to be considered, taking into account its societal relevance and the fact that it is being made object of some obligations under the AI Act and the Digital Service Act.

### 4.1 Education and Employment

In recent years there has been a steep increase in the global economic importance of AI, as reflected by more than doubling from 2020 to 2021 of private investments into AI and a 30-fold increase of the total number of AI patents since 2015 (Stanford University 2022). Moreover, the overwhelming majority of top results across technical benchmarks rely on using more training data, highlighting the importance of data and ethical standards for data for AI (Stanford University 2022). The increasing deployment of AI in various economic sectors is expected to also have notable effects on education and employment. Indeed, many advancements in AI, which are mostly driven by rapid progress of machine learning and its subfields, are represented by applications that automate work-related tasks and consequently affect the employment sector (Tolan, et al. 2021). Beyond that, these processes influence the education sector through two mechanisms. On the one hand, students need to be prepared for new skill sets for increasingly automated economies and societies. Secondly, it is expected that AI and related technological advancements are utilised to transform and enhance educational processes in the classroom and at the system level (Vincent-Lancrin and van der Vlies 2020).

The AI market of the education sector is expected to exhibit large growth rates.<sup>8</sup> AI has the ability to transform the way traditional educational systems work, increase organisation productivity, and encourage teachers and students of all abilities. In personalised education, AI assists in determining what a student knows and does not know. Based on this information, technology helps in creating a tailored study schedule that considers the knowledge gaps of learners. Intelligent teaching assistants will allow scholars to access guidance whenever they need without recurring to valuable teacher time. Thus, in this way, AI-based tools can help create better conditions where more learners have access to high quality and skills-focused education. Furthermore, AI technology offers opportunities and options for students that are unable to attend school due to varying reasons. Although there has always been a wealth of data in education, such as grades or administrative statistics on student absenteeism, the use of data to improve student learning, teacher instruction, and decision-making in educational administrations is relatively new as stakeholders in education swing back and forth between enthusiasm and scepticism about the trustworthy use of data for AI in education (OECD 2021).

---

<sup>8</sup><https://www.emergenresearch.com/industry-report/artificial-intelligence-in-the-education-sector-market>

The rapid advances in AI are not only affecting the labour market through the automation of tasks at the workplace but also causing a rapid increase of new forms of work such as online labour markets and digital platforms (Urzi Brancati, Pesole and Fernandez Macias 2020, Duch-Brown, et al. 2022), and an expansion of algorithmic management practices (Baiocco, et al. 2022). The benefits include improved business performance and working and living conditions. However, there are also concerns for the future of human work and employment. Algorithmic bias and reinforcement of (gender) stereotypes and exploitation of monitoring/surveillance possibilities are only few of many concerns. Finally, there are some fears that with machines improving their performance beyond human levels, human jobs may be at risk causing massive increases in unemployment.

During the workshop a panel of experts discussed the opportunities offered by standardisation to tackle bottlenecks and to identify gaps for pre-normative research. The panel was composed by Dee Masters, Barrister at the law office Cloisters in the UK, focusing on AI, Discrimination and Employment; Nikoleta Giannoutsou from the JRC with a focus digitalisation in education and Enrique Fernández-Macías from the JRC with a focus on emerging technologies and employment.

#### **4.1.1 State of the art, challenges and ongoing standardisation activities**

One specific aspect to the sectors of education and employment, is that they embed culture-dependent social relationships – between employers and employees, between teachers and students – with socially internalised definitions of the role that each agent plays in that relationship. Implementing AI applications into these structures and systems creates the AI-specific challenges around transparency but also risks affecting these relationships, which could create power imbalances that were not there before, or reinforce existing ones. These power imbalances also occur because of the unique issues of opacity and complexity that come with AI applications. Therefore, transparency is identified as a core concept to achieve inclusive, non-biased and trustworthy data for AI.

Key aspects of transparency are access to data and algorithms and the explainability of these algorithms, where different transparency standards would have to be set for data, algorithms and for the algorithms in education and employment.

Transparency of data use addresses questions such as *what* data is used, *how* data is used, and *what decisions* are made based on this data. There is a concern that large amounts of data are being collected about employees or students and processed to make decisions about them, without them being adequately informed. This causes issues around obfuscated data-ownership structures and infringements on privacy.

Transparency of algorithms is related to two issues. One is navigating the trade-off between allowing for access to the algorithm for third party audits and protecting intellectual property rights of the algorithm. The second issue deals with the problem that the increasing complexity of particularly deep learning algorithms inhibits their interpretability by non-experts. There is a generalised level of opacity which makes it hard to challenge decisions that are made algorithmically. Complexity also means that workers fail to understand how decisions are being made. They lack the tools, skills and competences to understand this, and this decreases their power with respect to the entity that controls the algorithm. Finally, transparency about the rules of the algorithm requires defined boundaries on what roles AI systems play in the workplace and in schools.

A related challenge is that of observability. Technically, it is possible to process data on students or workers and deploy corresponding AI applications without their involvement or even without their knowledge. However, the collection and storage of data and the use of algorithms should always be observable to affected stakeholders. Observability is key to the agency of affected stakeholders and empowers them in the protection of their fundamental rights to non-discrimination and privacy.

A further issue is the lack of inclusivity as datasets and algorithms struggle to represent and accommodate the needs and behaviours of groups that are outside the majority group which is considered the norm. AI relies mainly on pattern recognition and correlations which can only be observed with enough data. This works well for majority groups for which large amounts of data can be generated. But it is harder to accommodate and adjust AI for minority groups that look, speak and generally behave differently, which can perpetuate participation barriers for instance for disabled people.

There are issues about autonomy at work as workers who are subject to algorithmic management often refer to a feeling of lack of autonomy. There is a generalised feeling that the increasing use of algorithms in work places tends to generate power imbalances. Some of the power equilibria that have been reached over many years of work history are now being challenged by these new practices of algorithmic management. One

example of a system being challenged is collective bargaining and worker unionisation as workplaces are becoming decentralised through algorithmic management via digital platforms.

Finally, another challenge is related to the lack of a clear and common terminology that would enable cross-disciplinary and cross-sectoral collaboration on these challenges. This also concerns the definition of acceptable performance criteria and quantitatively measurable thresholds for standardisation.

#### **4.1.2 Pre-normative research gaps and standardisation opportunities**

The speakers proposed several potential measures to address these challenges.

To address the issue of opacity and complexity, one proposed solution was an individual right to understandable and personalised explanations from algorithmic decisions on a case-by-case basis. This explanation should contain information on what data has been looked at, how particular aspects have been weighted and how a particular conclusion has been reached. This would not only address the problem of opacity but also observability, as a comprehensive personalised explanation implicitly informs affected stakeholders when they are subjected to algorithmic decision making and consequently may nudge them to challenge these decisions.






To address the challenge of underrepresentation of potentially marginalised groups, the experts suggested further research on understanding this issue. That is, there is a need for understanding how AI tools treat for instance disabled people. This concerns the representativeness of the dataset, and if reasonable, accommodations can be made.

To address the issue of power imbalances, several data and AI governance mechanisms were proposed. This included a right for data reciprocity, transparency and interpretability. That is, employees, workers and students need to have rights to access the data that is being used to build AI that affects them. Further, social dialogue mechanisms between employers and employees, between AI providers, schools and teachers were proposed, in which employees and students should be given the right to co-determine the implementation of algorithms in workplaces and schools through representatives and support of external expert audits on bias evaluation and explainability. Another suggested measure to achieve control and regulation is the open-source model, which is a decentralised way of algorithm evaluation based on transparency and open dialogue. An additional proposed measure was the creation of mandatory external data trusts for workers' representatives and students to grant access to their data in a controlled way that respects data protection regulations but also allows for controlled access for audits.

All speakers agreed that there is a need for the setting of clear boundaries (i.e. red lines) for the use of algorithms in circumstances where infringement on fundamental rights could not be avoided by the setting of standards.

When conducting the exercise of mapping challenges and solutions to a matrix that considers the development stages of data for AI systems to standardisation concepts, it became clear that the most important and most urgent areas that needed to be addressed are Terminology and Measurement/Metrology. For instance, in the brainstorming session, transparency was identified as a core concept to achieve non-biased inclusive and trustworthy data for AI. However, it became clear that transparency in itself is a multifaceted and complex concept.

**Table 3 Overview of pre-normative research- and standards needs in selected standardisation categories and along the AI value chain of education and employment**

|   |  Terminology  |  Metrology  |  Performance Characterisation   |  Compatibility   |  Regulatory assessment                |
|---|--|--|--|---|--|
| <b>AI system deployment &amp; marketing</b> <ul style="list-style-type: none"> <li>– Regulatory assessment</li> <li>– Users</li> <li>– Transparency &amp; specification</li> <li>– Accountability/responsibility</li> <li>– Maintenance, post-market follow-up &amp; bias monitoring</li> <li>– Supply network</li> </ul> |  |  | <ul style="list-style-type: none"> <li>– Lack of autonomy</li> <li>– Trade-off privacy vs openness</li> <li>– Does AI do what it says?</li> </ul>                                | <ul style="list-style-type: none"> <li>– Transparency in decision making</li> <li>– General user instructions</li> <li>– Cross-user-standardisation</li> <li>– User training</li> </ul> | <ul style="list-style-type: none"> <li>– Deployment conditioned on explainability</li> <li>– Tech maintenance</li> </ul> |
| <b>AI system creation and production</b> <ul style="list-style-type: none"> <li>– Data sets &amp; algorithms (incl. bias)/models</li> <li>– Cybersecurity</li> <li>– System design &amp; integration</li> <li>– Upscaling &amp; evaluation</li> <li>– Quality control</li> </ul>  | <ul style="list-style-type: none"> <li>– Privacy</li> <li>– Transparency</li> </ul>  | <ul style="list-style-type: none"> <li>– Personalised explanations</li> <li>– Complexity</li> </ul>  | <ul style="list-style-type: none"> <li>– Hidden discrimination</li> <li>– Open source model</li> <li>– Power imbalance in information asymmetry</li> </ul>                       | <ul style="list-style-type: none"> <li>– Implementation of domain expertise</li> <li>– Transparency about data use</li> </ul>   | <ul style="list-style-type: none"> <li>– Participatory evaluation</li> </ul>   |
| <b>Data creation</b> <ul style="list-style-type: none"> <li>– Compilation, preparation, bias testing</li> <li>– Analysis, processing, labelling</li> <li>– Licensing &amp; restrictions</li> <li>– Sharing &amp; marketing</li> </ul>   | <ul style="list-style-type: none"> <li>– No clear common terminology</li> <li>– What data on workers?</li> <li>– Fuzzy fluid terminology</li> <li>– Education observability (model) vs limitations and importance</li> <li>– Stakeholder identification</li> </ul> | <ul style="list-style-type: none"> <li>– Representation of minorities in data</li> <li>– Accuracy</li> <li>– No established metric</li> <li>– Multi-perspective/sources data collaboration</li> <li>– Research on implications to identify key parameters</li> </ul> | <ul style="list-style-type: none"> <li>– Education observability (model) vs limitations and importance</li> <li>– Accuracy and completeness of representation in data</li> </ul> | –   | <ul style="list-style-type: none"> <li>– External data trust</li> </ul>  |

### 4.1.3 Prioritisation and conclusions

Consequently, one very important, urgent and potentially easy challenge to address, is to set a clear common terminology on AI and data-related matters, such as bias vs. discrimination (which has different meanings in tech or the legal context), transparency, and accuracy.

Another important and potentially feasible challenge is to draw red lines for certain tasks not to be performed by an AI alone, such as when assessing the performance of workers or students. The AI Act already sets a good example for that with its risk-based approach to regulating AI systems. Simultaneously, there is a need for personalised explanations to address the black box characteristic of many deep learning algorithms.

The the most important but also difficult challenge is to achieve a satisfactory level of transparency that fulfils all aspects that are related to this core concept.

## 4.2 Law enforcement and the public sector

The use of artificial intelligence in the law enforcement and public sector is especially sensitive, as its applications might have important consequences, ranging from legal (e.g. the arrest of a person, serve as a proof of crime in courts) to social (e.g. the surveillance of citizens in a public area, deny a public subsidy). AI systems must be trustworthy, and ensure an adequate use and governance of data throughout their lifecycle. It is also of the utmost importance to demonstrate their benefits to both the civil society and the law enforcement agencies (LEAs) in charge of using them. For example, multiple recent incidents with facial processing technologies causing discriminatory outcomes and privacy invasion (The New York Times 2020, Hupont, Tolan and Gunes, The landscape of facial processing applications in the context of the European AI Act and the development of trustworthy systems. 2022) have painted a highly negative picture of this technology. The parallel session on “*law enforcement and the public sector*” focused on analysing how standards could help to advance trustworthy AI in this sector, with a particular focus on data.

Four speakers participated in this parallel session: Patrick Grother from the US National Institute of Standards and Technology; Javier Rodríguez Saeta from the Spanish face recognition company Herta; Robin Allen lawyer at Cloisters and judge of the Crown Court (UK); and Rosalía Machín Prieto from the Spanish Ministry of Interior.

In the following we summarise the main discussions, findings and recommendations in terms of standardisation opportunities that were raised in the panel.

#### **4.2.1 State of the art, challenges and ongoing standardisation activities**

Law enforcement and the public sector are one of the most critical application areas for AI. The European proposal for the regulation of AI (European Commission 2021) contemplates these areas as “high risk” under certain intended uses, such as the remote biometric identification of people, the prediction of occurrence of a criminal offence, the detection of the emotional state of a person, and crime analytics. Having good quality data to train and evaluate AI systems for law enforcement is essential and, beyond these activities that are generally carried out off-line, it is also key to guarantee good performance and governance practices during operation and throughout the whole system’s lifecycle.

The panel identified the following main challenges in the sector that would need urgent attention:

- *Operational vs in-lab data gaps.* Ideally, the data used for the development of AI tools for law enforcement should be representative and consider all the real-world situations that might be encountered during operation. However, the deployment of these systems typically involves very different sensors: capturing data in real-time (e.g. cameras, scanners, fingerprint readers, X-rays), environmental conditions (e.g. lighting, weather conditions) and persons (e.g. in terms of ethnicity, geographical origin, age, gender, technical skills). It is therefore difficult to contemplate all the possible future operational scenarios at the AI system development stage. Moreover, collecting and annotating a large amount of real-world data is not always feasible in this sector for many different reasons, ranging from privacy issues (e.g. personal, biometric, sensitive data) to scarcity of data, e.g. few positive samples for deception detection, as in (Quijano-Sánchez, et al. 2018). For that reason, it is common that systems are developed based on simulated data, data generated ad-hoc in a lab, small datasets or data scrapped from the web (e.g. celebrity faces for training face recognition systems). As a result, the ecological validity of these systems is not guaranteed, and the process of transferring them from the lab to operational setting frequently comes at the cost of robustness, accuracy and trustworthiness.
- *The demographic bias problem.* While the whole AI field is strongly impacted by data biases (c.f. Section 3.1), this problem is particularly harmful in the law enforcement and public sector. For example, recent research has demonstrated that the performance of biometric systems used by LEAs vary greatly across demographics (Hupont and Fernández, Demogpairs: Quantifying the impact of demographic imbalance in deep face recognition 2019, Grother, Ngan and Hanaoka 2019, Wang and Deng 2021). Their accuracy typically decays for women, elder people and children, and this effect is higher in ethnicities that are under-represented in the training data (e.g. African, Asian for western algorithms, and Eastern Europe for some Chinese algorithms). This effect happens even with good quality data, data imbalance being the central problem. Additionally, there is no consensus in the field on how demographics should be modelled at the data level (e.g. using ethnic categories vs phenotypical descriptions such as skin colour tone), which makes the task of measuring biases even more challenging.
- *Obtaining data from trustworthy sources.* As mentioned above, the task of collecting and annotating data to train and validate AI systems for law enforcement is extremely sensitive, tedious, time-consuming and not always possible. For this reason, AI developers working for this sector frequently rely on datasets provided by third parties. Different types of data providers might come into play: datasets might be shared by administration or LEAs themselves through private agreements, publicly released by academic research groups (Williford, May and Byrne 2020, Sánchez, et al. 2020), or purchased to private companies specialised in the creation of datasets. In particular, the latter practice is becoming increasingly popular in the field, which could be of concern as there is currently no regulation or standard means of certifying the source of origin and reliability of the data.
- *Lack of explainability.* The increasing use of AI as a “black-box” approach has many benefits, such as an increase of robustness and accuracy, but the drawback is the loss of explainability and interpretability. This has negative implications for the trustworthiness of a system, something essential in law

enforcement and public sector contexts where users should be given informed knowledge to understand the system's outputs and decisions. For instance, an accused person needs to have access to AI systems used by LEAs to be able to challenge their use and accuracy (Phillips and Przybocki 2020). A way to tackle this problem is to equip black-box models with some explainability mechanisms (Guidotti, et al. 2018) (e.g. visualisations or approximations to simpler interpretable models), but the field of explainable AI is still in its infancy and more research is needed especially for this critical sector.

- *Understanding AI in the law enforcement context.* Police and judges are not yet adequately trained in what AI can and cannot do to help them in their daily duties, and what the benefits and issues are. It is also difficult for juries to fully understand AI-generated proofs presented in courts. More training in the field, with special emphasis on human rights and discrimination issues, would help to bring AI closer to these audiences and prevent eventual misuses/misinterpretations of AI systems' outputs.

Some standardisation and benchmarking efforts have been carried out in the last decade, but few of them tackle the previous challenges comprehensively. The following three are those considered to advance the state-of-the-art in the matter:

- *The NIST's biometric evaluation benchmarks.* The NIST's "Face Recognition Vendor Test" (FRVT<sup>9</sup>) is the world's largest evaluation benchmark for face recognition algorithms. Every year, tens of leading commercial vendors and research groups worldwide send their algorithms for evaluation, leading to a periodic publication of results that serves as a *de-facto* standard in the field. A large collection of evaluation metrics is presented, including a thorough assessment of demographic factors<sup>10</sup> and image quality issues<sup>11</sup>. The FRVT benchmark is composed by more than 100 million face photographs with well-balanced age/sex/ethnicity metadata, which makes it unique in the field. Another outstanding initiative implemented by the NIST is the "Biometric Technology Rally"<sup>12</sup>, which evaluates the performance of biometric technologies in a scenario that emulates a real airport where small groups of free-flowing people pass through a gate where a biometric system is operating. A large-scale recruitment of volunteers is carried out for the Rally, well-represented in terms of demographics, to evaluate not only technical aspects of the system, but also user experience and ethics.
- *ISO/IEC WD 19795-10.* The ISO/IEC 19795 is a standard on "Information technology — Biometric performance testing and reporting". It provides general principles for testing the performance of biometric systems and specifies performance metrics, requirements for the recording of test data, and requirements on test protocols. Its parts 2 "Testing methodologies for technology and scenario evaluation" and 6 "Operational testing" contribute to bridge the *operational vs in-lab* data gap. Part 10 "Quantifying biometric system performance variation across demographic groups", still under development, focus on the demographic bias problem.
- *CETSE Innovation Lab*<sup>13</sup>. The Spanish Ministry of Interior put into service in 2016 the Security Technology Centre ("Centro Tecnológico de Seguridad" - CETSE). The centre fosters the collaboration between LEAs, industry, administration and universities to test and develop the latest technology for law enforcement, including AI systems and data spaces. In the last years, they have put a special focus on AI and successfully deployed ground-breaking AI solutions at a very large scale. They are currently involved in the Starlight<sup>14</sup> European project ("Sustainable Autonomy and Resilience for LEAs using AI against High-priority Threats") where more than 40 key European stakeholders – including LEAs, technological companies and academic institutions – collaborate in the definition and analysis of AI-driven law enforcement scenarios.

#### 4.2.2 Pre-normative research gaps and standardisation opportunities

The above activities are undoubtedly pioneering the standardisation of good practices for a trustworthy adoption of AI in the law enforcement and public sector. However, many standardisation opportunities remain open as follows.

---

<sup>9</sup> <https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt>

<sup>10</sup> [https://pages.nist.gov/frvt/html/frvt\\_demographics.html](https://pages.nist.gov/frvt/html/frvt_demographics.html)

<sup>11</sup> [https://pages.nist.gov/frvt/html/frvt\\_quality.html](https://pages.nist.gov/frvt/html/frvt_quality.html)

<sup>12</sup> <https://www.dhs.gov/science-and-technology/biometric-technology-rally>

<sup>13</sup> <https://cetse.ses.mir.es/publico/cetse/idi.html>

<sup>14</sup> Starlight project. <https://www.starlight-h2020.eu/about>

It is first important to highlight that most current efforts focus on biometric applications exclusively, i.e., on applications involving the identification of individuals (e.g. face, fingerprint, iris recognition). While it is true that biometrics is one of the most well-established and mature AI-driven technologies in the field, there are many other applications that are gaining more and more adoption. Examples include tools for the prediction of recidivism, deception (Quijano-Sánchez, et al. 2018), financial fraud, control of disinformation (Martín, et al 2021), immigration (e.g. detection of unfaithful VISA applications), non-verbal behaviour analysis (Hupont and Chetouani, Region-based facial representation for real-time action units intensity detection across datasets 2019) and automatic voice transcription in courts/interrogatories (Negrão and Domingues 2021). There is therefore a need to broaden the focus beyond biometrics.

Considering this need to expand standardisation efforts beyond biometric systems, the speakers identified an additional set of important gaps to be bridged which are closely related to the challenges identified in the previous section. The sector would benefit from standardisation on operational testing/benchmarking methods and data interoperability, both in terms of technical aspects (e.g. metrics, data formats) and processes (e.g. evaluation protocols, communications, exchange of information). Related to data interoperability, having data spaces for LEAs is considered central for fostering trustworthy data sharing practices, not only among different LEAs, but also between LEAs and non-LEAs, both at national and international level.






Another problem considered particularly critical in the field is bias. Our speakers raised that this problem could be mitigated by reaching consensus on redlines and metrics to monitor bias throughout the whole AI system’s lifecycle. Standard guidelines on explainability would also contribute fighting the bias problem by providing meaningful explanations demonstrating that a decision is fair and non-discriminatory. This links to the need for a standard on AI-based evidence in courts, explicitly covering which type of evidence is to be accepted and which one is not, and under which circumstances.

Finally, there are two additional considerations that should be covered in the standards for this sector:

- Standards should take into account that they will be mainly adopted by SMEs. The speakers agreed that requirements for SMEs should be proportional with respect to those required to large companies.
- Standards should include processes for AI and human rights training for key users of the system (LEAs, judges, juries).

The following matrix summarises this discussion regarding open standardisation opportunities.

**Table 4 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the law enforcement and the public sector**

|  |  Terminology |  Metrology |  Performance Characterisation |  Compatibility |  Regulatory assessment                   |
|--|---|---|--|---|---|
| <b>AI system deployment &amp; marketing</b><br>– Regulatory assessment<br>– Users<br>– Transparency/specification<br>– Accountability/responsibility<br>– Maintenance, post-market follow-up & bias monitoring<br>– Supply network |   |   | – Operational testing methods and benchmarks.  |   | – Training LEAs on human rights issues.<br>– Training police and judges on AI.<br>– Take into account SMEs (proportionality). |
| <b>AI system creation and production</b><br>– Data sets & algorithms (incl. bias)/models<br>– Cybersecurity<br>– System design & integration<br>– Upscaling & evaluation<br>– Quality control                                      | – Common taxonomy on terms related to bias  | – Metrics and benchmarks for [data+model] bias evaluations.                                   |  | – Standards on interoperability.<br>– Standards on communications and cybersecurity.                | – Standards for evidences (e.g. in courts) and explainability.  |
| <b>Data creation</b><br>– Compilation, preparation, bias testing<br>– Analysis, processing, labelling<br>– Licensing & restrictions<br>– Sharing & marketing   |   | – Standardised metrics for biases in data.  | – Obtain data from trustworthy organisations   | – Standardised data spaces for LEAs.  | – Regulate data provider companies.   |



### **4.2.3 Prioritisation and conclusions**

Given the critical nature of the law enforcement and public sectors, and the rapid evolution of AI technologies, all the identified standardisation gaps would ideally need to be covered within the next 5-year timeframe. Nevertheless, the panel members identified three of them as top priority. The first one is the need for training LEAs on AI and human rights. A training program could be implemented with relatively little effort and in the short term, and would help ensuring that AI systems are used in a trustworthy way, taking advantage of their benefits but also considering their shortcomings. The other two standardisation needs that are deemed of high priority but that would be harder to operationalise due to the current lack of consensus on the matter are standards for bias measurement and mitigation, and standards for interoperability (LEAs-LEAs, LEAS-non LEAs) both at the data, communication and AI system level.

## **4.3 Finance**

Artificial intelligence in the AI sector has varying degrees of risk. It ranges from high risks such as essential private and public services, i.e. credit scoring denying citizen's opportunity to obtain a loan, to limited risks that refer to AI systems with specific transparency obligations, i.e. when using AI systems on financial information, where users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back.

A panel formed by three distinguished experts provided a brief overview of the state of art of AI in the finance sector before entering into a detailed discussion on data quality requirements and standardisation needs in the domain: Karen Croxson, Financial Conduct Authority of the United Kingdom (FCA); Andrea Caccia, chair CEN-CENELEC JTC 19 Block chain; Jörg Osterrieder, Zurich University of Applied Sciences.

### **4.3.1 State of the art, challenges and ongoing standardisation activities**

Financial services have to protect consumers, promote effective competition, and enhance integrity and resilience of the financial system. The regulatory landscape relies on standards for data quality, but a model is necessary to enable conformity assessment. Suitability of standards on data quality models should therefore be assessed for AI specific challenges. New standardisation activities should rely on existing work and not create AI specific models, to avoid unnecessary costs and complexity. Finance specific requirements should then leverage on this common approach to data quality.

The main goal of FCA in the UK is to regulate financial services, with the objective of protecting consumers, promote effective competition and enhance integrity and resilience of the financial system. Potential issues in AI for finance include, inter alia: algorithm bias potentially leading to discriminatory decisions worsening outcomes groups or exploitation of behavioural biases to identify and exploit consumers. Firms must have a good robust governance structure and be accountable for outcomes irrespective of technologies deployed. There is also a need for further debate to better understand the potential benefits and harms from AI. Further clarity is also needed on how current regulatory framework applies and how we can best support further safe AI adoption. There is a need for knowledge sharing and collaboration across regulators and opportunities to share the views of stakeholders. FCA is part of the Digital Regulation Cooperation Forum in the UK and will publish a Discussion Paper on AI jointly with the Bank of England later this year.






Standards for data quality models and measurement already exist, however their suitability for AI is not directly assessed. Data quality models should not be AI specific to avoid unnecessary compliance costs and data used for AI should be part of a bigger discussion on use of data in general – quality models should not be only AI dependent. The foreseen AI regulatory landscape relies on standards for data quality, but a model is necessary to avoid a qualitative approach that enables conformity assessment. The suitability of standards on data quality models should be assessed for AI specific challenges and new standardisation activities should rely on existing work and not create AI specific models to avoid unnecessary complexity. Finance specific requirements should leverage on this common approach to data quality.

The following topics were also addressed: the need to address data quality issues (accountability mechanism for data quality and model quality); define terminology for standard setting purposes (standards, guidelines, best practices to check for fairness and non/bias); consider algorithm auditing; consider a societal scientific agenda to look at how bias comes into decision making, independent of AI; and test model outcomes in the different stages to ensure that the right outcomes are obtained.

### 4.3.2 Pre-normative research gaps and standardisation opportunities

Standards for data quality models and measurements already exist, however their suitability for AI is not directly assessed. Data quality models should not be AI specific to avoid unnecessary compliance costs. The data used for AI is part of a bigger discussion on the use of data in general – quality models should not be only AI dependent. It is important to consider: the need for a societal scientific agenda to look at how bias comes into decision making, independent of AI; the need to test model outcomes in the different stages to ensure that the right outcomes are obtained.

**Table 5 Overview of pre-normative research- and standard needs in selected standardisation categories and along the AI value chain of the finance sector**

|  |  Terminology  |  Metrology                |  Performance Characterisation |  Compatibility |  Regulatory assessment |
|--|--|--|--|---|---|
| <b>AI system deployment &amp; marketing</b><br>– Regulatory assessment<br>– Users<br>– Transparency & specification<br>– Accountability/responsibility<br>– Maintenance, post-market follow-up & bias monitoring<br>– Supply network |  |  |  |   |   |
| <b>AI system creation and production</b><br>– Data sets & algorithms (incl. bias)/models<br>– Cybersecurity<br>– System design & integration<br>– Upscaling & evaluation<br>– Quality control  | – Big data creates more risk of bad data – risks related to poor data quality can translate into prudential risks.<br>– Define fairness, what does it mean in the context. | – Model for data quality useable for all sectors.<br>– Common approach based on existing standards needed. | – Testing model outcomes vs data quality check.<br>– We need different models for each application.<br>–       | –   | –   |
| <b>Data creation</b><br>– Compilation, preparation, bias testing<br>– Analysis, processing, labelling<br>– Licensing & restrictions<br>– Sharing & marketing   | – Identify AI/ finance challenges  | – Data representability.<br>– Training data is source of bias<br>–   | – Data provenance and characteristics  | –   | – Accountability for quality of data/model  |

### 4.3.3 Prioritisation and conclusions

Priorities addressing data quality issues in the financial sector concern the need for common approaches for accountability mechanisms for data quality and model quality; to define terminology for standard setting purposes (standards, guidelines, best practices to check for fairness and non/bias) and to consider algorithm auditing.

## 4.4 AI for media, including social media, content moderation, recommender systems

The session on AI for media, including social media, content moderation and recommender session, consisted in carrying out a technology assessment carried by two panellist: Symeon Papadopoulos from the Centre for Research and Technology Hellas and Jochen Leidner from the Coburg University.

### 4.4.1 State of the art, challenges and ongoing standardisation activities

The session on AI for media, including social media, content moderation and recommender systems linked the initiatives contemplated in the AI Act with two other legislative initiatives proposed by the European Commission: the Digital Services Act (DSA) and the Digital Markets Act (DMA). These two initiatives are meant to upgrade rules governing digital services in the EU in order to create a safer and more open digital space. The Commission made the proposals in December 2020 and on 25 March 2022. A political agreement was reached on the Digital Markets Act, and on 23 April 2022 on the Digital Services Act.

The DSA and DMA have two main goals:

- to create a safer digital space in which the fundamental rights of all users of digital services are protected;
- to establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally.<sup>15</sup>

In the context of AI applications using media, social media data employed in the media, or social media for content moderation or recommendation, it is important to assess the data quality aspects and the potential bias in the data employed for training. Content moderation can help eliminate offensive and toxic content, fake content and other negative types of content. At the same time, it should not limit access to a diversity of perspectives and opinions. Recommender systems, while they help to sift through large quantities of information to content that is relevant to the user, can also lead to focus on a single perspective and a decrease in the variety of information potentially available to users. Additionally, current business models of social media platforms rely on obtaining and maintaining users' attention for long periods of time. In some contexts, it has been shown that content that is suggested in these contexts becomes more emotionally charged and more polarised, leading to echo chambers and only intra-group feedback loops (Huszár, et al. 2022, Duan, et al. 2022). This may pose threats such as radicalisation, spread of mis- and disinformation, with potentially very dangerous consequences. Finally, social media data and large quantities of web data are used in training Natural Language Processing (NLP) deep learning models. Due to the heterogeneity of the content and users contributing to it, as well as their inherent subjectivity, such data is prone to different types of bias: gender, political affiliation, racial, topic-based, religious, etc. It is therefore important that the quality of the data, the bias it may contain are assessed and mitigated, so that NLP models that are trained on such data do not inherit and perpetuate the bias it contains (Sun, et al. 2019, Deven Santosh Shah, Schwartz and Hovy 2020, Czarnowska, Vyas and Shah 2021).

Starting from these initial reflections, both invited participants opened the session by stressing upon the need to define the notion of bias in AI when applied to media content and the importance of looking into the characteristics of data employed in content moderation and recommender systems. Whereas for the latter personal characteristics of users are paramount to the quality of the system and it is desirable for AI systems to be "biased" towards personal preferences, for the first, bias in moderation can lead to polarisation and an information bubble effect. This in turn can have serious consequences with regard to issues such as pluriperspectivism and objectivity of information in media, and lead to more critical issues such as disinformation and misinformation. A common observation was also that there are large differences in access to data between companies that own social media platforms in comparison to other industry players, as well as Academia and research institutions. This may be harmful for two reasons: firstly, unequal access can lead to smaller players not being able to create AI models that are as powerful and applicable. Secondly, lack of access to relevant datasets that can be used for benchmarking purposes can also make the task of assessing bias presence and bias mitigation very difficult, as no relevant data is available to train and test relevant models.

It was stressed that particular attention should be given to the link between individual and group fairness and the difficulty to quantify bias. In this context, it was underlined the fact that it is important to address the issue in a general manner and not dependent on the domain task.

In the same line of argumentation, it was underlined that access to relevant datasets (as size and content) is a key to understanding the potential bias present in the data, define it in a systematic manner and develop adequate methods to detect and mitigate it. This process would have to be done in accordance to the General Data Protection Regulation and privacy regulations. An appropriate definition is required in order to avoid concept drift. Given the complexity of the process and the various aspects it involves, the suggestion is to involve interdisciplinary teams to address this issue. The next steps suggested are the creation of relevant (as size and content) benchmarking datasets that are available to the entire research community and can be employed to test different types of bias. This should be followed by the development of methods and tools for auditing datasets in accordance to the aspects identified in the analysis of relevant datasets and the definition of corresponding guidelines on compliance. Both the private sector, as well as Academia and other research institutions should be involved in the development of methods and tools for bias detection and mitigation.

Media and news are sensitive domains as far as AI is concerned, given the potential risk of public perception manipulation. Quality assurance in data is a complex issue, as bias in data employed in AI systems for media

---






<sup>15</sup> <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

can have critical consequences. Standardisation should take into account existing best practices and find solutions as well as develop adequate new ones. The concept of good data also requires an adequate definition, to take into account quality issues, but also link data to its value when employed in AI applications.

Issues related to data access, particularly in the context of social media platforms data, which are available to companies owning the social platforms and only in small quantities to others, are also important. In the context of media applications of AI, privacy protection law should be at the forefront of technological development. Data quality and ethics audits need to be improved and encouraged with more use of open sources. Efforts to promote fairness, ethics and privacy protection are jurisdiction bound, but there is a general consensus on the need to protect these principles world-wide.

Finally, all discussants agreed that normative and standardisation acts should not hinder the capacity for innovation

**Table 6 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the media sector**

|   |  Terminology  |  Metrology   |  Performance Characterisation                                     |  Compatibility |  Regulatory assessment  |
|---|--|---|--|---|--|
| <b>AI system deployment &amp; marketing</b> <ul style="list-style-type: none"> <li>– Regulatory assessment</li> <li>– Users</li> <li>– Transparency &amp; specification</li> <li>– Accountability/responsibility</li> <li>– Maintenance, post-market follow-up &amp; bias monitoring</li> <li>– Supply network</li> </ul> | <ul style="list-style-type: none"> <li>– Define suitability of AI and data in the context of AI for media applications</li> </ul>                                  |   |  |   | <ul style="list-style-type: none"> <li>– Define new business models for data-based business</li> </ul>   |
| <b>AI system creation and production</b> <ul style="list-style-type: none"> <li>– Data sets &amp; algorithms (incl. bias)/models</li> <li>– Cybersecurity</li> <li>– System design &amp; integration</li> <li>– Upscaling &amp; evaluation</li> <li>– Quality control</li> </ul>  | <ul style="list-style-type: none"> <li>– Define fairness in group versus individual –focused applications</li> </ul>   | <ul style="list-style-type: none"> <li>– Ensure access to data from social media platforms and other relevant data sources</li> </ul>   | <ul style="list-style-type: none"> <li>– Create benchmarking models for bias detection and mitigation open to entire research community</li> </ul> |   | <ul style="list-style-type: none"> <li>– Create benchmarking models for auditing open to entire research community</li> </ul>                          |
| <b>Data creation</b> <ul style="list-style-type: none"> <li>– Compilation, preparation, bias testing</li> <li>– Analysis, processing, labelling</li> <li>– Licensing &amp; restrictions</li> <li>– Sharing &amp; marketing</li> </ul>   | <ul style="list-style-type: none"> <li>– Define terminology for data quality in AI for media</li> <li>– Define aspects for data quality in AI for media</li> </ul> | <ul style="list-style-type: none"> <li>– Create benchmarking datasets that are open for training and tuning to entire research community</li> <li>– Create labelling system for data provenance and data quality – compliance, representativeness aspects in line with privacy</li> </ul> |  |   | <ul style="list-style-type: none"> <li>– Create mechanisms for access to relevant data to be used for standardisation and auditing purposes</li> </ul> |

#### 4.4.2 Pre-normative research gaps and standardisation opportunities

Among the previous standardisation opportunities, the panel members identified four of them as high priority. Two of them could strongly leverage on existing pre-normative initiatives and be implemented with relatively small effort, namely 1) the definition of terminology regarding quality and fairness for data used in AI systems for media applications; 2) the definition of a checklist for aspects to be assessed for dataset quality; 3) the development of mechanisms to ensure open data access to relevant quantities of quality data for the training, testing and auditing of AI systems used in the media context; and 4) the creation of open benchmarking datasets and models for bias detection in AI applications used in the media context. Additionally, the definition of a new business model that is in line with ethical aspects, privacy and data quality is also an aspect that is particularly important in the context of AI systems for content moderation and

recommendations, but also more generally, AI systems for contexts where information shared and user engagement currently translate into revenue.

#### **4.4.3 Prioritisation and conclusions**

Concluding on the discussions of the panel, the main priorities and suggested next steps are:

- Creating mechanisms to ensure access to relevant media data to be used for bias aspect definitions, model training, testing and auditing; the creation of relevant benchmarking datasets and models to test, mitigate and audit potential bias in AI applications used in media content moderation and recommender systems.
- Definition of terminology for data quality in the context of AI systems used in media content moderation and recommender systems.
- Definition of relevant aspects of data quality in the context of AI systems for media, both for recommender systems (for individual versus collective use), as well as for content moderation.
- Definition of new business models corresponding to the new types of data-based businesses, incentivising revenue that is based on value obtained from use and not from user engagement time.

#### **4.5 Medicine and Healthcare**

It appears inevitable that AI applications will, within the coming years and decades, fundamentally change the way diseases and conditions are diagnosed and treated, the way patients are cared for in hospitals or in their own homes, and how we gain understanding of disease pathways, risk factors and effective novel (precision) therapies (see for instance (Rajpurkar, et al. 2022)). This is not surprising given the strong information and data dependence of medicine and healthcare, for example to understand intricate biomedical pathways relevant for pathogenesis in view of drug and therapy development or for managing complex decision flows in daily clinical practice, ideally tailored to individual, personalised needs. While there is a lot of public attention on specific AI applications (e.g. autonomous driving or intelligent personal assistants on smart phones), awareness of the variety of possible AI use cases in the health sector appears to be restricted mainly to the medical community and are not yet widely debated in society – despite their considerable ethical and socioeconomic impact. The number of AI applications in healthcare is illustrated by a recent ISO/IEC report which examined real-world use cases (either under development or already on the market) and which sampled the highest number of use cases in the healthcare sector (ISO/IEC 2021).

##### **4.5.1 State of the art, challenges and ongoing standardisation activities**

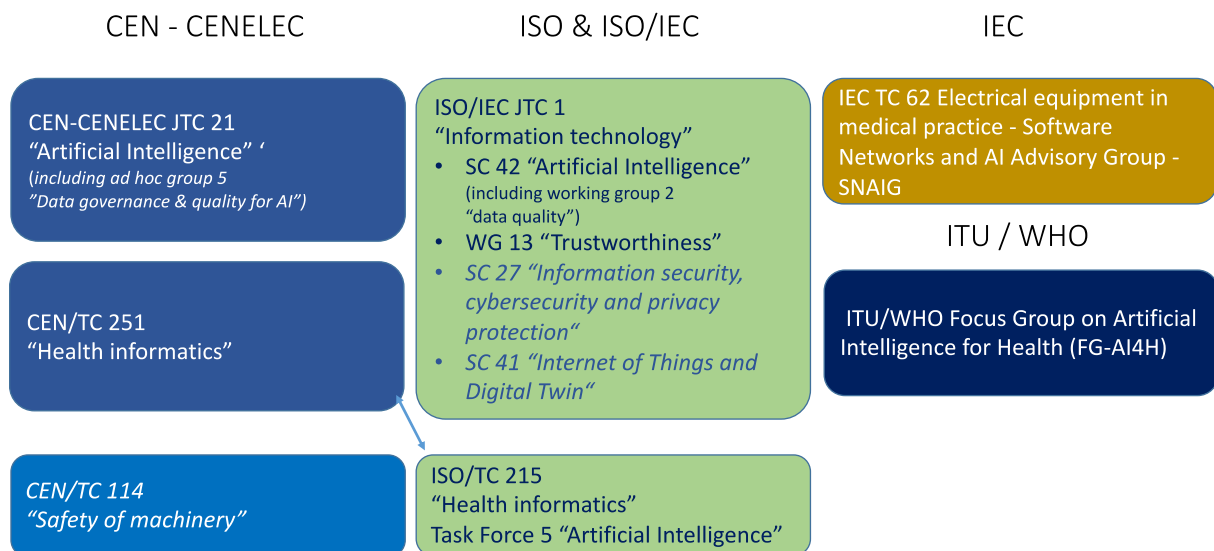
In the recent World Health Organisation guidance document on ethics and governance of AI for health (World Health Organisation 2021), AI health applications are discussed in the context of four categories. These seem helpful for structuring discussions on technical standardisation and priority setting of standardisation needs to satisfy regulatory requirements (e.g. conformity assessment). The categories are: 1) *healthcare* (e.g. diagnosis and prediction-based diagnosis, clinical care including risk identification and therapy optimisation, robotic surgery, health-wearable technology including closed-loop systems, decision-support systems), 2) *health system management* (e.g. administrative workflows, logistics), 3) *public health and public health surveillance* (e.g. monitoring of disease outbreaks, pandemic preparedness, health promotion), 4) *biomedical research* (e.g. drug repositioning, genomic medicine, big data exploitation including through anonymised or pseudonymised electronic health records). While some AI systems in these categories have already reached the market (see for instance list of devices published online by FDA (US Food and Drug Administration 2021)) and/or are used routinely (including in-house developments of companies), many others are still in the early developmental stage.

Many AI systems falling into above four categories would benefit from standards and guidance. Standardisation efforts from an EU perspective are particularly important for products that need to undergo conformity assessment (CE marking) prior to being placed on the market and which need to be subject to continuous market surveillance by the manufacturer. In such cases manufacturers, notified bodies and authorities, require (ideally EU harmonised) standards for assessing whether the essential legal requirements are fulfilled. Health-relevant products that may feature AI systems and fall under the “EU new approach to technical harmonisation and standards” (European Union 1985) involve conformity assessment based on EU harmonised standards. Importantly, under the draft EU AI Regulation (“AI Act” COM/2021/206 final (European Commission 2021), products that are regulated under specific legislations and which also require conformity assessment in agreement with applicable rules, are considered “high-risk”. These legislations are listed in

Annex II of the draft AI Act. Thus, this notion of “high-risk” is not derived from a criteria-based risk-benefit assessment, but based on reference to conformity assessment under sectorial product-specific legislation. Notably, many medical devices do not require conformity assessment involving a notified body since they pose low risks to patients and users and such products would not be considered under the draft AI Act. It remains to be seen how and if the final version of the AI Act will address risks of AI-based health products. Under the Medical Devices Regulation, risk classification of software (and thus AI systems) is outlined under rule 11 (Annex VIII, Chapter III) with risk classes (and hence associated conformity assessment requirements) ranging from I, IIa, IIb to III, depending on the intended purpose. Moreover, under the Medical Devices Regulation software is deemed an ‘active device’ (Article 2, definition 4) pointing to risk rules 9 and 10. Most medical device software is class IIa or higher and requires notified body involvement for the conformity assessment.

Standards, guidance and technical documents would help ensure effective and consistent conformity assessment, outlining clear benchmarks for developers in support of innovation. From an EU perspective, relevant standards should encapsulate the principles of trustworthiness of AI systems outlined by the Commission’s high-level expert group in 2019 (European Commission 2019). These principles apply to the entire product development cycle, i.e. from model design and training, over up-scaled production to placing the product on the market including post-market follow-up. These principles still need to be transposed into *actionable* standards. A first step in this direction may be the development of focused guidelines as tackled by the “Future AI” project (FUTURE-AI 2021). In addition, the emerging ‘standards landscape’ needs to cater eventually for a wide variety of sectors and use applications. In this context, CEN/CENELEC has provided a first roadmap in 2020, listing relevant definitions, ongoing standardisation activities, a use-case submission form to sample AI applications and proposed standardisation items (CEN-CENELEC 2020).

**Figure 3 A selection of international formal standardisation organisation groups and committees in the area of AI, health informatics and medical equipment. The arrow between the health informatics groups indicates that ISO/IEC and CEN/CENELEC groups are regularly exchanging information. In the context of this workshop, it is important to note that CEN-CENELEC JTC 21 has an ad hoc group on “data governance and quality for AI”, while ISO/IEC’s SC42 on AI has a dedicated working group on data quality. Other relevant standardisation groups are not shown (e.g. IEEE or VDE).**



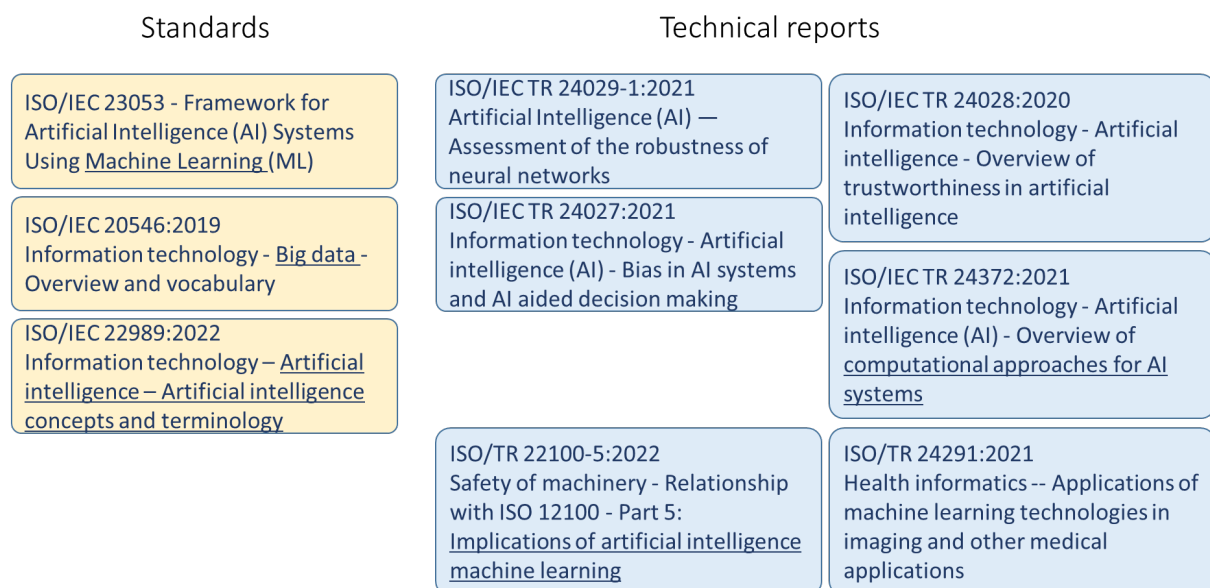
Against this background, the panel on AI in medicine and healthcare discussed the state of the art of standardisation with a view on healthcare and in particular data quality, exploring potential gaps regarding standards, guidance and technical reports as well as the types of standards needed (e.g. metrology, terminology, etc.) over the product development cycle.

The panel was composed of experts in artificial intelligence technology, medical devices software, standardisation and the science-policy interface, with expertise in legislative implementation and regulatory pathways in health technologies, biotechnology and toxicology. Alpo Värri is Research Director at Tampere University, Finland; Member and convenor of CEN/TC 251 on health informatics, Working Group II “technology and applications”; member of standardisation groups/committees of ISO and IEEE, and JTC1/SC42 dealing with AI; his research includes systems for pattern recognition in medicine. Koen Cobbaert is Senior Manager at

Philips in the area of “Quality, Standards & Regulations”; he is a standardisation expert in CEN/CENELEC JTC 21 “Artificial Intelligence” as well as a number of ISO Technical Committees on artificial intelligence, health informatics, quality management of medical devices and electrical equipment in medical practice. Thorsten Prinz works for VDE Health Germany and is consulting medical device manufacturers with AI-based software products regarding regulatory requirements of the medical devices regulation and on the emerging EU AI Regulation (“AI Act”); he is supporting the preparation of technical documentations for AI-based software medical devices and the implementation of quality management system processes for AI model development and evaluation. Sandra Coecke is project leader at the JRC and working on global harmonisation (OECD) of cell and tissue culture methods as well as promoting mathematical and AI models for health areas including food safety and pathogenic threats.

Following an introduction and brief discussion of relevant groups involved in standardisation (Figure 3) as well as on published documents (Figure 4), the panellists set out their views and reflections on challenges, solutions and possible next steps concerning data quality and standardisation of AI in medicine and healthcare.

**Figure 4 Relevant ISO/IEC publications (standards and technical reports) regarding both horizontal aspects of AI (e.g. robustness, bias, machine learning = ML) and health specific aspects (i.e. ML applications for imaging and other medical applications).**



At the outset of the session, all panellists presented their reflections on key issues in regard to AI standards and standardisation for healthcare applications:

Standardisation activities play a critical role in harvesting the potential opportunities of AI in healthcare and standards, while addressing trustworthiness and ensuring patient and user safety, should not stifle innovations and the EU's strong competitive position in regard to health technologies. Coordination between standardisation organisations, the scientific and regulatory community will be critical. The establishment of a joint working group AI in the summer of 2022 was a positive development that enabled analytics for health informatics between JTC1/SC42 and ISO/TC215. Attention should be given to innovations along the entire evidence pathway: researchers are often unaware of requirements, terminology and tools needed for bringing products to the market. Bridging various communities will be a key element for developing useful and readily applicable standards, particularly in a multidisciplinary field such as healthcare. A clear framework for risk assessment of AI systems used in healthcare may help clarifying and perhaps stratifying requirements under sectorial Regulation, supporting an innovation-friendly approach. Guidance documents developed through a multi-stakeholder process should be considered first-line solutions to enhance clarity, ensuring global impact in both high- and low-to-medium income countries. Given that the statistical and data processing foundations underlying AI systems are largely application-independent, standardisation should first focus on horizontal standards, e.g. on trustworthiness including data quality requirements, bias avoidance, intelligibility etc. In contrast, 'vertical' guidance could address the needs of specific sectors/use cases or applications. Finally, as an important next step, a pragmatic roadmap on the required standards, reports and guidance documents

should be developed by the standardisation community, ideally in close dialogue with the regulatory community.

There is a growing body of scientific literature identifying bias as one of the key challenges when deploying AI systems in healthcare. Issues with bias might undermine the acceptability of AI-supported healthcare solutions. This includes bias leading to healthcare inequity as exemplified by AI systems for cancer treatment or for COVID-19 diagnostics. Standards providing clear frameworks for avoiding bias will thus be key enablers for the safe use of AI. The number and complexity of relevant legislations that need to be considered when bringing an AI system to the market may be a huge challenge for AI developers and, in particular, small and medium sized enterprises. Specifically, the general data protection Regulation (GDPR = Regulation (EU) 2016/679) is reported as a potential roadblock to the market deployment of apps. In general, the requirements should be kept as simple as necessary in order to make the transition from research results into products as little burdensome as possible. Concerning the tackling of bias ongoing efforts include the TRIPOD-AI reporting guideline and the PROBAST-AI risk of bias tool for diagnostic and prognostic prediction models using AI (Collins, et al. 2015). TRIPOD-AI contains a check-list proposed to support a consistent approach for describing and possibly reducing bias. It may also help communicate residual bias in a transparent manner. In addition, reviewed and adopted 'quality labels' for data and algorithms used in AI might be a pragmatic solution. As next steps, it was identified the upcoming standardisation request (via a Commission implementing decision) which would set a clear framework for standardisation work. Standards would be required in approximately 10 areas. The CEN technical committee 251 on health informatics is currently reflecting on producing vertical healthcare-specific AI standards, complementing horizontal standards by CEN/CENELEC joint technical committee 21 on artificial intelligence. The latter group may adopt or adapt standards produced by ISO/IEC as appropriate.

Data to assess bias, representativeness and robustness are often identifiable and hence subject to legislated data protection standards that vary internationally. Developers/providers often lack access to suitable/sufficient data. Specific provisions in the draft EU AI Act regarding data accessibility may be problematic. The draft Act requires that authorities and notified bodies be given access to the entire data set (training, development/validation, testing). However, developers often have no access to the data that resides with patients or healthcare institutions and may thus not be in position to provide data access. Data are often shielded (e.g. data vaults, federated learning). The data quantity may be so vast that a physical transfer would result in high economical cost and carbon footprint (example: GPT-3 based AI systems). Finally, third country data protection legislation may prohibit data transfers. The requirement in the AI Act could thus have the effect that EU-produced AI-systems would not be able to use such data and that developers would have to resort to potentially smaller and less representative data sets (e.g. sub-populations, rare diseases, etc.) with negative effects on robustness. Bringing algorithms to the data is sometimes also not ideal, given the technical limitations of doing this at a sufficient scale. Concerning standardisation priorities, a balanced approach would be required concerning data use and the need to provide access to data sets; further, sub-group selection should be addressed; importantly, there should be method standards on establishing data metrics and describing data, in particular for raw (sensitive) data. In case the provider has no direct access to the data, the relevant third party should be able to provide, based on a standard, a sufficiently detailed description of the data and allow identification of bias/discrimination and need for further training and to warn users concerning bias sources. Such metrics might allow authorities and notified bodies to establish compliance, i.e. conformity with the essential requirements. To do this appropriately, the standard(s) should include inter alia an overview of possible metrics, measurement methodologies, selection of appropriate metrics, their thresholds and confidence levels for a statistically robust description.

There is a need for sufficient transparency: black box situations would undermine trust in AI systems and hence hinder effective deployment and marketing. Human agency is important: there are strong expectations concerning AI systems' capacity to address a variety of tasks. Sufficient consideration should therefore be given to the interface between an AI system and humans, ensuring maximum exploitation of the AI systems capacity while retaining human control and human agency in their use. Bias reduction would, in applications like medicine and health, require a multidisciplinary approach; examples include pre-clinical efficacy or toxicological assessment processes. Standards would be required to address bias and prejudiced assumptions both in algorithm development (algorithmic bias) and training data selection: AI has an enormous potential for medicine and healthcare, but, essentially, all data are generated by people – models output quality is limited by the quality of these data and addressing biases of gender, ethnicity, etc. should be the top priority for standardisation in view of creating inclusive solutions. As shown by the PSIS organ-on-a-chip example, multidisciplinary interaction and stakeholder collaboration are essential to achieve appropriate standards,



overcoming potential barriers of technology-to-market. Finally, new computational techniques (e.g. quantum computing) may be needed to tackle more complex problems in medicine, healthcare and life sciences.

There is a lack of sector-specific standards addressing a variety of parameters which would need to be taken into consideration when assessing data quality and appropriateness for a given application. These principles importantly include absence of bias but also balancedness of the data, their completeness, correctness, currentness, the inter- and intra-data set consistency, representativeness and, more importantly, the independence of training and testing data sets. Finally, the manufacturer should perform data profiling using statistical methods that demonstrates the appropriateness of the data for the intended purpose (e.g., with respect to the disease and patient population and the size of the data sets used). As a solution to address this “quality catalogue”, state-of-the-art measures of the properties listed above should be outlined in a data management document, which should form a central element of the AI-model development process and be part of the quality management system of the developer/manufacturer. Novel AI standards would need to fit into the landscape of already established standards in the health sector so as to avoid duplication, overlap and or confusion due to contradictory requirements and guidance. In the area of medical devices, relevant standards such as IEC 62304 and IEC 82304-1 should be considered, for example in regard to the safety classification of the software. In the interest of retaining an innovation-friendly approach, AI standards for AI systems falling under the MDR or the IVDR would need to be sufficiently clear while avoiding to further enhance the already high requirements set-out by the respective legislations.

#### **4.5.2 Pre-normative research gaps and standardisation opportunities**

To kick-start the panel discussion, four guiding questions were put forward by the moderator:

- What are the key challenges to be addressed to arrive at useful standards?
- What aspects of standardisation and guidance would need to be tackled?
- Can the diversity of application cases in the health sector be appropriately served by horizontal standards or are other elements needed?
- What role could guidance documents play in the standardisation process?

##### *a) Data set properties and data quality*

The following key elements were identified by the panel in relation to data properties and data quality:

*Terminology, data description, characterisation and (statistical) metrics:* There is a need for guidance and/or standards on data description and characterisation, including notably data statistics, metrics and symmetry of datasets. Terminology for describing data in a consistent manner was identified as a fundamental requirement.

*Modularisation of data quality aspects:* To support a qualitative and quantitative description of data quality in view of the intended purpose, a modularisation of data properties was outlined, which would need to be elaborated in terms of appropriate metrics. These data quality ‘modules’ consist of inter alia:

- *Legal compliance* of the data with applicable requirements;
- *Completeness and correctness:* are the data in the set indeed complete or are there elements missing and are the data correct (to avoid propagating errors of data transmission, compilation)?;
- *Currentness* (i.e. are the data sufficiently up-to-date to allow meaningful training of models?);
- *Inter and intra-data set consistency:* is there consistency within and between datasets (in particular of key variables/parameters), in particular where these have been derived by aggregating data automatically and/or manually from different sources and/or by different data specialists;
- *Representativeness:* are the data representative in view of the use case, e.g. risk, age stratification etc.;
- *Balancedness:* are the data balanced in view of achieving statistically meaningful desired (predictive) outcomes?

Both balancedness and representativeness are closely related to avoid bias.

- *Avoidance of bias*: tools for describing, measuring, and avoiding different forms of bias to the extent possible or, at least, be conscious and transparent about residual bias. Any residual bias stemming from the data employed should be described in the purpose, applicability and limitations of the final AI system.
- Is there a clear and conscious *separation of training and testing data*? Notably, this independence needs to stretch from initial development over internal improvement/validation stages to production and post-market monitoring/improvements.

*Anonymisation and pseudonymisation*: In view of provisions of data privacy and relevant legislation, the panel emphasised the need for clarity on how to achieve anonymisation and/or pseudonymisation, including identification of potential off-the-shelf solutions that could be used by developers. Importantly, such efforts should not obfuscate the collection of key parameters needed to understand and measure bias, balancedness and representativeness.

*Data set size*: The panel agreed that methodological guidance would be helpful on how to estimate data set size in view of the intended purpose, taking into account inter alia desired performance parameters (e.g. accuracy, confidence intervals) deemed sufficient for the given purpose. Potential real-world risk(s) posed, may also have to be factored into estimations of data size.

#### b) *Identification of standards*

When considering needs for standards and guidance over the development cycle of AI systems, the panellists agreed on the following items:

- In the “metrology” domain of standards, guidance and/or standards would be required to address state-of-the-art measurement protocols for defining metrics and (where feasible) thresholds. This would be required for the model / AI-system development stage, for the production stage where models are up-scaled and for the marketing stage including post-market surveillance / follow-up.
- In the “performance characterisation” domain, the panel identified a need for comprehensive guidance for developers/manufacturers on how to document all relevant aspects that lead to the development of a given AI system. Such guidance should outline all data quality elements (see detailed aspects identified above), relevant aspects during model creation and model training (including avoidance of algorithmic bias) and include also aspects of intelligibility (i.e. explainability and transparency), making also use of data measurement and description standards for developers and pure third party data providers. The latter was deemed important in cases where data cannot be physically moved, but where algorithms are brought to the data (residing on platforms, ‘data vaults’), e.g. through federated learning. Clear guidance on how to describe the purpose, use scenario(s), limitations and applicability of the AI system would enhance consistency, conscientiousness and transparency for developers and users alike.

In view of strengthening evidence on robustness and generalisability, such guidance should also address the reporting of external validation of AI systems, emphasising the utility of leveraging external data for assessing and confirming the utility of the system.

- Another important aspect highlighted related to standards and guidance on security considerations during AI system creation, including security aspects relating to general cybersecurity but also adversarial attacks and other AI-model specific threats that may tilt models into providing incorrect or undesired outputs.
- Considering the requirement of an adequate risk analysis and risk management of medical devices including AI systems regulated under the EU regulation on medical devices and conformity assessment options for in vitro diagnostic medical devices, the applicability of existing relevant risk management standards to AI specific risks would need to be ascertained (e.g. ISO 14971 or ISO/IEC 23894) as well as that of relevant guidance documents issued by notified bodies/standardisation bodies (e.g. BS 34971).

### **4.5.3 Prioritisation and conclusions**

To conclude the session, the panel agreed on key priorities for the next steps of standardisation.

On a general level, the panel agreed on three key priorities:

- i. *Horizontal standards and possible complementary guidance*: A priority for experts in standardisation and health will be the assessment of emerging ‘horizontal’ standards on AI in regard to their applicability and

sufficiency for health-related applications (in particular for products requiring conformity assessment). Where necessary, such assessment may lead to a gap analysis outlining needs for sector-specific guidance which could complement horizontal standards. There was agreement among the panellists that standard development should focus on horizontal aspects common to all AI applications irrespective of the sector. At the same time, specific sectorial needs could first be addressed through guidance on how to interpret horizontal standards for a given sectorial application (e.g. an AI-system for breast cancer diagnosis based on supervised machine learning using labelled data).






- ii. *As simple as possible, as detailed as necessary:* The panel agreed that, in general, the requirements set out in standards and guidance should be kept as simple as possible in order to limit any unnecessary burden when transiting from research results into products.
- iii. *Draw on existing documents and guidance:* The panel suggested that standardisation organisation should consider relevant guidance documents issued by other actors, specifically where such guidance relates to products subject to conformity assessment in relation to standards. For example the German Notified Bodies Alliance Questionnaire "Artificial Intelligence (AI) in medical devices" (German Notified Bodies Alliance 2022) lists relevant aspects to be considered for conformity assessment.

On a technical level, the panel identified four standardisation requirements as particularly important. These are shown in Table 7 which provides a matrix of the development pathway on an AI system (vertical) versus the types of standards (horizontal), capturing the priorities below:

- a. *Terminology:* as a fundamental requirement for effective communication between communities and experts as well as users of standards and guidance, the development of appropriate and sufficiently comprehensive terminology was determined by the panel as a priority. Terminology should address all elements along the development pathway, i.e. from data creation over AI system creation to AI system deployment and marketing. Recently ISO/IEC has published the standard. 22989:2022 on AI concepts and terminology which is expected to be endorsed by CEN-CENELE JTC 21. Further, there is ongoing work at IEEE concerning AI terminologies.
- b. *Guidance for measuring data quality:* Development of guidance or standards on measuring and describing quality and appropriateness of data in view of a given purpose, including statistical aspects. This should include aspects of avoidance of bias and discrimination as well as, where feasible, personalisation of devices. This aspect may be particularly relevant for the health sector, where devices may need to be tailored to individual needs and parameters. A comprehensive and modular data description and metrology standard could moreover lead to the development of standardised and widely accepted 'quality labels' for data ("data hygiene certificate", see (European Parliament 2020)). Such guidance or standards should address all three stages of the development pathway and could be linked to a guidance on describing AI systems (see below).
- c. *Guidance on how to document the essential elements of the AI system:* AI system developers would benefit from guidance for description of the essential elements of the final AI system, once being deployed/marketed. This should include the applicability domain and limitations of the system, potential non-applicability (to avoid risks from off-label use) as well as residual risks. Further, a sufficiently clear description of data quality including aspects of non-bias, bias mitigation and inclusiveness would support transparency, intelligibility and explainability. This could take the shape of a "data hygiene certificate", drawing on a modular approach for data quality description.
- d. *Guidance on safety and (cyber)security:* Guidance on how to evaluate safety, effectiveness, risks for AI systems used for health(care) purposes would be required. Further, the use of data involves vulnerabilities, namely to cyber-attacks. Adversarial 'data poisoning' could skew the data set at the stage of data creation and AI system development (i.e. model training based on data). This could lead to severe adverse outcomes for users of AI systems that were subject to such attacks. Thus, guidance is needed to support appropriate measures that minimise such risks.

The following matrix contrasts pre-normative research and standardisation needs of medicare AI systems along their development stages versus selected standardisation domains. The matrix shows the four key priorities for guidance and/or standards identified by the panel. Guidance on methodologies for data measurement protocols ("metrology domain") would support a robust description of AI systems once deployed (orange arrows).

**Table 7 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the medicine and healthcare sector**

|  |  Terminology |  Metrology              |  Performance Characterisation |  Compatibility  |  Regulatory assessment                                    |
|--|---|--|--|--|--|
| <b>AI system deployment &amp; marketing</b><br>– Regulatory assessment<br>– Users<br>– Transparency & specification<br>– Accountability/responsibility<br>– Maintenance, post-market follow-up & bias monitoring<br>– Supply network | ✓   | – Methods to establish state of the art data measurement protocols, data metrics and relevant thresholds | –  | – Guidance for documentation of AI system as deployed/marketed: applicability, limitations, risks, data description/data hygiene certificate | – Guidance for documentation of AI system as deployed/marketed: applicability, limitations, risks, data description/data hygiene certificate |
| <b>AI system creation and production</b><br>– Data sets & algorithms (incl. bias)/models<br>– Cybersecurity<br>– System design & integration<br>– Upscaling & evaluation<br>– Quality control  | ✓   | – Methods to establish state of the art data measurement protocols, data metrics and relevant thresholds | – Safety & cybersecurity aspects   | –  | –  |
| <b>Data creation</b><br>– Compilation, preparation, bias testing<br>– Analysis, processing, labelling<br>– Licensing & restrictions<br>– Sharing & marketing   | ✓   | – Methods to establish state of the art data measurement protocols, data metrics and relevant thresholds | – Safety & cybersecurity aspects   | –  | –  |

Applications of AI in medicine and healthcare hold an enormous promise to accelerate biomedical research, facilitate patient care, healthcare and clinical pathways and may facilitate gains regarding hospital administration and public health surveillance. With the healthcare sector under strain (e.g. pandemics, climate change, demographic developments), AI systems could be used for routine aspects in clinical practice, thus freeing healthcare workers of such tasks and allowing them to focus stronger on patient interaction or critical diagnostic and therapeutic decision making (keyword human agency), drawing on outputs provided by AI systems. AI systems may also enable personalisation of devices with enormous potential for patient well-being and considerable gains in quality of life.

For this vision to become reality, standards will play a crucial role. The panel agreed that horizontal standards addressing general aspects of statistical models used in AI systems should be assessed for their suitability for healthcare applications, with a focus on products requiring conformity assessment to gain market access. In case horizontal standards are not sufficient, specific healthcare-related guidance could complement the standard, explaining and expanding on specific aspects that cannot be made explicit in a horizontal standard aimed to support a variety of sectors.

The panel identified top priorities for standardisation: (I.) terminology to facility community-bridging and adoption and usefulness of standards by model developers/manufacturers, (II.) guidance on how to describe the critical elements of an AI model, including clear indications on metrics and statistical measures for data quality and appropriateness for a given purpose, (III.) gap analysis of horizontal standards in view of identifying needs for healthcare-specific standards.

#### 4.6 AI for Industrial Automation and Robotics

According to the latest figures of the World Robotics Report 2020 published by the International Federation of Robots (IFR), in 2019, there was a record of 2.7 million industrial robots operating in factories around the world – an increase of 12% in comparison with 2018. Asia remains the strongest market for industrial robots with China reaching 783,000 while active robots in the UK are just reaching 21,700 units. According to the same report, the robot density, measured as robots installed per 10,000 employees, Singapore, followed by South Korea and Japan show the highest shares. In Europe, the countries with the highest density are Germany, Sweden and Denmark, which rank after Japan at world level. As announced by the IFR, their

upcoming report on industrial robots will show an increase of those numbers for 2021, with 3 million industrial robots worldwide<sup>16</sup>.

In addition to industrial robots, service robots, both domestic/personal and professional ones, have been advancing rapidly, along with advances in computer vision, speech recognition, navigation systems and artificial intelligence (Lemaignan, et al. 2017). The relevant technical developments have made service robots able to interact seamlessly with humans and integrate in the human social environments, by perceiving the surrounding world and acting upon it. In this reciprocal process of perception and action, artificial intelligence programs can use data from the real world acquired through robots to improve their performance. Applications of service robots appear in multiple fields, such as in healthcare, agriculture, defence, education and entertainment. Most of service robots have unique designs and different degrees of automation – from full tele-operation to fully autonomous operation. Hence, the service robot segment is more diverse than the industrial robot segment (Duch-Brown, Rossetti and Haarbuerger 2021, Charisi, et al. 2021, Lee 2021).

In this context, although AI systems are making their mark on automation and robotics, mainly to manage the variability and unpredictability of the physical environment, the developments of AI in automation and robotics appears in a slower pace and in a narrower field of application than in non-embodied systems. As indicated at the latest report by IFR on Artificial Intelligence and Robotics, robots are complex systems that combine a number of hardware and software components typically integrated with other systems<sup>17</sup>. The complexity of the systems in combination with the often-complex environments robots exist require the establishment of concrete frameworks for data collection and management especially for long-term autonomy (Kunze, et al. 2018).

As such, standardisation in industrial automation and robotics has a relatively long tradition. The standard ISO13849-1:2015, for example, provides safety requirements and guidance on the principles for the design and integration of safety-related parts of control systems (SRP/CS), including the design of software<sup>18</sup>. The ISO10218-1:2011 standard specifies requirements and guidelines for the inherent safe design of a robotic system and refers to industrial robots only; however, the safety principles established with this standard can apply to non-industrial robots, as well<sup>19</sup>. The ISO/AWI PAS 5672 standard refers to robots, as collaborative applications and they test methods for measuring forces and pressures in quasi-static and transient contacts between robots and humans<sup>20</sup>. Lastly, the series of ISO/TC 299 standards in Robotics aim to ensure safety in industrial but also in service robots, excluding toys and military robots<sup>21</sup>. Similarly, the IEEE Standards Association in reference to the IEEE Robotics and Automation Society have a series of approved standards for robotics including the latest one IEEE 7007-2021 Ontological standard for ethically driven robotics and automation systems<sup>22</sup>.

The present session on AI in industrial automation and robotics brought together experts from relevant disciplines with the following goals:

- To present current and future needs and recommendations to address ethical concerns in the context of AI, especially in Industrial Automation and Robotics and the future AI Act;
- To map the existing and missing standardisation efforts;
- To indicate guidelines regarding data quality standards for AI models; and
- To propose and recommend steps to start or complement the process of drafting standards.

Against this background, the panel on AI for Industrial Automation and Robotics discussed the latest developments on data quality in the area of robotics and identified potential gaps regarding standardisation

---

<sup>16</sup> <https://ifr.org/ifr-press-releases/news/robot-sales-rise-again>

<sup>17</sup> [https://ifr.org/downloads/hidden/IFR\\_Artificial\\_Intelligence\\_in\\_Robotics\\_Position\\_Paper\\_V02.pdf?utm\\_source=CleverReach&utm\\_medium=email&utm\\_campaign=Paper+Download&utm\\_content=Mailing\\_12323895](https://ifr.org/downloads/hidden/IFR_Artificial_Intelligence_in_Robotics_Position_Paper_V02.pdf?utm_source=CleverReach&utm_medium=email&utm_campaign=Paper+Download&utm_content=Mailing_12323895)

<sup>18</sup> <https://www.iso.org/obp/ui/#iso:std:iso:13849:-1:ed-3:v1:en>

<sup>19</sup> <https://www.iso.org/standard/51330.html>

<sup>20</sup> <https://www.iso.org/standard/82488.html>

<sup>21</sup> <https://committee.iso.org/home/tc299>

<sup>22</sup> <https://standards.ieee.org/ieee/7007/7070/>

over the product development cycle while exchanged ideas on aspects of prioritisation of standardisation activities.

The panel was composed of experts in artificial intelligence technology: Abdil Amjad (Product Safety & Regulatory Engineer at Thermo Fisher Scientific in the Netherlands) works in field of product safety and regulatory with a specialisation in safety design of equipment in SEMI and robotics. He has worked on defining safety design requirements for hardware and software within the machinery sector, and on global regulatory requirements for wireless electrical and electronic systems. Aurélie Clodic (Research Engineer at the Laboratory for Analysis and Architecture of Systems and at the ANITI in Toulouse, France) works on human-robot collaborative task achievement and on robotics architecture design with a focused on decision-making and supervision dedicated to HRI. Emmanuel Kahembwe (Chief Executive Officer at Association for Electrical Electronic Information Technologies VDE UK/NI), has an expertise in the fields of AI, Robotics, Autonomous Systems, High Performance Systems and Games Development. He is the CEO of VDE UK, part of the VDE group and a member the ART/1 Committee on Artificial Intelligence at the British Standards Institution (BSI). Roland Behrens (Senior Scientist at Fraunhofer IFF in Germany) focuses on issues relating to the automated security clearance of collaborative robots and on validated models that accurately reflect the biomechanical response behaviour of humans to robot collisions.

#### **4.6.1 State of the art, challenges and ongoing standardisation activities**

Machine-learning algorithms that power robotic applications are advancing rapidly. Initially, algorithms had to be trained on large number of images of each target object in order to recognise it correctly. However, this is a time-consuming process. Therefore, significant development effort has been directed by AI software and systems providers to enable algorithms to learn with fewer, or none, tagged examples, which is achieved with semi-supervised or self-supervised methods. The knowledge accrued by the system in their applications, and by the end-user, naturally expands over time, reducing the training of an AI-based robot with the goal to enable the algorithm to quickly generalise from what it has already learned, applying existing knowledge to recognizing and manipulating new objects. However, currently, this does not imply that AI-driven robotic applications can be easily transferred from one environment to another.

As such the field of industrial automation and robotics has attracted special attention regarding standardisation. As mentioned in the previous paragraph there are several initiatives of standardisation that are specific to robotics. For example, the series of standards ISO/TC 299 Robotics have currently 37 standards published or being under development<sup>23</sup>. However, for inclusive, non-biased and trustworthy robotic systems there is a need for consideration of the quality of data that are collected for the training of a system as well as the management of the data a robot that functions in the physical environment collects.

This session started with a short presentation of the state-of-the-art on industrial robotics and automation by the four panellists.

Addressing the lack of unbiased data sets and data sets to support an AI safety product development which can be verified is important due to the uncertainties regarding the impact of the system on the relevant social structures when interaction with machine become the norm. The session also discussed the lack of newer verification and validation frameworks with regards to AI safety and implementations and the challenges with respect to safety assessment and verification due to close Human-Robot Interaction. In addition, there is also a lack of technical knowledge to assess AI safety. The introduction of guidelines and methodologies or frameworks for AI safety is of paramount importance. It was proposed to consider hybrid standardisation referencing and extracting methodology of other standards and training of workforce for collaborative efforts. Lastly, it was also proposed the preparation of data sets and the contribution towards regulatory committees for preparation and maintenance of data.

Concerning human-robot interaction, there are three main challenges in relation to standardisation in industrial automation and robotics. First, there is a need for special attention to data acquisition. For example, personal robots would be able to acquire data about people that directly interact with it but also data from the surrounding social environment. In such cases there needs to be mechanisms for selection of data according to the needs of the system and design systems that are able to forget, meaning that they would be deleted. Second, it is important of taking care of data sharing. Again, by using the example of the personal

---

<sup>23</sup> <https://www.iso.org/committee/5915511/x/catalogue/>

robot, the data that are collected by the robot during the interaction with a specific human might be sensitive and the human might not be willing to share. In that sense, the context should be taken into account in data collection. Lastly, it is also possible the manipulation of the human by the robot, which is of special concern to vulnerable populations such as children. In such cases, transparency is catalytic for the protection of the human agency. For the above-mentioned challenges, it was emphasised the need of training the professional academic sector for the creation of a pool of experts that would work together towards the mitigation of these challenges. This professional education could lead to the acquisition of a certification in order to maximise the conditions for transparency. Lastly, the education of the general public is needed regarding their interaction with AI-based social robots. This can be done through formal education in order to ensure that a minimum level of AI and robotics literacy is achieved for the general population.

There are concerns about the lack of awareness regarding standardisation in industrial automation and robotics, resulting on the need for an easily accessible portal to all available and applicable AI for Robotics standards. Safety is one of the most challenging aspects to address in physical robots; however, safety in robotics is still challenging to test and certify since there is a need for comparable performance and safety metrics and certificates of safety (proof of behaviour) for data-driven methods. As such, it was proposed that model-free data driven AI methods for robotics would require simulation testing, and simulation-to-real approaches. For this, infrastructure with standard simulators and testing environments are needed which would also facilitate automated testing approaches.

Concerning the safety in robotics with the use of AI-empowered sensors it was proposed model-based safety validation with the boundaries of existing standards (Saenz, et al. 2020). In this context it was suggested the introduction of “smart safety” with instantaneous certification of “micro”actions before their execution. Towards this direction, it was highlighted the need for accelerating standardisation procedures with agile digital standards that would replace the current tradition in standardisation with updates that take multiple years to achieve. For this, there is a need of clear and robust metrics to validate the reliability of AI-empowered safety sensors and the development of more model-based approaches such as in risk assessments.

#### **4.6.2 Pre-normative research gaps and standardisation opportunities**

During the session, one of the main concerns regarding standardisation was the issue of assessing and certifying the system’s robustness and human safety. This was discussed across all the three levels of the life cycle of product development, including data creation, the development of models and the deployment and operation of the resulting systems.






Technical specification of methods for ensuring system robustness are expected to play a key role, including those based on data-driven methods or model-based validation approaches. Simulation is considered a valuable tool that can facilitate the process of the assessment of the system that can in certain cases complement testing in the physical environment. Independent of the assessment method employed, the user should be in the loop, which often results in the combination of objective as well as subjective qualitative assessments. Smart safety was mentioned as one of the proposed methods, including the use of model-based methods to check whether robot actions and plans are within safety boundaries before they are executed. Special attention was given on the consideration of human fundamental rights especially for vulnerable populations in all the stages of the robot production life cycle when developing standards for Robotics and human-robot-interaction.

When categorising the value chain of AI systems, the data creation stage would be the first. For this stage the speakers emphasised the fact that the challenges start even with data collection. Physical robots that navigate the human environment might collect data that are not directly connected with the initial intended purpose, raising issues of privacy. This requires techniques for data cleaning and curation but also ways to eliminate unnecessary data according to existing regulations such as General Data Protection Regulation. Additional standards and technical specifications should support AI providers in applying effective mitigations for data bias and ensure the representativeness of usually under-represented populations.

The second stage along the AI systems value chain encompasses the AI system creation and production. In this stage the speakers referred to the need for metrics for the validation of the robustness and reliability of AI-robotic systems. For this, they proposed the creation and adoption of validation frameworks for AI robotics including, AI-based safety components. For the development of such frameworks, there is a need for common taxonomies and terminologies. There is a need to examine previously developed frameworks and the creation of novel ones and emphasise on the definition of objective measures to quantitatively assess the reliability of AI-robotics solutions. In this context, the speakers emphasised the need for the education and even

certification of professionals for the assessment of robotics solutions, including those involving human-robot-interaction.

**Table 8 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the industrial automation and robotics sector**

|  |  Terminology                            |  Metrology |  Performance Characterisation                           |  Compatibility   |  Regulatory assessment  |
|--|--|---|--|---|--|
| <b>AI system deployment &amp; marketing</b><br>– Regulatory assessment<br>– Users<br>– Transparency & specification<br>– Accountability/responsibility<br>– Maintenance, post-market follow-up & bias monitoring<br>– Supply network | – Lack of technical knowledge to assess safety<br>– Checklists for safety requirements<br>– Training workforce education | – Identification of standards directives  | – Transparency and documentation<br>– Vulnerable populations<br>– Impact of HRI  | – Risk management and assessment<br>– Novel frameworks and hybrid standardisation<br>– Extracting methodologies at of the other standards | – Formal verification methods<br>– Certification programs for conformity assessment<br>– Safety assessment verification<br>– Stakeholder collaboration<br>– Safety assessment verification |
| <b>AI system creation and production</b><br>– Data sets & algorithms (incl. bias)/models<br>– Cybersecurity<br>– System design & integration<br>– Upscaling & evaluation<br>– Quality control  |  | – Data preparation and maintenance<br>– Novel frameworks and hybrid standardisation         | – Smart safety<br>– Empowered safety sensors<br>– Assistive driving consistencies, sensors, testing, methods, safety checks, reliability | – Impact assessment of HRI<br>– Identification of standards directives  | – Training workforce education and certification<br>– System-level certification including sensors   |
| <b>Data creation</b><br>– Compilation, preparation, bias testing<br>– Analysis, processing, labelling<br>– Licensing & restrictions<br>– Sharing & marketing   | – Hybrid standardisation<br>– Data preparation and maintenance<br>– Identification of standards directives               | – Unbiased data<br>– Safety metric<br>– Forgetting data                                     | – Identification of criteria and thresholds<br>– Complex and unstructured environments<br>– Dependencies of the context                  | – Formal methods for robustness<br>– Human-in-the-loop interaction  |  |

The third stage describes the deployment of AI systems in the physical human environment. AI systems require a prior, and well documented risk assessment of the system across the entire lifecycle, the joint consideration of all technical requirements, including human safety and human oversight. In this phase, testing with actual users might reveal issues of either over-reliance or lack of trust in the robot. It was also mentioned that standards that ensure transparency and explainability of the system in an effective and user-friendly manner are relevant to address these potential issues. Special attention was given to the consideration of different applications as well as of the various contexts of operation in which a robotic platform might have to be tested. Also, especially in the case of social robots, standards are expected to include considerations towards the protection of vulnerable groups including children. Some of the above-mentioned concerns and proposals were proven to be transversal across different phases of the system development.

### 4.6.3 Prioritisation and conclusions

The final activity of the session was focused on the prioritisation of concrete actions towards the standardisation of AI and data in the context of industrial automation and robotics. The panellists agreed that action is needed to address the following points:

- Guidelines to navigate and adopt the different standards in the field including their interplay and checklists for robotics conformity procedures;
- Education and certification of professionals for the assessment of robotics solutions including human-robot-interactions;
- Model-based Safety validation/verification;
- Safety Verification of purely data driven, model-free systems.

Overall, standards that ensure human safety and the adherence of human fundamental rights, similar to other applications are of paramount importance, with a special attention on the aspects that relate to the



physical nature of robots. In this context, the robotics field bring specific concerns regarding not only their potential physical but also psychological effect on humans.

Lastly, it should be noted that during the discussion, it was apparent that many of the processes and technical approaches expected to be part of technical specifications are transversal and ideally should be applied throughout the entire life cycle of the robotic application or system. Some examples include:

- i. Risk management and assessment as a process that should apply during the AI lifecycle, including concept, data creation, model training, system development, deployment and operation;
- ii. Verification methods for accuracy, robustness, security and safety should, as well, be tailored to each step of the development process; and
- iii. Regarding the overall standardisation landscape, the speakers mentioned the need for guidance to help with the navigation and joint adoption of the different standards required for conformity.

## 5 Discussion on ways forward

In this exercise it was assessed data quality requirements for inclusive, non-biased and trustworthy artificial intelligence that would benefit from standardisation activities. Firstly, horizontal initiatives for data quality assessment and bias mitigation in research and industry were reviewed by assessing the creation and documentation of datasets for AI, followed by a review of data quality, bias examination and mitigation for already established AI models.

In a second phase it was discussed in six specific sectors the data quality requirements and potential standardisation needs in which artificial intelligence models play an exposed role. Those sectors are of the highest interest for the consumers and industry, and some can be high risk under certain contexts of use and, as such, subject to strict legal obligations before they can be put on the market: 1) Education and Employment; 2) Law Enforcement and the public sector; 3) Medicine and Healthcare.; 4) AI for Industrial Automation and Robotics; 5) Finance; and 6) AI for Media, including Social Media, content moderation, recommender systems.

There are several ongoing efforts to create and document harmonised datasets for AI. Some of them involve important industry players such as Airbus, Atos, Thalès, Valeo, Google, IBM and Bosch. Academic stakeholders are also increasingly paying attention to data quality issues and transparency practices. However, the culture of data transparency is not yet fully ingrained in the AI community, especially in the private sector, where these practices are sometimes perceived as being against the company's interests for different reasons: because they are time-consuming, the lack of unified taxonomy, methodologies and metrics, or the lack of incentives for data sharing. Standards could help bridge these gaps by providing well-defined data creation and documentation processes, harmonised metrics, benchmarks, tools and checklists to ease data creation, quality evaluation and maintenance. Moreover, the current regulatory landscape on data and AI (e.g. Data Act, Data Governance Act, AI Act) opens the door to the provision of standards aligned with legal requirements which would foster their wide adoption by the community. Some standards that are deemed urgent could rely on existing approaches and be implemented in a short term, such as checklists for data conformity and standards for data lineage mapping. Others that are equally important are nevertheless harder to operationalise because of the current lack of consensus in the AI community and the great implementation effort they require. This includes the building of standardised software tools for automatic data documentation and quality analysis, and the definition of a unified taxonomy (including labelling scheme) for AI data.

While reflecting on data quality and bias examination and mitigation in established AI systems, several aspects could benefit from standardisation, such as accounting for the provenance of data and testing the data. The speakers in the session agreed that currently a good definition of data quality for AI does not exist. Such a definition will need to include the aspects of quality AI is referring to, such as a definition of fairness, as well as a list of quality aspects to be assessed and a set of risk assessment criteria.

There is a gap in providing guidance on how to assess when a model works or fails: are there criteria for failure? This could be summarised in AI methodologies for testing the robustness of models, which ultimately assist to enlighten what kind of data goes into the models that are part of AI tools.

Regarding the quality of raw data, one needs to take into account how to assess quality. There is a need for specific tools, even when high quality data standards could assist to assess if the models have learned the right thing. Finally, there is a need for guidelines on how to describe bias mitigation, and their limits.

Data quality requirements and potential standardisation needs for artificial intelligence models play an exposed role in education and employment. Two main challenges were identified: firstly, transparency on how data for AI is used and secondly, how to avoid the creation and perpetuation of power-imbalances through AI. There was an agreement on the need to address the risk of algorithms structuring. Rules tend to build around the algorithm and not the other way around. Solutions include social dialogue mechanisms, data reciprocity and data validation. Further standardisation needs include metrology and terminology, before performance aspects can be tackled. Bias vs discrimination and transparency and accuracy are core concepts that are yet to be defined using a common terminology. Red lines need to be drawn for certain tasks, i.e. performance assessment for workers and students.

AI data quality requirements and potential standardisation needs for law enforcement go beyond face recognition, since there are many other AI tools used by law enforcement agencies (LEAs) that deserve equal attention. The common point of these applications is that the output generated by the AI (e.g. the name of an identified person, a license plate number, the detection of a cybercrime) might have important legal consequences (e.g. the arrest of a person, serve as evidence in courts). For this reason, along with a feeling of invasion of privacy and massive surveillance, citizens fear and have a bad perception of the use of AI in law enforcement. Therefore, this sector has to pay especial attention to trustworthiness, transparency and privacy issues. There are still many challenges to be solved to this end, but three of them are deemed of the utmost priority and would benefit from standard guidelines. The first one is the improvement of current AI systems, including the definition of metrics and evaluation protocols, to ensure they don't discriminate persons (e.g. because of gender, age, sex, disability and ethnicity). There is yet no consensus on how to measure bias and establish related redlines. Next, interoperability issues require important standardisation efforts, not only in terms of data but also regarding communications among organisations (LEAs-LEAs, LEAs-non LEAs, both at national and international levels). Finally, implementing a good training plan for LEA agents in the use of this technology is essential to guarantee that AI is used properly and in the benefit of society.

For the finance sector AI data quality requirements and potential standardisation needs were discussed, and concrete implications of improper use of data was seen as a key problem. Further, there was a certain difficulty noted for moving from the dependency of the process to an universal process. Certainly, there is a need of data quality including the quality model of big data such as for machine learning. It was recommended to start defining a common terminology, followed by an assessment of the consequences of bias in AI. Standard guidelines for model training and testing would be very useful to best approach models to avoid bias resulting from quantitative and qualitative elements.

AI systems for media, including social media, content moderation, recommender systems focus on information that contains text, videos, images, and audios. In this sector there is no definition agreed on what is the meaning of bias. In contrast to other sectors, there is a difference between individual and group AI. Since many data are privately owned the access to this data has led to discussions. Certainly more openness would be useful. In general, there is a need to create benchmarking datasets to understand and re-assess what benchmarking of good data sets means. Agreed common guidelines and standards to audit datasets can be useful in setting out definitions of best practices. A quality assurance is certainly a complex endeavour, since several business models in the media industry are closely linked to AI. There are potential risks for public perception. Off-the-shelf models could be offered to the research community including information on how models were created. These applications can serve as a start, to test the quality of models and to compare the applicability of some tasks versus others.

In the medicine and healthcare sector, AI data quality requirements face a regulatory jungle including difficult General Data Protection Regulation to comply with. Nevertheless there are AI requirements of data provision to authorities. A consensus require that all players should be included, also SMEs that play a particular role in the sector. AI models in the healthcare sector need to shift from black box to explainability and transparency. Potential standardisation needs were discussed, while it was noted that check standards have already been developed for health. It was recommended to envisage rather sooner than later terminology aspects of data characteristics, a need of harmonisation was mentioned, i.e. relevant elements for checking data. A strong focus for a standardisation action is however necessary for methodologies, i.e. how to define a good set of data, how to visualise data and how to overcome algorithm bias. Pre-normative research is needed to represent statistically data sets, without discrimination, and methods to establish re-combinable data metrics. Similarly to the other sections, the provision of guidance and to have a guided sector development has been identified.

In the industrial transformation and robotics sector, AI data quality requirements encounter a large variability of applications including opportunities but also challenges. It was discussed in particular challenges for standardisation, which include safety and system robustness as one of the most predominant aspects, calling for technical guidelines to address them. One of the promising AI technologies in the sector are simulations. In order to assure wide adoption of the standardisation deliverables, a last phase of product development needs to consider users, which may hitherto require technical guidance. Further aspects in data quality requirements for artificial intelligence in the industrial transformation and robotics sector that call for standardisation are smart safety, addressing human fundamental rights, and special clean data collection. Guidelines for stakeholders may be an early harvest, along with a common taxonomy. More challenging, but of utmost importance is how to achieve model-based certification or a data driven certification validation.

Artificial intelligence is employed in a diversity of thematic sectors at different levels. The approach of the AI Act to ensure quality for inclusive, non-biased and trustworthy AI systems has been horizontally designed by formulating general principles. To achieve this, a risk approach has been chosen, the higher the risk for life and health, the more quality and data transparency requirements are necessary. Although the horizontal approach is generally and widely supported, sector specific standards are seen necessary to complement the universal approach. Terminology, taxonomy and training on standardisation practice were identified to be at the forefront.

### **Ways Forward, how to combine sectorial information**

In the following we discuss how to bring ongoing discussions forward, and how to combine the sectorial information from the perspective of small medium enterprises, the policy perspective, legislators, consumer protection, fundamental rights, and basic science perspective.

What are the main challenges that can be address with standards and what are the main steps towards a trustworthy AI? This is a recurring question behind the AI Act . The JRC CEN-CENELEC mobilised the whole community around this topic. The immediate main challenges of the AI Act from a policy perspective is its own timeline with its foreseen application by 2025. This in consequence requires available standards ahead of this date and hitherto the launch of a solid mandate on time to the standard developing organisations. On the other hand, AI data quality requirements deal with societal issues, such as fundamental rights, ethical implications, and the inclusive involvement of different stakeholders. Hence, the process needs to ensure full inclusiveness in the standardisation process and an integration of the points of view in society. Furthermore, there is the mandate of a pre-consultation phase and the ability to work constructively among Member States and Standardisation Bodies. There is a need to identify the gaps to achieve priorities and plan the work for the next three years ahead. In terms of priorities among different domains, the AI Act has identified four main risk categories. The priority for standardisation is at the second risk level (high risk AI systems).

Have there been experiences from other domains? Data is certainly relevant in industrial and governmental sectors. AI deployment in these sectors is essential and there is a need of high quality of data sets particularly for interoperability and harmonisation. This is being achieved through standards, who play a key role in relation to AI. It is important that all third parties are involved in this process. As a reminder, the New Standardisation Strategy attempts to increase efficiency of the standardisation system, which identifies several milestones: the coordination among ESOs, a clear roadmap that gives concrete responsibilities to the ESOs, the participation of European experts at international level, addressing the problem of resources, and a well-structured plan which embodies a duly representation of EU values. Moreover, there should be a nuanced interplay between horizontal and vertical standardisation. On the one side general standards establish basic principles for AI, on the other side sectorial standards address specific needs that have to be planned and assessed. Ensuring an efficient cooperation with research activities in pre-normative research, it was recalled the standardisation strategy, the Code of Practice for researchers, which guides researchers on how to contribute to standardisation and the newly kicked-off Standardisation Booster, which facilitates coaching and guidance. In the future there will be other opportunities to coordinate better standardisation processes. The relation between R&I and standardisation is relatively high in the political agenda. For the moment the HD Booster is the reference, but there is also StandICT.eu. There might also arise needs to stronger coordination among European players within ISO. There is need to ensure that European participants will have a better dialogue. Nevertheless, all European National Standardisation Bodies are involved, and they seem to be well represented.

Non-discrimination is at the core of data quality requirements for inclusive, non-biased and trustworthy AI systems, consequently it needs to be guaranteed before any AI is deployed. Despite the complexity, many experts and international organisations are dedicated to provide solutions and translating legal requirements into practical applications. There is a need for clarity and to align initiatives based on own mandates, which

standardisation can provide. It is important to include and achieve consensus with civil society, as this will increase trust in the use of AI. Documenting and sharing data is crucial, especially since bias is a concern for discrimination. There is also a need for greater multi-disciplinarity in this context, as well as for agreeing on the definition of disciplinarity, which still is being discussed, advocating for a shorter but no narrower definition. Users want computational solutions, but this goes along with a trade-off of greater bias. Bias is a very broad term, as it includes political views, social background, sexual orientation, etc. It is important to identify bias. It is also important to collect information, however, this is limited by privacy rights. There is not only the need for justification, but also the assurance of safety handling.

AI systems are simply another type of product. Products that are autonomous. The French National Laboratory for Metrology and Testing (LNE) is providing services to the industry such as testing, evaluating, checking compliance of products and providing legal services such as certification and metrology. Metrology institutes such as LNE do normally not develop AI systems, but require tools for evaluation, metrics, scores, thresholds, and methods. Metrology institutes may design these tools, and in regards to regulatory requirements, they need to build testing methods to test compliance. There is a need for specialised data sets for assessing bias; this is not easy to build. Building reference databases is however an important task. It is therefore recommended to start with some top level characteristics, such as in health care, and to better start with methods, rather than defining thresholds or terminology. However, in many domains and for safety, thresholds and terminology can be very important too. But how the outcome of the workshop can help a better relation between data and AI? We should explore how we can design a standard to define trustworthiness. The AI Act led to the rise of a market of trustworthiness AI. Potential future standards need to address what are the characteristics of trustworthiness. In case of data, it is important to ensure trustworthiness of AI. Data transparency is very important, but is only one characteristic among many others. We must have top level properties of data that impact the trustworthiness of AI. It is all about providing clear definitions and tools.

We also need to reach an inclusive environment: Science, SMEs and users work together. An inclusive environment is possible and should be supported. A study on how data and data quality affects SMEs and standards has been published<sup>24</sup>, suggesting the participation in regulatory sandboxes and data quality. It will be difficult to achieve provision of error free data, hence pragmatism may lead to the acceptance that data is part of a long supply chain. Conformity assessment is recommended for data and to invest in meaningful work on data quality. The SME Alliance represents SMEs and most of them are from Europe. There is a new task group on data with 200 members, open to anyone who wants to engage. There are National, European and International standardisation committees. Interested individuals need to approach each their National standardisation body. Digital SME Alliance made a fact sheet<sup>25</sup> on their position advocating for error free and complete data sets. The Alliance is handing conformity assessment: they organise science, practitioners, standardiser initiatives. There is a need to be opened internationally, as well as for mapping requirements at international level. Data quality is perceived as a rather complex process.

AI systems may bring a lot of benefits in easing the workload of workers and may create more profits for companies, so that workers can focus on the core of their jobs. However, AI presents some risks too. Many jobs will be impacted by AI, others will be all transformed, and only with serious investments in training damage can be alleviated. Training would also help to mitigate a power unbalance between employers and employee. AI should be the subject of social dialogue. The recent published EU Strategy of Standardisation insists on the inclusion of representatives from civil society (workers, SMEs, consumers). There is a need to consider issues concerning human rights, such as bias related to employees, discrimination, privacy, data accuracy, and the possibility of using data out of context. Similarly, issues in relation to trustworthiness, reiterate the need to have also humans in control. There is an immediate need for addressing three types of standards: 1) definition (including aspects on bias, trustworthiness, etc.), which may be difficult, but even more difficult to find international consensus; 2) accuracy and 3) quality of data. How to lift this topics to ISO levels? Europe has specific values to protect, but is not an isolated continent. Different vocabulary – different cultures, how are workers by this circumstances impacted? Can we really have metrics when we have different cultures? There is universality when it comes to transparency, privacy, trust on the system if human is under control, data privacy. When it comes to bias, we may need a new concept on impact assessment. Nevertheless assessments on how high-risk applications are handled and tested.

---

<sup>24</sup> OECD (2021) The Digital Transformation of SMEs, 5. Artificial intelligence: Changing landscape for SMEs, OECD Publishing, Paris, <https://doi.org/10.1787/bdb9256a-en>

<sup>25</sup> DigitalSME (2021) AI Factsheet <https://www.digitalsme.eu/digital/uploads/DIGITAL-SME-FACT-SHEET-AI-ACT-FINAL.pdf>

## 6 Conclusion

AI systems have indeed the potential to improve our lives as well as the overall economic and societal welfare, leading to better healthcare services, safer and cleaner transport systems, better working conditions, higher productivity and new innovative products, services and supply chains. Artificial intelligence (AI) systems are being deployed every day, affecting every aspect of our lives with applications such as *recommender systems* for medical diagnostics, bank loan approval and CV filtering.

Recently, academic research on data quality in AI and machine learning has received increased attention. Bias in existing state-of-the-art AI models has been widely proven, raising concerns on societal consequences. Researchers are intensively working on methods to eliminate bias, starting from the input data and by curating AI model training. However, there is still a lot of work to be done, for example, no common approach to measure data quality has yet been defined.

The European Standardisation Strategy identifies as a priority the need for European Standards for data enhancing data interoperability, data sharing and data re-use, and Putting-Science-Into-Standards. This annual event is acknowledged by the Strategy as an important foresight exercise that explores standardisation needs linked to emerging technologies.

The objective of the workshop has been to map existing and missing standardisation efforts and to recommend steps to start the process of drafting new standards. From the workshop gathered AI practitioners and researchers that expressed current challenges and what methods have been developed to evaluate and mitigate bias present in AI. The ultimate objective is to initiate a discussion on how current and future standards can be used to mitigate bias in AI models.

The efforts to standardise AI systems are in full swing: standards in the field of AI can and will promote the transfer of technologies from research to market, and most importantly, a very important benefit European standards secure is the establishment of uniform and consensus-agreed requirements that support the implementation of the legislation that underpin our European ethics and values.

### **What could be the next steps?**

This event is very timely because through the EC-ESOs Task force Strategic Alignment Initiative, it is possible to bring into line the bottom-up approach (from experts, academia, researchers, businesses, societal stakeholders) and the top-down (from public policies and legislation) so that it is possible to work on common understanding and targets. One of its pilots is addressing AI including the development of a European Standardisation Roadmap (which will certainly use the outcome of this PSIS workshop).

Part of this alignment exercise consists on better and earlier communication for better anticipation. The current consultation on the draft Standardisation request for European Standards in support of the AI Act addressing safety and trustworthiness of AI systems can be a very concrete and successful example of the importance to anticipate to deliver timely solutions that can support the economy and the society.

Domain-specific characteristics may require specific standards that are adapted to the particularities and requirements of specific industries and applications. This requires cooperation between the experts and technical bodies focused on specific sectors to avoid overlaps and inconsistencies between horizontal and vertical standards.

There is also the possibility to exploit the collective European strength in ISO and IEC. CEN and CENELEC Members send experts to ISO/IEC JTC 1 SC 42 AI, which is just initiating the work on specific standards for data quality and bias.

It is also possible to seize opportunities to work with like-minded international partners to bring the European angle at the international level and to ensure Europe can act as a global standard-setter.

### **Call for action**

The European standardisation process reassures its inclusiveness and full representation of all relevant stakeholders, exploiting the connections we have with industry, including SMEs representatives, civil society, and academia.

This event has been an important occasion to build a stronger engagement of experts in AI standardisation who can be ready to contribute to the development of technical standards that are already needed to support the market deployment of AI systems in accordance with the EU's artificial intelligence act.

The discussion has now started and support has been committed to keep the community that gathered in this workshop up to date, with the aim to further engage in the development of European Standards, and share the next steps to support the timely availability of Data Standards for AI.

Points that will be taken to the Joint Technical Committee 21 on artificial intelligence (from the different sessions) include:

Standardisation opportunities/needs that can be taken up in the AI Standardisation Roadmap:

Standards for creating AI systems:

- Standards for *management* (data quality, system surety, 'balance not bias')
- Standards for *deployment* (understanding, trust, consistent application)
- Standards for *operational testing* (pre- and post-deployment)
- Standards for *data quality measurement* (metrics, KPI)
- Standards for certification of AI algorithms

*NB. prCEN/CLC/TR 17894 AI Conformity Assessment is under development*

Standards for deployment of AI systems:

- *Definition*: What is/are: AI, data, systems, terminology, risk, limitations, safety, human/AI interface, data quality.
- *Education & training* (AI semantics, meaning, intent, application and approach for designing and planning AI systems, data needs, bias measurement & mitigation...)
- *Design*: Standards for *oversight* of AI systems (by authorities) & Standards for *deployment* (users – industry, incl. SMEs)
- *Data*: creating datasets for AI: *Data provenance, collection & presentation* (minimum criteria, origin...), *using* and *sharing* data. Checklist for dataset conformity. Trusted data. Data maintenance and robustness.

Standards supporting take up and use of AI by industry:

- *All sectors*: *quality of off-the-shelf* AI models
- *SMEs* (standards enabling access to AI systems by all types of organisations and entities)
- *Law enforcement* (gatekeeping, facial recognition, image quality...)
- *Finance*: data 'fairness', decision making processes based on AI, objectivity and quantitative measures
- *Media*: specific aspects of bias in AI for media, comparison of different AI models, benchmarking of datasets, definitions of bias and quality specific for media, codification of existing best practices, testing for bias in data and benchmarking.
- *Healthcare*: regulatory situation (GDPR, MDR), data-related provisions to Regulatory Authorities – legal compliance, diverse range of use cases, harmonisation of terminology, *anonymisation* of data, '*state-of-the-art*' *measurement protocols*, *documenting AI systems*, personalisation needs.
- *Industrial Automation & Robotics*: Safety of AI systems (safety by design) and metrics, diverse range of applications, robot/human interface and safety, respect for human rights incl. *privacy*, terminology (will support users to contribute to standardisation)

## 7 References

- Afzal, S, C Rajmohan, M Kesarwani, and S Mehta. 2021. *Data readiness report*. In *2021 IEEE International Conference on Smart Data Services (SMDS)*.
- Baiocco, S, E Fernández-Macías, U Rani, and A Pesole. 2022. *The algorithmic management of work and its implications in different contexts (No. 2022/02)*. JRC Working Papers Series on Labour, Education and Technology.
- Bäuerle, A, AA Cabrera, F Hohman, M Maher, D Koski, X Suau, and D Moritz. 2022. "Symphony: Composing Interactive Interfaces for Machine Learning." *CHI Conference on Human Factors in Computing Systems (April)*: 1-14.
- Birhane, A, and VU Prabhu. 2021. "Large image datasets: A pyrrhic win for computer vision?" *IEEE Winter Conference on Applications of Computer Vision (WACV) IEEE* January: 1536-1546.
- CEN-CENELEC. 2020. *CEN-CENELEC Focus Group Report: Road Map on Artificial Intelligence (AI)*. Accessed August 29, 2022. <https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/>.
- Charisi, V., R. Compano, N. Duch-Brown, E. Gomez-Gutierrez, D. Klenert, M. Lutz, R. Marschinski, and Torrecilla-Salinas. 2021. *What future for European robotics?* Sevilla: Joint Research Centre.
- Chmielinski, KS, S Newman, M Taylor, Josh Joseph, KThomas, J Yurkofsky, and YC Qiu. 2022. "The dataset nutrition label (2nd gen): Leveraging context to mitigate harms in artificial intelligence." *arXiv preprint (arXiv)*. arXiv:2201.03954.
- Collins, G. S., J. B. Reitsma, D. G. Altman, and K. G. Moons. 2015. "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement." *BMJ* 7594.
- Czarnowska, Paula, Yogarshi Vyas, and Kashif Shah. 2021. "Quantifying Social Biases in NLP: A Generalization and Empirical Comparison of Extrinsic Fairness Metrics." *Transactions of the Association for Computational Linguistics* 9: 1249-1267. doi:[https://doi.org/10.1162/tacl\\_a\\_00425](https://doi.org/10.1162/tacl_a_00425).
- Deven Santosh Shah, H, Andrew Schwartz, and Dirk Hovy. 2020. "Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics. 52.
- Duan, Zening, Jianing Li, Josephine Lukito, Kai-Cheng Yang, Fan Chen, Dhavan V Shah, and Sijia Yang. 2022. "Algorithmic Agents in the Hybrid Media System: Social Bots, Selective Amplification, and Partisan News about COVID-19." *Human Communication Research* 48.
- Duch-Brown, N, E Gomez-Herrera, F Mueller-Langer, and S Tolan. 2022. *Market power and artificial intelligence work on online labour markets*. *Research Policy*, 51(3), 104446.
- Duch-Brown, N., F. Rossetti, and R. Haarbuerger. 2021. *Evolution of the EU market share of robotics: Data and methodology*. Luxembourg: Publications Office of the European Union. doi:10.2760/292931.
- European Commission. 2019. *European Commission's High-Level Expert Group on AI. ALTAI - The Assessment List on Trustworthy Artificial intelligence*. Luxembourg: European Commission. <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-listtrustworthy-artificial-intelligence>.
- European Commission. 2021. *European Commission's proposal for a Regulation on Artificial Intelligence*. Luxembourg: European Commission. [43](https://digital-</a></p></div><div data-bbox=)

strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

- . 2021. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.
- European Parliament. 2020. *Panel for the future of Science and Technology: Artificial intelligence: from ethics to policy*. Accessed August 29, 2022. [https://www.europarl.europa.eu/stoa/en/document/EPRS\\_STU\(2020\)641507](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2020)641507).
- European Union. 1985. *Resolution of 7 May 1985 on a new approach to technical harmonization and standards*. 7 May. [https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31985Y0604\(01\)](https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31985Y0604(01)).
- FUTURE-AI. 2021. *FUTURE-AI: Best practices for trustworthy AI in medicine*. Accessed August 29, 2022. <https://future-ai.eu/>.
- Gebru, T, J Morgenstern, B Vecchione, J Wortman Vaughan, H Wallach, H Daume III, and K Crawford. 2018. "Datasheets for Datasets." *arXiv e-prints, arXiv-1803*. 1803. Accessed 09 27, 2022. <https://arxiv.org/abs/1803.09010>.
- German Notified Bodies Alliance. 2022. *Questionnaire Artificial Intelligence (AI) in medical devices*. Accessed August 29, 2022. <https://www.ig-nb.de/>.
- Grother, P, M Ngan, and K Hanaoka. 2019. "Face recognition vendor test (FRVT): Part 3, demographic effects." National Institute of Standards and Technology, Gaithersburg, MD.
- Guidotti, R, A Monreale, S Ruggieri, F Turini, F Giannotti, and D Pedreschi. 2018. "A survey of methods for explaining black box models." *ACM computing surveys (CSUR)* 51 (5): 1-42.
- Hameed, Mazhar, and Felix Naumann. 2020. "Data preparation: A survey of commercial tools." *Hameed, Mazhar, and Felix Naumann. ACM SIGMOD Record* 49 (3): 18-29.
- Holland, S, Ahmed Hosny, S Newman, Joseph, Joshua, and Kasia Chmielinski., Joshua Joseph, and Kasia Chmielinski. 2018. *The dataset nutrition label: A framework to drive higher data quality standards* (arXiv preprint). arXiv:1805.03677.
- Hupont, I, M Micheli, B Delipetrev, E Gómez, and J Soler Garrido. 2022. *Documenting high-risk AI: an European regulatory perspective*. (European Commission). Accessed 09 27, 2022. [https://www.techrxiv.org/articles/preprint/Documenting\\_high-risk\\_AI\\_an\\_European\\_regulatory\\_perspective/20291046](https://www.techrxiv.org/articles/preprint/Documenting_high-risk_AI_an_European_regulatory_perspective/20291046).
- Hupont, I, S Tolan, and H Gunes. 2022. "The landscape of facial processing applications in the context of the European AI Act and the development of trustworthy systems." *Sci Rep* 12 10688.
- Hupont, Isabelle, and Carles Fernández. 2019. "Demogpairs: Quantifying the impact of demographic imbalance in deep face recognition." *14th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE.
- Hupont, Isabelle, and M Chetouani. 2019. "Region-based facial representation for real-time action units intensity detection across datasets." *Pattern Analysis and Applications* 22 (2): 477-489.
- Hupont, Isabelle, E Gomez, Songul Tolan, L Porcaro, and A Freire. 2022. "Monitoring Diversity of AI Conferences: Lessons Learnt and Future Challenges in the DivinAI Project." *arXiv preprint*. doi:arXiv:2203.01657.



- Huszár, F, SI Ktena, C O'Brien, L Belli, A Schlaikjer, and M Hardt. 2022. "Algorithmic amplification of politics on Twitter." *Proc Natl Acad Sci USA*. (PNAS) 119 (1): 119. doi:10.1073/pnas.2025334119.
- Hutchinson, B, A Smart, E Denton, C Greer, O Kjartansson, and M Mitchell. 2021. *Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O. & Mitchell, M. (2021, March). Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. In 2021 ACM Conference on Fairness, Ac.*
- ISO. 2022. *ISO/IEC CD 5259-1 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples*. 12 09. <https://www.iso.org/standard/81092.html>.
- ISO/IEC. 2022. *ISO/IEC TR 24027:2021 Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*. 12 09. <https://www.iso.org/standard/77607.html>.
- . 2021. *ISO/IEC TR 24030:2021 Information technology — Artificial intelligence (AI) — Use cases*. Accessed 09 28, 2022. <https://www.iso.org/standard/77610.html>.
- Kumar, SA, MM Nasralla, I García-Magariño, and H Kumar. 2021. "A machine-learning scraping tool for data fusion in the analysis of sentiments about pandemics for supporting business decisions with human-centric AI explanations." *PeerJ Computer Sci*.
- Kunze, L., N. Hawes, T. Duckett, M. Hanheide, and T. Krajník. 2018. "Artificial intelligence for long-term robot autonomy: A survey." *IEEE Robotics and Automation Letters* (3): 4023-4030.
- Lee, I. 2021. "Service robots: a systematic literature review." *Electronics* (10): 2658.
- Lemaignan, S., M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami. 2017. "Artificial cognition for social human-robot interaction: An implementation." *Artificial Intelligence* (247): 45-69.
- Madaio, Michael, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. "Co-designing checklists to understand organizational challenges and opportunities around fairness in AI." *CHI Conference on Human Factors in Computing Systems*. 1-14.
- Martín, A, J Huertas-Tato, A Huertas-García, G Villar-Rodríguez, and D Camacho. 2021. *FacTeR-Check: Semi-automated fact-checking through Semantic Similarity and Natural Language Inference*. preprint, preprint.
- Matthew, Arnold, RKE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Alexandra Mojsilovic, Ravi Nair, Natesan Ramamurthy, Alexandra Olteanu, and David Piorkowski. 2019. "AI FactSheets: Increasing trust in AI services through supplier's declarations of conformity." *IBM Journal of Research and Development*, (4/5): 6-10.
- Mitchell, M, Simone Wu, A Zaldivar, P Barnes, L Vasserman, B Hutchinson, E Spitzer, and T Gebru. 2019. "Model cards for model reporting." *Conference on fairness, accountability, and transparency*,. 220–229.
- Mitchell, Margaret , Simone Wu, Andrew Zaldivar,, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2018. *Model cards for model reporting. In Conference on fairness, accountability, and transparency*.
- Negrão, M, and P Domingues. 2021. "SpeechToText: An open-source software for automatic detection and transcription of voice recordings in digital forensics." *Forensic Science International: Digital Investigation* 38: 301223.

- OECD. 2022. "Framework for Classification of AI Systems: a tool for effective AI policies." Accessed 09 27, 2022. <https://oecd.ai/en/classification>.
- . 2021. *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, OECD Publishing, Paris, <https://doi.org/10.1787/589b283f-en>.
- Phillips, P. Jonathon, and Mark Przybocki. 2020. "Four principles of explainable AI as applied to biometrics and facial forensic algorithms." *arXiv* (arXiv preprint) arXiv:2002.01014. arXiv:2002.01014.
- PricewaterhouseCoopers. 2020. *2020 AI Predictions Report*. Accessed 10 15, 2022. <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-business-survey.html>.
- Quijano-Sánchez, L, F Liberatore, J Camacho-Collados, and M Camacho-Collados. 2018. "Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police." *Knowledge-Based Systems* 149: 155-168.
- Rajpurkar, P, E Chen, O Banerjee, and EJ Topol. 2022. "AI in health and medicine." *Nat Med*. 28 (1): 31-38. doi:10.1038/s41591-021-01614-0.
- Reale, N, M Nasrabadi, and R Chellapa. 2016. "An analysis of the robustness of deep face recognition networks to noisy training labels." *IEEE Global Conference on Signal and Information Processing* 1192-1196.
- Saenz, J., R. Behrens, E. Schulenburg, H. Petersen, O. Gibaru, P. Neto, and N. Elkmann. 2020. "Methods for considering safety in design of robotics applications featuring human-robot collaboration." *The International Journal of Advanced Manufacturing Technology* (107): 2513-2331.
- Sánchez, FL, I Hupont, S Tabik, and F Herrera. 2020. "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects." *Information Fusion* 64: 318-335.
- Song, H, M Kim, D Park, Y Shin, and JG Lee. 2020. "Learning from noisy labels with deep neural networks: A survey." *IEEE Transactions on Neural Networks and Learning Systems*.
- Stanford University. 2022. *Stanford University Human-Centered Artificial Intelligence (2022). Artificial Intelligence Index Report 2022*. <https://aiindex.stanford.edu/report/>.
- Sun, Tony, Andrew Gaut, Shirlyn Tang, Mai ElSherief, Yuxin Huang, Diba Mirz, and Jieyu Zhao. 2019. "Mitigating Gender Bias in Natural Language Processing: Literature Review." *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics. 1630-1640.
- The New York Times. 2020. "Another arrest, and jail time, due to a bad facial recognition match." *The New York Times*, 29 12. Accessed 09 28, 2022. <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>.
- Tolan, S, A Pesole, F Martínez-Plumed, E Fernández-Macías, J Hernández-Orallo, and E Gómez. 2021. *Measuring the occupational impact of AI: tasks, cognitive abilities and AI benchmarks*. *Journal of Artificial Intelligence Research*, 71, 191-236.
- Urzi Brancati, M.C., A Pesole, and E Fernandez Macias. 2020. *New evidence on platform workers in Europe*. EUR 29958 EN, Publications Office of the European Union, Luxembourg, JRC118570.
- US Food and Drug Administration. 2021. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. Accessed 09 28, 2022.

<https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>.

- Vincent, N, and B Hecht. 2021. "Data and its (dis) contents: A survey of dataset development and use in machine learning research." *Patterns* 2 (11): 100388.
- Vincent-Lancrin, S, and R van der Vlies. 2020. *Trustworthy artificial intelligence (AI) in education: Promises and challenges*. OECD Education Working Papers, 218.
- Wang, F, L Chen, C Li, S Huang, F Wang, Y Chen, C Qia, and C Change Loy. 2018. "The devil of face recognition is in the noise." *European Conference on Computer Vision (ECCV)* (September).
- Wang, M, and W Deng. 2021. "Deep face recognition: A survey." *Neurocomputing* 429: 215-244.
- Williford, JR, BB May, and J Byrne. 2020. "Explainable face recognition." *European Conference on Computer Vision* (Springer) 248-263.
- World Health Organisation. 2021. *Ethics and governance of artificial intelligence for health*. Accessed 09 28, 2022. <https://www.who.int/publications/i/item/9789240029200>.
- Zhu, Zheng, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, et al. 2021. "A Benchmark Unveiling the Power of Million-scale Deep Face Recognition." *WebFace260M: Conference on Computer Vision*.

## List of abbreviations and definitions

AI – Artificial Intelligence  
CCMC – CEN CENELEC Management Centre  
CE - Manufacturer affirms the good's conformity with European health, safety, and environmental protection standards  
CEN - European Committee for Standardization  
CENELEC - European Committee for Electrotechnical Standardization  
CETSE - Centro Tecnológico de Seguridad  
CV – Curriculum vitae  
DNA - Deoxyribonucleic acid is a double helix carrying genetic instructions  
EHR - Electronic Health Records  
ESO - European Standards Organizations (ESO) under Regulation 1025/2012  
EU – European Union  
FCA - Financial Conduct Authority of the United Kingdom  
FDA – United States Food and Drug Administration  
FRVT - Face Recognition Vendor Test  
IBM - International Business Machines Corporation  
IEC - International Electrotechnical Commission  
IEEE - Institute of Electrical and Electronics Engineers  
IFR - international federation of robots  
ISO- International Organization for Standardization  
IVDR – EU conformity assessment options for in vitro diagnostic medical devices.  
JRC – Joint Research Centre  
JTC – Joint Technical Committee  
LEA - Law Enforcement Agencies  
NGO – Non-Governmental organisation  
NIST – US National Institute of Standards and Technology  
NLP - Natural Language Processing  
OECD - Organisation for Economic Co-operation and Development  
PROBAST - tool to assess the risk of bias and applicability of prediction model studies.  
PSIS – Putting Science Into Standards Workshop  
SME – Small and Medium Sized Enterprise  
StandICT.eu - EU framework project with the central goal of ensuring a neutral, reputable, pragmatic and fair approach to supporting European presence in the international ICT standardisation  
TRIPOD-AI - development of a reporting guideline for diagnostic and prognostic prediction studies based on artificial intelligence  
UK – United Kingdom  
UNESCO - United Nations Educational, Scientific and Cultural Organization  
US – United States  
VDE - Verband der Elektrotechnik, Elektronik und Informationstechnik is one of Europe's largest technical-scientific associations

## List of figures

|  |    |
|--|----|
| Figure 1 Type of workshop participating organisations.....   | 4  |
| Figure 2 Level of experience in standardisation among the participants.....  | 4  |
| Figure 3 A selection of international formal standardisation organisation groups and committees in the area of AI, health informatics and medical equipment. The arrow between the health informatics groups indicates that ISO/IEC and CEN/CENELEC groups are regularly exchanging information. In the context of this workshop, it is important to note that CEN-CENELEC JTC 21 has an ad hoc group on “data governance and quality for AI”, while ISO/IEC’s SC42 on AI has a dedicated working group on data quality. Other relevant standardisation groups are not shown (e.g. IEEE or VDE)..... | 26 |
| Figure 4 Relevant ISO/IEC publications (standards and technical reports) regarding both horizontal aspects of AI (e.g. robustness, bias, machine learning = ML) and health specific aspects (i.e. ML applications for imaging and other medical applications).....   | 27 |

## List of tables

|  |    |
|--|----|
| Table 1 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain for creating and documenting datasets.....                | 11 |
| Table 2 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain for data quality and bias examination and mitigation..... | 13 |
| Table 3 Overview of pre-normative research- and standards needs in selected standardisation categories and along the AI value chain of education and employment.....                         | 17 |
| Table 4 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the law enforcement and the public sector.....         | 20 |
| Table 5 Overview of pre-normative research- and standard needs in selected standardisation categories and along the AI value chain of the finance sector.....                                | 22 |
| Table 6 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the media sector.....                                  | 24 |
| Table 7 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the medicine and healthcare sector.....                | 32 |
| Table 8 Overview of pre-normative research and standards needs in selected standardisation categories and along the AI value chain of the industrial automation and robotics sector.....     | 36 |

## **Annexes**

### **Annex 1. Workshop agenda**

Putting Science Into Standards. Workshop on Data quality requirements for inclusive, non-biased and trustworthy AI. Online on 8 and 9 June 2022

14:00 -14:05 *Welcome and introduction* Alexandra Balahur (moderator)

14:05 -14:25 *Opening*, Bernard Magenmann (Deputy Director-General JRC), Stefano Calzolari (President CEN)

14:25 -14:35 *Data in the context of AI Act*, Lucilla Sioli (Director DG CNECT)

14:35 -14:45 *Ensuring an ethical use of AI technologies to the benefit of humanity* Gabriela Ramos (Assistant Director-General, UNESCO)

14:45-14:55 *OECD AI data quality principles on trustworthiness, human rights and democratic values*, Karine Perset (Head of AI unit OECD)

14:55-15:05 *Algorithmic bias considerations US standardization initiative for identifying and managing bias in AI*, Ansgar Koene (Chair, IEEE P7003 Working Group), Reva Schwartz (principal investigator for AI bias, NIST)

15:05 -15:15 *Data requirements as outlined in the AI Act*, Gabriele Mazzini (Team leader AI act, DG CNECT)

15:15-15:35 *Overview on standardisation road mapping on ESO and ISO level*, Sebastian Hallensleben (Chair, CEN-CENELEC JTC 21 AI), Patrick Bezombes (Chair, ISO/IEC JTC 1/SC 42/AG 3 AI standardization roadmap)

15:35-15:45 Break

### **BLOCK I PARALLEL SESSION**

DAY 1 *Horizontal initiatives for data quality assessment and bias mitigation in research and industry*

15:45-17:15

*Creating and documenting datasets for AI*, Felix Naumann (Hasso-Plattner-Institut), Emmanuel Kahembwe (VDE), Kasia Chmielinski (Dataset Nutrition label), Flora Dellinger (Confiance.ai), Rapporteurs: Isabelle Hupont Torres, Josep Soler Garrido

*Data quality and bias examination and mitigation in AI*, Rasmus Adler (Fraunhofer IESE), Francisco Herrera (Univ. Granada), David Reichel (FRA), Fred Morstatter (ISI), Rapporteur: Maurizio Salvi, Alexandra Balahur

DAY 1

09:00-09:15 *Welcome to Day 2* Alexandra Balahur (moderator)

### **BLOCK II PARALLEL SESSIONS**

*Data quality and bias mitigation needs and practices in selected sectors*

09:15-10:30 *Education and Employment*, Dee Masters (Cloister), Nikoleta Giannoutsou (JRC), Enrique Fernandez Macias (JRC), Rapporteur: Songül Tolan, Matteo Sostero

*Law Enforcement and the public sector*, Patrick Grother (NIST FRVT), Javier Rodríguez Saeta (Herta), Robin Allen (Cloister), Rosalía Machín Prieto (Gov. Spain), Rapporteur: Isabelle Hupont Torres

*Finance*, Karen Croxson (FCA UK), Andrea Caccia (Chair CEN CENELEC JTC 19 Blockchain), Jörg Osterrieder (Zurich University of Applied Sciences), Rapporteur: Maurizio Salvi

10:30-10:45 Break

DAY 2

10:45-12:00 *AI for Media, including Social Media, content moderation, recommender systems*, Symeon Papadopoulos (Centre for Research and Technology Hellas), Maja Pantic (Imperial College, London), Manuel Gomez Rodriguez (MPI), Jochen Leidner (Coburg University), Rapporteur: Alexandra Balahur

*Medicine and Healthcare*, Sandra Coecke (JRC), Thorsten Prinz (VDE Health), Alpo Värri (Convenor CEN/TC 251 Health Informatics WG2), Koen Cobbaert (Philips), Rapporteur: Claudius Griesinger

*AI for Industrial Automation and Robotics*, Aurélie Clodic (ANITI), Roland Behrens (Fraunhofer IFF), Emmanuel Kahembwe (Univ Edinburgh), Adil Amjad (Thermo Fisher Scientific), Rapporteur: Vasiliki Charisi, Isabelle Hupont Torres

12:00-13:30 Lunch break

## **PLENARY**

13:30-14:15 Flash Summaries of parallel sessions

14:15-15:45 *Panel discussion on ways forward*, Emilia Gómez (JRC, moderation), Emilia Tantar (CEN and CENELEC JTC 21), Salvatore Scalzo (DG CNECT), Antonio Conte (DG GROW), David Reichel (FRA), Agnès Delaborde (LNE), Philippe Saint-Aubin (CFDT, ETUC expert in JTC 21)

15:45-16:00 Closing, Elena Santiago Cid (CEN CENELEC Director General)

## **Annex 2. Workshop Advisory board members**

Werner Bailer (Austrian Standards, ASI), Arne Berre (Standards Norway, SN), Patrick Bezombes (ISO/IEC, JTC 1/SC 42/AHG 5 AI, standardization landscape and roadmap), Thierry Boulance (European Commission, DG CNECT), Tatjana Evas (European Commission, DG CNECT), Anders Friis-Christensen (European Commission, JRC), Ashok Ganesh (CEN-CENELEC), Chiara Giovannini (ANEC), Emilia Gomez (European Commission, JRC), Sebastian Hallensleben (CEN-CENELEC, JTC 21 AI), Laurens Hernalsteen (CEN-CENELEC, JTC 21 AI), Kim Skov Hilding (CEN-CENELEC, JTC 21 AI), Filipe Jones Mourao (European Commission, DG, CNECT), Marijan Kamenjašević (Croatian Standards Institute, HZN), Matthew King (European Commission, JRC), Irina Orsich (European Commission, DG CNECT), Andrea Raffaelli (Small Business Standards, SBS), Salvatore Scalzo (European Commission, CNECT), Emilia Tantar (Small Business Standards, SBS), Fabio Taucer (European Commission, JRC)

## **GETTING IN TOUCH WITH THE EU**

### **In person**

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online ([european-union.europa.eu/contact-eu/meet-us\\_en](https://european-union.europa.eu/contact-eu/meet-us_en)).

### **On the phone or in writing**

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: [european-union.europa.eu/contact-eu/write-us\\_en](https://european-union.europa.eu/contact-eu/write-us_en).

## **FINDING INFORMATION ABOUT THE EU**

### **Online**

Information about the European Union in all the official languages of the EU is available on the Europa website ([european-union.europa.eu](https://european-union.europa.eu)).

### **EU publications**

You can view or order EU publications at [op.europa.eu/en/publications](https://op.europa.eu/en/publications). Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre ([european-union.europa.eu/contact-eu/meet-us\\_en](https://european-union.europa.eu/contact-eu/meet-us_en)).

### **EU law and related documents**

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex ([eur-lex.europa.eu](https://eur-lex.europa.eu)).

### **Open data from the EU**

The portal [data.europa.eu](https://data.europa.eu) provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth



## The European Commission's science and knowledge service

Joint Research Centre

### JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



**EU Science Hub**  
[joint-research-centre.ec.europa.eu](https://joint-research-centre.ec.europa.eu)



@EU\_ScienceHub



EU Science Hub - Joint Research Centre



EU Science, Research and Innovation



EU Science Hub



EU Science



Publications Office  
of the European Union