

'Putting Science Into Standards' workshop

Welcome! We will start soon



slido.com #Standards4AI









Welcome and introduction

Alexandra BALAHUR

JRC moderator



Audience interaction





- Select the **Day 1: Main room** on Slido
- Zoom chat only technical questions to host
- Camera and audio OFF

Programme – 8 June



14:00 – 14:05	WELCOME AND INTRODUCTION, Alexandra BALAHUR, JRC - moderator
14:05 - 14:25	 OPENING Bernard MAGENHANN, 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 - 14:40	ENSURING AN ETHICAL USE OF AI TECHNOLOGIES TO THE BENEFIT OF HUMANITY, Gabriela RAMOS, Assistant Director- General, UNESCO
14:40- 14:50	OECD AI DATA QUALITY PRINCIPLES ON TRUSTWORTHINESS, HUMAN RIGHTS AND DEMOCRATIC VALUES, Karine PERSET, Head of AI unit, OECD
14:50-15:05	 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, DG CNECT Team Leader for the AI Act, Gabriele MAZZINI, DG CNECT Team Leader for the AI Act
© CEN-CENELEC 2021	 OVERVIEW ON STANDARDISATION ROAD MAPPING IN EUROPEAN AND INTERNATIONAL STANDARDS ORGANIZATIONS Sebastian HALLENSLEBEN, Chair, CEN-CENELEC JTC 21 AI Patrick BEZOMBES, ISO/IEC JTC 1/SC 42/AG 3 (AI standardization roadmapping)

Programme – 8 June



PARALLEL SESSIONS HORIZONTAL INITIATIVES FOR DATA QUALITY ASSESSMENT AND BIAS MITIGATION IN RESEARCH AND INDUSTRY 15:45-17:15 **CREATING AND DOCUMENTING DATASETS FOR AI** DATA QUALITY AND BIAS EXAMINATION AND **MITIGATION IN AI** Felix NAUMANN, Hasso-Plattner-Institut **Emmanuel KAHEMBWE,** VDE **Rasmus ADLER,** Fraunhofer IESE ٠ Kasia CHMIELINSKI, Dataset Nutrition label Francisco HERRERA, Univ Granada ٠ Flora DELLINGER, Confiance.ai David REICHEL, FRA ٠ Fred MORSTATTER, ISI **Rapporteurs**: **Isabelle Hupont Torres,** JRC **Rapporteurs**: Maurizio Salvi and Alexandra Balahur, JRC



Bernard MAGENHANN, JRC Deputy DG Stefano CALZOLARI, CEN President





Data in the context of the AI Act

Lucilla SIOLI

Director DG CNECT







ENSURING AN ETHICAL USE OF AI TECHNOLOGIES TO THE BENEFIT OF HUMANITY

Gabriela RAMOS

Assistant Director-General, UNESCO





OECD AI Data quality principles on trustworthiness, human rights and democratic values

Karine PERSET

Head of AI Unit OECD







IEEE P7003 Standard for Algorithmic bias considerations

Ansgar KOENE

Chair IEEE P7003 Working Group



The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems





IEEE 7000-2021: Model Process for Addressing Ethical Concerns During System Design **IEEE 7001-2022**: Transparency of Autonomous Systems **IEEE P7002**: Data Privacy Process **IEEE P7003**: Algorithmic Bias Considerations IEEE P7004: Child and Student Data Governance **IEEE P7005:** Employer Data Governance **IEEE P7006:** Personal Data AI Agent Working Group **IEEE 7007-2022:** Ontological Standard for Ethically Driven Robotics and Automation Systems **IEEE P7008**: Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems **IEEE P7009**: Fail-Safe Design of Autonomous and Semi-Autonomous Systems **IEEE 7010-2020**: Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being **IEEE P7011:** Process of Identifying and Rating the Trustworthiness of News Sources **IEEE P7013:** Benchmarking of Automated Facial Analysis Technology



IEEE STANDARDS ASSOCIATION

Standards

Products & Services Te

Technologies & Initiatives

Participate Search

Q

MAC ADDRESS

BUY STANDARDS

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)

Developing metrics and processes towards the implementation of a certification methodology addressing transparency, accountability and algorithmic bias

< Industry Connections

The Ethics Certification Program for Autonomous and Intelligent Systems



About

The goal of The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) is to create specifications for certification and marking processes that advance transparency, accountability and reduction in algorithmic bias in Autonomous and Intelligent Systems (A/IS).

Algorithmic Bias Considerations



- ► All non-trivial^{*} decisions are biased
- We seek to minimize bias that is:
 - Unintended
 - *Unjustified
- as defined by the context where the system is used.

*Non-trivial means the decision has more than one possible outcome and the choice is not uniformly random.

Overview of IEEE P7003









Thank you for your attention



Dr. Ansgar Koene ansgar.koene@ieee.org

www.cencenelec.eu



Tag us @standards4EU





Towards a Standard for Identifying and Managing Bias in Artificial Intelligence

Reva SCHWARTZ

Principal investigator for AI bias, NIST



Towards a Standard for Identifying and Managing Bias in Artificial Intelligence* National Institute of Standards and Technology Information Technology Laboratory

Reva Schwartz, Apostol Vassilev, Kristen Greene, Lori Perine (NIST ITL) Andrew Burt, Patrick Hall (BNH)



National Institute of Standards and Technology U.S. Department of Commerce *NIST Special Publication 1270, March 2022; https://doi.org/10.6028/NIST.SP.1270

Taxonomy of Al Bias

Key takeaways

Risks from AI are

realcontribute to real-world harmsdifficult to measure.

- There is much more to bias than datasets and algorithms.
- Governance and related structural and cultural practices are essential for transforming the way we think about and approach AI risks such as bias.



What's missing?





Current focus on computational/statistical bias obfuscates the other two categories



Currently, lifecycles track processes and ensure delivery of high-performing functional technology—but do not identify or manage harms

Transforming Culture – Socio-technical Systems Approach

Takes into consideration the larger social system in which AI operates, its purpose and potential impacts

- Manage bias within/connected to specific operational context
 - utilize broader set of perspectives and expertise
 - apply human-centered design to AI systems
- Apply the **scientific method** to AI systems
- Set up **governance** structures for the people who build and maintain AI systems
- Consideration of **limitations** from an impact and values-based perspective



Balancing representation in datasets is merely a first step

Mathematical approaches alone are not able to capture information related to contextuality, contestability, and procedurality.

Are there suitable datasets *for the purpose* of the various applications, domains and tasks for which the AI system is being developed and deployed?



THANK YOU



Contact us via email at <u>ai-bias@list.nist.gov</u>

For more info on the NIST AI Bias work, visit https://www.nist.gov/artificialintelligence/ai-fundamental-researchfree-bias



Data in the context of the AI Act

Gabriele MAZZINI

DG CNECT Team Leader for the AI Act



Article 8 - Compliance with the requirements





- 1. High-risk AI systems shall comply with the requirements established in this Chapter.
- The intended purpose of the high-risk AI system and the risk management system referred to in Article 9 shall be taken into account when ensuring compliance with those requirements.

Article 3(12) – intended CENELEC CENELEC Commission

'intended purpose' means the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation;

- Key concept to AI Act proposal
 - ► Common to EU product legislation → the product shall be safe and compliant when used in accordance with the product information
 - Relevant for classification as high-risk (Art. 6 and Annex III and amendment thereof)
 - ▶ Relevant for compliance with requirements (always taking into account intended purpose)
 - Relevant for risk management

▶ ...

Article 9 - Risk management system



- Continuous iterative process throughout the entire lifecycle of AI system
 - identification and evaluation of known and foreseeable risks associated with AI system, including after placing on the market (through post-market monitoring)
 - suitable risk management measures:



"due consideration to the effects and possible interactions resulting from the combined application of the requirements" and taking "into account the generally acknowledged state of the art"

Art. 10 - Data and data governance



Appropriate data governance and management practices

- ► Relevant **design choices**
- ► Data collection
- ► Data **preparation and processing operations** (e.g. annotation, labelling, cleaning, enrichment, aggregation)
- ► Formulation of relevant assumptions (what data shall measure/represent)
- Assessment of availability, quantity and suitability of the data sets needed
- ► Examination of **possible bias**
- Identification of possible data gas and shortcomings (and how they can be addressed)

High-quality data

Art. 10 - Data and data governance



JICH AST

Relevant, representative, free of errors, complete and with appropriate statistical properties, including as regards the persons on which the AI system is intended to be used

High-quality data

Take into account **specific geographical, behavioural or functional setting** within which the AI system is intended to be used

presumption of compliance if AI system has been trained and tested on data concerning that specific geographical, behavioural and functional setting

Legal basis to process special categories of personal data for the purposes of ensuring bias monitoring, detection and correction



Thank you for your attention

gabriele.mazzini@ec.europa.eu



Overview on standardization road mapping on ESO and ISO level

Sebastian HALLENSLEBEN Chair CEN-CLC JTC 21 AI Patrick BEZOMBES Vice-chair CEN-CLC JTC 21 AI Convenor ISO/IEC JTC 1/SDC 42/AG 3 (AI roadmapping)



SC 42 work program



✤ <u>35 standards published or under development and more to come</u>

- > 2 foundational standards on terminology (publication in the coming weeks)
- > 1 standard on "AI management system" (under development)
- 11 standards on trustworthiness characteristics:
 - Safety, Robustness, Controllability , Explainability, Quality, Bias...
- 4 International Standards projects on Data quality (early stage)

✤ <u>Gaps:</u>

- Some trustworthiness characteristics are not addressed (yet): AI oversight, Sustainability, Accuracy, Fairness,...
- Verifiable technical specifications, metrics and controls for trustworthiness characteristics
- Conformity assessment schemes

Future potential work item/area:

Strategies for Mitigating Ethical and Societal Concerns (considered)

▶

SC 42 work items related to data quality and bias



			rustworthiness	afety	scurity	uality	Trustwo	rthiness outrolability	<pre>cplainability</pre>	ransparency	las	ccuracy	ata (Quality and reference architecture)	Da	esting and evaluation	/nthetic data	Sustain ability	onformity assessment	
	Stage	Target publication date	-	Š	Š	ď	Å	Ŭ	£	Ē	ä	Ă	Ő	ā	Ĕ	S	Ш	Ŭ	
SO/IEC DTR 24368	50.00	juil-22															х		Overview of ethical and societal concerns
SO/IEC TR 24027:2021	60.60	2021									Х								Bias in Al systems and Al aided decision making
SO/IEC AWI 5259-1	20.20	juil-23																	Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples
SO/IEC AWI 5259-2	20.00	janv-24				X							X						Data quality for analytics and machine learning (ML) — Part 2: Data quality measures
SO/IEC AWI 5259-3	20.00	janv-23				x							X						Data quality for analytics and machine learning (ML) — Part 3: Data quality management requirements and guidelines
SO/IEC AWI 5259-4	20.00	juil-23											X						Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework

Stage code Stage

- 0 Preliminary
- 10 Proposal
- 20 Preparatory
- 30 Committee
- 40 Enquiry
- 50 Approval
- 60 Publication

C (

JTC 21 standardization global approach



Support European Union legislation:

- Current priority: harmonized standards to underpin the AI Act (by Autumn 2024)
- Further work (already started):
 - standards to support data regulation (data governance, data space...) where relevant for AI
 - standards to support sustainability regulation (Green Deal) where relevant for AI

Support European industry needs:

- Simplicity and clarity of AI standards, minimization of effort to confom to AI Act
- Minimization of divergence between European and international level
- Coverage of organisational, process, product and engineering perspectives

What makes "good" European AI standards?

- **No reinventing the wheel**: Use of ISO-IEC standards whereever possible
- Supporting innovation and competition:
 - Actionable(*) standards or « Make it simple for innovation »
 - Coherent with other standards and frameworks ETSI, ISO/IEC, OECD, UNESCO, IEEE, GAIA-X, CAI / Council of Europe, ...

Satisfying European specificities and requirements:

- > EU values and principles, including protections for citizens and consumers
- > EU AI Act, with its timeline
- Strategic goal: EU should be a global standard setter, not just a standard taker

^(*) Actionable:

- Comprehensive
- Operationalizable specifications and metrics
- Innovation-friendly



Steps towards AI standards supporting the AI Act





Anticipate the AI Act \rightarrow JTC 21 AI standardization roadmap





Request for AI standards by the European Commission in anticipation of the approval of the AI Act

Step 3

Joint CEN-CENELEC and ETSI response and integration into work programmes

Step 4

On a case-by-case basis, CEN-CENELEC and ETSI will -

- adopt/adapt ISO-IEC standards, or
- develop homegrown standards

Examples of current JTC 21 work to support the AI Act



(*) Trustworthiness characteristics:

fairness...

Robustness, explanability, accuracy, safety,

AI Terminology

Adopt/adapt ISO-IEC standards (ISO/IEC FDIS 22989 and ISO/IEC FDIS 23053)

AI management system & AI risk management

- □ Adopt/adapt ISO-IEC standards (ISO/IEC CD 42001 and ISO/IEC FDIS 23894)
- Preliminary work on « AI risk catalogue and risk management » as potential homegrown initiative

AI Trustworthiness characteristics (*)

Numerous initatives, many standards, but gaps, partial coverage of the EU requirements, time development concerns

Preliminary work on « Overarching unified approach on trustworthiness characteristics» as potential homegrown initiative

AI conformity assessment

- Analysis launched
- □ Concerns regarding the competencies of notified bodies, for high-risk AI systems certification



Analysis to be launched

Audience interaction







Let's take a break!





COFFEE BREAK

Parallel sessions start at 15:45 Make sure to connect in time!