ISO/IEC JTC 1/SC 42 "Artificial intelligence"
Secretariat: **ANSI**
Committee manager: **Benko Heather Ms.**

WG 3 Convenor Closing Report - October 2024

Document type	Related content	Document date	Expected action
Meeting / WG report	Meeting: Le Chesnay-Rocquencourt (France) 7 Oct 2024	2024-10-11	**INFO**

SC 42 (Artificial Intelligence)
WG 3 (Trustworthiness)

Opening Plenary Report, Versailles, October 2024

David Filip, Convenor SC 42/WG 3

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS

Final actionable items for the closing plenary

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

- Proposal to form a liaison
  - ISO/TC 145, nominating Tom Lebrun (Canada)
- NP ballots
  - [RoadmapAHG] → A Template for Documentation of ethical implications of an AI application NP TS
  - [Reliability] PWI 42118 → [Reliability] NP 42118
- Change of track
  - ISO/IEC AWI 25059 [SQuaRE4AI] 2nd ed 24 → 36 months
  - ISO/IEC AWI 25029 [Nudge] 18 → 36 months
- Scope clarification for [Nudge] [see next page]
- Editor re-appointments as necessary
  - Change of PWI 18966 editor Takashi Egawa (Japan) → Nicholas Poirier (Canada)

Scope Editorial Change - AI enhanced nudging (25029)

Old Scope

The document provides definitions, concepts, guidelines, and methodologies to address AI-enhanced nudging mechanisms by organizations. The document also provides requirements for designing Responsible AI-enhanced nudging mechanisms, key indicators, both horizontally (by industry and sectors) and vertically (by applications and technologies), processes for self-assessment or/and third-party audits to build a trustworthy ecosystem. The document will support organizations that develop or use AI-enhanced nudges, or any entity interested in the protection of the civil society or individuals , including consumers and workers.

New Scope

This document provides definitions, concepts, guidelines, and a **methodology** to address AI-enhanced nudging mechanisms.

It also specifies requirements for designing responsible AI-enhanced nudging mechanisms and key indicators. **This document is applicable horizontally.** This document supports organizations that develop or use AI-enhanced nudges, as well as entities interested in the protection of **individuals, groups of individuals and society as a whole**, including consumers and workers.

- [Changed because it provides a consistent body of methodology not a number of competing methodologies]
  ~~methodologies~~ > **methodology**
- [Used a more precise verb but added "also" for readability]
  ~~The document also provides~~ → **It also specifies**
- [simplified to avoid confusion]
  ~~both horizontally (by industry and sectors) and vertically (by applications and technologies~~ _> This document is applicable horizontally.
- [Deleted not to violate CA neutrality]
  ~~processes for self-assessment or/and third-party audits to build a trustworthy ecosystem~~
- [Changed to align with 42005]
  ~~civil society or individuals~~ > **individuals, groups of individuals and society as a whole**

Acknowledgements

Congratulations to Olivier (ISO/IEC TS 25058:2024), Takashi and Riccardo (ISO/IEC TR 5469:2024) and Xiaoqui (ISO/IEC TS 8200:2024) for reaching publication!

Thanks to:

all editors, AHG convenors, and experts for their hard work between the plenary weeks;

Domenico Natale (SC 7 liaison) for invaluable input to 25059, 25058, and other discussions.

Agenda

Recommendations to the closing plenary

Group Overview and Officers

Meeting History

Progress of Work

- Projects Status
- Study Items
- Assigned Tasks

Key Issues

Publications

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

TR 24028:2020 .. Overview of trustworthiness in Artificial Intelligence [Overview] **Stage 60.60**

TR 24029-1:2021 .. Assessment of the robustness of neural networks -- Part 1: Overview [RNN-1] **Stage 60.60**

TR 24027:2021 .. Bias in AI systems and AI aided decision making [Bias] **Stage 60.60**

TR 24368: 2022 .. Overview of ethical and societal concerns [Ethics] **Stage 60.60**

23894:2023 .. Guidance on risk management [RiskMgmt] **Stage 60.60**

24029-2:2023 .. Assessment of the robustness of neural networks -- Part 2: Methodology for the use of formal methods [RNN-2] **Stage 60.60**

25059:2023 .. (SQuaRE) – Quality model for AI systems [SQuaRE4AI] Stage **Stage 90.92**

TR 5469:2024 .. Functional safety and AI systems [FSafety] **Stage 60.60**

TS **25058** [was: 5471] .. Guidance for quality evaluation of AI systems [QuEG4AI], **Stage 60.60**

TS 8200:2024 .. Controllability of automated artificial intelligence systems [Control], **Stage 60.60**

- Editors: Orit Levin [Overview] Arnault Ioualalen [RNN-1 & RNN-2]; Adam Leon Smith, [Bias & SQuaRE4AI]; Viveka Bonde [Ethics]; Peter Deussen [RiskMgmt]; **Olivier Blais [QuEG4AI]; Xiaoqi Cao [Control]; Takashi Egawa and Riccardo Mariani [FSafety]**;

Group Overview and Officers

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence

WG 3 Trustworthiness of AI

CD TS 6254 .. Objectives and approaches for explainability of ML models and AI systems [XAI] Stage 30.60→50.00 **[9 months extension granted, CRM concluded with strong consensus, preparing DTS, editorial issues before DTS launch]**

CD 12792 .. Transparency taxonomy of AI systems [Transparency], Stage 40.60→50.00 **[DIS CRM in progress]**

DTS 12791 .. Treatment of unwanted bias in classification and regression machine learning tasks [Bias], Stage 60.00 **[2 DTS ballot → publication]**

AWI TR 42106 .. Overview of differentiated benchmarking of AI system quality characteristics [Layers], Stage 20.00→30.00 [CD progression consensus]

AWI 24029-3 .. Assessment of the robustness of neural networks -- Part 3: Methodology for the use of formal methods [RNN-3], Stage 20.00 **[changed track to 36 months by CIB ballot]**

AWI TS 2243 .. Guidance on addressing societal concerns and ethical considerations [Ethics], Stage 20.00→30.00 [CD progression from Versailles]

AWI 42105 .. Guidance for human oversight of AI systems [Oversight], Stage 20.00 [CD progression expected within a month after Versailles]

AWI 25059 .. [2nd edition] Quality model for AI systems [WD in progress][consider changing track to 36 months]

AWI 25029 .. AI-enhanced nudging [Nudging], Stage 20.00 [WD in progress]

PWI 17866 .. Best practice guidance for mitigating ethical and societal concerns [Ethics], Stage 00.00 **[kept open in Seoul]**

PWI 18966 .. Oversight of AI systems [Oversight], Stage 00.00 **[produced AWI 42105, kept open to incubate an NP on governance and management aspects of human oversight → SC 40 consultation → NP out of Versailles]**

PWI 42108 .. Operational design domain (ODD) for AI systems [Form 4 → CIB]

PWI 42118 .. Reliability of AI systems **[proposing NP ballot in Versailles]**

PWI 42117 .. Trustworthiness fact labels for AI systems [early stages]

PWI 24029-5 .. Applicability of the methodology to other artificial intelligence algorithms [in progress]

Officers

- Convenor: David Filip | Secretary: Aditya Mohan (NSAI) | Roadmapping AHG Co-Convenors: David Wotton & Harm Ellens
- Editors: Arnault Ioualalen [RNN-3]; Adam Leon Smith, [Bias]; Viveka Bonde [Ethics]; Jaeho Lee [XAI], Olivier Blais [SQuaRE4AI]; Rania Wazir [Transparency]; Takashi Egawa [Oversight]; Nisheeth Srivastava [Layers]; Enrico Panai [Nudging]; C Anantaram [Reliability];

Meeting History (including associated CRMs)

38+3 (41) teleconferences held since the 13th Plenary in Seoul in April 2024

4 [SQuaRE4AI] – (2nd ed.) AWI

3 [Nudge] – AWI [parallel with JTC 21]

2 [RNN-3] – AWI

0+0 [XAI] – CD CRM implementation → DTS

0+0 [Bias] – 2 DTS ballot → Publication

0+3 [Transparency] – DIS → DIS CRM

3 [Layers] – WD → CD

5 [Ethics] – WD development

6 [Oversight] – WD development and NP preparation

3 [ODD] – NP preparation → decision making to roadmap → CIB

5 [Reliability] – NP preparation

1 [TFL] – initial discussions

6 [RoadmapAHG] – [NIST gap analysis AhG→]developing Trustworthiness characteristics matrix, **PWI**s, and **NP**s

Progress of Work – Project Status

Project Plans

- ISO/IEC CD TS 6254 – [XAI]
  - **Target and limit dates after change of track approved: DTS 2023-04-14, TS 2023-12-30 [2024-02-11]**
  - Current Stage: CD CRM → DTS
  - **Project made great progress after F2F meetings in Berlin and Vienna. CRM concluded with strong consensus, but complex implementation**
  - **Slowed down on implementation and editorial issues before launching DTS**
  - **9 month extension granted in Vienna, new limit date 2024-11-10**

Progress of Work – Project Status

Project Plans

- ISO/IEC 1st DTS CRM→ 2nd DTS 12791 – [Bias]
  - Plan and target dates after change of track: Approved **2022-02-14**, WD 2022-03-15, CD 2022-12-31 DTS 2023-08-22, TS 2024-12-30 [2025-02-14]
  - Current Stage: 2nd DTS succeeded 28:1:11 → publication
  - Track changed to 36 months to allow for 2nd DTS with CEN-CENELEC/JTC 21

Progress of Work – Project Status

Project Plans

- ISO/IEC CD 12792 – [Transparency]
  - Plan and target dates: Approved **2022-02-15**, WD 2021-04-15, CD 2023-02-15, DIS 2023-12-30 [2024-02-14], FDIS 2024-07-07 [2024-11-04], IS 2025-02-14 [2025-02-14]
  - Current Stage: **DIS (left OSD as OSD doesn't support parallel balloting with CEN)**
  - About 2 months delay caused by CD admin issues
  - DIS CRM expected to finish in Versailles → FDIS

Progress of Work – Project Status

Project Plans

- ISO/IEC PWI 17866 – [Ethics]
  - Kept open in Berlin and Vienna, consider closing
    - **Produced AWI TS 2243**
    - **If no other project idea emerges in Versailles should be closed**

Progress of Work – Project Status

Project Plans

- ISO/IEC AWI TS 22443 – [Ethics]
  - Plan and target dates: Approved **2023-09-05**, CD 2024-08-29, **TS 2025-07-31 [09-05]**
  - Progression to CD expected in Versailles

Progress of Work – Project Status

Project Plans

- ISO/IEC AWI TR 42106 – [Layers]
  - Plan and target dates: Approved **2023-05-04**, CD 2024-10-24, **TR 2025-02-28 [05-04]**
  - CD progression agreed before Versailles
  - Exploration of potential normative sequel moved under [TFL]

Progress of Work – Project Status

Project Plans

- ISO/IEC PWI 18966 – [Oversight]
  - **Kept open in Berlin and Vienna**
  - **Bottom up [~Controllability + human~]**
    - **Building on technical and terminology prerequisites from [Control] ISO/IEC TS 8200:2024 → AWI 42105**
  - **Top down [~Governance and management implications of human oversight~]**
    - **Building on ISO/IEC 38507:2022, key liaison with SC 40**
    - **Likely to propose NP to the closing plenary**
    - **[WRT SC 40 enhanced liaison or JWG?]**

Progress of Work – Project Status

Project Plans

- ISO/IEC AWI 42105 – [Oversight]
  - Plan and target dates: Approved **2023-08-20**, CD 2024-09-29, DIS 2025-02-28 [08-20] **IS 2026-03-31 [08-20]**
  - Developing WD on OSD. Tracking regional regulatory requirements and recommendations
  - Aiming for CD within a month after Versailles

Progress of Work – Project Status

Project Plans

- ISO/IEC AWI 24029-3 – [RNN-3]
  - Plan and target dates [after change of track]: Approved **2023-08-21**, **CD 2024-11-30?**, DIS 2024-07-31 [2025-08-21], **IS 2025-07-31 [26-08-21]**
  - WD development on OSD last cycle and this week
  - Slower than expected WD progress (2 project meetings cancelled due to lack of contributions). Need more robust expert engagement
  - Track changed to 36 month via CIB between plenaries

Progress of Work – Project Status

Project Plans

- ISO/IEC AWI 25059 (2nd ed.) – [SQuaRE4AI]
  - Plan and target dates: Approved **2024-03-05**, **CD 2024-08-31**, DIS 2025-02-28 [03-05], **IS 2025-08-31 [2026-03-05]**
  - 2nd edition of 25059:2023
    - To update SC 7 normative references
    - To add a 3rd quality model for AI services
  - Good WD but complex situation consider change of track to 36 months
  - Current track still feasible if CD reached by end of 2024

Progress of Work – Project Status

Project Plans

- ISO/IEC AWI 25029 – [Nudge]
  - Plan and target dates: Approved **2024-04-04 [04-29]**, **CD 2024-06-05**, DIS 2025-01-31, **IS 2026-01-31**
  - Joint development with ISO lead of a project started at CEN-CENELEC/JTC 21
  - Aggressive timeline, challenging NP ballot comments
    Can track be changed? Given that the project has been previously developed in JTC 21?
  - Consider change of track to 24 or even 36 months
  - **Manual transfer from CEN OSD to ISO OSD, wrong editor info from JTC 21**
  - 1st couple of international meetings with acting editor, major changes to WD agreed
  - Editorial scope clarification needed
  - Enrico Panai confirmed as Project Editor

Progress of Work – Project Status

Project Plans

- ISO/IEC PWI 42108 – [ODD]
  - Recommended a CIB ballot to support NP progression .. *Terminology and concepts for domain engineering*
    - **CIB ballot closed on 23 September**
    - **[new directives→] Project will autoclose if CIB and NP succeed.. Will be kep open in case CIB or NP fail..**

Progress of Work – Project Status

Project Plans

- ISO/IEC PWI 42118 – [Reliability]
  - Good progress on Form 4 refinement
  - **Likely to prose NP ballot to closing plenary**

Progress of Work – Project Status

Project Plans

- ISO/IEC PWI 42117 – [TFL]
  - Platform to brainstorm CA labeling ideas for Trustowrthiness characteristics and to communicate them with CASCO

  - **No plenary action likely in Versailles**
  - **More meetings next cycle**

Progress of Work – Project Status

Project Plans

- ISO/IEC PWI 24029-5 – [RNN-5]
  - Platform to brainstorm CA labeling ideas for Trustowrthiness characteristics and to communicate them with CASCO
  - **Possible NP ballot recommendation to the closing plenary**

Progress of Work – Study Items

See [WG 3 N25]

Developing a matrix method to display coverage of trustworthiness characteristics (coordinating with AG 3)

Security and **Privacy** - following AHG 3, PWIs and NPs in SC 27

Recently developed approved items address

Functional Safety, Explainability, Quality Evaluation, Controllability, Oversight, Bias, Transparency

Also developing an overall quality/trustworthiness model based on the SC 7 SQuaRE methodology [SQuaRE4I], [QUEG4AI]

New gap analysis exercise driven by NIST RMF → new characteristics proposed for WG 3 trustworthiness matrix, likely to look at Singapore regulation

Outstanding characteristics include: Safety (in general), Robustness of AI (as opposed to NN only)

Currently running 6 PWI projects apart from general roadmapping activity

[Ethics], [Oversight], [ODD], [Reliability], **[RNN-5], [TFL]**

Roadmapping activity developed proposal for Risk management of generative AI, CIB to launch NP succeeded 23:6:11, i.e. 79.3% mandate to launch the NP

Planning to recommend some NP ballots

[Oversight] → [Governance and management implications of human oversight] NP or NP TS [in enhanced liaison or mode 5 with SC 40]

[RoadmapAHG] → [~ethical implications documentation template~] NP

[RoadmapAHG] → [~watermarking~] NP

Key Issues

- Concept and terminology coordination issues across projects and WGs
  - Oversight (WG 3) vs Human-machine teaming (WG 4)
  - explainability, interpretability, transparency, automation, quality model for data, engineering life cycle, verification & validation
- Gap analysis based on EU AI Act, US NIST RMF, Singapore fwk → need to continue transforming identified characteristics into NP proposals
- Need of **verifiable requirements standards** suitable to be harmonized with regulations <-> Vienna-Frankfurt agreements based relationship with CEN-CLC/JTC 21

Expected actionable items for the closing plenary

- Planning to recommend NP ballots
  - [RoadmapAHG] → [~watermarking~] NP
  - [RoadmapAHG] → [~ethical implications documentation template~] NP
  - [Oversight] PWI 18966 → [~governance and management implications of human oversight~] NP TS or NP 42105-2
  - [Reliability] PWI 42118 → [Reliability] NP 42118
  - [RNN-5] PWI 24029-5 → [RNN-5] NP 24029-5
- Change of track
  - [SQuaRE4AI] 2nd ed 24 → 36 months
  - [Nudge] 18 to 24 or 36 months [if possible in parallel development]
- Scope clarification for [Nudge]
- Close [Ethics] PWI
- Editor re-appointments as necessary

Final actionable items for the closing plenary

- Planning to recommend NP ballots
  - ~~[RoadmapAHG] → [~watermarking~] NP~~ [needs more work] →
    → Form liaison with ISO/TC 145 Tom Lebrun (Canada) to be appointed as liaison officer
  - [RoadmapAHG] → A Template for Documentation of ethical implications of an AI application NP TS
  - [Oversight] PWI 18966 → ~~[~governance and management implications of human oversight~] NP TS or NP 42105-2~~ → [needs more work] →
    Change of PWI 18966 editor Takashi Egawa (Japan) → Nicholas Poirier (Canada)
  - [Reliability] PWI 42118 → [Reliability] NP 42118
  - ~~[RNN-5] PWI 24029-5 → [RNN-5] NP 24029-5~~ → [needs more work]
- Change of track
  - [SQuaRE4AI] 2nd ed 24 → 36 months
  - ISO/IEC AWI 25029 [Nudge] 18 → 36 months
- Scope clarification for [Nudge]
- ~~Close [Ethics] PWI~~ [continued discussions until May 2025]
- Editor re-appointments as necessary

SC 42 – Artificial Intelligence