



SC 42's Joint Working Group 5

Natural Language Processing

Lauriane Aufrant

– *Co-Convenor ISO/IEC JTC 1/SC 42/JWG 5*

SC 42 Workshop @ Seoul – 24/04/2024

JWG 5 “Natural Language Processing”

Created in **October 2023**, kick-off in **January 2024**

Collaboration from 2 committees:

- ▶ ISO/IEC JTC 1/SC 42 “Artificial Intelligence”
- ▶ ISO/TC 37 “Language and terminology”

Led by 2 Co-Convenors:

- ▶ SC 42 side: Lauriane Aufrant (France)
- ▶ TC 37 side: Avashlin Moodley (South Africa)

Two work items:

- ▶ TR 23281 on NLP tasks: machine translation, speech recognition, question answering, information retrieval, OCR, sentiment analysis, speaker recognition...
- ▶ AWI 23282 on NLP evaluation

The European context: CEN/CLC/JTC 21 & the AI Act

The AI Act: a European regulation on AI, supported by standards



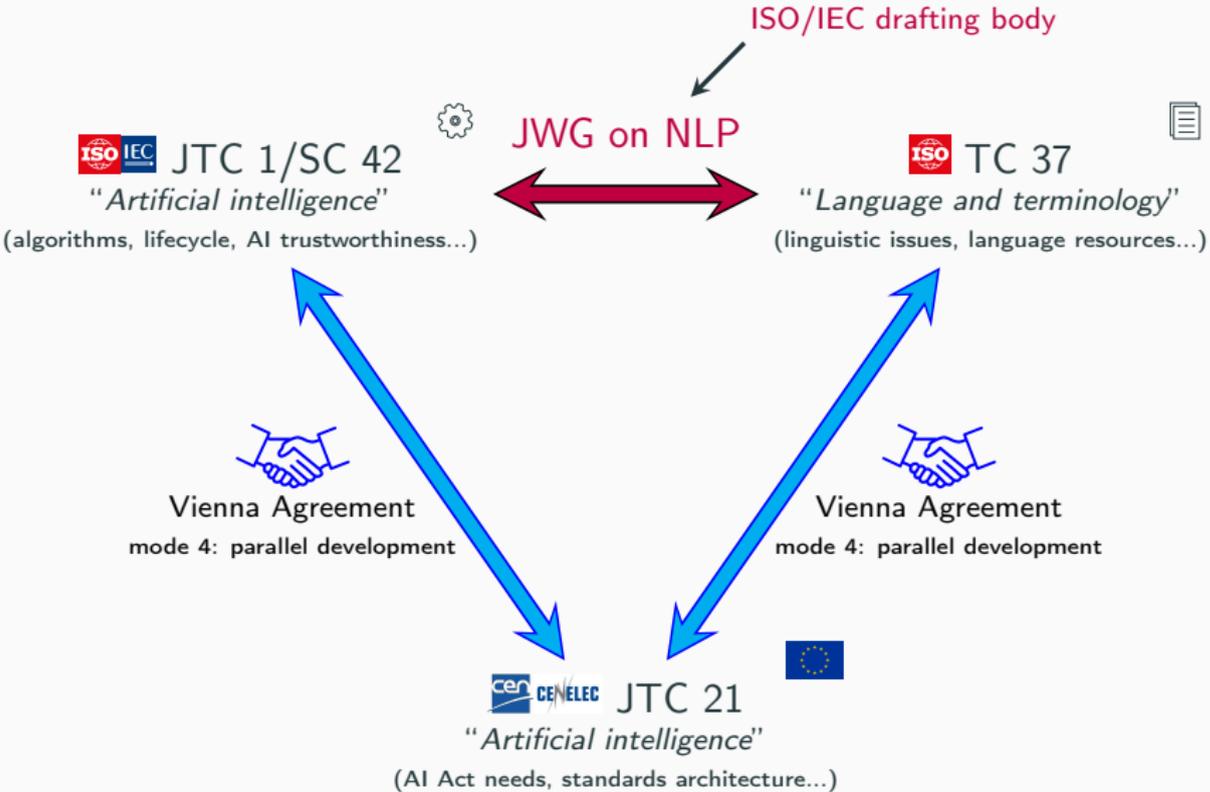
European Commission's Standardization Request to JTC 21:

1. risk management
2. **data quality & governance**
3. record keeping & logging
4. **transparency**
5. human control
6. **accuracy**
7. **robustness**
8. **cybersecurity**
9. quality management
10. conformity assessment

↔ AI Act adopted in **2024**, entry into force **mid-2025** for general-purpose AI, **mid-2026** for high-risk AI systems

↔ Standards for high-risk AI systems to be published by **mid-2025**

A 3-way collaboration



Incubating an NLP standardization roadmap

Filling gaps with respect to SC 42 standards (AI) + TC 37 standards (language) + *de facto* NLP standards:

- ▶ **Foundational:** **tasks (landscape & terminology)**, domains (taxonomy & criteria), documentation (needs, templates, incorporating academic checklists...)
- ▶ **Evaluation:** **metrics (incl. NLP-specific)**, **good practices, guidelines for qualitative / human evaluation**, thresholds, **test sets...**
- ▶ **Quality beyond evaluation:** multilinguality (concepts & criteria: how many? how diverse? model vs system?), frugality (data), efficiency (compute), speed (measures), robustness (concepts, criteria, measures, taxonomy...), bias (concept, detection, mitigation), interpretability & explainability...
- ▶ **Data and formats:** annotation formats (expanding TC 37 work to other tasks), I/O formats (for *system* interoperability), annotation guidelines, annotation process (good practices, quality control, IAA & adjudication...)
- ▶ **Approaches and methods:** taxonomy of NLP methods (relying on a taxonomy of AI methods)

ISO/IEC TR 23281 *Artificial intelligence – Overview of AI tasks and functionalities related to natural language processing*

Editor: Lauriane Aufrant (France)

- ▶ Landscaping tasks (+ other functionalities, e.g. domain adaptation) in the whole NLP spectrum
 - ↔ including speech, sign languages, multimodality...
- ▶ Addressing terminological conflicts: ambiguity of “*question answering*”, “*information extraction*”, “*named entity*”...
 - ↔ seeking technical clarity on inputs/outputs
- ▶ Insights on combining modular tasks into higher-level functions
- ▶ Identification of existing standards (incl. *de facto*) per task

Timeline: start 10/2022 (in JTC 21), in parallel development since 02/2024, targeted publication 10/2024

International Standard on NLP evaluation

ISO/IEC AWI 23282 *Artificial intelligence – Evaluation methods for accurate natural language processing systems*

Editor: Rania Wazir (Austria)

- ▶ Defining NLP-specific metrics + other evaluation protocols (human, qualitative...) in a non-ambiguous way
- ▶ Mapping tasks (from TR) to evaluation metrics/protocols
- ▶ Calibration, interpretation of measures
- ▶ Considerations for metric implementation, under-specified formulas... to ensure reproducibility & comparability
- ▶ Requirements on evaluation resources, e.g. non-trivial test set

Timeline: start 01/2024, consultation mid-2024, ballot Q4 2024, publication mid-2025 – *in sync with the EU AI Act timeline*

Chatbots & related systems:

- ▶ Large Language Models belong to NLP
- ▶ Generative AI belongs *in part* to NLP

↔ strong interplay with other working groups: terminology, testing, methods, trustworthiness, use cases...

NLP for standardization...?

- ▶ Assistance for drafting
- ▶ Translating standards
- ▶ Supporting compliance
- ▶ ... any suggestion?

Thank you for your attention!