ISO Form 4
NEW WORK ITEM PROPOSAL (NP)

Circulation date: 2024-10-29 **Closing date for voting:** 2025-01-27	**Reference number:** ISO/IEC NP TS 25569 ISO/IEC JTC 1/SC 42 **N 1952**
Proposer ISO/IEC JTC 1/SC 42	
Secretariat ANSI	

A proposal for a new work item within the scope of an existing committee shall be submitted to the secretariat of that committee.
A proposal for a new project committee shall be submitted to the Central Secretariat, which will process the proposal in accordance with ISO/IEC Directives, Part 1, Clause 2.3.
Guidelines for proposing and justifying new work items or new fields of technical activity (Project Committee) are given in ISO/IEC Directives, Part 1, Annex C.
IMPORTANT NOTE: Proposals without adequate justification and supporting information risk rejection or referral to the originator.

☒ The proposer confirms that this proposal has been drafted in compliance with Annex C of ISO/IEC Directives, Part 1.

PROPOSAL

(to be completed by the proposer, following discussion with committee leadership if appropriate)

TITLE **English title:** Artificial Intelligence -- Implementation guidance on de-identification of data used in Machine Learning (ML) **French title:** *(In the case of an amendment, revision or a new part of an existing document, show the reference number and current title)*
SCOPE This document provides implementation guidance on de-identifying data used in machine learning. The guidance includes methods that can be used during machine learning model development.

PURPOSE AND JUSTIFICATION

For a Machine Learning (ML) model to be effective, the richness of data used to train the Machine learning algorithms is crucial. It is equally important to ensure through use of appropriate de-identification methods, that the data use in ML does not lead to violation of data protection regulations.
De-identification of data will be needed for various reasons including facilitation of the below:
a) Non-disclosure of personal data during model use
b) Prevent identity theft while using biometric information such as face and voice samples
c) Avoid exposure of personal information from targeted attacks such as membership inference, model inversion, and data extraction attacks
d) Eliminate the need to have legal basis for processing by converting them to non-personal data or not collecting the personal data
While by design we don’t expect ML to store data - only the patterns and relationship through attributes such as parameter, weightages, but in reality, the raw data memorization is not uncommon and there were number of instances reported where AI model regurgitated personal information from training data such as email address, telephone number etc. Stable diffusion for instance reportedly memorizes individual images from their training data and emit them at generation time (see: https://arxiv.org/abs/2301.13188)
Although there are means to remove data such as machine unlearning, inductive graph unlearning, and approximate data deletion, it will be constrained by our incomplete knowledge on presence of specific data within the model which we want ML to forget/delete.
The proposed technical specification is expected to build further on ISO/IEC 27091, ISO/IEC 20889, ISO/IEC 27559 and other evolving standards to recommend methods to de-identify data and procedures for embedding these in the AI life-cycle, in the context of ML, while balancing the value and utility. It is easy to remove quasi-identifiers to de-identify data but such exercise invariably renders data useless for ML since it is only through such characteristic attributes that algorithms derive rich patterns that is central to AI. While ISO/IEC 20889 and ISO/IEC 27559 lists de-identification methods, they do not give implementation guidance in the AI life cycle environment.
Following are some of the standards which has reference to De-Identification, however there is no implementation guidance in any of the existing or under development standards:
ISO/IEC 5259-1:2024, Clause 5.3.3.3 states datasets used for ML and analytics can contain PII, which should be protected in accordance with applicable requirements throughout all stages of the DLC model. De-identification techniques can be used to remove PII but production data used to make predictions about individuals can still contain or be linkable to PII.

ISO/IEC 27559:2022 proposes use of de-identification techniques in order to support compliance with regulatory requirements and relevant privacy principles.

ISO 25237:2017 in Clause 5 highlights the use of de-identification to reduce privacy risks in wide variety of situations.

ISO/IEC 27701 in clause B.1.4.5 states that the organization should define and document data minimization objectives and what mechanisms (such as de-identification) are used to meet those objectives.

ISO/IEC 22989:2022 in clause 5.10 necessitates the use of de-identification or other processes, which can be required if the dataset includes personally identifiable information (PII) or is associated with individuals or organizations, before the data can be used by the AI system.

ISO/IEC 23053:2022 in clause 8.3 includes reference to de-identification in relation to data preparation stage. It also states although the acquired data can have been de-identified earlier, additional de-identification can be required because of data processing (e.g. joining datasets) in this stage.

The following standards (published or under development) with relation to the proposed technical specifications exists. However, there is no overlap with any of existing standards (published or under development):

ISO/IEC 20889:2018 'Privacy enhancing data de-identification terminology and classification of techniques

This document provides a description of privacy-enhancing data de-identification techniques, to be used to describe and design de-identification measures in accordance with the privacy principles in ISO/IEC 29100, intended for privacy enhancement, without any guidance specific to AI or ML and the scope excludes unstructured data such as images, audio, free form text, video etc all of which are extensively used in machine learning. Moreover, the standard was published in 2018, when the maturity of AI technologies was at relative infancy compared to what it is today.

ISO/IEC WD 27091 'Cybersecurity and Privacy – Artificial intelligence – Privacy protection'

This document provides guidance for organizations to help organizations identify privacy risks throughout the AI system lifecycle, recommend mitigation actions and establish mechanisms to evaluate the consequences of and treat such risks. The standard seems to be not intended to provide implementation guidance in the context of data used in ML, which is what the proposed NWIP will provide.

Moreover, the development and recommendation of the methods will not only be focused on de-identification but also minimizing loss of data value which is crucial for success of AI models. Overall, we expect ISO/IEC 27091 and this proposed technical specification to complement each other.

Sustainable Development Goals (SDGs)

Goal 9: Industry, Innovation, and Infrastructure

Preparatory work

☐ A draft is attached ☒ An outline is attached ☐ An existing document serving as the initial basis is attached

The proposer is prepared to undertake the preparatory workrequired:

☒ Yes ☐ No

If a draft is attached to this proposal:

Please select from one of the following options:

☒ The draft document can be registered at Preparatory stage (WD – stage 20.00)
☐ The draft document can be registered at Committee stage (CD – stage 30.00)
☐ The draft document can be registered at enquiry stage (DIS – stage 40.00)

If the attached document is copyrighted or includes copyrighted content:

☐ The proposer confirms that copyright permission has been granted for ISO to use this content in compliance with the ISO/IEC Directives, Part 1 (see also the Declaration on copyright).

Is this proposal for an ISO management System Standard (MSS)?

☐ Yes ☒ No

Note: If yes, this proposal must have an accompanying justification study. Please see the Consolidated Supplement to the ISO/IEC Directives, Part 1, Annex SL or Annex JG

Indication of the preferred type to be developed

☐ International Standard ☒ Technical Specification

☐ Publicly Available Specification *

* While a formal NP ballot is not required to start developing a PAS (no eForm04), the NP form may provide useful information for the committee P-members to consider when deciding to initiate a Publicly Available Specification.

Proposed Standard Development Track (SDT – to be discussed by the proposer with the committee manager or ISO/CS)

☐ 18 months ☐ 24 months ☒ 36 months

Draft project plan (as discussed with committee leadership)

Proposed date for first meeting: 2025-02-14

Dates for key milestones: Circulation of 1st Working Draft (if any) to experts: 2025-03-31

Committee Draft consultation (if any): 2026-01-15

DIS submission*:

Publication*: 2028-01-15

* Target Dates for DIS submission and Publication should be set a few weeks ahead of the limit dates automatically determined when selecting the SDT.

NOTE: ISO/Meetings and ISO/Projects allow you to register and continuously update the meeting dates and project target dates during the development of the project.

Known patented items (see ISO/IEC Directives, Part 1 for important guidance)

☐ Yes ☒ No

If "Yes", provide full information as annex

Co-ordination of work: To the best of your knowledge, has this or a similar proposal been submitted to another standards development organization?

☐ Yes ☒ No

If “Yes”, please specify which one(s):

Listing of relevant documents (such as standards and regulations) at international, regional and national level

Identification and description of relevant affected stakeholder categories (Please see ISO CONNECT)

Benefits/Impacts/Examples

Industry and commerce - large industry	Industry can adopt AI with ease of compliance if de-identification methods are adopted
Industry and commerce - SMEs	Industry can adopt AI with ease of compliance if de-identification are adopted
Government	Government/State can adopt AI for larger benefit of society and public good with ease of compliance if de-identification methods are adopted
Consumers	Consumers privacy will be safeguarded
Labour	Workplace privacy will not be impacted due to use of AI
Academic and research bodies	
Standards application businesses	
Non-governmental organizations	
Other (please specify)	Citizens, and public at large will not have to become victim of technology paternalism

Liaisons:

A listing of relevant external international organizations or internal parties (other ISO and/or IEC committees) to be engaged as liaisons in the development of the deliverable.

Joint/parallel work:

Possible joint/parallel work with:

□ IEC (please specify committee ID)

□ CEN (please specify committee ID)

□ Other (please specify)

A listing of relevant countries which are not already P-members of the committee.

Note: The Committee Manager shall distribute this NP to the ISO members of the countries listed above to ask if they wish to participate in this work

Proposed Project Leader (name and e-mail address)

Srinivas Poosarla
srinivasp@infosys.com

Name of the Proposer
(include contact information)

Heather Benko
hbenko@ansi.org

This proposal will be developed by:

☒ An existing Working Group: ISO/IEC JTC 1/SC 42/WG 2 Data

□ A new Working Group:

(Note: establishment of a new Working Group requires approval by the parent committee)

□ The TC/SC directly

□ To be determined:

Supplementary information relating to the proposal

☒ This proposal relates to a new ISO document

□ This proposal relates to the adoption as an active project of an item currently registered as a Preliminary Work Item

□ This proposal relates to the re-establishment of a cancelled project as an active project

Other:

Maintenance agencies (MA) and registration authorities (RA)

□ This proposal requires the designation of a maintenance agency. If so, please identify the potential candidate:

□ This proposal requires the designation of a registration authority. If so, please identify the potential candidate:

NOTE: Selection and appointment of the MA or RA are subject to the procedure outlined in ISO/IEC Directives, Part 1, Annex G and Annex H.

☒ Annex(es) are included with this proposal (provide details)

Additional information/question(s)