ISO FORM 4

NEW WORK ITEM PROPOSAL (NP)

DATE OF CIRCULATION:
Click here to enter a date.

CLOSING DATE FOR VOTING:
Click here to enter a date.

REFERENCE NUMBER:
ISO/IEC PWI 47559

PROPOSER:

☐ ISO member body:
Click or tap here to enter text.

☒ Committee, liaison or other:
ISO/IEC JTC 1/SC 42

☐ **WITHIN EXISTING COMMITTEE**
Document Number: Click or tap here to enter text.
Committee Secretariat: Click or tap here to enter text.

☐ **PROPOSAL FOR A NEW PC**

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.

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

English title

Artificial Intelligence — Privacy Preservation of training data for ML

French title

(Please see ISO/IEC Directives, Part 1, Annex C, Clause C.4.2).
In case of amendment, revision or a new part of an existing document, please include the reference number and current title

SCOPE
(Please see ISO/IEC Directives, Part 1, Annex C, Clause C.4.3)

This document provides guidelines and requirements on privacy preservation including de-identification methods for training data used in machine learning.

PURPOSE AND JUSTIFICATION

(Please see ISO/IEC Directives, Part 1, Annex C and additional guidance on justification statements in the brochure Guidance on New Work)

For an ML model to be effective, the quality of data used to train the Machine learning algorithms plays crucial role. It is equally important to ensure the data use does not lead to privacy violation and unauthorised disclosure of confidential information.

Privacy violation and contravention to applicable privacy laws could occur for various reasons including the below:

a) Regurgitate personal data during model use thereby revealing personal information

b) Inability to exercise rights which is a regulatory requirement from GDPR and such laws – right to delete, correct, and even ascertaining whether ones' data has been used in training

c) Not having appropriate legal basis for processing – violation of GDPR article 6. Notice is insufficient, consent is not adequate and legitimate interest is also not admissible, as we saw from recent cases

d) Biometric information such as face and voice samples (through others' recordings found in the public domain) may be misused for identity theft

e) Exposure of personal information from targeted attacks such as membership inference, model inversion, and data extraction attacks

While by design we don't expect ML to store data - only the patterns and relationship thru 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

Although there are means to remove data such 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 standard being developed will recommend methods to ensure privacy preservation including de-identification techniques and procedure for embedding these in the AI life-cycle, in the context of ML for both structured and unstructured data, while keeping the value loss minimal for various types of use cases. Effective privacy preservation must balance privacy with data utility. Currently no specific standard focuses exclusively on the privacy of training data in machine learning. This gap highlights the need for a dedicated framework that addresses the unique challenges associated with machine learning data, helping organizations better protect individual privacy and enhance the security of their data-driven initiatives

The following standards (published or under development) with relation to this proposed NP exist. 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, and establish mechanisms to evaluate the consequences of and treat such risks. The standard is not focused only on data de-identification but overall privacy aspects of AI.

Moreover, the development and recommendation of right methods will not only be focused on privacy but also loss of data value which is crucial for success of AI models.

Following are some of standards which has reference to De-Identification:

ISO/IEC FDIS 5259-1, 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.

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

IS/ISO/IEC 23053:2022 in clause 8.3 includes de-identification: 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.

(Please use this field or attach an annex)

PROPOSED PROJECT LEADER (name and email address)

Srinivas Poosarla; srinivasp@infosys.com

PROPOSER (including contact information of the proposer's representative)

ISO/IEC JTC 1/SC 42

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

PROJECT MANAGEMENT

Preferred document
☒ International Standard
☐ Technical Specification
☐ Publicly Available Specification*

* While a formal NP ballot is not required (no Form04), 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

Proposed date for first meeting: Click here to enter a date.

Proposed TARGET dates for key milestones

- Circulation of 1st Working Draft (if any) to experts: Click here to enter a date.
- Committee Draft consultation (if any): Click here to enter a date.
- DIS submission*: Click here to enter a date.
- Publication*: Click here to enter a date.

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

It is proposed that this DOCUMENT will be developed by:

☒ An existing Working Group, add title ISO/IEC JTC 1/SC 42 WG2 Data

☐ A new Working Group Click or tap here to enter text.
(Note that the establishment of a new Working Group requires approval by the parent committee by a resolution)

☐ The TC/SC directly

☐ To be determined

☒ 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: Click or tap here to enter text.

Additional guidance on project management is available here.

PREPARATORY WORK

☐ A draft is attached

☐ An existing document serving as the initial basis is attached

☒ An outline is attached
Note: at minimum an outline of the proposed document is required

The proposer is prepared to undertake the preparatory work required:

☒ 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 clause 2.13 of ISO/IEC Directives, Part 1 (see also the Declaration on copyright).

RELATION OF THE PROPOSAL TO EXISTING INTERNATIONAL STANDARDS AND ON-GOING STANDARDIZATION WORK

To the best of your knowledge, has this or a similar proposal been submitted to another standards development organization or to another ISO committee?

☐ Yes ☒ No

If Yes, please specify which one(s)

☐ The proposer has checked whether the proposed scope of this new project overlaps with the scope of any existing ISO project

☐ If an overlap or the potential for overlap is identified, the proposer and the leaders of the existing project have discussed on:
   i. modification/restriction of the scope of the proposal to avoid overlapping,
   ii. potential modification/restriction of the scope of the existing project to avoid overlapping.

☐ If agreement with parties responsible for existing project(s) has not been reached, please explain why the proposal should be approved
   Click or tap here to enter text.

☐ Has a proposal on this subject already been submitted within an existing committee and rejected? If so, what were the reasons for rejection?
   Click or tap here to enter text.

This project may require possible joint/parallel work with
☐ IEC (please specify the committee) Click or tap here to enter text.
☐ CEN (please specify the committee) Click or tap here to enter text.
☐ Other (please specify) Click or tap here to enter text.

Please select any UN Sustainable Development Goals (SDGs) that this proposed project would support (information about SDGs, is available at www.iso.org/SDGs)

☐ GOAL 1: No Poverty
☐ GOAL 2: Zero Hunger
☐ GOAL 3: Good Health and Well-being
☐ GOAL 4: Quality Education
☐ GOAL 5: Gender Equality
☐ GOAL 6: Clean Water and Sanitation
☐ GOAL 7: Affordable and Clean Energy
☐ GOAL 8: Decent Work and Economic Growth
☐ GOAL 9: Industry, Innovation and Infrastructure
☐ GOAL 10: Reduced Inequality
☐ GOAL 11: Sustainable Cities and Communities
☐ GOAL 12: Responsible Consumption and Production
☐ GOAL 13: Climate Action
☐ GOAL 14: Life Below Water
☐ GOAL 15: Life on Land
☐ GOAL 16: Peace, Justice and strong institutions

N/A GOAL 17: Partnerships for the goals

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 less regulatory hurdles if data is deidentified
Industry and commerce – SMEs	Industry can adopt AI with less regulatory hurdles if privacy is preserved
Government	Government/State can adopt AI for larger benefit of society and public good with less regulatory hurdles if privacy is preserved
Consumers	Consumers privacy will be safeguarded
Labour	Workplace privacy will not be impacted due to use of AI
Academic and research bodies	Click or tap here to enter text.
Standards application businesses	Click or tap here to enter text.
Non-governmental organizations	Click or tap here to enter text.
Other (please specify)	Citizens, and public at large will not have to become victim of technology paternalism

Listing of countries where the subject of the proposal is important for their national commercial interests (Please see ISO/IEC Directives, Part 1, Annex C, Clause C.4.8)

Click or tap here to enter text.

Listing of external international organizations or internal parties (other ISO and/or IEC committees) to be engaged in this work (Please see ISO/IEC Directives, part 1, Annex C, Clause C.4.9)

Listing of relevant documents (such as standards and regulations) at international, regional and national level (Please see ISO/IEC Directives, Part 1, Annex C, Clause C.4.6)
Click or tap here to enter text.

ADDITIONAL INFORMATION

Maintenance Agencies (MAs) and Registration Authorities (RAs)

☐ This proposal requires the designation of a maintenance agency.
If so, please identify the potential candidate:
Click or tap here to enter text.

☐ This proposal requires the designation of a registration authority.
If so, please identify the potential candidate
Click or tap here to enter text.

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.

Known patented Items (Please see ISO/IEC Directives, Part 1, Clause 2.14)

☐ Yes ☒ No

If Yes, provide full information as an annex

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