DRAFT International Standard

ISO/IEC DIS 26641

Information technology — Security profiles for Natural Language Interaction Protocol (NLIP)

ICS: 35.030; 35.240.01

ISO/IEC JTC **1**

Secretariat: **ANSI**

Voting begins on:
2026-05-04

Voting terminates on:
2026-07-27

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENTS AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

This document is circulated as received from the committee secretariat.

FAST TRACK PROCEDURE

Reference number
ISO/IEC DIS 26641:2026(en)

© ISO/IEC 2026



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2026

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Conformance ... **2**
5 Notational conventions ... **2**
6 Overview ... **2**
7 Security profiles ... **3**
7.1 Agent Security Profile 1 (Mandatory) ... 3
7.1.1 Transport Layer Security ... 3
7.1.2 Identity & Authentication: NLIP Agent Authentication ... 3
7.1.3 Agent Lifecycle: Ethical-by-Design ... 3
7.2 Agent Security Profile 2 ... 3
7.2.1 Identity & Authentication: NLIP Client Authentication ... 4
7.2.2 Runtime & Behaviour: NLIP Client Authorization ... 4
7.2.3 Runtime & Behaviour: Privilege Misuse Prevention ... 4
7.2.4 Observability & Audit: Agent Behaviours ... 4
7.2.5 Observability & Audit: Client Throttling and DoS ... 4
7.2.6 Observability & Audit: Data Governance and Privacy ... 4
7.2.7 Agent Lifecycle: Secure-by-Design Disclosure ... 4
7.3 Agent Security Profile 3 ... 4
7.3.1 Prevention of Prompt Injection (including Indirect Prompt Injection) ... 4
7.3.2 Runtime Agent Behaviour Controls ... 4
7.3.3 Information Leakage ... 5
Annex A (informative) **NLIP Secure-by-Design Disclosure Template** ... **6**
Annex B (informative) **NLIP Ethical-by-design disclosure template/text** ... **7**
Bibliography ... **9**

© ISO/IEC 2026 – All rights reserved

Foreword

© ISO/IEC 2026 – All rights reserved

Introduction

The technology of Artificial Intelligence has the potential to be truly transformative to society. Despite some limitations, the technology is capable of many functions, including but not limited to answering questions, translating, describing and summarizing multi-modal content, generating new content, and summarizing large volumes of information. This enables the creation of intelligent agents that can use AI to analyze data and provide new services.

A much bigger boost to the social benefits of AI technology can be obtained by interactions among different intelligent agents, which may be under the control of different organizations and users. The interaction among intelligent agents can unlock new economic and social value, just like the interactions among various Internet-based services was enabled with the advent of the web browser.

We need a standard common protocol that is used by humans to interact with an intelligent agent, and for intelligent agents to interact with each other.

Ecma Technical Committee 56 is defining this Standard. Four Standards define the standard protocol called Natural Language Interaction Protocol (NLIP), its binding over HTTPS/REST, its binding over websockets and its binding over AMQP.

Security is an important aspect of any system that implements a protocol. In order to properly secure the implementations of NLIP while maintaining interoperability, a set of recommendations for security management is needed. This Standard provides that requisite set of recommendations.

This recommendation is targeted at security engineers, site reliability engineers (SREs), architects, Chief Information Security Officers (CISO), and compliance teams deploying NLIP agents at enterprise-scale. This Standard defines a set of standard security profiles which should be used in NLIP implementations.

Details for the security guidelines implementation can be found in an exemplar living document at URL https://github.com/nlip-project/security_guidelines.

© ISO/IEC 2026 – All rights reserved

Information technology — Security profiles for Natural Language Interaction Protocol (NLIP)

1 Scope

This Standard establishes the Agent Security Profiles for NLIP as a standalone Ecma Standard. Each Security Profile defines a set of mandatory security requirements for an agent implementing NLIP. Conformance to this Standard is mandatory for any implementation claiming NLIP conformance.

Details on how the security guidelines described in this Standard should be implemented are out of scope of this Standard.

This Standard does not cover aspects related to internals of foundation-models, safeguards for physical data centers, or issues related to national export controls.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ECMA-404, *The JSON Data Interchange Syntax, 2nd edition (December 2017)*

ECMA-431, *Binding of the Natural Language Interaction Protocol (NLIP) over HTTP/HTTPS*

ECMA-432, *Binding of the Natural Language Interaction Protocol (NLIP) over WebSocket*

ECMA-433, *Binding of the Natural Language Interaction Protocol (NLIP) over AMQP*

IETF RFC 2119, *Key words for use in RFCs to Indicate Requirement Levels*[1)]

IETF RFC 8446, *The Transport Layer Security (TLS) Protocol Version 1.3*[2)]

3 Terms and definitions

3.1
security countermeasures and objectives (SCO)
a SCO is a set of security goals relevant to agentic systems and measures, along with approaches and strategies to achieve the goals

3.2
agent security profile
an agent security profile is a pre-defined checklist of *security countermeasures and objectives (SCO)* (3.1) in agent development and deployment

3.3
red teaming
red teaming is the practice of simulating real-world attackers to test an agentic system's defences, focusing on its ability to detect, respond, and recover rather than just finding vulnerabilities

1) https://datatracker.ietf.org/doc/rfc2119.

2) https://datatracker.ietf.org/docrfc8446.

© ISO/IEC 2026 – All rights reserved