2023.09.05
Version 4.0

OVON Identification & Authentication Requirements for Conversational Assistants

The Open Voice Network
Architecture Work Group of the Technical Committee
5 September 2023

TABLE OF CONTENTS

1.0 Purpose

The Open Voice Network proclaims as its motto, “voice technology worthy of user trust”. The purpose of this document is to describe a set of attributes that conversational assistants may exhibit to allow for their identification and authentication in support of that trustworthiness. These requirements are presented independently from any technology or human processes that may be developed as implementations. Indeed, it is possible that some of these requirements will prove to be unfeasible at this time but realizable at some future point.

Note that this document only pertains to identification and authentication of conversational assistants, and not to analogous mechanisms for identification and authentication of human users.

1.1 Definitions

Authentication: the process of verifying an asserted identification (qv); in this context, the ability to demonstrate that a conversational assistant is verifiably who it claims to be.

Conversation: a joint activity in which two or more participants (human or automated) use linguistic forms and non-verbal signals (i.e., gestures) to communicate to achieve an outcome that meets a shared goal.

Conversational assistant: a digital participant in a conversation (qv); a piece of technology (hardware/software) that accepts natural language requests expressed using one or more of text, voice, and graphics from the user, processes the requests, and presents results to the user using one or more of text, voice, and/or graphics..

Identification: a set of attributes that serve to uniquely and unambiguously distinguish one entity (here, a conversational assistant) from all others.

TrustMark: an initiative of the Open Voice Network to promote the principles and core values of trustworthy conversational AI.

Trustworthiness: a subjective human judgment based on available attributes that a conversational assistant will correctly perform the task it was designed for and will not engender undesirable side effects.

1.2 Scope

This document describes requirements for *identification* and *authentication* of conversational assistants, i.e. *automated* participants in a voice-based system. Initial requirements for judging *trustworthiness* are also presented.

This document only pertains to identification and authentication of conversational assistants, and not to analogous mechanisms for identification and authentication of human users.

This document contains both *normative* elements – characteristics which *must* be satisfied by conversational assistants – and *informative* elements – explanations, illustrations, examples, etc. which serve to guide the human reader of this document, as well as possible enhancements or extensions to the normative characteristics. This document uses *shall* and *may* for these elements respectively.

1.3 Identification Requirements

1. Normative: Each conversational assistant *shall* be uniquely identified by one or more identification mechanisms.
    - a. Informative: this identification mechanism *may* be based on existing paradigms such as, but not limited to, metadata, registry, directory.
2. Normative: Each conversational assistant *shall* possess the following attributes used primarily for identification:
    - a. Normative: Name of the conversational assistant
    - b. Normative: Name of responsible person or organization — Author or Owner and their contact information
    - c. Normative: Cyberspace location
        - i. Informative: This location *may* be specified using mechanisms such as Uniform Resource Location (URL), Uniform Resource Name (URN), other address scheme to be defined
    - d. Normative: Government assigned id
        - i. Informative: If the author or owner is a business entity, this id *may* be based upon existing schemas such as employer identification number (EIN) in USA, Business number (BN) in Canada, company number in UK and Australia, and VAT identification number (in the European Union).
        - ii. Informative: if the business entity possesses multiple ids (for example multinational corporations), a decision process *may* be invoked to decide which id to use.
        - iii. Informative: if the author or owner is an individual or non-business entity, this id *may* be based upon other schemas.

        iv. Informative: if the author or owner does not possess, or chooses not to use, a government-issued id, other mechanisms *may* exist to assign an id.

3. Informative: The identification of a conversational assistant *may* also include (non-exhaustive) the following attributes:
    a. List of purposes / intents / actions that can be fulfilled
    b. Metadata about content the assistant can provide
    c. List of communication protocols that are supported
    d. List of encryption protocols that are supported and their keys (or other mechanisms)
    e. Dialog history / context expected at activation
    f. Negotiation patterns accepted at activation
    g. Mechanisms for access control

1.4 Repository Requirements

4. Informative: A set of conversational assistant attributes [see above] *may* be collected into a repository.
    a. Informative: This repository *may* reuse principles and/or technology from existing paradigms such as DNS or ICANN.
    b. Informative: The characteristics of the set of attributes have not yet been specified.

1.5 Authentication Requirements

5. Normative: Mechanisms *shall* exist to verify that the identification attributes of a conversational assistant are valid and correct.
    a. Informative: These mechanisms *may* reuse principles and/or technology from existing initiatives in analogous domains (e.g. web browsing).
    b. Informative: These mechanisms *may* cover tamper detection and spoofing detection.
    c. Informative: Next steps to specify these requirements include:
        i. Adapt and adopt existing authentication protocols and standards.
        ii. Perform gap analysis to determine what is covered by current standards to identify new standards.

1.6 Trustworthiness Requirements

6. Informative: Conversational assistants *may* exhibit trustworthiness attributes including (non-exhaustive):
    a. TrustMark certification
    b. Certificates from a certificate authority (e.g. security, NIST, ISO, etc.)

7. Informative: Mechanisms *may* exist to evaluate functional trustworthiness (i.e. the assistant will correctly perform the task for which it was designed and not engender undesirable side effects) including (non-exhaustive):
    a. Traceability of ownership and modifications; to be specified to address the following questions:
        i. Is this regarding chain of custody?
        ii. Traceability of ownership over time?
        iii. Traceability of end-to-end fulfillment of request (entire delegation journey)?
        iv. Management of the history of the interaction?
    b. Transparency of algorithms and actions
    c. Escalation process in case of dysfunction
    d. Evaluation of audit records of the assistant and its producer
    e. Evaluation of an assistant's availability (uptime / capacity / etc.)
    f. Evaluation of the accuracy of an assistant's output; this *may* be static or dynamic, single point and/or trends
    g. Evaluation by a recognized certification organization; how this may work is TBD
    h. Traceability of information – sources and destinations
8. Informative: Functional trustworthiness *may* be evaluated statically and/or continuously in real time. A cost-benefit tradeoff *may* be used to determine the need for systematically evaluating trustworthiness.
9. Informative: Mechanisms *may* exist to support subjective trustworthiness of voice assistants (i.e. reassurance of the human user) including (non-exhaustive):
    a. Evaluation of user reviews covering aspects such as usability, successful outcomes, etc.
    b. Tracking of frequency of assistant use (e.g. hits, popularity, etc.)
    c. Tracking of frequency of connection to a particular assistant from another assistant