AI for Good
Global Summit

Introduction Slides

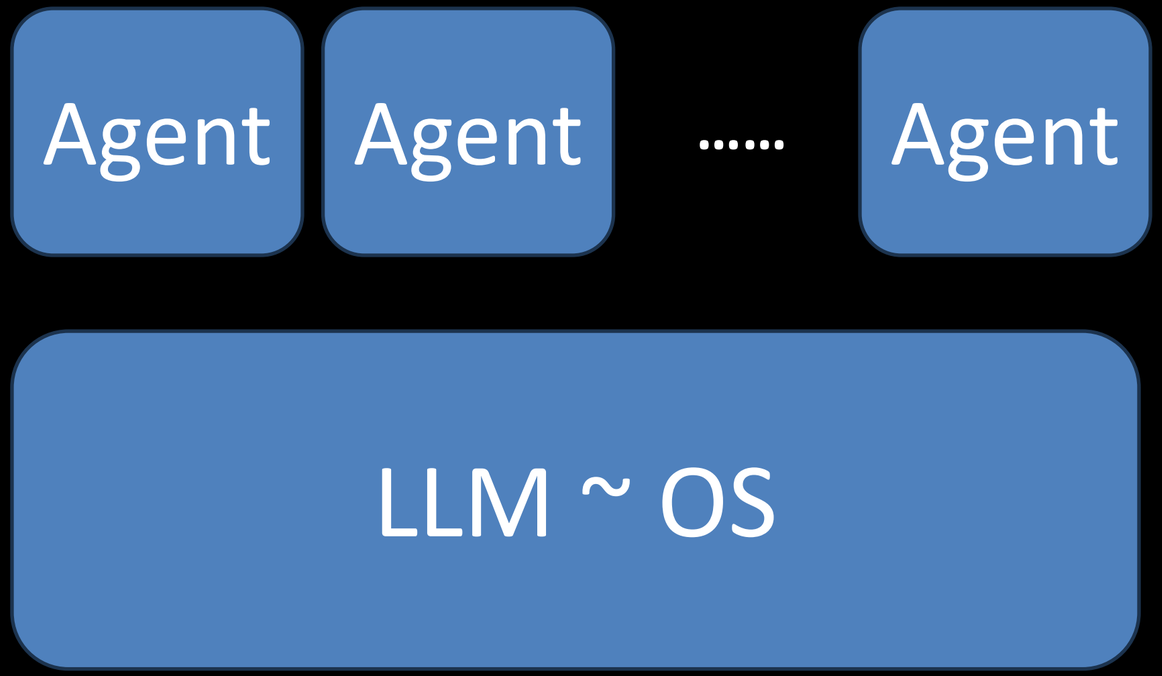
Mr. Kai WEI
Director, CAICT AI Institute

11 July 2025
Geneva, Switzerland

AI for Good
Global Summit

Characteristics of AI agents

CAICT



Perception

recognition, comprehension, interaction and reasoning

Planning

planning, scheduling and optimization

Action

execution in the digital space or in the physical environment

Memory

short-term and long-term memory

IEC ISO ITU

AI for Good Global Summit

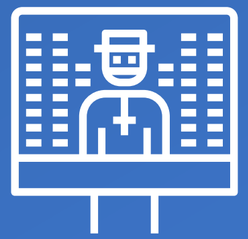
Challenges

CAICT

Reliability

- The root cause of agent reliability issues lies in model hallucination.
- To relief of the hallucination is very hard.

Safety & Security

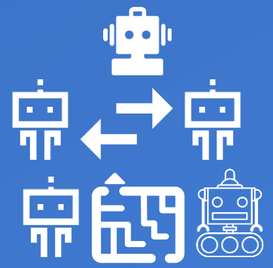


- Confronted with prompt injection and other emerging attacks.
- Traditional security risks such as sensitive personal information and privacy breaches are intensifying.

Interoperability

- Agent communication requires unified interface standards.
- The development of MCP and A2A ecosystems still requires standards for interoperaability.

Operations & Maintenance



- The interactions between multiple agents increased the system complexity.
- It will significantly decrease the observability and increase the difficulty of O&M.

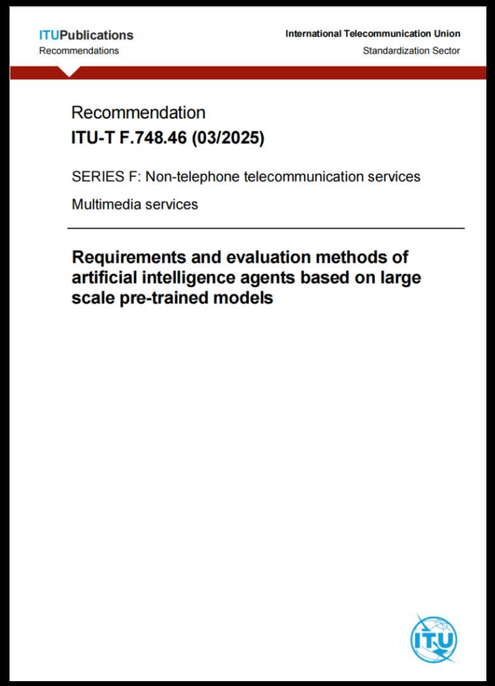
IEC ISO ITU

AI for Good Global Summit

Standards for AI agent security

CAICT

ITU-T F.748.46 @ SG21



ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

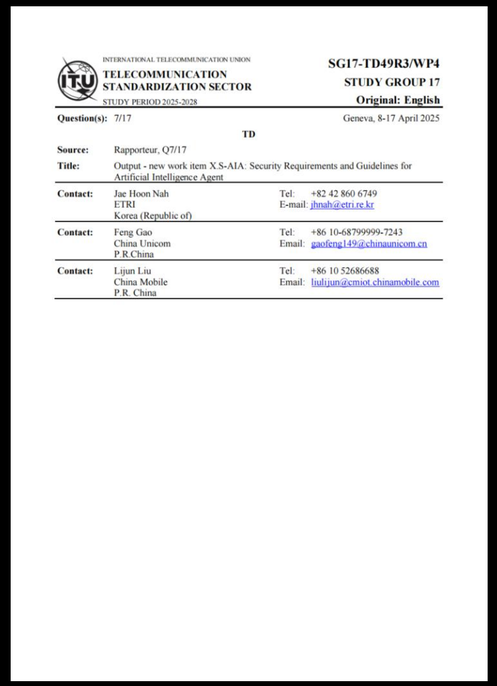
Recommendation
ITU-T F.748.46 (03/2025)

SERIES F: Non-telephone telecommunication services
Multimedia services

Requirements and evaluation methods of artificial intelligence agents based on large scale pre-trained models

- Define the **capabilities and evaluation methods** for general AI agents.
- Propose an evaluation baseline for **agent-based product and applications**.

Security Req. of AI Agent @ ITU-T SG17



INTERNATIONAL TELECOMMUNICATION UNION
TELECOMMUNICATION STANDARDIZATION SECTOR
STUDY PERIOD 2025-2028

SG17-TD49R3/WP4
STUDY GROUP 17
Original: English

Question(s): 7/17 — Geneva, 8-17 April 2025

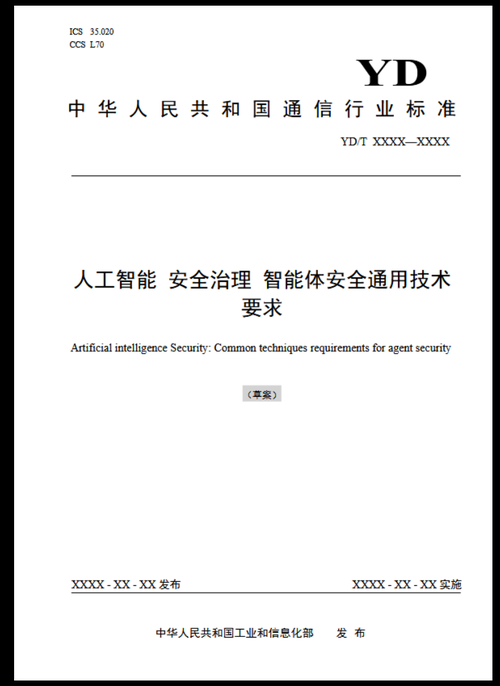
TD

Source: Rapporteur, Q7/17
Title: Output - new work item X.S-AIA: Security Requirements and Guidelines for Artificial Intelligence Agent

Contact: Jae Hoon Nah, ETRI, Korea (Republic of) — Tel: +82 42 860 6749, E-mail: jhnah@etri.re.kr
Contact: Feng Gao, China Unicom, P.R.China — Tel: +86 10-68799999-7243, Email: gaofeng149@chinaunicom.cn
Contact: Lijun Liu, China Mobile, P.R. China — Tel: +86 10 52686688, Email: liulijun@cmiot.chinamobile.com

- Analyze security risks systematically across an agent's **perception, planning, memory and action** stages.
- Propose **lifecycle-wide risk protection** requirements.

General Req. of AI Agent security@ MIIT TC1



ICS 35.020
CCS L70

YD

中 华 人 民 共 和 国 通 信 行 业 标 准

YD/T XXXX—XXXX

人工智能 安全治理 智能体安全通用技术要求

Artificial intelligence Security: Common techniques requirements for agent security

（草案）

XXXX - XX - XX 发布 — XXXX - XX - XX 实施

中华人民共和国工业和信息化部 发布

- Define the general requirements for **security capability of AI agent**.
- Establish a framework for **system interaction security** specifications.

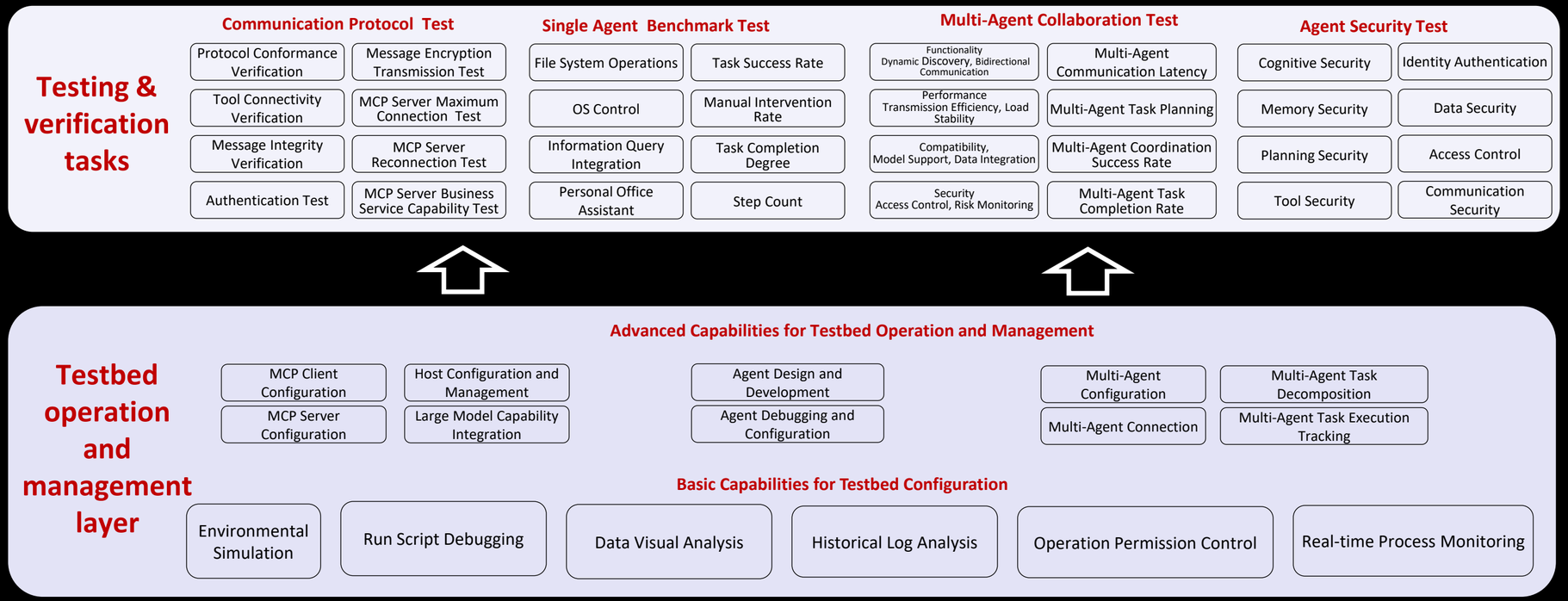
IEC ISO ITU

AI for Good Global Summit

CAICT's Trusted AI Agent Testbed

CAICT

CAICT is building a comprehensive AI Agent testbed for testing & verification of reliability, interoperability and security.



Source: Framework of Trusted AI Agent Testbed, CAICT, June 2025

IEC ISO ITU