Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 15 March 2026

B. Sarikaya
Unaffiliated
R. Schott
Deutsche Telekom

11 September 2025

AI Agents for 6G Requirements and Implementation Approaches
draft-sarischo-6gip-aiagent-requirements-00

Abstract

This document provides requirements for 3GPP Artificial Intelligence/ Machine Learning (AI/ML) Agents for the 6th generation mobile network, or 6G. Requirements depend on the types and application areas of agents. We describe each type and state their requirements. AI Agent implementation efforts, how APIs can be discovered, how inter-domain and intra domain AI Agents can be discovered using DNS lookup are explained.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 15 March 2026.

Copyright Notice

Table of Contents

1. Introduction

Artificial Intelligence (AI) has historically been defined as the science and engineering to build intelligent machines capable of carrying out tasks as humans do. Inspired from the way human brain works, machine learning (ML) is defined as the field of study that gives computers the ability to learn without being explicitly programmed. Since it is believed that the main computational elements in a human brain are 86 billion neurons, the more popular ML approaches are using "neural network" as the model. Neural networks (NN) take their inspiration from the notion that a neuron's computation involves a weighted sum of the input values. A computational neural network contains the neurons in the input layer which receive some values and propagate them to the neurons in the middle layer of the network, which is also called a "hidden layer". The weighted sums from one or more hidden layers are ultimately propagated to the output layer, which presents the final outputs of the network.

Recurrent neural network (RNN) models are a type of deep neural networks which use sequential data feeding. The input of RNN consists of the current input and the previous samples. RNN models recently have been replaced with parallel processing [LLMPaper] and the transformer architecture and they are being used in the natural language processing task on mobile devices, e.g., language modeling, machine translation, question answering, word embedding, and document classification. The resulting system is commonly called Large Language Model (LLM).

AI Agents play a crucial role in modern telecommunications by enabling intelligent automation, decision-making, and adaptive network management. These agents are software-driven entities that leverage artificial intelligence, including machine learning and natural language processing, to interact with users, applications, and network components. In a 6G environment, AI agents enhance network efficiency by dynamically optimizing resources, predicting network conditions, and facilitating providing seamless communication between services. By integrating Large Language Models (LLM), AI agents can understand complex requests, translate them into actionable insights, and orchestrate 3GPP services (e.g. communication service, sensing service, AI-related) network capabilities and functions autonomously, ultimately improving user experience when consuming the 3GPP services, operational efficiency, and service innovation.

This document aims to present the types of AI agents 6G network needs and the requirements needed to support each case. We discuss implementation issues next.

In a related work, [aiagent6g] attempts to analyze the agent protocol requirements and relevant enabling technologies based on 6G mobile communication system specific characteristics.

On the other hand [aiagentusecase] introduces use cases and requirements on AI Agents in 6G networks. It attempts to elaborate on the requirements for high performance communication, security and energy efficiency.

2. 6G Network AI Agents

The following use cases and requirements are from [TR22.870].

2.1. AI Agents collaboration with third-party AI using LLM

A third-party application (e.g. a smart city traffic management system) AI Agent sends a text-based request or query to the 6G network. The request is processed by an AI agent in the 6G network that leverages LLMs and the network's advanced capabilities (e.g. sensing, real-time data processing, telemetry, analytics, and others) to provide a response or perform an action. This interaction mimics how users interact with chatbots like ChatGPT, but it is tailored for network-specific tasks and applications.

* Requirements

The network shall be able to support secure means to expose its services to the authorised third-party AI agent based on its intent.

The network shall be able to take into account information related to user mobility context, subscription information when invoking 3GPP services based on user intent(s).

2.2. AI Agents for Artificial General Intelligence

Autonomous agents (AI agents) have long been recognized as a promising approach to achieving artificial general intelligence (AGI), which is expected to accomplish tasks through self-directed planning and actions. In recent years, these agents, leveraging the capabilities of LLMs, are expected to effectively perform diverse tasks in social science, natural science, and engineering, among others. AI agents can take on various forms, such as embodied intelligent robots, virtual assistants, and autonomous systems (e.g. drones).

* Requirements

The network shall support trusted network access for 3rd party AI agent and support a mechanism to expose 3rd party AI agent' s attributes (e.g. related users, sensing capabilities, AI capabilities, service features) to other 3rd party AI agents.

The network shall be able to support security identification for 3rd party AI agents provided by authorized 3rd party associated with a user (e.g. AI agents belonging to a customer).

The network shall support mechanisms for 3rd party AI agents to provide/register their attributes (e.g. sensing capabilities, AI capabilities, service features, associated authorized users) to 6G network, and discover other authorized 3rd party AI agents to achieve collaborative task.

The network shall provide means to support efficient and secure communication between 3rd party AI agents over a wide area in a group considering the diverse lifetime of tasks.

2.3. AI Agents on Device

AI agent on device will be popular in 6G era, due to the fast development on device-based computing power and model light-weighting.

* Requirements

The system shall provide a suitable means for an AI agent application on UE to invoke some 3GPP services (e.g. IMS service).

The system shall provide an efficient way to expose information (e.g. change of QoS) to the application on the UE.

The network shall be able to support the message exchange between the AI agent application on different UEs considering the diverse capabilities supported by different AI applications (e.g. AI agents applications).

2.4. Collaborative AI Agents

AI Agents can perform tasks for or represent e.g. devices, persons, drones, or cars. These AI Agents may be either implemented in a UE or in the network. By offloading tasks to the network, devices can save on complexity and energy consumption. Furthermore, an AI Agent in the network can still represent a device, person, drone or car, when that device, person, drone or car is not reachable, e.g. because of radio conditions or battery outage. Offload can happen towards a local/edge network but can also be to a nearby other device with more processing capabilities.

* Requirements

The system shall support hosting of large amounts of AI agent applications managed and controlled by the 6G core network and/or multiple AI Agent applications on a UE.

The system shall support secure interoperability between AI Agents and between AI Agents and applications to achieve a collaborative task.

2.5. Home Robots

Home robots will engage in household chores based on preconfigured models, such as sweeping the floor, vacuuming, folding clothes, washing dishes, and organizing rooms. They will also take care of family members by monitoring health data, reminding medication, and dialling emergency calls. Additionally, they will socialize and entertain with humans while interacting with other smart devices to create a more intelligent ecosystem. All of this aims to bring us a more convenient, comfortable, and safe family life. All or part of the AI inference services are provided by home robots.

* Requirements

The network shall be able to provide AI service (e.g. AI model inference) to a UE.

The system shall be able to support negotiation of the service performance (e.g. latency, inference accuracy), between UE and 6G network, when providing AI service (e.g. AI model inference).

The network shall be able to support mechanism to guarantee the service performance (e.g. latency, inference accuracy) when providing AI service (e.g. AI model inference).

2.6. Built-in Intelligent Communication Assistant

Empowered by the rapid development of the AI technology, the service providers are able to provide personalized and enriched services to their users when making daily routines within their homes, at their workplaces, in stores, at restaurants, as well as traveling for work or leisure. These kinds of personalized services are widely enjoyed by the customers. For example, a lot of countries are facing a major challenge in providing care support for senior citizens due to their rapidly ageing population and declining old-age support. The capability to introduce AI techs to provide more personalized and real-time communication services would be a great help.

* Requirements

The network (e.g. in conjunction to IMS) shall be able to provide intelligent communication assistant service to users.

The network shall support charging information collection for the intelligent communication assistant service.

The network (e.g. in conjunction to IP Multimedia Subsystem, IMS) shall be able to support the interaction and collaboration between different user' s intelligent communication assistants, e.g. during an IMS calling service, both calling and callee parties are using intelligent communication services.

The network (e.g. in conjunction to IMS) shall be able to support the intelligent communication assistant to use operator native capabilities (e.g. AR rendering, XR rendering in service hosting environment, SMS or voice).

3. Implementation Issues

In the section above Section 2 we described various kinds of AI agents 6G network needs. In this section we will look at AI Agent implementation efforts so far.

General purpose AI Agents like a travel AI Agent, loan handling, shopping for clothing AI agent are discussed in [aiprotocol]. These types of AI Agents can be built using Large Language Models, like GPT (Generative Pre-Trained Transformer)-4o, Gemini, Anthropic, etc. There are also open source ones like Llama.

Like LLM tools, AI Agents work on prompt-completion mode, they get prompts and they reply with completions. Designing AI Agents for specific tasks is developing to be an engineering practice.

We will shortly describe the steps involved:

Protocols involved include IP, TCP, UDP, QUIC for host communication, HTTP, SIP and RTP at the application layer. Above all that are the protocols for AI Agents. So far Model Context Protocol (MCP) [MCP] and Agent to Agent protocol [Agent2agent] which operate at the HTTP level and are expected to be standardized.

AI Agent to API communication. Agents provide services to user by invoking APIs either in the same domain or another domain. AI Agent design includes to teach the agent what they are and give them enough information to know how to use them appropriately.

User to AI Agent communication could be via various means like a phone call, a chat like by a chat app or via video. Several types of media are involved simultaneously using protocols, tools established currently and widely used.

AI Agent to user communication is similarly involving well established techniques like email, voice or chat when both the user and agent are on the same administrative domain. If not, administrative configuration is required to allow the systems to communicate with each other using protocols like SIP.

AI Agent implementation points to many areas in user to AI Agent, AI Agent to API and AI Agent to AI Agent where new protocol work is needed, these are discussed next.

3.1. Future Work

In AI Agent to API case, the discovery of APIs can happen using a well known URI and a link relation. [RFC9727] defines the api-catalog well-known URI to which HTTPS GET request to the Publishers site returns an API catalog document.

In the case of user to AI Agent, AI Agent discovery can be made using DNS [ajand]. For that purpose a DNS TXT record is specified. Providers advertise their agent service by publishing a single DNS TXT record at _agent.<domain> such as

text _agent.example.com. 300 IN TXT "v=aid1;u=https://api.example.com/mcp;p=mcp;a=pat;s=Example AI Tools"

which advertises a remote AI Agent called mcp. Local agents can be advertised using Docker.

Agents can be discovered using DNS lookups querying TXT records giving the domain name. If the query succeeds and the protocol is supported the client can start using the AI Agent with the protocol.

Authentication and authorization [RFC6749], [RFC6819] are to be discussed later.

4. Security Considerations

Security considerations of 6G AI Agents is TBD.

5. IANA Considerations

There are no IANA considerations for this document.

6. Acknowledgements

7. References

7.1. Normative References

[RFC6749] Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, DOI 10.17487/RFC6749, October 2012, <https://www.rfc-editor.org/rfc/rfc6749>.

[RFC6819] Lodderstedt, T., Ed., McGloin, M., and P. Hunt, "OAuth 2.0 Threat Model and Security Considerations", RFC 6819, DOI 10.17487/RFC6819, January 2013, <https://www.rfc-editor.org/rfc/rfc6819>.

[RFC9727] Smith, K., "api-catalog: A Well-Known URI and Link Relation to Help Discovery of APIs", RFC 9727, DOI 10.17487/RFC9727, June 2025, <https://www.rfc-editor.org/rfc/rfc9727>.

7.2. Informative References

[Agent2agent] Google, "Agent2Agent (A2A) Protocol", June 2025, <https://google.github.io/A2A/>.

[aiagent6g] Yang, C., Huang,, H., Akhavain, A., Liu, F., An,, X., Xing,, W., Li, J., Wang, A., and Y. Wencong,, "Requirements and Enabling Technologies of Agent Protocols for 6G Networks", Work in Progress, Internet-Draft, draft-hw-ai-agent-6g-00, 19 July 2025, <https://datatracker.ietf.org/doc/html/draft-hw-ai-agent-6g-00>.

[aiagentusecase] Yu, M., Wang, A., Li, J., and Z. Li, "AI Agent Use Cases and Requirements in 6G Network", Work in Progress, Internet-Draft, draft-yu-ai-agent-use-cases-in-6g-01, 7 July 2025, <https://datatracker.ietf.org/doc/html/draft-yu-ai-agent-use-cases-in-6g-01>.

[aiprotocol] Rosenberg, J. and C. F. Jennings, "Framework, Use Cases and Requirements for AI Agent Protocols", Work in Progress, Internet-Draft, draft-rosenberg-ai-protocols-00, 5 May 2025, <https://datatracker.ietf.org/doc/html/draft-rosenberg-ai-protocols-00>.

[ajand] Agent Community, "Agent Identity and Discovery (AID)", August 2025, <https://docs.agentcommunity.org/aid/>.

[LLMPaper] al., A. V. et., "Attention is all you need", August 2017, <https://arxiv.org/html/1706.03762v7>.

[MCP] Anthropic, "Model Context Protocol", June 2025, <https://modelcontextprotocol.io/>.

[TR22.870] 3rd Generation Partnership Project, "Study on 6G Use Cases and Service Requirements", June 2025, <https://www.3gpp.org/ftp/Specs/archive/22_series/22.870/22870-031.zip>.

[TS22.261] 3rd Generation Partnership Project, "Service Requirements for the 5G System", December 2024.

Authors' Addresses

Behcet Sarikaya
Unaffiliated
Email: sarikaya@ieee.org

Roland Schott
Deutsche Telekom
Deutsche-Telekom-Allee 9
64295 Darmstadt
Germany
Email: Roland.Schott@telekom.de