AI Co-pilots for Intent-driven Wireless Networks

Sridhar Rajagopal and Eran Pisek
Ennoia Technologies, Inc.
{sridhar, eran}@ennoia-tech.com

1. Problem statement

Wireless networks are inherently complex to build, test, deploy, and debug. As mobile technology evolves from 4G through to now 6G, this complexity continues to grow—driven by the need for backward compatibility, co-existence with legacy systems, and the integration of new features and services. RAN also represents a high-cost, high-skill domain. Its infrastructure spans a wide array of components, including radios, routers, switches, servers, and increasingly, GPUs—often sourced from multiple vendors.

In parallel, the operational expertise required to test and deploy RAN and mobile systems, both in labs and in the field, is significant. This contributes to high OPEX. Despite advances in automation, many integration and testing issues still require lengthy debug cycles—often stretching over several weeks—due to the system's inherent complexity.

AI offers a compelling solution to these challenges. By training AI models to understand the complexities of RAN infrastructure and UEs, we can enable intelligent co-pilots that, for example, assist with RAN integration, issue identification, and automated resolution. This approach can significantly shorten troubleshooting timelines, reduce the need for site visits, and ultimately improve engineering productivity and time-to-market.

2. Proposed solution

We have developed a software platform, called Ennoia Connect Platform (ECP) that hosts a RAN hardware (HW) specific model repository and has been trained on RAN HW APIs available on the RAN infrastructure. The platform is exposed via a co-pilot interface, where commands are given in natural language to control/configuration of the various RAN HW components and to analyze and interpret the output coming from the HW infrastructure. The infrastructure could be radios, servers, switches, test equipment, emulators – any RAN hardware that exposes an API to control the infrastructure. Note that most, if not all, HW today expose APIs for test automation and use. In scenarios where the hardware is not directly accessible, we translate the user intent to generate configuration files for the management system controlling the hardware. We train AI models on the specific RAN hardware and then use intent for configuration and analysis of the specific hardware.

We provide two types of AI co-pilot functions on our ECP platform:

a. CAT (Chat and Test) – which is used to control and configure the infrastructure.
b. AID (AI-based Debug) – which is used for interpretation and analysis of the output.

Figure 1 describes the ECP block diagram. The ECP uses Small Language Models (SLMs) that can run on a local machine or edge infrastructure for data privacy while supporting public & private cloud models as well if required for fast AI development.

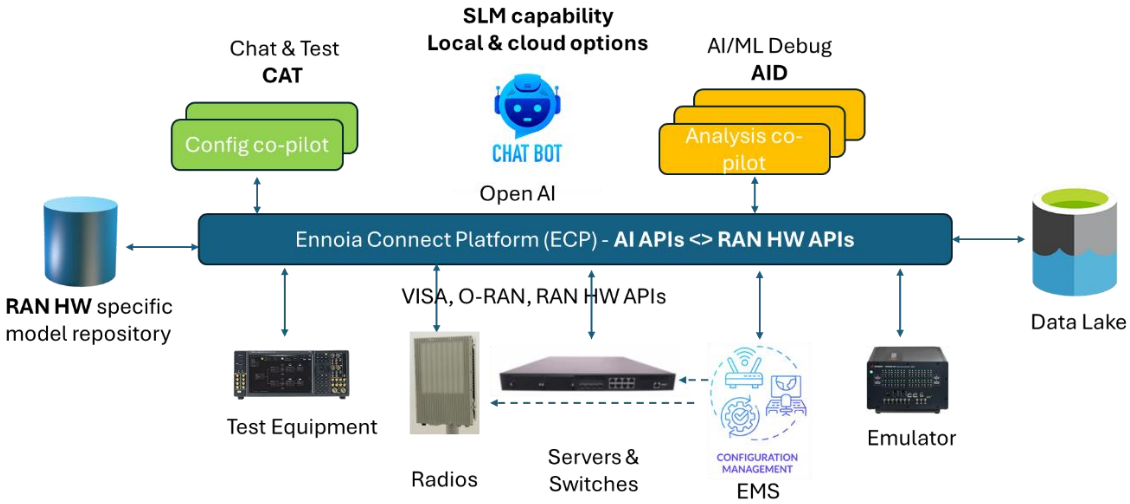


Figure 1 – Ennoia Connect Platform (ECP)

2.1 Innovation

The following are the ECP innovations:

1. We provide an HW abstraction layer between the AI APIs and the RAN HW APIs using the ECP.
2. We enable new hardware integration via generative AI where the APIs are extracted from the hardware documentation and auto integrated into API library used by the connect platform.
3. Unlike most intent-driven approaches which focus on gathering data stored in a data lake and modifying configurations based on intent (SW-only approach), we enable a new "hardware in the loop" approach for intent driven configuration and analysis. In this approach, the hardware is part of the AI model training and validation.
4. We provide AI-based configuration and analysis applications for various RAN HW that enable new capabilities such as interference detection and O-RAN fronthaul PCAP analysis.

2.2 Methodology

We have trained AI models on RAN hardware such that the user-intent is understood by the hardware. There are multiple features available in the ECP for configuration, analysis, plotting results, saving output, calibration, doing software updates, report generation using GenAI etc., which are enabled by the CAT and AID co-pilot functions. For example, a user-intent such as "Change the radio to operate in Band 77 with bandwidth 100 MHz" is translated to a series of configurations in real-time, which is then executed automatically by the ECP. This is provided by the CAT function in the ECP, which configures the HW. Also, a user-intent such as "Is the radio functioning properly?" executes the AID function in the ECP, which analyzes the output of the radio and informs whether the output is per expectations.

2.3 Development plan

Our product development plan includes three steps:

1. Manual integration and model training of new RAN infrastructure:
    a. We manually integrate new RAN infrastructure by understanding the APIs.
    b. We have already completed this step for multiple RAN infrastructure such as a few O-RAN radios and test equipment using VISA and O-RAN fronthaul APIs.
2. Developing new AI applications for analysis and debug:
    a. We develop new AI applications for analysis and debug of the output of the RAN HW.
    b. Initial applications include interference detection, PCAP analysis of O-RAN hardware and using genAI for generating summary reports.
3. Auto integration of new RAN infrastructure using generative AI:
    a. We use genAI to automatically integrate new RAN infrastructure where the AI learns the APIs by digesting the product manuals and then trains the models to convert intent into appropriate APIs for the models.

2.4 Outcome & Results

We have an initial release for our ECP product. Figure 2 describes the initial connection with a prototyping spectrum analyzer hardware called tinySA (www.tinysa.org) to the ECP. We have also integrated our product with commercial grade test equipment. The AI co-pilot identifies the hardware, and the user communicates with the spectrum analyzer in a natural language as shown in Figure 2.

Intent: How to configure the tinySA

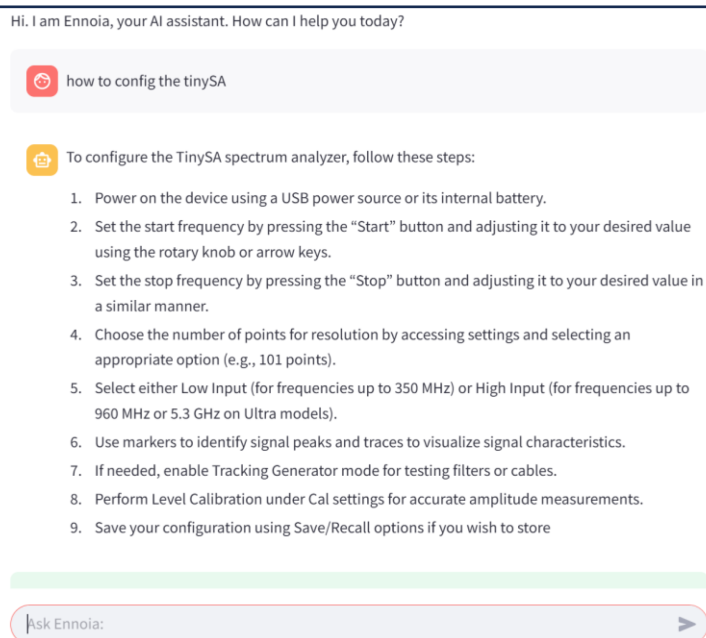


Figure 2 – AI Query example for how to configure the spectrum analyzer.

The example in Figure 3 shows the AI co-pilot screen, where the AI co-pilot is working with the SLM model offline and after finding the hardware infrastructure (tinySA) to communicate with. The SLM is running on the local GPU of a laptop, enabling data privacy and local AI compute. The user can ask the AI co-pilot to perform certain tasks and analyze and plot the output. Both manual and automatic options are available. In Figure 3, the user asks to change the start and stop frequencies to be between 600MHz and 900MHz.

Intent: Set the start frequency to be 600MHz and the stop frequency to 900MHz

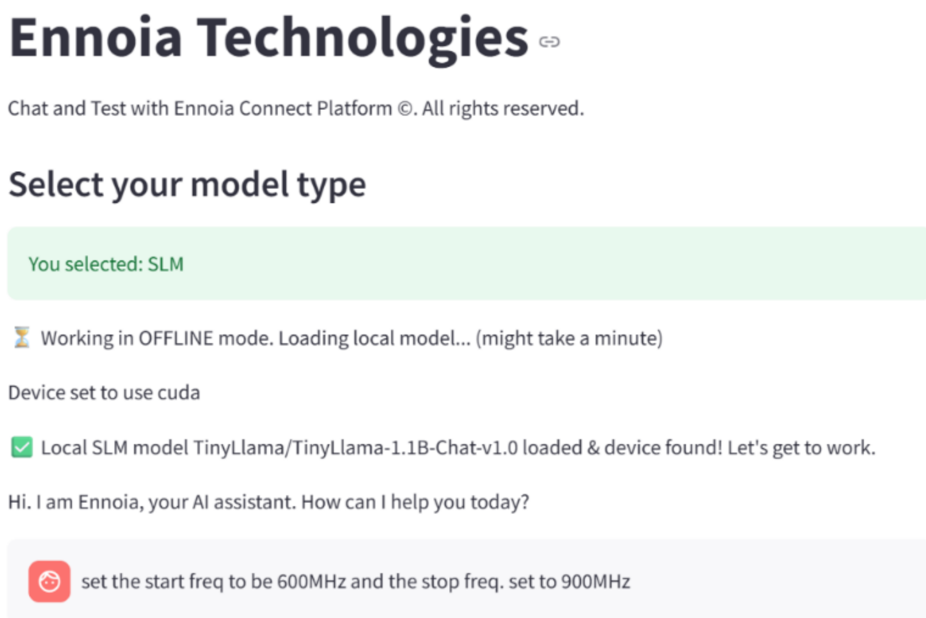
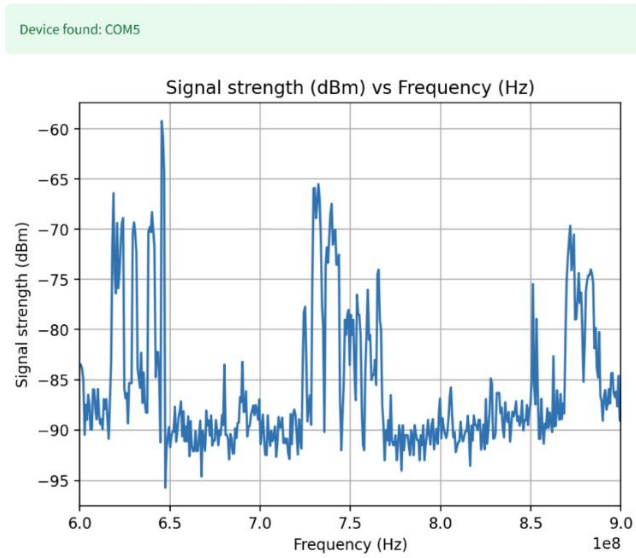


Figure 3 – AI Co-pilot to control RAN hardware using user-intent (SLM Mode)

An example of the output of the user intent execution is shown in Figure 4. When a spectrum analysis hardware (used in this example) is queried for frequency spectrum between 600 MHz and 900 MHz, the output returns a plot of the spectrum. In addition, the output of the spectrum is analyzed, and several operator bands are identified as seen in the spectrum. Thus, the output is made meaningful to the user instead of merely relaying raw data from the hardware.

Intent: Plot and analyze 600MHz to 900MHz spectrum



	operator	strength	technology	service	frequency_range	band_3GPP	source
0	T-Mobile	-68.22	LTE, 5G NR	Mobile	618 - 631	Band 71	Internal Database
1	T-Mobile	-69.55	LTE, 5G NR	Mobile	622 - 641	Band 71	Internal Database
2	T-Mobile	-63.75	LTE, 5G NR	Mobile	630 - 724	Band 71	Internal Database
3	T-Mobile	-67.66	LTE	Mobile	646 - 742	Band 12	Internal Database
4	T-Mobile	-71.28	LTE	Mobile	738 - 750	Band 12	Internal Database
5	AT&T (FirstNet)	-74.03	LTE	Mobile	753 - 869	Band 14	Internal Database
6	AT&T, Verizon	-71.52	LTE	Mobile	851 - 881	Band 5	Internal Database
7	AT&T, Verizon	-74	LTE	Mobile	877 - 885	Band 5	Internal Database

Ask Ennoia:

Figure 4 – Spectrum analysis output using intent (600MHz – 900 GHz)

ECP also supports advanced functions such as the use of ML and Gen AI for interference detection in an O-RAN radio. Using the fronthaul PCAP, the AI co-pilot can further detect issues such as interference and can generate a detailed report of the signal characteristics as well as confirm existence of interference as shown in Figure 5.

Intent: Analyze the downloaded pcap and report any interference

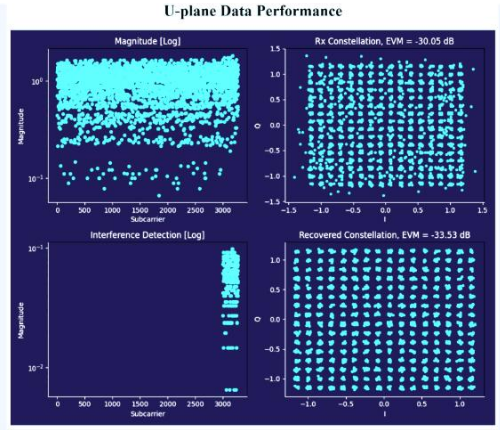
Report Header

O-RAN Fronthaul Packet Analysis Report

Item	Value
Filename	oran_uplane_output_interf_awgn_new.pcap
Date	2025-05-24 22:27:37
Status	✕
Issues	Found Interference

Fronthaul Data Summary

Variable	Value	Description
Sub-carrier spacing (KHz)	30 KHz	Defines spacing between sub-carriers.
Number Of Antennas	4	Represents antennas configured in this setup.
Max Frames	1	Indicates the max number of frames transmitted.
DL Direction	Uplink	Specifies the traffic direction.
U-Plane Packet Type	U-plane	Denotes user plane packets carrying user data.
Number Of PRBs	273	Count of Physical Resource Blocks for transmission.
Bandwidth Frequency (MHz)	98.28 MHz	Total bandwidth allocated for transmission.
Interference	1	Indicates if interference was found in the packet.



Detailed Analysis

1. Sub-carrier Spacing - The value of 30 KHz indicates the spacing between sub-carriers in the frequency domain. This spacing is crucial for defining the granularity of the available bandwidth and can directly affect both latency and spectral efficiency during transmission.

2. Number of Antennas - With 4 antennas configured, the system checks for multiple signal streams, enhancing signal reliability and increasing throughput capabilities compared to single-antenna setups. The use of multiple antennas can also improve coverage and mitigate fading effects.

3. Max Frames - The value of 1 for the Max Frames indicates a minimal transmission setting, possibly designed to support low-latency applications. This could be appropriate for scenarios where immediate data delivery is critical but may limit the capacity in high-demand situations.

4. DL Direction - Observing the direction as Uplink points to a configuration focused on user-generated data being sent to the network. This sets the context for the transmission efficiency and the resource allocation on the fronthaul link, highlighting its role in overall uplink operations.

5. U-Plane Packet Type - The designation as U-plane confirms the packets in focus are for user data transmission, encapsulating and carrying essential user information over the network. This encapsulation is critical for ensuring that user data is correctly processed and routed through the network architecture.

6. Number Of PRBs - The count of 273 Physical Resource Blocks indicates a substantial allocation of resources for the transmission process. A higher number of PRBs usually contributes to increased data rates and better overall service quality for users connected to the network.

7. Bandwidth Frequency - The allocated frequency of 98.28 MHz illustrates the total bandwidth available for transmission, which directly influences the data rate and capacity of the network. A higher bandwidth supports greater throughput and user demands.

8. Interference - The indication of 1, signifying found interference, is concerning as it suggests potential degradation of signal quality or loss of data integrity. Identifying the source and nature of this interference is critical for maintaining optimal network performance.

In conclusion, while the overall fronthaul data indicates substantial infrastructure capabilities and configurations, the presence of interference necessitates further investigation to ensure optimal performance and reliability of the network.

Prepared by: Ennoia's AI O-RAN Fronthaul Packet Analyzer

Figure 5 - Interference Detection on O-RAN Fronthaul

3. Impact analysis

We have two main contributions in our proposal:

1. We demonstrate the feasibility of user-intent driven AI co-pilots to build, test, deploy and debug AI-RAN system infrastructure using AI models specifically trained on the hardware. Based on business case analysis done for 3 Tier-1 operators, we expect that our platform will reduce time duration for issue analysis and resolution by 50%, decrease site visits by 30% and improve productivity and time-to-market with AI-RAN, especially in multi-vendor environments, saving operators significant OPEX cost.
2. We enable wireless systems AI developers to focus on the AI innovation without having to learn the details of the underlying hardware and RAN protocol aspects, thereby lowering the bar for AI-RAN innovation.

Our goal is to expose our AI-co-pilots solutions to a wider variety of AI-RAN infrastructure and mobile systems to enable faster integration and validation of AI intent-based wireless technologies.