

Large Perceptive Models

Dr. Lina Bariah

Digital Future Institute,
Khalifa University

The first AI wave taught machines to read, write, summarize, and reason

LLM

Text became
the universal
interface for AI

Multimodal models

Images, audio, and video
entered the loop

What comes next?

AI that understands
environments, systems,
and signals

Language was
the beginning —
not the endpoint

- The future of AI will live inside networks, cities, factories, transportation, and critical infrastructure
- Those worlds do not speak in paragraphs. They speak in streams, states, events, and signals.

The world is not text

A building **speaks**
through **sensors**

A city **speaks** through
mobility patterns

A network speaks
through **telemetry**
and **radio emissions**

A factory **speaks**
through **events, delays,**
and **interference**

Perceptive AI

Images

What is visible?

Audio

What is heard?

Sensors

What changed?

RF

What is invisible?

Topology

How is it connected?

Context

Why does it matter?

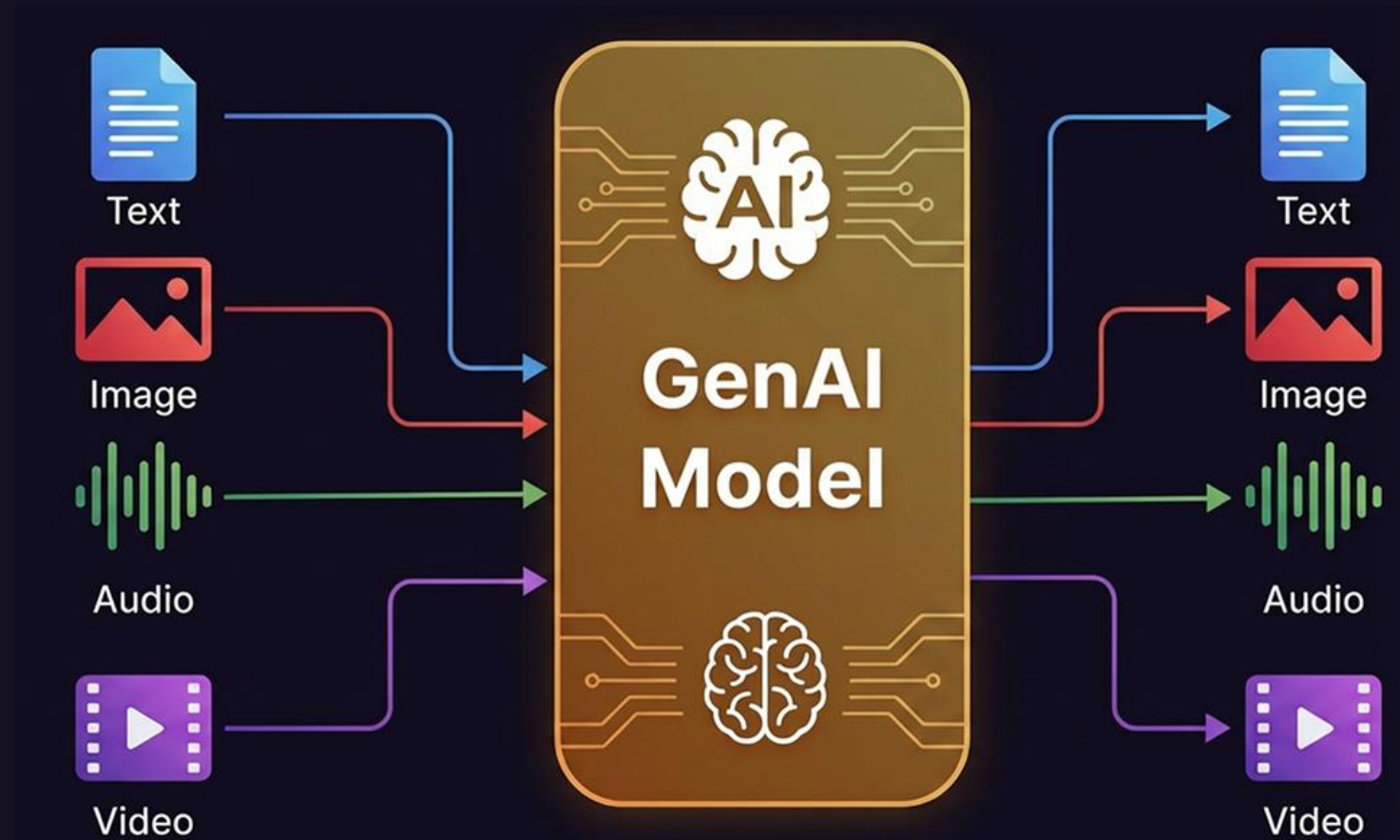
An intelligent machine must sense
reality before it can shape it

Beyond Text: Multimodal AI

Multimodal AI processes and integrates information from various data types: text, images, audio, video

Provides a more comprehensive understanding and robust outputs compared to unimodal AI

Mirrors human information processing for more intuitive and effective applications



From language intelligence to environment intelligence

Language

Documents,
code, dialogue

Meaning from
text

Multimodal

Images, audio,
video

Meaning across
media

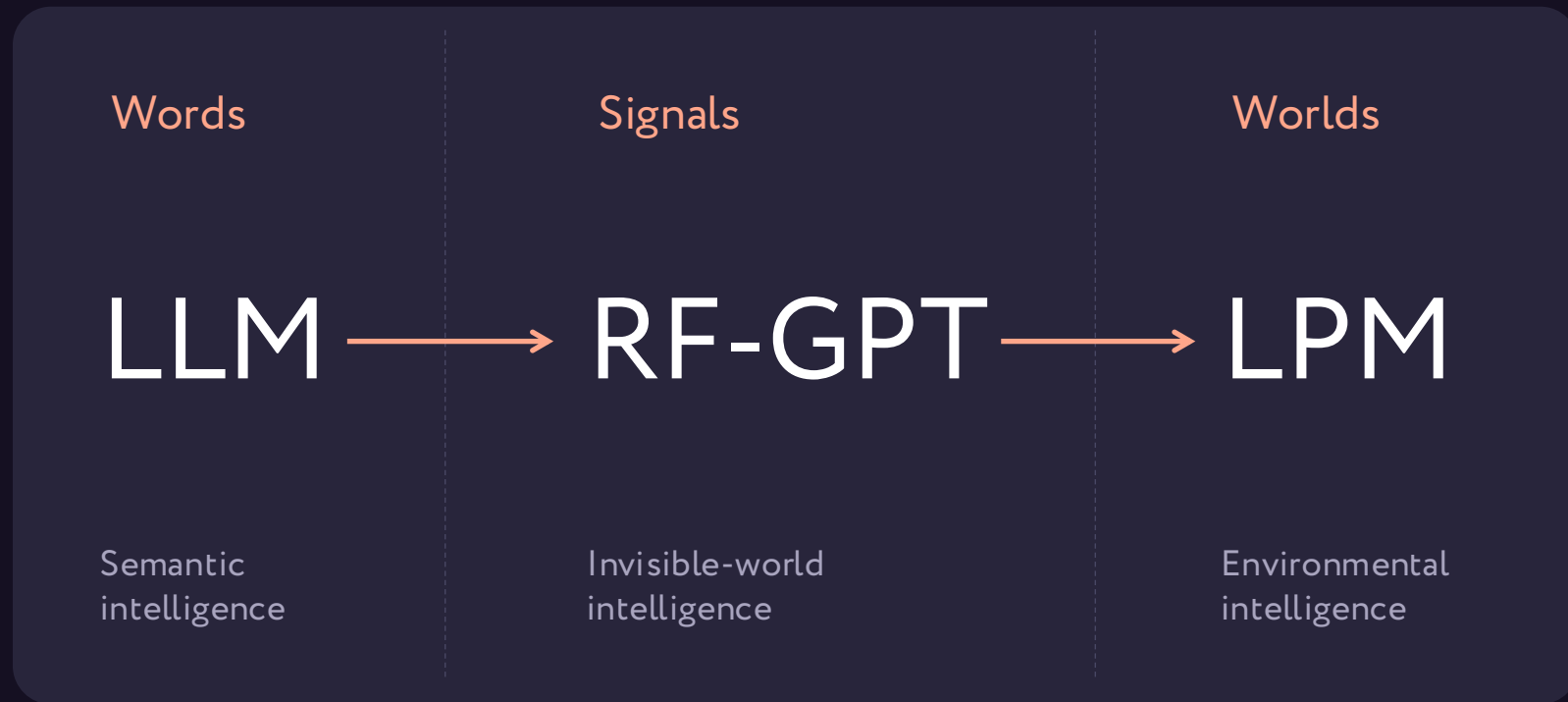
Perceptive

Systems, states,
scenes, signals

Meaning from
operational reality

The next generation of models will not only answer questions. They will diagnose environments and act

From words
to signals
to worlds



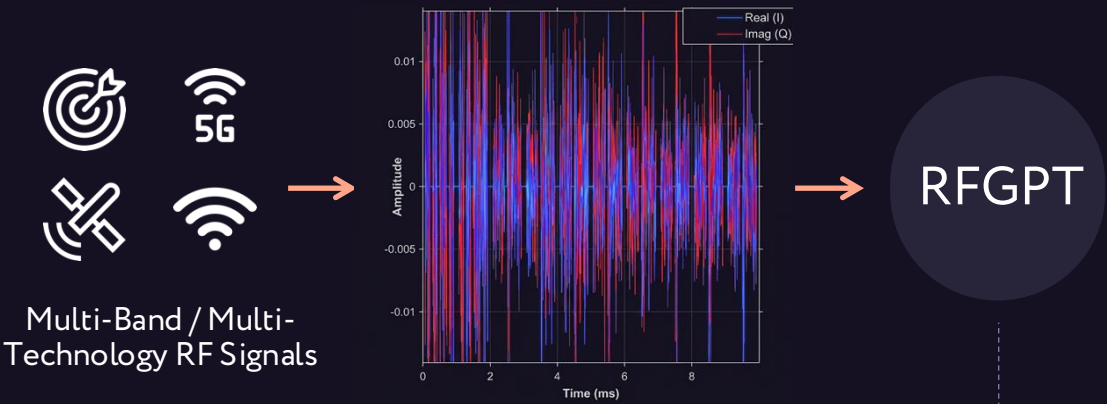
This is not just a sequence of models.
It is a widening definition of intelligence

RF-GPT

Invisible World Intelligence

RFGPT is an AI foundation model that understands, classifies, and interprets RF signals across technologies and spectrum bands

AI RF Intelligence Engine		
Protocol structure Analyze protocol layers of RF protocols (eg.. LTE, NR-SSB, WFI POCCH, Control Channel)	RF environment diagnostics Analyze anomalies , and can understand RF conditions	Wireless technology Classify technologies (G6, LTE, Wi-Fi, Bluetooth, IoT bands) operational in
Spectrum occupancy Detects PRB allocations, channel occupation, and spectrum utilization patterns	Temporal RF activity Track RF activities considering durations, bursts and timing petterns	Signal & modulation Analyze modulation schemes, MCS, waveform properties for performance insights



AI RF Intelligence Engine



Talk to the Signals!



Describe this RF scene



This capture shows a 5G-NR downlink in FR-1 with two active user equipment (UE1 and UE2). Both transmissions share the same carrier but occupy separate PRB allocations.



Do you observe other wireless technologies?



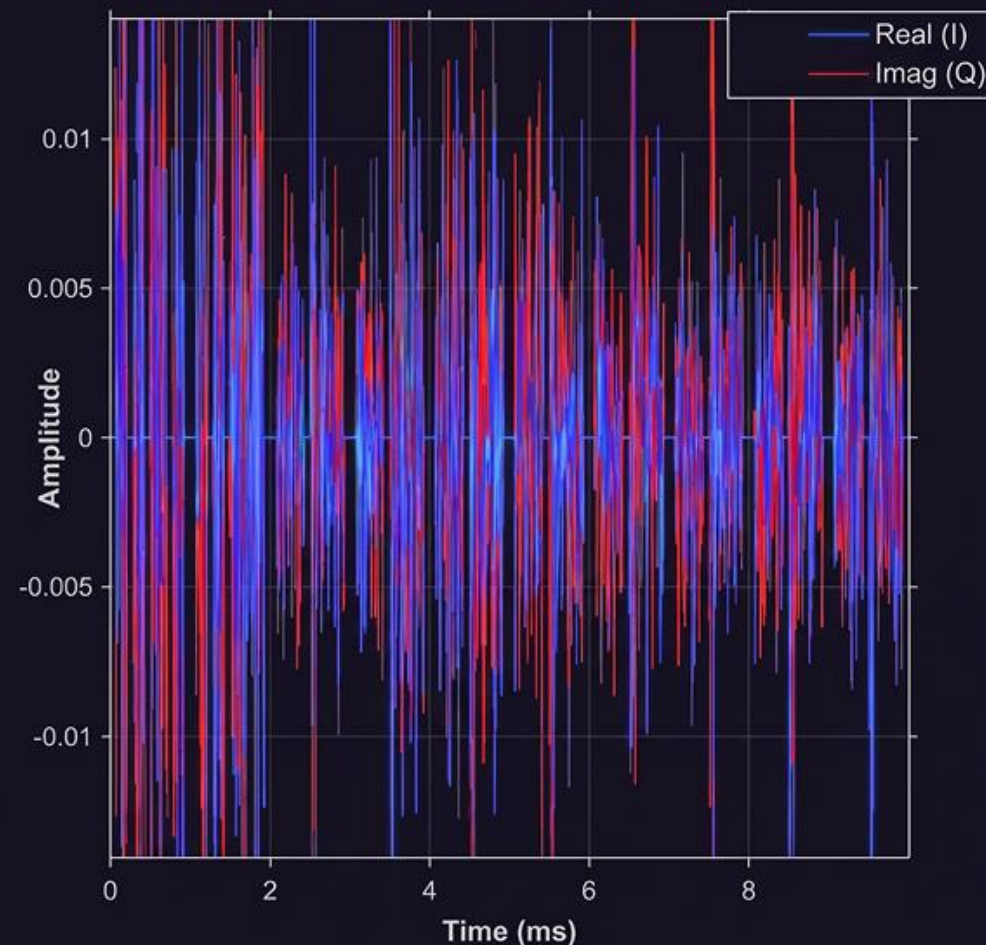
No. Only 5G-NR signals are present.
There is no LTE, Wi-Fi, Bluetooth, or other wireless technology detected.



Any anomaly in the spectrum? Do the transmissions overlap?



No anomalies detected. The SSB, PDCCH, and DM-RS signals follow expected NR patterns, and the power levels indicate a low-SNR scenario. The two UE transmissions overlap in time but remain separate in frequency, each occupying its own PRB segment.



Understand the Signals!



What are the users likely doing based on this RF activity?



Periodic uplink bursts from UE1 → consistent with video conferencing (likely real-time encoding). - Downlink-heavy traffic with large transport blocks for UE2 → indicative of video streaming or content consumption

Combining RF scheduling, buffer status reports, and modulation patterns:

- UE1 is likely engaged in an interactive application (Zoom/Teams-like)
- UE2 is passively consuming high-throughput media.

Predict the Signals!



What will happen to this RF scene in 10 seconds?



This is a dense urban 5G-NR FR1 deployment with moderate mobility. Based on Doppler spread and scheduling patterns, I predict:

UE1 will experience a 3~5 dB SINR drop due to an incoming blockage (likely a moving vehicle).

The gNB scheduler will reallocate PRBs toward UE2, which shows more stable channel conditions.

A beam switch is expected within ~6 ms to maintain link reliability.

Act on the Signals!



Optimize this RF scene for latency-critical applications.

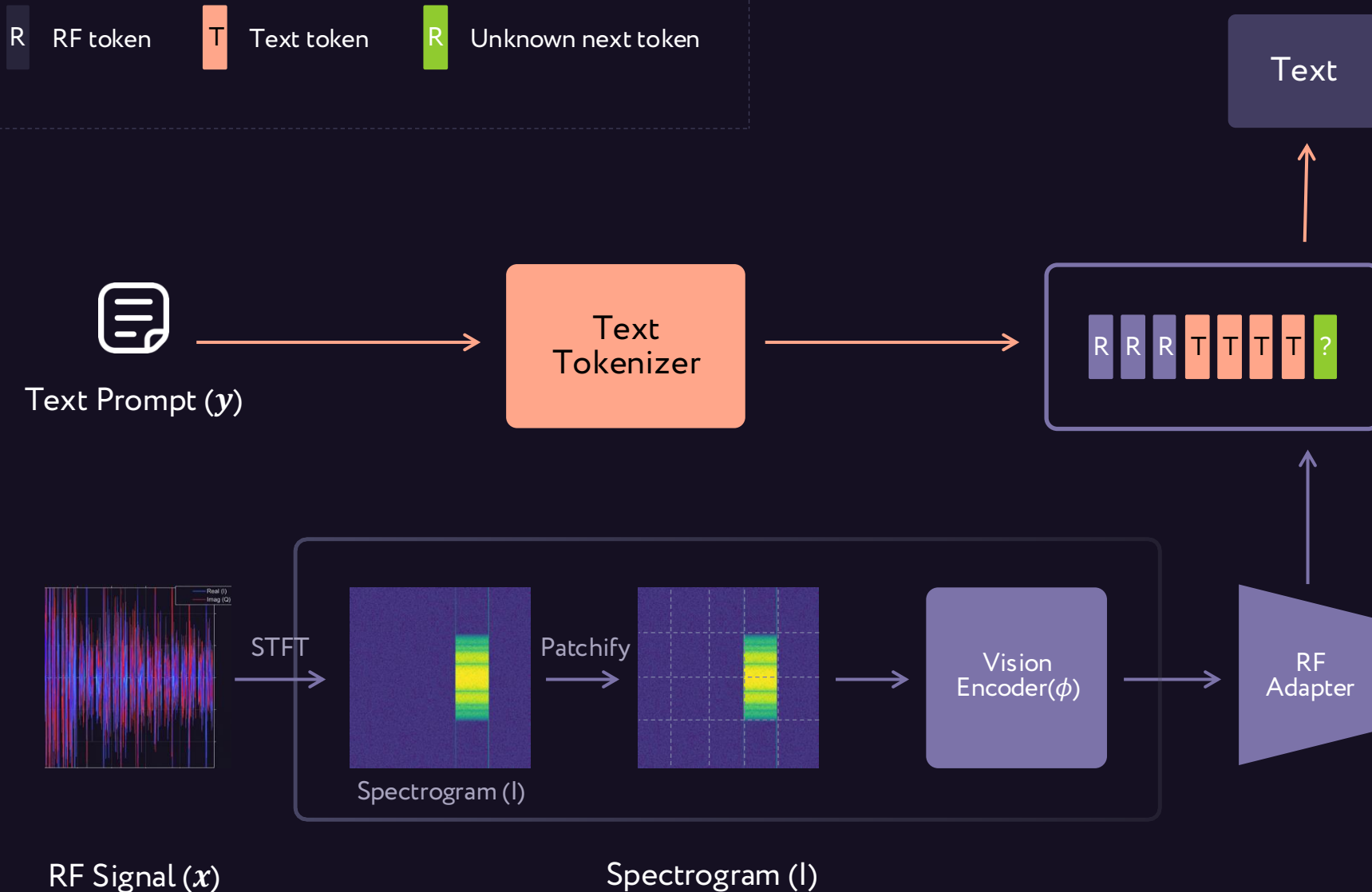


Reassign UE1 to higher-priority scheduling (URLLC-like profile).
Reduce MCS slightly to improve decoding reliability under low SNR.
Trigger beam refinement to stabilize channel conditions.
Allocate contiguous PRBs to reduce fragmentation delay.

Projected outcome: 25~40% latency reduction for UE1;
Slight throughput tradeoff (~8%) for UE2

How RF-GPT Works?

H. Zou, Y. Tian, B. Wang, L. Bariah, C. Huang S. Lasaulce, and M. Debbah, "RF-GPT: Teach AI to See the Wireless World", (Under Review) IEEE TWC



Live Demo

Implementation



Laptop +
RTX 4090

GPU FFT +
AI Model

USRP
B210 SDR

70 MHz - 6 GHz
56 MSPS max

Log-Periodic
Antenna

Wideband Reception
380 MHz - 6 GHz

AI Model

Comparison

Same 806 MHz signal
analyzed by two
different vision models

Llama 3.2 Vision (11B)

Generic vision-language
model No RF domain training

VS

RFGPT (7B)

Fine-tuned on RF spectrograms
Domain-specific knowledge



Judge: Gemini Pro — evaluated both outputs for accuracy and usefulness 

“

The signal is a narrowband signal with
bandwidth of approximately 20.0
MHz. **The signal appears to be clean
and free from significant noise.**

“

3G UMTS downlink (TM1 32-DPCH), 5
MHz UMTS carrier, energy spread across
carrier, medium SNR, continuous time
burst, **no overlapping signals**

LPM

Environmental Intelligence



Camera

Smell



RF

Taste



Temperature

Contact



Pressure

Sound



Large Perceptive Model

Enabling Intelligent Perception & Understanding

Design principles for the next generation of LPM

Physics-aware representations

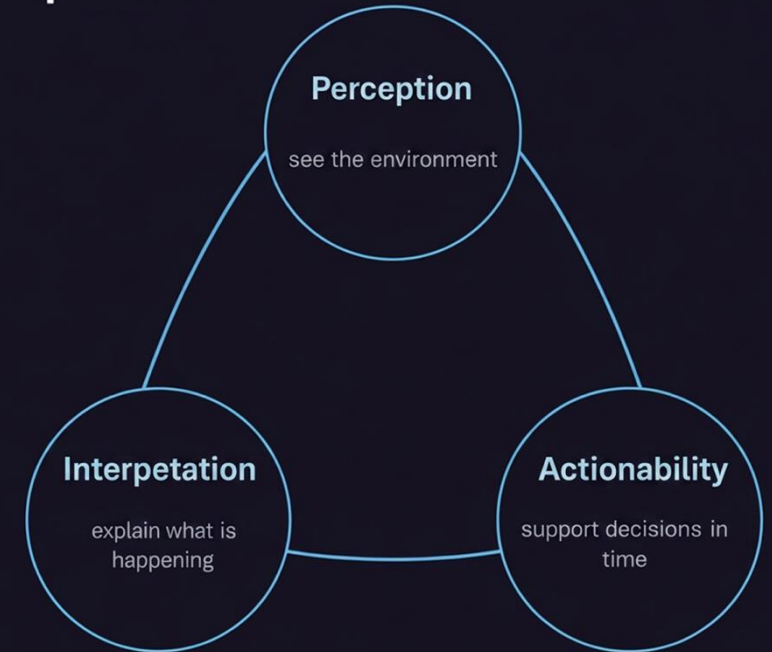
Cross-modal alignment between sensors, topology, and language

Natural-language explanation without losing signal fidelity

Edge-to-cloud deployment and real-time constraints

Trust, observability, and human oversight

The winning model will combine three properties:



Where this leads next

Physical AI

Robots, sensors, and
environments connected
through perceptive models

AI-native networks

Networks that
understand, explain,
and adapt in real time

Invisible intelligence

Systems that perceive what
humans cannot directly see

New interfaces

From dashboards to dialogue
with complex environments

Five takeaways

- 01 Large Language Models were the **first chapter**, not the whole **book**
- 02 Large Perceptive Models extend AI from **text** to **environments**
- 03 **IoT** is a natural testbed for **perceptive intelligence**
- 04 **RF-GPT** shows that AI can learn to **understand the invisible spectrum**
- 05 The next frontier of AI is not only what **machines can say** — it is what **machines can perceive**

Thank you!