

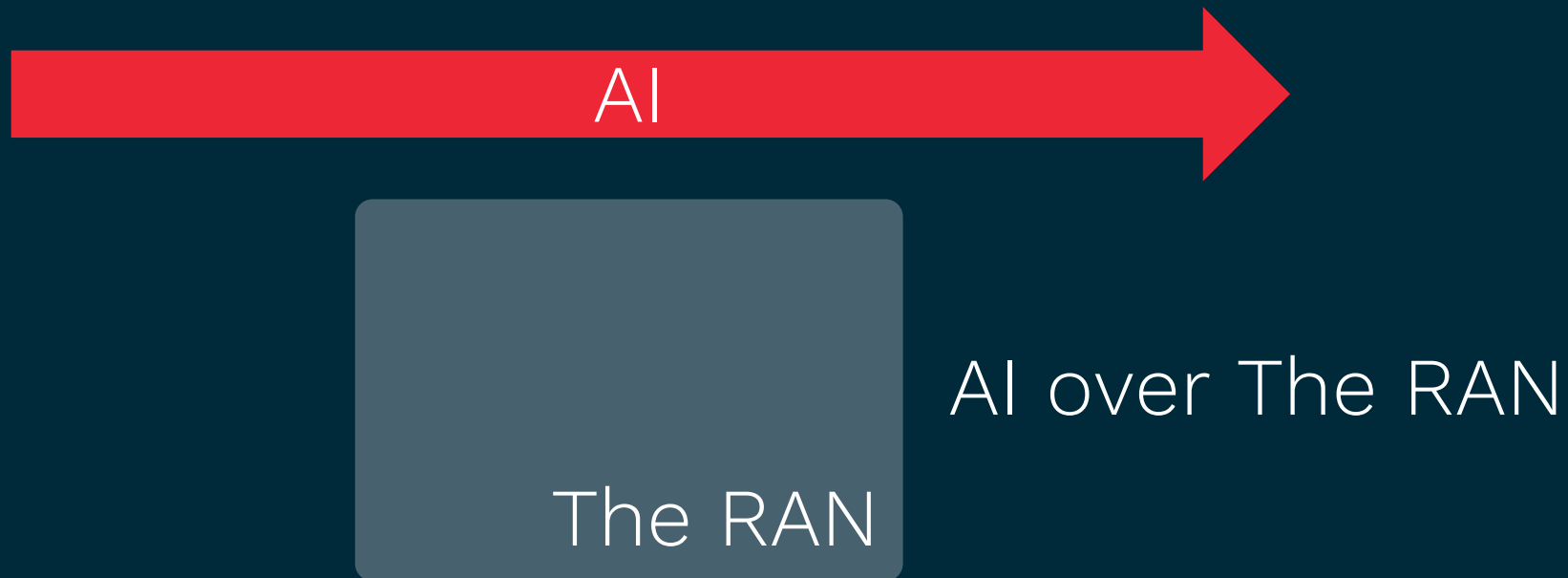


Our World.
Squared.

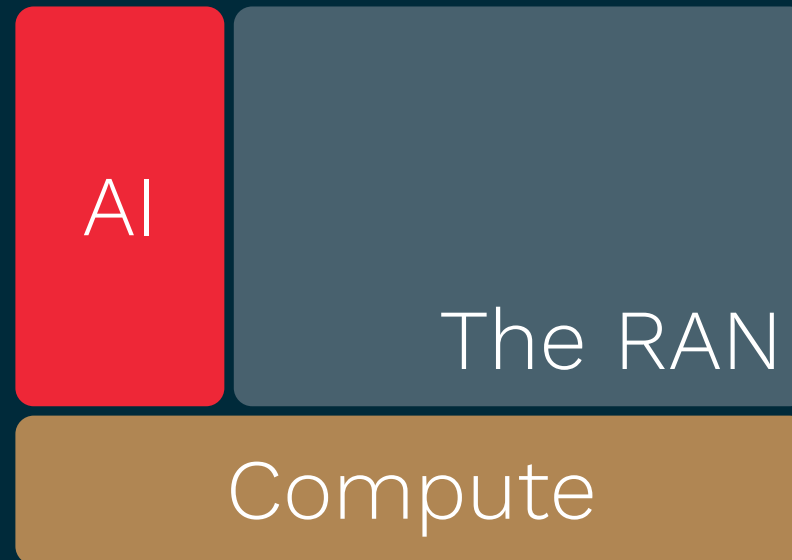
Putting AI Properly in The RAN

Peter Claydon – CEO

Artificial Intelligence and The RAN

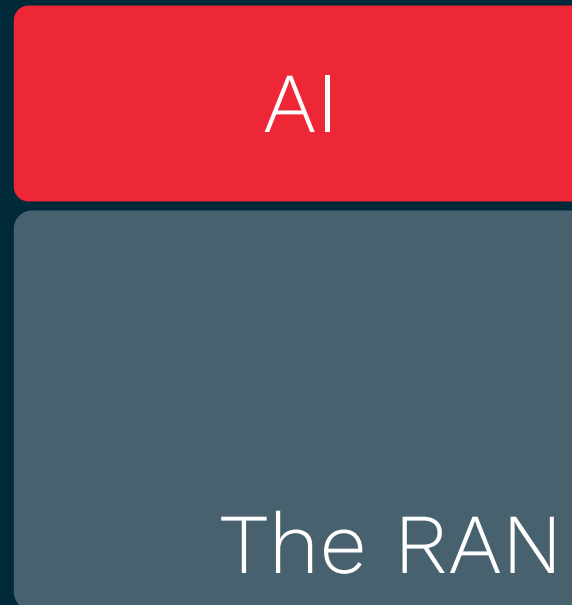


Artificial Intelligence and The RAN



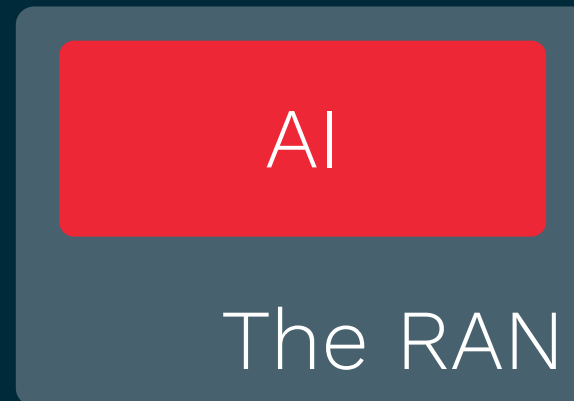
AI and The RAN

Artificial Intelligence and The RAN



AI for The RAN

Artificial Intelligence and The RAN



AI in The RAN

AI in Telecoms – The Myth...



“challenges for AI in Telecom”

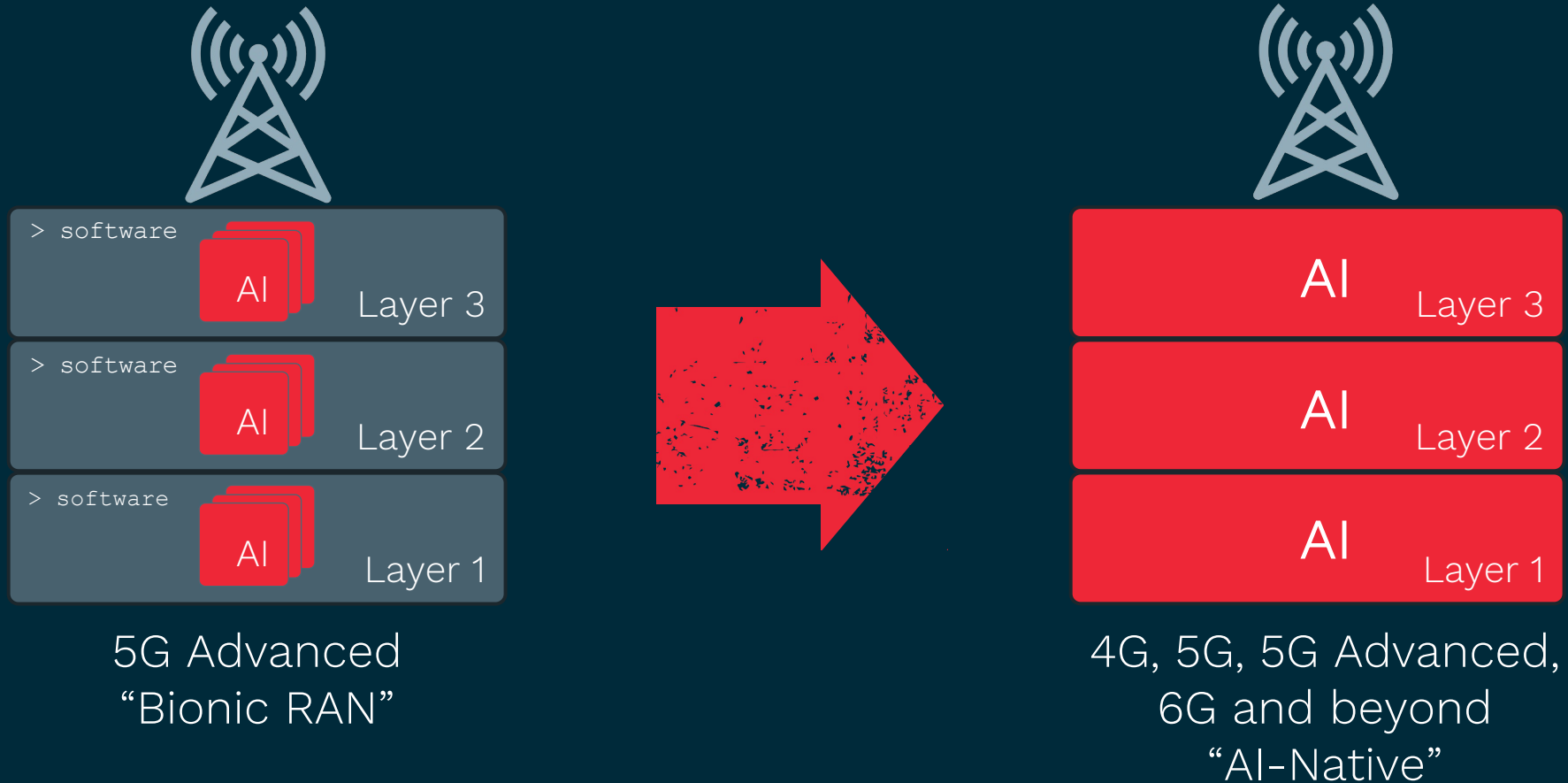
“ROI-driven”

“driving growth through AI”

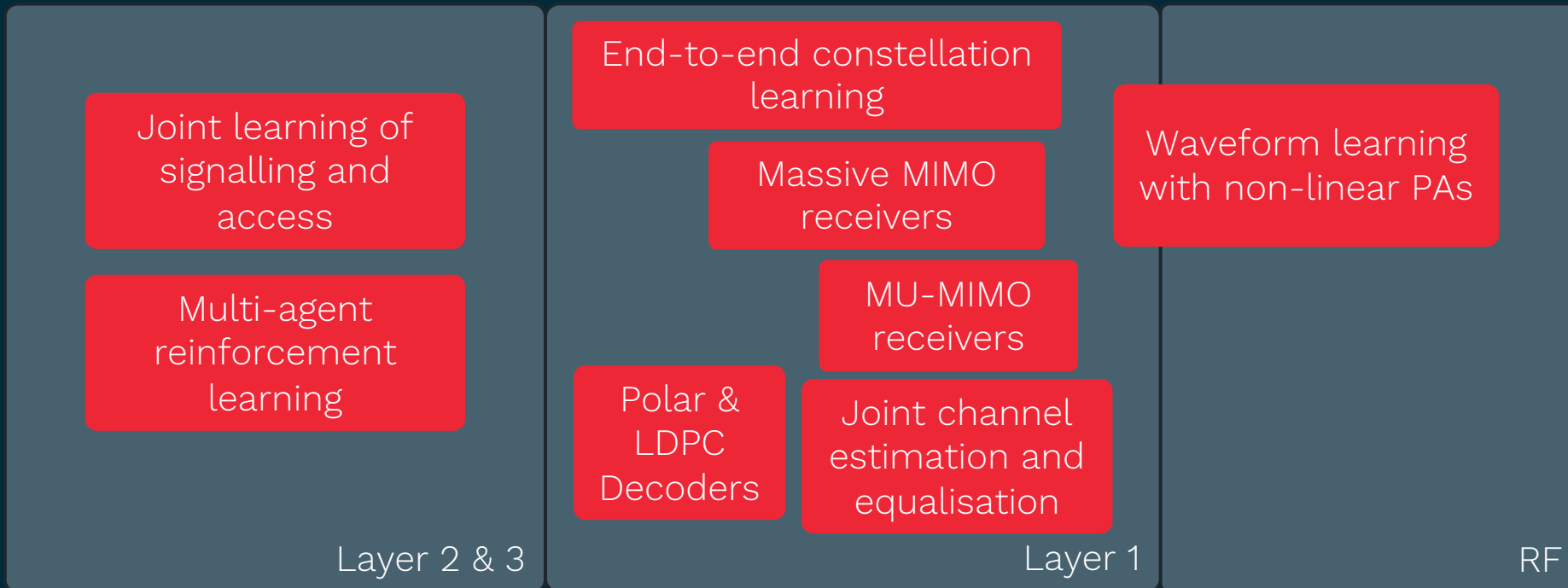
“AI-enhanced services”

Mostly vague and very high-level.

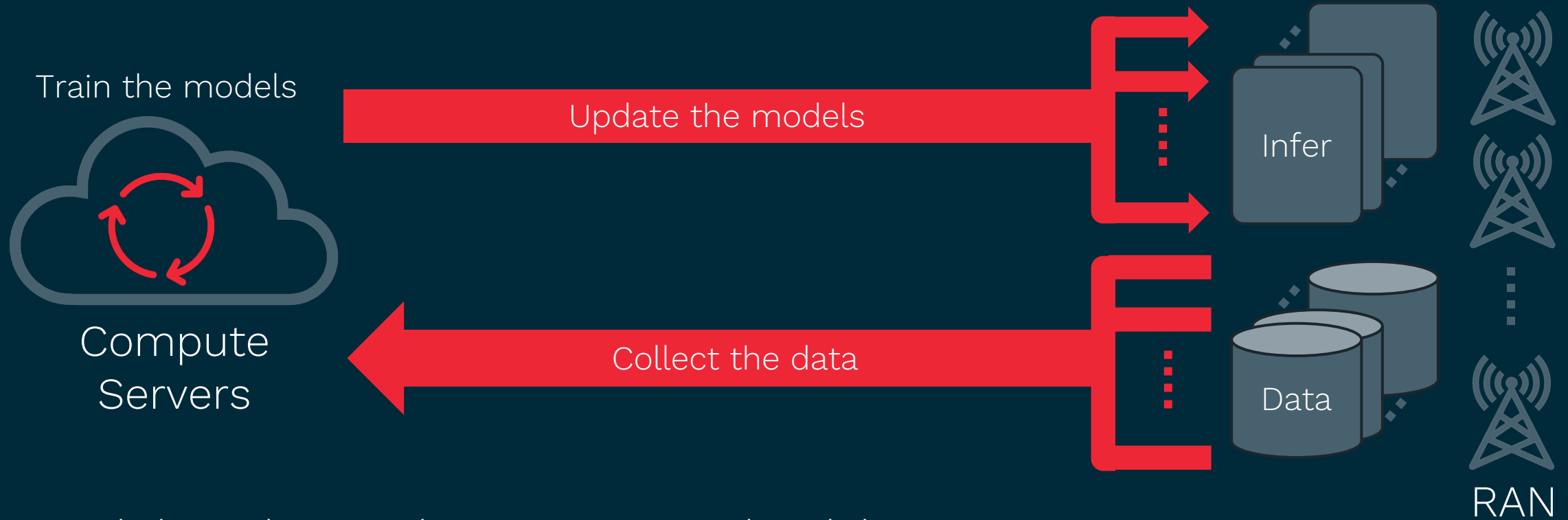
Specifics: AI-Ingress to AI-Native



AI in Baseband: the story so far...



AI-model Life-cycle Management



And then there's the monitoring, the debugging, etc., etc.

A “Life Support System” with negative CAPEX & OPEX impact

Understanding the Base Station Challenge

Baseband processing

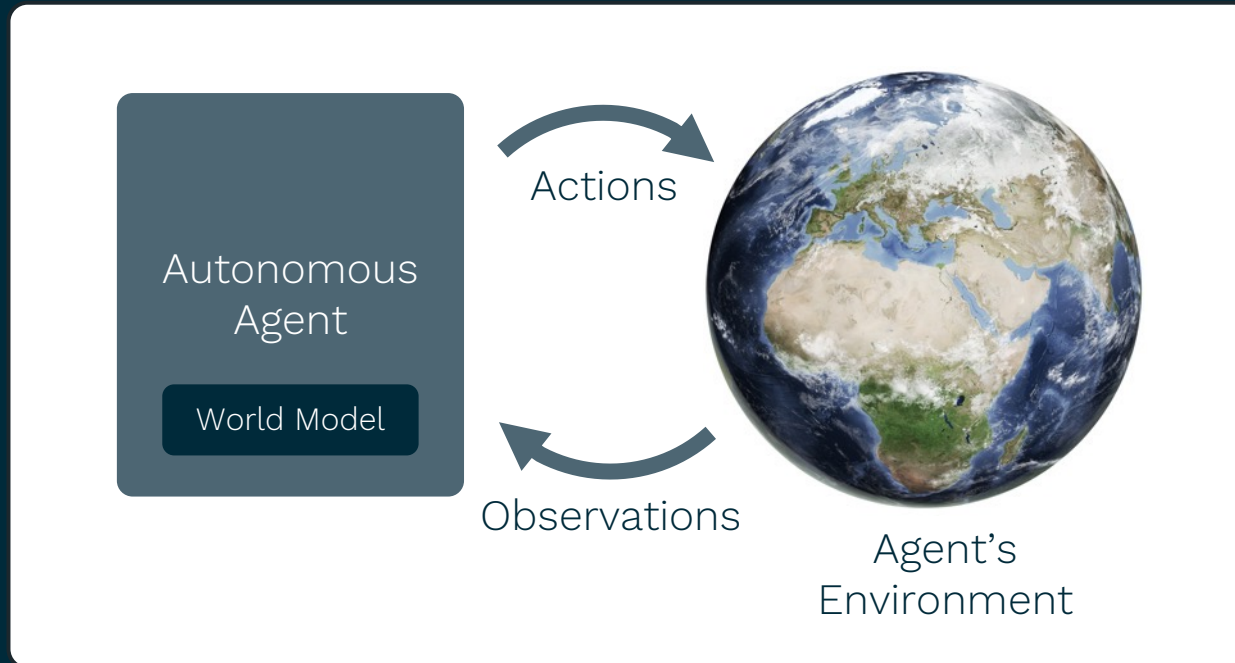
- High throughput
- Ultra-low latency
- Deterministic behaviour
- A mixture of types of processing
- All the parts must act together as a system

Often needs to be mast-mounted (cf ULPI)

- To achieve required performance
- Hence needs to be low power, passively cooled

General-purpose AI chips are not suitable

What actually is an “agent”?



Cybernetic Loop

- Wiener, N. (1948)

A good regulator needs a “World Model”

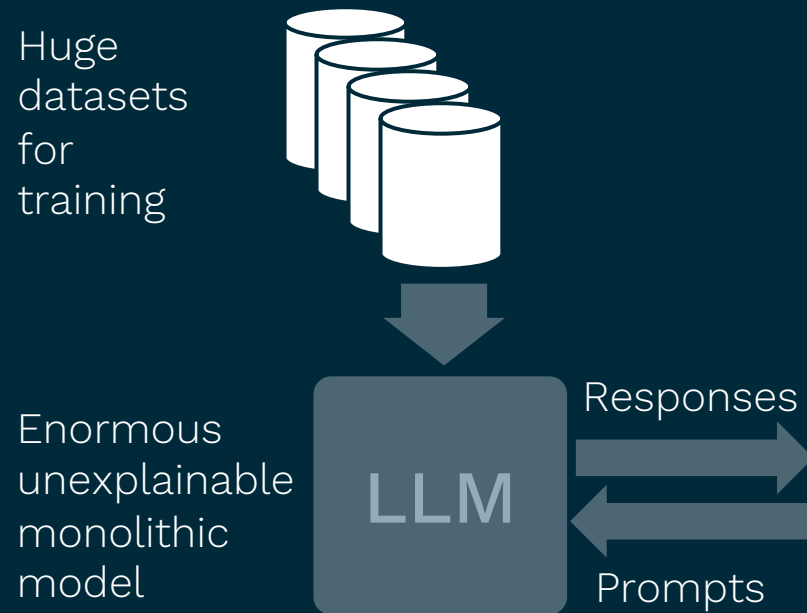
- Conant, R. C., & Ashby, W. R. (1970)

A system that would adapt its policy if it found that how its actions influenced the world had changed

- Google DeepMind (Kenton et al., 2022)

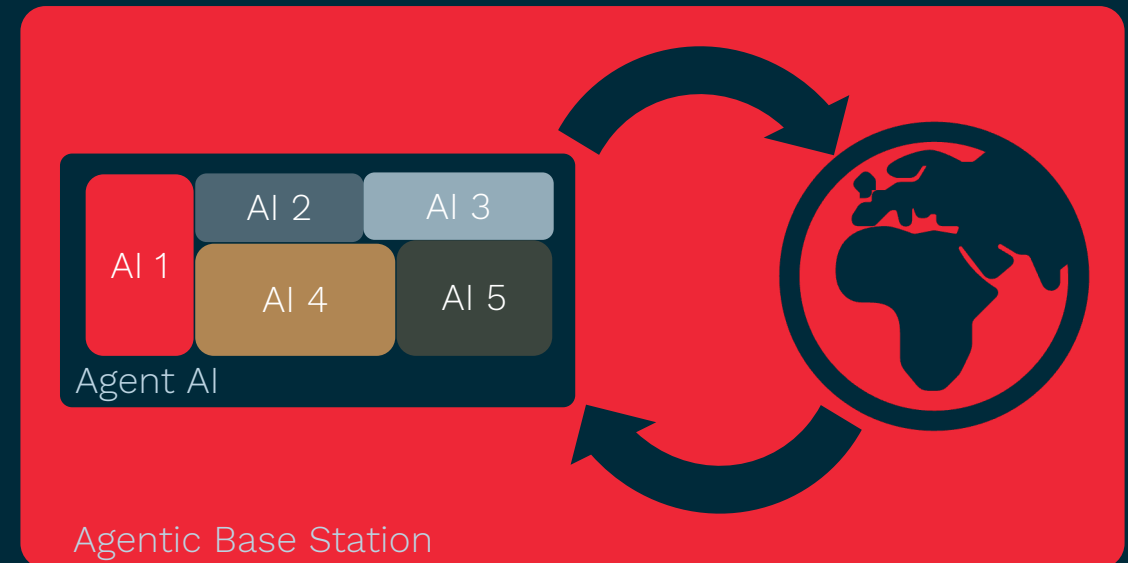
From LLMs to AI-Native RAN

Large Language Models



Low throughput: 100s words/s
Latency insensitive
“Biological real-time”

Real-Time Embedded Composite AI



High throughput: Gigabits/s
Latency sensitive: ~ microseconds
“Deterministic machine real-time”

When is an agent not an agent?

*“Calling Current LLM Wrappers “Agents” is Like Calling Excel Macros a Programming Language Revolution”**

*<https://hackernoon.com/calling-current-llm-wrappers-agents-is-like-calling-excel-macros-a-programming-language-revolution>

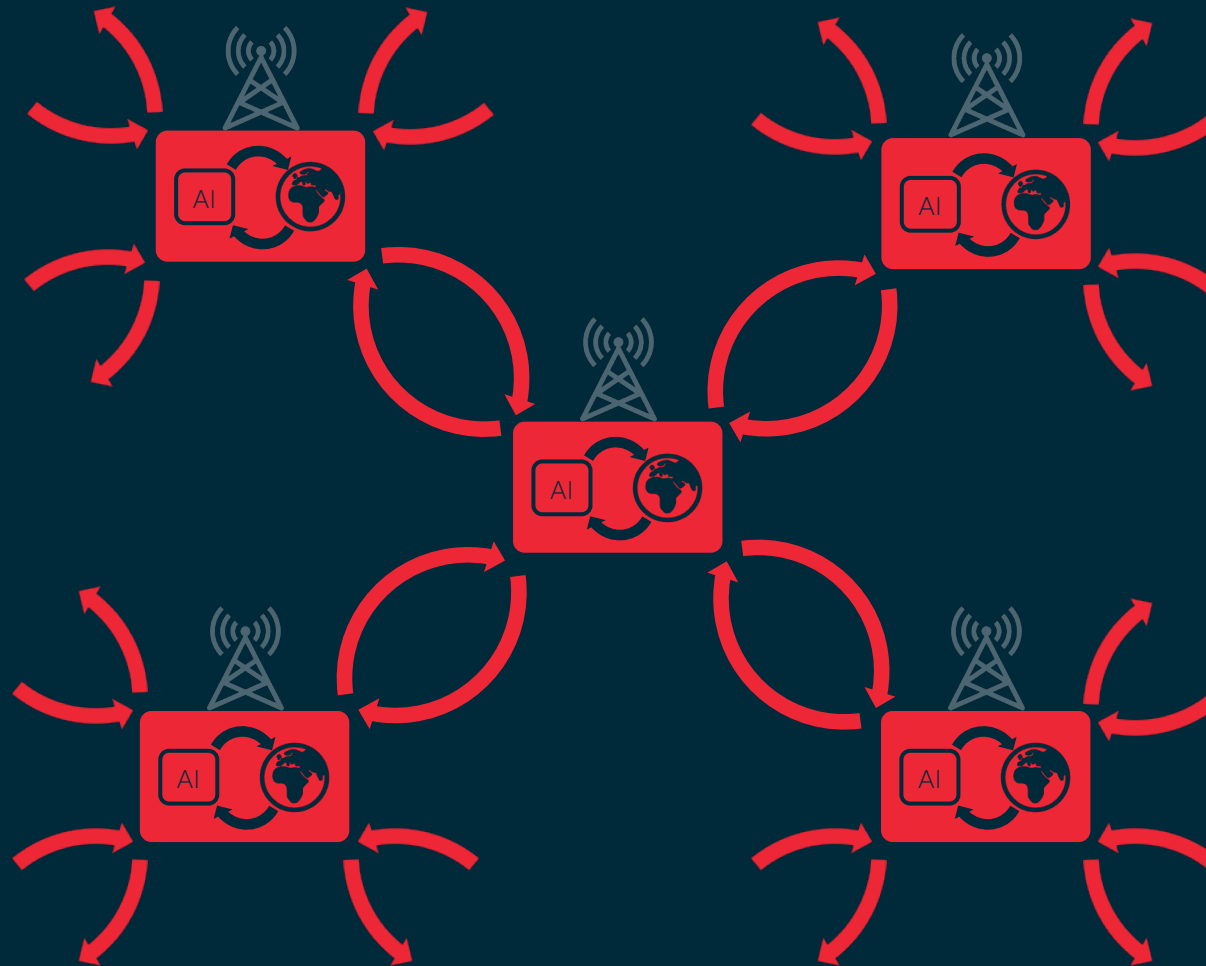
Processing for Agentic Base Stations

Just the right amount of:

- Inference acceleration
- DSP hardware acceleration
- DSP software performance
- I/O type and bandwidth
- Processing performance
- Cost
- Power consumption



Collaborative Multi-Agent RAN



Making AI-Native a Reality

Real-time embedded signal processing in baseband is hard

- And represents a very different problem for AI

It's challenging to add current AI to the RAN edge at scale

- Scaling requires a new approach with local learning on local data

Baseband processing cannot all take place in the cloud

- Or even in an air-cooled unit at the cell site

Base stations acting as true autonomous agents scale!



Our World.
Squared.

info@ransemi.com

ransemi.com