Qualcomm

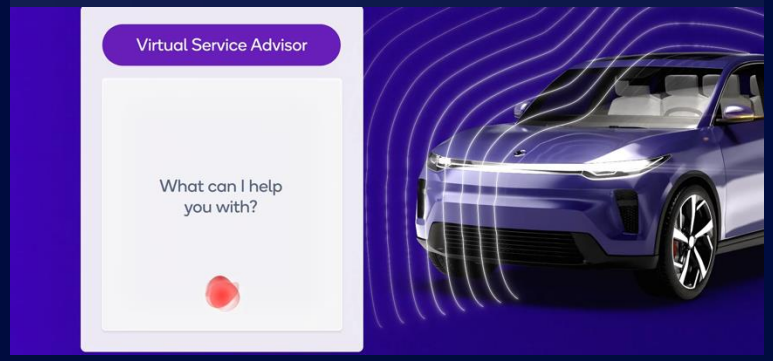
AI is changing the future

Luiz Tonisi
President, Qualcomm Latin America

AI is changing the future

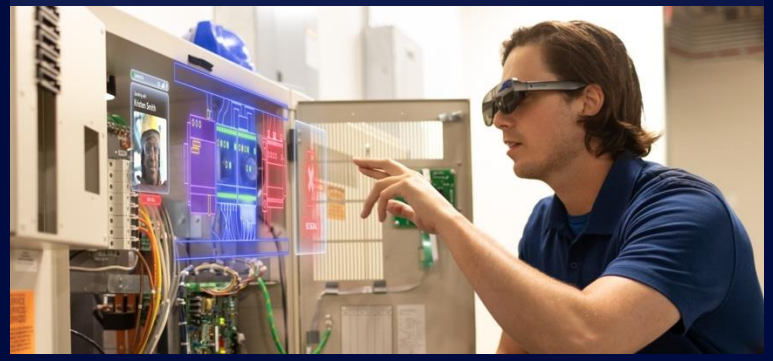
HANDSETS

AUTOMOTIVE



PC

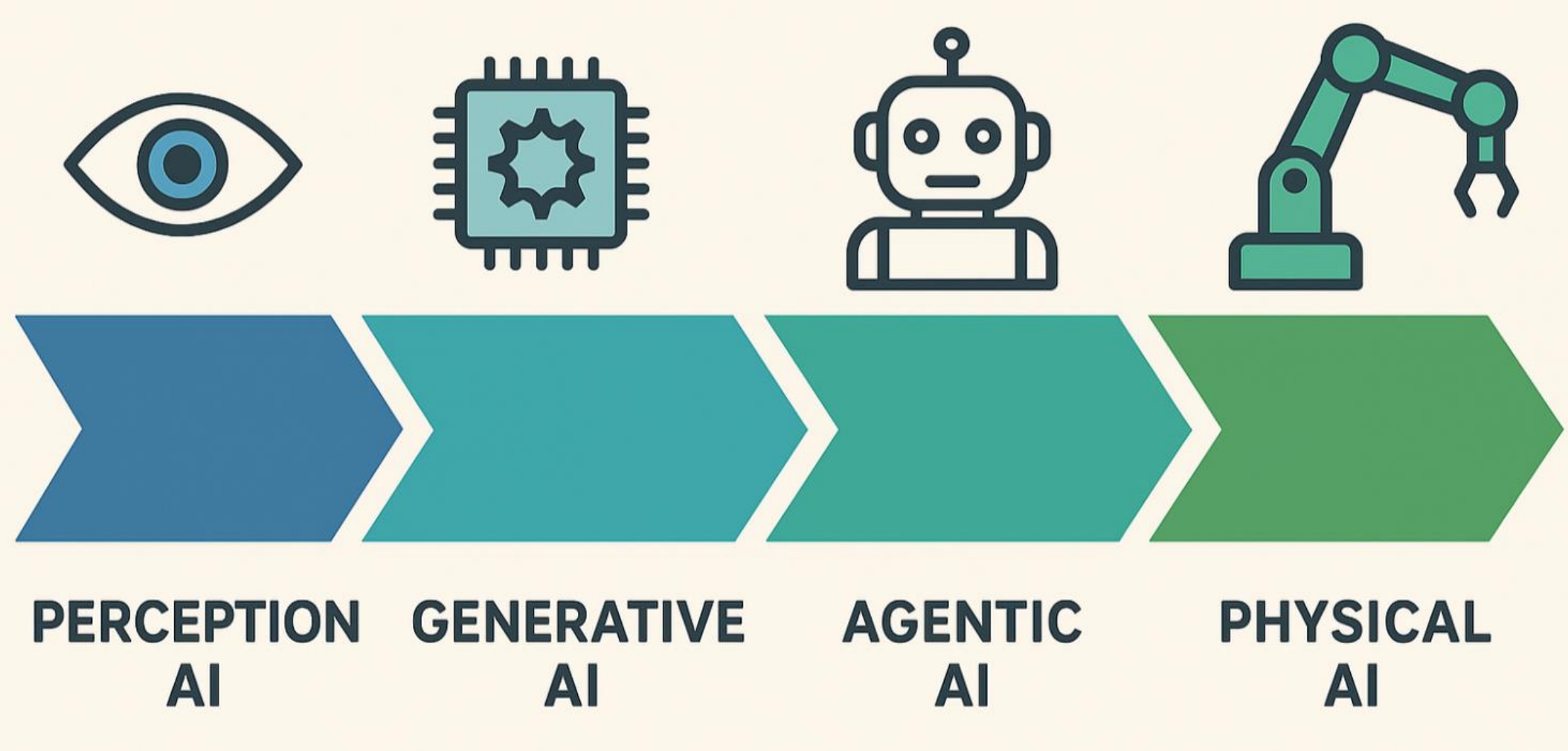
XR



INDUSTRIAL

NETWORKING

The AI industry itself is rapidly evolving

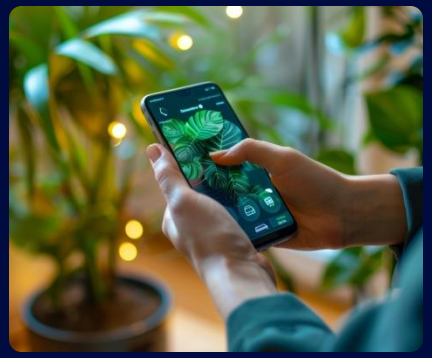


AI is the new UI

CREATING A NEW UI ACROSS DEVICES

Voice/audio

Text/touch

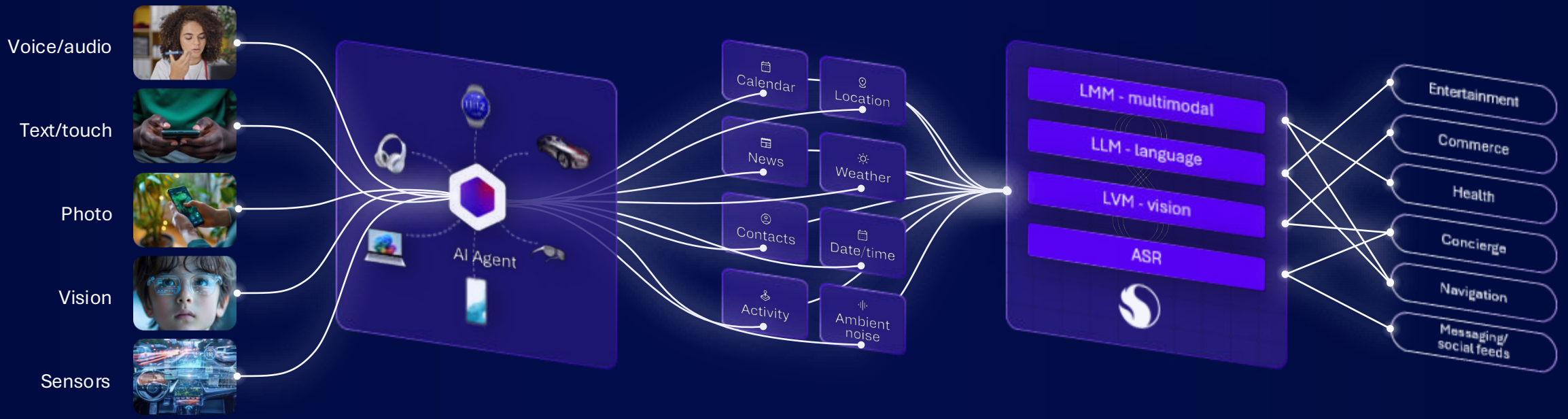


Photo

Vision

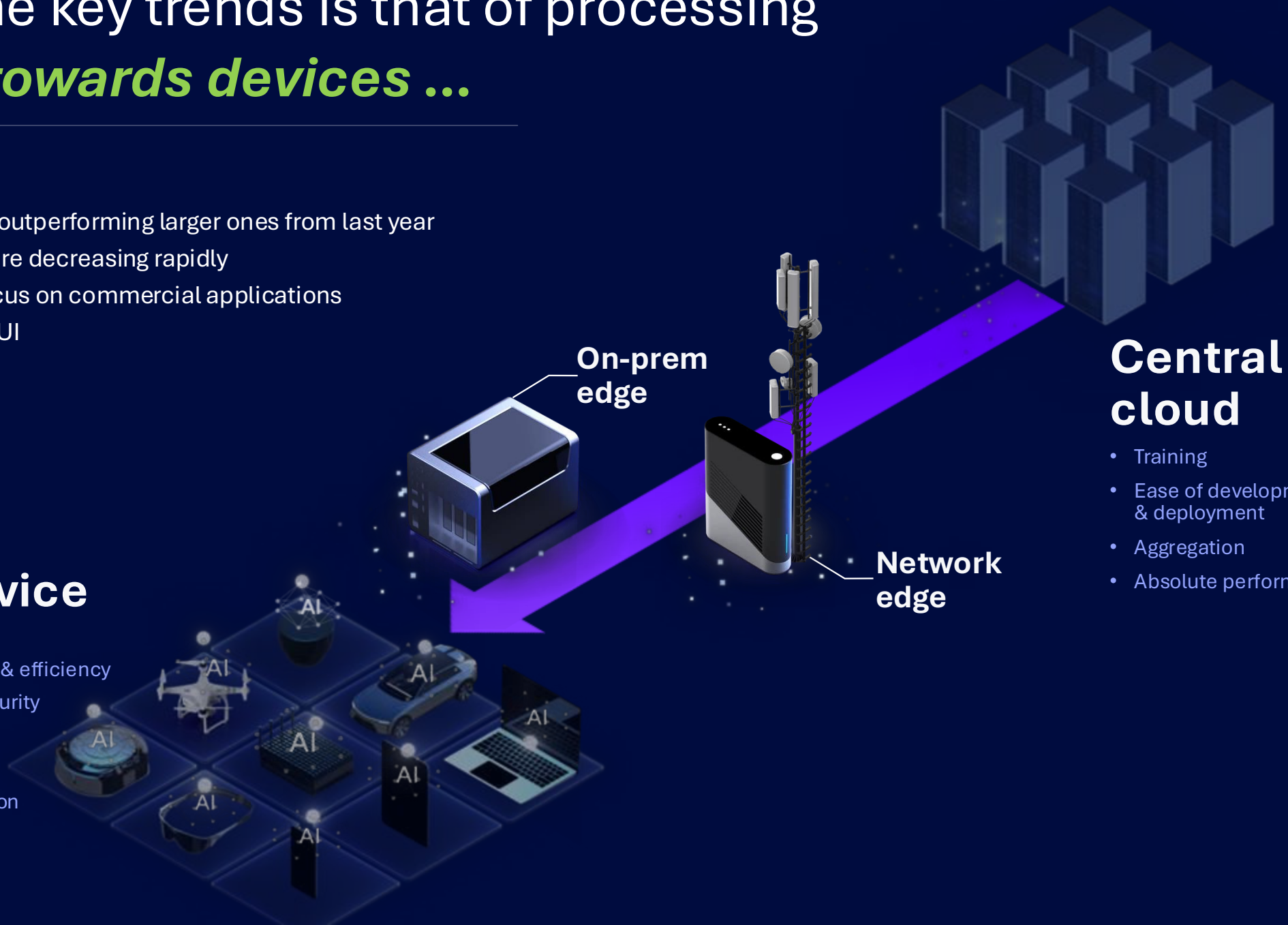
Sensors

Supporting key use cases



One of the key trends is that of processing *moving towards devices ...*

- New models outperforming larger ones from last year
- Model sizes are decreasing rapidly
- Increased focus on commercial applications
- AI is the new UI



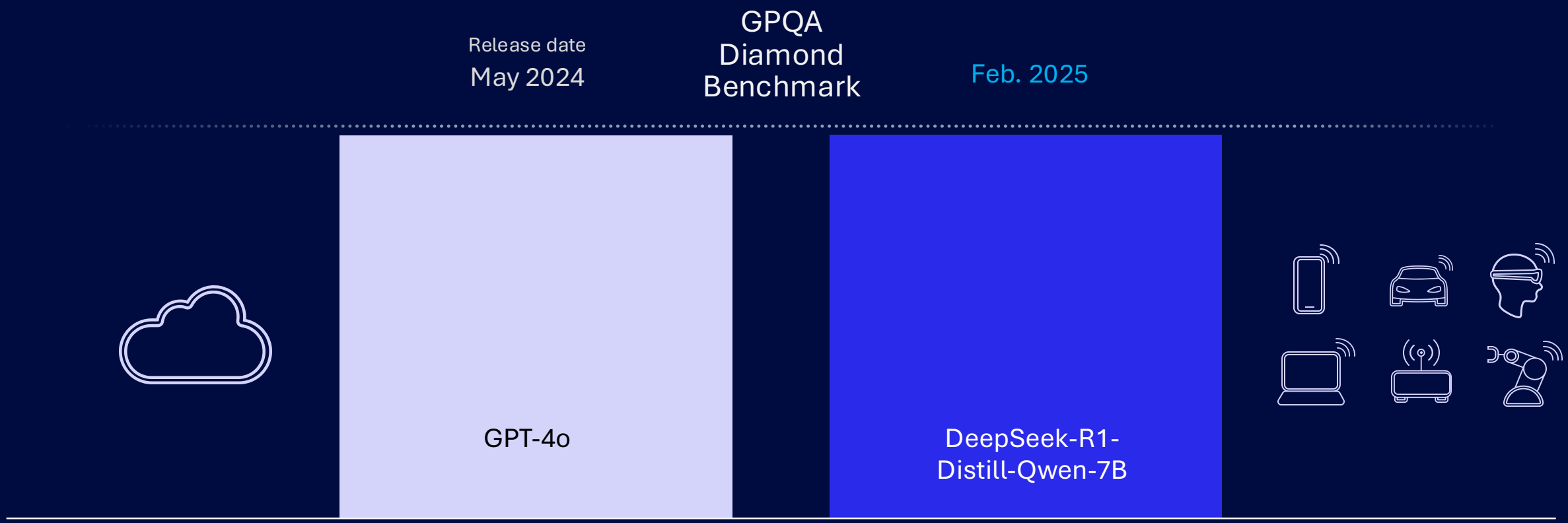
Central cloud

- Training
- Ease of development & deployment
- Aggregation
- Absolute performance

On-device

- Inference
- Performance & efficiency
- Privacy & security
- Immediacy
- Reliability
- Personalization

The rise of on-device AI model capabilities is, in fact, **fast accelerating** in 2025



- *GPQA Diamond é um subconjunto de alta qualidade do benchmark GPQA (Graduate-Level Google-Proof Q&A), focado nas 198 questões mais desafiadoras. Ele é projetado para avaliar o raciocínio de nível de pós-graduação em modelos de linguagem, com perguntas que exigem profundo conhecimento e raciocínio, em vez de apenas recuperação de fatos.*
- *Gethub, Feb. 2025; evaluation results. The maximum generation length is set to 32,768 tokens. For benchmarks requiring sampling, we use a temperature of 0.6, a top-p value of 0.95, and generate 64 responses per query to estimate pass@1*

Cloud inferencing costs hinder GenAI **scalability….**

	OpenAI	Meta	ANTHROPIC	MISTRAL AI_	Alphabet
CLOUD MODEL	GPT-4o	Llama 3.1 405B	Claude 3.5 Sonnet	Mistral Large 2	Gemini 1.5 Pro
CLOUD INFERENCING PRICE PER 1M TOKENS	$4.38	$5.13	$6.00	$3.00	$2.19

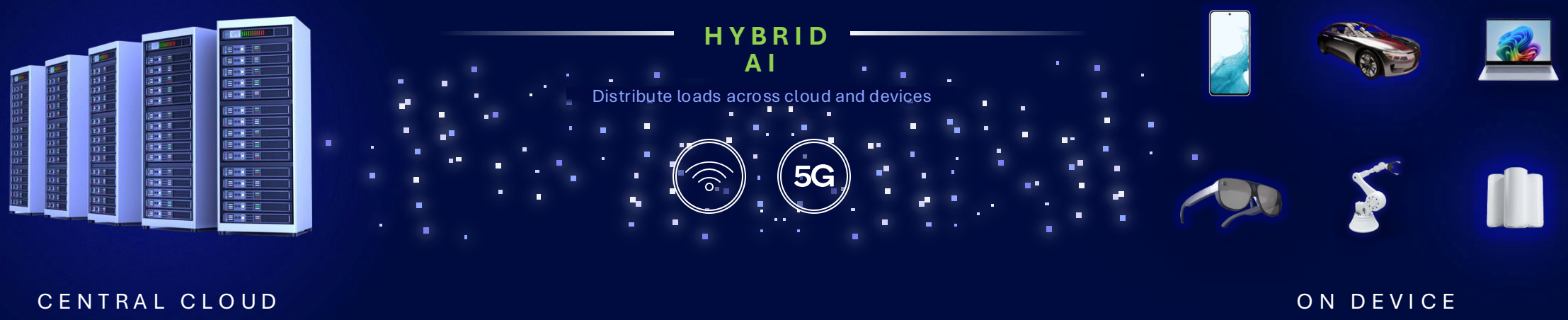
CENTRAL CLOUD

Artificial Analysis, Nov. 2024.

....while Edge AI brings **critical cost-efficiency opportunities**

"[AI inferencing on devices] reduces operational costs for developers and app providers"

- IDC, Oct. 2024



Train on cloud Run on Qualcomm

← **Post**

Elon Musk ✓ 𝕏
@elonmusk

Inscrever-se ···

It’s an easy prediction of where things are headed.

Devices will just be edge nodes for AI inference, as bandwidth limitations prevent everything being done server-side.

Traduzir post

> **X Freeze** ✓ @amXFreeze · 21 de ago
> xAI's long term plan is to be a edge node running AI inference to generate pixels and audio
>
> No more traditional OS or apps but just AI rendering everything directly

What's next

INCREASING MODEL SIZE

1-10B+ parameters

20-60B+ parameters

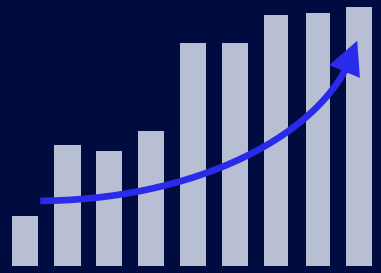
13-20B+ parameters

1-4B+ parameters

1-7B+ parameters

1-10B+ parameters

INCREASING QUALITY



INCREASING CAPABILITIES

Longer context

Personalization

Multi-modality

Concurrent models

AI is also transforming Robotics

robotics

THE MEGA TRENDS

- Labor shortages due to 4D work
- E-commerce fulfillment
- Robots learn & work with humans
- Digital twins

THE CHANGE CATALYSTS

- AI & LLMs are redefining what is possible on the robot
- Improved sensing & CV
- Demand for automation
- Ecosystem expansion

Use Cases

Lawnmowers | Retail Robots | Service robots | Warehousing | Off-highway | Industrial automation

Industry Verticals

CONSUMER HOME & LIFE	COMMERCIAL RETAIL	LOGISTICS	OFF-HIGHWAY	AUTOMATION	PROFESSIONAL SERVICES
Robot vacuums Companion robots Education robots Surveillance	Cleaning Shelf scanning Concierge Stocking	Moving pallets Sorting Pick-n-place Delivery	Agriculture Construction Mining Last mile delivery	Manufacturing Pick-n-place Conveyors Humanoids	Lawn mowers Polish, sand, etc. Surveillance Concierge in malls, airports

Qualcomm has a protagonist role in AI enablement across community of developers, enterprises & telcos in Brazil

Developer Enablement

- **Brazil is top#3 Qualcomm AI Hub community of users globally!**

- **Qualcomm spearheading dozens of AI Developer enablement programs in the country**
- **AI-focused Hack-a-Tons**

AI on Premises

- **Qualcomm driving AI deployment across B2B Verticals in Brazilian market**
  - Energy, Mining, Retail, Industrials, Logistics
- **Brazil is a promising Edge AI market in global scenario**
  - Est. US$ 0.8 B (2026 *)
- **Use-Cases**
  - Machine vision for manufacturing, Smart Surveillance, Retail personalization, Energy grid optimization, Autonomous logistics

Enabling Telcos

- **Qualcomm has been crafting strategic AI partnerships with LATAM telcos**
  - Telco AI Services
    - Edge AI computing
    - AI+Connectivity (ex: AIGW, RAN Orchestration/Automation – EdgeWise)
    - AI software stack
    - Industries use-case enablement
    - Design and implementation of monetization and commercial model

Sources: Grand View Research, MarketsAndMarkets, Mobility foresight

Key takeaways

- **AI is rapidly evolving:** The AI industry is undergoing significant changes, with advancements in model capabilities and applications
- **On-device AI:** AI processing is moving towards devices, offering benefits like performance, efficiency, privacy, security, immediacy, reliability, and personalization
- **Quality of small models**: The quality of small models suitable for on-device AI is rapidly increasing, making them more effective for various applications
- **Cost-efficiency:** Edge AI brings relevant cost-efficiency opportunities by distributing loads across cloud and devices, reducing costs for developers and app providers
- **Diverse model choices:** There is a wide range of models available for edge AI, including both traditional and proprietary models
- **Future trends:** The future of AI includes increasing model quality, personalization, longer context, concurrent models, and multi-modality capabilities
- **AI as the new UI:** AI is becoming the new user interface, supporting key use cases like text creation, content creation, code generation, live translation, photo/video editing, and productivity
- **Increasing compute needs:** There is a growing need for compute power, AI capable of running models on the edge, and safety to allow robots to work alongside humans

Thank you

Follow us on:

For more information, visit us at qualcomm.com and qualcomm.com/blog