

CREATE
CONNECT
LIVE
inspire

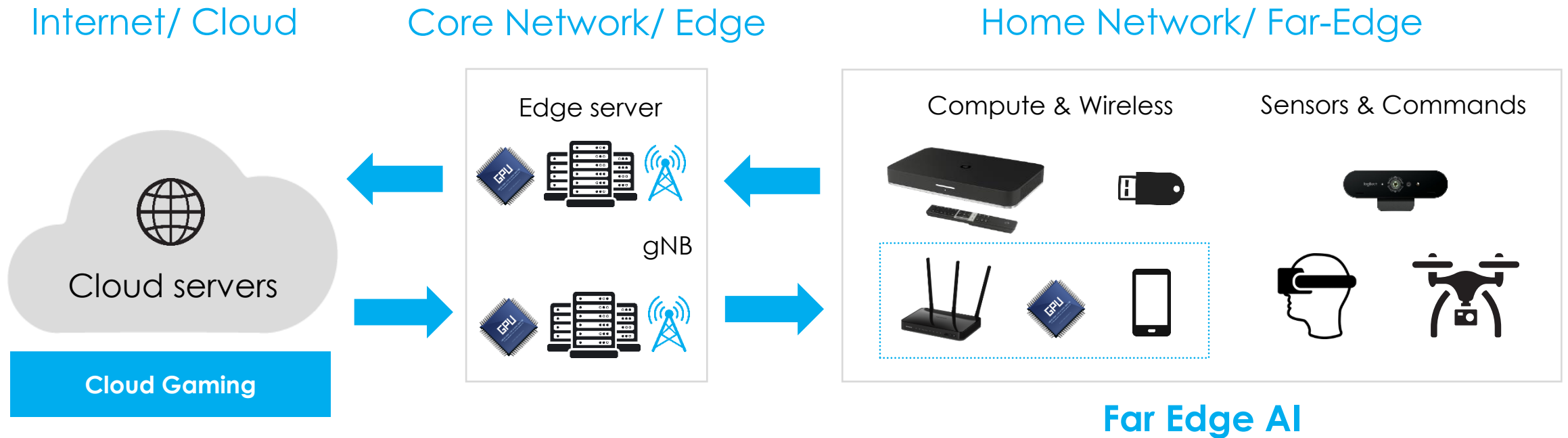
Cloud Gaming with Far Edge AI

In partnership with:



interdigital™

How does Edge AI Operate in an Edge-Cloud Architecture?



Edge AI supports a plethora of wireless verticals, from gaming to healthcare, while distributing tasks across fog, far edge, edge, and backend cloud.

The Vision for Far Edge AI



AI on generic devices maintains very limited, heterogenous, and volatile resources to leverage available local compute resources within the home...

Enabled by:

- Emerging Edge AI accelerators like Google TPU, Intel VPU, and Nvidia GPU
- Forthcoming open platforms & standards
- Recent advancements in distributed deep learning/ inference

To support:

- Market growth in consumer electronics (CE) and AI inference
- AI innovation in vertical applications, such as for ultra-responsive apps



Technology



Edge AI refers to AI and machine learning models on devices with limited resources, while far Edge AI is more limited, with heterogenous and volatile resources.

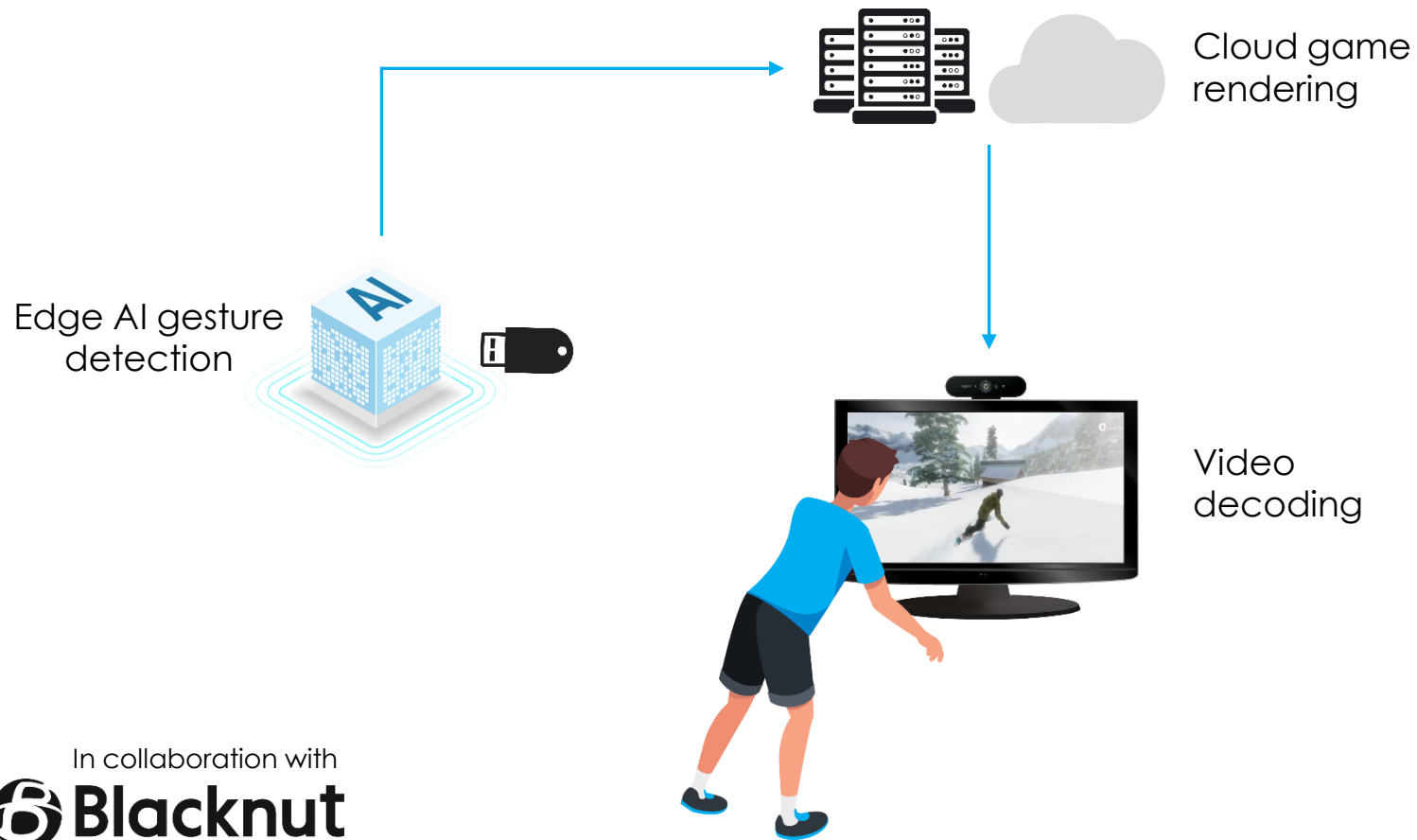
The technology enables **new capabilities**

- ✓ Real-time video analysis on home and local devices
- ✓ Dynamic adaptation to available compute resources
- ✓ Shared AI models managed via in-home AI hub

And delivers **significant benefits**

- ✓ Low latency analyses deliver interactive experiences like speech enhancement, text generation, video stream modification, and user interface
- ✓ Edge and home AI offer solutions to privacy, ensuring audio and video data don't leave the home perimeter
- ✓ Utilizing local available compute resources reduce cloud costs
- ✓ Preserves (uplink) bandwidth

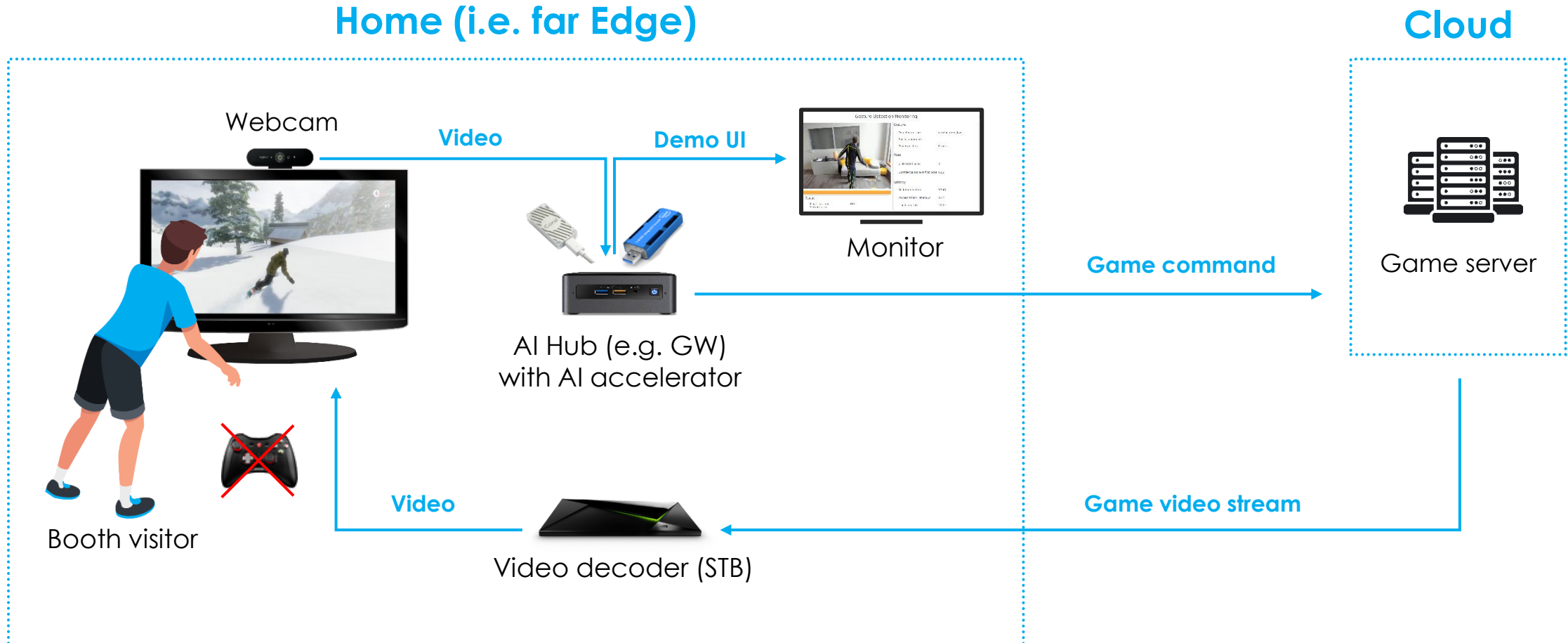
Application of Edge AI to Snowboarding Cloud Game



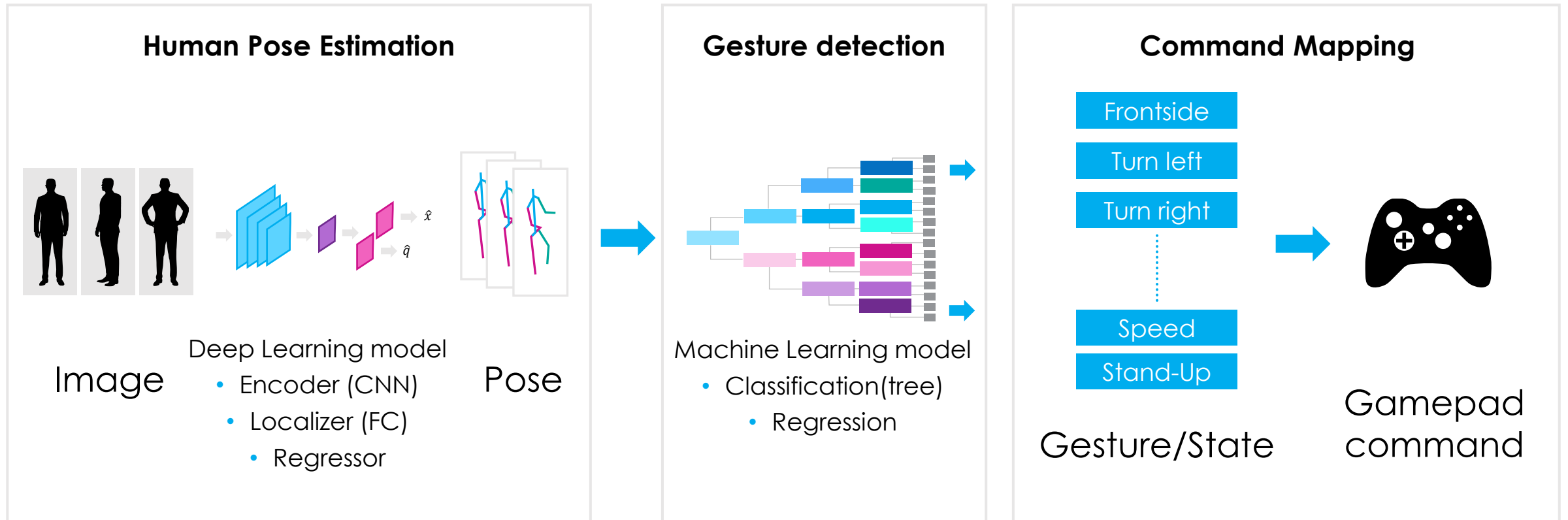
Gesture Detection Monitoring

Gesture	
Detected gesture	carving_right_low
Special command	-
Player position	Regular
Pose	
# detected poses	1
Confidence score of first pose	0.59
Latency	
ML inference time	17.40
Process time (Frame I/O)	1.77
True frame rate	30.04
Status	
Player tracking	YES
Server status	--

Setup of Cloud Gaming with Far Edge AI



AI Pipeline



Performance Tradeoffs of Cloud vs Edge

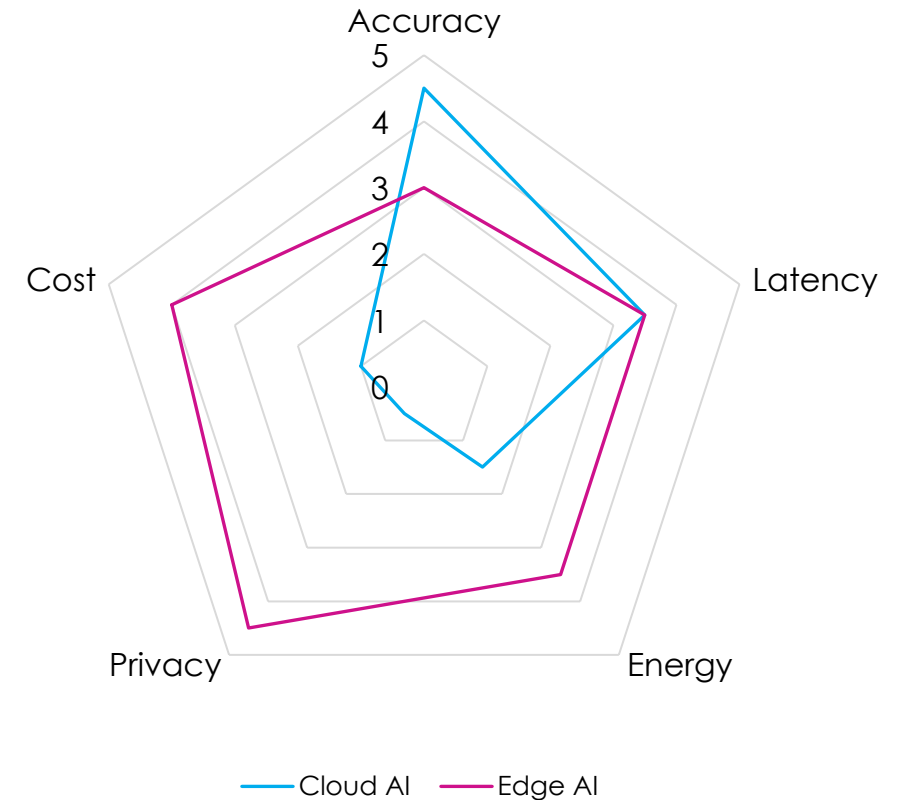
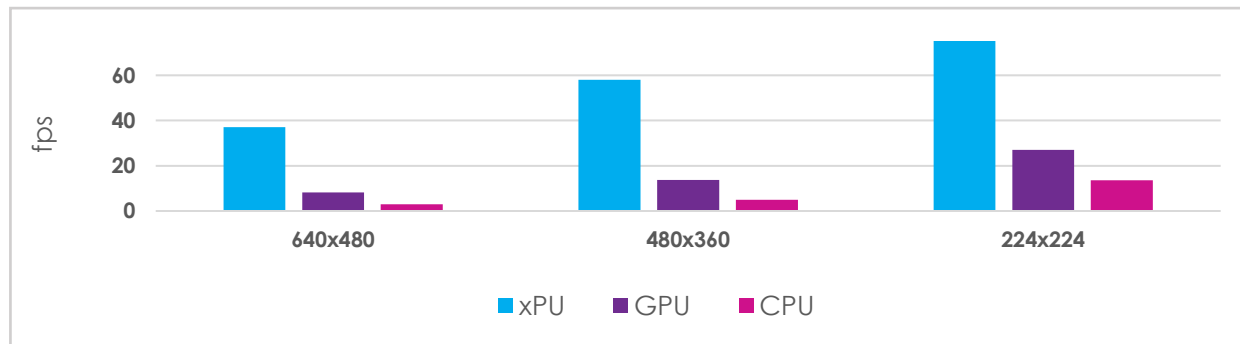


Accuracy | Latency | Energy | Privacy | Cost

Edge AI must be flexible

Accuracy and inference latency are highly dependent on

- Hardware (CPU, GPU, ASIC)
- Input data (resolution) and Model (Architecture, Quantization)
- Framework (RT, Lite)



*Highest value is better