

Edge Computing on the Edge

DeiC Konference 2021

Anders E. Kalør, 4. november 2021
aek@es.aau.dk



AALBORG
UNIVERSITET

Computing flytter sig mod kanten af netværket

Deloitte Insights

By topic ▾ By sector ▾ Spotlight ▾

Article

Gaining an intelligent edge: Edge computing and intelligence could propel tech and telecom growth

TMT Predictions 2021

15 minute read · 07 December 2020

Chris Arkenberg United States
Sanket S. Nesaraj United States
Ariane Bucaille France
Dan Littmann United States

See more

The intelligent edge is poised to propel tech and telecom companies toward the next generation of connectivity and efficiency, driving another wave of industry growth.

Rising from decades of instrumentation, automation, and connectivity, the intelligent edge is maturing into a revolutionary set of capabilities that are already transforming some of the largest technology communications companies on the planet. Although market estimates vary considerably, Deloitte predicts that in 2021, the global market for the intelligent edge will expand to US\$12 billion, continuing its compound annual growth rate (CAGR) of around 35%. Expansion in 2021 will be driven primarily by hyperscale cloud providers optimizing their infrastructure and service offerings. These highly capitalized leaders are establishing the use cases and best practices that may make it easier for companies across multiple industries to attain the capabilities of the intelligent edge. By 2023, 70% of enterprises may likely run some amount of processing at the edge. As one leading graphics processing unit (GPU) manufacturer has stated, "We're about to enter a phase where we're going to create an internet that is thousands of times bigger than the internet that we enjoy today."⁴

Though challenges and headwinds exist, we believe that the intelligent edge is poised to transform the computing landscape, propelling the world's largest technology companies toward the next generation of connectivity and

Learn more
View TMT Predictions

ericsson.com Products and solutions ▾ Discover ▾ Transforming enterprises ▾ Future technologies ▾ About us ▾

5G and cloud: How telecom can architect the next cloud era

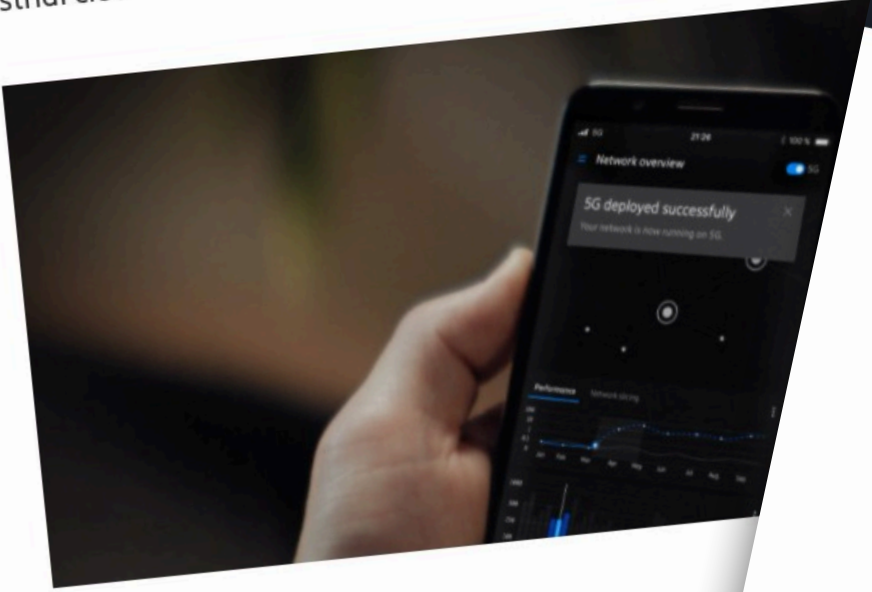
HOME > ERICSSON BLOG > 5G AND CLOUD: HOW TELECOM CAN ARCHITECT THE NEXT CLOUD ERA

5G and the massive increase of compute power in the network will transform the telecom industry like never before. Service providers, offering the most value add in emerging edge ecosystems, stand to gain most from lucrative new pathways into 5G cloud revenue streams. However, securing tomorrow's market means making inroads already today. Below, I explain how service providers can gain an edge in the race to the industrial cloud and become key architects of the next cloud era.

FEB 04, 2021 | 4 min.

Erik Ekudden
Group CTO and Head of Technology & Strategy

CATEGORY
Cloud and virtualization 5G



Moving the compute power from a device to the network edge gives an indication of what is to come for next generation interfaces. This breakthrough, which will ultimately enable 5G use cases, is one of the many use cases we are developing in our innovation lab in Santa Clara, California, with partners Qualcomm and Intel. However, this is the first step. To make such demanding use cases scale worldwide, we need to work to increase the interoperability of 5G and cloud infrastructure stacks.

Evolving to the high performance network edge

We call this the **network compute fabric**, and service providers will play a critical role in enabling and orchestrating this new computing paradigm, one which will have a substantial impact on digitalization of the world, like **Industry 4.0**.

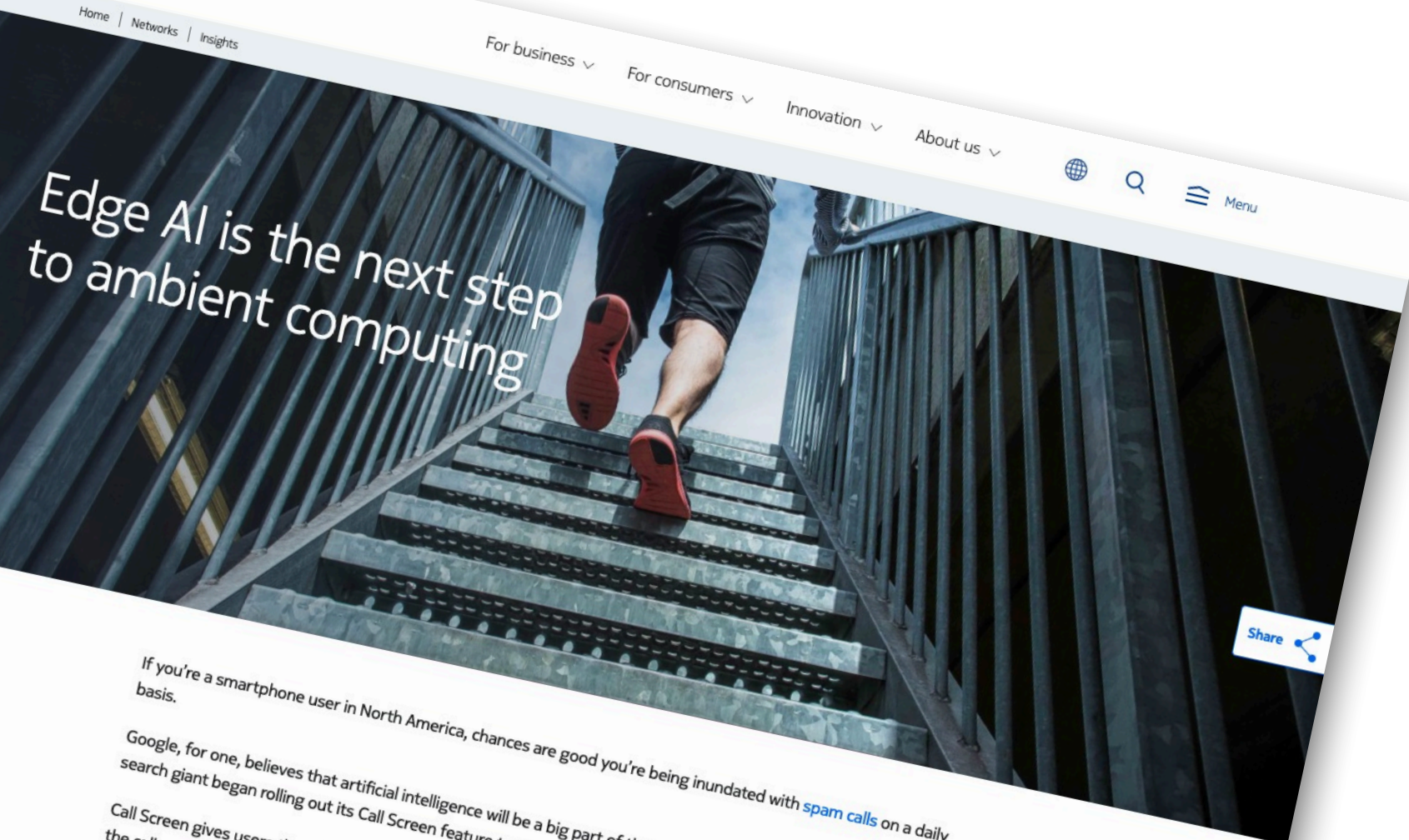
In this computing paradigm, connectivity, compute and execution will form a single unified, integrated execution environment for distributed applications –

NOKIA

Home | Networks | Insights

For business ▾ For consumers ▾ Innovation ▾ About us ▾

Edge AI is the next step to ambient computing



If you're a smartphone user in North America, chances are good you're being inundated with spam calls on a daily basis.

Google, for one, believes that artificial intelligence will be a big part of the solution to this problem, which is why the search giant began rolling out its Call Screen feature to Pixel devices in 2018.

Call Screen gives users the option to intercept incoming calls with their own robot assistant. The AI answers and asks the caller questions, then relays answers to the user through real-time text transcription. The user can request further information, reject the call, send it to voicemail or accept it.

It's not a be-all and end-all solution, but the company believes it's a step in the right direction.

"We're trying to think of using technology in a way where it helps us, similar to if we had another person helping us," says Yossi Matias, vice-president of engineering at Google. "Hopefully we're able to use technology to get control back over some of our incoming calls."

Call Screen, which Google upgraded in December to automatically answer suspected spam calls without even bothering the user, doesn't rely on traditional artificial intelligence. Rather, it's made possible by edge AI – advanced language processing capability resides on the smartphone itself, rather than on a server in the cloud.

The phone itself thus processes the call without having to access a server, speeding the process to the point of near instantaneousness.

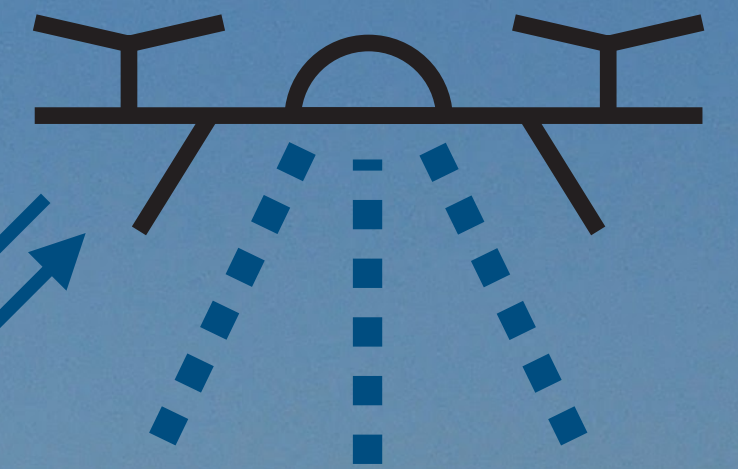
The cost, the profits, and the benefits

Om mig

- Ph.d.-studerende ved Aalborg Universitet (Connectivity-sektionen)
- Forsker i fremtidens (trådløse) kommunikationsteknologier:
 - Teknologier til at understøtte Internet of Things
 - Integration mellem kommunikation og computing
 - Intelligent kommunikation ved brug af AI/ML



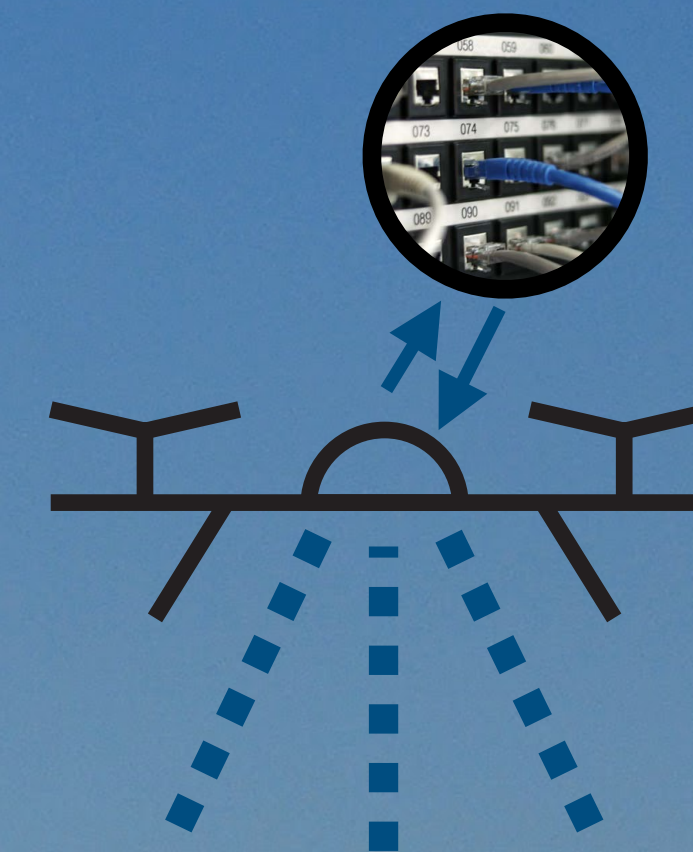
Traditionelt er skyen placeret centralt, langt fra enhederne



I edge computing flyttes computing tættere på enheden



Helt op til enheden



Hybrid



Hvorfor edge computing?

- Høj **latenstid** mellem enheden og skyen
- Bedre **skalerbarhed** når antallet af enheder og datamængden stiger
- Potentielt højere **pålidelighed**
- Mere **privatliv** til enheden



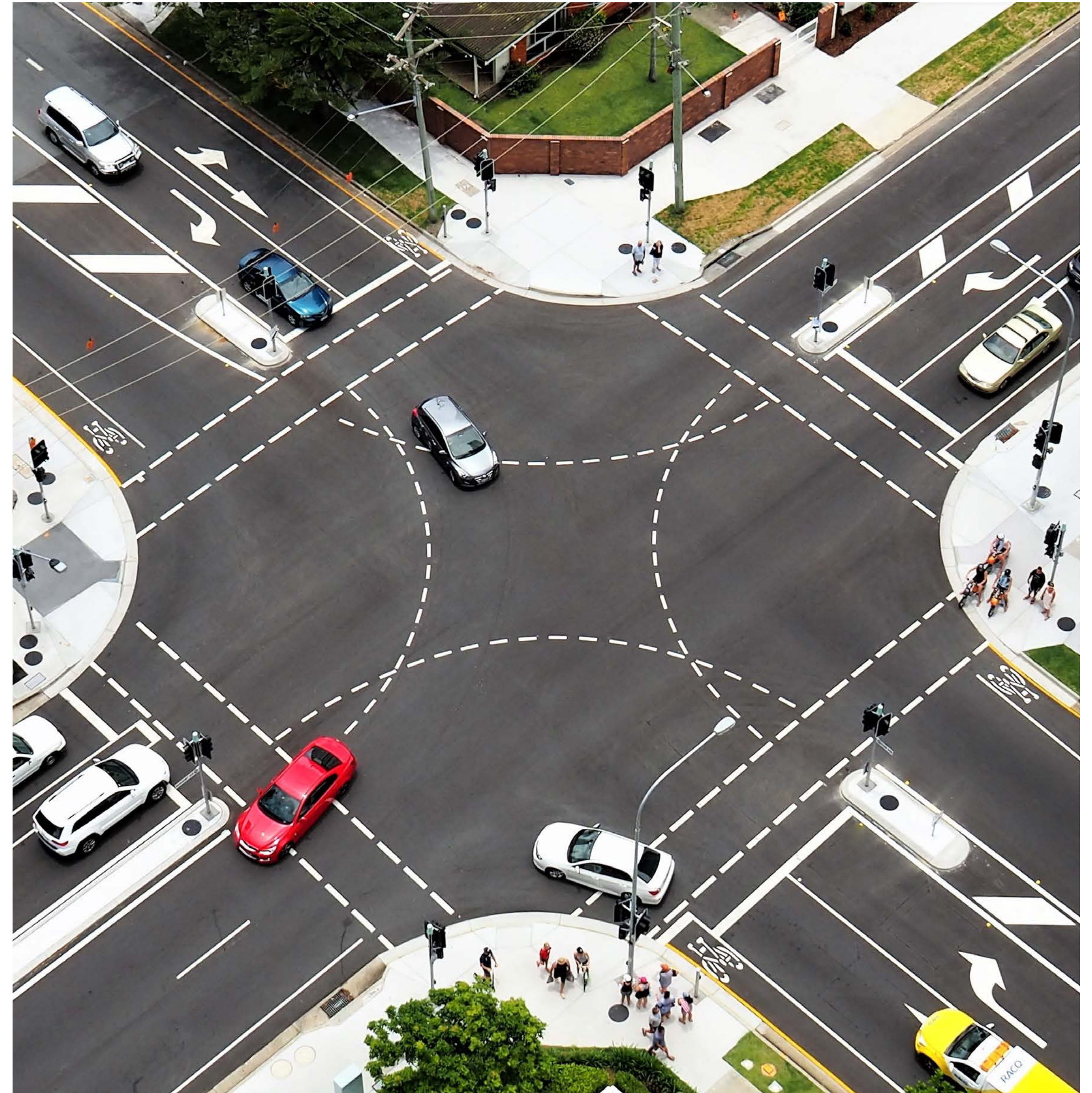
Latenstid

- At minimere latenstiden har været et centralt fokus i 5G (og er i 6G)
- Forbindelsen til skyen udgør i stigende grad en flaskehals
- Latenstiden kan minimeres ved at placere "en sky" hos teleudbyderen, eller helt ude ved mobilmasten



Skalerbarhed, pålidelighed

- Flere ting bliver forbundet til Internettet og afhænger af AI/ML/lign.
- Mængden af data stiger
- Maskering af cloud (u)pålidelighed



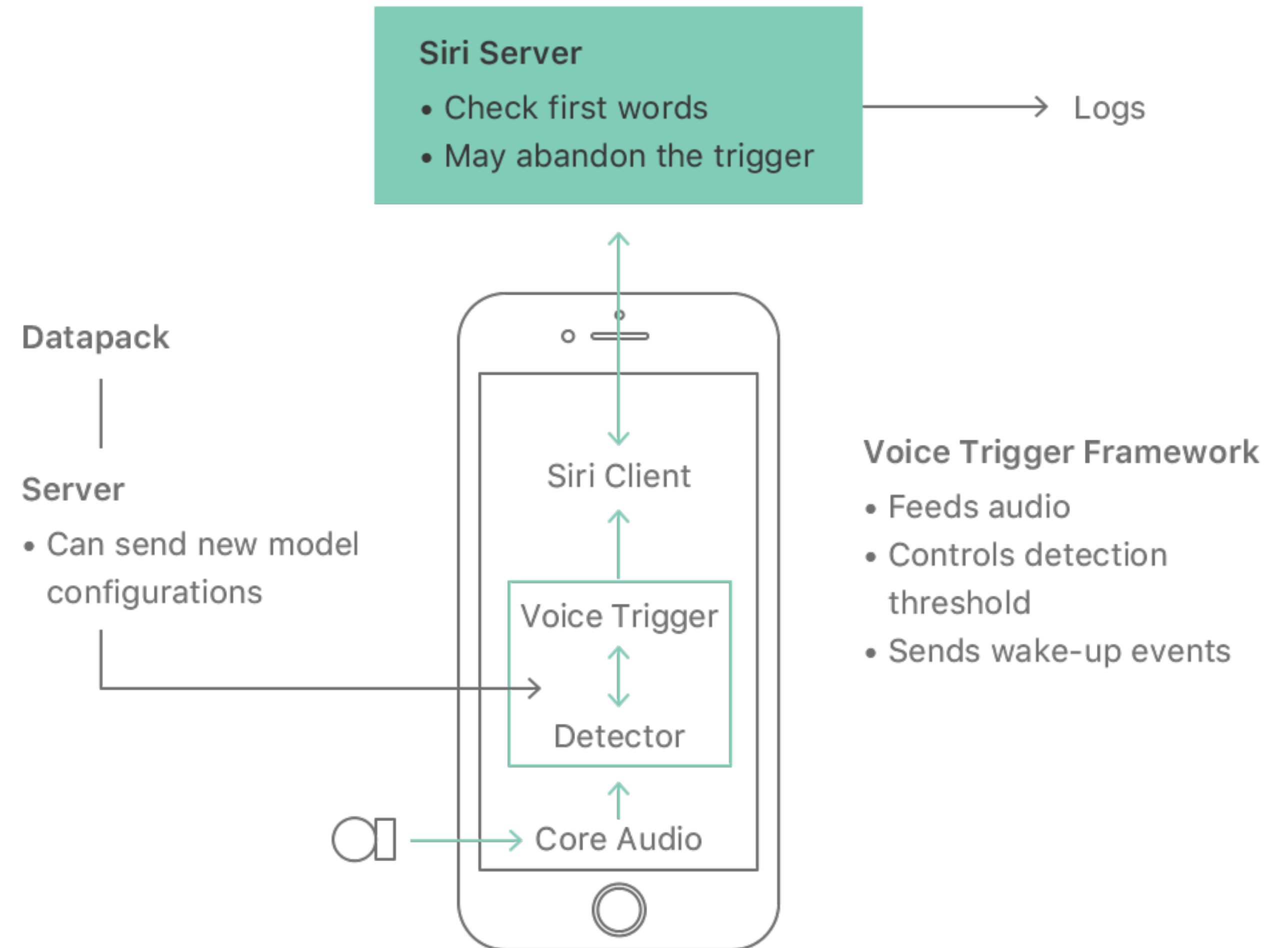
Privatliv

- IoT er blevet en del af vores hverdag
- Vi ønsker ikke at billeder, tale, osv. skal sendes til skyen

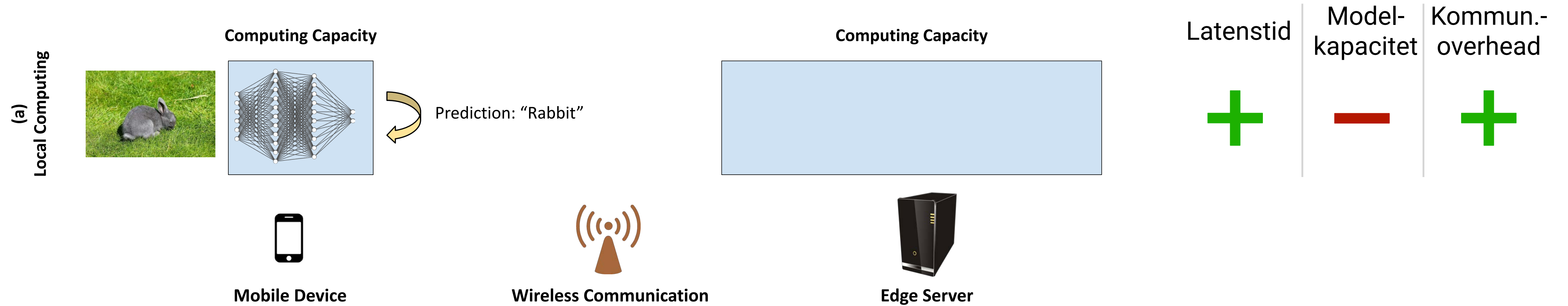


Apple's Siri

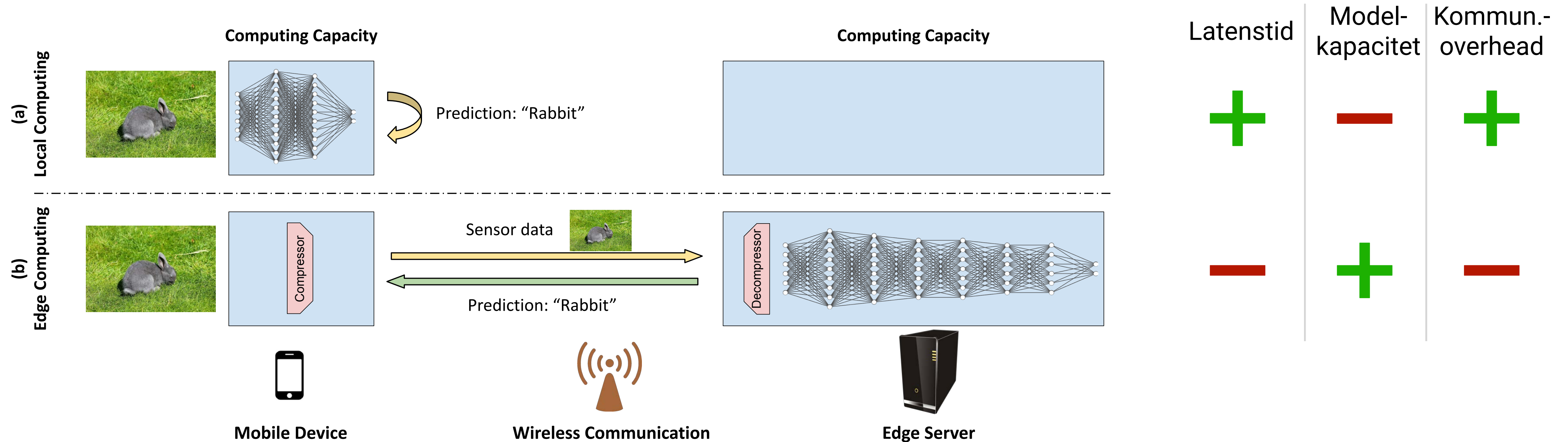
- Lytter konstant efter "Hej Siri"
- Kun hvis det detekteres, sendes lydoptagelsen til en server i skyen
- Serveren kører optagelsen igennem en bedre model
- Lydoptagelserne i skyen bruges til træning



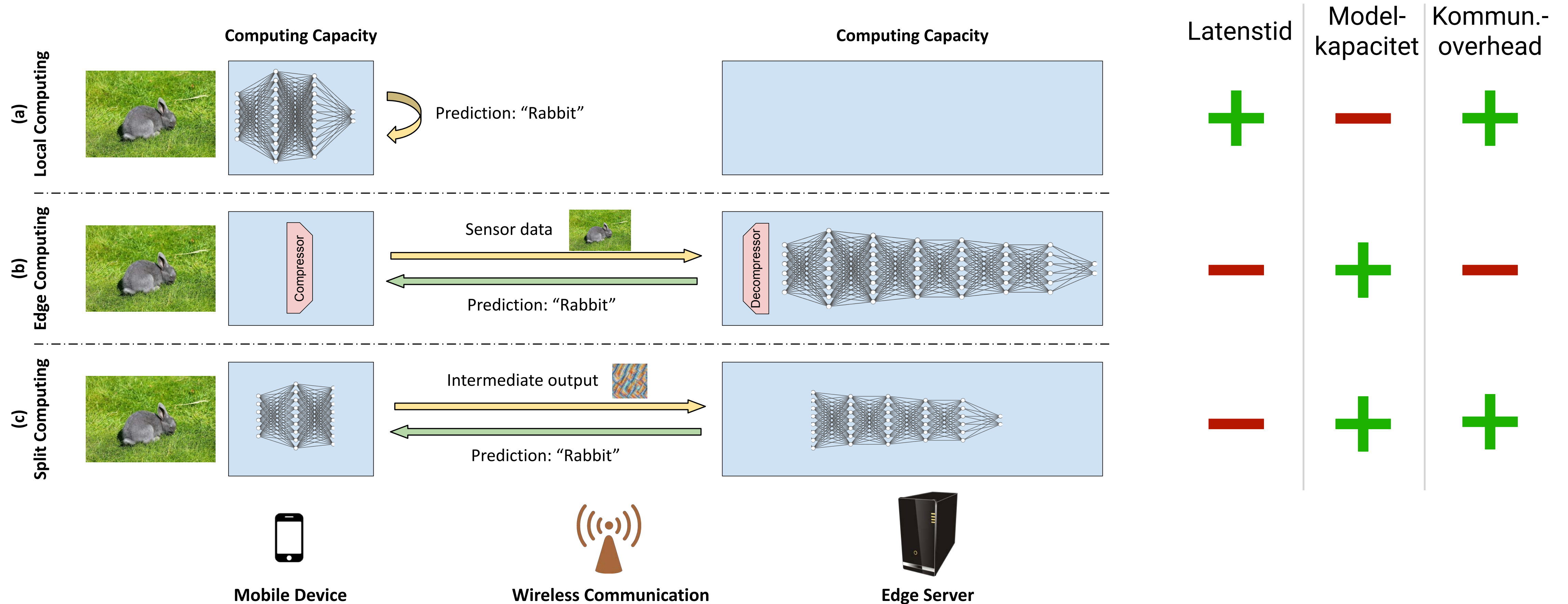
Fordeling mellem enhed og server



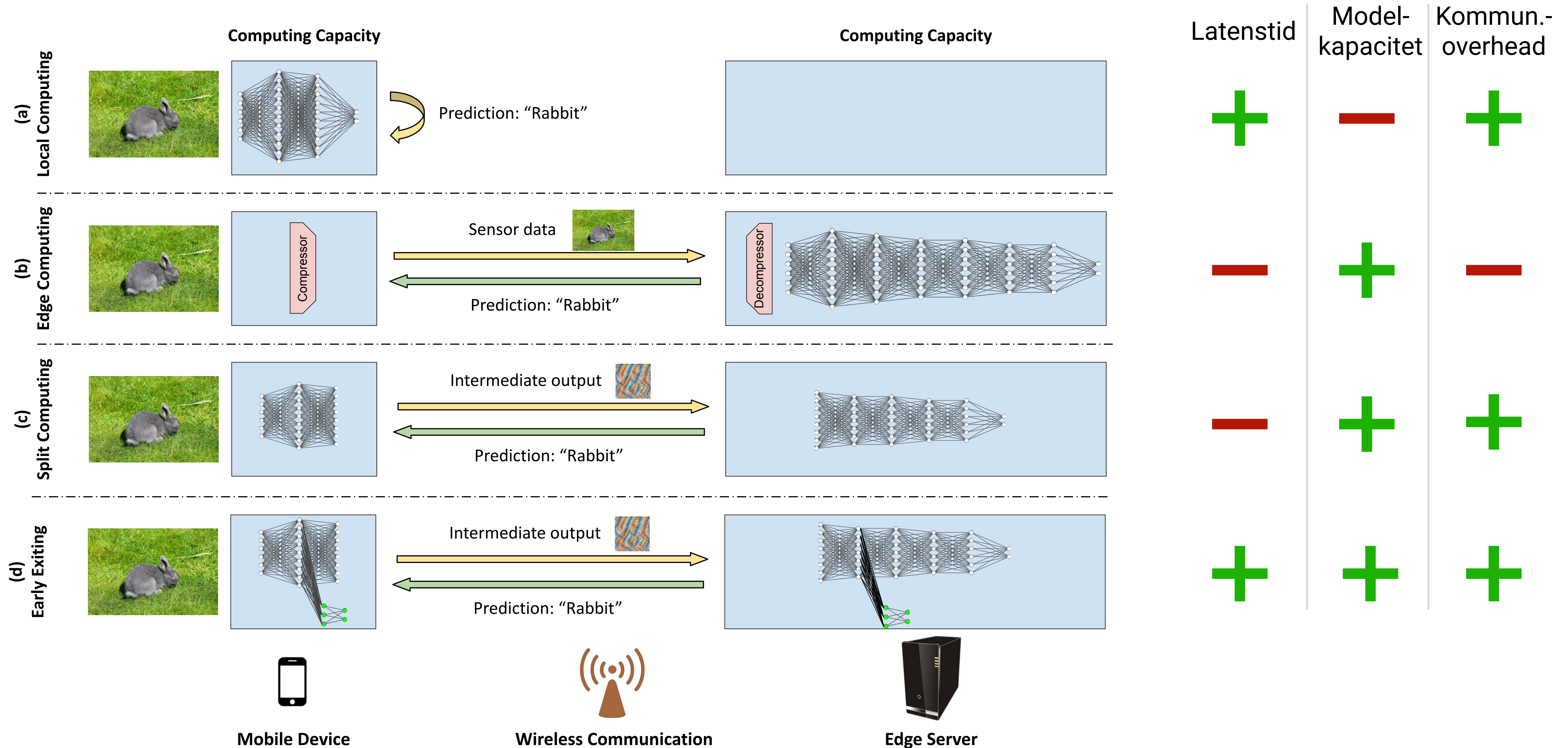
Fordeling mellem enhed og server



Fordeling mellem enhed og server

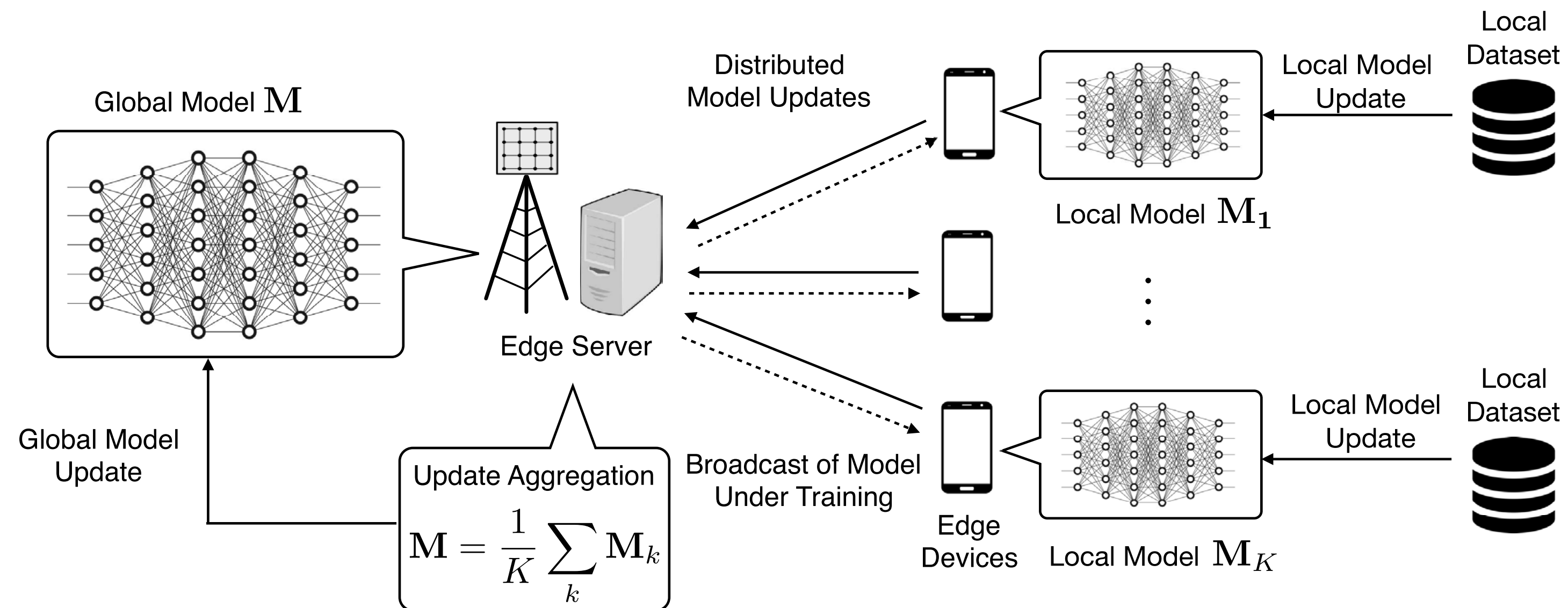


Fordeling mellem enhed og server



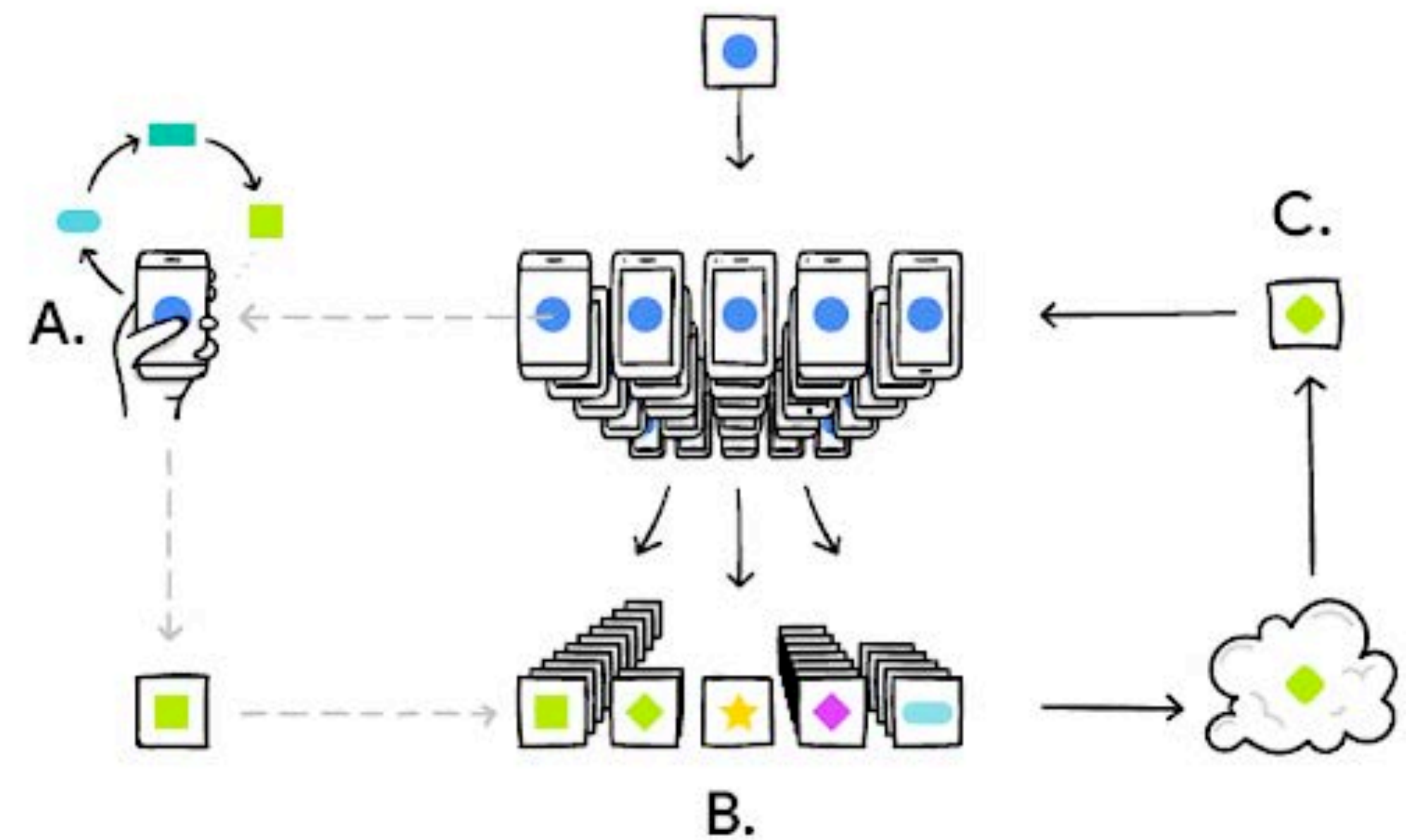
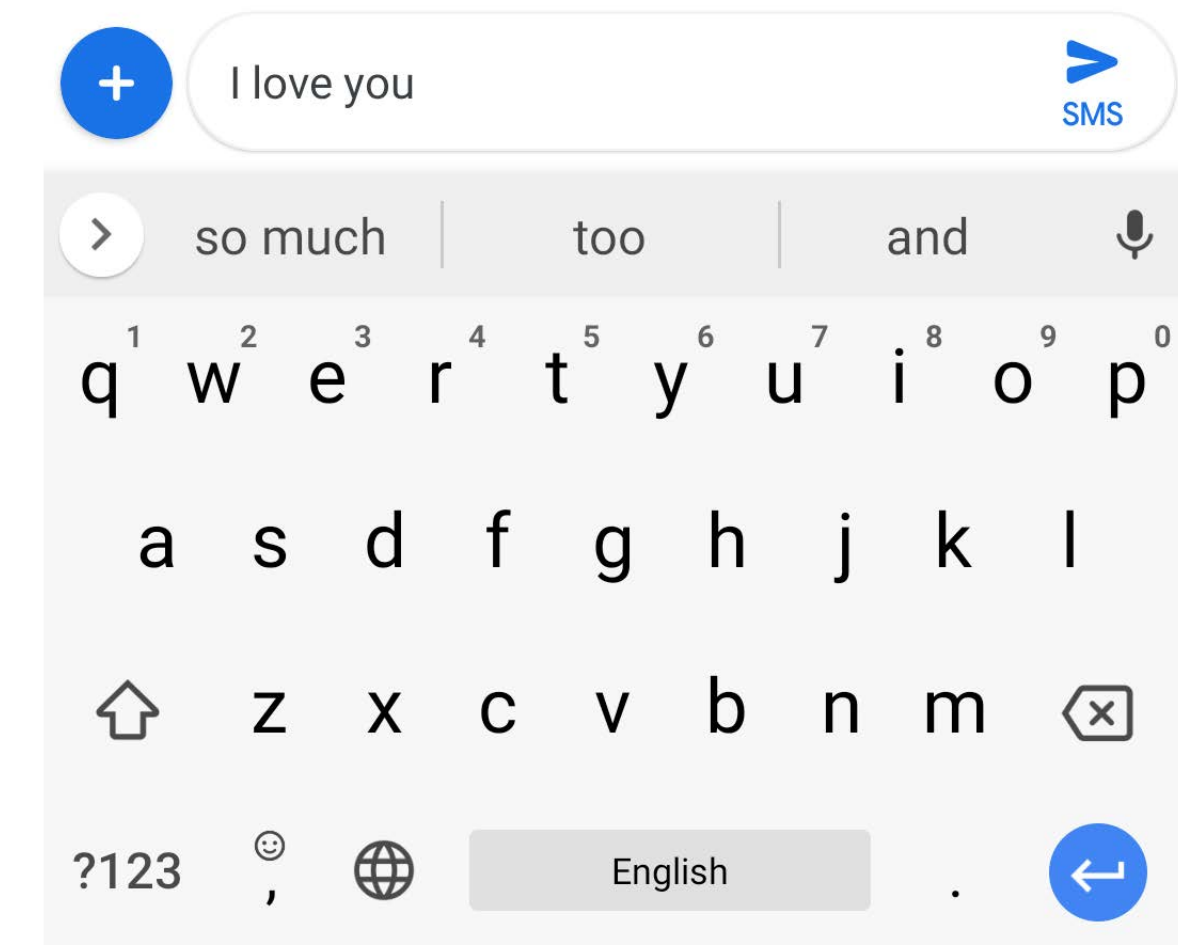
Federated learning

Hvis dataen er hos enhederne, hvordan skal modellerne så trænes?



Google's Gboard tastatur

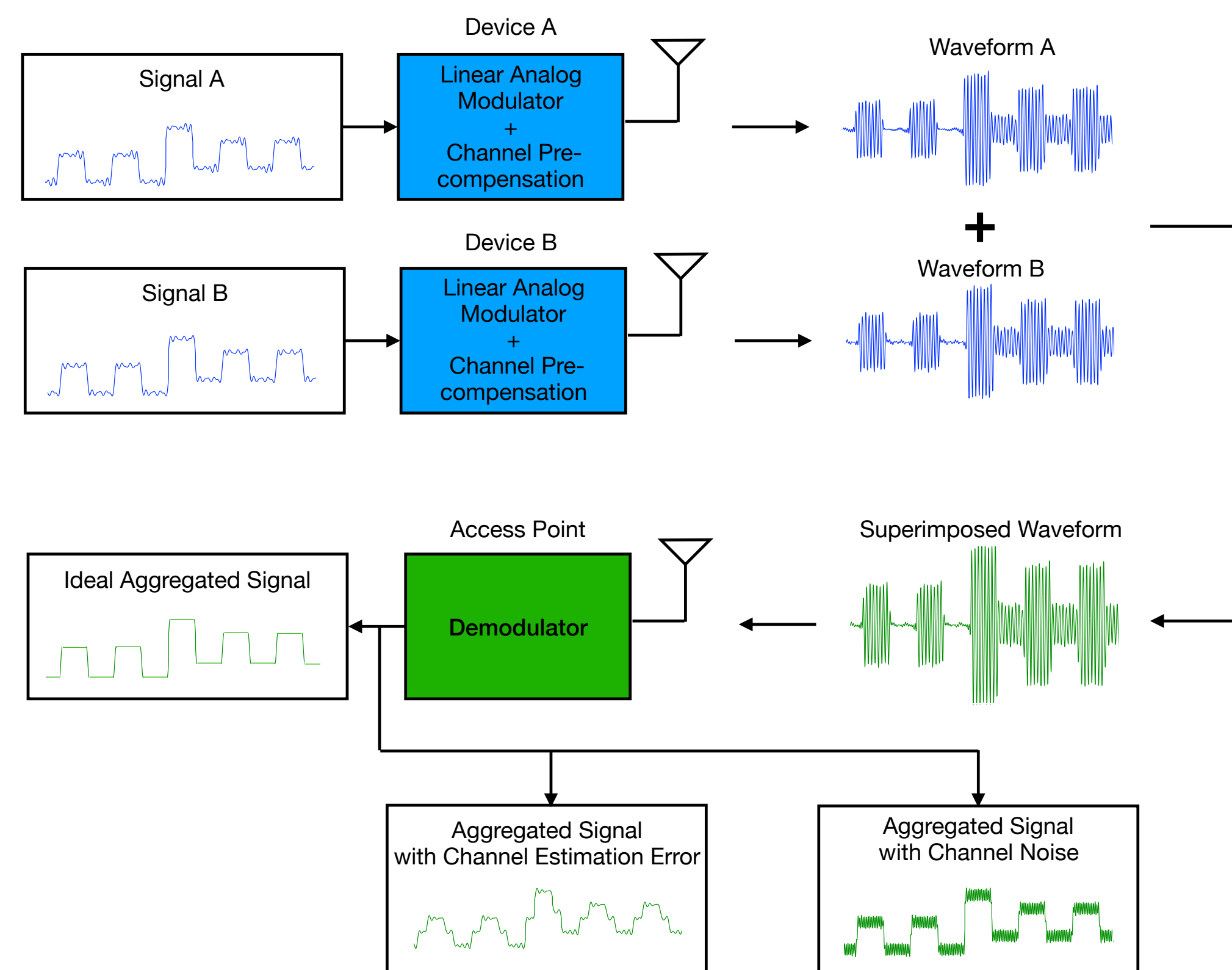
- A. Personalisering
- B. Aggregering af modeller (gradienter)
- C. Modelopdatering



Integration af kommunikation og edge computing

Federated learning over kommunikationskanal

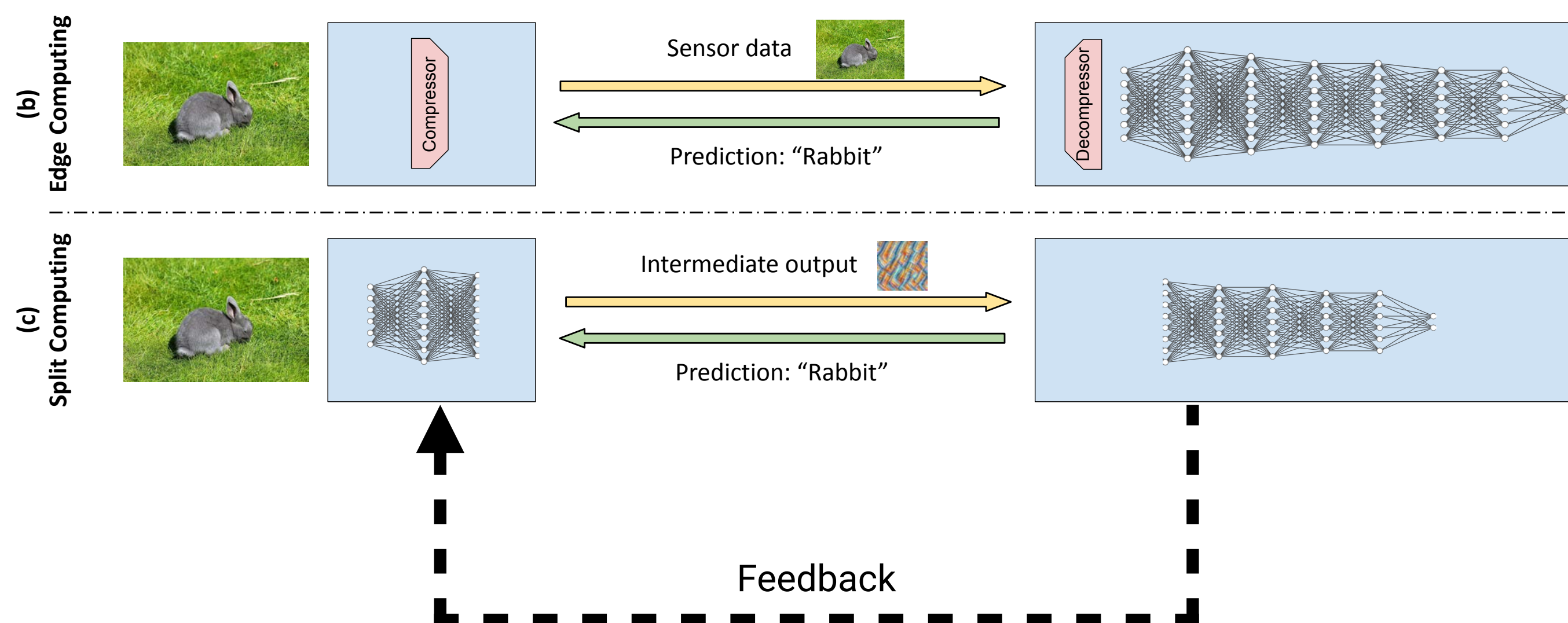
- Kan vi udnytte det trådløse medie til effektivt at aggregere modeller?
- I teorien meget effektivt
- Kræver nøjagtig synkronisering



Integration af kommunikation og edge computing

Statistisk inferens over kommunikationskanal

- Traditionelt er kommunikationssystemer designet som en "bit-pibe"
- Hvad hvis vi ikke ved, hvad modtageren skal bruge, for at kunne dekode billedet?
- Hvad hvis vi har mulighed for feedback?



Opsummering

- Der er mange **fordele** ved at flytte skyen nærmere enheden
- Hvordan laver vi **opgavefordeling mellem enheden og skyen/skyerne?**
- Udviklingen er i **fuld gang**, særligt inden for ML/AI
- Integrationen mellem edge computing og kommunikation åbner **nye muligheder**

