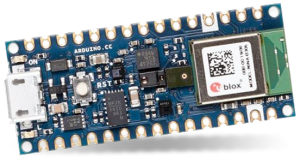
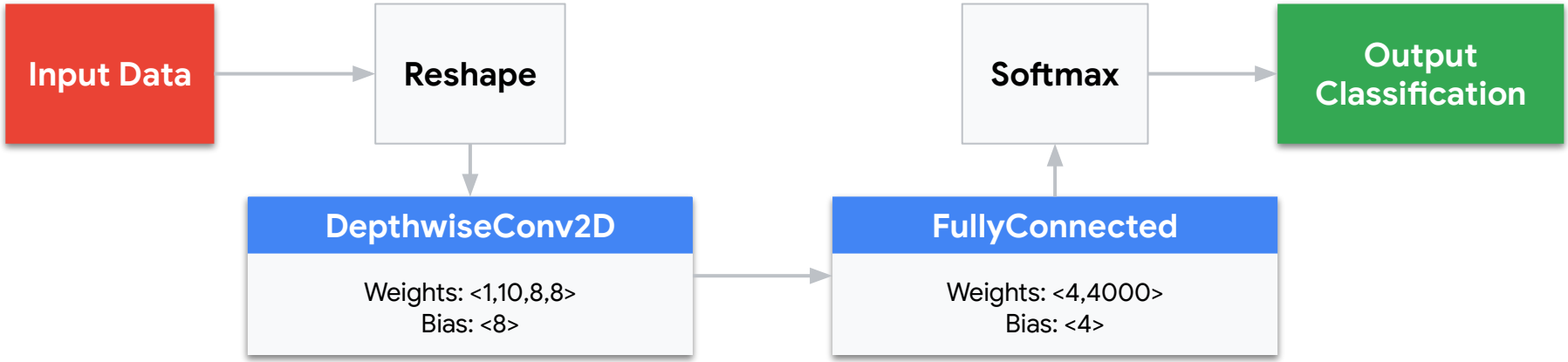
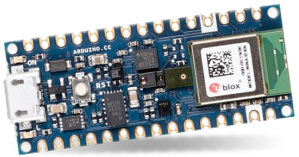
Cascade Architectures

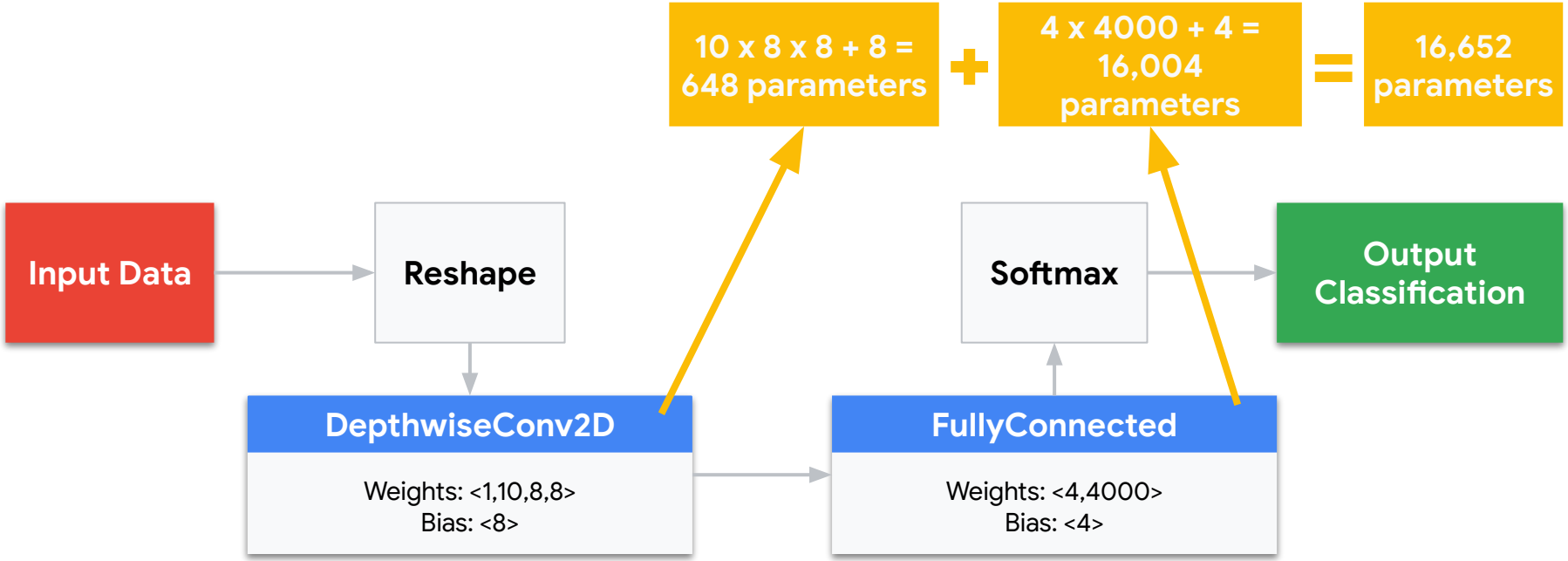


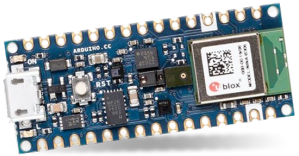
Our board [Course 3 Kit] only has **256KB** of RAM (memory)





Our board [Course 3 Kit] only has **256KB** of RAM (memory)

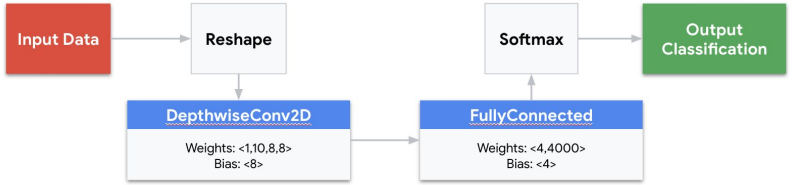




Our board [Course 3 Kit] only has **256KB** of RAM (memory)

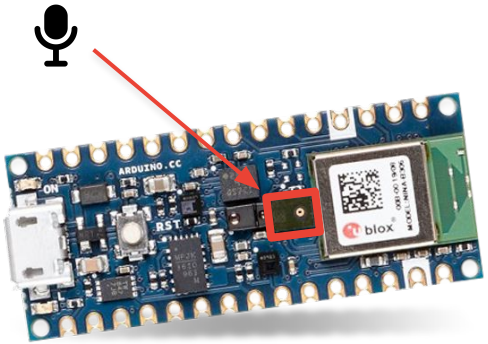
Trade-offs

- Limited **vocabulary**
- Limited **accuracy**
- Limited **user experience**

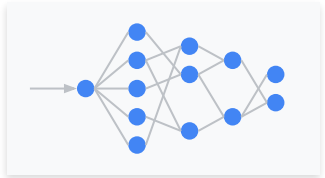


“Cascade” detection: a multi-stage model

1 Continuously listen on the microcontroller

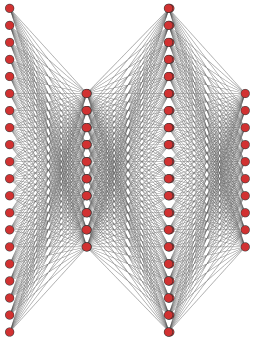


2 Process the data with **TinyML** at the edge



3 Send the data to the cloud when triggered

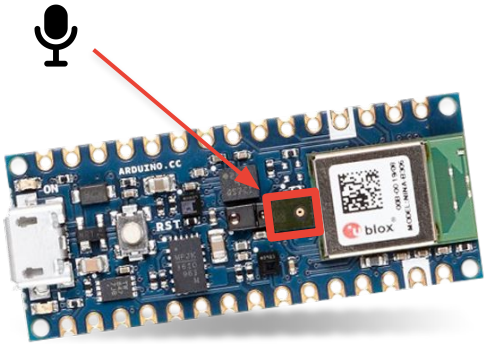
4 Process the full speech data with a large model in the cloud

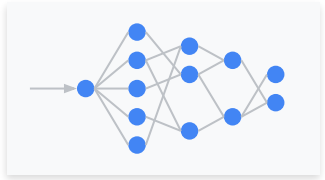


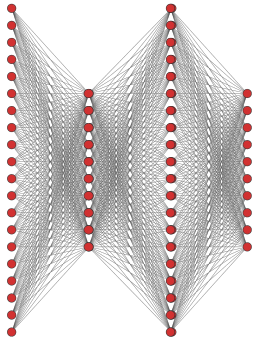
“Cascade” detection: a multi-stage model

4 Process the full speech data with a large model in the cloud

2 Process the data with **TinyML** at the edge





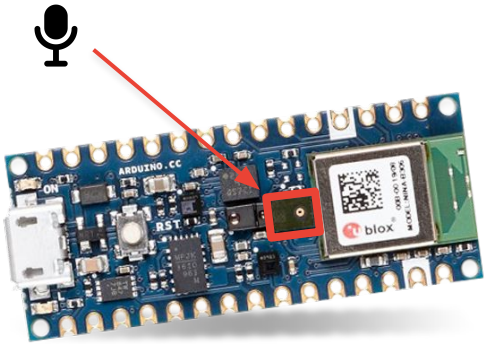


3 Send the data to the cloud when triggered

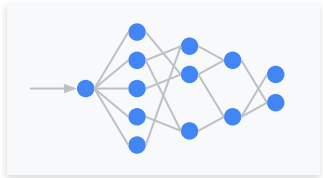
1 Continuously listen on the microcontroller

“Cascade” detection: a multi-stage model

1 Continuously listen on the microcontroller



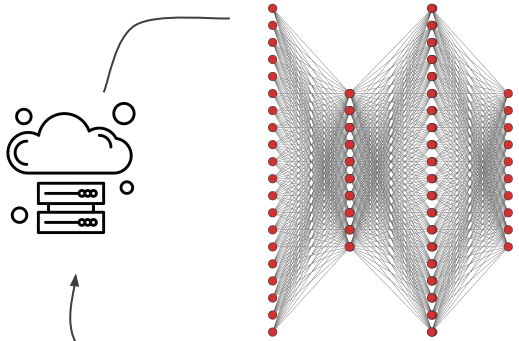
2 Process the data with **TinyML** at the edge



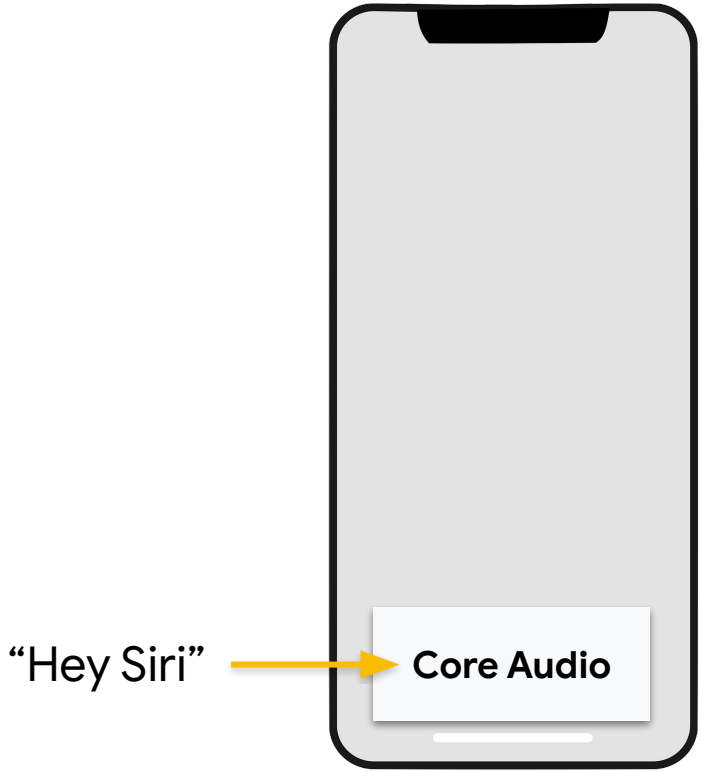
3 Process on a secondary larger model on a larger local device

4 Send the data to the cloud when triggered

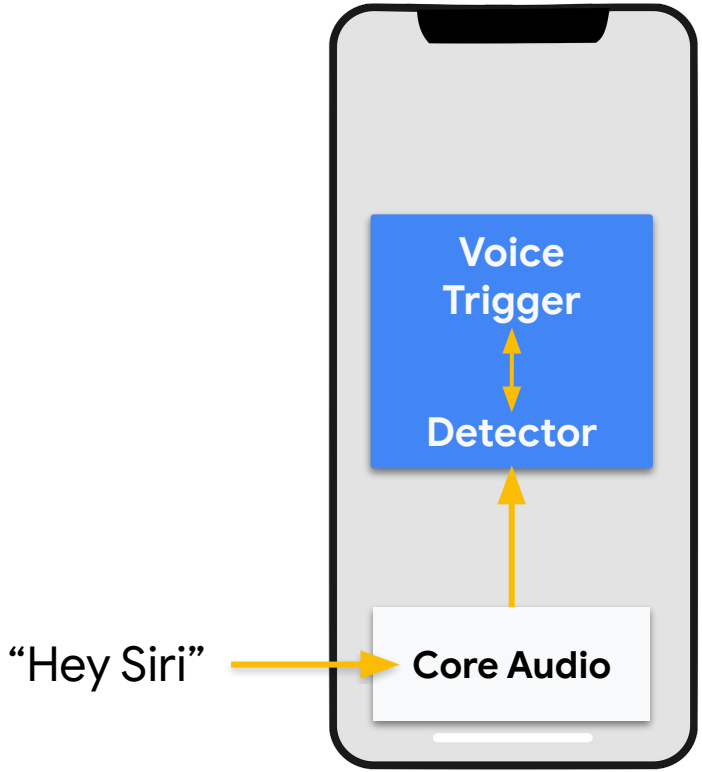
5 Process the full speech data with a large model in the cloud



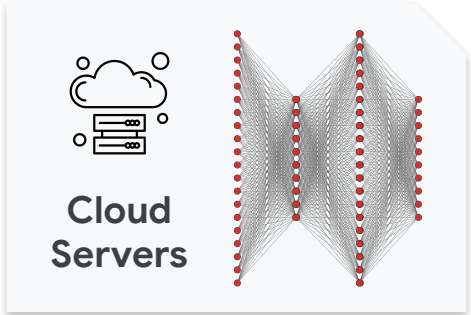
“Hey Siri”

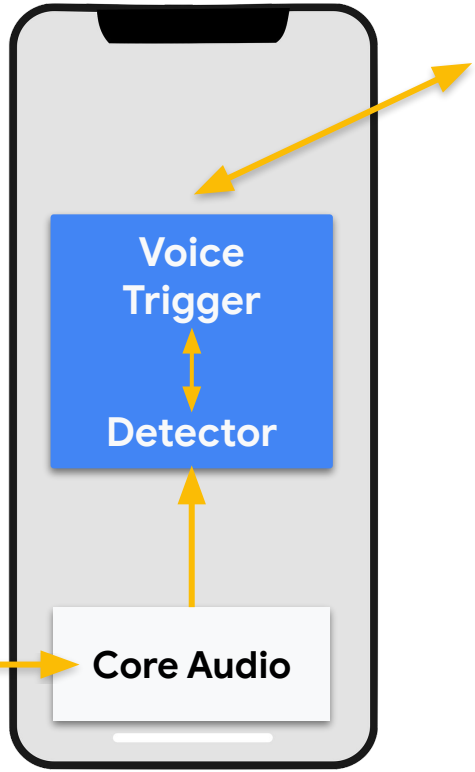


“Hey Siri”



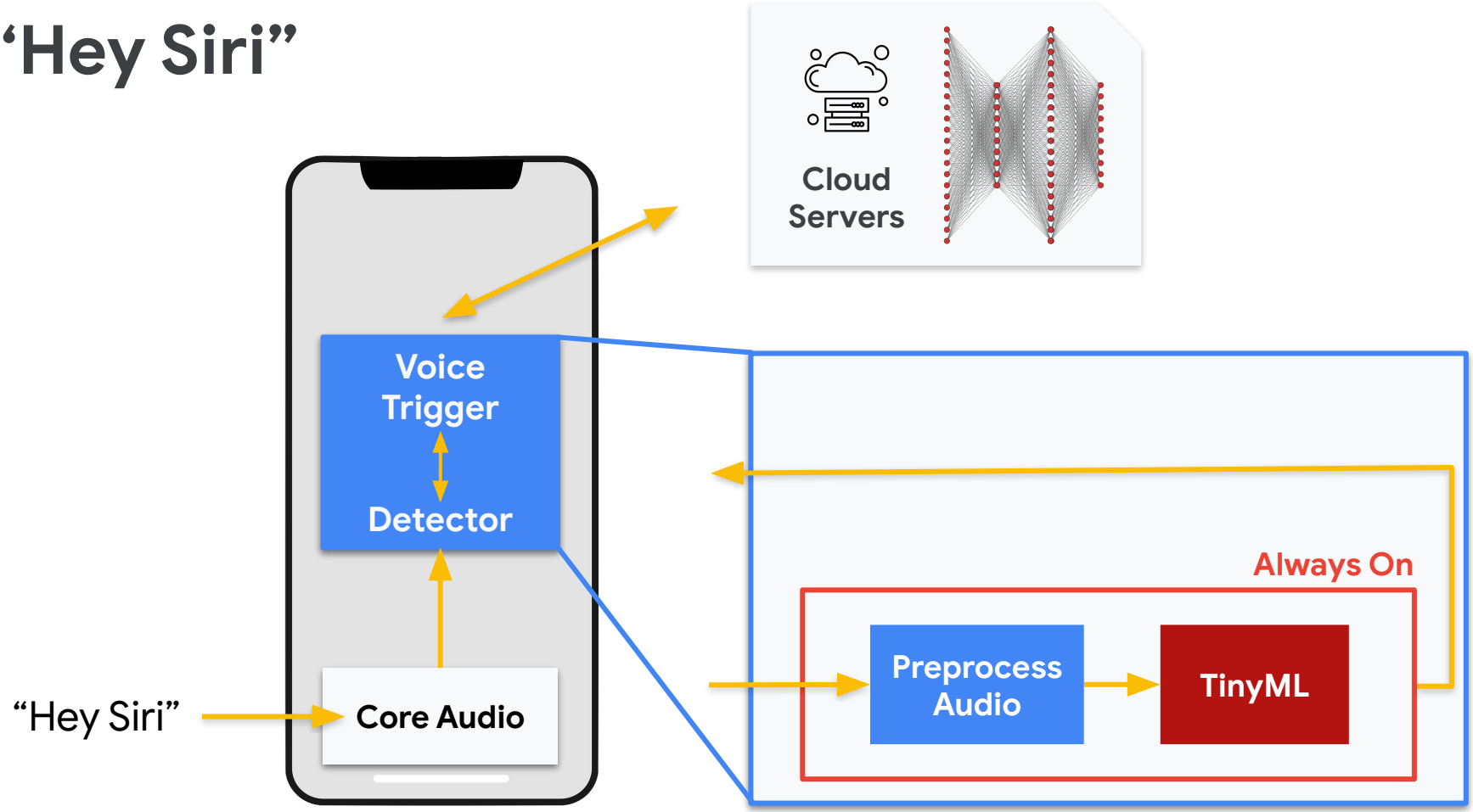
“Hey Siri”



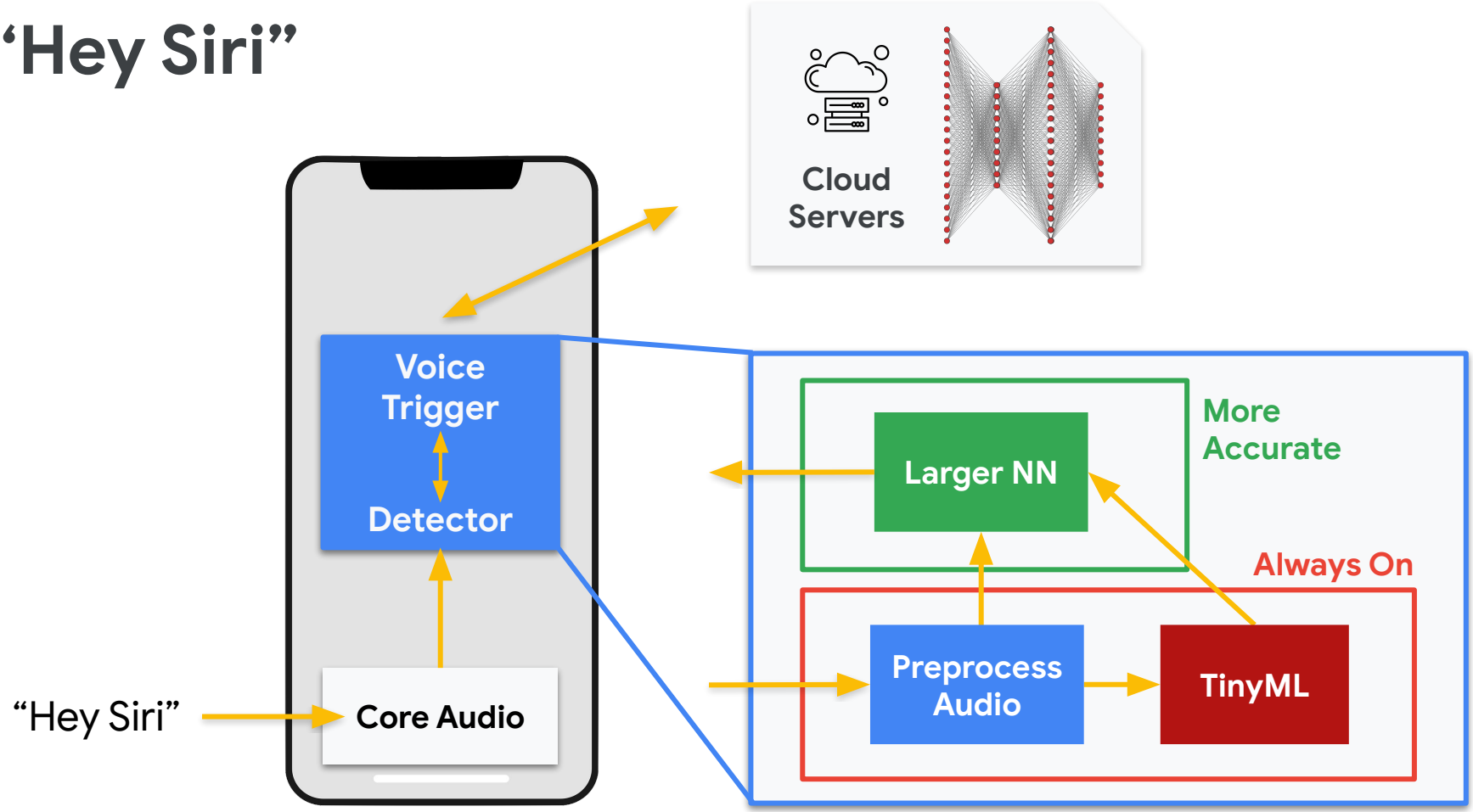


“Hey Siri”

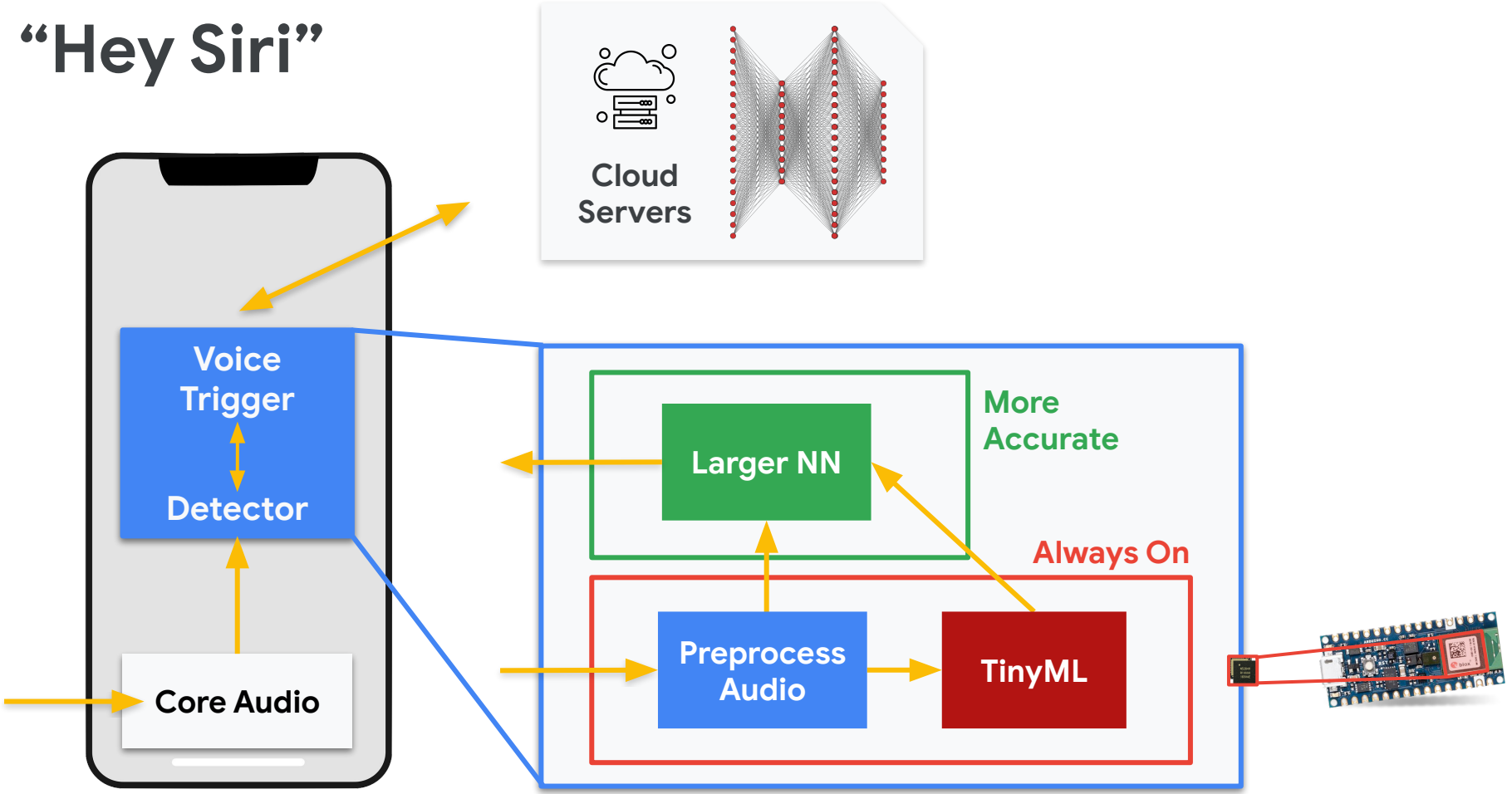
“Hey Siri”


Cloud
Servers
Voice
Trigger
Detector
Always On
Preprocess
Audio
TinyML
“Hey Siri”
Core Audio

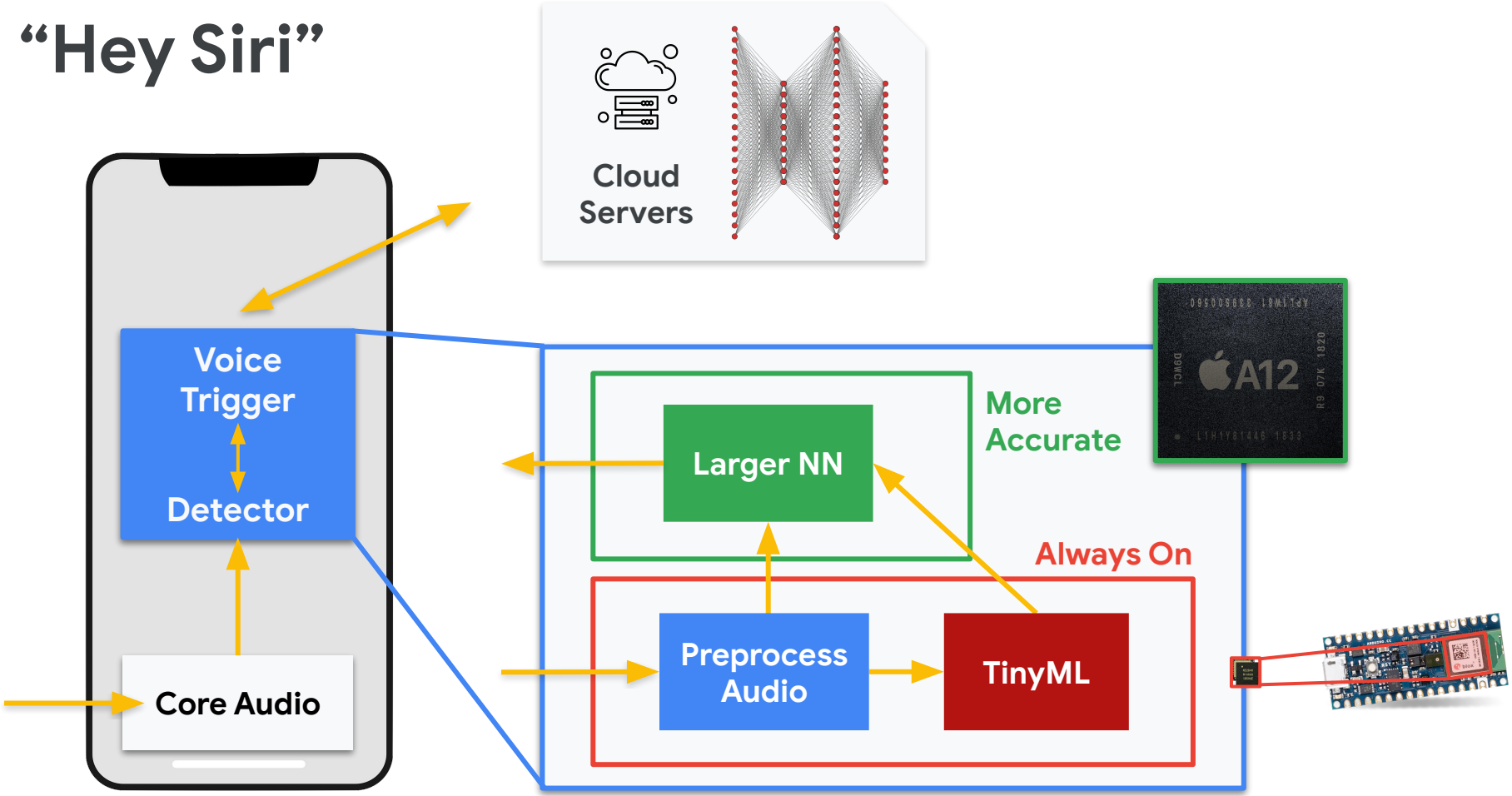
“Hey Siri”


Cloud Servers
Voice Trigger
Detector
Core Audio
“Hey Siri”
Larger NN
More Accurate
Always On
Preprocess Audio
TinyML

“Hey Siri”

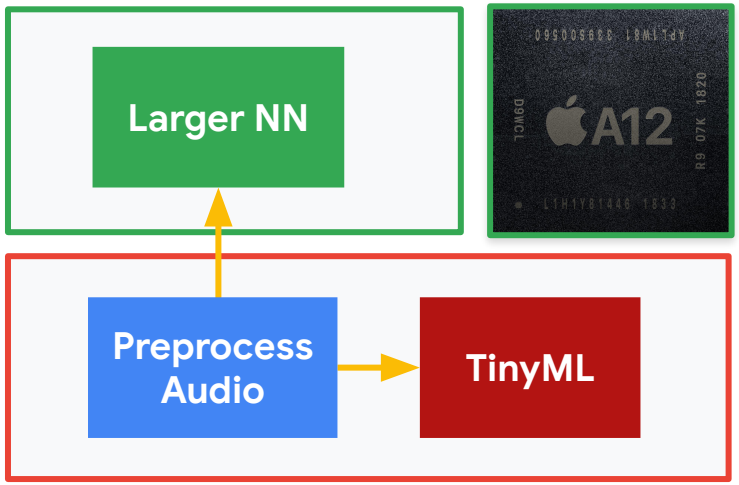

Cloud
Servers
Voice
Trigger
Detector
Core Audio
Larger NN
More
Accurate
Always On
Preprocess
Audio
TinyML

“Hey Siri”


Cloud
Servers
Voice
Trigger
Detector
Core Audio
Larger NN
More
Accurate
Always On
Preprocess
Audio
TinyML
A12

“Hey Siri”

More Accurate



Always On