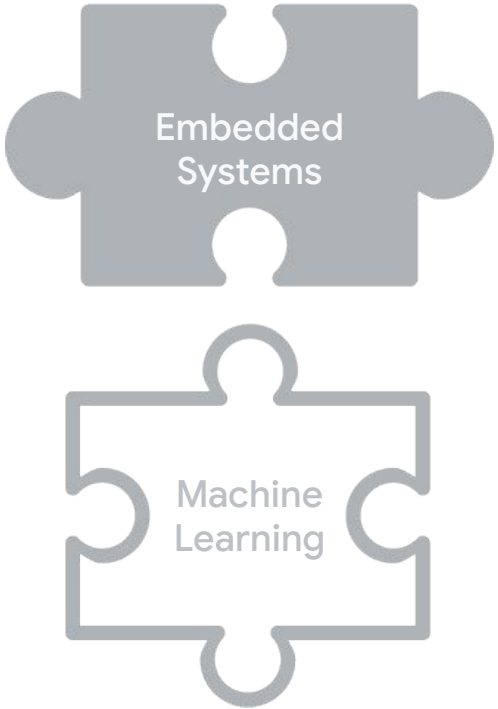
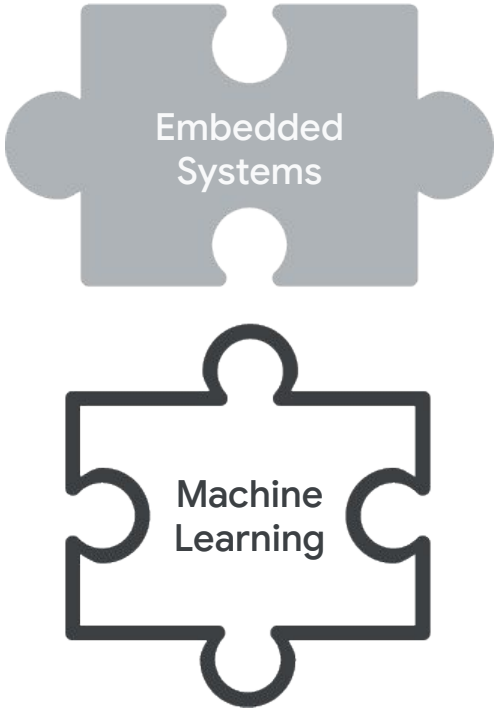
# What are the Challenges for TinyML?

Part C

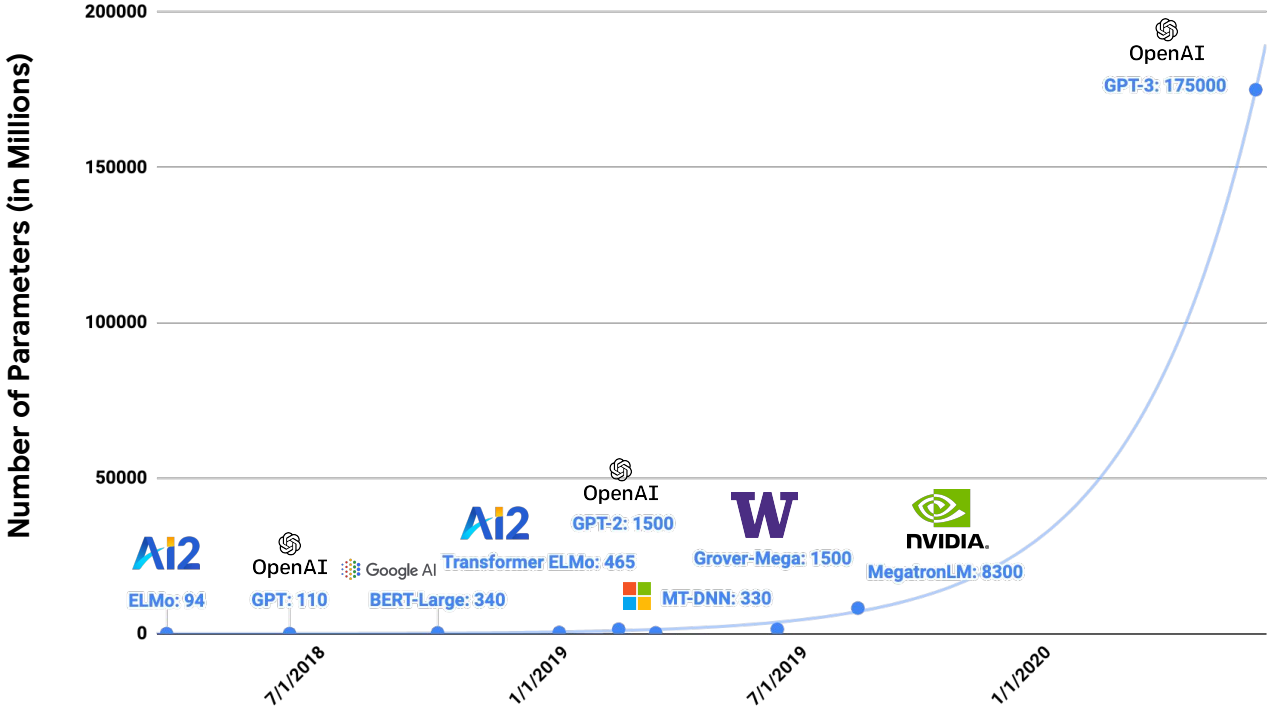



# TinyML

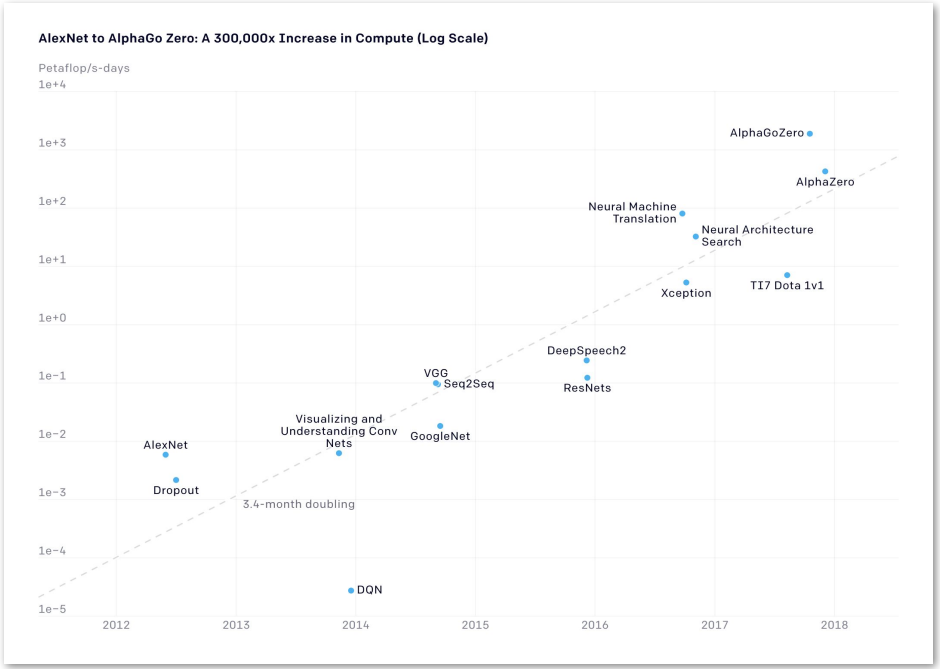



TinyML

# ML Model Size Growth



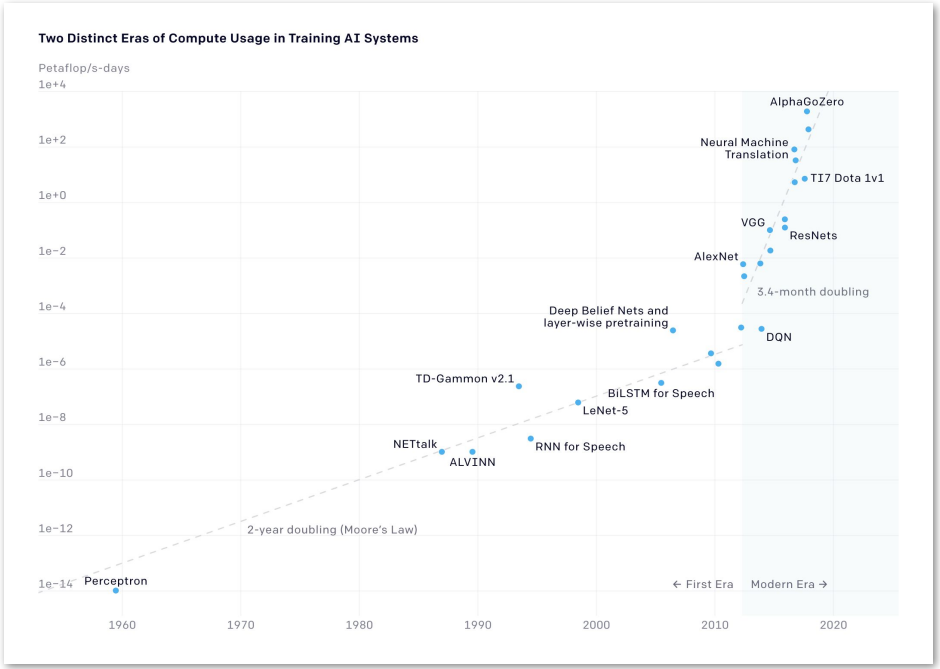
# ML Compute Needs (2012 to Present Day)




Source: https://openai.com

In recent years, **computing needs grew by 300,000x** to train the machine learning models that are widely deployed in the industry

# ML Compute Needs (from the 1960s)




Source: https://openai.com

**In recent years,** the amount of computing needed has grown remarkably fast.

Compute requirements are **doubling nearly every 3 to 4 months**

# Cloud TPU

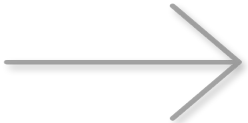
Source: Google

Source: Google

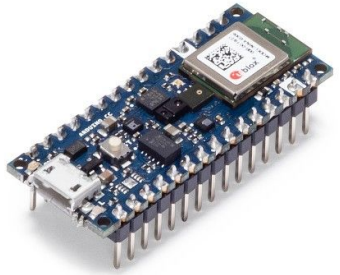
Source: Google

Source: Google

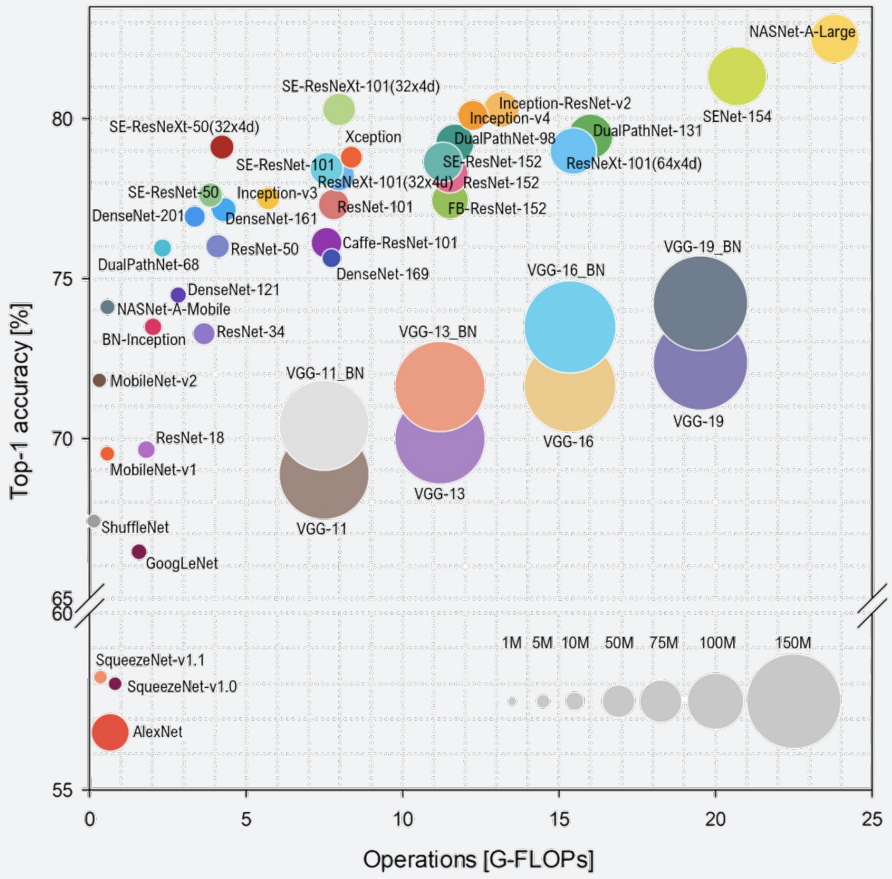
Cloud TPU



**TinyML**



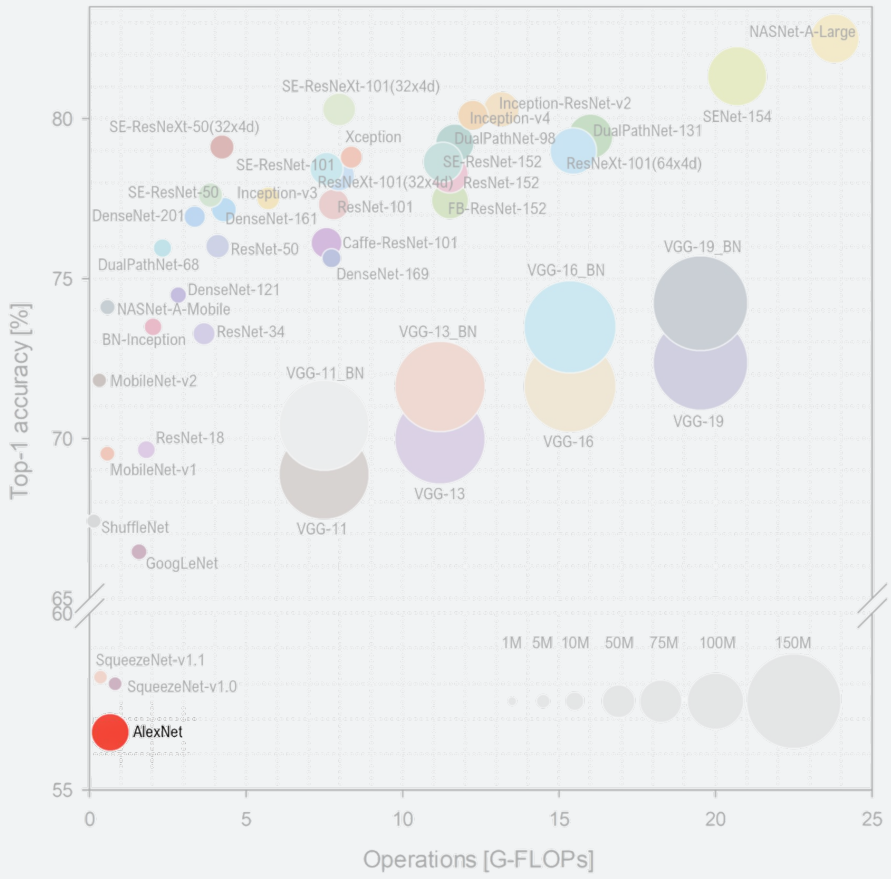
# ML Model Evolution




**Source:** S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018

# ML Model Evolution

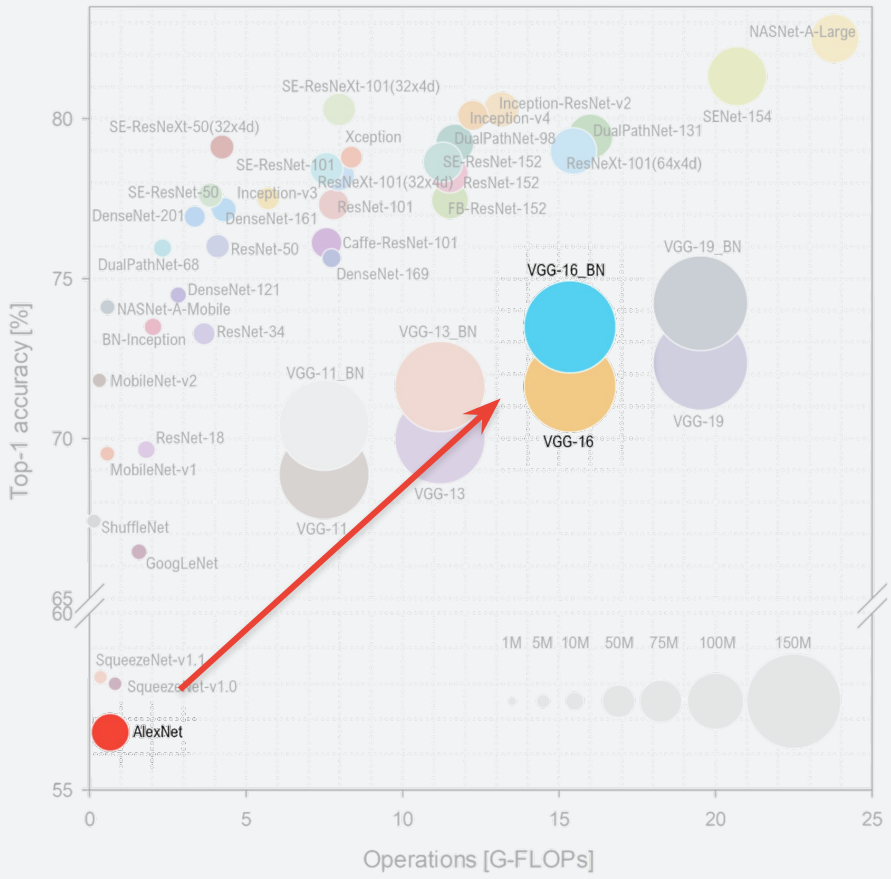
- **AlexNet** (2012)
  - 57.1% accuracy
  - 61MB in size




**Source:** S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018

# ML Model Evolution

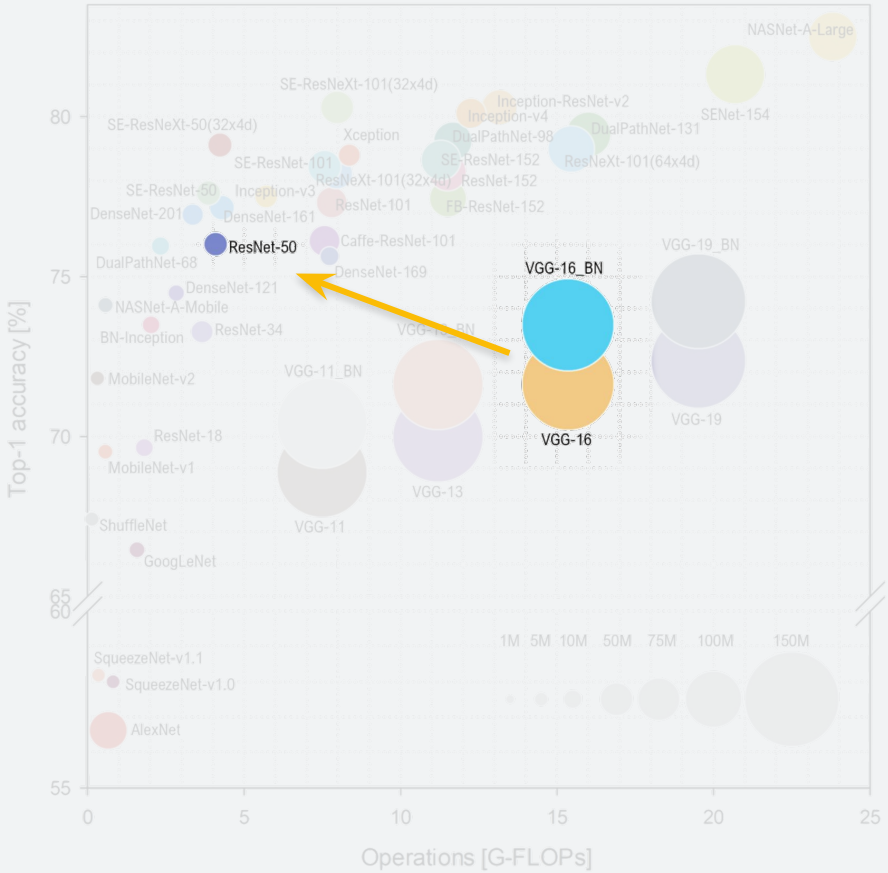
- **VGGNet** (2014) [VGG-16]
  - **71.5%** accuracy
  - **528MB** in size




**Source:** S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018

# ML Model Evolution

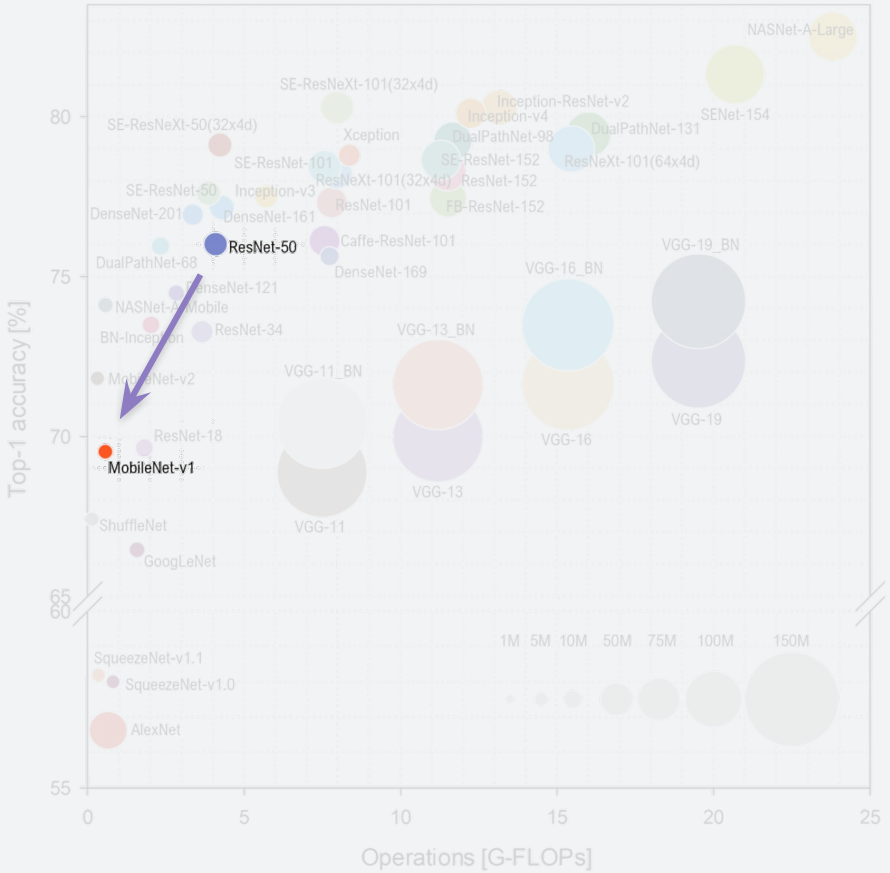
- **ResNet** (2015)
  - **75.8%** accuracy
  - **22.7MB** in size




**Source:** S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018

# ML Model Evolution

- **MobileNet** (2015)
  - ***MobileNetv1***
    - **70.6%** accuracy
    - **16.9MB** in size




**Source:** S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018

# ML Model Evolution

- **MobileNet** (2015)
  - ***MobileNetv1***
    - 70.6% accuracy
    - 16.9MB in size

**Source:** S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018

# Problem:

Our board (in your kit for Course 3) only has **256KB** of RAM (memory) yet ***MobileNetv1*** needs **16.9MB**!