tinyML Talks

Enabling Ultra-low Power Machine Learning at the Edge

“Getting Started with TinyML”

Pete Warden - Google
March 31, 2020

www.tinyML.org
tinyML Talks Sponsors

Sponsorships now available – contact Bette@tinyML.org for info
Next tinyML Talk

<table>
<thead>
<tr>
<th>Date</th>
<th>Presenter</th>
<th>Topic / Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 14</td>
<td>Sek Chai CTO &amp; founder of Latent.ai</td>
<td>Adaptive AI for a Smarter Edge</td>
</tr>
</tbody>
</table>

Webcast start time is 8 am Pacific time
Approximate length is 30 minutes

Please contact talks@tinyml.org if you are interested in presenting
Pete Warden is the technical lead of the TensorFlow mobile and embedded team at Google, and was previously CTO of Jetpac (acquired in 2014).
Stay Safe!
We’re in a pandemic

It feels weird to be thinking about technology when we’re in a crisis!

Everything I’m talking about will still be relevant when we emerge, so feel free to ignore it for now in favor of real priorities and catch up when you have time.

That said, being stuck in lockdown is boring, so I want to share some of the most interesting problems I know about.

All that you need should be orderable online too.
What is this talk about?

Getting you up and running with some practical embedded machine learning projects.

I’ll show you everything you need to create your own devices that make sense of the world around them.

No soldering, programming, or PhD required!
What is TinyML?

The usual definition is running machine learning on embedded devices at an average of less than one milliwatt in power.

This power requirement is important because it allows unattended devices on batteries or energy harvesting.

Here we’ll stretch the definition temporarily to include MCUs that use 10’s of mWs, since they’re easier to work with and widely available.
What hardware do you need?

Arduino Nano BLE Sense 33

store.arduino.cc/usa/nano-33-ble-sense

- $30
- Easy to program
- BLE built-in
- 10’s of milliwatts
- Not easy to battery power
- Small form-factor
What hardware do you need?

SparkFun Edge board

[sparkfun.com/products/15170](https://sparkfun.com/products/15170)

- $15
- 1 to 2 milliwatts, thanks to Ambiq
- Arduino support in-progress
- Battery slot
- Small form-factor
What hardware do you need?

AdaFruit AI Board

[adafruit.com/product/4317](adafruit.com/product/4317)

- $45
- 10’s of milliwatts
- Display
- Rechargeable battery support
- Arduino-compatible
What hardware do you need?

Discovery STM32F746G


- $56
- 100's of milliwatts
- Lots of peripherals
- Large display
- No battery support
- Big!
What hardware do you need?

All of these boards are great!

Don’t forget a USB Micro cable too if you don’t have one.

You’ll also need a laptop you can plug into. Mac, Linux, Windows all work with Arduino, and Chromebooks can too with a bit of fiddling.
Software
What software do you need?

I’m the technical lead for TensorFlow Lite Micro, so obviously I recommend checking out tensorflow.org/lite/microcontrollers!

Lots of vendors like ST Micro, NXP, CEVA and others also have their own libraries optimized for their platforms.

There’s a lot of great work happening in this space though, so do also look at projects like:

- [github.com/uTensor/uTensor](https://github.com/uTensor/uTensor)
- [edgeimpulse.com](https://edgeimpulse.com)
- [cartesiam.ai](https://cartesiam.ai)
What is TensorFlow Lite Micro?

It’s Google’s open source library for running machine learning on microcontrollers and DSPs, available at github.com/tensorflow/tensorflow/tree/master/tensorflow/lite/micro

It fits in less than 20KB of binary footprint, and has no operating system, malloc/free or C library dependencies, so it can run on bare metal.

TensorFlow Lite takes models from TensorFlow’s training environment and converts them into optimized representations that can be run on small devices.

It’s used internally for always-on applications, and has attracted industry support, with optimized kernels for Arm Cortex M, Cadence Tensilica Hifi, Synopsys ARC chips, and more.
The most common IDE for embedded development is Arduino. You can use the software with many boards supplied by other hardware manufacturers.

TensorFlow Lite Micro is an official library and you’ll find it in the library manager.

It comes with four examples, showing how to run speech, accelerometer, and image machine learning models on a microcontroller.
What can you do with your system?

Out of the box, TensorFlow Lite Micro on Arduino and other supported devices has several tutorials showing you how to build practical applications:

- Voice recognition. Listen out for a “wake word”.
- Magic wand. Detect gestures using an accelerometer
- Person detector. Analyze a camera image to spot people.

There are a lot of other possibilities for you to build your own solutions to real-world problems. These include:

- Enviromental. Wildlife tracking from bird songs to camera traps.
- Industrial. Predictive maintenance to save money and downtime for businesses.
How can you learn more?

tinymlbook.com

This is a 500 page book written by me and @dansitu covering end to end training and deployment of ML models with TFLite Micro.

It has chapters on all the examples, showing you step by step how to get up and running on a variety of different boards.

The first 6 chapters are available as a free PDF.
How can you learn more?

I’m @petewarden on Twitter, or you can email at petewarden@google.com with questions.

The tinyML organization has regular meetups (online during the pandemic) and lots of other great resources like this webinar series. Check out their forums!

The TFLite Micro working group has a chatroom and monthly video call, for people interested in developing with the software.
Please take the 5 question poll and continue the conversation @ tinyML.org/forums
Next tinyML Talk

<table>
<thead>
<tr>
<th>Date</th>
<th>Presenter</th>
<th>Topic / Title</th>
</tr>
</thead>
</table>
| April 14| Sek Chai
CTO & founder of Latent.ai | Adaptive AI for a Smarter Edge |

Webcast start time is 8 am Pacific time
Approximate length is 30 minutes

Please contact talks@tinyml.org if you are interested in presenting
Copyright Notice

This presentation in this publication was presented as a tinyML® Talks webcast. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyML.org