tinyML Summit
Enabling Ultra-low Power Machine Learning at the Edge

Products and applications enabled by tinyML

March 28 – 29, 2023

www.tinyML.org
Responsible Design of Edge AI

A Pattern Approach for Detecting and Mitigating Bias

Wiebke Hutiri
Delft University of Technology

TinyML Summit 2023

TU Delft
The Edge AI Dream

I really need a cup of coffee before I go to work.

No problem, the nearest café is 2 km away, route has been calculated.
1 Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.
2 Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.
VOETTEKST WIJZIGEN

1. Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.

2. Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.

The Edge AI Reality

Image credit: AI Addict YouTube | San Jose, USA
VOETTEKST WIJZIGEN

1. Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.

2. Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.

What could possibly go wrong?
VOETSTEKST WIJZIGEN

1 Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.

2 Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.

What could possibly go wrong?

THE ACCENT GAP
We tested Amazon’s Alexa and Google’s Home to see how people with accents are getting left behind in the smart-speaker revolution.

Al-generated Voice Firm Clamps Down After 4chan Makes Celebrity Voices for Abuse

4chan members used Eleven Labs to make deepfake voices of Emma Watson, Joe Rogan, and others saying racist, transphobic, and violent things.

Al-generated Voice Firm Clamps Down After 4chan Makes Celebrity Voices for Abuse

How I Broke Into a Bank Account With an Al-generated Voice

Banks in the U.S. and Europe tout voice ID as a secure way to log into your account. I proved it’s possible to trick such systems with free or cheap Al-generated voices.

Why biased AI can be damaging to your health

Tesla behind eight-vehicle crash was in ‘full self-driving’ mode, says driver

SF Firefighters Smash Cruise Self-Driving Taxi Window To Stop It From Driving Over Hose
VOETTEKST WIJZIGEN

Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.

Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.
What if We Embraced Responsible Design of Edge AI?
What is “Responsible Design” Anyways?

Responsible Design:
Recognising the Impact of How We Design

Aligning Design With What Society Values
What Does Society Value?

I'm the hungriest

My mom paid for it

I worked the hardest

It was my idea

4 kids sharing 9 slices of pizza. Who should get the last piece?
Humans care about fairness.

It was my idea
My mom paid for it
I worked the hardest
I'm the hungriest
Humans care about fairness.
1. Avoid being in a position of disadvantage
2. Avoid putting others in a position of disadvantage
VOETTEKST WIJZIGEN

1 Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.

2 Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.

THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

What Does Society Value?
What Does Society Value?

Article 2: No Discrimination

Everyone is entitled to all the rights and freedoms, without distinction of any kind, such as race, color, sex, language, religion, political or other opinions, national or social origin, property, birth, or another status.

Furthermore, no distinction shall be made on the basis of the political, jurisdictional, or international status of the country or territory to which a person belongs.
Bias ⇄ Fairness, Non-discrimination and Inclusion

Image credit: Johnny Miller Unequal Scenes | Rio de Janeiro, Brazil
Bias ⇔ Fairness, Non-discrimination and Inclusion

Bias:

Image credit: Johnny Miller Unequal Scenes | Rio de Janeiro, Brazil
Bias ≠ Fairness, Non-discrimination and Inclusion

Bias:
A skewed or slanted perspective.

AI Model: No Loan For You
ME: Why

AI:
“I said it right, just said it.”
Keh diya na. Bas, keh diya.

Image credit: Johnny Miller Unequal Scenes | Rio de Janeiro, Brazil
Bias ➔ Fairness, Non-discrimination and Inclusion

Bias:
Systematically assigns undesirable outcomes to individuals or groups on grounds deemed unreasonable

AI Model: No Loan For You
ME: Why
AI:

“I said it right, just said it.”

Image credit: Johnny Miller Unequal Scenes | Rio de Janeiro, Brazil
Bias ≠ Fairness, Non-discrimination and Inclusion

Bias > data problem

Image credit: Johnny Miller Unequal Scenes | Rio de Janeiro, Brazil
Bias ⇝ Fairness, Non-discrimination and Inclusion

Bias in on-device ML workflows

Image credit: Johnny Miller Unequal Scenes | Rio de Janeiro, Brazil
VOETTEKST WIJZIGEN

1 Ga naar de tab 'Invoegen' en klik op 'Kop- en voettekst'.

2 Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op 'Overal toepassen' om de gegevens op elke slide toe te passen.

There is no magic wand for fixing bias, but...
There is no magic wand for fixing bias, but...

We can recognise the impact of how we design and:
There is no magic wand for fixing bias, but...

We can recognise the impact of how we design and:

1. Detect bias
There is no magic wand for fixing bias, but...

We can recognise the impact of how we design and:

1. Detect bias
2. Take mitigating actions
There is no magic wand for fixing bias, but…

We can recognise the impact of how we design and:

1. Detect bias
2. Take mitigating actions
3. Be transparent about the design context
Detecting and Mitigating Bias in On-device ML

Know-how on how to detect and mitigate bias is not readily available.
Detecting and Mitigating Bias in On-device ML

**Know-how** on how to detect and mitigate bias is **not readily available**.

**Difficult to navigate and prioritise** algorithmic fairness techniques.
Detecting and Mitigating Bias in On-device ML

Know-how on how to detect and mitigate bias is not readily available.

Difficult to navigate and prioritise algorithmic fairness techniques.

Efforts to detect and mitigate bias are ineffective and even counterproductive.
Common Bias Evaluation Failures

1. Little consideration for who is affected
2. Little regard for what should be measured
3. No consideration for what metrics are capable of measuring
4. Limited consideration of what is being compared (i.e. which groups, and why)
Design patterns capture proven design experience in generalisable templates so that they can be reused in future design projects.
Example

Label Distribution Across Groups

“Analyse the label distribution to ensure that labels are balanced across groups.”
How to Use Design Patterns

Benchmark Evaluation Dataset Recipe

Purpose:

*Ensure that the benchmark datasets used for evaluation are representative, appropriate and sufficient.*
How to Use Design Patterns

Benchmark Evaluation Dataset Recipe

- Evaluation datasets
- Metadata approximation (optional)
- Label distribution across groups
- Data quantity distribution across groups
- Data source
## Design Pattern Purpose and Scope

<table>
<thead>
<tr>
<th>Pattern Scope</th>
<th>Dataset</th>
<th>Evaluation</th>
<th>Pre-processing</th>
<th>In-processing</th>
<th>Post-processing</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data quantity distribution across groups</td>
<td>Continuous testing</td>
<td>Decorrelating sensitive features</td>
<td>Adversarial bias mitigation</td>
<td>Output label redistribution</td>
<td>Visualisation</td>
</tr>
<tr>
<td></td>
<td>Label distribution across groups</td>
<td>Disaggregated base metrics</td>
<td>Feature editing</td>
<td>Imposing fairness constraints on the optimization objective</td>
<td>Threshold selection &amp; optimization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metadata approximation</td>
<td>Error tolerance</td>
<td>Sampling and reweighing</td>
<td>In-processing model selection</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluation datasets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fairness assumptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group-based bias measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Individual bias-measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Meta-measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Types of harm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Design Choices That Propagate Bias in On-device ML

Hutiri et. al. (2023) “Tiny, always-on and fragile: Bias propagation through design choices in on-device ML” just accepted ACM Transactions on Software Engineering & Methodology
Benefits of Using Design Patterns

Detect and mitigate bias *better* and *faster*
Benefits of Using Design Patterns

Detect and mitigate bias *faster* and *better*

Quickly upskill know-how on bias
Benefits of Using Design Patterns

- Detect and mitigate bias faster and better
- Quickly upskill know-how on bias
- Common language on bias for teams
Benefits of Using Design Patterns

- Detect and mitigate bias *faster* and *better*
- Quickly upskill know-how on bias
- Common language on bias for teams
- Surface and choose between design alternatives
Benefits of Using Design Patterns

- Detect and mitigate bias *faster and better*
- Quickly upskill know-how on bias
- Common language on bias for teams
- Surface and choose between design alternatives
Towards Fair and Non-discriminatory Edge AI
Towards Fair and Non-discriminatory Edge AI
1 Ga naar de tab ‘Invoegen’ en klik op ‘Kop- en voettekst’.
2 Typ de gewenste voettekst in, in het aangegeven tekstkader. Klik vervolgens op ‘Overal toepassen’ om de gegevens op elke slide toe te passen.

Towards Fair and Non-discriminatory Edge AI
Towards Fair and Non-discriminatory Edge AI

Integrate

Implement

Expand
Towards Fair and Non-discriminatory Edge AI

Integrate

Implement

Expand

Collaborate
Towards Fair and Non-discriminatory Edge AI

Wiebke Hutiri
w.toussaint@tudelft.nl
@wiebkehutiri
Copyright Notice

This presentation in this publication was presented at the tinyML® Summit (March 28 - 29, 2023). 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