ETHICS OF LANGUAGE ENCODING

ELIZA WELLS, EMBEDDED ETHI**CS**
10/21/21

1

WHO AM I?

- Eliza Wells, elizaw@mit.edu

- Philosophy PhD student at MIT

- Research interests: moral and social philosophy

- Graduate Fellow for Embedded EthiCS @ Harvard

GOALS FOR TODAY

1. Consider how language encoding systems can cause harms

2. Understand and be able to identify instances of two specific kinds of harms: *representational* and *allocative*

3. Think through how to balance avoiding those harms with avoiding technical costs using the principle of *reasonable accommodation*
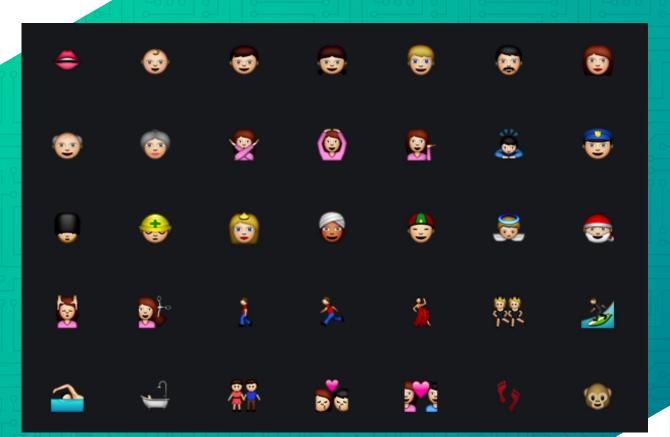
4. Apply these concepts using case studies

UNDERSTANDING HARMS

- Individuals and groups can suffer many different kinds of harms.
  - Physical harm, economic harm, sexual harm, political harm, emotional harm…
- Sometimes, individuals suffer as a result of their group membership.
- Harms can be caused without intent.

Does this set of emojis cause any harms?

If so, to whom?

How?

Apple's original emoji keyboard, 2008

Kate Crawford

REPRESENTATIONAL HARM

When a system reinforces the subordination of some group on the basis of identity.

Harmful stereotypes or misconceptions can be **reinforced** by representation **or** lack of representation.

- Google Photos algorithm that labeled Black faces as "gorillas"
- Image search results for "CEO" that depict only white men

HARMS AND LANGUAGE ENCODING

- What kinds of harms did non-English speakers face before Unicode?

Kate Crawford

ALLOCATIVE HARM

When a system withholds resources from certain groups.

Examples of potential resources: money, jobs, opportunities, education, health care, loans, etc.

- Voter ID laws
- Gender wage gap
- Rising costs of higher education

HARMS AND LANGUAGE ENCODING

- In what ways did Unicode alleviate some of those harms?

- Which harms still remain? Have any new ones been created?

AVOIDING SOME HARMS CAN CAUSE OTHER HARMS.

Technical costs of systems that can represent more languages have included…

- Storage costs

- Transmission costs

- Increased possibility of error and ambiguity

HOW DO WE WEIGH AVOIDING TECHNICAL COSTS AGAINST AVOIDING ALLOCATIVE AND REPRESENTATIONAL HARMS?

REASONABLE ACCOMMODATION

- A reasonable accommodation is an adjustment to a system that allows individuals with certain needs to participate in that system on **an equal basis with others**, without imposing an undue burden on those providing the accommodation.

- The denial of reasonable accommodation is discrimination.

REASONABLE ACCOMMODATION

Often, people need accommodations not because they need to compensate for a deficiency, but because the original system was not designed with them in mind.

EXAMPLES OF REASONABLE ACCOMMODATION

- Making buildings accessible for wheelchair users
- Extending exam times
- Gender neutral restrooms
- Breaks during the work day for prayer

UNDUE BURDEN

- An accommodation results in an undue burden when it requires "significant difficulty and expense," considered in context.

- Examples:

  - Hiring an ASL interpreter, *Searls v. Johns Hopkins Hospital*, 2016

  - Changing employee duties, *Treadwell v. Alexander*, 1983

  - Providing an audio transcription service, *Dobard v. San Francisco Bay Area Rapid Transit Authority*, 1993

ONE WAY OF MINIMIZING BURDENS: LESS COSTLY TECHNICAL SOLUTIONS

BACK TO PROFESSOR KOHLER!

CASE STUDIES

1. TOTO

ᗞᘎᗞᐱᘎ ᙊᕐᓂᓃᕐ ᐁᘎᓃᗞᕐ ᑎᐱ (x2)
ᑎᐱᑎᐱᐁᕐᗞᑎᕐ ᐁᘎᓃᕐᕐᓇ ᗞᗞᐁ
ᗞᘎᗞᐱᘎ ᙊᕐᓂᓃᕐ ᐁᘎᓃᗞᕐ ᑎᐱ

ᗞᐱᐸᐱᐁᕐᐸᕐᓇ ᐁᘎᓃᗞᕐ ᗞᐱᐸᕐᓇ ᐁᕐᔑᒐᐱ ᑎᐱᐸᕐᓇ ᔑᙊᗞᑌᗞᐁᐱ (x2)
ᒐᔑᓃᐸᕐᓇ ᙊᐱᘎᑌ ᓇᗞᓃᐱ ᐁᘎᓃᗞᕐ ᑎᐱ (x2)

ᗞᘎᗞᐱᘎ ᙊᕐᓂᓃᕐ ᐁᘎᓃᗞᕐ ᑎᐱ (x2)
ᑎᐱᑎᐱᐁᕐᗞᑎᕐ ᐁᘎᓃᕐᕐᓇ ᗞᗞᐁ
ᗞᘎᗞᐱᘎ ᙊᕐᓂᓃᕐ ᐁᘎᓃᗞᕐ ᑎᐱ

Toto is a critically endangered language spoken by a village of 1,500 people in India. Speakers are decreasing because children speak Bengali at school and most jobs require Hindi or English. The Toto script for writing the language was developed in 2015. Since 2015, Toto speakers have expressed a renewed interest in writing their own language and a sense of status in having their own script.

2. KLINGON

Klingon was developed by Marc Okrand in 1985 for a Star Trek movie. Klingon is estimated to have 60 fluent speakers (none as a first language), and there is a play, an opera, and several books almost entirely in Klingon. There is even a Duolingo course for learning it.

3. MAYA SCRIPT

The Maya script was used from at least the 3rd century BCE until Spanish colonization in the 16th century. It is no longer used by any modern speakers, and most texts written in it were destroyed by the conquistadors. However, it has historical and cultural importance. The script has more than 300 characters which can be combined in a number of difficult ways.
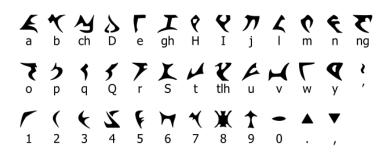
DISCUSSION

What are the most significant harms at stake in choosing whether or not to encode these scripts in Unicode? What should computer scientists do?

Toto

Klingon

Maya

BEYOND ENCODING

Encoding languages in Unicode does not guarantee their use. Computer scientists must also develop fonts and use those fonts on operating platforms.

- **Cherokee** was encoded in Unicode in 1999, but didn't appear in Google or Apple products until 2011.

- **N'Ko** was encoded in 2006, but could not be used by its speakers until 2012 because it was not rendered on the Windows platforms they used.

REASONABLE ACCOMMODATION AND COMPUTER SCIENCE

When considering a design decision, ask:

- Who will benefit from this decision?
- Who will it harm?
  - In what ways?
- What will it take to avoid that harm?
  - Will different harms be caused?
- Does that constitute an undue burden?
  - Who has the resources to shoulder this burden?
  - Who has the responsibility to shoulder it?

THANK YOU!

Evaluation: **https://tinyurl.com/CS61EmbEF21**