

ETHICS OF LANGUAGE ENCODING

ELIZA WELLS, EMBEDDED ETHICS

10/21/21

WHO AM I?

- ① Eliza Wells, elizaw@mit.edu
- ① Philosophy PhD student at MIT
- ① Research interests: moral and social philosophy
- ① 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

Image from <https://emojipedia.org/apple/iphone-os-2.2/>



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?

2B820 — 1.1 TC-2121 弋	2B834 — 1.10 V4-4033 砧	2B848 2.9 GZJW-00036 南	2B85C 4.9 GZJW-00056 帥
2B821 — 1.3 GDC-002 丕	2B835 — 1.10 GZJW-00080 冪	2B849 2.9 V4-404A 钟	2B85D 4.10 V4-4082 隰
2B822 — 1.4 V4-4022 寺	2B836 — 1.11 TD-2138 耆	2B84A 2.13 GZJW-00667 勑	2B85E 4.11 GZJW-00672 飭
2B823 — 1.4 UTC-00318 东	2B837 — 1.12 V4-4039 禕	2B84B 2.14 GZJW-00668 靱	2B85F 4.14 GZJW-00674 無
2B824 — 1.5 GZJW-00006 芳	2B838 — 1.12 V4-403A 清	2B84C 2.14 V4-404D 卷	2B860 4.14 V4-4053 嬾
2B825 — 1.5 UTC-00006 丙	2B839 — 1.12 V4-403B 能	2B84D 2.19 V4-404E 嬾	2B861 4.27 GZJW-00673 霧
2B826 — 1.6 GX-C-2000.04 所	2B83A — 1.12 V4-403C 慮	2B84E 2.21 V4-404F 彘	2B862 乙 5.3 GZJW-00040 卮
2B827 — 1.6 V4-4027 軒	2B83B — 1.12 V4-4037 耆	2B84F 4.2 TC-2136 夕	2B863 乙 5.4 GZJW-00044 攷
2B828 — 1.6 GZJW-00048 皂	2B83C — 1.13 V4-4033 堯	2B850 4.4 V4-4060 旂	2B864 乙 5.5 JK-46509 丸



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?

2B820 — 1.1 弋 TC-2121	2B834 — 1.10 砵 V4-4033	2B848 2.9 南 GZJW-00036	2B85C 4.9 𠂇 GZJW-00056
2B821 — 1.3 丕 GIDC-002	2B835 — 1.10 冪 GZJW-00080	2B849 2.9 𠂇 V4-404A	2B85D 4.10 𠂇 V4-4082
2B822 — 1.4 寺 V4-4022	2B836 — 1.11 𠂇 TD-2138	2B84A 2.13 𠂇 GZJW-00657	2B85E 4.11 𠂇 GZJW-00672
2B823 — 1.4 东 UTC-00318	2B837 — 1.12 𠂇 V4-4039	2B84B 2.14 𠂇 GZJW-00668	2B85F 4.14 𠂇 V4-4074
2B824 — 1.5 𠂇 GZJW-00006	2B838 — 1.12 𠂇 V4-403A	2B84C 2.14 𠂇 V4-404D	2B860 4.14 𠂇 V4-4053
2B825 — 1.5 𠂇 UTC-00006	2B839 — 1.12 𠂇 V4-403B	2B84D 2.19 𠂇 V4-404E	2B861 4.27 𠂇 GZJW-00673
2B826 — 1.6 𠂇 GX-2000.04	2B83A — 1.12 𠂇 V4-403C	2B84E 2.21 𠂇 V4-404F	2B862 Z 5.3 𠂇 GCY-00040
2B827 — 1.6 𠂇 V4-4027	2B83B — 1.12 𠂇 V4-4037	2B84F 4.2 𠂇 TC-2136	2B863 Z 5.4 𠂇 GZJW-00044
2B828 — 1.6 𠂇 GZJW-00048	2B83C — 1.13 𠂇 V4-4033	2B850 4.4 𠂇 V4-4060	2B864 Z 5.5 𠂇 JK-45009

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 HARMES?**

REASONABLE ACCOMMODATION

- ① 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.
- ② 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

ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ (x2)
ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ
ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ

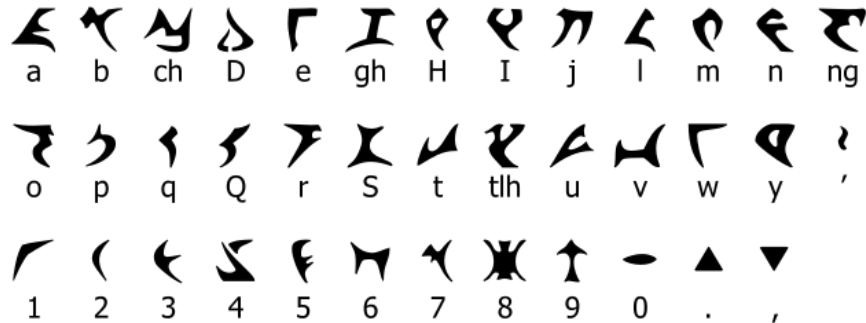
ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ (x2)
ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ (x2)

ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ (x2)
ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ
ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ ᑭᑦᑲᑦᑲᑦ

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

ዮጋዳሊግ ሥጋዕጌ ፕሌዮጌ በሌ (x2)
በሌበሌፕሮጋጋጌ ፕሌዮዮ ያጋፕ
ዮጋዳሊግ ሥጋዕጌ ፕሌዮጌ በሌ
ዳሌጋሌፕጋጋጌ ፕሌዮጌ ዳሌጋጌ ፕጌጌሌ በዘጋግ
ጌጌጋጋጌ ሥሌፕህ ጆጆጋሌ ፕሌዮጌ በሌ (x2)
ዮጋዳሊግ ሥጋዕጌ ፕሌዮጌ በሌ (x2)
በሌበሌፕሮጋጋጌ ፕሌዮዮ ያጋፕ
ዮጋዳሊግ ሥጋዕጌ ፕሌዮጌ በሌ

Klingon

a	b	ch	D	e	gh	H	I	j	l	m	n	ng
o	p	q	Q	r	S	t	tlh	u	v	w	y	'
1	2	3	4	5	6	7	8	9	0	.	,	

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:

- ① Who will benefit from this decision?
- ① Who will it harm?
 - ① In what ways?
- ① What will it take to avoid that harm?
 - ① Will different harms be caused?
- ① Does that constitute an undue burden?
 - ① Who has the resources to shoulder this burden?
 - ① Who has the responsibility to shoulder it?

THANK YOU!

Evaluation: <https://tinyurl.com/CS61EmbEF21>

