

# **ETHICS, CHOICES, AND VALUES**



# **HELLO AGAIN!**

**I'm Katie Creel, the Embedded EthiCS Fellow.  
You can always get in touch at [kcreel@stanford.edu](mailto:kcreel@stanford.edu)  
Or plan a visit to my office hours at  
[calendly.com/kathleencreel](https://calendly.com/kathleencreel)**

# San Jose Mercury News

25 cents

Serving Northern California Since 1851

March 1, 1989

## Enter. What is ethical in computer use? Return

By Tom Philp  
Mercury News Staff Writer

A Stanford computer scientist and a philosopher are developing the university's first course to get students to examine the ethical implications of their use of computers.

The broad-ranging course, to be taught this spring, will deal with topics ranging from the outbreak of computer

viruses to privacy issues of electronic bulletin boards. While some universities have developed courses to help students prepare for the rapidly changing computer world, no other university in Silicon Valley — or the Bay Area — now offers such a course.

"We're not trying to give them the answers," said Terry Winograd, the associate professor of computer science

who is developing the course. "We're trying to get them to do good thinking."

Among the questions to be pondered: should students freely share copyrighted software? Should they be concerned if their work has military applications? Should they submit a project on deadline if they are concerned that potential bugs could ruin others' work?

For two years, Stanford has offered a

seminar on computer ethics, but it was for fewer than a dozen students. But the new course, which can satisfy a curriculum requirement for computer science undergraduate students, will probably be several times larger.

"The hope is, we can take students who are currently more oriented in

See ETHICS, Page 8A

The course will address issues like invasion of privacy, ownership of computer programs, and the risks they are introducing to people's lives

# OUTLINE





# Homework 7b: Analyzing Data Bias

## Using Matplotlib

**Due: 11:55pm (Pacific Daylight Time) on Wednesday, May 26th**

BASED ON PROBLEMS BY COLIN KINCAID, MONICA ANUFORO, JENNIE YANG, NICK BOWMAN, JULIETTE WOODROW, CHRIS PIECH, MEHRAN SAHAMI, AND KATHLEEN CREEL.

CS198 Advertisement: Become a teacher! At Stanford we welcome section leaders from every walk of life -- and we teach students from all corners of the university. Have you thought about applying to section lead? [Learn more.](#)

In this assignment, we have done almost all of the data processing for you. We have organized the data files and stored them into one json which you will load and then use when writing your code. In this assignment, you will write code to plot the data in an interesting way. Plotting and visualizing across professor gender and review quality reveals interesting trends about human language usage. We hope that you will be able to use this exercise in data visualization to also think critically about the underlying biases that exist in online datasets! The end product of this assignment is a complete application that will help you dig deep into our provided dataset while answering important social and ethical questions along the way.

QUALITY

5.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

DIFFICULTY

1.0

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



# **WHAT CAN WE LEARN FROM THIS DATA?**

**In order to know how to use a dataset  
appropriately, we need to examine it for patterns of  
bias.**

# What is in the Dataset?





A photograph of three spelunkers standing at the entrance of a cave. They are silhouetted against a bright, sunlit green valley. Each spelunker has a headlamp on, which is turned on, illuminating their faces and the surrounding cave walls. The cave interior is dark, while the exterior view is a vibrant, sun-drenched landscape with dense green vegetation. The text 'You Have the Power to Find Out!' is overlaid on the right side of the image in white, bold, sans-serif font, with each word on a separate grey rectangular background.

**You Have  
the Power  
to Find  
Out!**



A photograph of three spelunkers in a dark cave, looking out through a large opening at a bright, sunlit forest. The spelunkers are silhouetted against the light, and their headlamps are on. The forest outside is lush and green, with sunlight filtering through the trees.

You Have  
the Power  
to Find  
Out!

Values &  
Habits

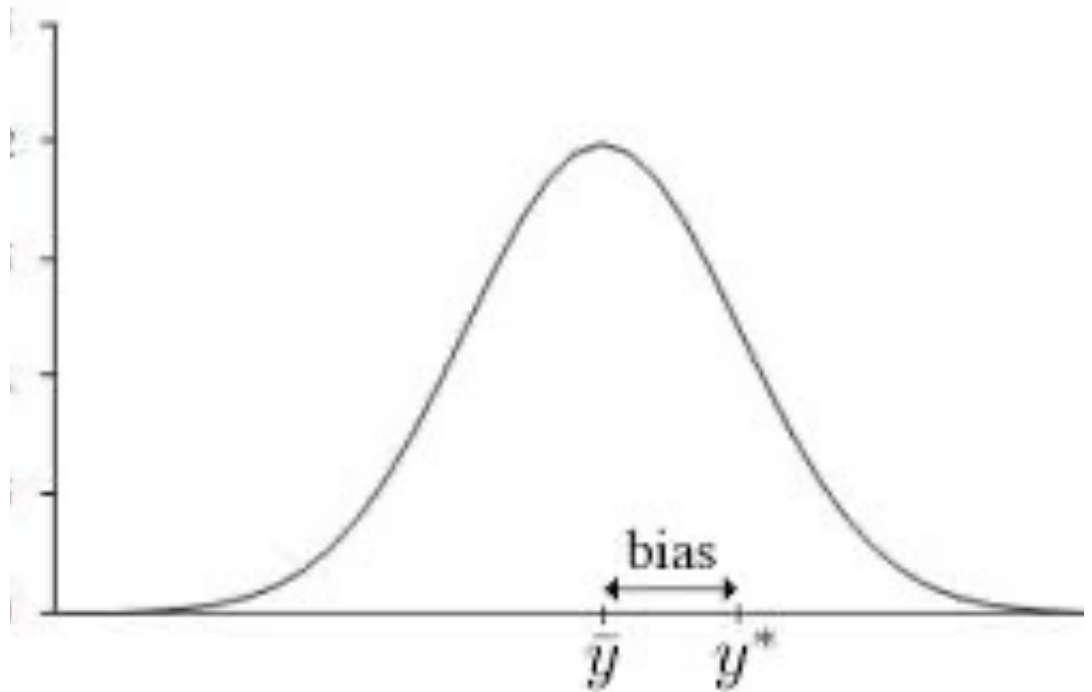
Ethics  
Skills

CS  
Skills

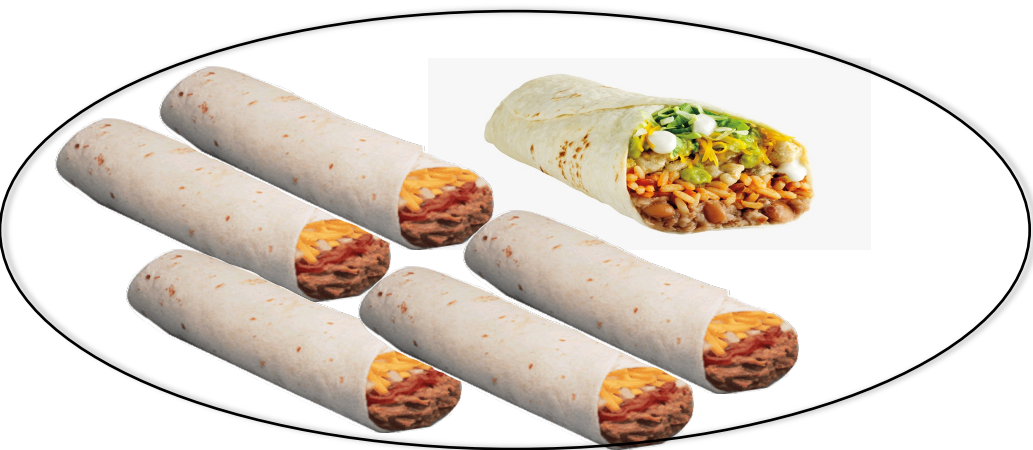
# WHAT IS BIAS?

Statistical bias is the difference between measured results and “true” value.

This is the “neutral” or statistical meaning of the word bias. You will see it often in discussions of patterns in data.

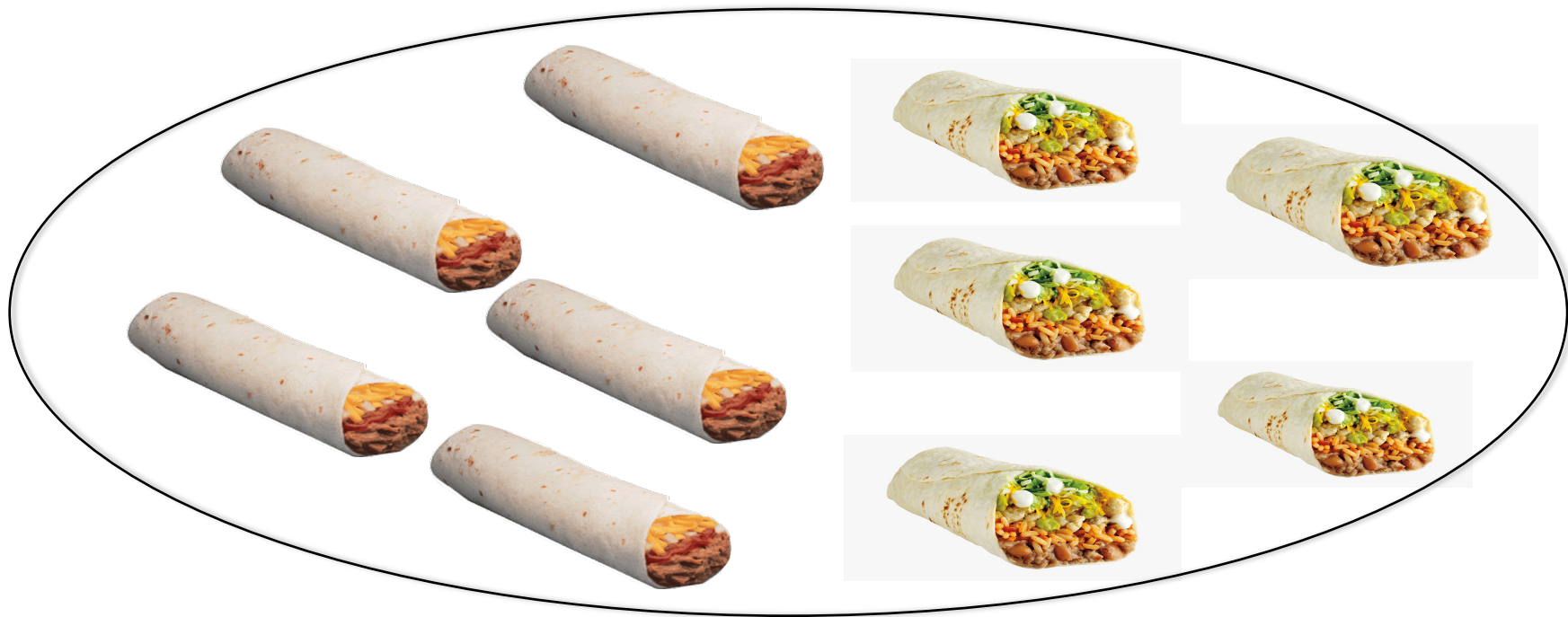


# EXAMPLE: SAMPLING BIAS





# EXAMPLE: SAMPLING BIAS



What Kinds of Bias Raise  
Ethical Concerns?



# DISCRIMINATORY BIAS IN DATA

**Biased measurement or classification**

**+ use of that bias to compound existing injustice or to fail to treat all as having equal moral worth (Hellman 2020)**

**=> *Discriminatory or Unfair Bias***

**HOW DOES  
DEFINING BIAS  
HELP US  
UNDERSTAND OUR  
PROFESSOR  
RATINGS DATA?**

# I HAVE DATA ABOUT PEOPLE! NOW WHAT? CHECKING BIAS

Decide how to use the data given the bias

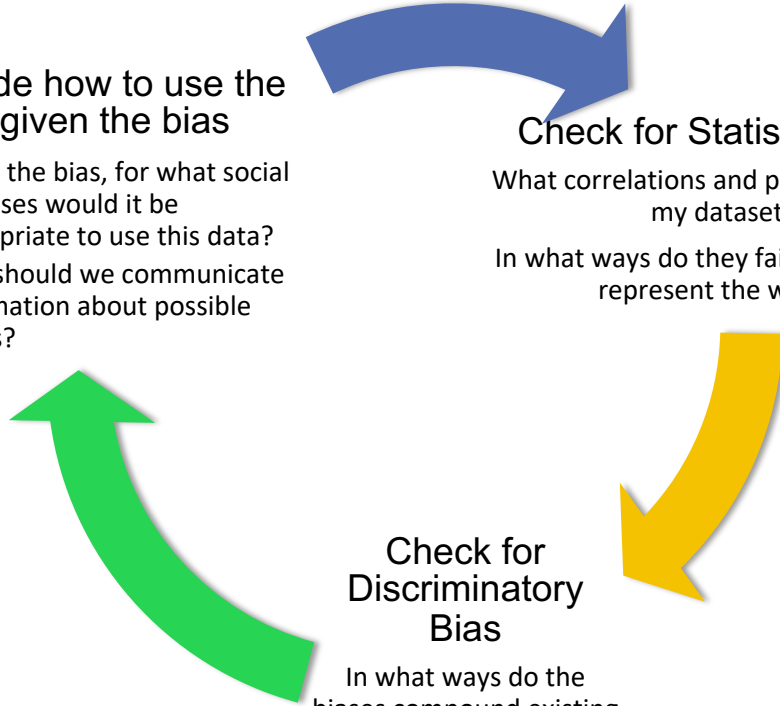
- Given the bias, for what social purposes would it be appropriate to use this data?
- How should we communicate information about possible biases?

Check for Statistical Bias

- What correlations and patterns exist in my dataset?
- In what ways do they fail to accurately represent the world?

Check for Discriminatory Bias

- In what ways do the biases compound existing injustice or fail to treat all people as being of equal moral worth?



# DATASHEETS FOR DATASETS (2020)

Accompany each dataset with a “datasheet” describing:

- **Motivation**
- **Composition**
- **Collection Process**
- **Preprocessing/cleaning/labeling**
- **Uses**
- **Distribution**
- **Maintenance**

TIMNIT GEBRU, Google

JAMIE MORGENSTERN, Georgia Institute of Technology

BRIANA VECCHIONE, Cornell University

JENNIFER WORTMAN VAUGHAN, Microsoft Research

HANNA WALLACH, Microsoft Research

HAL DAUMÉ III, Microsoft Research; University of Maryland

KATE CRAWFORD, Microsoft Research; AI Now Institute

The machine learning community currently has no standardized process for documenting datasets, which can lead to severe consequences in high-stakes domains. To address this gap, we propose *datasheets for datasets*. In the electronics industry, every component, no matter how simple or complex, is accompanied with a datasheet that describes its operating characteristics, test results, recommended uses, and other information. By analogy, we propose that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses, and so on. Datasheets for datasets will facilitate better communication between dataset creators and dataset consumers, and encourage the machine learning community to prioritize transparency and accountability.

**WHAT  
FRAMEWORKS CAN  
WE USE TO  
EVALUATE FAIR  
DISTRIBUTIONS?**

# Equality of Opportunity:

everyone has same opportunity to develop skills needed for the job, apply for the job, and get promoted.





**Parity: Everyone is equally likely to be a good teacher, so we should expect to end up with number of good teachers (and high rankings) proportionate to population.**



# OTHER DEFINITIONS OF FAIRNESS (IN CS 109!)

## Simulating loan decisions for different groups

Drag the black threshold bars left or right to change the cut-offs for loans.  
Click on different preset loan strategies.

### Loan Strategy

Maximize profit with:

**MAX PROFIT**

No constraints

**GROUP UNAWARE**

Blue and orange thresholds are the same

**DEMOGRAPHIC PARITY**

Same fractions blue / orange loans

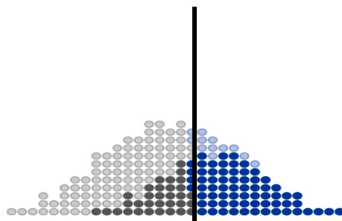
**EQUAL OPPORTUNITY**

Same fractions blue / orange loans to people who can pay them off

### Blue Population

0 10 20 30 40 50 60 70 80 90 100

loan threshold: 59

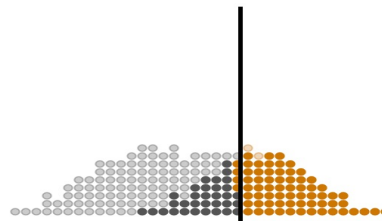


denied loan / would default  granted loan / defaults   
denied loan / would pay back  granted loan / pays back 

### Orange Population

0 10 20 30 40 50 60 70 80 90 100

loan threshold: 53



denied loan / would default  granted loan / defaults   
denied loan / would pay back  granted loan / pays back 

# **FAIRNESS BEYOND THE NUMBERS**

QUALITY

5.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

DIFFICULTY

1.0

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



QUALITY

5.0

DIFFICULTY

1.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



QUALITY

5.0

DIFFICULTY

1.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that ~~has never done~~ anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



AWESOME

QUALITY

5.0

DIFFICULTY

1.0



CS1



AWESOME

May 21st, 2015

Attendance: **No** Mandatory Grade: **A** Textbook: **No** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

CLEAR

GREAT TEACHER



# DESCRIPTIVE VS NORMATIVE LANGUAGE

Descriptive  
language

"is"

"was"

what people did

what happened

Normative  
language

"right"

"wrong"

"good"

"bad"

"should"

"should not"



# Descriptive or Normative?

QUALITY

5.0



CS101



AWESOME

21st, 2015

Attendance: **Not Mandatory** Grade: **A** Texts:

CLEAR

DIFFICULTY

1.0

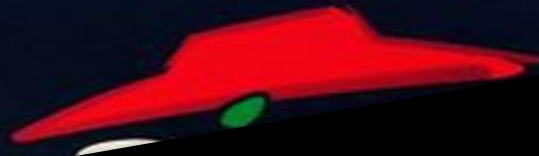
This class was awesome. A beginner like me that had never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

GREAT TEACHER





**Descriptive**



**Normative**

# THICK NORMATIVE TERMS

**“Thick” normative  
terms are both  
descriptive AND  
normative**

## **EXAMPLES:**

**Cowardly :: Cautious**

**Polite :: (?)**

**Rude :: (?)**

**Chill :: (?)**

**Kind :: (?)**

**etc**

# **THICK NORMATIVE TERMS & FAIRNESS**

- **We compare people in many ways, not just numerically**
- **Thick normative terms express morally or aesthetically “loaded” judgments**

# **DESCRIPTIVE OR NORMATIVE?**

**Mark Zuckerberg on whether Facebook would fact-check false claims about election suppression:**

- 1. “We have a different policy, I think, than Twitter on this.”**
- 2. “You know, I just believe strongly that Facebook shouldn’t be the arbiter of truth of everything that people say online.”**
- 3. “I think in general private companies probably shouldn’t be—or, especially these platform companies—shouldn’t be in the position of doing that.**

# **DESCRIPTIVE OR NORMATIVE?**

**Not surprising  
that statements  
setting the policy  
for platforms  
would be  
normative.**

**What about the  
programs behind  
the platforms  
themselves?**

**Do programs like  
the ones you are  
writing contain  
normative claims  
or values?**

# **HOW ARE VALUES EMBEDDED IN DESIGN?**

# **1. PROBLEM FORMULATION EMBEDS VALUES**

**Formulating a problem means  
describing the desired solution as good  
or worthy of being done.**



# **PROBLEM FORMULATION STATEMENTS ARE NORMATIVE**

**Formulating a problem means describing the desired solution as good or worthy of being done.**

**What is the problem to be solved?**

**For whom is this a problem? Who would benefit from its solution?**

**Who can agree that this is a problem worth solving?**

# WHAT IS THE PROBLEM TO BE SOLVED?

**“HOMELESS PEOPLE  
ARE SLEEPING HERE  
AND WE (WHO IS WE?)  
WANT THEM TO STOP”**



# WHAT IS THE PROBLEM TO BE SOLVED?

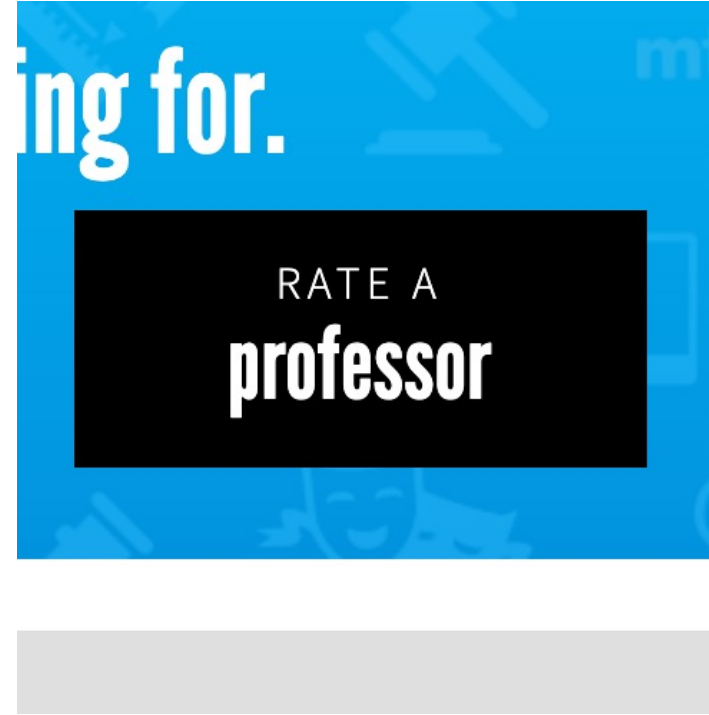
**PEOPLE ARE SLEEPING  
HERE AND WE (WHO IS  
WE?) WANT THEM TO  
STOP”**



**“SOME PEOPLE IN OUR  
COMMUNITY DON'T HAVE  
A PLACE TO SLEEP AND  
WE (WHO IS WE?) THINK  
THEY SHOULD”**



# 1. PROBLEM FORMULATION EMBEDS VALUES



# EXAMPLE: FORMULATING A RATINGS PROBLEM



- **What are the problem(s) to be solved?**
- **For whom are these problems?**
- **Who would benefit from their solution(s)?**
- **For each problem, who can agree that the problem is worth solving?**

# EXAMPLE: FORMULATING A SEARCH PROBLEM



- **What are the problem(s) to be solved?**
- **For whom are these problems?**
- **Who would benefit from their solution(s)?**
- **For each problem, who can agree that the problem is worth solving?**

# 2. CHOICE OF DATA EMBEDS VALUES

CS 106A Mid-Quarter Evaluation

\* Required

Instructor Feedback

These first few questions come from Mehran Sahami, Chris Piech, and Juliette Woodrow. Your anonymous feedback will be reviewed by them, so please be specific and give suggestions for improvement.

What has Mehran been doing so far that has worked well for you? What should he continue doing? \*

Your answer

What has Mehran been doing so far that has not worked well for you? Is there anything that he should stop or start doing?

Your answer

What has Chris been doing so far that has worked well for you? What should he

**Surveys are cheap  
to run**

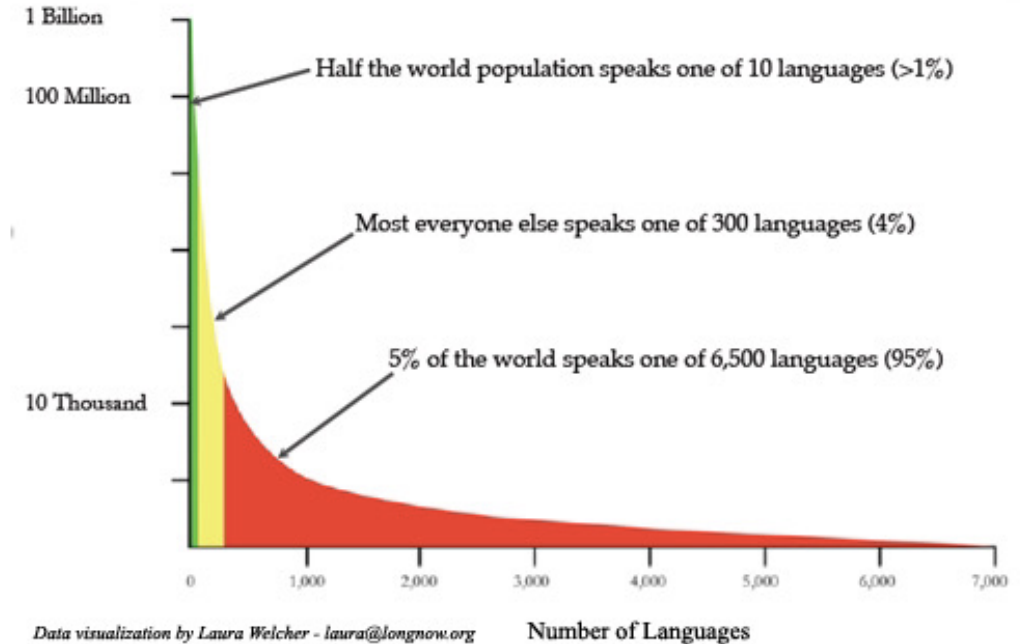
**They measure  
opinions**

**What are other  
ways to measure  
quality of  
professors?**

# **REPRESENTATION, LANGUAGE, AND DATA**

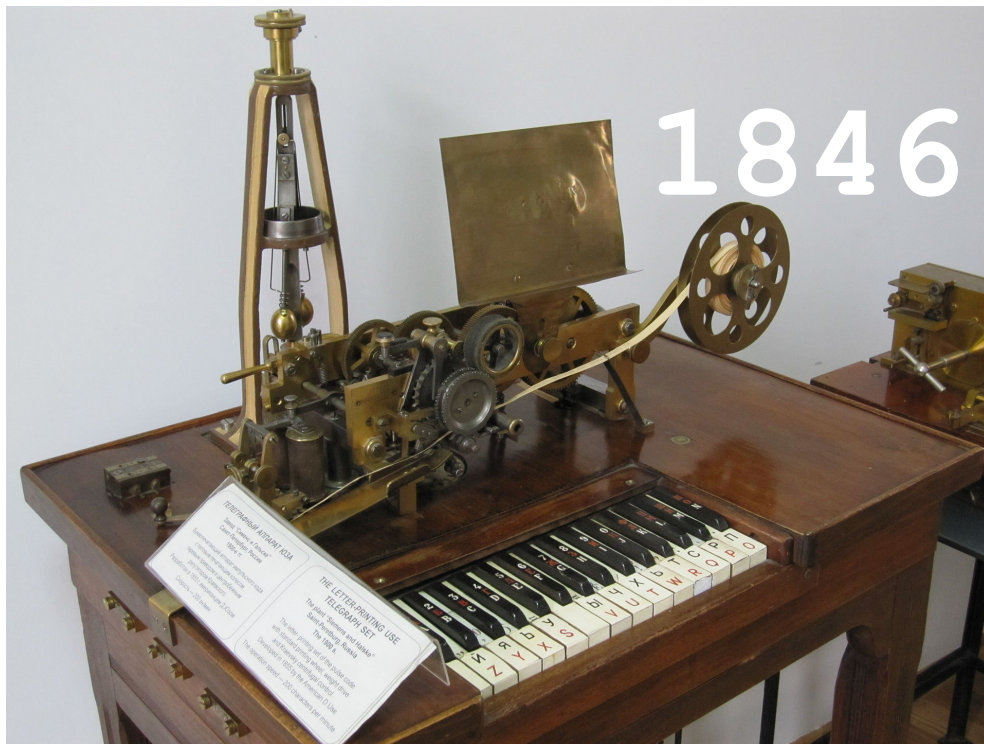


# LINGUISTIC REPRESENTATION & THE LONG TAIL



***A Very Very Brief  
History of Symbol Representation  
in Communications Technology ...***

# Telegraph Typewheel, Covers 56 Characters

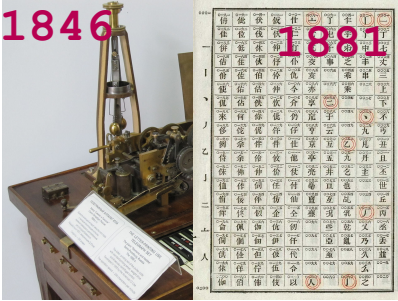


# 4 Digit Character-to-Code Mapping for the Telegraph

〇〇一	〇一八 俯	〇一六 礎	〇一四 伏	〇一二 板	〇一〇 仕	〇〇九 上	〇〇六 了	〇〇四 乍	〇〇二 丁	〇〇〇 一
一〇一	〇一八 佳	〇一六 位	〇一四 伐	〇一二 伏	〇一〇 仁	〇〇九 亡	〇〇六 地	〇〇四 乎	〇〇二 了	〇〇〇 丁
一〇二	〇一八 併	〇一六 低	〇一四 休	〇一二 仰	〇一〇 行	〇〇九 元	〇〇六 子	〇〇四 中	〇〇二 七	〇〇〇 中
一〇三	〇一八 倍	〇一六 住	〇一四 伙	〇一二 仔	〇一〇 仆	〇〇九 事	〇〇六 之	〇〇四 丰	〇〇二 丈	〇〇〇 丈
一〇四	〇一八 份	〇一六 佐	〇一四 伯	〇一二 伶	〇一〇 仇	〇〇九 亥	〇〇六 乖	〇〇四 串	〇〇二 三	〇〇〇 三
一〇五	〇一八 使	〇一六 佑	〇一四 估	〇一二 仲	〇一〇 今	〇〇九 亦	〇〇六 乘	〇〇四 上	〇〇二 上	〇〇〇 上
一〇六	〇一八 侃	〇一六 估	〇一四 伙	〇一二 介	〇一〇 亨	〇〇九 亨	〇〇六 亨	〇〇四 下	〇〇二 下	〇〇〇 下
一〇七	〇一八 來	〇一六 何	〇一四 你	〇一二 批	〇一〇 仍	〇〇九 荒	〇〇六 于	〇〇四 不	〇〇二 不	〇〇〇 不
一〇八	〇一八 侈	〇一六 佻	〇一四 佻	〇一二 仔	〇一〇 章	〇〇九 云	〇〇六 九	〇〇四 丐	〇〇二 丐	〇〇〇 丐
一〇九	〇一八 例	〇一六 余	〇一四 伴	〇一二 伴	〇一〇 仕	〇〇九 京	〇〇六 互	〇〇四 凡	〇〇二 丑	〇〇〇 丑
一一〇	〇一八 待	〇一六 余	〇一四 佻	〇一二 佻	〇一〇 他	〇〇九 亭	〇〇六 五	〇〇四 九	〇〇二 丹	〇〇〇 且
一一一	〇一八 侏	〇一六 佛	〇一四 仲	〇一二 价	〇一〇 仗	〇〇九 亮	〇〇六 井	〇〇四 乞	〇〇二 主	〇〇〇 丕
一一二	〇一八 徂	〇一六 作	〇一四 伺	〇一二 任	〇一〇 付	〇〇九 曼	〇〇六 亘	〇〇四 也	〇〇二 世	〇〇〇 世
一一三	〇一八 侑	〇一六 佻	〇一四 伴	〇一二 仿	〇一〇 仙	〇〇九 亘	〇〇六 互	〇〇四 亘	〇〇二 丘	〇〇〇 丘
一一四	〇一八 侑	〇一六 佻	〇一四 似	〇一二 企	〇一〇 全	〇〇九 况	〇〇六 乳	〇〇四 丙	〇〇二 丙	〇〇〇 丙
一一五	〇一八 命	〇一六 佩	〇一四 伽	〇一二 佻	〇一〇 仞	〇〇九 些	〇〇六 乾	〇〇四 父	〇〇二 丞	〇〇〇 丞
一一六	〇一八 侗	〇一六 侗	〇一四 佻	〇一二 伊	〇一〇 仞	〇〇九 亞	〇〇六 亂	〇〇四 乃	〇〇二 丟	〇〇〇 丟
一一七	〇一八 供	〇一六 佻	〇一四 但	〇一二 佻	〇一〇 代	〇〇九 亟	〇〇六 亟	〇〇四 久	〇〇二 亟	〇〇〇 亟
一一八	〇一八 依	〇一六 伴	〇一四 仔	〇一二 伍	〇一〇 令	〇〇九 令	〇〇六 令	〇〇四 么	〇〇二 么	〇〇〇 么
一一九	〇一八 伽	〇一六 伯	〇一四 佈	〇一二 伎	〇一〇 以	〇〇九 人	〇〇六 丁	〇〇四 之	〇〇二 之	〇〇〇 之



1846



USASCII code chart

1963

		Column							
Row	Bit								
	b <sub>7</sub>	b <sub>6</sub>	b <sub>5</sub>	b <sub>4</sub>	b <sub>3</sub>	b <sub>2</sub>	b <sub>1</sub>	b <sub>0</sub>	
0	0	0	0	0	0	0	0	0	0
0	0	1	2	3	4	5	6	7	
0	0	0	0	0	NUL	DLE	SP	0	@
0	0	0	1	1	SOH	DC1	!	1	A
0	0	1	0	2	STX	DC2	"	2	B
0	0	1	1	3	ETX	DC3	#	3	C
0	1	0	0	4	EOT	DC4	\$	4	D
0	1	0	1	5	ENO	NAK	%	5	E
0	1	1	0	6	ACK	SYN	&	6	F
0	1	1	1	7	BEL	ETB	'	7	G
1	0	0	0	8	BS	CAN	(	8	H
1	0	0	1	9	HT	EM	)	9	I
1	0	1	0	10	LF	SUB	*	:	J
1	0	1	1	11	VT	ESC	+	;	K
1	1	0	0	12	FF	FS	.	<	L
1	1	0	1	13	CR	GS	-	=	M
1	1	1	0	14	SD	RS	.	>	N
1	1	1	1	15	SI	US	/	?	O



1991

U+20E0	U+1502	U+2866	U+943E	U+0590	U+280F
/	U+2B1C	U+1490	U+2639	U+052B	U+13EA

Adopt a Character +

Emoji +

Basic Info +

News

Events

Connect +

Membership +

Press

Everyone in the world should be able to use their own language on phones and computers.


























[LEARN MORE ABOUT UNICODE](#)

U+0904	U+05E9	U+0E07	U+1035	U+0378	U+0937
--------	--------	--------	--------	--------	--------

# Unicode!



- Adopt a Character +
- Emoji +
- Basic Info +
- News
- Events
- Connect +
- Membership +
- Press

 U+30ED	 U+15D2	 U+2666	 U+043E	 U+2690	 U+2B07
 U+FF0F	 U+2B1C	 U+1F490	 U+2639	 U+263B	 U+13EA
 U+1F643	<p>Everyone in the world should be able to use their own language on phones and computers.</p> <p><a href="#">LEARN MORE ABOUT UNICODE</a></p>				
 U+0994	 U+05E9	 U+0E07	 U+1D25	 U+51F8	 U+0B37
 U+203F	 U+1F60A	 U+03CE	 U+127F	 U+060F	 U+1557

# Unicode!

The screenshot shows the Unicode website interface. At the top left is the 'Unicode' logo. Below it is a grid of various characters, each with its corresponding Unicode code point (e.g., U+30ED, U+1302, U+2666, U+043E, U+2690, U+2B07). A large yellow callout box is overlaid on the grid, containing the text: 'Everyone in the world should be able to use their own language on phones and computers.' Below this text is a blue link that says 'LEARN MORE ABOUT UNICODE'. On the left side of the page, there is a vertical navigation menu with the following items: 'Adopt a Character', 'Emoji', 'Basic Info', 'News', 'Events', 'Connect', 'Membership', and 'Press'. Each item has a blue plus sign next to it.

**Unicode**

U+30ED U+1302 U+2666 U+043E U+2690 U+2B07

U+FF0F U+2B1C U+1F490 U+2639 U+263B U+13EA

U+1F643

U+1F9E1

[LEARN MORE ABOUT UNICODE](#)

U+0994 U+05E9 U+0E07 U+1D25 U+5178 U+0B37

U+203F U+1F60A U+03CE U+127F U+060F U+1557

Adopt a Character +

Emoji +

Basic Info +

News

Events

Connect +

Membership +

Press



# **UNICODE: 1 MILLION SYMBOLS**

**FULL COVERAGE FOR 90  
LANGUAGES**

**BASIC COVERAGE FOR 200  
LANGUAGES**

# UNICODE: SUCCESSFUL PROCESS FOR IMPROVEMENT

## WELCOME TO THE SCRIPT ENCODING INITIATIVE

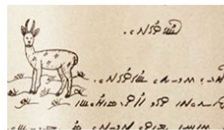
The Script Encoding Initiative (SEI), established in the [UC Berkeley Department of Linguistics](#) in April 2002, is a project devoted to the preparation of formal proposals for the encoding of scripts and script elements not yet currently supported in Unicode (ISO/IEC 10646).

[Unicode](#) is the universal computing standard specifying the representation of text in all modern software. To date, Unicode has largely focused on the major modern scripts, particularly those scripts most widely used in business. Some minority and historic scripts have already been encoded, as well as historic characters of the major modern scripts.



Over [100 scripts remain](#) to be encoded. Minority scripts are still used in parts of South and Southeast Asia, Africa, and the Middle East. Unencoded scripts include Kpelle and Loma. Scripts of historical significance include Book Pahlavi, Large Khitan, and Jurchen. Even for major modern scripts there are many difficult historical issues remaining to be addressed: for example, the

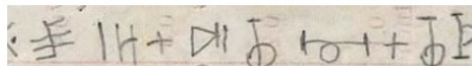
The goal of the SEI project is to fund the preparation of script proposals that will be successfully approved by the Unicode Technical Committee and WG2 (ISO/IEC 10646) without requiring extensive revision or involvement of the committee itself.



A secondary goal to encourage the creation of freely-available Unicode-conformant fonts. This will help to promote widespread adoption and implementation of the scripts.

By providing funding for proposal authors, drawn from faculty and graduate students as well as other experts, the Script Encoding Initiative represents a concerted effort to tackle the remaining scripts and remaining script issues. The project will be assisted by a Unicode Vice President to assure that the proposals meet requirements of the Unicode Technical Committee and of the international standards community. To date, the project has helped get over [70 scripts encoded](#).

The Script Encoding Initiative project is of world-wide importance, for minority and historic scripts. For a minority language, having its script included in the universal character set will help to promote native-language education, universal literacy, cultural preservation, and remove the linguistic barriers to participation in the technological advancements of computing. For historic scripts, it will serve to make communication easier, opening up the possibilities of online education, research, and publication.



# UNICODE IS A SUCCESS STORY IN (AT LEAST) TWO WAYS

## **Inclusive and Representative**

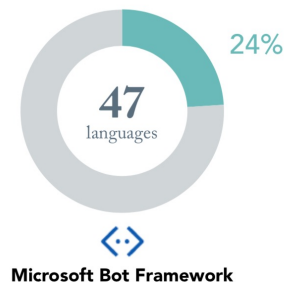
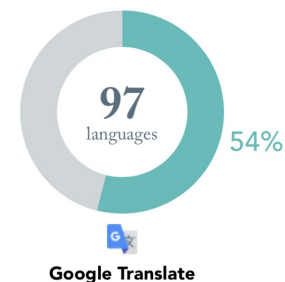
- **Covers languages spoken by at least 95% of people in the world.**

## **Successful Process for Improvement**

- **Unicode has an open-source process by which scholars and speakers of small languages can propose additional scripts.**

# NLP & REPRESENTATION

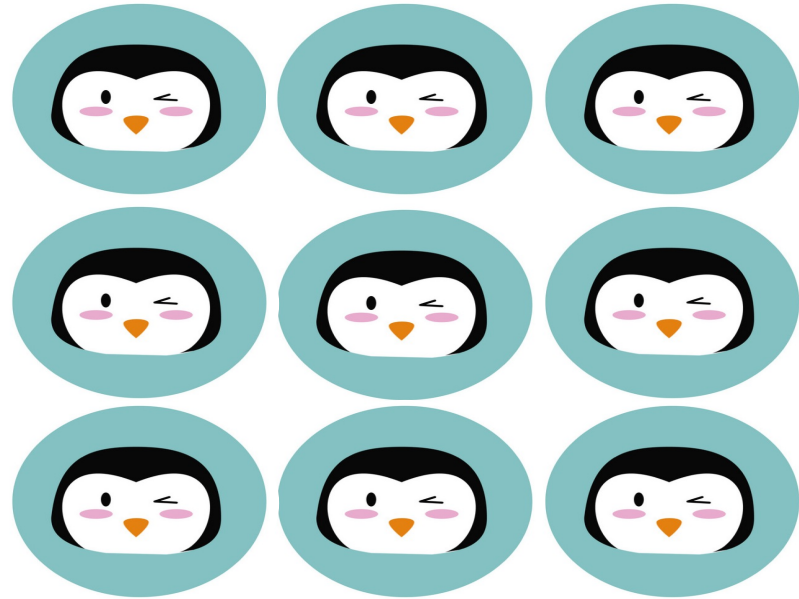
Compare successes of Unicode with web-based NLP, text and translation services, which work best for well-resourced languages and people



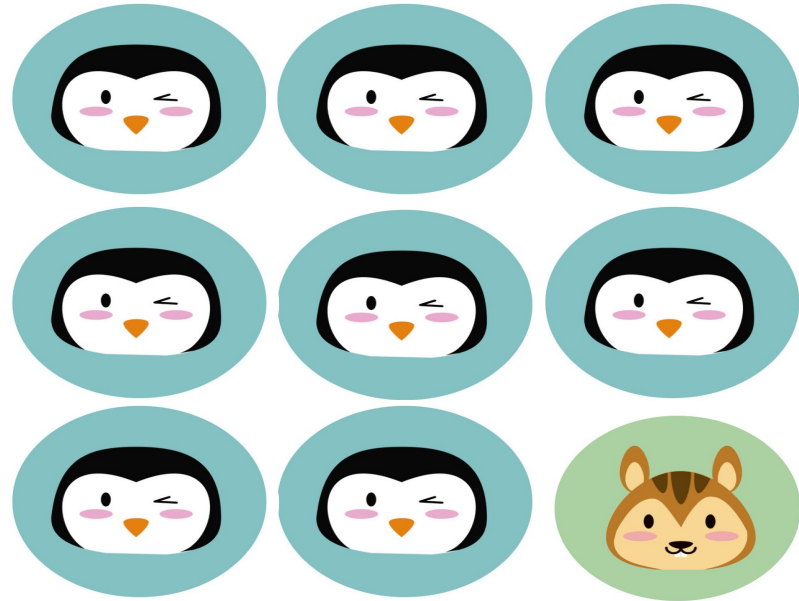
Percentage of those living on less than \$1.50/day whose language is covered by the tool

SOURCE: World Bank, Ethnologue and software provider websites

# REPRESENTATION, SURVEYS, AND LONG TAILS IN DATA

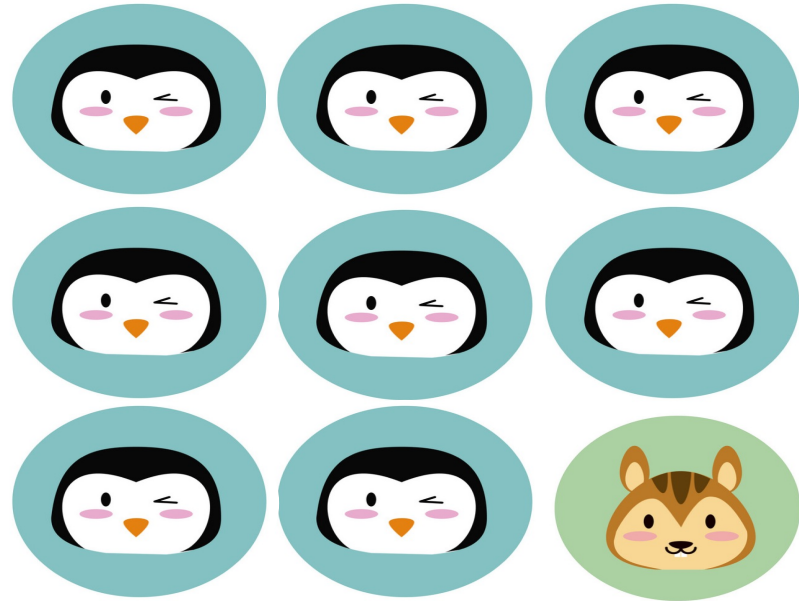


# REPRESENTATION, SURVEYS, AND LONG TAILS IN DATA



If I had a different experience, my numerical ratings may not affect the survey data – but they matter!

# REPRESENTATION, SURVEYS, AND LONG TAILS IN DATA



I want to see myself represented in search results, but other people are using different terms to describe me!

**WHAT KIND OF HARM IS  
LACK OF  
REPRESENTATION?**



# Distributional or Allocative Harm:

How should things or outcomes be distributed?



# Equality of Opportunity:

**Everyone has the Same Access to Pursue the Good.**



# Equality of Outcome:

Everyone gets the same good things  
(and the same responsibilities)



# REPRESENTATIONAL HARMS

- **Am I represented in this system?**
- **Can I express myself in it?**
- **Does this system represent me, my culture, and my self-expression?**



**THANK YOU!**

**You can always get in touch at [kcreel@stanford.edu](mailto:kcreel@stanford.edu)  
Or plan a visit to my office hours at  
[calendly.com/kathleencreel](https://calendly.com/kathleencreel)**