

# Ling/CSE 472: Introduction to Computational Linguistics

---

4/18

Societal Impact of NLP

# Overview

---

- Stakeholder-focused typology of risks of NLP & voice technology
- Value Scenarios
- Reading questions

# Typology

---

- A systematic classification of phenomena, along one or more dimensions
- Helps to explore the space of possibilities
- Helps to understand relationships across categories

Prev work: Hovy & Spruitt 2016,  
Lefevre-Halftermeyer et al 2016

# Hovy & Spruitt 2016

## “The Social Impact of Natural Language Processing”

---

- Survey of some types of issues
- Importantly raised awareness of the discussion within English-language NLP circles
- Introduced concepts of:
  - Exclusion, Overgeneralization, Bias confirmation, Topic Overexposure, Dual use
  - Illustrated with NLP-specific examples of negative impacts
- Not exhaustive, not a typology

# Guiding principles: Sociolinguistics

(e.g. Labov 1966, Eckert & Rickford 2001)

---

- Variation is the natural state of language
  - Variation in pronunciation, word choice, grammatical structures
- Status as ‘standard’ language is a question of power, not anything inherent to the language variety itself
  - Language varieties & features associated with marginalized groups tend to be stigmatized
- Meaning, including social meaning, is negotiated in language use
- Our social world is largely constructed through linguistic behavior

# Guiding principles: Value sensitive design

---

- Value sensitive design (Friedman et al 2006, Friedman & Hendry 2019):
  - Identify stakeholders
  - Identify stakeholders' values
  - Design to support stakeholders' values

# Stakeholder-centered typology

---

**Table 1 Typology of possible harms of language technology**

	<b>Direct stakeholders</b>	<b>Indirect stakeholders</b>
Tech use	User, by choice	Harm to individual
	User, not by choice	Harm to community
Tech development	Annotator or crowdworker	Unwitting data contributor

(D'Arcy & Bender 2023, p.58)





# W Language technologies you use, but not by choice

Total Results: 0

Powered by  **Poll Everywhere**

Start the presentation to see live content. For screen share software, share the entire screen. Get help at [pollev.com/app](https://pollev.com/app)







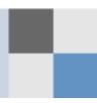

# Language technologies used by others that affect you

W

Total Results: 0

Powered by  **Poll Everywhere**

Start the presentation to see live content. For screen share software, share the entire screen. Get help at [pollev.com/app](https://pollev.com/app)



# Direct stakeholders: By choice

---

- *I choose to use this voice assistant, dictation software, machine translation system...*
  - ... but it doesn't work for my language or language variety
    - Suggests that my language/language variety is inadequate
    - Makes the product unusable for me
  - ... but the system doesn't indicate how reliable it is
    - Users reliant on machine translation/auto-captioning for important info left in the dark about what they might be missing

# Direct Stakeholders: By Choice

---

- *I choose to use this information retrieval system...*
  - ... but the presentation of the results juxtaposed against the question I have in mind obscures important information

Busy & Sick At the Same Time? Schedule Online Or Walk Into An Urgent Care Clinic. We're Proud To Offer A Refreshingly Friendly, Fast and Convenient Experience. Open 8a-8p Daily.

Call (206) 430-7570

Ad · www.justanswer.com/ask-nurse

### Ask Online Nurses | 24-Hour Online Nurse Helpline | JustAnswer.com

When You Don't Feel Good, Stay Home & Ask RNs from Bed. Chat Online 24/7. A Doctor Will...

# 855.520.9500

#### 24-Hour Nurse Care Line

Our nurses can help you decide whether you need to see a doctor, go to an **urgent care clinic** or **emergency** room or **care** for yourself at home. Call 855.520. 9500 to speak with one of our nurses now.

UW Medicine > aco > welcome-guide  
**Urgent Care | UMP Plus – UW Medicine Accountable Care Network**

About Featured Snippets Feedback

#### PEOPLE ALSO ASK

Is there a Ask a Nurse hotline I can call? ▾

What is a 24/7 nurse line? ▾



medicare nurse hotline

ALL NEWS SHOPPING MAPS IMAGES

# 1-800-633-4227

You can also call 1-800-MEDICARE (1-800-633-4227). TTY users can call 1-877-486-2048.

Medicare.gov > pdf PDF  
**10969- Medicare & Home Care - Medicare.gov**

About Featured Snippets Feedback

#### PEOPLE ALSO ASK

Does Medicare have a nurse line? ▾

What is the phone number to ask a nurse? ▾

Is there someone I can call for medical advice? ▾

What is the Medicare benefits helpline? ▾

Feedback

Medica > wellness > medicaid > nurs...

What is the Medicare benefits helpline? ▾

Feedback

Medica > wellness > medicaid > nurs...

### Nurse Line - Medica

NurseLine™ by HealthAdvocate<sup>SM</sup> services are available for Medicare and Medicaid members 24-hours a day, seven days a week.

(866) 715-0915

Aetna Medicare > live-well > ask-nu...

### Aetna Medicare Nurse Hotline: Ask a Nurse 24/7 | Aetna Medicare

Dec 11, 2019 · Call 1-800-556-1555 (TTY: 711) anytime. If you need urgent or emergency care, call 911 and/or your doctor immediately.

UHC.com > health-care-tools > 24-h...

### Your 24-Hour Nurse | UnitedHealthcare

With 24-Hour Nurse, you can get advice from a registered nurse - anytime, 24/7. Just call and you can ask a nurse your questions – ...

Q1Medicare.com > PartD-Get-Help-...

### 1-800-MEDICARE Helpline - Medicare and You Handbook - A Guide to Medicare

At Medicare, we are always working to improve our service to you. The 1-800- MEDICARE (1-800-633-4227) helpline has a speech- ...

# Direct stakeholders: Not by choice

---

- *My screening interview was conducted by a virtual agent*
- *I can only access my account information via a virtual agent*
- *Access to a 911 system requires interaction with a virtual agent first*
  - ... but it doesn't work or doesn't work well for my language variety
    - I scored poorly on the interview, even though the content of my answers was good
    - I can't access my account information or 911

# Direct stakeholders: Not by choice

---

- *LM (language modeling) technology can now generate very real sounding text, in English at least* (Radford et al 2019)
  - ... but which is not grounded in any actual relationship to facts
    - I mistake the text for statements made by a human publicly committing to them
    - I become more distrustful of all text I see online
- Language models trained on ‘standard’ or ‘official’ sounding documents will sound ‘standard’ or ‘official’.

# Direct stakeholder, Tech Development

---

- Exploitative labor practices (Fort et al 2011; Gray & Suri, 2019; Perrigo 2023)
- Psychological harm from traumatic content work (Lefeuve-Halftermeyer et al 2016)

# Indirect stakeholders: Harm to individual

---

- *Someone searched for me online*
  - ... but the ethnicity associated with my name triggered display of negative ads including my name (Sweeney 2013)
- *Someone searched for critics of the government*
  - ... and found my blog post/tweet
- *Someone put my words into an MT system*
  - ... which got the translation wrong and led the police to arrest me (*The Guardian*, 24 Oct 2017; <https://bit.ly/2zyEetp>)



# Indirect stakeholders: Subject of query

---

## Facebook

- Sor Facebook translates 'good morning' into 'attack them', leading to arrest

ative

- Sor Palestinian man questioned by Israeli police after embarrassing mistranslation of caption under photo of him leaning against bulldozer



# Indirect stakeholders: Harm to individual

---

- *Someone designed a system to classify people by identity characteristics according to linguistic features*
  - Information I thought I was presenting only in some venues is made available in others
  - Identity characteristics are attributed to me that are false, but believed (over and above my own assertions) because “the system said so”

# Indirect stakeholders: Harm to communities

---

- *Virtual assistants are gendered as female and ordered around*
- *Systems are built using general webtext as a proxy for word meaning or world knowledge*
  - ... but general web text reflects many types of bias (Bolukbasi et al 2016, Caliskan et al 2017, Gonen & Goldberg 2019)
    - My restaurant's positive reviews are underrated because of the name of the cuisine (Speer 2017)
    - My resume is rejected because the screening system has learned that typically "masculine" hobbies correlate with getting hired
    - My image search reflects stereotypes back to me

# Indirect stakeholders: Harm to communities

- *System knowledge*

The image shows a Google search interface for the term "doctor". The search bar contains the word "doctor". Below the search bar are navigation tabs for "All", "Maps", "Images" (which is selected), "News", "Videos", "More", "Settings", "Tools", "Collections", and "Safe". There are also filter buttons for "female", "cartoon", "clip art", "patient", "stethoscope", "animated", and "world".

The search results are displayed in a grid of seven images, each with a caption and a source URL:

- Image 1:** A smiling male doctor in a white coat with a stethoscope. **Caption:** Well Connection | MyBlue [myblue.bluecrossma.com](http://myblue.bluecrossma.com)
- Image 2:** A male doctor in a light blue shirt sitting at a desk with a laptop. A green checkmark icon is overlaid on the image. **Caption:** How to Use HP Print and Scan Doctor fo... [support.hp.com](http://support.hp.com)
- Image 3:** A male doctor in a white coat with arms crossed. **Caption:** How to Spot a Bad Doctor | MD ... [mdmag.com](http://mdmag.com)
- Image 4:** A male doctor in a white coat examining a baby held by a woman. **Caption:** Tips for Choosing a Doctor - Scripps He... [scripps.org](http://scripps.org)
- Image 5:** A group of six diverse doctors in white coats standing together. **Caption:** Looking for a Doctor - LCMS Member ... [lcmedsoc.org](http://lcmedsoc.org)
- Image 6:** A male doctor in blue scrubs examining a young girl who is holding a stuffed animal. **Caption:** Why Does the Doctor Do That ... [webmd.com](http://webmd.com)
- Image 7:** A male doctor in a white coat with a stethoscope, looking confused with his hands out. **Caption:** Do doctors understand test results ... [bbc.com](http://bbc.com)

name

ed

# Indirect stakeholders: Tech development

---

- *ASR doesn't caption my words as well as others'*
  - My contributions are rendered invisible to search engines
- *Language ID systems don't identify my dialect*
  - Social-media based disease warning systems fail to work in my community (Jurgens et al 2017)
- *My creative output or social media posts are appropriated as training data*
  - Without my consent, without compensation

# Who's job is this?

---

- **Speech/language tech researchers & developers:** build better systems, promote systems appropriately, educate the public
- **Procurers:** choose systems/training data that match use case, align task assigned to speech/language tech system with goals
- **Consumers:** understand speech/language tech system output as the result of pattern recognition, trained on some dataset somewhere
- **Members of the public:** learn about benefits and impacts of speech/language tech and advocate for appropriate policy
- **Policy makers:** consider impacts of pattern matching on progress towards equity, require disclosure of characteristics of training data

# How can we empower people to do those jobs?

---

- Documentation of data sets and models trained on them (Thursday)
- Methodologies for thinking through how technology might interact with social systems (e.g. value sensitive design)
  - identifying the people/communities likely to be impacted
  - eliciting their input (e.g. Diverse Voices <https://techpolicylab.uw.edu/project/diverse-voices/> )
  - thinking through scenarios (e.g. <http://www.envisioningcards.com/>, value scenarios)

# Value Scenarios

---

- Design Scenarios (Rosson & Carroll 2003): Tell the story of how the product, once developed, will be used. Focus on user, typically with happy outcomes.
- Value Scenarios (Nathan et al 2007): Tell the story of how the product, deployed pervasively over time, will impact society. Focus on both users and other stakeholders, imagine what could go wrong.



# Value Scenarios: Elements

---

- Stakeholders
- Pervasiveness
- Time
- Systemic Effects
- Value Implications

# Envisioning cards

---

- Think of a language technology that is in use now or could be in the near future
- Use the prompt on your envisioning card to imagine what could happen if the technology is widely adopted
- Discuss with partner
- If time: share out

# Reading questions

---

- Does 'appropriation' in this context mean the use of a technology that is outside what it is supposed to be used for?
- The two value scenarios seem to center around pretty ambitious/advanced technology? What might a value scenario have looked like for current technology that is already widely used? How visible are these systemic effects?
- Do companies often do an evaluation of the consequences of their technologies, similar to what is demonstrated with the value scenarios? If so, at what stages of the design process?

# Reading questions

---

- The two Value Scenarios posted in this article seem awfully pessimistic, even though there is a fair shot that the scenario they present could play out. Are all Value Scenarios typically this gloomy? I'm sure there are other more positive scenarios that could play out in both of the situations presented, but not fluffed up to the extent of the SBD scenarios that make it seem like nothing could go wrong. I also don't doubt the usefulness of these kinds of techniques for considering the consequences a certain action could have, but I would almost think that if they were this gloomy no one would ever make a decision.

# Reading questions

---

- The article says that value scenarios are meant to have a dark/pessimistic perspective (or design noir, as they call it) in order to counter the over-the-top optimism that comes with most discussion of new technologies. I understand that sentiment, but these two examples seem aggressively cynical--could a value scenario include a second outlook with a brighter perspective? Maybe presenting two sides or versions of a similar story, that way people can compare and decide which elements seem most likely? Since they intend this to be used in public discussions about the ethics of new technologies, it seems like a creator of a scenario could benefit from having multiple perspectives.

# Reading questions

---

- When we say that a "noir portrayal" counterbalances a tendency to focus on positive aspects, do we mean that these Value Scenarios are meant to be neutral overall? Or are they meant to have a negative tone so that we can specifically focus on only the harmful aspects of the technology?

# Reading questions

---

- In the reading, it discussed about two scenarios that are very hard to look at. The impact of both of them is bad to say the least. This reminded of the launch of Apple's AirTag. The product is a perfect functional product that did allow many users to track their items and find them when they are lost. However, the more serious problem is that this device can be used to stalk people. Even though Apple did try to stop that, it wasn't successful in many sense. I am wondering whether or not we should have legislations that force a un-launch of a product or services if the harm outweigh the benefit especially related to criminal offences.

# Reading questions

---

- I was most confused on how LLMs are trained. Bender, Gebru et al. mentions how NVIDIA's MegatronLM had 8.3B parameters but what does this really mean? What is a parameter? Also the dataset size of 174 GB seems small. I would have thought that LLMs are trained on much more text than 174GB worth. Lastly, what is a high-level overview of how LLMs like those mentioned in the article are trained?



# Reading questions

---

- I'm curious if there exist any studies as to the positive (if any) environmental impacts of LLMs through any reduced energy consumption their operation enables. For example, the use of a chatbot customer service system precludes the need for a team of customer service agents driving gas-powered vehicles to work every day and using more electronic devices - can we quantify/ballpark estimate the potential benefits such a system could provide?

# Reading questions

---

- Regarding size not guaranteeing diversity: Is there a world where we can categorize this data by groups and adjust the proportions so we get a more ideal spread within our training data that is representative of real life (this could be done by discarding less-diverse material to get better proportions)? The goal with this method being a newfound ability to use larger corpuses regardless of their existing biases by making them a little less large so that they are optimally diverse.

# Reading questions

---

- On "The Risk of Racial Bias in Hate Speech Detection" - what are some steps we could potentially take to ensure that automatic hate speech detection models can account for differences in individual sensitivities to offensive language, rather than relying solely on a fixed set of criteria that may not be applicable/relevant to everyone?

# NLP/Compling in the news

---

- <https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/>
- <https://www.caidp.org/cases/openai/>
- <https://twitter.com/60Minutes/status/1647742247444553732?s=20>
- <https://medium.com/@emilymenonbender/google-ceo-peddles-ai-hype-on-cbs-60-minutes-4a0e080ef406>