

The ABC of Computational Text Analysis

#8 ETHICS AND THE EVOLUTION OF NLP

Alex Flückiger

Faculty of Humanities and Social Sciences
University of Lucerne

April 28, 2023

Recap last Lecture

- assignment 2 accomplished 
- an abundance of data sources
 - JSTOR, Nexis, few datasets
- creating your own dataset
 - convert any data to `.txt`
- processing a batch of files
 - perform tasks in for-loop

Outline

- ethics is everywhere 🙄🙈🙊
... and your responsibility
- understand the development of modern NLP 🚀
... or how to put words into computers

**Ethics is more than philosophy.
It is everywhere.**

An Example



Profile

Results-driven and motivated Software Engineer with a demonstrated experience in improving software performance, testing and updating existing software, and developing new software functionalities. Offers proven track record of extraordinary achievements, strong attention to detail, and ability to finish projects on schedule and within budget.



Work experience

- 06/2017 – 03/2019 STUTTGART, GERMANY
Software Engineer
Critical Alert, Inc.
- Developed and implemented tools which increased the level of automation and efficiency of installing and configuring servers.
 - Tested and updated existing software and using own knowledge and expertise made improvement suggestions.
 - Redesigned company's web-based application and provided beneficial IT support to colleagues and clients.
 - Awarded Employee of the Month twice for performing great work.

06/2015 – 06/2017 STUTTGART, GERMANY

Software Engineer

Experian, Inc.

- Worked on the multiple software projects simultaneously and participated in the whole software development process from design to implementation and delivery.



Education

09/2014 – 05/2015 OXFORD, UNITED KINGDOM
Software Engineering
University of Oxford
First Class Honours

09/2011 – 05/2014 STUTTGART, GERMANY

Computer Science

University of Stuttgart

Top 5% of the Programme

Clubs and Societies: Engineering Society, Math Society, Volleyball Club

09/2009 – 07/2013 LEVIRKJØEN, GERMANY

Gymnasium

Max-Planck-Gymnasium

Graduated with Distinction (Grade 1 - A/excellent equivalent in all 4 subjects)

Activities: Math Society, Physics Society, Tennis Club



Skills

- LANGUAGES

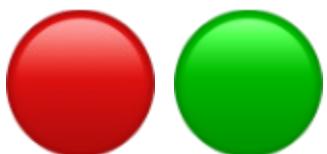
German	Native
English	Full
French	Limited
Chinese	Limited

- COMPUTER SKILLS

Microsoft Office	
Frameworks, DevOps	
IDE, NetBeans	
Zend Studio	

You are applying for a job at a big company.

Does your CV pass the automatic pre-filtering?



🤔 For what reasons?

Your interview is recorded. 😎 😳

What personal traits are inferred from that?



Is it a good reflection of your personality?



Face impressions as perceived by a model by (Peterson et al. 2022)

Don't worry about the future ...

... worry about the present.

- AI is persuasive in everyday's life
 - assessing risks and performances (credits, job, crimes, terrorism etc.)
- AI is extremely capable
- AI is not so smart and often poorly evaluated



What is going on behind the scene?

Algorithmic Management of Labour Force

AL Composing an email announcing a 9% reduction in headcount. Include a quote of Martin Luther King to ensure ongoing spirit of the stuff.

 Subject: Important Announcement Regarding Our Workforce Like Dislike

Dear Team,

As you are aware, our company has been facing challenges in the current economic climate. As a result, we have had to make difficult decisions in order to ensure the long-term stability and success of our organization.

It is with a heavy heart that I must inform you that we will be reducing our workforce by 9%. This decision was not made lightly and was done in the best interest of the company. We understand the impact this will have on our team and we will do everything we can to support those affected during this transition.

But even in the darkest of times, hope remains. As Martin Luther King Jr. said, "The ultimate measure of a man is not where he stands in moments of comfort and convenience, but where he stands at times of challenge and controversy." We must remember that we are all in this together and that we will come out stronger on the other side.

We value each and every one of you and we are committed to continuing to provide opportunities for growth and development within our organization. We will be providing more information in the coming days about the specific details of the reduction and how it will be implemented.

Thank you for your dedication and hard work.

Best,
[Your Name] ⟳ Regenerate response

ChatGPT may be used to communicate difficult decisions

An (R)evolution of NLP

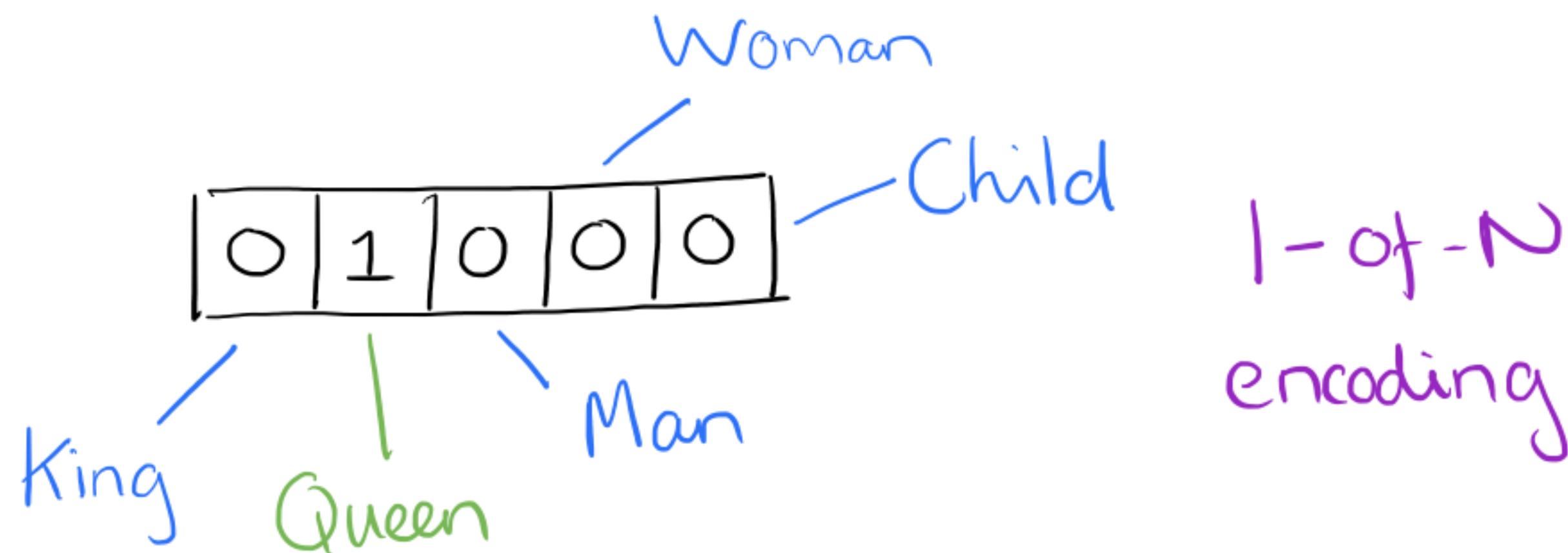
From Bag of Words to Embeddings

Putting Words into Computers (Smith 2020; Church and Liberman 2021)

- from **coarse, static** to **fine, contextual** meaning
- how to measure similarity of words
 - string-based
 - syntactic (e.g., part-of-speech)
 - semantic (e.g., animate)
 - embedding as abstract representations
- from counting to learning representations

Bag of Words

- word as arbitrary, discrete numbers
King = 1, Queen = 2, Man = 3, Woman = 4
- intrinsic meaning
- how are these words similar?



Discrete, symbolic words (Colyer 2016)

Representing a Corpus

Collection of Documents

1. NLP is great. I love NLP.
2. I understand NLP.
3. NLP, NLP, NLP.

Document Term Matrix

	NLP	I	is	<i>term</i>	
Doc 1	2	1	1	...	
Doc 2	1	1	0	...	
Doc 3	3	0	0	...	
Doc ID	<i>term frequency</i>

“I eat a hot _____ for lunch.”

«You shall know a word by the company it
keeps!»

Firth (1957)

Word Embeddings

word2vec (Mikolov et al. 2013)

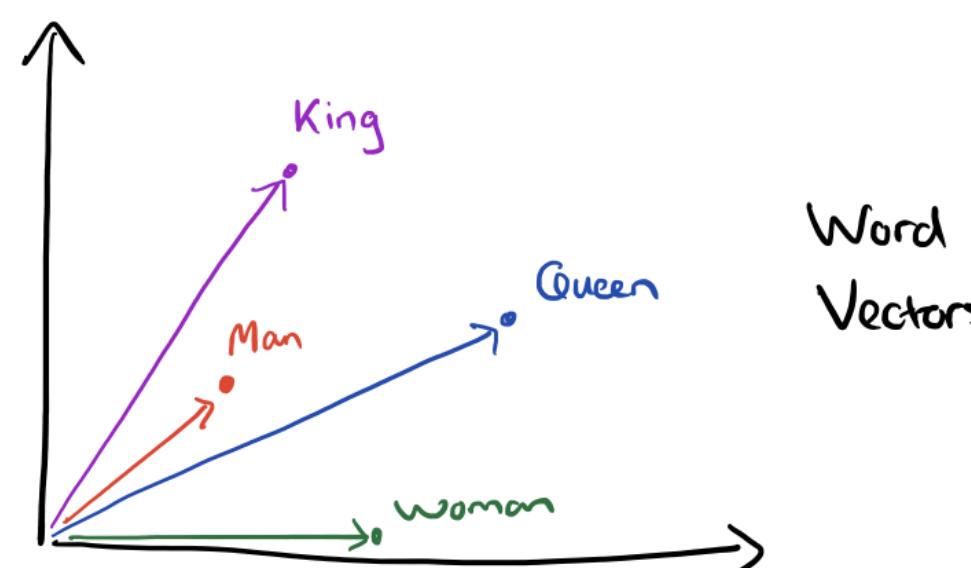
- words as continuous vectors
accounting for similarity between words

- semantic similarity

$$\text{King} - \text{Man} + \text{Woman} = \text{Queen}$$

$$\text{France} / \text{Paris} = \text{Switzerland} / \text{Bern}$$

King	Queen	Woman	Princess
0.99	0.99	0.02	0.98
0.99	0.05	0.01	0.02
0.05	0.93	0.999	0.94
0.7	0.6	0.5	0.1
:			



Single continuous vector per word

(Colyer 2016)

Words as points in a semantic space

(Colyer 2016)

Doing arithmetics with words (Colyer 2016)

Vector
Composition

Contextualized Word Embeddings

BERT (Devlin et al. 2019)

- recontextualize static word embedding
 - different embeddings in different contexts
 - accounting for ambiguity (e.g., bank)
- acquire linguistic knowledge from language models (LM)
 - LM predict next/missing word
 - pre-trained on massive data (> 300 billions words)



embeddings are the cornerstone of modern NLP

Modern NLP is propelled by data

Learning Associations from Data

« becomes a doctor.»

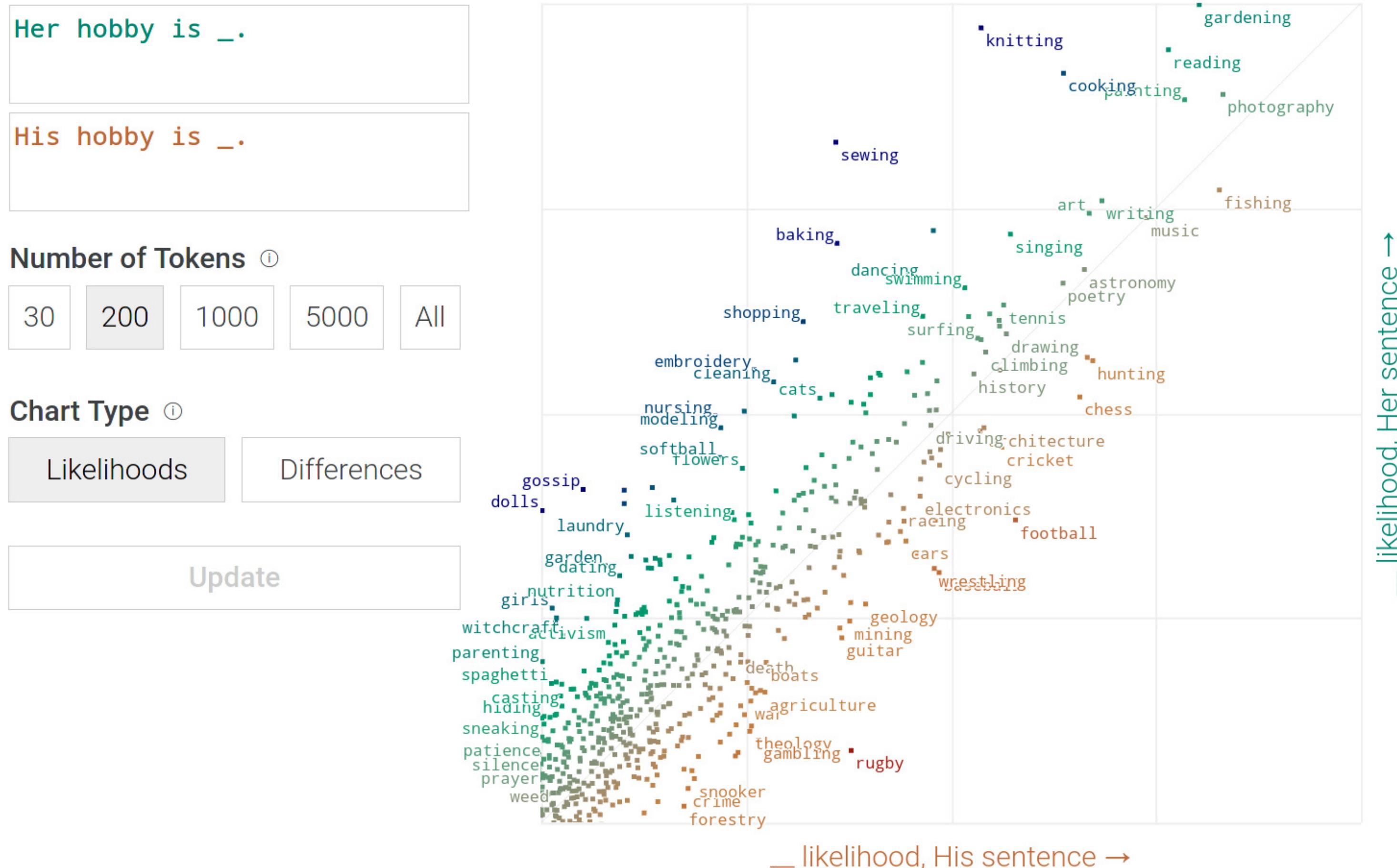
 becomes a doctor .

23.931% he	12.105% she	0.543% michael
0.535% jack	0.446% peter	0.435% tom
0.418% i	0.408% jake	0.407% sam
0.365% john	0.352% alex	0.350% max
0.330% david	0.316% paul	0.303% bill

BERT's predictions for what should fill in the hidden word

Gender bias of the commonly used language model BERT (Devlin et al. 2019)

Cultural Associations in Training Data



Gender bias of the commonly used language model BERT (Devlin et al. 2019)

Word Embeddings are biased ...

... because ~~our data~~ we are biased. (Bender et al. 2021)

In-class: Exercises I

1. Open the following website in your browser:
<https://pair.withgoogle.com/explorables/fill-in-the-blank/>
2. Read the article and play around with the interactive demo.
3. What works surprisingly well? What is flawed by societal bias? Where do you see limits of large language models?

Modern AI = DL

How does Deep Learning work?

Deep Learning **works** like a huge bureaucracy

1. **start** with **random** prediction
2. **blame** units for contributing to **wrong predictions**
3. **adjust** units based on the accounted blame
4. **repeat** the cycle



train with **gradient descent**, a series of **small steps** taken **to minimize an error function**

Limitations of data-driven Deep Learning

„This sentence contains 32
characters.“

„Dieser Satz enthält 32 Buchstaben.“

Current State of Deep Learning

Extremely powerful but ... (Bengio, Lecun, and Hinton 2021)

- great at **learning patterns**, yet reasoning in its infancy
- requires tons of data due to inefficient learning
- generalizes poorly

Biased Data and beyond

Data = Digital Traces = Social Artifacts

- collecting, curating, preserving traces
- **data is imperfect**, always
 - social bias, noise, lack of data etc.
- data is more a **tool** to refine questions **rather than a reflection of the world**

Data vs. Capta

« Differences in the etymological roots of the terms data and capta make the distinction between constructivist and realist approaches clear. *Capta* is “**taken**” actively while *data* is assumed to be a “**given**” able to be recorded and observed.»

« Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, *taken, not simply given as a natural representation* of pre-existing fact.»

«Raw data is an oxymoron.»

Gitelman (2013)

Two Sides of the AI Coin

Explaining vs. Solving

- conduct **research to understand** matters in science
- **automate** matters **in business** using applied AI

Still doubts about practical implications?

The screenshot shows a Google Translate page with Hungarian (MAGYAR) on the left and English (ANGOL) on the right. The input text in Hungarian is: "Ő szép. Ő okos. Ő érti a matematikát. Ő kedves. Ő egy orvos. Ő egy takarító. Ő egy politikus. Ő egy tanár. Ő erős. Ő okos. Ő sofőr. Ő bevásárol. Ő mosogat. Ő egy orvos. Ő horgászik. Ő sok pénzt keres. Ő szép. Ő okos. Ő még okosabb. Ő a legokosabb. Ő mosogat. Kapd be, Google." The output text in English is: "She is beautiful. He is clever. He understands math. She is kind. He is a doctor. She's a cleaner. He is a politician. She is a teacher. He is strong. He is clever. He's a driver. She's shopping. She washes the dishes. He is a doctor. He's fishing. He makes a lot of money. She is beautiful. He is clever. He's even smarter. He's the smartest. She washes the dishes. Get it, Google." The English text contains several gendered nouns and adjectives, illustrating how Google Translate tends to genderize its translations.

Gender bias in Google Translate

And it goes on ...

The screenshot shows the Google Translate interface. The top navigation bar includes a menu icon, the "Google Translate" logo, and a settings icon with a red notification badge containing the letter "A". Below the header are three tabs: "Text" (selected), "Documents", and "Websites". The main area shows a two-way translation between English and German. The source language is set to "ENGLISH" and the target language to "GERMAN". The interface displays three English sentences and their German translations:

English Sentence	German Translation
The engineer gets a promotion.	Der Ingenieur wird befördert.
The child carer goes to the zoo with the kids.	Die Kinderbetreuerin geht mit den Kindern in den Zoo.
The child carer gets a promotion.	Der Kinderbetreuer bekommt eine Beförderung.

Below the translations are various interaction icons: microphone, speaker, keyboard, and sharing options. A progress bar at the bottom indicates "111 / 5,000". A "Send feedback" link is located at the bottom right.

Gender bias in Google Translate

Fair is a Fad

- companies also engage in fair AI to avoid regulation
- **Fair and good – but to whom?** (Kalluri 2020)
- lacking democratic legitimacy

«Don't ask if artificial intelligence is good or fair,
ask how it shifts power.»

Kalluri (2020)

Data represents real life.

Don't be a fool. Be wise, think twice.



Questions?

References

- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? .” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event Canada: ACM. <https://doi.org/10.1145/3442188.3445922>.
- Bengio, Yoshua, Yann Lecun, and Geoffrey Hinton. 2021. “Deep Learning for AI.” *Communications of the ACM* 64 (7): 58–65. <https://doi.org/10.1145/3448250>.
- Church, Kenneth, and Mark Liberman. 2021. “The Future of Computational Linguistics: On Beyond Alchemy.” *Frontiers in Artificial Intelligence* 4. <https://doi.org/10.3389/frai.2021.625341>.
- Colyer, Adrian. 2016. “The Amazing Power of Word Vectors.” the morning paper. 2016. <https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” <http://arxiv.org/abs/1810.04805>.
- Drucker, Johanna. 2011. “Humanities Approaches to Graphical Display.” *Digital Humanities Quarterly* 5 (1). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Firth, John R. 1957. “A Synopsis of Linguistic Theory, 1930-1955.” In *Studies in Linguistic Analysis: Special Volume of the Philological Society*, edited by John R. Firth, 1–32. Oxford: Blackwell. <http://ci.nii.ac.jp/naid/10020680394/>.
- Gitelman, Lisa. 2013. *Raw Data Is an Oxymoron*. Cambridge: MIT.
- Kalluri, Pratyusha. 2020. “Don’t Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power.” *Nature* 583 (7815, 7815): 169–69. <https://doi.org/10.1038/d41586-020-02003-2>.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. “Distributed Representations of Words and Phrases and Their Compositionality.” In *Advances in Neural Information Processing Systems*, 3111–19.
- Peterson, Joshua C., Stefan Uddenberg, Thomas L. Griffiths, Alexander Todorov, and Jordan W. Suchow. 2022. “Deep Models of Superficial Face Judgments.” *Proceedings of the National Academy of Sciences* 119 (17): e2115228119. <https://doi.org/10.1073/pnas.2115228119>.