

# The ABC of Computational Text Analysis

*#1 INTRODUCTION +  
WHERE IS THE DIGITAL REVOLUTION?*

Alex Flückiger

Faculty of Humanities and Social Sciences  
University of Lucerne

03 March 2022

~~#COVID-19~~ 🤔

- Back to normal? How was it, though?
- Let me know when you have some special needs

# Outline

1. digital revolution or hype?
2. about us
3. goals of this course

# AI: A non-standard Introduction



# The world has changed, hasn't it?



*A symbolic image of artificial intelligence (HWZ)*

# An Era of Big Data + AI

# Group Discussion

## What makes a computer looking intelligent?

AI is a moving target with respect to ...

- human capabilities
- technological abilities

# Transfer of Human Intelligence

## from static machines to more flexible devices

- mimicking intelligent behavior
  - reading + seeing + hearing
  - speaking + writing + drawing
- a sense of contextual perception
- many degrees of freedom

# Seeing like a Human?



An image segmentation with Facebook's Detectron2 (Wu et al. 2019)

# Speaking like a Human?

---

## Meena Conversation 4

---

**Human:** *Hi!*  
**Meena:** *Hey! How's it going?*  
**Human:** *How's what going?*  
**Meena:** *Pretty good. How's your day been?*  
**Human:** *Excellent. I spent most of trying to learn to surf.*  
**Meena:** *Sweet! What kind of surfing?*  
**Human:** *The kind where I try to catch a wave and fail*  
**Meena:** *Haha. Do you have a surfboard?*  
**Human:** *I borrowed a friend's, to see if I like it. I'm still a beginner.*  
**Meena:** *Ah. I'm sure you'll get the hang of it.*  
**Human:** *Appreciate the vote of confidence :) Have you ever been surfing?*  
**Meena:** *I have! It's actually a ton of fun.*  
**Human:** *Did you find it hard at first?*  
**Meena:** *A little bit, yes. But once I got the hang of it, it was really fun.*  
**Human:** *That's good to know. Do you go surfing regularly?*  
**Meena:** *Yup. I live in southern Arizona, so there's plenty of surfing to be had.*

---

Chatting with [Google's Meena](#) (Adiwardana et al. 2020)



*Not really, Arizona is not by the sea.*

# Beyond Perception and Unimodality





Generated Images by a Neural Network

<https://thisxdoesnotexist.com/>

Give me *more!*



# Trend towards Multimodality



“a hedgehog using a calculator”



“a corgi wearing a red bowtie and a purple party hat”



“robots meditating in a vipassana retreat”



“a fall landscape with a small cottage next to a lake”



“a surrealist dream-like oil painting by salvador dali of a cat playing checkers”



“a professional photo of a sunset behind the grand canyon”



“a high-quality oil painting of a psychedelic hamster dragon”



“an illustration of albert einstein wearing a superhero costume”

*Breakthrough by combining language processing and image generation with GLIDE (Nichol et al. 2021)*

# Deepfakes? It is real!



“a man with red hair”



“a girl hugging a corgi on a pedestal”

# Demos of *intelligent* Text Processing

## Can you disenchant them?

- sentiment analysis
- question answering
- creating images from text
- text generation
- chat bot
- machine translation
- zero-shot classification
- image captioning

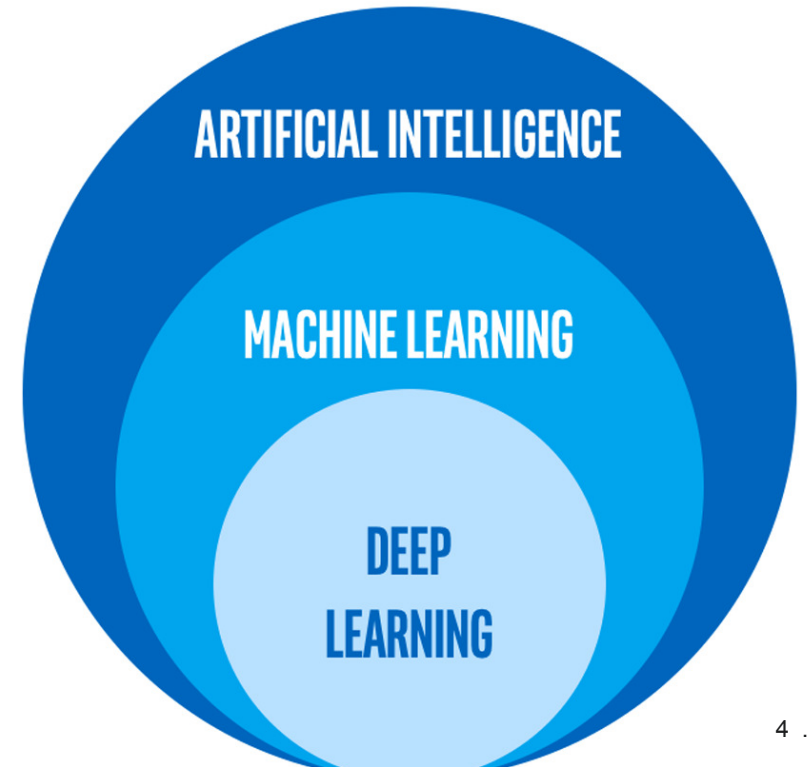
# Artificial Intelligence

## Subfields

- Natural Language Processing (NLP)
- Computer Vision (CV)
- Robotics

# How does Computer Intelligence work?

- interchangeably (?) used concepts  
Artificial Intelligence (AI), Machine Learning (ML), **Deep Learning** (DL)
- generalize **patterns** from lots of data  
more recycling than genuine intelligence  
theory agnostically
- supervised **training** is the most popular  
pairs of input data and outcome



# AI Hype in a Nutshell

```
AI = from humankind import solution
```





# Why this matters for Social Science



# Computational Social Science

## data-driven research

- **computational social science** (Lazer et al. 2009)  
Digital Humanities, Computational History, Data Science
- **highly interdisciplinary**
- **early computational history already in 1960s** (Graham, Milligan, and Weingart 2015)

# Group Discussion

What kind of data is there?

What data is relevant for social science?

- **data as traces of social behaviour**  
tabular, text, image
- **datafication**  
sensors of smartphome, digital communication
- **much of human knowledge compiled as text**

# About the Mystery of Coding

coding is like...

- cooking with recipes
- superpowers

A woman dressed as Superman is flying through a bright blue sky filled with fluffy white clouds. She is wearing the iconic red and blue suit with a yellow 'S' shield on her chest and a red cape that is billowing behind her. Her arms are outstretched, and she has a determined expression. A semi-transparent grey rectangular box is overlaid on the center of the image, containing the text 'Women have coding powers too!' in white, bold, sans-serif font.

**Women have coding powers too!**

# Where the actual Revolution is

Coding is a **superpower** ...

- flexible
- reusable
- reproducible
- inspectable
- collaborative

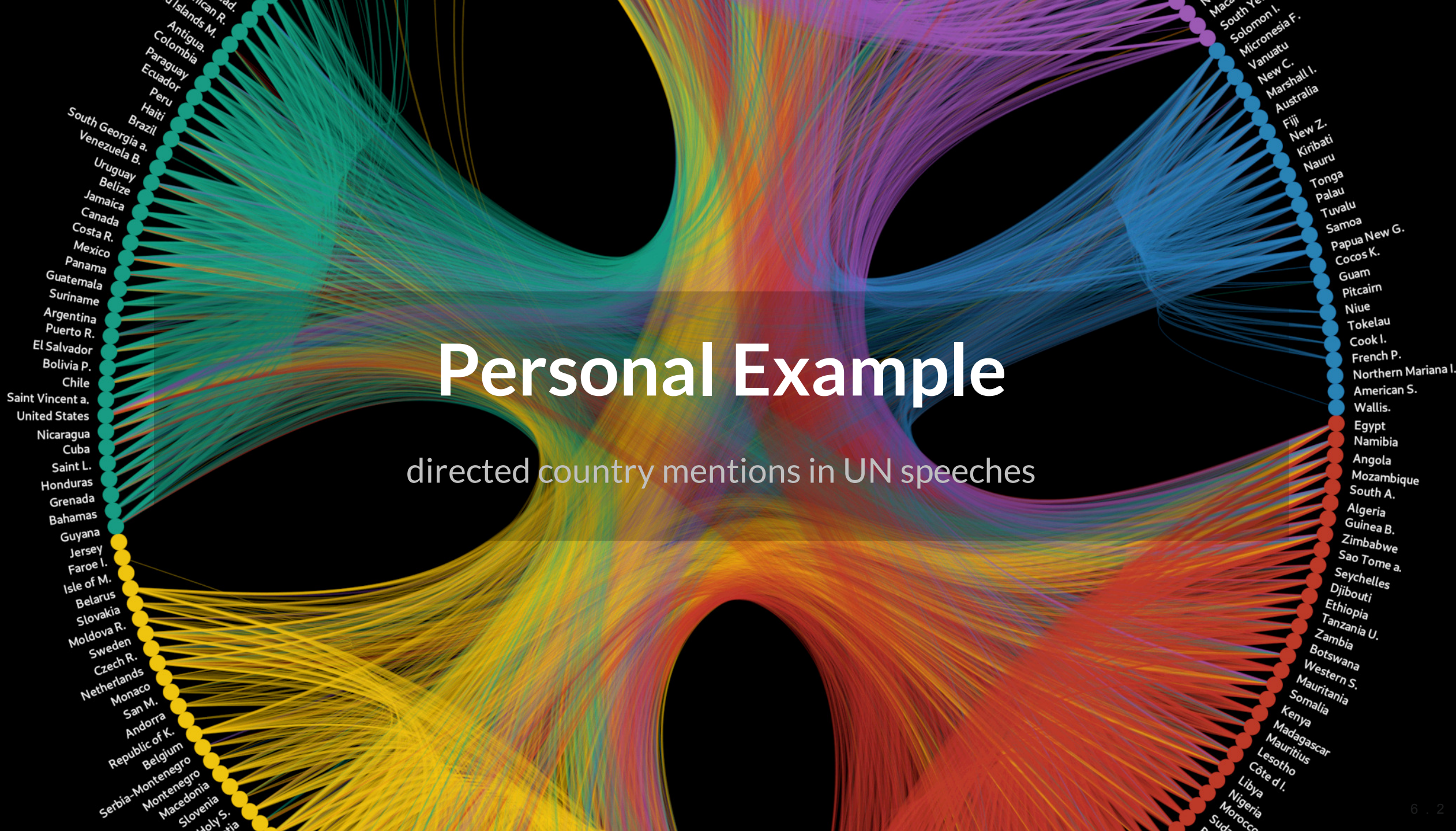
... to tackle complex problems on scale

# About us



# Personal Example

directed country mentions in UN speeches



# Goals of this Course



# What you learn

- **computationally analyze**, interpret, and visualize **texts**  
command line + Python
- **digital literacy** + scholarship
- problem-**solving** capacity

# Learnings from previous Courses

- too much content, too little **practice**
- programming can be overwhelming
- **learning** by **doing**, doing by **googling**

# Levels of Proficiency

1. **awareness** of today's computational potential
2. **analyzing** existing datasets
3. **creating** + analyzing new datasets
4. applying advanced **machine learning**

# What I teach

- computational **practises**
- **critical perspective** on technology
- lecture-style introductions
- hands-on coding sessions
- discussions + experiments in groups

# Topics

## techniques

- text processing
- extracting and aggregating information
- creating simple visualizations
- optical character recognition (OCR)
- scraping files

## data

- using existing resources
- creating new resources



*inputs are more than welcome!*

# Provisional Schedule



Date	Topic
03 March 2022	Introduction + Where is the digital revolution?
10 March 2022	Text as Data
17 March 2022	Setting up your Development Environment
24 March 2022	Introduction to the Command-line
31 March 2022	Basic NLP with Command-line
07 April 2022	Learning Regular Expressions
14 April 2022	Working with (your own) Data
21 April 2022	<i>no lecture (Osterpause)</i>
28 April 2022	Ethics and the Evolution of NLP
05 May 2022	Introduction to Python
12 May 2022	NLP with Python
19 May 2022	NLP with Python + Working Session
26 May 2022	<i>no lecture (Christi Himmelfahrt)</i>
02 June 2022	Mini-Project Presentations + Discussion

TL;DR 

You will be tech-savvy...

...yet no programmer applying fancy machine learning

# Requirements

- no technical skills required   
self-contained course
- laptop (macOS, Win10, Linux)   
update system  
free up at least 15GB storage  
backup files



# Grading

- **3 exercises during semester**  
no grades (pass/fail)
- **mini-project with presentation**  
backup claims with numbers  
work in teams  
data of your interest
- **optional: writing a seminar paper**  
in cooperation with Prof. Sophie Mützel

# Organization

- seminar on Thursday from 2.15pm - 4.00pm
- course website [KED2022](#) with slides + information
- readings on [OLAT](#)
- communication on [OLAT Forum](#)
  - forum for everything except personal
  - subscribe to notifications
  - direct: [alex.flueckiger@doz.unilu.ch](mailto:alex.flueckiger@doz.unilu.ch)

# Who are you?

Please fill out this questionnaire





Questions?

# Reading

## Required

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323(5915):721–23.

(via OLAT)

## Optional

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Open Draft Version. Under contract with Imperial College Press.

[online](#)



# References

- Adiwardana, Daniel, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, et al. 2020. "Towards a Human-like Open-Domain Chatbot." <http://arxiv.org/abs/2001.09977>.
- Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Open Draft Version. Under contract with Imperial College Press. <http://themacroscope.org>.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science." *Science* 323 (5915): 721–23. <https://doi.org/10.1126/science.1167742>.
- Nichol, Alex, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. "GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models." <http://arxiv.org/abs/2112.10741>.
- Wu, Yuxin, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. *Detectron2*. Meta Research. <https://github.com/facebookresearch/detectron2>.