

# The ABC of Computational Text Analysis

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

Alex Flückiger  
Faculty of Humanities and Social Sciences  
University of Lucerne

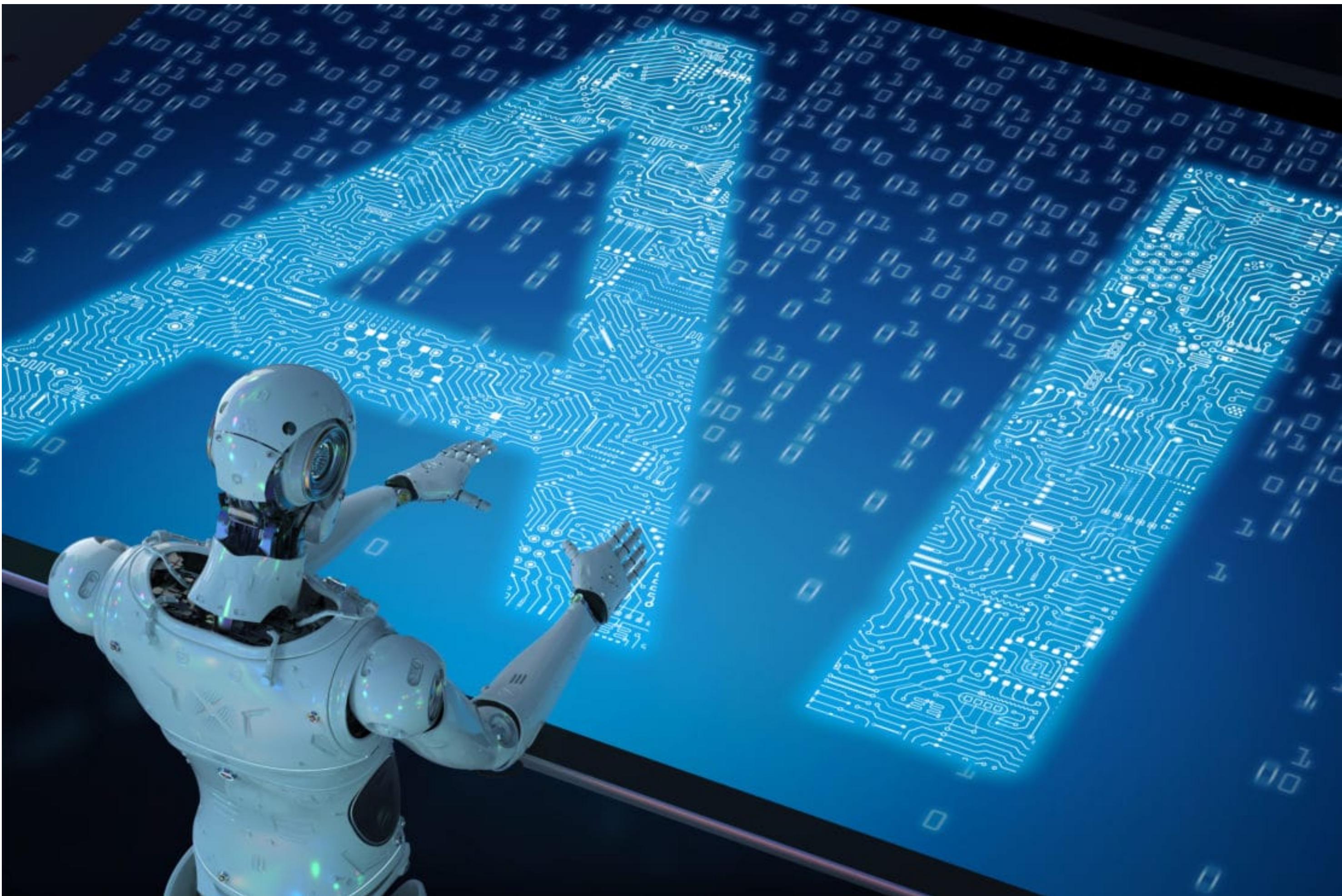
23 February 2023

# 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?



# 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 by Facebook's Detectron2 (Wu et al. 2019)

# Speaking like a Human?

## Speech-to-Text (STT)

Recognizing speech regardless of language, accent, speed, noise etc.

- Check out [samples](#) of Whisper (Radford et al. 2022)
- Check [demo](#) for Swiss German (Plüss et al. 2021)

## Text-to-Speech Synthesis (TTS)

Personalizing voice given an audio sample of 3s

- Check out [samples](#) of VALL-E (Wang et al. 2023)



# Generative and Multimodal AI

# Outsmarting Humans?

# ChatGPT is amazing but ...

... it is also a stochastic parrot.



(Bender et al. 2021)

# Can you disenchant ChatGPT?

## Experiment with ChatGPT

- What works (surprisingly) well?
- When does it fail?

# These People do not exist

Generated Images by a Neural Network (Karras et al. 2020)



Generate *more!*

# Trend towards Multimodality



A storefront with 'Muse' written on it, in front of Matterhorn Zermatt.



A surreal painting of a robot making coffee.



A cake made of macarons in a unicorn shape.



Three dogs celebrating Christmas with some champagne.

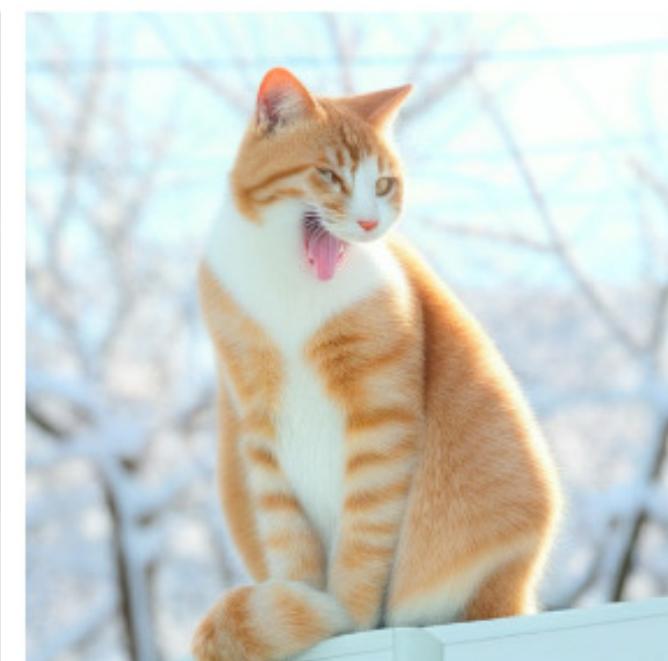
*Breakthrough by combining language processing and image generation with Muse* (Chang et al. 2023)

# Deepfakes? Yes, they are real!

Input image



Editing output



A Shiba Inu

A dog holding  
a football in its  
mouth

A basket of oranges

A photo of a cat  
yawning

A photo of a vase  
of red roses

*Editing pictures with Muse using natural language* (Chang et al. 2023)

# Video is just the last barrier...

**Synthesize any content with ever increasing quality**

- Checkout this [demo trailer](#) for authentic dubbing
- Use words and images to synthesize new videos with [Gen-1](#) (Esser et al. 2023)



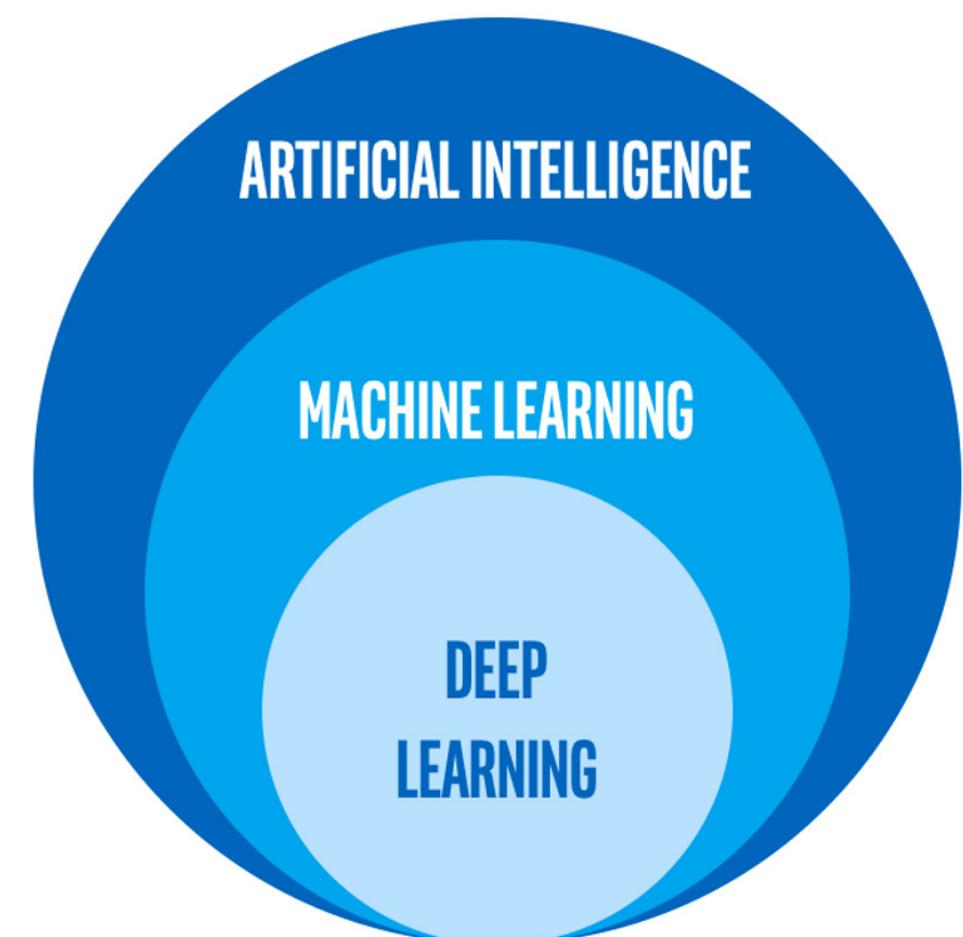
# Artificial Intelligence

## 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)
- learn **patterns** from lots of data  
more recycling than genuine intelligence  
theory agnostically
- supervised **training** is the most popular  
learn relation between input and output



# AI is also Hype

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

All is different to Human  
intelligence

829451, 0.07418429, 0.66673773, 0.98018585, 0.16763814, 0.86710376, 0.55951162, 0.33785509, 0.02626346, 0.47175728, 0.23067162, 0.2773619, 0.1501815, 0.26310512, 0.42061658, 0.77389495, 0.38098379, 0.08868848, 0.46058002, 0.50690262, 0.59905786, 0.77119195, 0.68336732, 0.60541317, 0.4903216, 0.43235588, 0.61449073, 0.24023924, 0.49408374, 0.78123944, 0.33895859, 0.84212152, 0.9432899, 0.217333, 0.35219669, 0.05423672, 0.5926178, 0.72210584, 0.83532963, 0.76463754, 0.16937548, 0.90732891, 0.91315041, 0.10762946, 0.88444707, 0.37388686, 0.76169685, 0.52041133, 0.845968, 0.98120302, 0.83087297, 0.11270352, 0.64186353, 0.04767055, 0.0485364, 0.12084652, 0.16909768, 0.79760446, 0.23634279, 0.98309046, 0.10103919, 0.47973376, 0.77044871, 0.37635039, 0.98989451, 0.42299366, 0.80863832, 0.33989656, 0.14969653, 0.24072135, 0.38481632, 0.07041355, 0.5559498, 0.29417609, 0.05121623, 0.27335799, 0.11510317, 0.12436115, 0.94166874, 0.43521343, 0.01574713, 0.47895682, 0.81542824, 0.55192919, 0.29218659, 0.09174559, 0.42847419, 0.25093282, 0.06408784, 0.89203029, 0.46310012, 0.90179843, 0.80303815, 0.13008473, 0.19678015, 0.99704098, 0.1055286, 0.59973374, 0.45370415, 0.79636357, 0.07858557, 0.13911537, 0.52951605, 0.18976998, 0.5819224, 0.87121155, 0.38853649, 0.87593348, 0.2224633, 0.87825389, 0.29712081, 0.69115835, 0.92883539, 0.66834701, 0.69986855, 0.95275904, 0.87533499, 0.69071387, 0.84586047, 0.46363871, 0.5078105, 0.92830235, 0.25230561, 0.04264319, 0.26313922, 0.09366894, 0.46002723, 0.97870525, 0.14294762, 0.10765255, 0.20673146, 0.47924256, 0.273429, 0.03229992, 0.59000182, 0.079867, 0.09900018, 0.24006025, 0.92607372, 0.86469472, 0.00259478, 0.90787125, 0.81865272, 0.11118454, 0.144895, 0.13555582, 0.28708137, 0.67201359, 0.27947508, 0.95646008, 0.5290694, 0.01663762, 0.72596267, 0.15579892, 0.99991186, 0.46833689, 0.3101146, 0.83803446, 0.1398754, 0.10693991, 0.54277255, 0.03858724, 0.37531373, 0.65770202, 0.70266899, 0.58797343, 0.43545992, 0.46832795, 0.2227064, 0.29304854, 0.80212152, 0.78558867, 0.11125733, 0.55933486, 0.47233518, 0.98970719, 0.07911327, 0.31502887, 0.33420011, 0.30765105, 0.2987607, 0.54778241, 0.69637408, 0.05521608, 0.46128916, 0.12033237, 0.25060302, 0.26122896, 0.78490366, 0.03564119, 0.88678632, 0.89983917, 0.1236243, 0.73271648, 0.73036699, 0.90168095, 0.88147008, 0.11252588, 0.11121773, 0.29328988, 0.18093192, 0.87142164, 0.44432315, 0.54101688, 0.3396327, 0.80099756, 0.0710542, 0.85922124, 0.20631035, 0.19995689, 0.66998863, 0.89721627, 0.48586768, 0.51054386, 0.52999152, 0.22840026, 0.3104349, 0.21021156, 0.07461448, 0.5823082, 0.66050858, 0.98566374, 0.06621486, 0.38370994, 0.96703583, 0.46197723, 0.01985645, 0.09053363, 0.31743002, 0.60602532, 0.6651861, 0.1802219, 0.3075303, 0.5197277, 0.80948972, 0.88438402, 0.0817725, 0.12409288, 0.37352616, 0.19829569, 0.2625787, 0.72964129, 0.9915773, 0.9741447, 0.880090, 0.11729, 0.5197277, 0.727187, 0.741666, 0.07141, 0.06819, 0.50980963, 0.3335821, 0.42666803, 0.26057, 0.021405, 0.7150503, 0.6555555, 0.222062, 0.319997, 0.604774, 0.6486743, 0.054214, 0.03917386, 0.62607333, 0.63521213, 0.4588726, 0.580145, 0.951113, 0.85636, 0.244052, 0.18433, 0.771029, 0.42111, 0.376176, 0.22341028, 0.6692516, 0.98866529, 0.61815532, 0.67766941, 0.67400353, 0.19405055, 0.10825144, 0.8036886, 0.38477034, 0.95783714, 0.68756525, 0.81158131, 0.775627, 0.7560764, 0.32132667, 0.62019336, 0.53422944, 0.92716892, 0.45036567, 0.78308935, 0.14640807, 0.24196815, 0.24140617, 0.68670988, 0.8662444, 0.46714092, 0.47612112, 0.21580927, 0.4307106, 0.5845609, 0.3828225, 0.14634585, 0.34055906, 0.23548654, 0.57817003, 0.9388773, 0.6745013, 0.58375266, 0.91247882, 0.53869737, 0.1429730, 0.3010174, 0.1222801, 0.724284, 0.704626, 0.47868176, 0.72358401, 0.75207502, 0.51178962, 0.31408877, 0.54548918, 0.52262982, 0.9995404, 0.1153715, 0.56617, 0.2007197, 0.6318279, 0.97240863, 0.70431727, 0.78710566, 0.3450736, 0.9487154, 0.30532446, 0.1414234, 0.17664814, 0.15982977, 0.9787621, 0.88858635, 0.67116541, 0.88924914, 0.99750128, 0.34335852, 0.327706, 0.45444521, 0.86250113, 0.51038866, 0.03444767, 0.88713232, 0.98610034, 0.57338999, 0.21504094, 0.08631724, 0.91087582, 0.10086746, 0.26256, 0.98554742, 0.93512302, 0.09476145, 0.3475696, 0.48662246, 0.35574585, 0.02906274, 0.62512557, 0.72779561, 0.29036812, 0.87022702, 0.8434812, 0.23813711, 0.88883248, 0.37898002, 0.32673627, 0.59645067, 0.7234791, 0.93755561, 0.19861349, 0.20419817, 0.03589282, 0.30884502, 0.5734598, 0.19269938, 0.73404843, 0.28849353, 0.93539801, 0.49899802, 0.4781033, 0.2223823, 0.38173886, 0.29305289, 0.67520215, 0.56602555, 0.36684341, 0.33974298, 0.7031872, 0.78609146, 0.4564601, 0.88942624, 0.55393394, 0.90686845, 0.42534314, 0.02765051, 0.93154215, 0.67071709, 0.3774014, 0.3325974, 0.74690781, 0.58257695, 0.56617443, 0.16470804, 0.25931036, 0.86503644, 0.16349719, 0.85850551, 0.36944745, 0.11652745, 0.294993, 0.10651657, 0.14863165, 0.4398803, 0.50674525, 0.27422641, 0.10323303, 0.39854659, 0.60515043, 0.78513435, 0.03993646, 0.87657498, 0.2131398, 0.24072836, 0.51311985, 0.98022862, 0.38983557, 0.86793385, 0.76403785, 0.42765071, 0.87729656, 0.07965547, 0.17378203, 0.35764046, 0.2637499, 0.99506722, 0.62058809, 0.31682388, 0.76524575, 0.42814961, 0.97169199, 0.41251292, 0.9400806, 0.4151149, 0.56583137, 0.39667195, 0.24459122, 0.18682743, 0.77831586, 0.76857211, 0.40289284, 0.55618379, 0.96411916, 0.32744293, 0.98256465, 0.92688416, 0.72177531, 0.06135222, 0.2225768, 0.98614231, 0.07953396, 0.94117132, 0.17013064, 0.63020399, 0.46456359, 0.48314658, 0.12407727, 0.01275128, 0.78190186, 0.68115999, 0.3843999, 0.95577906, 0.94229118, 0.9885269, 0.97376953, 0.50043274, 0.37493048, 0.54529709, 0.57576211, 0.55868575, 0.42363751, 0.9832678, 0.332968, 0.82423524, 0.38948823, 0.16359862, 0.27052009, 0.24686862, 0.69082872, 0.56517825, 0.79584692, 0.94172521, 0.6666855, 0.74609502, 0.52457943, 0.02849603, 0.07278765, 0.29675732, 0.40164173, 0.72768733, 0.77835769, 0.21474951, 0.80307205, 0.88074336, 0.1119304, 0.04230572, 0.4339425, 0.23318203, 0.48700466, 0.14080441, 0.15602402, 0.45488153, 0.2059769, 0.46624392, 0.34737895, 0.75438115, 0.35114957, 0.88838527, 0.56651005, 0.27184795, 0.84393332, 0.2346668, 0.15257668, 0.45093505, 0.47686787, 0.92540229, 0.55753801, 0.85730468, 0.18668896, 0.42050033, 0.7196028, 0.54035591, 0.7964559, 0.50385513, 0.47367413, 0.74437083, 0.22563819, 0.68055333, 0.97841587, 0.99648022, 0.06667557, 0.10857445, 0.0668146, 0.94106309, 0.53926666, 0.75440388, 0.71448794, 0.21650415, 0.69543818, 0.63877204, 0.64120112, 0.78008187, 0.93427463, 0.4444064, 0.3816038, 0.12269671, 0.7188782, 0.09066216, 0.14913997, 0.34626546, 0.74917799, 0.00397471, 0.81133724, 0.56197427, 0.86438034, 0.50195518, 0.52223558, 0.53809704, 0.50945166, 0.83124344, 0.90233734, 0.6017624, 0.34630596, 0.83825859, 0.87407229, 0.17935847, 0.03967245, 0.07094562, 0.5010168, 0.73287606, 0.53367008, 0.41585371, 0.4550718, 0.70436542, 0.01598325, 0.92547509, 0.05948388, 0.59772483, 0.77019315, 0.40899488, 0.0000000

# Why this matters for Social Science

# Computational Social Science

## data-driven research

- **computational social science** (Lazer et al. 2009; Salganik 2017)  
Digital Humanities, Computational History, Data Science
- **highly interdisciplinary**
- **machine learning empowers researchers** (Lundberg, Brand, and Jeon 2022)
- **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 smartphone, digital communication
- much of human knowledge compiled as text

# About the Mystery of Coding

coding is like...

- cooking with recipes
- superpowers



Women have coding  
powers too!

# Where the actual Revolution is

Coding is a **superpower** ...

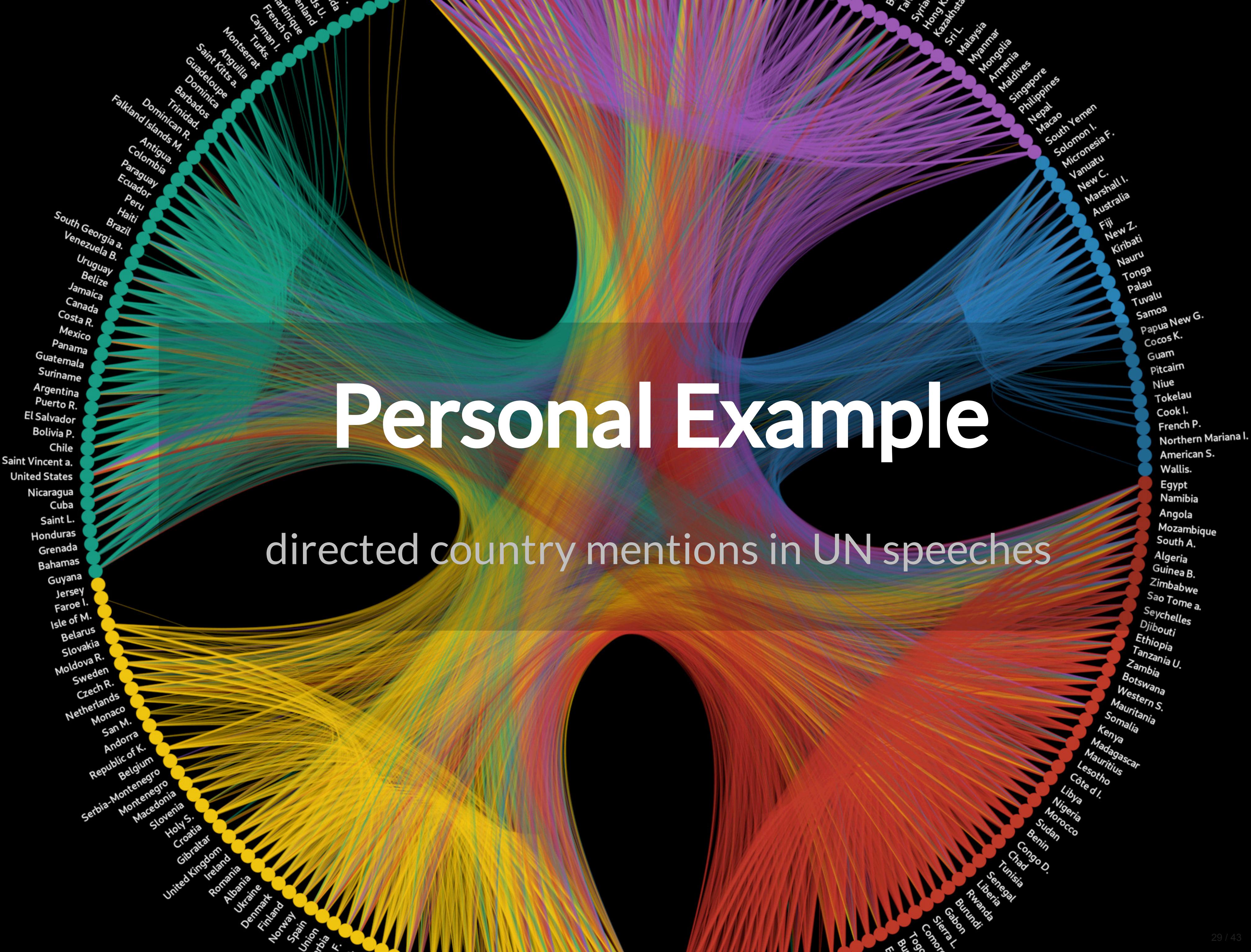
- flexible
- reusable
- reproducible
- inspectable
- collaborative

... to tackle complex problems on scale

# About us

# Personal Example

# directed country mentions in UN speeches



# Goals of this Course

# What you learn

- collect and curate **data**
- **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**

# How I teach

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

# Provisional Schedule

Date	Topic
23 February 2023	Introduction + Where is the digital revolution?
02 March 2023	Text as Data
09 March 2023	Setting up your Development Environment
16 March 2023	Introduction to the Command-line
23 March 2023	Basic NLP with Command-line
30 March 2023 (Zoom)	Learning Regular Expressions
06 April 2023 (Zoom)	Working with (your own) Data
<del>13 April 2023</del>	<i>no lecture (Osterpause)</i>
20 April 2023	Ethics and the Evolution of NLP
27 April 2023	Introduction to Python + VS Code
04 May 2023	Data Analysis of Swiss Media
11 May 2023	NLP with Python
<del>18 May 2023</del>	<i>no lecture (Christi Himmelfahrt)</i>
25 May 2023	NLP with Python II + Working Session
01 June 2023	Mini-Project Presentations + Discussion



There are two digital lectures via Zoom.

TL;DR 

You will be tech-savvy...  
...yet no programmer applying fancy machine learning

# Requirements

- no technical skills required 
- self-contained course
- laptop (macOS, Win11, Linux) 
- update system
- free up at least 15GB storage
- backup files

# Grading



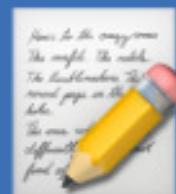
- **3 exercises during semester**  
no grades (pass/fail)
- **mini-project with presentation**  
backup claims with numbers  
work in teams  
data of your interest
- **optional: writing a seminar paper**  
in cooperation with Prof. Sophie Mützel

# Organization

- seminar on Thursday from 2.15pm - 4.00pm
  - additionally, streaming via Zoom
- course website **KED2023** 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

- 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>.
- Chang, Huiwen, Han Zhang, Jarred Barber, A. J. Maschinot, Jose Lezama, Lu Jiang, Ming-Hsuan Yang, et al. 2023. “Muse: Text-To-Image Generation via Masked Generative Transformers.” January 2, 2023. <https://doi.org/10.48550/arXiv.2301.00704>.
- Esser, Patrick, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, and Anastasis Germanidis. 2023. “Structure and Content-Guided Video Synthesis with Diffusion Models.” February 6, 2023. <https://doi.org/10.48550/arXiv.2302.03011>.
- Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian’s Macroscopic*. Open Draft Version. Under contract with Imperial College Press. <http://themacroscopic.org>.
- Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. “Analyzing and Improving the Image Quality of StyleGAN.” March 23, 2020. <https://doi.org/10.48550/arXiv.1912.04958>.
- 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>.
- Lundberg, Ian, Jennie E. Brand, and Nanum Jeon. 2022. “Researcher Reasoning Meets Computational Capacity: Machine Learning for Social Science.” *Social Science Research* 108 (November): 102807. <https://doi.org/10.1016/j.ssresearch.2022.102807>.
- Plüss, Michel, Lukas Neukom, Christian Scheller, and Manfred Vogel. 2021. “Swiss Parliaments Corpus, an Automatically Aligned Swiss German Speech to Standard German Text Corpus.” June 9, 2021. <https://doi.org/10.48550/arXiv.2010.02810>.
- Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. “Robust Speech Recognition via Large-Scale Weak Supervision.” December 6, 2022. <https://doi.org/10.48550/arXiv.2212.04356>.