

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~~ 🤔

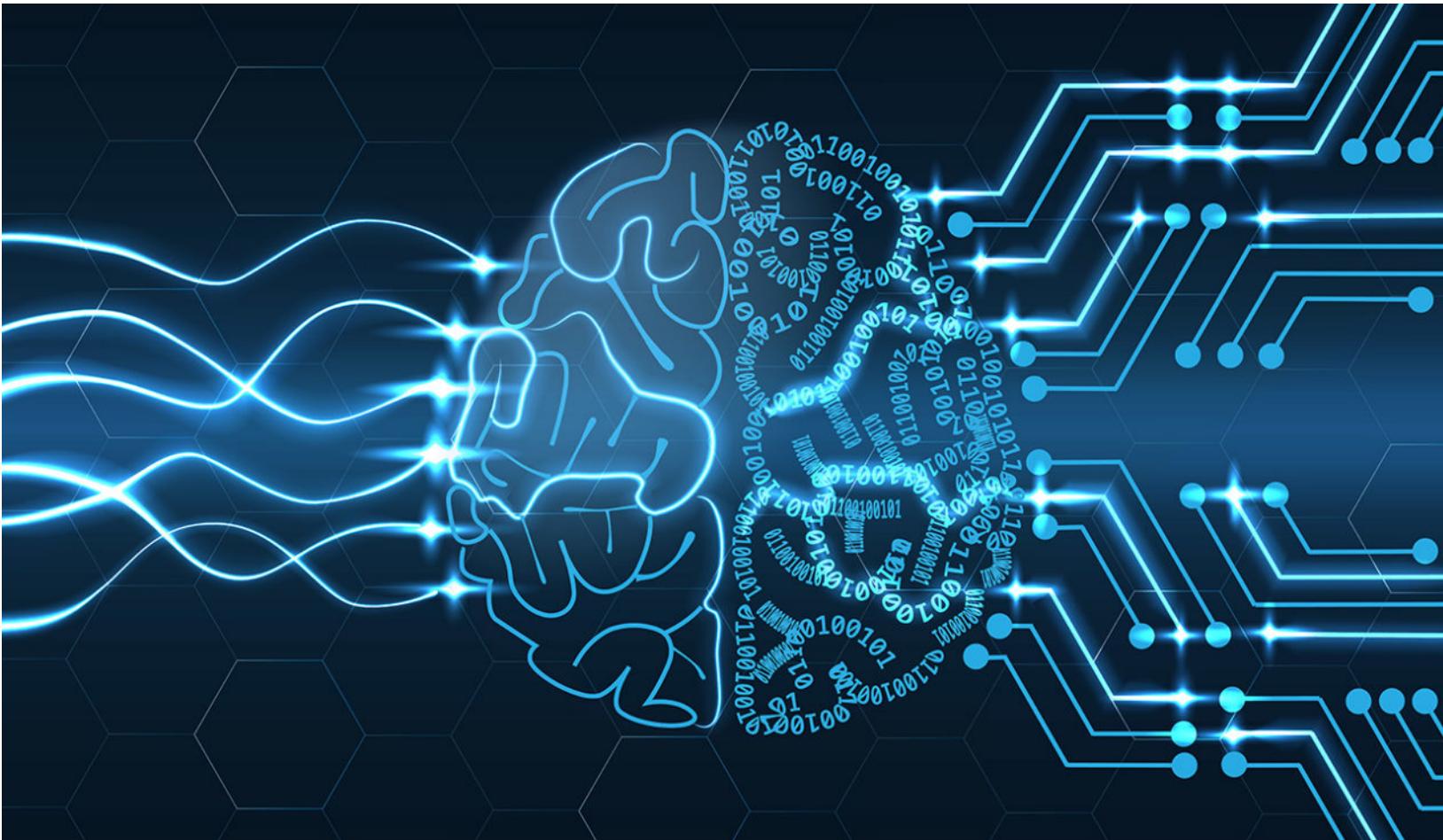
- Back to normal? How was it, though?
- Let me know when you have some special needs

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 dalí 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

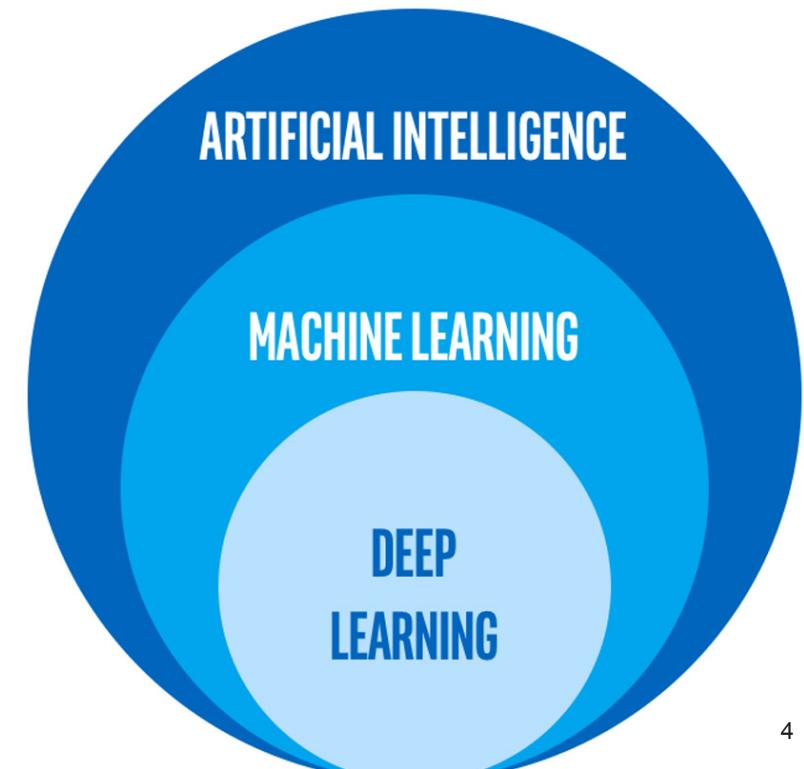
Artificial Intelligence

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
```

This is how current AI looks like

168772, 0.15200756, 0.38829451, 0.07418429, 0.66673773, 0.98018585, 0.16763814, 0.86710376, 0.55951162, 0.33785509, 0.02626346, 0.47175728, 0.23067162, 0.2773619, 0.11454822, 0.80790298, 0.51
734878, 0.11679081, 0.06501815, 0.26310512, 0.42061658, 0.77389495, 0.38098379, 0.08868848, 0.46058002, 0.50690262, 0.59905786, 0.77119195, 0.68336732, 0.60541317, 0.45165225, 0.81831575, 0.31
654246, 0.5173906, 0.06903216, 0.43235588, 0.61449073, 0.24023924, 0.49408374, 0.78123944, 0.33895859, 0.84212152, 0.9432899, 0.217333, 0.35219669, 0.05423672, 0.54828346, 0.33879539, 0.11
479458, 0.32023267, 0.58926178, 0.72210584, 0.83532963, 0.76463754, 0.16937548, 0.90732891, 0.91315041, 0.10762946, 0.88444707, 0.37388686, 0.76169685, 0.52041133, 0.81258545, 0.53919545, 0.41
647513, 0.73535647, 0.1745968, 0.98120302, 0.83087297, 0.11270352, 0.64186353, 0.04767055, 0.0485364, 0.12084652, 0.16909768, 0.79760446, 0.23634279, 0.98309046, 0.19054919, 0.41255494, 0.11
164376, 0.66417077, 0.60103919, 0.47973376, 0.77044871, 0.37635039, 0.98989451, 0.42299366, 0.80863832, 0.33989656, 0.14969653, 0.24072135, 0.38481632, 0.07041355, 0.89827435, 0.03549646, 0.91
712072, 0.45380595, 0.42559498, 0.29417609, 0.05121623, 0.27335799, 0.11510317, 0.12436115, 0.94166874, 0.43521343, 0.01574713, 0.47895682, 0.81542824, 0.55192919, 0.21376429, 0.83996156, 0.11
62373, 0.70650595, 0.78218659, 0.09174559, 0.42847419, 0.25093282, 0.06408784, 0.89203029, 0.46310012, 0.90179843, 0.80303815, 0.13008473, 0.19678015, 0.99704098, 0.10606426, 0.466846, 0.51
806803, 0.67975773, 0.96055286, 0.59973374, 0.45370415, 0.79636357, 0.07858557, 0.13911537, 0.52951605, 0.18976998, 0.5819224, 0.87121155, 0.38853649, 0.87593348, 0.00340097, 0.88076484, 0.41
727308, 0.38427574, 0.67224633, 0.87825389, 0.29712081, 0.69115835, 0.92883539, 0.66834701, 0.69986855, 0.95275904, 0.87533499, 0.69071387, 0.84586047, 0.46363871, 0.50473321, 0.92366178, 0.51
732677, 0.29516387, 0.45078105, 0.92830235, 0.25230561, 0.04264319, 0.26313922, 0.09366894, 0.46002723, 0.97870525, 0.14294762, 0.10765255, 0.20673146, 0.47924256, 0.05069769, 0.0026216, 0.21
910217, 0.00194285, 0.70273429, 0.03229992, 0.59000182, 0.079867, 0.09900018, 0.24006025, 0.92607372, 0.86469472, 0.00259478, 0.90787125, 0.81865272, 0.11118454, 0.179951, 0.36991889, 0.61
286492, 0.04814166, 0.1744895, 0.13555582, 0.28708137, 0.67201359, 0.27947508, 0.95646008, 0.5290694, 0.01663762, 0.72596267, 0.15579892, 0.99991186, 0.46833689, 0.31885293, 0.27407871, 0.61
048462, 0.10562578, 0.59101146, 0.83803446, 0.1398754, 0.10693991, 0.54277255, 0.03858724, 0.37531373, 0.65770202, 0.70266899, 0.58797343, 0.43545992, 0.46832795, 0.99461625, 0.9568131, 0.51
350683, 0.33606992, 0.66227064, 0.29304854, 0.80212152, 0.78558867, 0.11125733, 0.55933486, 0.47233518, 0.98970719, 0.07911327, 0.31502887, 0.33420011, 0.30765105, 0.43608741, 0.8278588, 0.41
871894, 0.8583858, 0.93987607, 0.54778241, 0.69637408, 0.05521608, 0.46128916, 0.12033237, 0.25060302, 0.26122896, 0.78490366, 0.03564119, 0.88678632, 0.89983917, 0.14878763, 0.969423, 0.21
843313, 0.62452422, 0.58236243, 0.73271648, 0.73036699, 0.90168095, 0.88147008, 0.11252588, 0.11121773, 0.29328988, 0.18093192, 0.87142164, 0.44432315, 0.54101688, 0.30702835, 0.31892811, 0.19
356658, 0.06219605, 0.24396327, 0.80099756, 0.0710542, 0.85922124, 0.20631035, 0.19995689, 0.66998863, 0.89721627, 0.48586768, 0.51054386, 0.52999152, 0.22840026, 0.82319605, 0.05364585, 0.61
546367, 0.51318129, 0.18104349, 0.21021156, 0.07461448, 0.5823082, 0.66050858, 0.98566374, 0.06621486, 0.38370994, 0.96703583, 0.46197723, 0.01985645, 0.09053363, 0.38859547, 0.3160512, 0.91
417637, 0.4798936, 0.44743002, 0.60602532, 0.66518613, 0.48021519, 0.30752338, 0.55497277, 0.80948972, 0.88438402, 0.08217725, 0.12409288, 0.37352616, 0.19829569, 0.84911379, 0.52988915, 0.81
978047, 0.59791818, 0.27625787, 0.72964129, 0.99157725, 0.39747549, 0.84112019, 0.36373839, 0.54243265, 0.72739437, 0.74116662, 0.01778544, 0.21686499, 0.50980963, 0.33859405, 0.81940022, 0.81
752336, 0.61757252, 0.96335821, 0.42666803, 0.26057488, 0.0214015, 0.71505503, 0.46556965, 0.42720562, 0.31995973, 0.60427749, 0.64886783, 0.14547148, 0.03917386, 0.65569092, 0.30585473, 0.91
113604, 0.76622595, 0.37607333, 0.62512123, 0.4588226, 0.05808458, 0.95404138, 0.88581666, 0.29140852, 0.11844632, 0.77760193, 0.79442171, 0.06767761, 0.22341028, 0.69664054, 0.70206938, 0.01
718003, 0.40667715, 0.02692516, 0.9884629, 0.5181563, 0.04716941, 0.67400353, 0.11405055, 0.1082114, 0.03688, 0.8471034, 0.9578314, 0.38751625, 0.81150131, 0.93641847, 0.53451291, 0.81
020479, 0.25554169, 0.64775627, 0.756176, 0.321356, 0.0401846, 0.512944, 0.927582, 0.415057, 0.830815, 0.464180, 0.74141451, 0.68701388, 0.87195675, 0.81932992, 0.71
16652, 0.510018, 0.01662444, 0.46714012, 0.761614, 0.15092, 0.301068, 0.515669, 0.8830245, 0.453485, 0.405190, 0.2414854, 0.781100, 0.883013, 0.66448737, 0.59813936, 0.01
517088, 0.27328868, 0.80745013, 0.58375266, 0.91247882, 0.53869737, 0.14297307, 0.53071786, 0.22521863, 0.42502184, 0.70026626, 0.47868176, 0.72358401, 0.75207502, 0.57294178, 0.97597007, 0.91
096594, 0.15940259, 0.76178962, 0.31408877, 0.54548918, 0.52262982, 0.99954042, 0.17529745, 0.66767177, 0.82007397, 0.63618879, 0.97240863, 0.70431727, 0.78710566, 0.88038121, 0.88483492, 0.91
030175, 0.22472301, 0.09450736, 0.9487154, 0.30532446, 0.1414234, 0.17664814, 0.15982977, 0.97876273, 0.88858635, 0.67116541, 0.88924914, 0.99750128, 0.34335852, 0.33949754, 0.4574135, 0.51
522579, 0.07446491, 0.6527706, 0.45444521, 0.86250113, 0.51038866, 0.03444767, 0.88713232, 0.98610034, 0.57338999, 0.21504094, 0.08631724, 0.91087582, 0.10086746, 0.04856606, 0.12266554, 0.41
36731, 0.01896768, 0.4426256, 0.98554742, 0.93512302, 0.09476145, 0.3475696, 0.48662246, 0.35574585, 0.02906274, 0.62512557, 0.72779561, 0.29036812, 0.87022702, 0.81238442, 0.3138893, 0.11
219462, 0.1678605, 0.15434812, 0.23813711, 0.88883248, 0.37898002, 0.32673627, 0.59645067, 0.7234791, 0.93755561, 0.19861349, 0.20419817, 0.03589282, 0.30884502, 0.53066089, 0.34080309, 0.21
207962, 0.5442396, 0.89734598, 0.19269938, 0.73404843, 0.28849353, 0.93539801, 0.49899802, 0.4781033, 0.2223823, 0.38173886, 0.29305289, 0.67520215, 0.56602555, 0.36687071, 0.49260917, 0.61
938278, 0.28331053, 0.93684341, 0.33974298, 0.7031872, 0.78609146, 0.4564601, 0.88942624, 0.55393394, 0.90686845, 0.42534314, 0.02765051, 0.93154215, 0.67071709, 0.30005018, 0.77899115, 0.91
524857, 0.53693479, 0.09774014, 0.3325974, 0.74690781, 0.58257695, 0.56617443, 0.16470804, 0.25931036, 0.86503644, 0.16349719, 0.85850551, 0.36944745, 0.11652745, 0.56587223, 0.26369511, 0.71
842509, 0.62628536, 0.60294993, 0.10651657, 0.14863165, 0.4398803, 0.50674525, 0.27422641, 0.10323303, 0.39854659, 0.60515043, 0.78513435, 0.03993646, 0.87657498, 0.21632641, 0.86778243, 0.01
44423, 0.59041854, 0.88131398, 0.24072836, 0.51311985, 0.98022862, 0.38983557, 0.86793385, 0.76403785, 0.42765071, 0.87729656, 0.07965547, 0.17378203, 0.35764046, 0.24408659, 0.29837998, 0.61
456205, 0.19835696, 0.98637499, 0.99506722, 0.62058809, 0.31682388, 0.76524575, 0.42814961, 0.97169199, 0.41251292, 0.9400806, 0.4151149, 0.56583137, 0.39667195, 0.98073902, 0.54011904, 0.91
350726, 0.46031599, 0.83459122, 0.18682743, 0.77831586, 0.76857211, 0.40289284, 0.55618379, 0.96411916, 0.32744293, 0.98256465, 0.92688416, 0.72177531, 0.06135222, 0.39145084, 0.00926561, 0.31
717906, 0.94337418, 0.77225768, 0.98614231, 0.07953396, 0.94117132, 0.17013064, 0.63020399, 0.46456359, 0.48314658, 0.12407727, 0.01275128, 0.78190186, 0.68115999, 0.35187262, 0.69508548, 0.31
516399, 0.0820243, 0.85843999, 0.95577906, 0.94229118, 0.9885269, 0.97376953, 0.50043274, 0.37493048, 0.54529709, 0.57576211, 0.55868575, 0.42363751, 0.9832678, 0.92408923, 0.44920966, 0.31
338962, 0.19098382, 0.4832968, 0.82423524, 0.38948823, 0.16359862, 0.27052009, 0.24686862, 0.69082872, 0.56517825, 0.79584692, 0.94172521, 0.6666855, 0.74609502, 0.54789156, 0.74805648, 0.21
443603, 0.98945893, 0.56457943, 0.02849603, 0.07278765, 0.29675732, 0.40164173, 0.72768733, 0.77835769, 0.21474951, 0.80307205, 0.88074336, 0.1119304, 0.04230572, 0.4464518, 0.58364118, 0.31
619102, 0.29209694, 0.27339425, 0.23318203, 0.48700466, 0.14080441, 0.15602402, 0.45488153, 0.2059769, 0.46624392, 0.34737895, 0.75438115, 0.35114957, 0.88838527, 0.50659062, 0.36747274, 0.41
104512, 0.76891909, 0.19651005, 0.27184795, 0.84393332, 0.2346668, 0.15257668, 0.45093505, 0.47686787, 0.92540229, 0.55753801, 0.85730468, 0.18668896, 0.42050033, 0.72344403, 0.83930597, 0.91
490435, 0.16136975, 0.34196028, 0.54035591, 0.7964559, 0.50385513, 0.47367413, 0.74437083, 0.22563819, 0.68055333, 0.97841587, 0.99648022, 0.06667557, 0.10857445, 0.0932683, 0.88942261, 0.91
386458, 0.90469276, 0.0268146, 0.94106309, 0.53926666, 0.75440388, 0.71448794, 0.21650415, 0.69543818, 0.63877204, 0.64120112, 0.78008187, 0.93427463, 0.4444064, 0.36709441, 0.70256521, 0.91
79161, 0.27795564, 0.4816038, 0.12269671, 0.7188782, 0.09066216, 0.14913997, 0.34626546, 0.74917799, 0.00397471, 0.81133724, 0.56197427, 0.86438034, 0.50195518, 0.81087441, 0.34517867, 0.01
821386, 0.67496579, 0.40223558, 0.53809704, 0.50945166, 0.83124344, 0.90233734, 0.6017624, 0.34630596, 0.83825859, 0.87407229, 0.17935847, 0.03967245, 0.07094562, 0.55174193, 0.74978665, 0.51
398436, 0.38600355, 0.8301068, 0.73287606, 0.53367008, 0.41585371, 0.4550718, 0.70436542, 0.01598325, 0.92547509, 0.05948388, 0.59772483, 0.77019315, 0.40899488, 0.0556999, 0.47415863, 0.1

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?

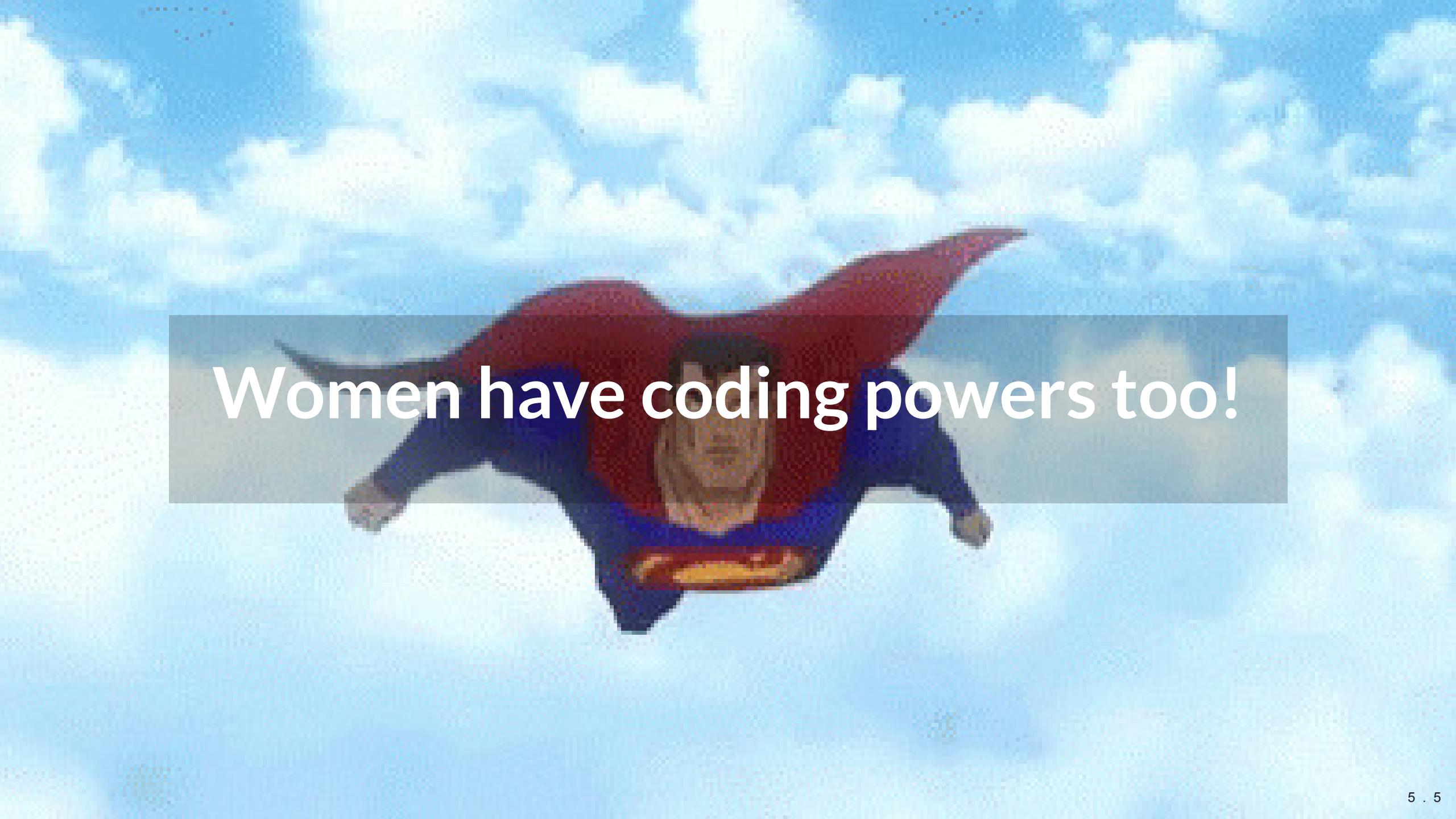
What data is relevant for social science?

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

About the Mystery of Coding

coding is like...

- cooking with recipes
- superpowers

A woman with long dark hair is shown from the waist up, wearing a red and blue superhero-style costume with a gold belt and a long, flowing red cape. She is flying through a bright blue sky filled with white and grey clouds. Her arms are outstretched to her sides, and she has a determined expression. A large, semi-transparent rectangular box covers the middle portion of the image, containing the text.

Women have coding powers too!

Where the actual Revolution is

Coding is a **superpower** ...

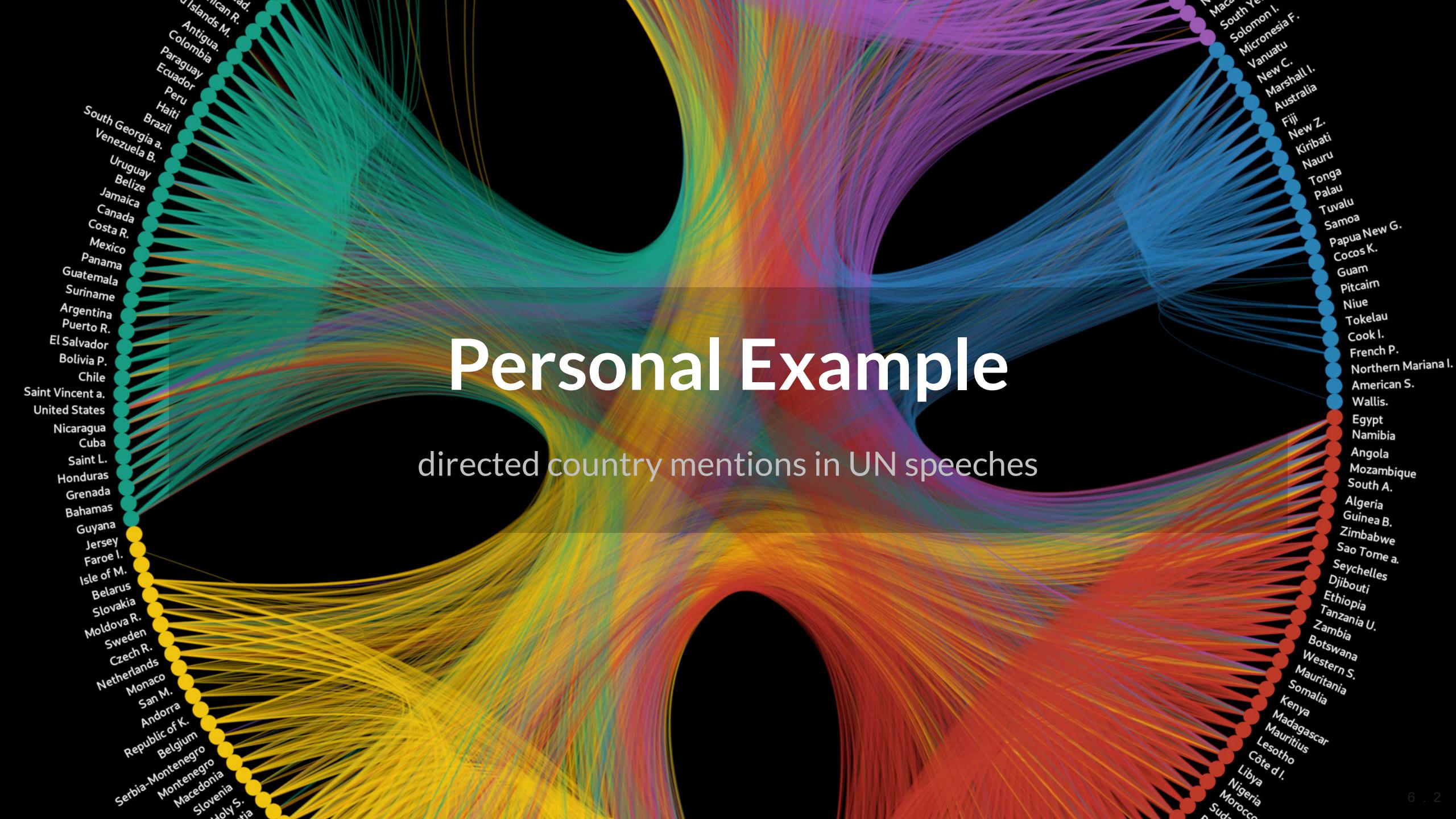
- flexible
- reusable
- reproducible
- inspectable
- collaborative

... 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

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.” February 27, 2020.
<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.” December 22, 2021. <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>.