

The ABC of Computational Text Analysis

#1 INTRODUCTION +
WHERE IS THE DIGITAL REVOLUTION?

Alex Flückiger
Faculty of Humanities and Social Sciences
University of Lucerne

20 February 2025

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 device 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
 - perception: reading + seeing + hearing
 - generation: speaking + writing + drawing
 - moving in the physical world
- flexibility and contextual adaptability
- reproducing any media form

Seeing like a Human?



An image segmentation by Facebook's [Detectron2](#) (Wu et al. 2019)

Hearing and Speaking like a Human?

- **Speech-to-Text** 
robust understanding complex environments (e.g. language, accent, noise) (Radford et al. 2022)
- **Text-to-Speech** 
respect emotions beyond correct pronunciation
- **Speech-to-Speech** 
[voice translation](#) by SeamlessM4T v2 (Duquenne et al. 2023)
[voice cloning](#) by VALL-E (Wang et al. 2023)

Simulating (multiple) Humans?

Beyond linear conversations

- Generate podcasts based on any text
- Generate songs following instructions



Outsmarting Humans?

Debunk some myths around ChatGPT

- is a brand, **large-language models (LLM)** is the technology
- generates fluent text, **not necessarily truthful**
- is highly **useful**, although it understands little
- what is tough for humans **might be easy for the model**; and vice-versa
- is **English**-focused, multi-linguality is limited
- generates **non-reproducible** outputs
- generated text **cannot be detected** (except verbatim parts)
- yesterday's **version** might be different than today's

Where does the smartness come from?

Number of words exposed

- ~100'000'000 for a typical 13-year old kid (Timiryasov and Tastet 2023)
- ~4'300'000'000 words in entire Wikipedia
- >15'000'000'000'000 for current LLMs 🤯 (Grattifiori2024?)

An LLM is amazing but ...

... it is also a stochastic parrot. 

(Bender et al. 2021)

The LLM apocalypse

A post-apocalyptic take on education (Mollick 2023)

- ChatBots are another tool like Google Search
- ChatBots are your personal tutor
 - productivity tool to become faster and better
 - help you to think, not replacing thinking
 - don't trust blindly
 - be transparent

LLMs are a tool, learn how to use it 

The Future of Search

- ChatBots challenge classic search engines

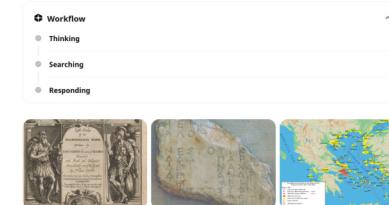
answer with source attribution instead of ranked snippet
blurring the line between search and generation

[Microsoft Copilot](#), [Google Gemini](#), [you.com](#)

- Agents pursuing more and complex tasks

ChatBots are an interface rather than a single model

- explain the background, action, and consequences of the peloponnesian war



Background of the Peloponnesian War

The Peloponnesian War (431–404 BC) was a significant conflict between two of the most powerful city-states in ancient Greece: [Athens](#) and [Sparta](#), along with their respective allies. The roots of the war can be traced back to the [Pentecontaetia](#), a period of tension following the Persian Wars, during which Athens emerged as a dominant naval power and began to expand its influence through the Delian League. This expansion fostered fear and resentment among other Greek city-states, particularly Sparta and its allies, who viewed Athenian imperialism as a direct threat to their autonomy and security.

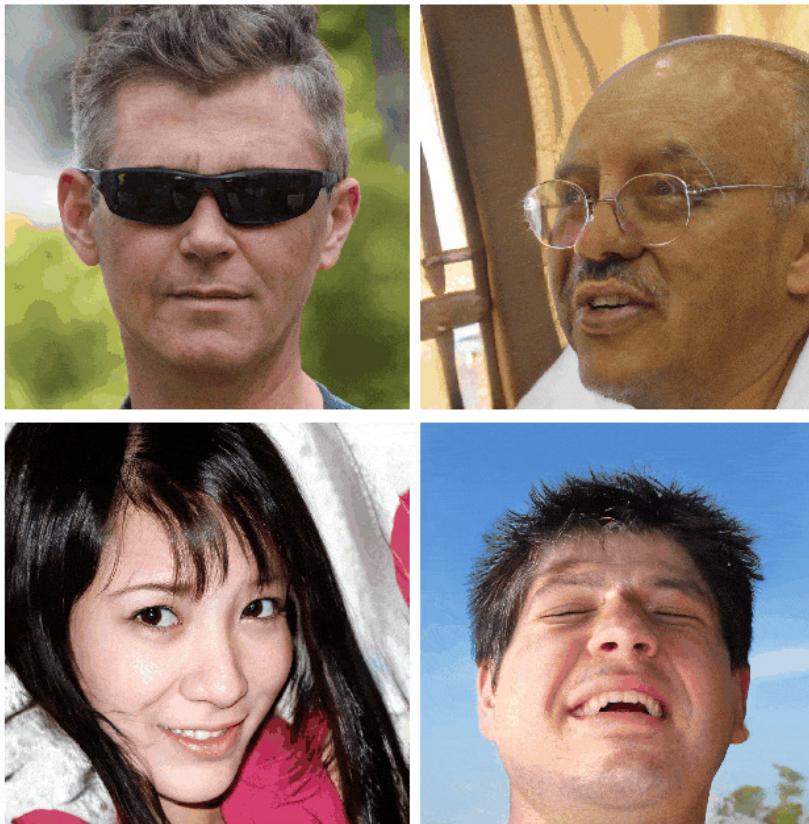
The immediate causes of the war included disputes over trade and alliances, as well as the growing power of Athens, which was perceived as aggressive and expansionist. Tensions escalated when Athens refused to dismantle its defensive walls, which had been a point of contention for Sparta. The conflict was characterized by a struggle for power and dominance in the Greek world, leading to a series of military engagements that would last nearly three decades.

More complex query on [you.com](#)

Trends towards Generative and Multimodal AI

These people do not exist

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



Faces generated by [StyleGAN](#). Generate [more](#) faces!

Multimodality and guidance

Guided generation of text, audio, images, video

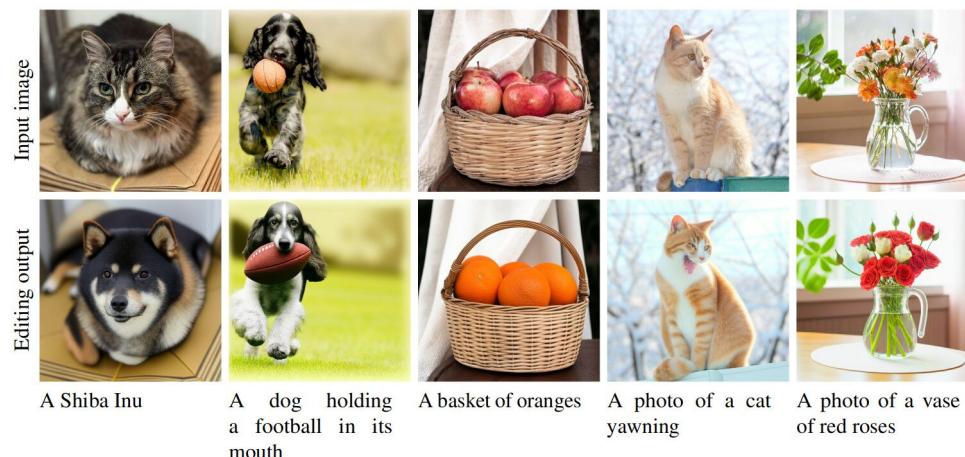
Prompt: Elephant amigurumi walking in savanna, a professional photograph, blurry background



State-of-the-Art image generation model Imagen3. (Imagen-Team-Google et al. 2024)
Free demo [on Gemini platform](#).

Interact with images using text prompts

- **Generate** (Imagen-Team-Google et al. 2024)
- **Edit** (Sheynin et al. 2023)
e.g. virtual **try-on of clothes**
- **Enhancing**
- **Explain**



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

Erase or edit reality

For your Instagram or Politics



Modify pictures thoroughly in [Google Photos](#)

From Image to Video Generation



Synthesize any content with ever increasing quality (Bar-Tal et al. 2024; Brooks et al. 2024)

- use words and images to synthesize videos
Veo 2 constitutes the current research frontier
- real-world dubbing for movies



Performance Google Veo 2 vs OpenAI Sora
(end vs start 2024)

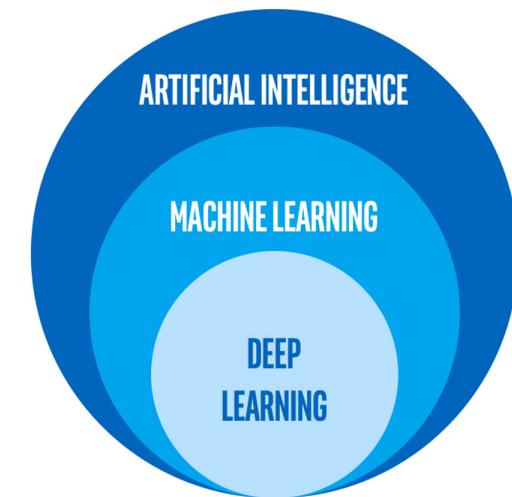
Artificial Intelligence

(Converging) 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
```

AI is different to Human
Intelligence

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, 0.51734878, 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, 0.3654246, 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, 0.1479458, 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, 0.4647513, 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, 0.16164376, 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, 0.9712072, 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, 0.1662373, 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, 0.58086803, 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, 0.47272308, 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, 0.5732677, 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, 0.29910217, 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, 0.6286492, 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, 0.60048462, 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, 0.53350683, 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, 0.48871894, 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, 0.28843313, 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, 0.6356658, 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, 0.546367, 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, 0.417637, 0.4798936, 0.44743002, 0.60602532, 0.617861, 0.24021519, 0.375337, 0.55497277, 0.80948972, 0.843842, 0.0821775, 0.12409288, 0.37352616, 0.19829569, 0.84911379, 0.52988915, 0.8978047, 0.59791818, 0.27625787, 0.72964129, 0.1172, 0.077549, 0.81101, 0.37338, 0.04238, 0.273947, 0.711166, 0.077844, 0.218049, 0.5180963, 0.33859405, 0.81940022, 0.87572336, 0.61757252, 0.96335821, 0.42666803, 0.600548, 0.21401, 0.175050, 0.405696, 0.272052, 0.195953, 0.614277, 0.648683, 0.144440, 0.0317386, 0.65569092, 0.30585473, 0.91113604, 0.76622595, 0.37607333, 0.63521213, 0.4588226, 0.05808458, 0.95404138, 0.88581666, 0.25140532, 0.11844633, 0.77760293, 0.79442171, 0.0657761, 0.22341028, 0.69664054, 0.70206938, 0.00771803, 0.40667715, 0.02692516, 0.98866529, 0.61815532, 0.67766941, 0.67400353, 0.19405055, 0.10825144, 0.8036886, 0.38477034, 0.95783714, 0.68756525, 0.81158131, 0.93641847, 0.53451291, 0.89202479, 0.25554169, 0.64775627, 0.7560764, 0.32132667, 0.62019336, 0.53422944, 0.92716892, 0.45036567, 0.78308935, 0.14640807, 0.24196815, 0.24140617, 0.68670988, 0.87195675, 0.81932992, 0.716652, 0.510018, 0.01662444, 0.46714092, 0.47612112, 0.21580927, 0.43071, 0.5816609, 0.382825, 0.14634585, 0.34055906, 0.23548654, 0.57817003, 0.9388773, 0.66448737, 0.59813936, 0.517088, 0.27328868, 0.80745013, 0.58375266, 0.91247882, 0.53869737, 0.14297, 0.05371, 0.622216, 0.4501, 0.02616, 0.47868176, 0.72358401, 0.75207502, 0.57294178, 0.97597007, 0.9096594, 0.15940259, 0.76178962, 0.31408877, 0.54548918, 0.52262982, 0.99954, 0.01729, 0.56177, 0.00739, 0.61879, 0.97240863, 0.70431727, 0.78710566, 0.88038121, 0.88483492, 0.93030175, 0.22472301, 0.09450736, 0.9487154, 0.30532446, 0.1414234, 0.17664314, 0.15982977, 0.9737627, 0.88838633, 0.6716547, 0.88924914, 0.99750128, 0.34335852, 0.33949754, 0.4574135, 0.52522579, 0.074446491, 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, 0.4336731, 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, 0.219462, 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, 0.20279762, 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, 0.62893278, 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, 0.93524857, 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, 0.7842509, 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, 0.044423, 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, 0.6246205, 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, 0.97350726, 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, 0.33717906, 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, 0.316516399, 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, 0.044920966, 0.338962, 0.19098382, 0.4832968, 0.82423524, 0.38948823, 0.16359862, 0.27052009, 0.24686862, 0.56517825, 0.79584692, 0.94172521, 0.6666855, 0.74609502, 0.54789156, 0.2443603, 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, 0.3619102, 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, 0.404512, 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, 0.0940435, 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, 0.97386458, 0.904269276, 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, 0.79161, 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, 0.821386, 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, 0.74973565, 0.398436, 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, 2021; Salganik 2017)
Digital Humanities, Computational History, Data Science
- new, interdisciplinary approaches to long-standing questions
- 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**
- **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 (ChatGPT?!)**

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

-  There will be 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



- **2 assignments 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 [KED2025](#) with slides + information
- readings on [OLAT](#)
- communication on [OLAT Forum](#)
 - forum for everything except personal
 - subscribe to notifications
 - direct: alex.flueckiger@doz.unilu.ch

Registration via UniPortal

In order to acquire credits for this course, registration via UniPortal within the registration period is mandatory.



Registration period: 5th February – 1st March 2025

Assignment #1



- get/submit via OLAT
 - starting tonight
 - deadline: 1 March 2025, 23:59
- discuss issues on OLAT forum

Course Website





Questions?

References

- Bar-Tal, Omer, Hila Chefer, Omer Tov, Charles Herrmann, Roni Paiss, Shiran Zada, Ariel Ephrat, et al. 2024. "Lumiere: A Space-Time Diffusion Model for Video Generation." January 23, 2024. <https://doi.org/10.48550/arXiv.2401.12945>.
- 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>.
- Brooks, Tim, Bill Peebles, Connor Homes, Will DePue, Yufei Guo, Li Jing, David Schnurr, et al. 2024. "Video Generation Models as World Simulators." <https://openai.com/research/video-generation-models-as-world-simulators>.
- 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>.
- Duquenne, Paul-Ambroise, Brian Ellis, Hady Elsahar, Justin Haaheim, John Hoffman, Hirofumi Inaguma, Christopher Klaiber, et al. 2023. "Multilingual Expressive and Streaming Speech Translation."
- 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>.
- Imagen-Team-Google, Jason Baldridge, Jakob Bauer, Mukul Bhutani, Nicole Brichtova, Andrew Bunner, Lluis Castrejon, et al. 2024. "Imagen 3." December 13, 2024. <https://doi.org/10.48550/arXiv.2408.07009>.
- 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, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. 2021. "Meaningful Measures of Human Society in the Twenty-First Century." *Nature* 595 (7866, 7866): 189–96. <https://doi.org/10.1038/s41586-021-03660-7>.