LLMs in Academia: Best Practices and Common Pitfalls

Mathieu Gravey

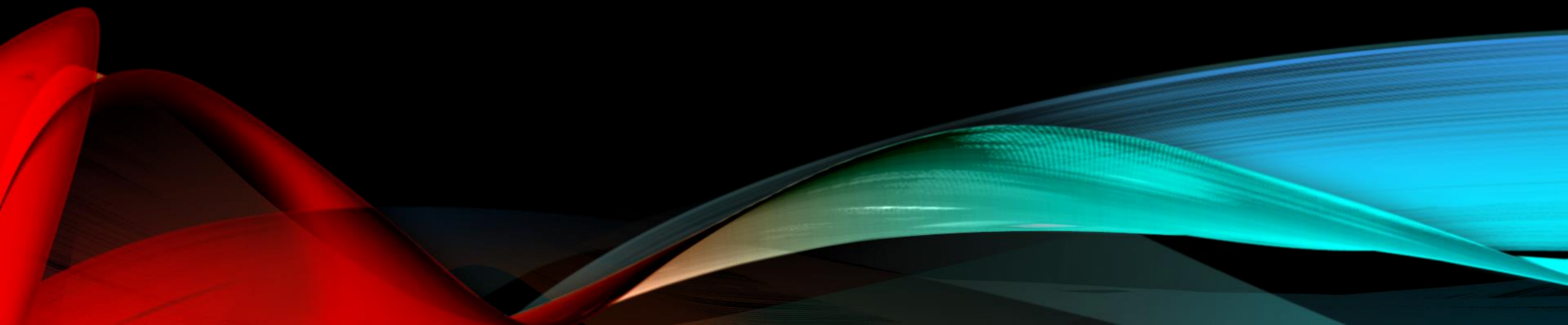
ÖAW
AUSTRIAN ACADEMY OF SCIENCES

What is your added value ?

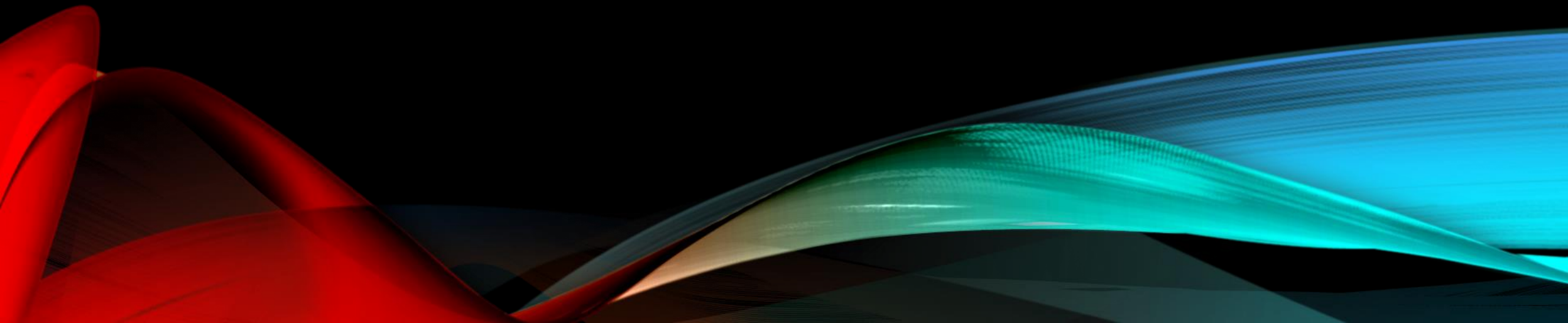
- Who thinks it's their unique writing voice?
- Who thinks we don't care
  it's the science that speaks?

"ChatGPT is an advanced AI-powered **text generator**, designed to assist with diverse tasks by understanding and producing human-like language."

Produced by ChatGPT



The origins



How to produce text

1. Take a sample
2. Try to produce something similar

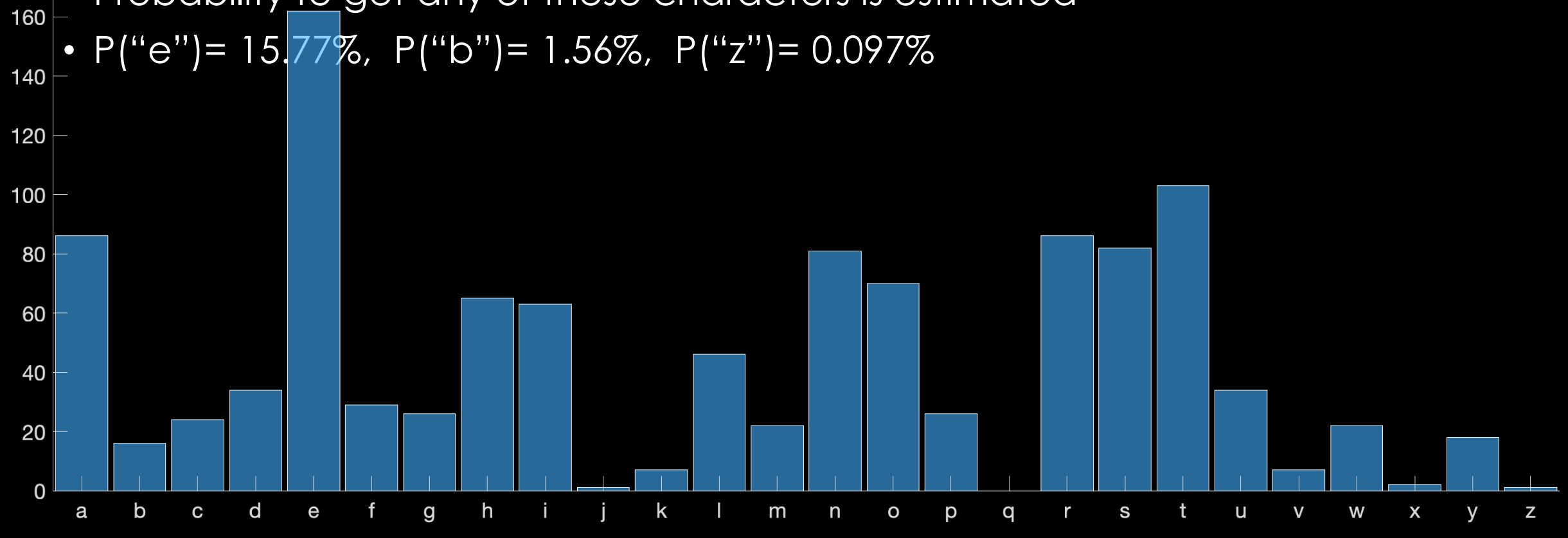
Example of stochastic text

A hundred years ago there were one and a half billion people on Earth. Now, over six billion crowd our fragile planet. But even so, there are still places barely touched by humanity. This series will take to the last wildernesses, and show you the planet and its wildlife as you have never seen them before. Imagine our world without sun. Male Emperor Penguins are facing the nearest that exists on planet Earth - winter in Antarctica. It's continuously dark and temperatures drop to minus seventy degrees centigrade. The penguins stay when all other creatures have fled because each guards a treasure: A single egg rested on the top of its feet and kept warm beneath the downy bulge of its stomach. There is no food and no water for them, and they will not see the sun again for four months. Surely no greater ordeal is faced by any animal. As the sun departs from the Antarctic it lightens the skies in the far north. It's March and light returns to the high Arctic, sweeping away four months of darkness. A polar bear stirs. She has been in her den the whole winter. Her emergence marks the beginning of spring. After months of confinement underground she slides down the slope. Perhaps to clean her fur, perhaps for sheer joy. Her cubs gaze out of their bright new world for the very first time. The female calls them, but this steep slope is not the easiest place to take your first steps.

Training image: Planet Earth(BBC)

Statistics at order 0

- Count all characters in the text
- Probability to get any of these characters is estimated
- P(“e”)= 15.77%, P(“b”)= 1.56%, P(“z”)= 0.097%



Simulation at order 0

cpstaii. hseiao eel.gt oesrw hl hi ew oatd rtoioe,th epehi l bripa yit liesu iraut ni nmi n wetiah.retnttogtfa,hna ea grstet ien.ry mwnfthihinin sual f eie nlaeran ennn ons.sammh yiuihtaldsonien nr aathacodawe pe mlc pdy ibtlen ohn s hh hyernla shnllgt h e rh uawfftdrhpaeudna srsxa an ltes if g.uhte f lesae e. asntemthoil debitisfonueee nnbn dtnh e.hcfnhns eto f.trotr cy s r e tg ewb batrf reet h nsestt ea s lunaesc g f pdptmh th cfh scnof npetnfrt e a tnthibr psrstoeasndnfdt ea.srerar ggginsdheohti giprrnihnprrrsileeraucn no.hre irr utero.gfofih onepn nehr yttowtmwsc ootogiopt otso,s tusereitnn ene gfhcebas ptsval iszdgtolhhtsw aodamtwee ocyeolpel pnuhntoo r ylnfa lleh rtn gtrnd th e r aocene tr tossybntne ta'gl nchclwernonruntmefireihron to nwe nbeoashrnoin.shms nohdh,ei t aiseemi aes..r eaunrra tse cesnee h idoohlhleu ihkh orenmerhsslpni etaifodeldtswewtnrrhn n snr o hgn pin lasoi at 'rohge noresltrnzh trtdb b nrtoytmtmr re.un rsmb i umoonveui lu 'e lhssimn ewnujo nurt oasorwp deteaeo wta tar t,fnotoenss ssahsra nnet oeve ealtgtpen ahintaa rhfohenpu rr et.eofwtfegshpg teo iselnr n etstntptsongeemc aenretefaos eimlhht t eei moueltuidd eriarl lh ltnhoawaesynffeoahrantgrlef: v epwri tin.afhlsayt-n ehgtrhue ei cne

Training image: Planet Earth(BBC)

Statistics at order 1

- The text can be used to analyze all the pairs of letters, and to define the probabilities to find any configuration such as
- P(“aa”) , P(“ab”), P(“ac”), etc.
- One can then compute conditional probabilities
- P(“ab” / ”a”)
- And use it for a simulation algorithm

Simulation at order 1

c, wig. ay. aimoy then thethe then caur wdet achetherig. auig d withs. ac, cay thst cr ay. cheng. acay wilaimouithethe ay c, d the acac, ac, therilaiths croropethstheng. wengay cherorkimorkigauroy. the the cr wd ailitherkig wit ay thethstherknfuigauthe wenfuouithe cauoy. wd che cherknfurkslig thstherken d aig ay t wilay d thetherourke wd cherouthslauimerkenfay. wig. aurksths wd cachs thethenfac, detheths crkn acrilime wig. ay thengautherkithetheng. wer auoy d wd ailaigacaigauourither d ay. acrithethengacauroy d ths aurorofout d crknslacrk t auop aimaigay deths. d wd derouig c, thet then ay ay wethstheroy crke d auilaigaig. cerk achenfacaigac, ay. d acer wigac, thens auopetherofoy. ths t wd the crkig d de cethens ailacrorour ths aimouig thensthe d cauths. caigaithenfuroy. wimerk aurk ths d wdenfacaimouroy aimeropet d wd dens wd ay acacacrouthethen wilacroy. d weths ay the d ailililauourk auiths d thethe acaigaurken wd cheng d werks. thethe c, ay cr de theng d cauouths d cay caur thetherkig. ay thetherk ay d d aigacailay. therimourkethslay. ailailay d d d d thenfoy d thenstherks. werkimopet c, aig thet aigac, cauithen witherkilauit thslaithst aurkn aithen wimay wimofuigay. aig. ther the ay. ac, d theng. the aigailig. d t d cenfuimorke wde cet acrig. d aurouoy. a

Training image: Planet Earth(BBC)

Simulation at order 2

cret nor em, beet lionthe overne pland onthe place ove perne ov willionf tomarts she plane plan he overns not newing onfirs sloplach. he places net not lionthe onthe planderne ove plane onfirs she onfirsteps. he pe. to tic, sloplace places not not lidesteet norld overns she plach againthe placesses nor dowd ane onf ing overnew youn he onfirsteet lideragai tirsteps. he pe. to tomarthe onftre placesteps new wing aging ont not net norld againg an he plach and a plane onfirsteens net lionthe perns sloperades sloperatur ean anderagint not not liont lionfirs norldloplan a she plach a pe. tirs nor emacesse pe. he plach a sloplane plan he perne pe. he pe. timarthe places nor dowdestep slope. he onfthe peraginthe plan he pe. he onf to tomarthe ont norldlight nor dowd aging overagint not not ligromace pe. he overathe pe. he overatur desteen aging agai a pe. he ove plandes she ont lionftrs she perns she onthe onfteps. he place pe. he ov wing onf ticauint lionthe overns norld onthe pe. he ont lionfirsteent net nor ear em, beet lionthe plach. he ov youn agaze onfteps she pe. ticale plane overns norldlionfirs she peradloplace onf timar emace ov weraging ove pe. he pe. he planewy againg ove ove onthe pe. he ont nor ear dowd a sloperns nor emarts she onftre ov your earthe per

Training image: Planet Earth(BBC)

Simulation at order 3

creturnesses, and ther joy depter joy darkng ot surely nor perace to ther joy. ther montinuo ther months of censuae:t ret antarctic,ic but of conths of ther joy ping other months of ther joy. ther joy. ther montinusun alls frop of ther crowd on ther jon perature: ther months of cent undernearts from tope is north -rch guarkserigh and ther joy dept surely denguins ther months of conths of ther jonfergrneater montige. ther months of ther montiget and ther joy. her montin,ing a porar brgrnessest sun agaie will new perhaps ther montinuouslope.hit's cent un deptep slope.haps frop to ther montin,agill neater. aftedp slope is north - perhaps ther crowd its sto ther creater. af hirst sun dhe slope eacing of ther months of ther crout of contigrteaping a peracimal. after.ens ther crowd out eaca.serigroun allst winter montigrout excalf world will new wildern her months of cergrness. she vn her montig, ther cromack stomther monting. her crown ther months of cer creather montin,mal. af her creasurely north - perath - raturns ther joy ping other creasiest planet each gas barested befor ther montinuo might new perhaps frath -ragile.wildergrness. a sing a sinning of ther months of cent uight nfine antarctic bul months of ther monting away winter.ehe slope.hit's marctic before.

Training image: Planet Earth(BBC)

Simulation at order 4

creturns the ferytur months of the feryturns to calescallsothere os cent tnd tee slope is no grze of sontinuo sleepislnot the farstiststaypfhur frigh after.mher dor the facen the fer tubts stoeasieep so tike.planet and sheeplnis no waker cubs gare stirl take yop oo the facing ow eorth. norto. of continus seveasirely this sere:ia light new a solace colls to the far nor she hingle emsrid far nonths of sor penguins and show your first time. the wasi steepinn awl emergrncentigrodnd lhe vlope. penguins are fleand kepa winter ooy. her emergess. sur, and sheep and khe fer crown ohe is no gake to cleater months of sonths.of continuo serinn. a treatures hown ous mot stoep slope.is not steps the facen in antarctica. the far beigraeel juy ani al. cs to call otderneath the wasieep so take to tae bentiur.fiuemen ala as the whnter ories iee in alls chnfing.oafpole is no wake yopr woyld for fur nonfingmto cale emperaaps for for fur, and kheebegg rf cedto the whon confing. aftar menths of spope.is not this seriest placested on the were and kema and kemale cill planet and kept under. her emer cur, perhaps from the were our first this series iee our first this steep and khe fer tures have fled on the wasieet und sheep so time. the ferytfirst stom confing. hfr crowd our fragile

Training image: Planet Earth(BBC)

Simulation at order 7

MEMORY ERROR !!!

256 x 256 x 256 x 256 x 256 x 256 x 256 =
72’058’000’000’000’000 possibilities (potentially)

QS simulation order 2

A hundred years ago there were one and a half billion people on Earth. Now, over six billion crowd our fragile planet. But even so, there are still places barely touched by humanity. This series will take to the last wildernesses, and show you the planet and its wildlife as you have never seen them before. Imagine our world without sun. Male Emperor Penguins are facing the nearest that exists on planet Earth - winter in Antarctica. It's continuously dark and temperatures drop to minus seventy degrees centigrade. The penguins stay when all other creatures have fled because each guards a treasure: A single egg rested on the top of its feet and kept warm beneath the downy bulge of its stomach. There is no food and no water for them, and they will not see the sun again for four months. Surely no greater ordeal is faced by any animal. As the sun departs from the Antarctic it lightens the skies in the far north. It's March and light returns to the high Arctic, sweeping away four months of darkness. A polar bear stirs. She has been in her den the whole winter. Her emergence marks the beginning of spring. After months of confinement underground she slides down the slope. Perhaps to clean her fur, perhaps for sheer joy. Her cubs gaze out of their bright new world for the very first time. The female calls them, but this steep slope is not the easiest place to take your first steps.

Training image: Planet Earth(BBC)

cr

QS simulation order 2

A hundred years ago there_were_one and a half billion people on Earth. Now, over six billion crowd our fragile planet. But even so, there_are_still places barely touched by humanity. This series will take to the last wildernesses, and show you the planet and its wildlife as you have never seen them before. Imagine our world without sun. Male Emperor Penguins are_facing the nearest that exists on planet Earth - winter in Antarctica. It's continuously dark and temperatures drop to minus seventy degrees centigrade. The penguins stay when all other credtures have fled because each guards a treasure: A single egg rested on the top of its feet and kept warm beneath the downy bulge of its stomach. There_is no food and no water for them, and they will not see the sun again for four months. Surely no greater ordeal is faced by any animal. As the sun departs from the Antarctic it lightens the skies in the far north. It's March and light returns to the high Arctic, sweeping away four months of darkness. A polar bear stirs. She has been in her den the whole winter. Her emergence marks the beginning of spring. After months of confinement underground she slides down the slope. Perhaps to clean her fur, perhaps for sheer joy. Her cubs gaze out of their bright new world for the very first time. The female calls them, but this steep slope is not the easiest place to take your first steps.

Training image: Planet Earth(BBC)

cre

QS simulation order 2

A hundred years ago there were one and a half billion people on Earth. Now, over six billion crowd our fragile planet. But even so, there are still places barely touched by humanity. This series will take to the last wildernesses, and show you the planet and its wildlife as you have never seen them before. Imagine our world without sun. Male Emperor Penguins are facing the nearest that exists on planet Earth - winter in Antarctica. It's continuously dark and temperatures drop to minus seventy degrees centigrade. The penguins stay when all other creatures have fled because each guards a treasure: A single egg rested on the top of its feet and kept warm beneath the downy bulge of its stomach. There is no food and no water for them, and they will not see the sun again for four months. Surely no greater ordeal is faced by any animal. As the sun departs from the Antarctic it lightens the skies in the far north. It's March and light returns to the high Arctic, sweeping away four months of darkness. A polar bear stirs. She has been in her den the whole winter. Her emergence marks the beginning of spring. After months of confinement underground she slides down the slope. Perhaps to clean her fur, perhaps for sheer joy. Her cubs gaze out of their bright new world for the very first time. The female calls them, but this steep slope is not the easiest place to take your first steps.

Training image: Planet Earth(BBC)

cret

QS simulation order 2

A hundred years ago there were one and a half billion people on Earth. Now, over six billion crowd our fragile planet. But even so, there are still places barely touched by humanity. This series will take to the last wildernesses, and show you the planet and its wildlife as you have never seen them before. Imagine our world without sun. Male Emperor Penguins are facing the nearest that exists on planet Earth - winter in Antarctica. It's continuously dark and temperatures drop to minus seventy degrees centigrade. The penguins stay when all other creatures have fled because each guards a treasure: A single egg rested on the top of its feet and kept warm beneath the downy bulge of its stomach. There is no food and no water for them, and they will not see the sun again for four months. Surely no greater ordeal is faced by any animal. As the sun departs from the Antarctic it lightens the skies in the far north. It's March and light returns to the high Arctic, sweeping away four months of darkness. A polar bear stirs. She has been in her den the whole winter. Her emergence marks the beginning of spring. After months of confinement underground she slides down the slope. Perhaps to clean her fur, perhaps for sheer joy. Her cubs gaze out of their bright new world for the very first time. The female calls them, but this steep slope is not the easiest place to take your first steps.

Training image: Planet Earth(BBC)

cret

But wait, not all combination exist

- Like I have no sample of “cersdyvc”

  ➔ cersdyvc

- Tokenization !! We don’t look to characters but to word.

- Each “word” (set of characters) is encoded as avec high dimensional vector

“OpenAI's large language models process text using tokens, which are common sequences of characters found in a set of text. The models learn to understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.”

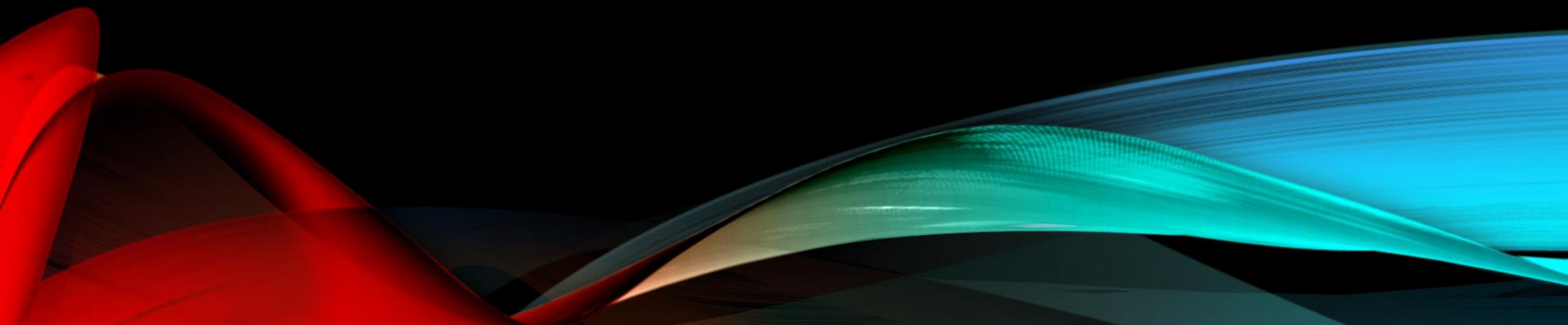
https://platform.openai.com/tokenizer

Tokenization example

[32, 11779, 2101, 5288, 1354, 1504, 1001, 326, 261, 6375, 12842, 1665, 402, 16464, 13, 6549, 11, 1072, 7429, 12842, 19829, 1039, 68256, 17921, 13, 3072, 1952, 813, 11, 1354, 553, 2928, 9610, 35815, 41119, 656, 41006, 13, 1328, 5594, 738, 2304, 316, 290, 2174, 77833, 268, 11, 326, 2356, 481, 290, 17921, 326, 1617, 40214, 472, 481, 679, 3779, 6177, 1373, 2254, 13, 56304, 1039, 2375, 2935, 7334, 13, 28507, 75402, 158456, 553, 20511, 290, 35210, 484, 13031, 402, 17921, 16464, 533, 13655, 306, 154109, 13, 7744, 40371, 8883, 326, 28295, 9440, 316, 40335, 124706, 18210, 2427, 5213, 973, 13, 623, 16517, 104885, 5092, 1261, 722, 1273, 43509, 679, 68442, 2236, 2454, 59641, 261, 47849, 25, 355, 4590, 16102, 127354, 402, 290, 2344, 328, 1617, 11059, 326, 13185, 9144, 39397, 290, 1917, 88, 7750, 684, 328, 1617, 31866, 13, 3274, 382, 860, 4232, 326, 860, 3411, 395, 1373, 11, 326, 1023, 738, 625, 1921, 290, 7334, 2418, 395, 4242, 5503, 13, 122240, 860, 10740, 173466, 382, 29280, 656, 1062, 13983, 13, 1877, 290, 7334, 2834, 8661, 591, 290, 176692, 480, 4207, 696, 290, 79186, 306, 290, 4150, 16173, 13, 7744, 7561, 326, 4207, 7377, 316, 290, 1932, 63651, 11, 78027, 4194, 4242, 5503, 328, 44420, 13, 355, 39618, 16387, 420, 14168, 13, 3627, 853, 1339, 306, 1335, 1786, 290, 6062, 13655, 13, 6526, 84618, 22891, 290, 10526, 328, 12860, 13, 6311, 5503, 328, 105731, 43889, 1770, 33885, 1917, 290, 56134, 13, 30391, 316, 4687, 1335, 21779, 11, 12951, 395, 55047, 15917, 13, 6526, 18538, 82, 71350, 842, 328, 1043, 13712, 620, 2375, 395, 290, 1869, 1577, 1058, 13, 623, 15806, 11666, 1373, 11, 889, 495, 48432, 56134, 382, 625, 290, 39356, 2475, 316, 2304, 634, 1577, 10331, 13]

- “ ordeal” ➔ 173466
- ” Antarctic”➔ 176692

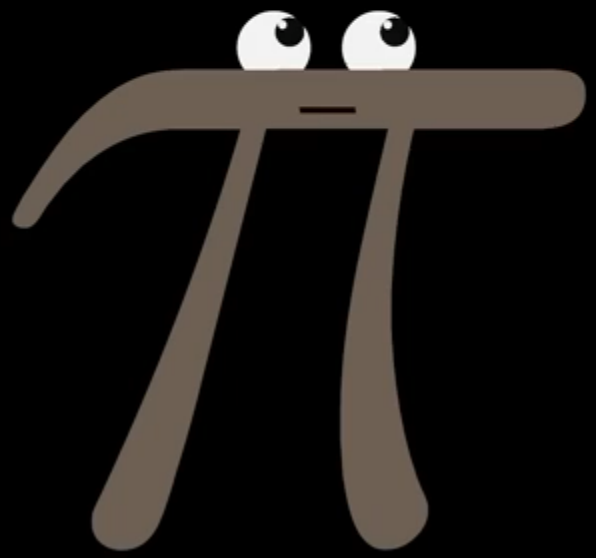
Introduction to Language Models and Transformers



From token to embedding

- Embedding

$$x \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$$



3Blue1Brown

The power of Attention

- "I **mean** to say, it's pretty **mean** to **mean** that the **mean meaning** of **'mean**' only **means** the **mean** as in average, which would be a **mean** reduction!"

- How to give sense to this various "mean" → Attention

- Attention, compute the new vector based on the others
  → the word conditional to the others

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer,

25 August 2003 League of Extraordinary Gentlemen: Sean Connery is one of the all time greats and I have been a fan of his since the 1950's. I went to this movie because Sean Connery was the main actor. I had not read reviews or had any prior knowledge of the movie. The movie surprised me quite a bit. The scenery and sights were spectacular, but the plot was unreal to the point of being ridiculous. In my mind this was not one of his better movies it could be the worst. Why he chose to be in this movie is a mystery. For me, going to this movie was a waste of my time. I will continue to go to his movies and add his movies to my video collection. But I can't see wasting money to put this movie in my collection

I found this to be a charming adaptation, very lively and full of fun. With the exception of a couple of major errors, the cast is wonderful. I have to echo some of the earlier comments -- Chynna Phillips is horribly miscast as a teenager. At 27, she's just too old (and, yes, it DOES show), and lacks the singing "chops" for Broadway-style music. Vanessa Williams is a decent-enough singer and, for a non-dancer, she's adequate. However, she is NOT Latina, and her character definitely is. She's also very STRIDENT throughout, which gets tiresome. The girls of Sweet Apple's Conrad Birdie fan club really sparkle -- with special kudos to Brigitta Dau and Chiara Zanni. I also enjoyed Tyne Daly's performance, though I'm not generally a fan of her work. Finally, the dancing Shriners are a riot, especially the dorky three in the bar. The movie is suitable for the whole family, and I highly recommend it.

Judy Holliday struck gold in 1950 withe George Cukor's film version of "Born Yesterday," and from that point forward, her career consisted of trying to find material good enough to allow her to strike gold again. It never happened. In "It Should Happen to You" (I can't think of a blander title, by the way), Holliday does yet one more variation on the dumb blonde who's maybe not so dumb after all, but everything about this movie feels warmed over and half hearted. Even Jack Lemmon, in what I believe was his first film role, can't muster up enough energy to enliven this recycled comedy. The audience knows how the movie will end virtually from the beginning, so mostly it just sits around waiting for the film to catch up. Maybe if you're enamored of Holliday you'll enjoy this; otherwise I wouldn't bother. Grade: C

Once in a while you get amazed over how BAD a film can be, and how in the world anybody could raise money to make this kind of crap. There is absolutely No talent included in this film - from a crappy script, to a crappy story to crappy acting. Amazing...

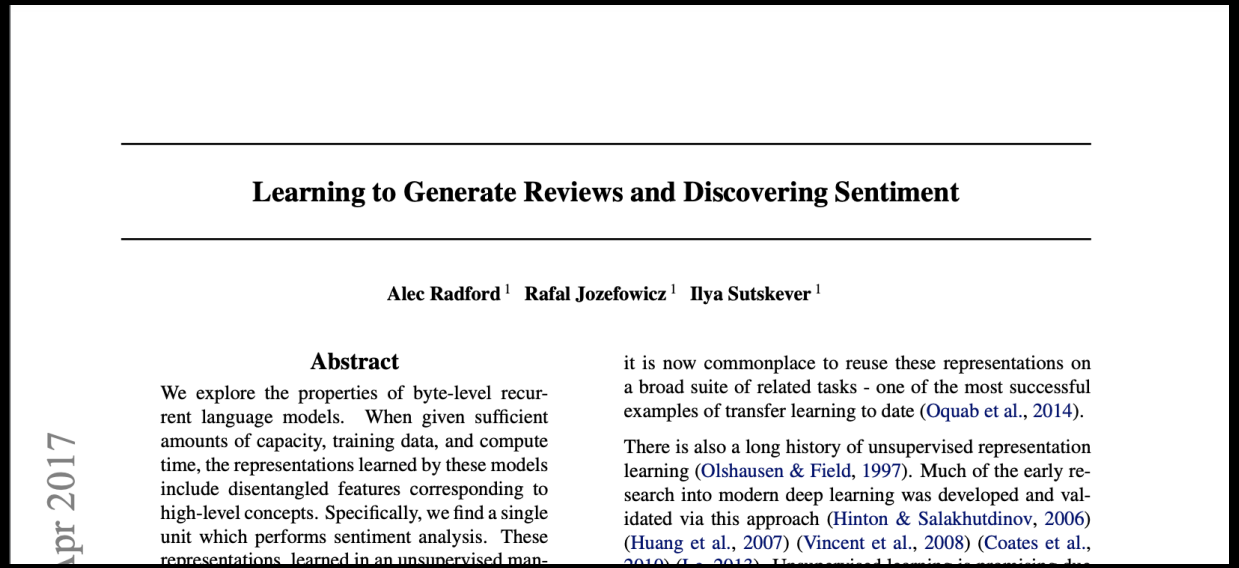
Team Spirit is maybe made by the best intentions, but it misses the warmth of "All Stars" (1997) by Jean van de Velde. Most scenes are identic, just not that funny and not that well done. The actors repeat the same lines as in "All Stars" but without much feeling.

God bless Randy Quaid...his leachorous Cousin Eddie in Vacation and Christmas Vacation hilariously stole the show. He even made the awful Vegas Vacation at least worth a look. I will say that he tries hard in this made for TV sequel, but that the script is so NON funny that the movie never really gets anywhere. Quaid and the rest of the returning Vacation vets (including the orginal Audrey, Dana Barron) are wasted here. Even European Vacation's Eric Idle cannot save the show in a brief cameo.... Pathetic and sad...actually painful to watch...Christmas Vacation 2 is the worst of the Vacation franchise.

Figure 4. Visualizing the value of the sentiment cell as it processes six randomly selected high contrast IMDB reviews. Red indicates negative sentiment while green indicates positive sentiment. Best seen in color.

Learn concept

- In 2017 OpenAI Announced discovering a hidden diemention about sentiments.



Learning to Generate Reviews and Discovering Sentiment

Alec Radford [1] **Rafal Jozefowicz** [1] **Ilya Sutskever** [1]

Abstract

We explore the properties of byte-level recurrent language models. When given sufficient amounts of capacity, training data, and compute time, the representations learned by these models include disentangled features corresponding to high-level concepts. Specifically, we find a single unit which performs sentiment analysis. These representations, learned in an unsupervised man-

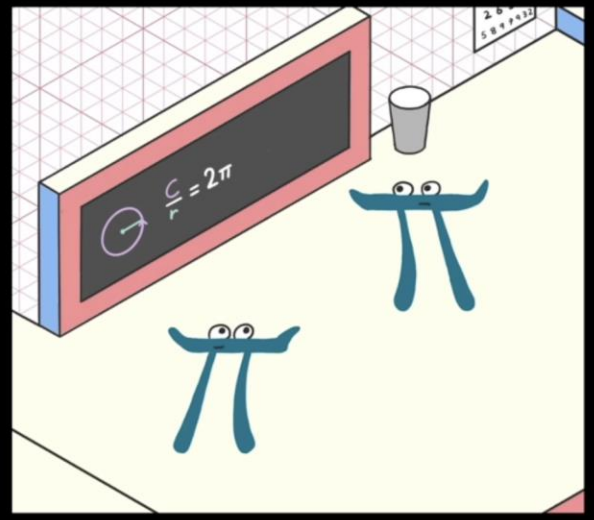
it is now commonplace to reuse these representations on a broad suite of related tasks - one of the most successful examples of transfer learning to date (Oquab et al., 2014).

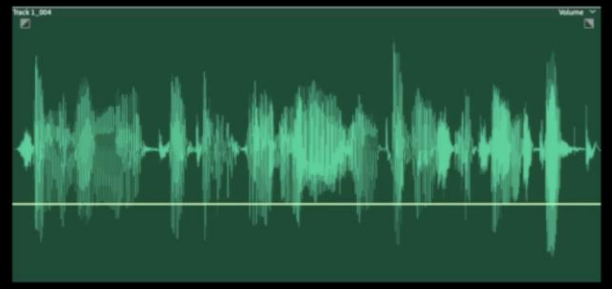
There is also a long history of unsupervised representation learning (Olshausen & Field, 1997). Much of the early research into modern deep learning was developed and validated via this approach (Hinton & Salakhutdinov, 2006) (Huang et al., 2007) (Vincent et al., 2008) (Coates et al.,

pr 2017

Final Sampling

Behold, a wild pi creature,
foraging in its native _____





Transformer

land	22%
forest	9%
country	5%
habitat	4%
forests	4%
soil	4%
territory	2%
woods	2%
lands	1%
waters	1%
woodland	1%
grass	1%
⋮	

Retrieval-Augmented Generation (RAG)

Combine **information retrieval** (fetch facts) with **language generation** (compose text)

Make LLMs *fact-grounded*, *up-to-date*, and *domain-aware*.

Dos and Don'ts of ChatGPT

When to use it ?

- Two schools

- Try a little everything !
  - Especially if you would appreciate someone else doing it 😆

General rules

- Never assume that it knows !!! NEVER
  - Even if it look like
  - Tell everything, pass the maximum of information
  - Tell it how to answer, even with examples
- It’s a tool !!
  - You take full responsibility of it’s use
- Do NEVER put names or personal information
- Don’t be too polite with GPT.
  - Just say to do it!

Different scenarios

- You know how to do it
  - Have fun ! Simply quickly supervise EACH TIME

  ⚠ it's not because it works once that it will continue to work !!!

  ➔ if you use it a lot you will know, when to use it, when to double/triple check

- You don't know how to do it
  - Proceed with caution

- Ideas generation
- Data/information retrieval ➔ retrieval-augmented generation (RAG)
- Critical opinion

Make an abstract

It's a text machine !!

➔Use it to make text!!

- Simply dump your idea
  - Unstructured
  - Text
  - Paper
  - Any type of related content

  ➔ask for an abstract !

  "Now make me an abstract of X words for the Y conference. And propose me a title as well"

Get title ideas

- Explain what you are looking for
- Maybe you want to have an acronym that has as a meaning or a specific acronym
- Ask for more sample ”give me 10 variants”

Example

- "Optimizing ChatGPT for Academic Research: Essential Dos and Don'ts"
- **"ChatGPT in Academia: Best Practices and Common Pitfalls"**
- "Mastering ChatGPT for Scholarly Work: Dos and Don'ts"
- "Using ChatGPT for Academic Success: Key Dos and Don'ts"
- "Academic Excellence with ChatGPT: What Works and What Doesn't"
- "Enhancing Research with ChatGPT: Dos and Don'ts for Academics"
- "Leveraging ChatGPT in Academia: A Practical Guide to Dos and Don'ts"
- "The Academic's Guide to ChatGPT: Dos and Don'ts for Effective Use"
- "ChatGPT for Researchers: Navigating Dos and Don'ts in Academia"
- "Academic Efficiency with ChatGPT: Dos and Don'ts for Scholars"

Data representation

- Data sample: https://github.com/fivethirtyeight/data

Emails

- Prompt:
  improve this:
- https://chatgpt.com/c/d3f63623-ca7c-4e9b-979a-551661dac650

Proof reading

- Prompt:
  I need you to do proofreeding, expect quality, but at the same time minimize the amount of change comparing to the original. I will provide you section after section, please answer with the proofread version and not extra text around. It needs to look like expert-quality proofreading

- Change the tone

Coding / analysis

- Convert code from one language to another
- Debug
- Analysis the code:
  - Prompt:
    do you see something weird in this code ?

The future of ChatGPT

- Thinking mode ➔ thinking before generating (there is multiple level)

Can generate idea competition before answering

And more

Speak and restructure !!

A “neutral” opinion / biases

- Ask if a text is biased.

- Explore social biases
  - GPT-3 was trained on (300 billion tokens)~ 400 million page’s
  - A machine learning is a regression (best fit) ~average
  - Therefore, we can explore how this text (the society) are biased

Gain knowledge along the way

- Answer to emails for an event
- You can start a thread and each email you answer will bring knowledge after 10 emails, it can usually answer most of the question automatically

Search engine

- What to say ?
- RAG over internet !

Changes in one year

For references

- Too bad
- Not made for this !!
- Reference
  - https://www.perplexity.ai/
  - https://consensus.app/
- Paper generator
  - https://storm.genie.stanford.edu/ (experimental)

→ **Deep research**, make in less than 30 min one of the best literature review

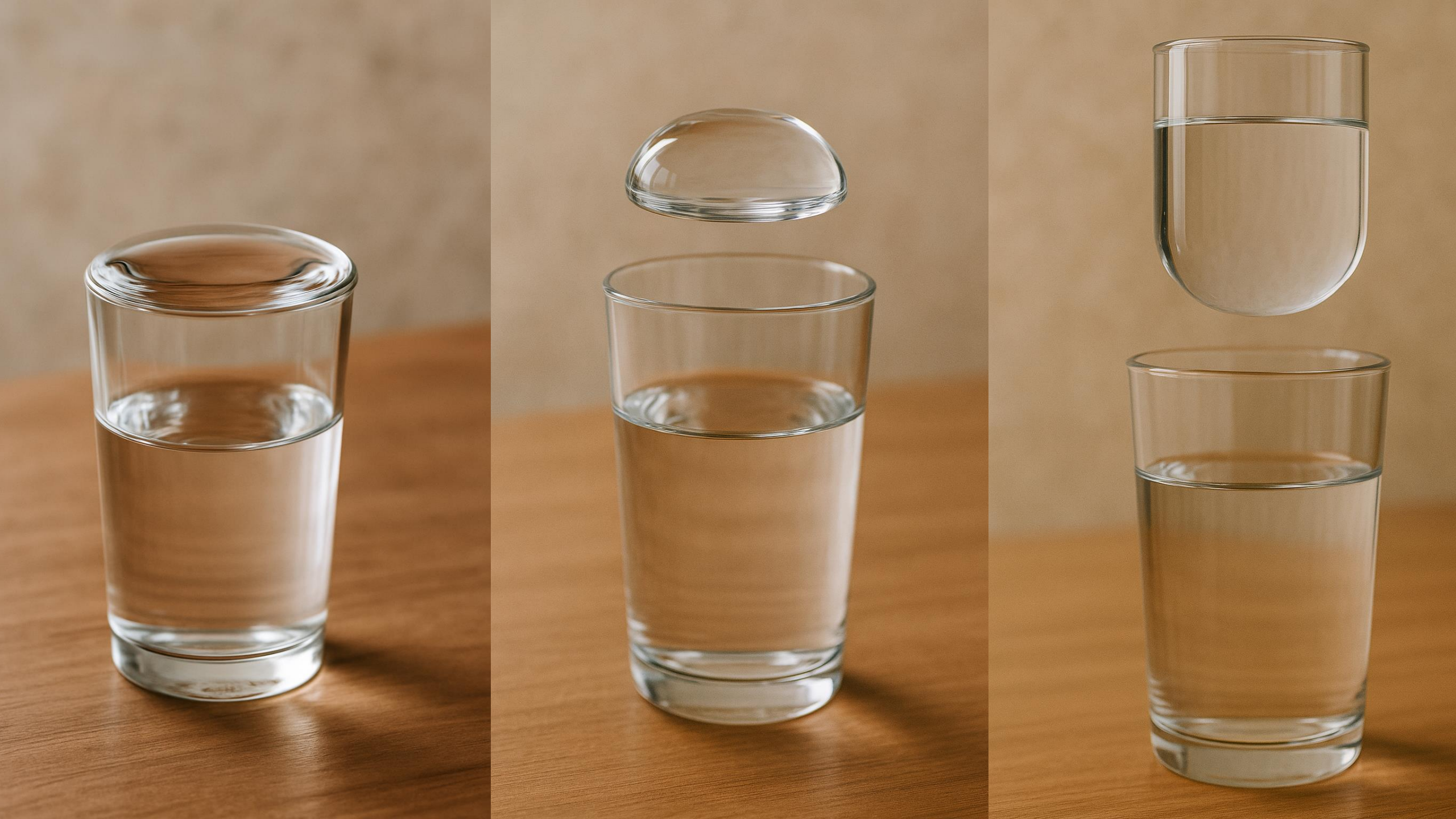
it prefers open publication 😉

ChatGPT extra feature

- Dall.e → *gpt-image*
  - Image generator, ok to make a quick carton, logo, illustration image, but nothing scientific…. But we are coming closer and closer
- GPTs
  - customized ChatGPT, tailored for some specific tasks, using custom instruction, an potentially specific document.
  - → loose momentum
- Canvas
  - Work on a document, paper, code … → Agent !
- Call as service
  - Can create custom function that ChatGPT can call online → MCP

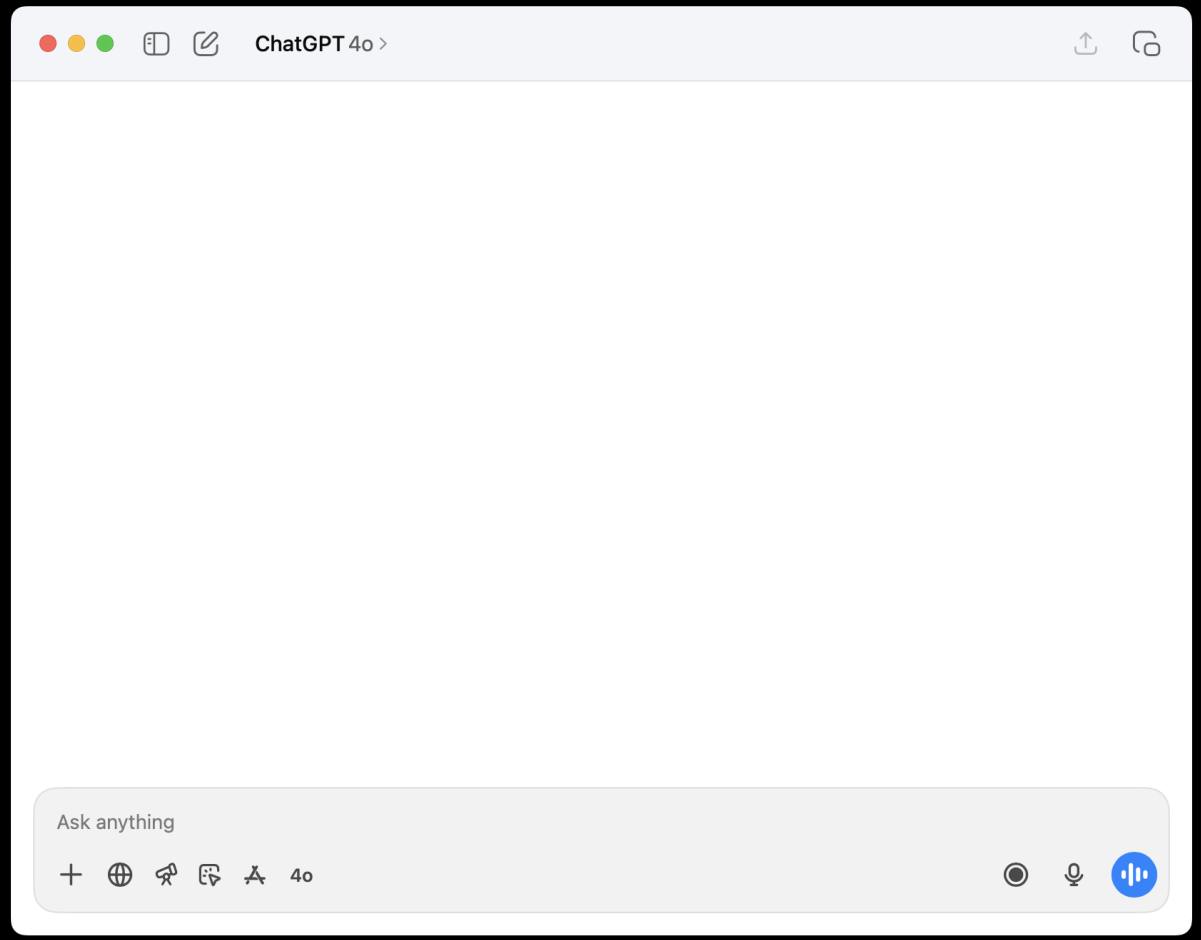
Now

- Images generator,
  - can now support image modification, style etc.
  - local change, only at the selection
- Text !!
- And overall, more realistic, even physically realistic



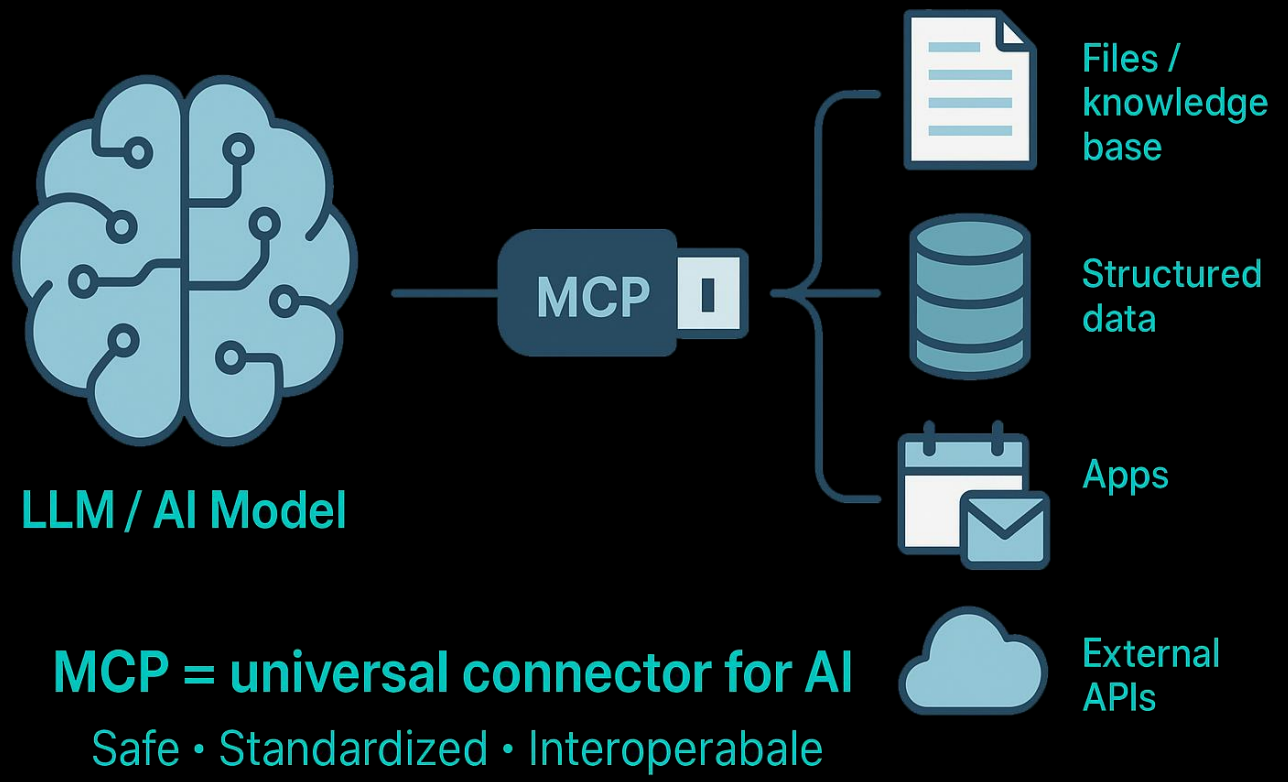
Speech to text

- Now Integrated



Model Context Protocol (MCP)

- MCP is like a "universal language" that lets AI models connect safely and easily to other tools, data, or apps — without needing custom code each time.



Agents

A single word for lot of concepts

LLM + Reasoning + Memory + Tools + Planning + Autonomy = Agent

- LLM provides intelligence,
- tools provide capabilities,
- planning and memory provide continuity,
- autonomy ties it all together into behavior.

Use an Agent when the task is complex, interactive, and requires multi-step reasoning with external tools.

Other platforms

Software	Developer	Context Length	Image Generation	Code Performance	Key Strengths	Limitations
ChatGPT	OpenAI	~1 M tokens, some model only	✅ Yes	✅ Excellent	Multimodal, large ecosystem, strong coding & reasoning	Subscription-based, variant limits
Claude	Anthropic	Up to 1 M tokens	❌ No	✅ Best	Long context, safe reasoning, strong text & code understanding	No image gen, slower rollout
Gemini	Google DeepMind	Up to 2 M tokens (1M in public)	✅ Yes	✅ Good	Fully multimodal (text + image + video), Google integration	Privacy & regional limitations
DeepSeek	DeepSeek AI	~128 K tokens	✅ Yes	✅ Strong	Fast, cost-efficient, good code generation	Less global support, young ecosystem
Mistral	Mistral AI	~128 K tokens	❌ No	✅ Strong	Open-weight models, great research & code use	Limited multimodal features

Do it yourself

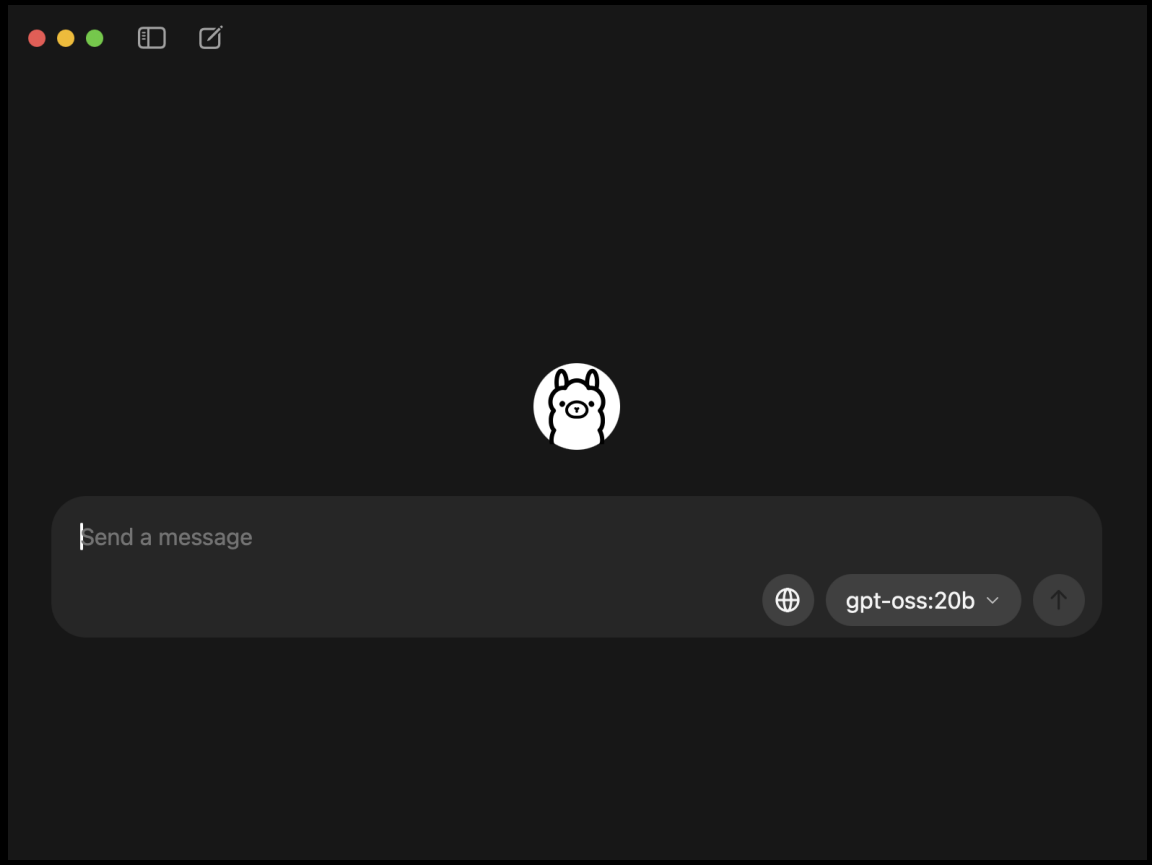
- Host yourself the model,
  - on a laptop or better 😀

  But No cloud

- Example ollama

Various model:

  - gpt-oss
  - Llama
  - gemma
  - Deepseek-R1



Local LLM Platforms — Comparison Overview

Software	Interface	Models Supported	Deployment	Performance	API Support	Notable Features
Ollama	CLI / REST API	GGUF (LLaMA, Mistral, Gemma, etc.)	Local (Mac, Win, Linux)	⚡ Optimized	✅	Lightweight, scriptable
LM Studio	GUI + Local API	GGUF	Local (Mac, Win, Linux)	⚡ Similar	✅	User-friendly UI, model manager
GPT4All	GUI	GGUF / GPT4All models	Local (Mac, Win, Linux)	⚡ Efficient	❌	Privacy-focused, offline
vLLM	API only	Transformers	Server / GPU	⚡⚡⚡ Production-grade	✅	Enterprise inference engine

Parameters

Parameter	Typical Range	Effect	Tip
Temperature	0–2	Controls randomness / creativity	0.2–0.7 for logic, >1 for ideas
Top-p	0–1	Keeps top-probability tokens	0.8–0.95 for balanced output
Top-k	10–100	Limits to *k* most likely tokens	Often off if using Top-p
Max Tokens	Model-dependent	Max length of output	Stay below context limit
Presence Penalty	–2 to +2	Boosts topic variety	+0.5 to +1 for diversity
Frequency Penalty	–2 to +2	Reduces repetition	~0.5 to 1.0 is typical
Repetition Penalty	1–2	Penalizes repeats	~1.2 common
Stop Sequences	Custom	Defines where to stop	Use in multi-turn chats
System Prompt	—	Sets role/personality	Keep short & stable
Seed	Integer	Makes output reproducible	Fix for experiments

Hands-on Session