

Sissevaade suurte keelemudelite põnevasse maailma

Andres Kull





Keelemudelite areng (1)

NLP (Natural Language Processing)

- teksti klassifitseerimine
 - nimega entiteetide tuvastamine (named entity recognition)
 - tekstikogumite klasterdamine
 - tekiotsing andmekogumitest
 - sentiment analüüs
 - keeltepaaridevaheline tõlkimine
 - etc.
-
- spetsiifilised mudelid iga vajaduse jaoks
 - piiratud täpsus ja rakendatavus
 - meenuta Google inglise - eesti tõlkeid

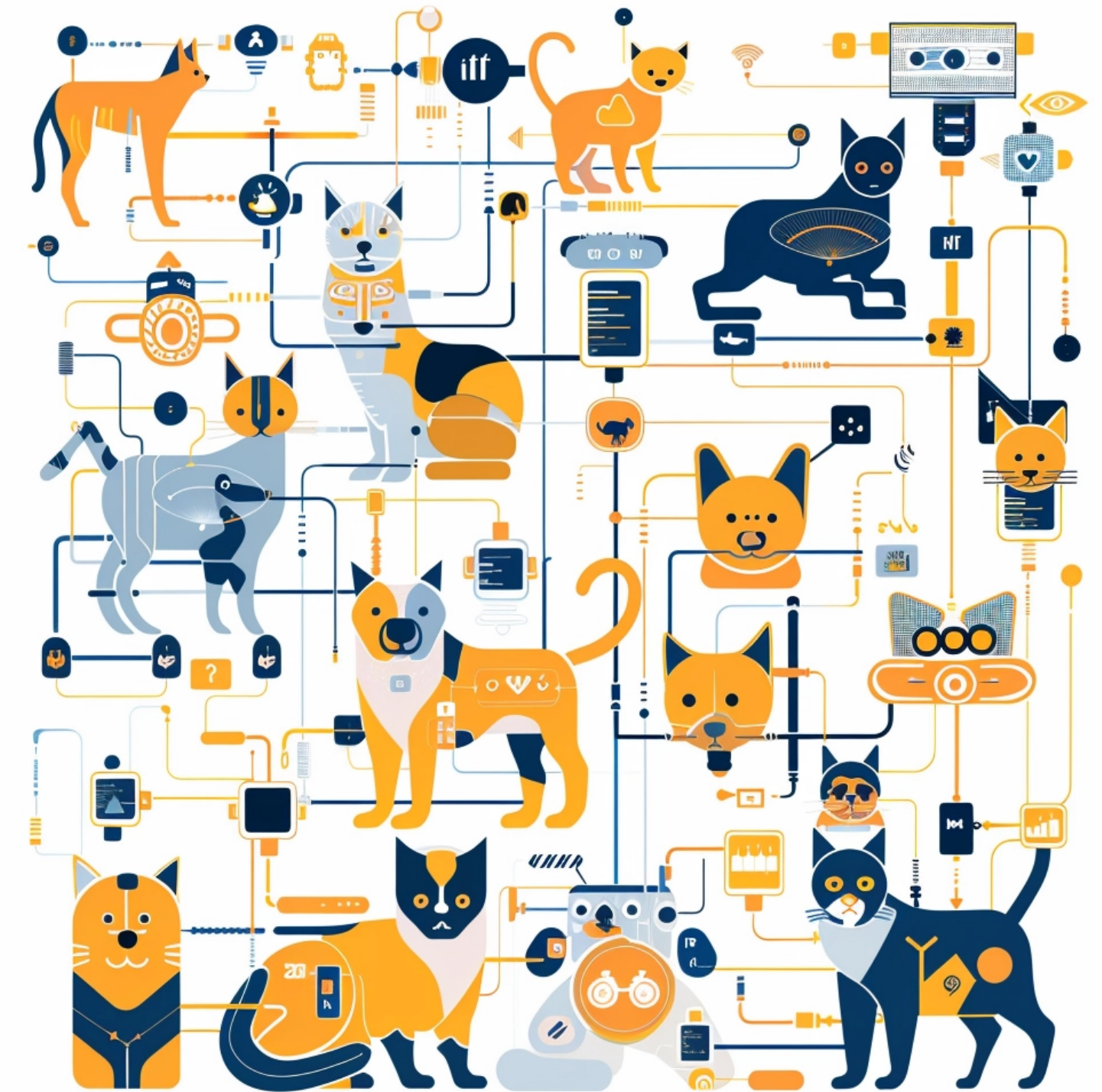




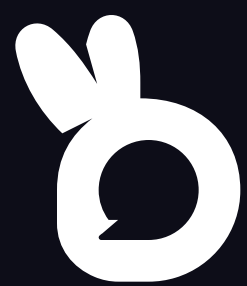
Keelemudelite areng (2)

2012

- süvaõpe (deep-learning on neural networks)
- AlexNet mudeli edu ImageNet pildituvastusvõistlusel
- Piltidelt kasside ja koerte tuvastamine enneolematu täpsusega



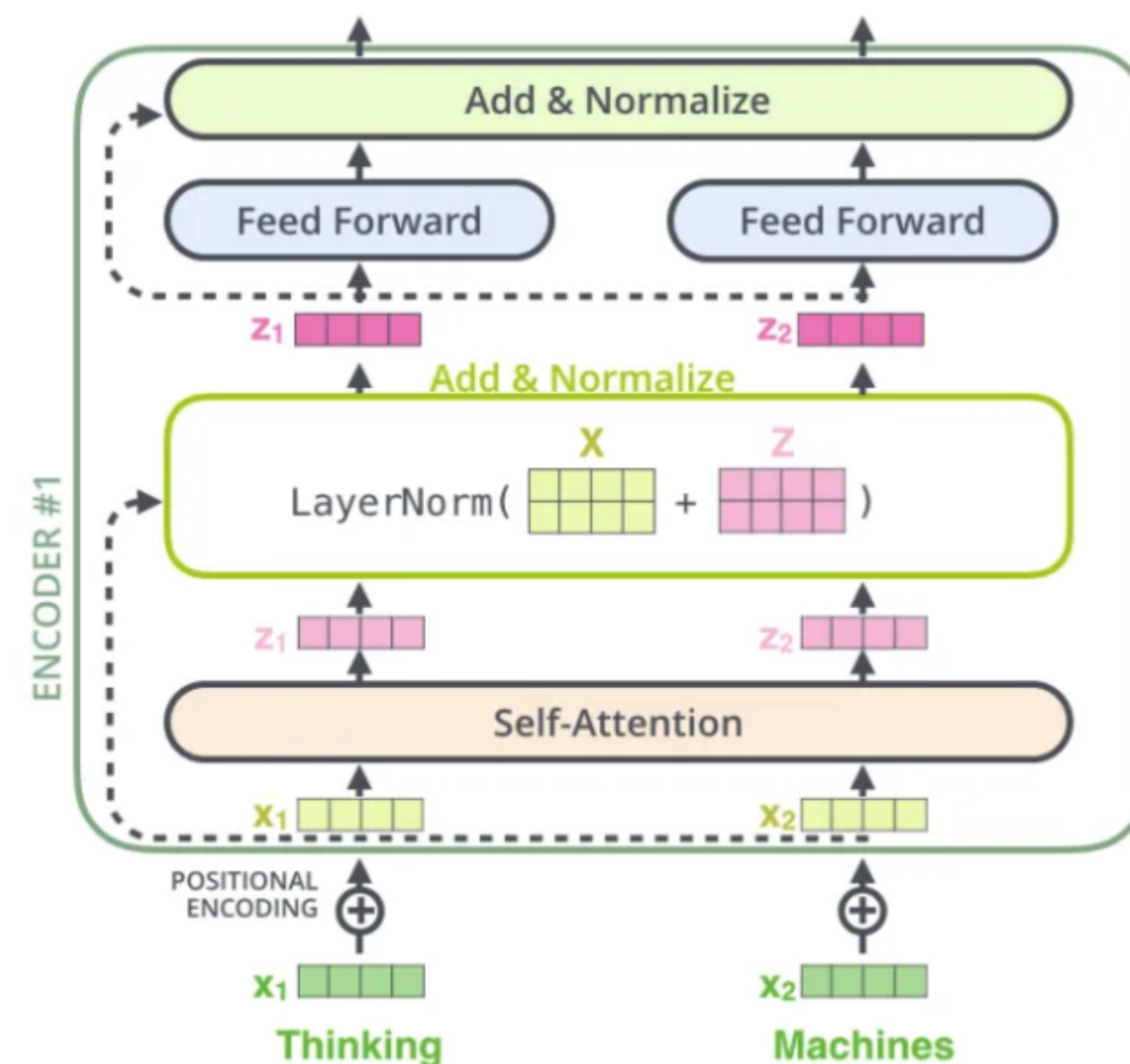
Süvaõpe tõi uue hingamise ka keelemudelite arengusse



Keelemudelite areng (3)

2017

- “Attention is All You Need”
- Transformer mudelid tähelepanu mehhanismiga
- võimaldasid mudelitel keskenduda olulisele ja mõista pikka teksti
- Oluline edasimineku võrreldes seniste state-of-the-art mudelitega (RNN; LSTM)

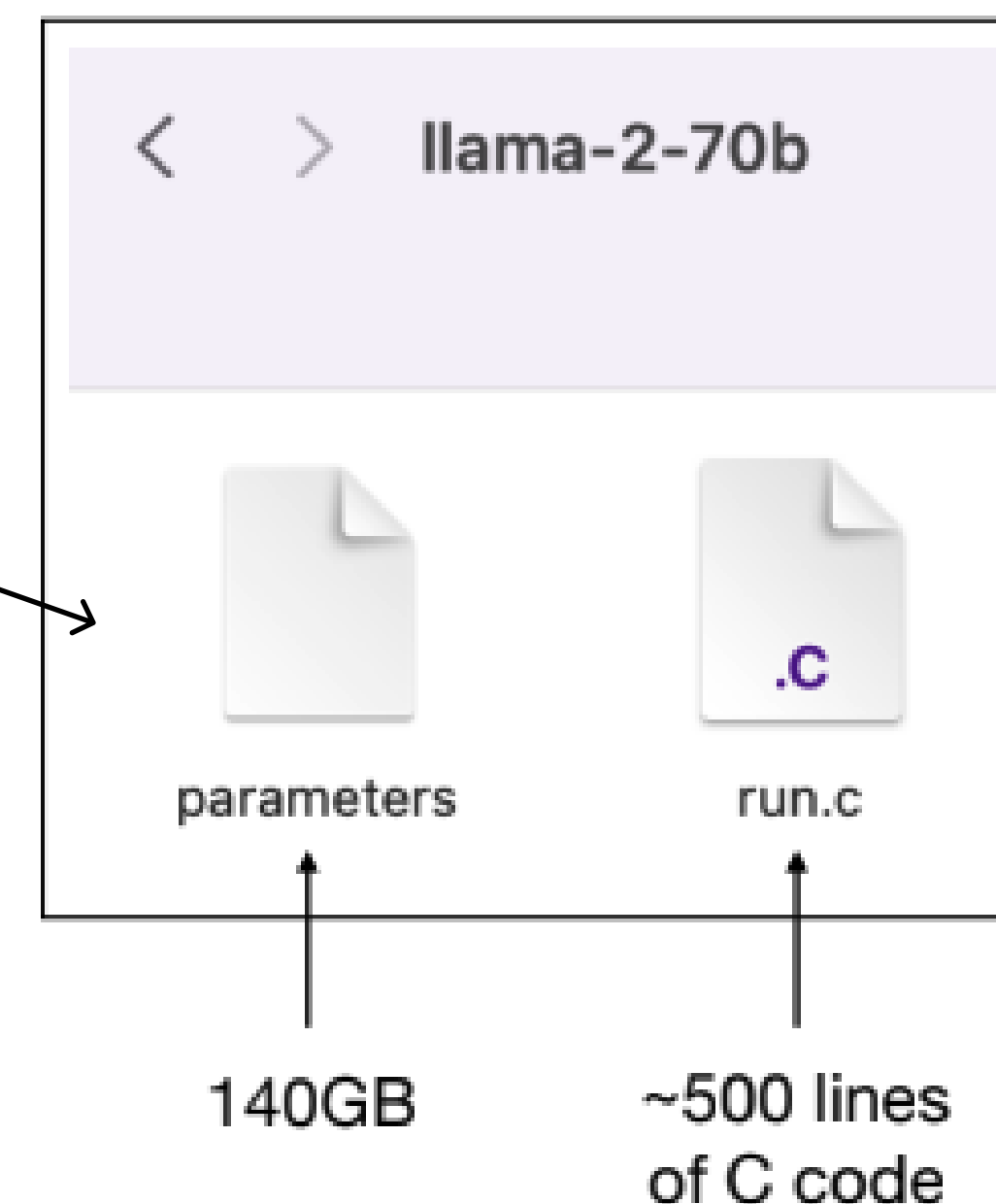
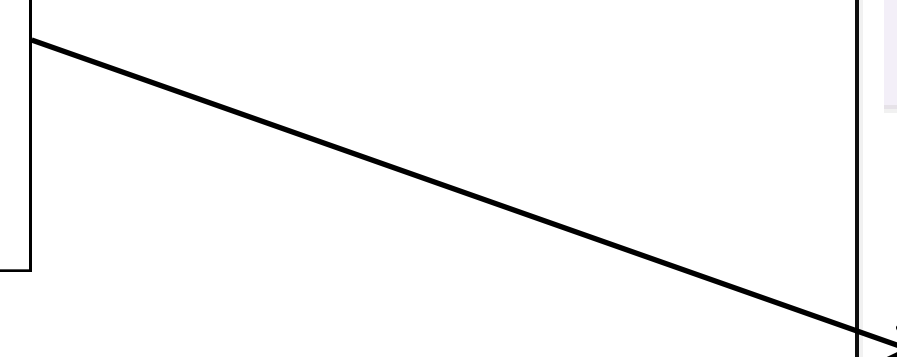




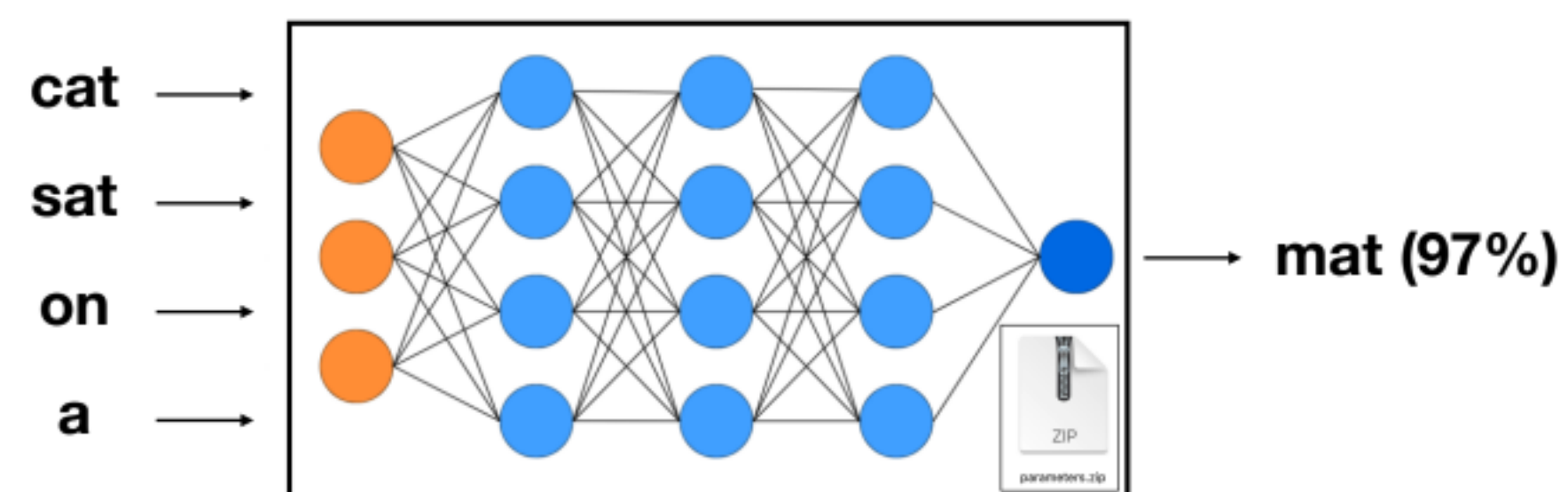
Mis on suur keelemudel (LLM)?



6000 GPU
12 päeva
~\$2M



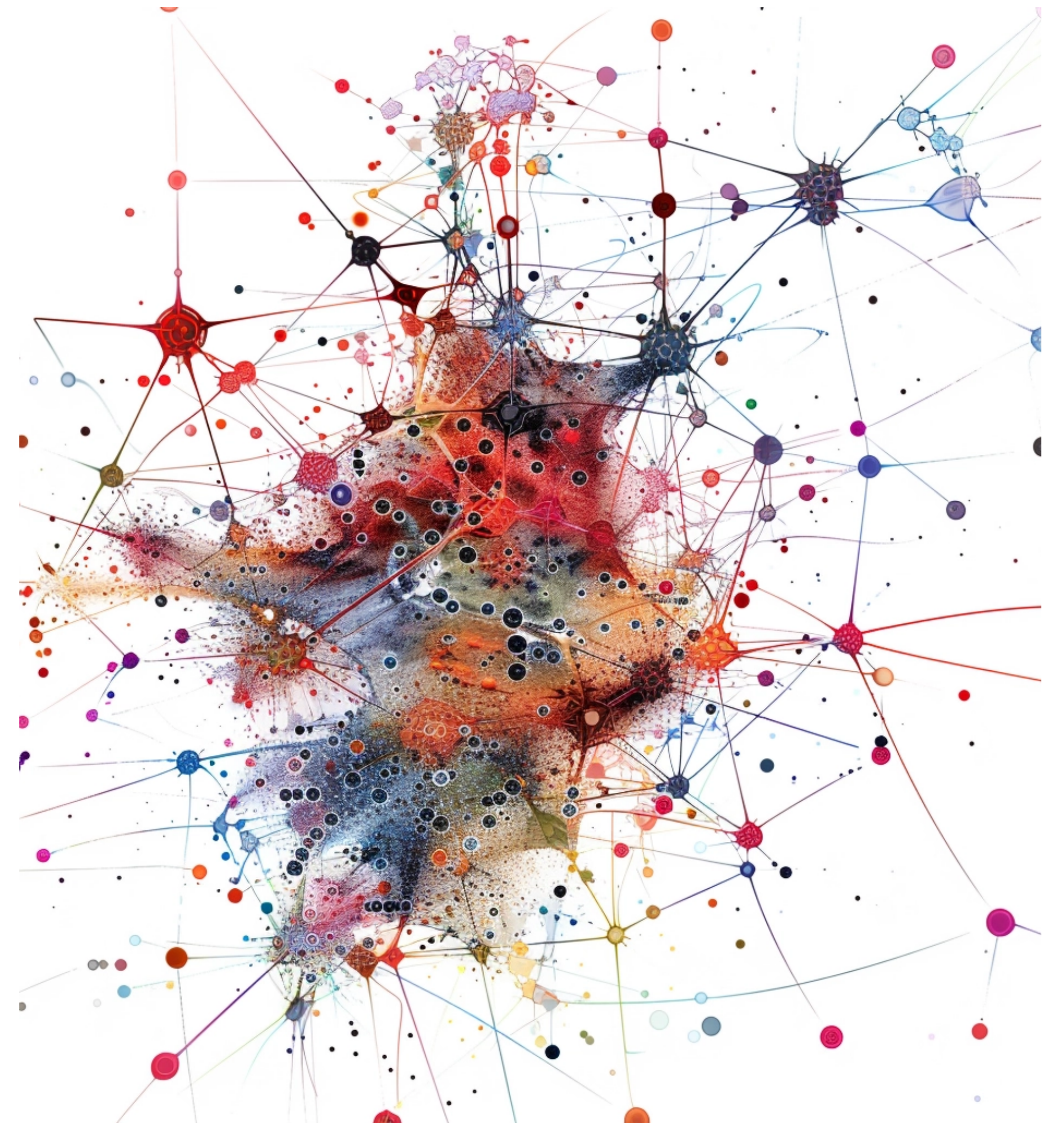
- sügav närvivõrk
- treenitud kogu internetis saadaoleva tekstuaalse info abil
- treenitud ennustama järgmist tekstilõiku
- omandab meeletu faktide ja seoste kogumi
- isetekkeliste omaduste ilmumine

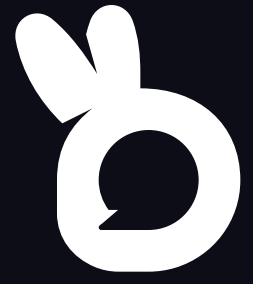




Isetekkelised omadused

- tekstide tõlkimine
- keeleline loovus ja sisuloome
- kontekstitundlikkus ja järeldusvõime
- teabe kokkuvõtmine
- küsimustele vastamine
- dialoogisüsteemid
- emotsioonide ja sentimentide tuvastamine
- koodi genereerimine ja tõlgendamine

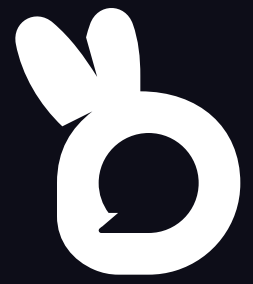




Baasmudelite võrdlus

Chatbot Arena Leaderboard by huggingface.co

Rank ▲	🌐 Model ▲	★ Arena Elo ▲	📊 95% CI ▲	🗳️ Votes ▲	Organization ▲	License ▲	Knowledge Cutoff ▲
1	GPT-4-1106-preview	1251	+5/-5	45291	OpenAI	Proprietary	2023/4
2	GPT-4-0125-preview	1251	+6/-6	15251	OpenAI	Proprietary	2023/12
3	Claude-3-Opus	1233	+9/-7	5246	Anthropic	Proprietary	2023/8
4	Bard..(Gemini..Pro)	1203	+6/-8	12623	Google	Proprietary	Online
5	GPT-4-0314	1185	+5/-5	24689	OpenAI	Proprietary	2021/9
6	Claude-3-Sonnet	1180	+10/-8	5259	Anthropic	Proprietary	2023/8
7	GPT-4-0613	1161	+5/-5	39845	OpenAI	Proprietary	2021/9
8	Mistral-Large-2402	1155	+6/-6	9746	Mistral	Proprietary	Unknown
9	Mistral-Medium	1147	+5/-4	22171	Mistral	Proprietary	Unknown
10	Qwen1.5-72B-Chat	1147	+4/-5	15288	Alibaba	Qianwen LICENSE	2024/2
11	Claude-1	1146	+5/-6	20833	Anthropic	Proprietary	Unknown
12	Claude-2.0	1127	+6/-5	13679	Anthropic	Proprietary	Unknown
13	Mistral-Next	1124	+5/-6	11875	Mistral	Proprietary	Unknown
14	Gemini.Pro...(Dev..API)	1118	+6/-7	11496	Google	Proprietary	2023/4
15	Claude-2.1	1116	+4/-5	31815	Anthropic	Proprietary	Unknown



Assistentide treenimine

Baasmudel



<USER>

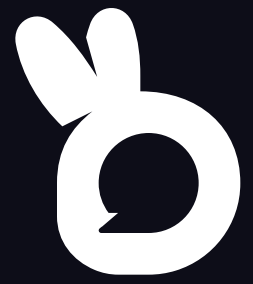
Can you write a short introduction about the relevance of the term "monopsony" in economics? Please use examples related to potential monopsonies in the labour market and cite relevant research.

<ASSISTANT>

"Monopsony" refers to a market structure where there is only one buyer for a particular good or service. In economics, this term is particularly relevant in the labor market, where a monopsony employer has significant power over the wages and working conditions of their employees. The presence of a monopsony can result in lower wages and reduced employment opportunities for workers, as the employer has little incentive to increase wages or provide better working conditions...

Peenhäälestamine (fine-tuning)

1. Kirjutada hulgaliselt kasutaja-assistent stsenaariumeid +100k
2. Peenhäälestada baasmudel nendega
3. Tulemuseks on **assistendi mudel**
4. Jooksutada hulgaliselt evaluatsioone
5. Paigaldada toodangusse
6. Monitoorida, koguda ebatäpsusi
7. GOTO step 1



Tekstide tõlkimine



You

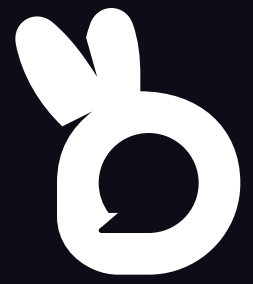
Translate into finnish "mitu jalga on jalgrattal?"



ChatGPT

Kuinka monta jalkaa polkupyörässä on?

- Mudelit ei ole spetsiaalselt õpetatud tõlkima
- Mudel on ise õppinud seosed eri keelte ja mõistete vahel
- Väikesed keeled ei ole enam probleemiks



Keeleline loovus ja sisuloome



You

Mis on suur keelemudel? Selgita nii, et 6 aastane saaks aru?



ChatGPT

Suur keelemudel on nagu suur arvutiprogramm, mis oskab sõnadega mängida. Kujuta ette, et see on nagu väga tark raamat, mis on lugenud läbi kõik lood, raamatud ja veebilehed maailmas, ning nüüd oskab see ise jutte rääkida, küsimustele vastata või isegi mõistatusi lahendada. Kui sa küsid sellelt programmi-sõbralt, mis on tema lemmikloom, võib ta sulle rääkida loo koerast või kassist, isegi kui tal päriselt lemmiklooma ei ole. See on justkui maagiline raamat, mis oskab mõelda ja rääkida nagu inimene.





Keeleline loovus ja sisuloome



You

Mis on suur keelemudel? Selgita nii, et ka automaatikainsener aru saaks.



ChatGPT

Suur keelemudel on arvutiprogramm, mis suudab mõista ja genereerida inimkeelt, kasutades selleks suuri andmehulkasid. Kujutage ette, et see on nagu väga arenenud tüüpi tarkvara, mis on "õppinud" keelt, analüüsidest tohutul hulgal tekste internetist – alates uudisartiklitest ja blogipostitustest kuni raamatuteni.

See töötab põhimõttel, mis sarnaneb sellega, kuidas teie loote ja kohandate juhtimissüsteeme: mudel "treenitakse" andmete abil, et õppida keele struktuuri, grammatikat ja semantikat, täpselt nagu te programmeerite juhtimissüsteemi teatud käitumist õppima andurite ja sisendite põhjal.

Näiteks, kui te küsite sellelt mudelilt, kuidas kirjutada efektiivset juhtimissüsteemi koodi, suudab see pakkuda teile vastust, analüüsidest sarnaseid küsimusi ja vastuseid, mida ta on õppinud. See võimaldab mudelil toimida nagu virtuaalne assistent, mis võib aidata lahendada keerukaid probleeme või genereerida tehnilist teksti, tuginedes sellele, mida ta on "õppinud" oma treeningandmetest.



Kontekstitundlikkus ja järeldusvõime

You

Kuidas automaatikainsenerid saaksid LLM-e oma töös rakendada, et parandada süsteemide juhtimist või optimeerida protsesse? Palun kirjelda ühte rakendusnäidet tuues välja eelised ja väljakutsed.

 ChatGPT

Automaatikainsenerid saavad LLM-e rakendada näiteks protsesside optimeerimisel ja süsteemide tõrkeotsingul, kasutades selleks keelemudelite võimet töödelda ja analüüsida suures koguses tehnilisi dokumente, logifaile ja kasutajate tagasisidet.

Rakendusnäide: Tõrkeotsingu Automaatika Süsteemide Jaoks

LLM-i saab integreerida süsteemi tõrkeotsingu platvormiga, kus mudel analüüsib masinate logifaile ja tehnilisi dokumente, et tuvastada probleeme ja soovitada parandusmeetmeid. Insenerid saavad esitada süsteemi kirjeldusi ja sümptomeid loomulikus keeles, mille peale LLM genereerib asjakohased juhised või soovitused probleemi lahendamiseks.

Eelised:

- **Efektivsus:** Kiirendab tõrkeotsingut, võimaldades inseneridel keskenduda keerukamatele ülesannetele.
- **Teabe kättesaadavus:** Suudab töödelda ja analüüsida suurt hulka dokumente, pakkudes


Väljakutsed:

- **Andmete kvaliteet:** Mudeli täpsus sõltub kättesaadavate andmete kvaliteedist ja asjakohasusest.
- **Spetsiifilisus:** LLM-id võivad genereerida üldisi soovitusi, mis ei pruugi alati konkreetse süsteemi või probleemi puhul olla piisavalt spetsiifilised.
- **Usaldusväarsus:** Mudeli pakutud lahenduste usaldusväarsuse tagamine nõuab pidevat kontrolli ja valideerimist inseneride poolt.


Selle rakendusnäite kaudu saavad automaatikainsenerid kasutada LLM-e, et muuta tõrkeotsing ja süsteemide hooldus efektiivsemaks ja tõhusamaks, kuigi see nõuab teatud ettevalmistust ja järelkontrolli, et tagada lahenduste asjakohasus ja usaldusväarsus.




Teabe kokkuvõtmine

 You

tee lühikokkuvõte sellest wikipedia lehest https://et.wikipedia.org/wiki/Automaatika_kateeder

 Copilot





Great! You can always review your permissions in Settings.

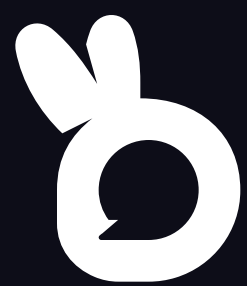
 **Notification and App Settings**

_____ This conversation may have info that's not public. Conversation will not be saved after this point. _____

Siin on lühikokkuvõte sellest wikipedia lehest ¹:

- Automaatika kateeder oli **Tallinna Polütehnilises Instituudis (TPI) Hanno Sillamaa** juhtimisel moodustatud kateeder, mis töötas aastatel **1960 – 1992**.
- Kateeder õpetas erialasid **automaatika ja telemehaanika ning automatiseeritud juhtimissüsteemid**, mis olid NSV Liidus kasutusel olnud mõisted, mis tähendasid kompleksset spetsialiseerumist automaatikavahendite ja automatiseeritud süsteemide ning telemeetria ja kaugjuhtimise alal.
- Kateeder oli seotud ka **küberneetika** arendamisega Eestis ja NSV Liidus ning osales mitmetes rahvusvahelistes projektides ja koostöös.
- Kateeder lõpetas tegevuse **1992. aastal**, kui TPI reorganiseeriti **Tallinna Tehnikaülikooliks (TTÜ)** ja automaatika kateeder liideti **arvutisüsteemide instituudiga**.

Learn more [1](#)  et.wikipedia.org [2](#)  docs.google.com [3](#)  et.wikipedia.org [4](#)  et.wikipedia.org



Koodi genereerimine

ussimäng > ussimäng.py > ...

```
1 import pygame
2 import time
3 import random
4
5 pygame.init()
6
7 white = (255, 255, 255)
8 black = (0, 0, 0)
9 red = (213, 50, 80)
10 green = (0, 255, 0)
11 blue = (50, 153, 213)
12
13 dis_width = 600
14 dis_height = 400
15
16 dis = pygame.display.set_mode((dis_width, dis_height))
17 pygame.display.set_caption('Ussimäng')
18
19 clock = pygame.time.Clock()
20
21 snake_block = 10
22 snake_speed = 15
23
24 font_style = pygame.font.SysFont('arial', 16)
25
26 def our_snake(snake_block, snake_list):
27     for x in snake_list:
28         pygame.draw.rect(dis, black, x)
29
30 def message(msg, color):
31     mesg = font_style.render(msg, True, color)
32     dis.blit(mesg, [dis_width/2, dis_height/2])
33
```



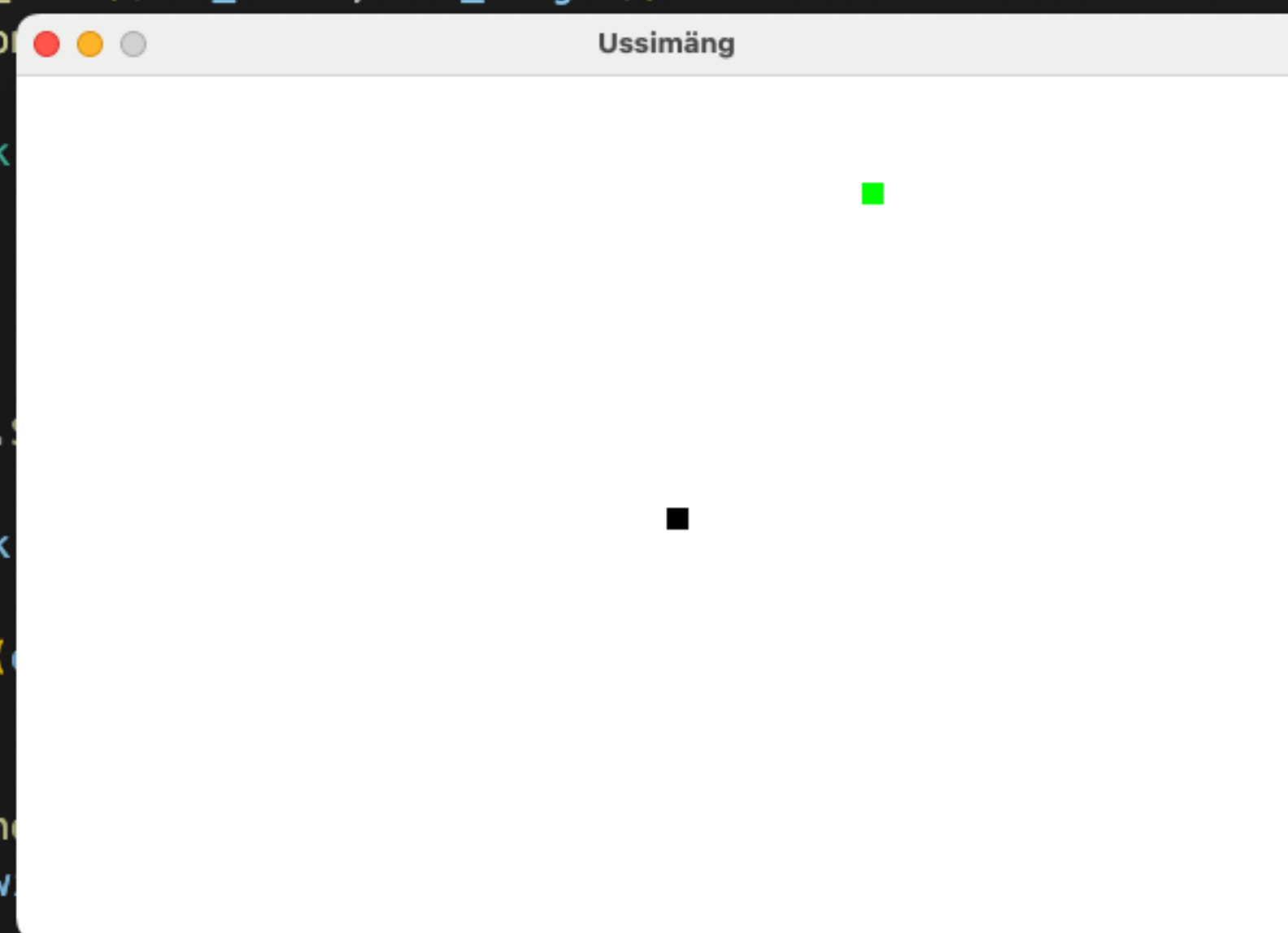
You

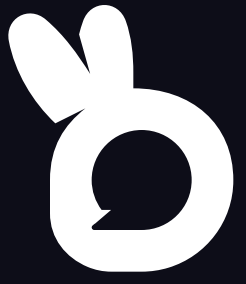
Genereeri ussimängu kood pythonis



ChatGPT

Alljärgnevalt on toodud lihtne ussimängu Pythoni kood, kasutades `pygame` teeki. Enne koodi jooksutamist veenduge, et olete `pygame` oma süsteemi paigaldanud. Seda saab teha käsurealt käsklusega `pip install pygame`, kui Python ja pip on teie süsteemis juba olemas.





You

Kas sa oskad kirjutada ussimängu ka R keeles? Proovi!



ChatGPT

R-keel ei ole tavaliselt esimene valik mängude arendamiseks, kuna see on rohkem suunatud statistilisele analüüsile ja andmetöötlemisele. Siiski on võimalik luua lihtsamaid mänge ka R-keeles, kasutades graafilisi pakette nagu `ggplot2` või interaktiivseid graafikuid loovaid `shiny` rakendusi. Siiski tuleb märkida, et võrreldes teiste programmeerimiskeeltega, nagu Python koos `pygame` teegiga, on R-keele võimalused piiratumad ja tulemused ei pruugi olla sama sujuvad.

Ussimängu loomiseks R-keeles peame leidma loovaid viise, kuidas kasutada olemasolevaid graafilisi võimalusi. Järgnev näide on pigem teoreetiline ja lihtsustatud, näidates, kuidas võiks alustada:

r

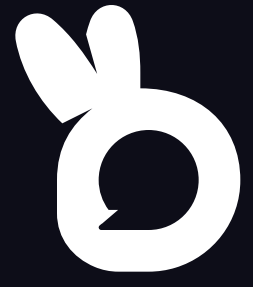
Copy code

```
# See kood ei pruugi töötada otse R-i keskkonnas ilma täiendavate kohandusteta

library(ggplot2)

# Mänguvälja algseadistus
game_area <- matrix(0, nrow = 20, ncol = 20)
snake <- data.frame(x = 10, y = 10)
food <- data.frame(x = sample(1:20, 1), y = sample(1:20, 1))

# Joonista mänguvälja
plot_game <- function(snake, food) {
  game_plot <- ggplot() +
    geom_tile(data = as.data.frame(which(game_area == 1, arr.ind = TRUE)), aes(x = rc
    geom_point(data = snake, aes(x = x, y = y), color = "blue") +
    geom_point(data = food, aes(x = x, y = y), color = "red") +
    xlim(0, 21) +
    ylim(0, 21) +
    theme_minimal() +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
  print(game_plot)
}
```



Igamehe LLM-d

Create a custom GPT:

< New GPT
• Draft

Create Configure

Name
Name your GPT

Description
Add a short description about what this GPT does

Instructions
What does this GPT do? How does it behave? What should it avoid doing?

Conversation starters
x

Knowledge
Upload files

Capabilities
 Web Browsing
 DALL-E Image Generation
 Code Interpreter ⓘ

Actions
Add actions

GPTs app store:



Botastico - SaaS for building chatbots



Allows everyone, in just a few minutes

- to build their own customer facing chatbots
- using their own data
- embed the chatbot widget on their website

The virtual assistant works for you

- 24/7/365
- speaks all languages

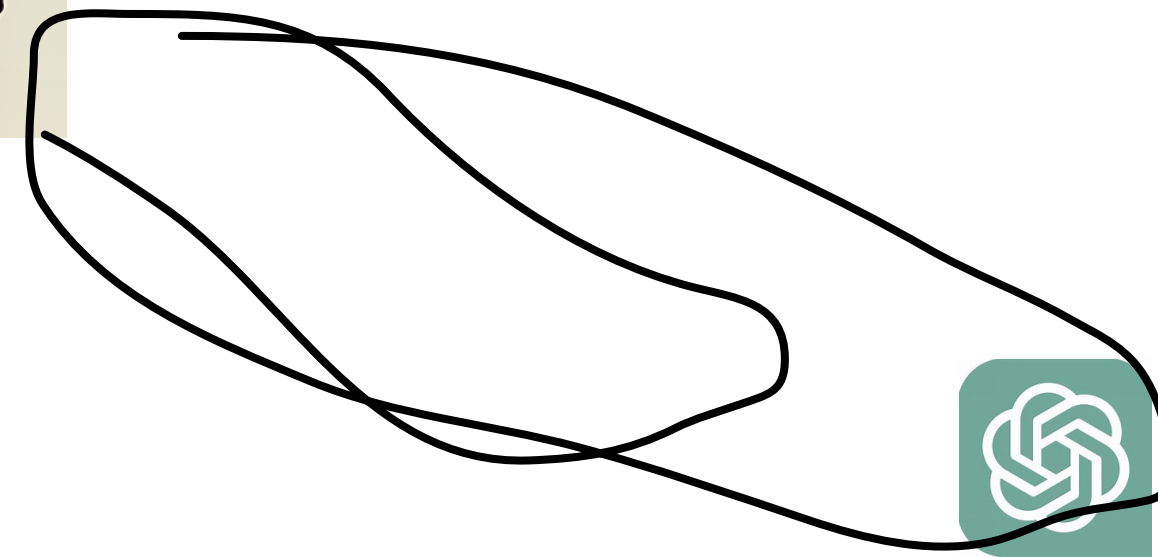
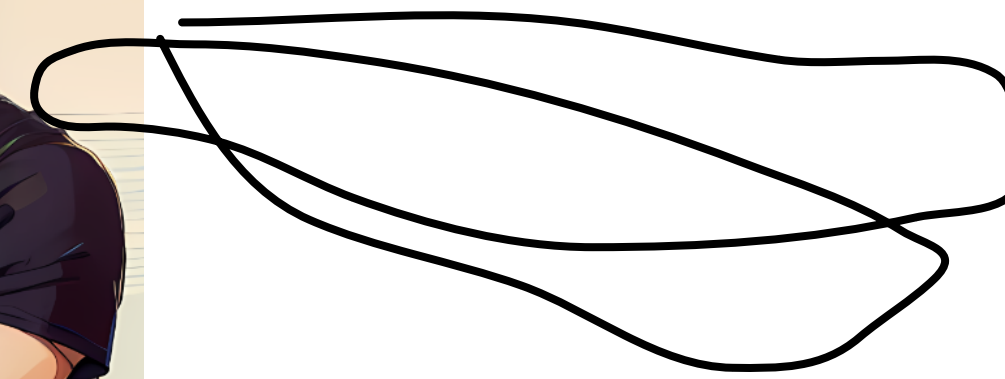
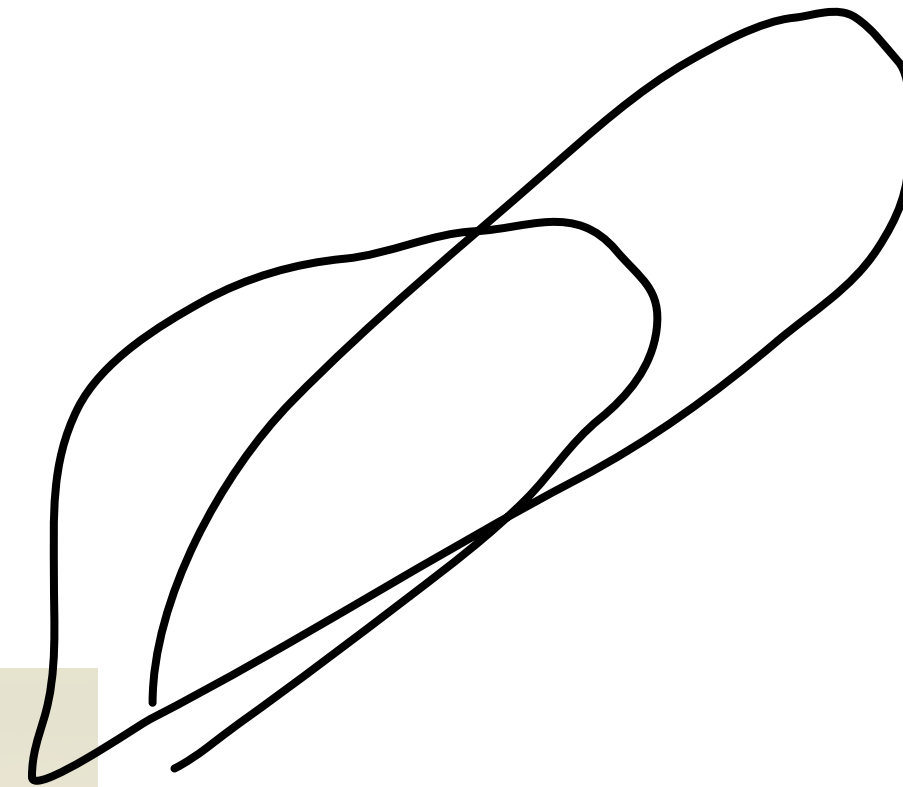
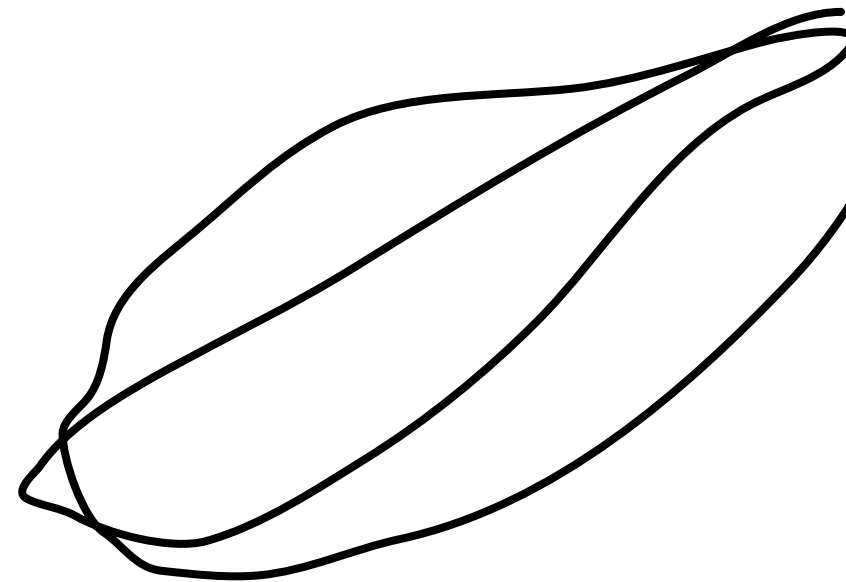
7 months of pair programming with ChatGPT

How can I create my chatbot? ▾

What features do you offer? ▾

Hi there! Ask me about me 🤖

How do we work?





MMM - multimodaalsed mudelid

Mudel on treenitud lisaks tekstidele

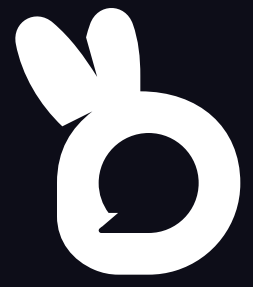
- piltide
- videote
- audisalvestustega

Mudel saab aru piltidest

- võib küsida piltide kohta küsimusi
- ja saada relevantseid vastuseid

Mudel oskab genereerida

- tekstist pilte - Midjourney, Openai DALL-E
- tekstidest videosid - Openai SOMA
- tekstist audiot
- tekstist ja inimese pildist rääkiva või laulva inimese video



It's Q&A time



**Thanks to Midjourney and ChatGPT
for the illustrations!**