# ChatGPT alapok

Gyires-Tóth Bálint


Bálint Gyires-Tóth, PhD

Egyetemi docens

Budapesti Műszaki és Gazdaságtudományi Egyetem

NVIDIA Deep Learning Institute

Oktató és egyetemi nagykövet

NAV Mesterséges Intelligencia Munkacsoport

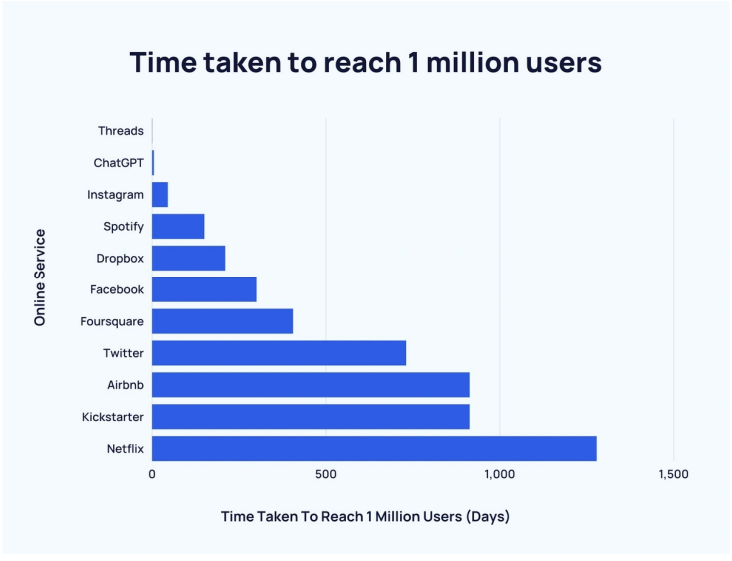
Külső szakértő

Mélytanulás kutatás, fejlesztés, oktatás

# Ki próbált már valamilyen MI eszközt?

# Ki alkalmaz napi szinten MI eszközt?

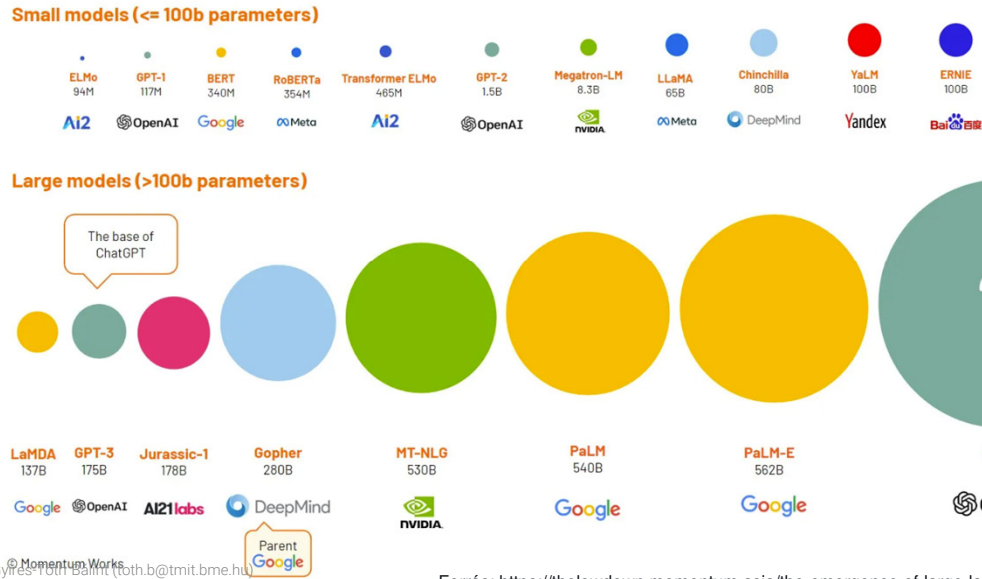
# Új feladat esetén kinek jut MI eszköz elsőnek eszébe?


Time taken to reach 1 million users
Online Service
Threads
ChatGPT
Instagram
Spotify
Dropbox
Facebook
Foursquare
Twitter
Airbnb
Kickstarter
Netflix
0
500
1,000
1,500
Time Taken To Reach 1 Million Users (Days)

# Nagy nyelvi modellek alkalmazási lehetőségei

Gyires-Tóth Bálint (toth.b@tmit.bme.hu)

# Large Language Models are becoming very large indeed










Forrás: https://thelowdown.momentum.asia/the-emergence-of-large-language-models-llms/

| 10-hatvány | magyarul (hosszú skála) | US, modern angol (rövid skála) |
|---|---|---|
| $10^0$ | egy | one |
| $10^1$ | tíz | ten |
| $10^2$ | száz | hundred |
| $10^3$ | ezer | thousand |
| $10^6$ | millió | million |
| $10^9$ | milliárd (ezermillió) | billion |
| $10^{12}$ | billió (ezermilliárd) | trillion |
| $10^{15}$ | billiárd (ezerbillió) | quadrillion |
| $10^{18}$ | trillió | quintillion |
| $10^{21}$ | trilliárd (ezertrillió) | sextillion |

Forrás: https://hu.wikipedia.org/wiki/A_t%C3%ADz_hatv%C3%A1nyai


175 milliárd hangya ≈ 35 000 km

# Modell méretek

**Table 1**. Large-scale models released in recent years.

| Name | #Param | Date |
|---|---|---|
| GPT [1] | 110 million | Jun 2018 |
| BERT [2] | 340 million | Oct 2018 |
| GPT-2 [3] | 1.5 billion | Feb 2019 |
| Megatron-LM [4] | 8.3 billion | Sep 2019 |
| Turing-NLG [5] | 17 billion | Feb 2020 |
| GPT-3 [6] | 175 billion | May 2020 |
| Switch Transformer [7] | 1.6 trillion | Jan 2021 |
| BaGuaLu [8] | 174 trillion | April 2021 |
| Megatron-Turing NLG [9] | 530 billion | Jan 2022 |
| PaLM [10] | 540 billion | April 2022 |

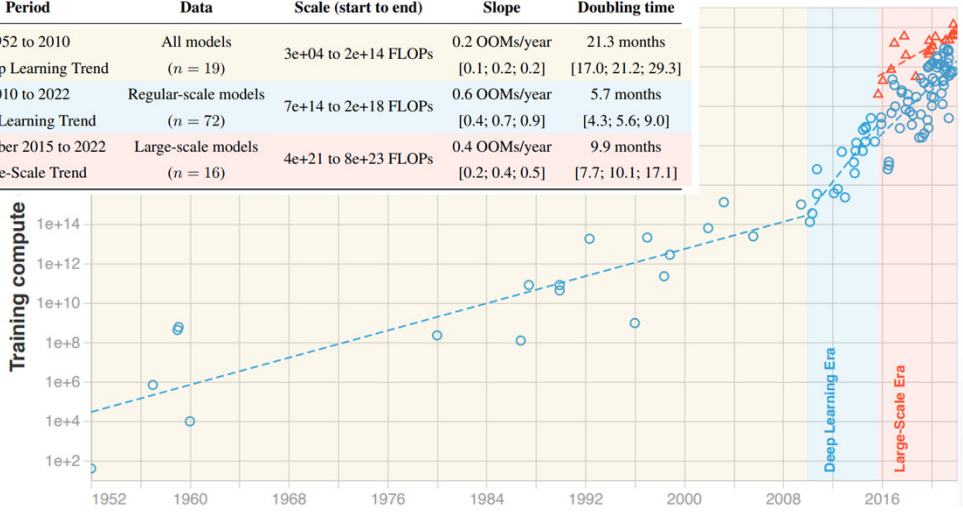
**Table 2**. Estimated training time on one NVIDIA V100 GPU.

| Name | #Param | Dataset Size | Time |
|---|---|---|---|
| GPT [1] | 110M | 4GB | 3 days |
| BERT [2] | 340M | 16GB | 50 days |
| GPT-2 [3] | 1.5B | 40GB | 200 days |
| GPT-3 [6] | 175B | 560GB | 90 years |

Forrás: https://github.com/qhliu26/BM-Training

More huge models: ViT-G/14 (CV, 2.4B), V-MoE (CV, 15B), ViT-22B (CV, 22B), Stable Diffusion (CV, 890M), Flamingo (NLP-CV, 80B), M6 (multimodal, 10T), stb.

| Period | Data | Scale (start to end) | Slope | Doubling time |
|---|---|---|---|---|
| 1952 to 2010<br>Pre Deep Learning Trend | All models<br>($n = 19$) | 3e+04 to 2e+14 FLOPs | 0.2 OOMs/year<br>[0.1; 0.2; 0.2] | 21.3 months<br>[17.0; 21.2; 29.3] |
| 2010 to 2022<br>Deep Learning Trend | Regular-scale models<br>($n = 72$) | 7e+14 to 2e+18 FLOPs | 0.6 OOMs/year<br>[0.4; 0.7; 0.9] | 5.7 months<br>[4.3; 5.6; 9.0] |
| September 2015 to 2022<br>Large-Scale Trend | Large-scale models<br>($n = 16$) | 4e+21 to 8e+23 FLOPs | 0.4 OOMs/year<br>[0.2; 0.4; 0.5] | 9.9 months<br>[7.7; 10.1; 17.1] |

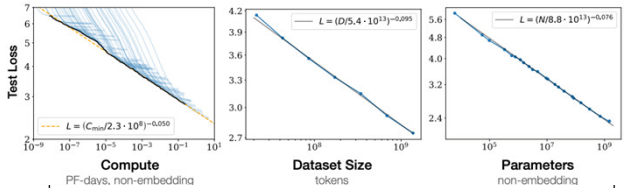


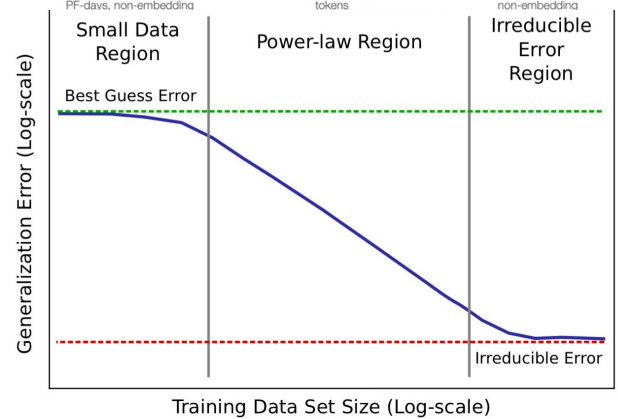

Source: Sevilla, Jaime, et al. "Compute trends across three eras of machine learning." *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022.

# A DEEP LEARNING ERŐTÖRVÉNYE

Hestness, Joel, et al. "Deep learning scaling is predictable, empirically." arXiv preprint arXiv:1712.00409 (2017).

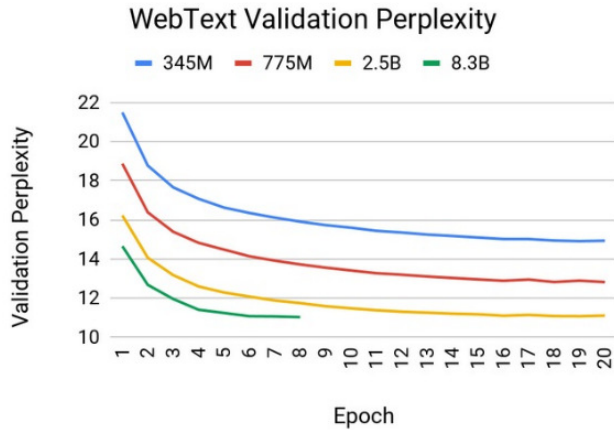
*„These power-law learning curves* ***exists across all tested domains, model architectures, optimizers, and loss functions****. Further, within each domain,* ***model architecture and optimizer changes only shift the learning curves*** *but* ***do not affect the power-law exponent****—the "steepness" of the learning curve."*



# Nagy modellek előnyei

- *Nagyobb pontosság*
- *Adathatékonyság + few-shot learning*
- *Erőforrás-hatékonyság*
- *Hiperparaméterek jelentősége drasztikusan csökken*



# Kis nyelvi modellek

- Kis (párszáz millió, vagy néhány milliárd paraméter) modellek tanítása
- Célspecifikus adatbázissal finomhangolás

Tulajdonságok:

- Erőforrás hatékony
- Feladat specifikus


Gyires-Tóth Bálint (toth.b@tmit.bme.hu)

# Új paradigma: LLM

- Nagy (milliárd paraméterű) modellek tanítása
- Utasításokkal és kérdés-válaszokkal való finomhangolás

Tulajdonságok:

- Zero- és few-shot tanulás
- Általános modell sokféle feladatra
- Nagy számításigény
- Külső adatforrások felhasználása (Internet, dokumentumtár)

Nagy nyelvi modellek

Utasítás követő nyelvi modellek

Társalgó nyelvi modellek

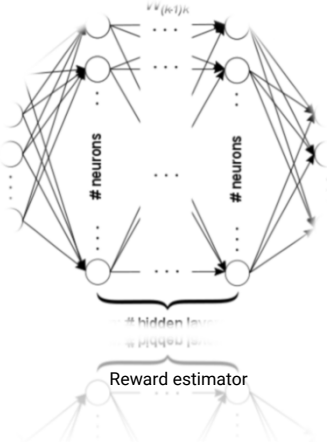
RAG rendszerek

# A ChatGPT jellegű rendszerek működése

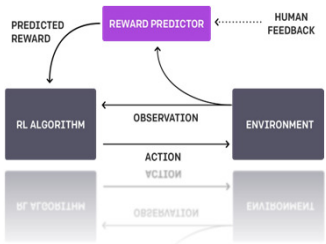
Gyires-Tóth Bálint (toth.b@tmit.bme.hu)

# Komponensek

Large Language Model




Reward estimator




Megerősítéses tanulás

# Training language models to follow instructions with human feedback

**Long Ouyang*** **Jeff Wu*** **Xu Jiang*** **Diogo Almeida*** **Carroll L. Wainwright***

**Pamela Mishkin*** **Chong Zhang** **Sandhini Agarwal** **Katarina Slama** **Alex Ray**

**John Schulman** **Jacob Hilton** **Fraser Kelton** **Luke Miller** **Maddie Simens**

**Amanda Askell†** **Peter Welinder** **Paul Christiano*†**

**Jan Leike*** **Ryan Lowe***

OpenAI

## Abstract

Making language models bigger does not inherently make them better at following a user's intent. For example, large language models can generate outputs that are untruthful, toxic, or simply not helpful to the user. In other words, these models are not *aligned* with their users. In this paper, we show an avenue for aligning language models with user intent on a wide range of tasks by fine-tuning with human feedback. Starting with a set of labeler-written prompts and prompts submitted through the OpenAI API, we collect a dataset of labeler demonstrations of the desired model behavior, which we use to fine-tune GPT-3 using supervised learning. We then collect a dataset of rankings of model outputs, which we use to further fine-tune this supervised model using reinforcement learning from human feedback. We call the resulting models *InstructGPT*. In human evaluations on our prompt distribution, outputs from the 1.3B parameter InstructGPT model are preferred to outputs from the 175B GPT-3, despite having 100x fewer parameters. Moreover, InstructGPT models show improvements in truthfulness and reductions in toxic output generation while having minimal performance regressions on public NLP datasets. Even though InstructGPT still makes simple mistakes, our results show that fine-tuning with human feedback is a promising direction for aligning language models with human intent.

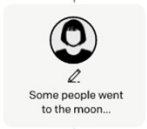
Ouyang, Long, et al. "Training language models to follow instructions with human feedback." *Advances in Neural Information Processing Systems* 35 (2022): 27730-27744.

Step 1

**Collect demonstration data, and train a supervised policy.**

A prompt is sampled from our prompt dataset.


A labeler demonstrates the desired output behavior.




This data is used to fine-tune GPT-3 with supervised learning.

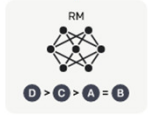
Step 2

**Collect comparison data, and train a reward model.**

A prompt and several model outputs are sampled.


A labeler ranks the outputs from best to worst.


This data is used to train our reward model.




Step 3

**Optimize a policy against the reward model using reinforcement learning.**

A new prompt is sampled from the dataset.


The policy generates an output.


The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

Source: https://openai.com/research/instruction-following

# Instruct adatbázis példák

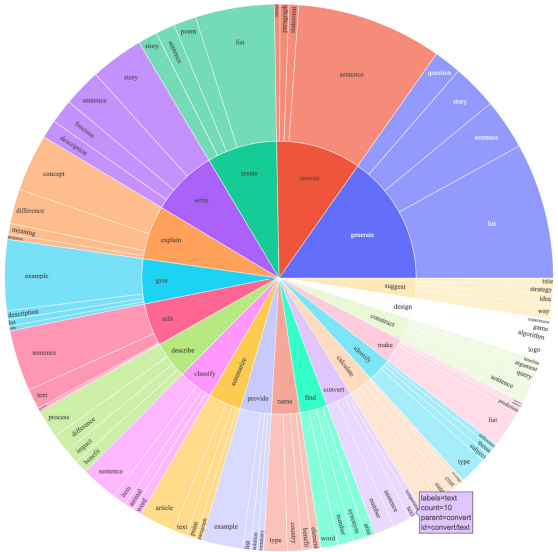
Minőségében lehet különbség:

- Kézi
- Manuális fordítás
- Automatikus fordítás
- Utasítás követő LLM-el készült adatbázis

Példák:

- https://arxiv.org/pdf/2203.02155.pdf 26. oldal
- https://huggingface.co/datasets/HuggingFaceH4/instruction-dataset
- https://huggingface.co/datasets/databricks/databricks-dolly-15k
- https://crfm.stanford.edu/2023/03/13/alpaca.html

# Stanford Alpaca



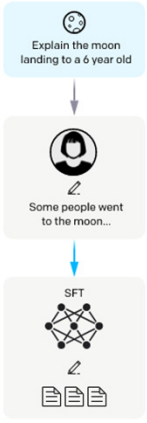
Step 1

**Collect demonstration data, and train a supervised policy.**

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3 with supervised learning.



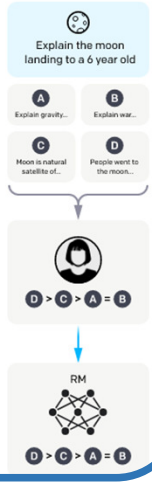

Step 2

**Collect comparison data, and train a reward model.**

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.




Step 3

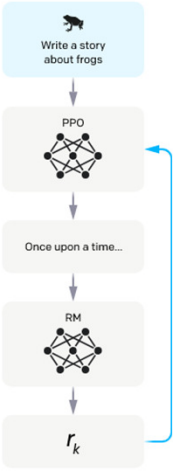
**Optimize a policy against the reward model using reinforcement learning.**

A new prompt is sampled from the dataset.

The policy generates an output.

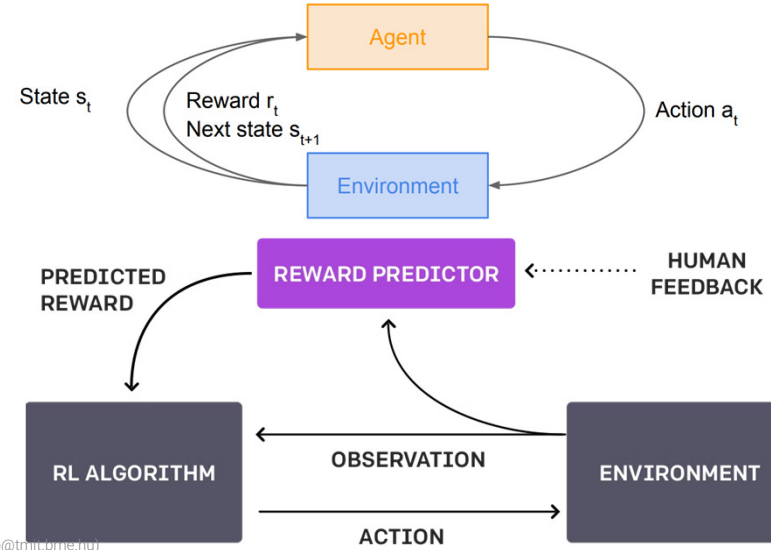
The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.




Source: https://openai.com/research/instruction-following

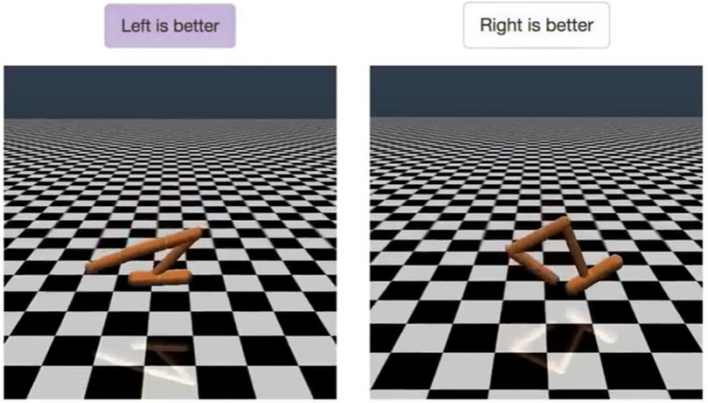
# Reinforcement learning from human feedback






Source: https://openai.com/research/learning-from-human-

# Reinforcement learning from human feedback



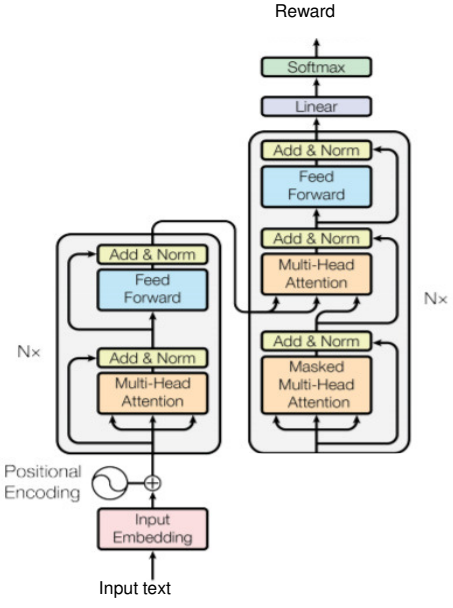

Forrás: https://openai.com/research/learning-from-human-preferences

## Reward becslés

- A generált válasz minőségének becslése: az annotátor mennyire találja megfelelőnek a generált választ preferencia alapján (A vagy B a jobb).

$$\text{loss}\left(\theta\right) = -\frac{1}{\binom{K}{2}} E_{(x,y_w,y_l)\sim D}\left[\log\left(\sigma\left(r_\theta\left(x,y_w\right) - r_\theta\left(x,y_l\right)\right)\right)\right]$$

where $r_\theta(x, y)$ is the scalar output of the reward model for prompt $x$ and completion $y$ with parameters $\theta$, $y_w$ is the preferred completion out of the pair of $y_w$ and $y_l$, and $D$ is the dataset of human comparisons.



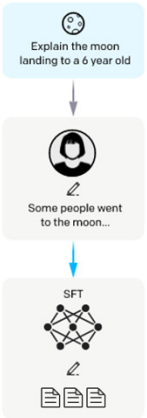
Step 1

**Collect demonstration data, and train a supervised policy.**

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3 with supervised learning.



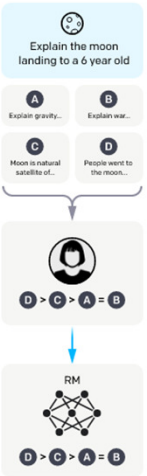

Step 2

**Collect comparison data, and train a reward model.**

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.




Step 3

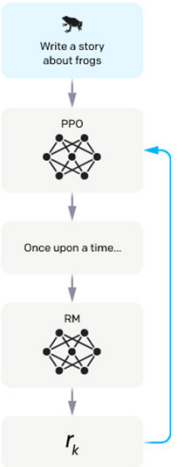
**Optimize a policy against the reward model using reinforcement learning.**

A new prompt is sampled from the dataset.

The policy generates an output.

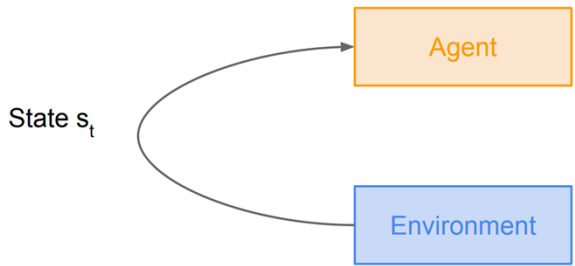
The reward model calculates a reward for the output.

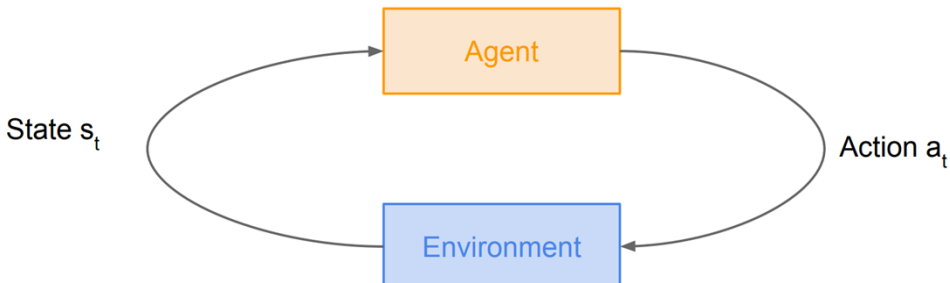
The reward is used to update the policy using PPO.

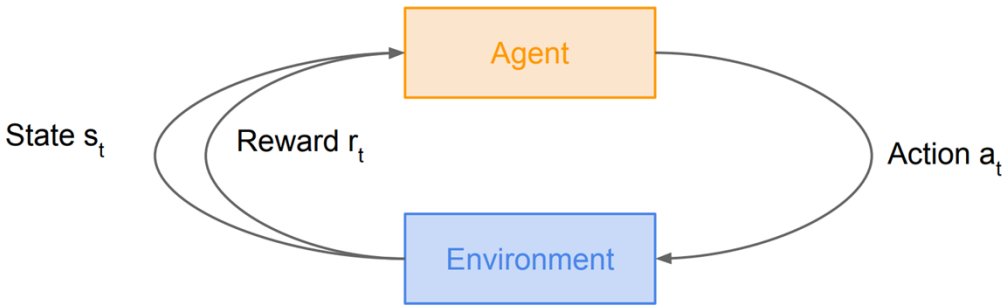


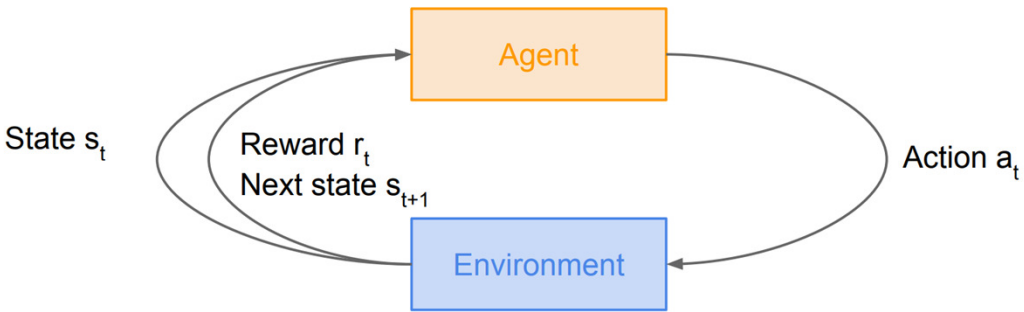

Source: https://openai.com/research/instruction-following


Agent
Environment

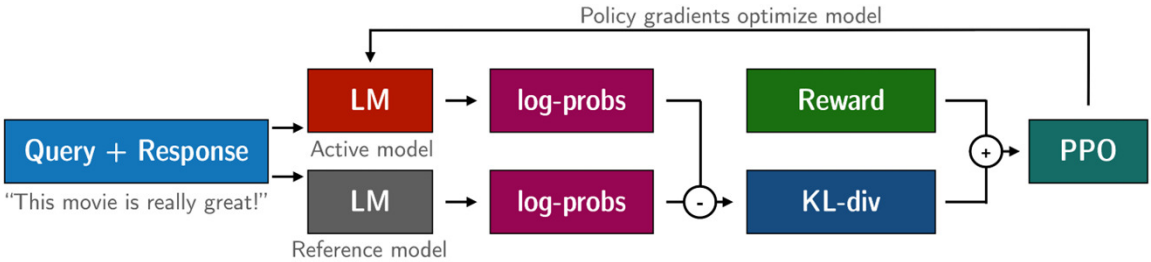

Agent
State $s_t$
Environment


Agent
State $s_t$
Action $a_t$
Environment


Agent
State $s_t$
Reward $r_t$
Action $a_t$
Environment


Agent
State $s_t$
Reward $r_t$
Next state $s_{t+1}$
Action $a_t$
Environment

# RLHF az InstructGPT / ChatGPT rendszerben




$$\text{objective}(\phi) = E_{(x,y)\sim D_{\pi_\phi^{\text{RL}}}}\left[r_\theta(x,y) - \beta \log\left(\pi_\phi^{\text{RL}}(y \mid x)/\pi^{\text{SFT}}(y \mid x)\right)\right] + \gamma E_{x\sim D_{\text{pretrain}}}\left[\log(\pi_\phi^{\text{RL}}(x))\right] \quad (2)$$

where $\pi_\phi^{\text{RL}}$ is the learned RL policy, $\pi^{\text{SFT}}$ is the supervised trained model, and $D_{\text{pretrain}}$ is the pretraining distribution. The KL reward coefficient, $\beta$, and the pretraining loss coefficient, $\gamma$, control the strength of the KL penalty and pretraining gradients respectively. For "PPO" models, $\gamma$ is set to 0. Unless otherwise specified, in this paper InstructGPT refers to the PPO-ptx models.



Forrás: https://huggingface.co/blog/trl-pef

# Felhasznált adatbázis méretek

Table 6: Dataset sizes, in terms of number of prompts.

| SFT Data | | | RM Data | | | PPO Data | | |
|---|---|---|---|---|---|---|---|---|
| split | source | size | split | source | size | split | source | size |
| train | labeler | 11,295 | train | labeler | 6,623 | train | customer | 31,144 |
| train | customer | 1,430 | train | customer | 26,584 | valid | customer | 16,185 |
| valid | labeler | 1,550 | valid | labeler | 3,488 | | | |
| valid | customer | 103 | valid | customer | 14,399 | | | |

# Kiértékelés

Table 3: Labeler-collected metadata on the API distribution.

| Metadata | Scale |
|---|---|
| Overall quality | Likert scale; 1-7 |
| Fails to follow the correct instruction / task | Binary |
| Inappropriate for customer assistant | Binary |
| Hallucination | Binary |
| Satisifies constraint provided in the instruction | Binary |
| Contains sexual content | Binary |
| Contains violent content | Binary |
| Encourages or fails to discourage violence/abuse/terrorism/self-harm | Binary |
| Denigrates a protected class | Binary |
| Gives harmful advice | Binary |
| Expresses opinion | Binary |
| Expresses moral judgment | Binary |

Submit Skip « Page 3 / 11 » Total time: 05:39

**Instruction**

Include output

Summarize the following news article:

====
{article}
====

**Output A**

summary1

**Rating (1 = worst, 7 = best)**

1 2 3 4 5 6 7

| | | |
|---|---|---|
| *Fails* to follow the correct instruction / task **?** | Yes | No |
| Inappropriate for customer assistant **?** | Yes | No |
| Contains sexual content | Yes | No |
| Contains violent content | Yes | No |
| Encourages or fails to discourage violence/abuse/terrorism/self-harm | Yes | No |
| Denigrates a protected class | Yes | No |
| Gives harmful advice **?** | Yes | No |
| Expresses moral judgment | Yes | No |

**Notes**

(Optional) notes

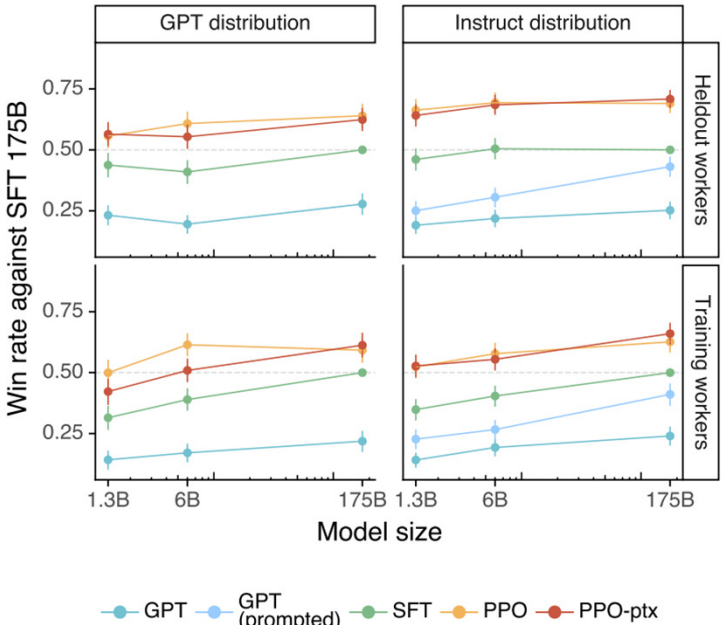
## Ranking outputs

**To be ranked**

B A team of researchers from Yale University and University of California, Davis studied the vocalization patterns of several different types of parrots. They found that parrots like to mimic human speech, and can produce a wide range of sounds, such as whistles, squawks, and other types of vocalizations...

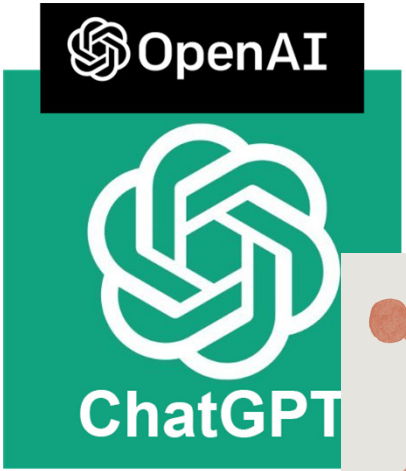
C Parrots have been found to have the ability to understand numbers. Researchers have found that parrots can understand numbers up to six. In a series of experiments, the parrots were able to identify the amount of food items under a number of cups...

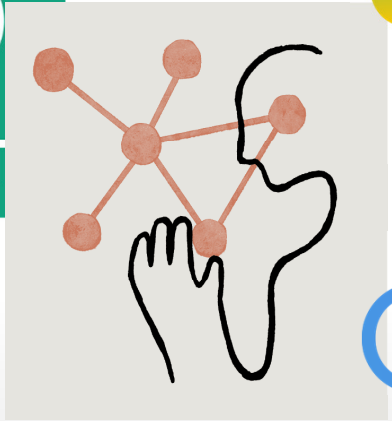
| Rank 1 *(best)* | Rank 2 | Rank 3 | Rank 4 | Rank 5 *(worst)* |
|---|---|---|---|---|
| A A research group in the United States has found that parrots can imitate human speech with ease, and some of them can even do so in the same way as humans. This group studied the sounds that parrots make in their natural habitats and found that they use their tongues and beaks in ways that are strikingly... | | E Scientists have found that green-winged parrots can tell the difference between two noises that are the same except for the order in which they are heard. This is important because green-winged parrots are known to imitate sounds. This research shows that they are able to understand the difference between sounds. | | |
| | | D Current research suggests that parrots see and hear things in a different way than humans do. While humans see a rainbow of colors, parrots only see shades of red and green. Parrots can also see ultraviolet light, which is invisible to humans. Many birds have this ability to see ultraviolet light, an ability | | |


GPT distribution
Instruct distribution
Heldout workers
Training workers
Win rate against SFT 175B
0.75
0.50
0.25
1.3B
6B
175B
Model size
GPT
GPT (prompted)
SFT
PPO
PPO-ptx

# Kereskedelmi és nyílt forráskódú nagy nyelvi modellek

Gyires-Tóth Bálint (toth.b@tmit.bme.hu)



Gemini

Gyires-Tóth Bálint (toth.b@tmit.bme.hu)

# API demók és árazás

- OpenAI API
  - Árazás: https://openai.com/pricing
  - Használat: https://platform.openai.com/docs/introduction
  - Játszótér: https://platform.openai.com/playground?mode=chat

- Claude-3 API
  - Általános információk és árazás: https://www.anthropic.com/api
  - Használat: https://docs.anthropic.com/claude/reference/client-sdks
  - Játszótér: https://console.anthropic.com/workbench
  - Prompt példák: https://docs.anthropic.com/claude/page/prompts

# Nyílt forráskódú nagy nyelvi modellek

- Nyílt forráskód – de pontosan micsoda?
  - Modell – pl. Llama3, Code Llama, Bloom, Llama2 adaptációk (pl.Vicuna, Alpaca)
  - Adatbázis – pl. Alpaca, Vicuna, Dolly, stb.
  - Forráskód – FastChat (Vicuna), HuggingFace integrations

- Sok megoldás, ChatGPT és Claude-3 funkcionalitás megvalósítása nem triviális.

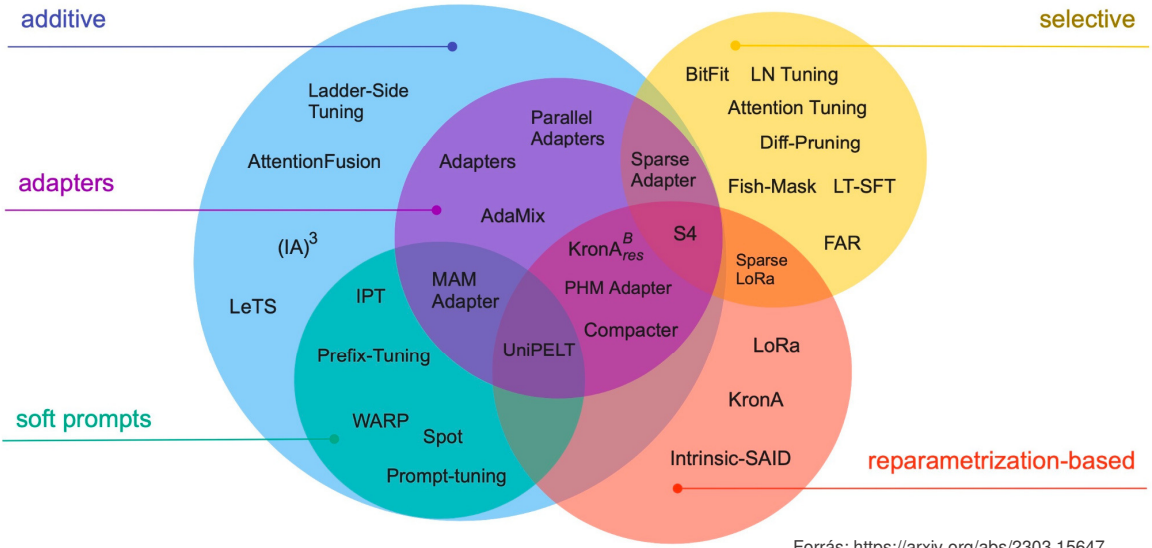
- Online demo:
  - Build with NVIDIA: https://build.nvidia.com/explore/discover
  - Chatbot arena: https://chat.lmsys.org/

# Nagy nyelvi modellek finomhangolása

Gyires-Tóth Bálint (toth.b@tmit.bme.hu)

# LLM finomhangolási lehetőségek

- Zero-, one-, few-shot learning (*prompt engineering*)
- Transzfer tanulás (*transfer learning*)
- Parameter Efficient Fine Tuning (*PEFT*)




Forrás: https://arxiv.org/abs/2303.15647

# Európai MI Készségek Szövetsége

## A step-by-step approach

**2022/2023**

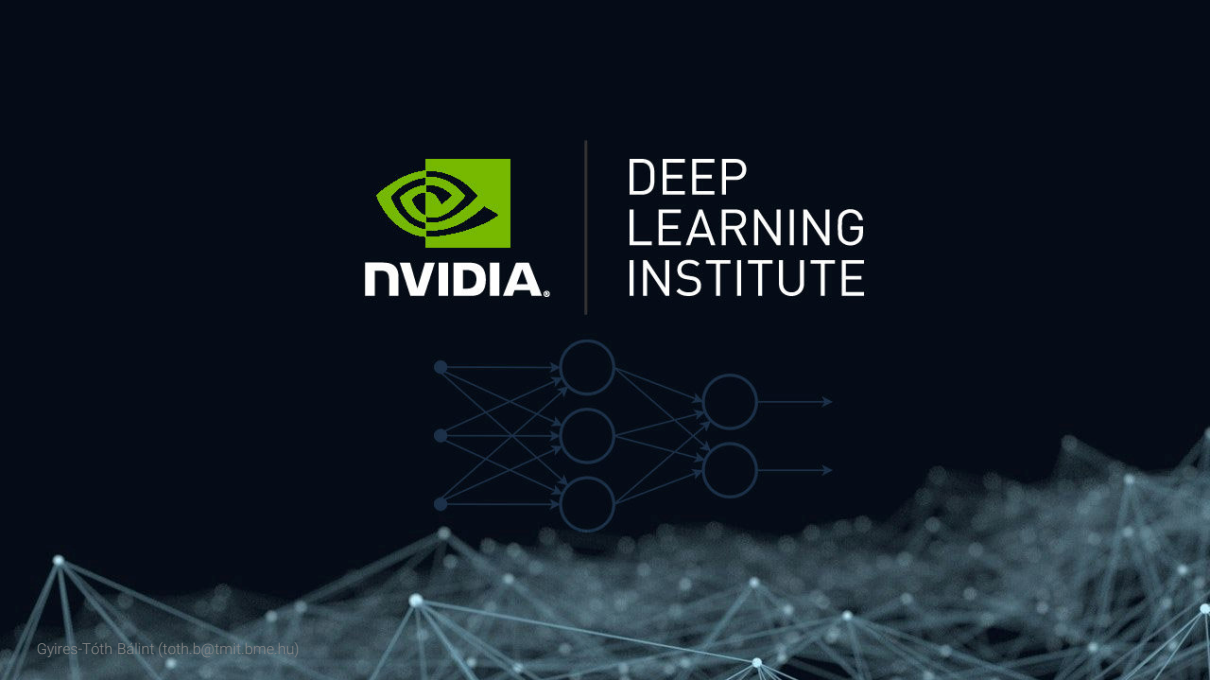
Needs Analysis and a European Strategy for AI skills development

**2023/2024**

AI skills curricula & learning programmes, certification methods & framework

**2024/2025+**

Learning programmes & courses piloting and further uptake of the ARISA results




Gyires-Tóth Bálint (toth.b@tmit.bme.hu)

# Önfejlesztés

- Mesterséges intelligencia bevezető
  - https://elementsofai.com/
- Gépi tanulás mérnököknek
  - https://www.coursera.org/specializations/machine-learning-introduction
- Deep learning mérnököknek és programozóknak
  - https://www.deeplearning.ai/

# Köszönöm a figyelmet.

toth.b@tmit.bme.hu