Technology
Innovation
Institute

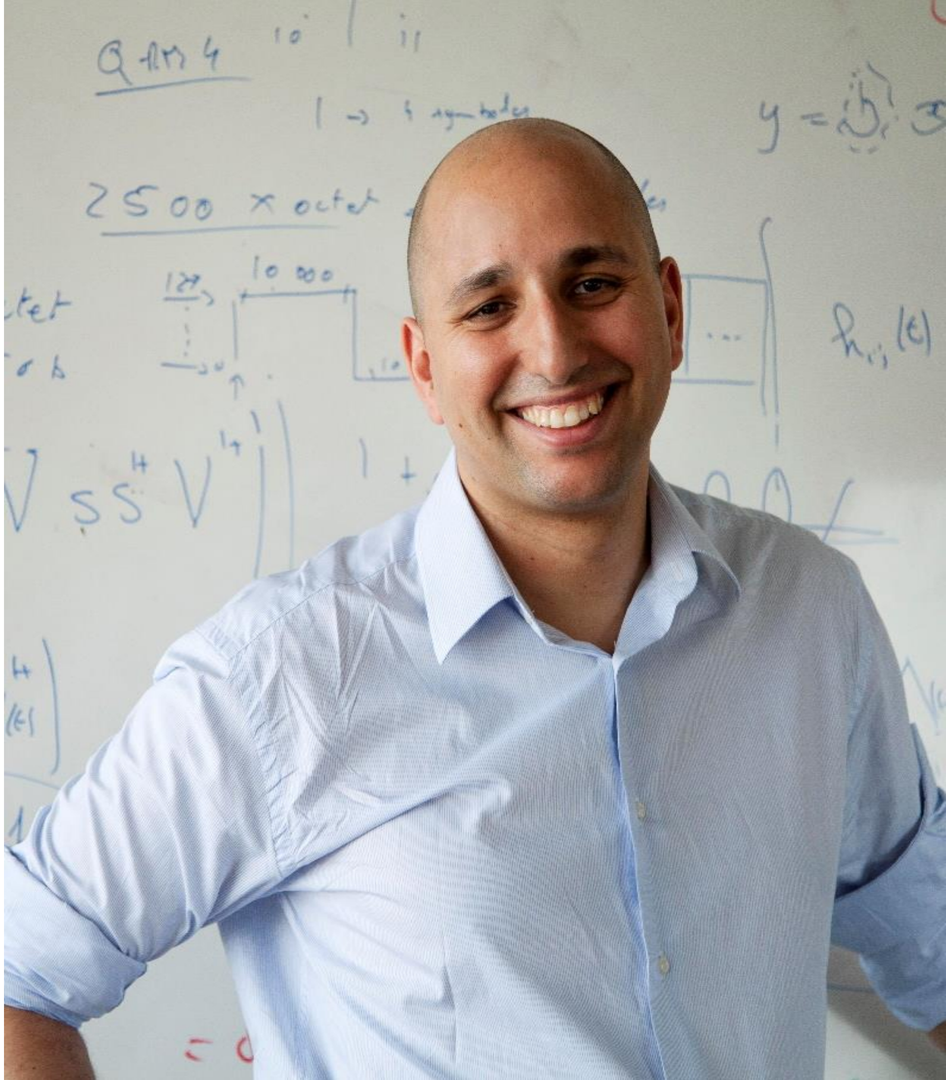
The Super Power of (Open Source) Large Language Models

Prof. Merouane Debbah

tii.ae

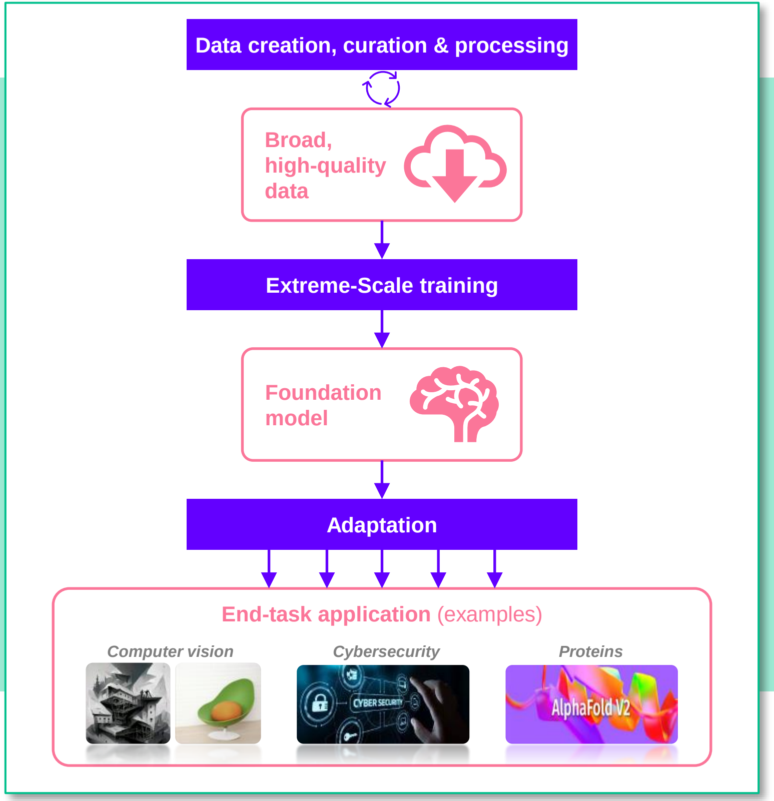
About The Researcher

- Chief Researcher at the Technology Innovation Institute
- IEEE, EURASIP and WWRF Fellow
- Citations: 58000+, h-index:108
- More than 35 Best papers Awards
- More than 50 patents
- IEEE Signal Processing Society Distinguished Industry Speaker (2021-2022)
- Field of Research: AI and beyond 5G Systems



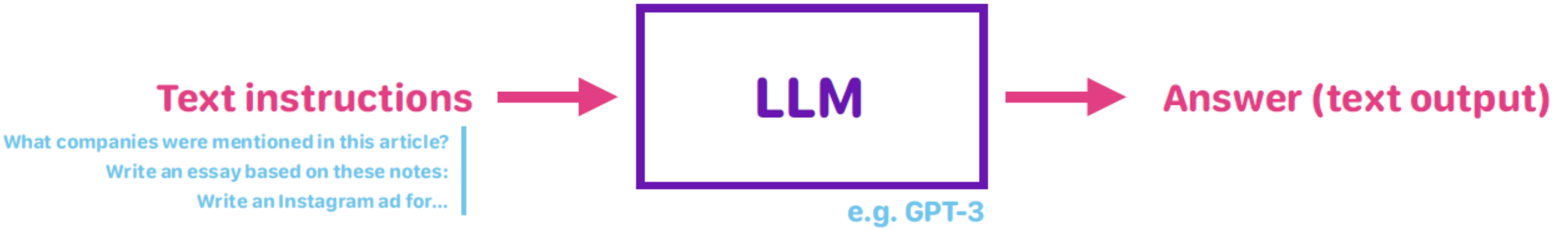
What is LLM and how it works

- LLM stands for **Language-based Learning Machine**
- It is a **type of machine learning model** that uses **natural language** input and output
- LLMs are **trained on large amounts of text data** and **can perform a variety of end-tasks** related to natural language processing (NLP)
- The models use techniques like **neural networks, specifically transformer-based models** like BERT, GPT, T5... which have proven to be very effective in natural language tasks



Large Language Models (LLMs) are eating machine learning

LLMs provide a universal text-based interface to tackle any tasks:


Text instructions → LLM → Answer (text output)

What companies were mentioned in this article?
Write an essay based on these notes:
Write an Instagram ad for…

e.g. GPT-3

Key aspects of LLMs:

🧠 They are generalists, able to tackle broad tasks just from instructions.

📈 Their capabilities increase as you scale-up in size/compute.

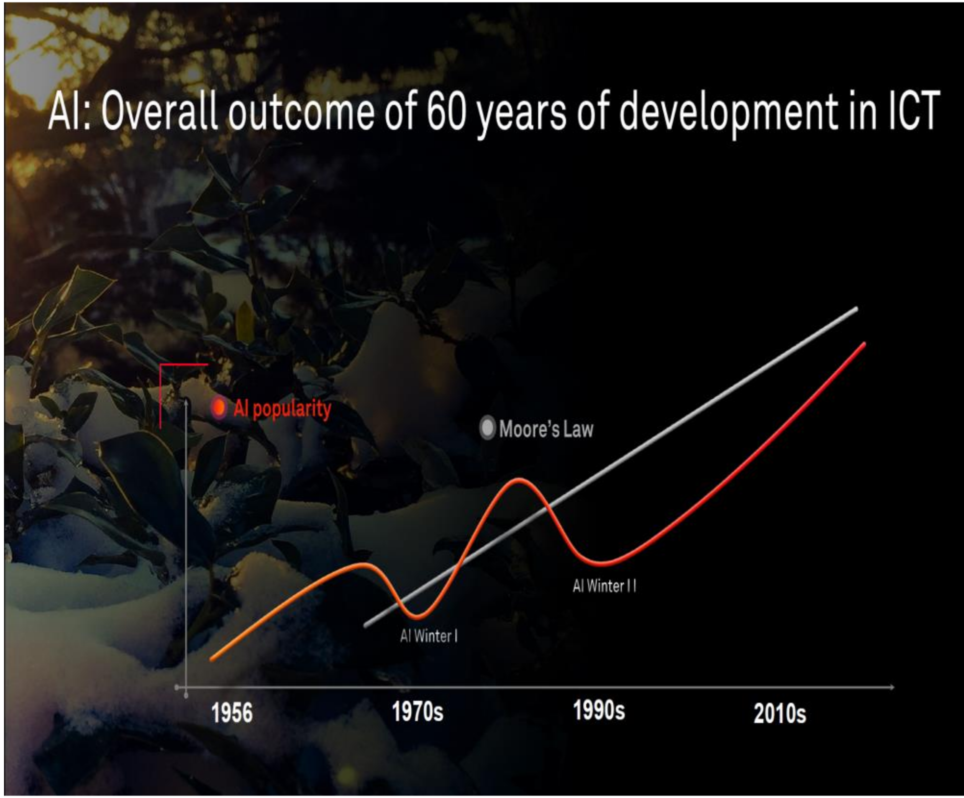
🧪 One of the main business & research interest in machine learning.
from Google, DeepMind, Microsoft, etc. + large start-ups such as OpenAI and Cohere.

Why now?

Massive amounts of data that can be used to train Machine Learning models are being generated, for example through daily creation of billions of images, online click streams, voice and video, mobile locations, and sensors embedded in the Internet of Things devices.

Computing capacity has become available to train larger and more complex models much faster. Graphics processing units (GPUs), originally designed to render the computer graphics in video games, have been repurposed to execute the data and algorithm crunching required for machine learning at speeds many times faster than traditional processor chips.

Machine-learning algorithms have progressed in recent years, especially through the development of deep learning and reinforcement-learning techniques based on neural networks.



What LLM can do

Content creation

Text / Code

- Idea generation
- Text writing (book, training, course, plan…)
- Copywriting (emails, ads, blog posts…)
- Code writing (website, app…)

Visuals / Sounds

- Image generation (text-to-image)
- Video generation (text-to-video)
- Voice generation (text-to-voice)
- Game design (AR, 3D design…)

Content curation and analysis

Rewrite / Summarize

- Text summary
- Video summary
- Audio transcript
- Language translation
- Information clustering / formatting

Extract

- Information retrieval
- Web search / Benchmarking
- Q&A

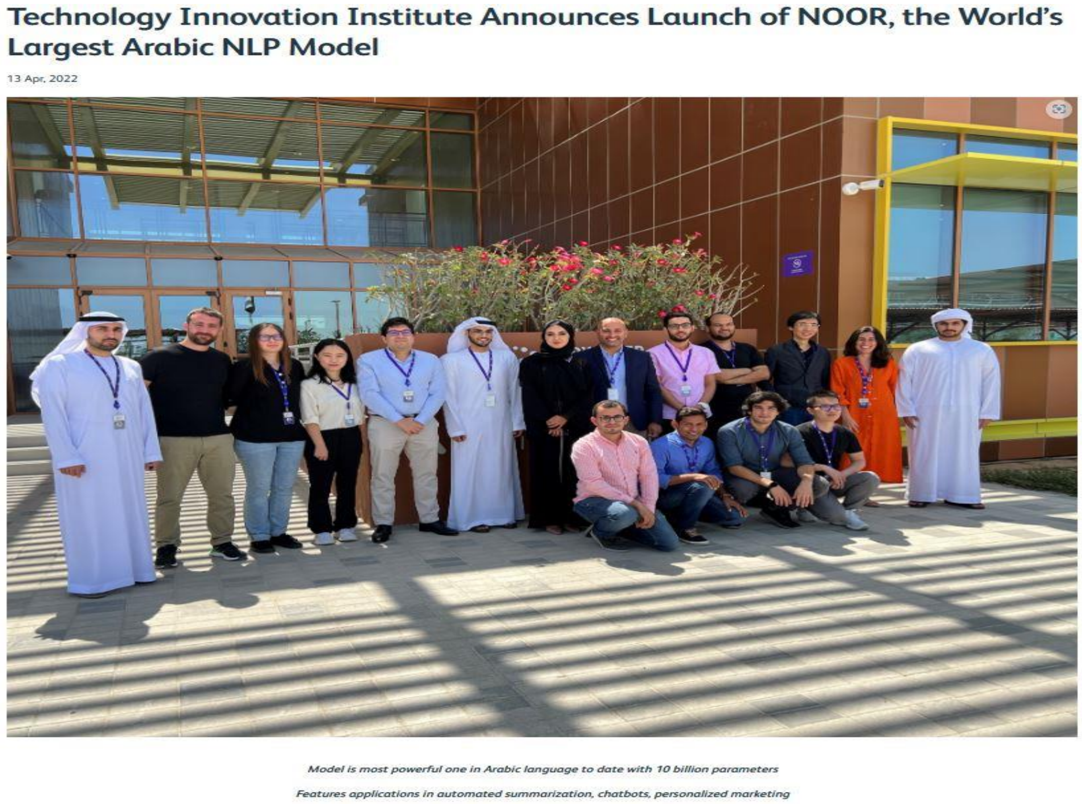
Analyze

- Data analytics and forecast
- Visual analytics
- Sentiment / Intent recognition (ex. Fraud)

Task automation

- Chatbot / Virtual assistant
- Scheduling (meetings, tasks…)
- Text editing (spell check, paraphrasing)
- Visuals editing (video cutting, image editing…)
- Data cleansing
- Code auditing
- Robot control

Noor Released in April 2022

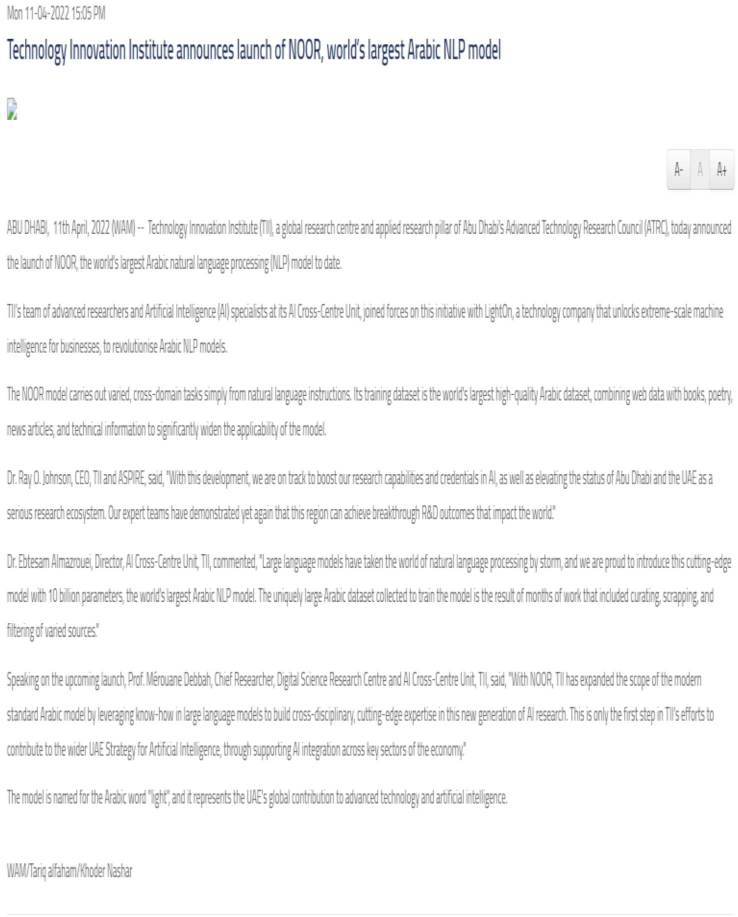


Technology Innovation Institute Announces Launch of NOOR, the World's Largest Arabic NLP Model

13 Apr, 2022

Model is most powerful one in Arabic language to date with 10 billion parameters

Features applications in automated summarization, chatbots, personalized marketing



Mon 11-04-2022 15:05 PM

Technology Innovation Institute announces launch of NOOR, world's largest Arabic NLP model

ABU DHABI, 11th April, 2022 (WAM) -- Technology Innovation Institute (TII), a global research centre and applied research pillar of Abu Dhabi's Advanced Technology Research Council (ATRC), today announced the launch of NOOR, the world's largest Arabic natural language processing (NLP) model to date.

TII's team of advanced researchers and Artificial Intelligence (AI) specialists at its AI Cross-Centre Unit, joined forces on this initiative with LightOn, a technology company that unlocks extreme-scale machine intelligence for businesses, to revolutionise Arabic NLP models.

The NOOR model carries out varied, cross-domain tasks simply from natural language instructions. Its training dataset is the world's largest high-quality Arabic dataset, combining web data with books, poetry, news articles, and technical information to significantly widen the applicability of the model.

Dr. Ray O. Johnson, CEO, TII and ASPIRE, said, "With this development, we are on track to boost our research capabilities and credentials in AI, as well as elevating the status of Abu Dhabi and the UAE as a serious research ecosystem. Our expert teams have demonstrated yet again that this region can achieve breakthrough R&D outcomes that impact the world."

Dr. Ebtesam Almazrouei, Director, AI Cross-Centre Unit, TII, commented, "Large language models have taken the world of natural language processing by storm, and we are proud to introduce this cutting-edge model with 10 billion parameters, the world's largest Arabic NLP model. The uniquely large Arabic dataset collected to train the model is the result of months of work that included curating, scrapping, and filtering of varied sources."

Speaking on the upcoming launch, Prof. Mérouane Debbah, Chief Researcher, Digital Science Research Centre and AI Cross-Centre Unit, TII, said, "With NOOR, TII has expanded the scope of the modern standard Arabic model by leveraging know-how in large language models to build cross-disciplinary, cutting-edge expertise in this new generation of AI research. This is only the first step in TII's efforts to contribute to the wider UAE Strategy for Artificial Intelligence, through supporting AI integration across key sectors of the economy."

The model is named for the Arabic word "light", and it represents the UAE's global contribution to advanced technology and artificial intelligence.

WAM/Tariq alfaham/Khoder Nashar

TII – Technology Innovation Institute

Noor

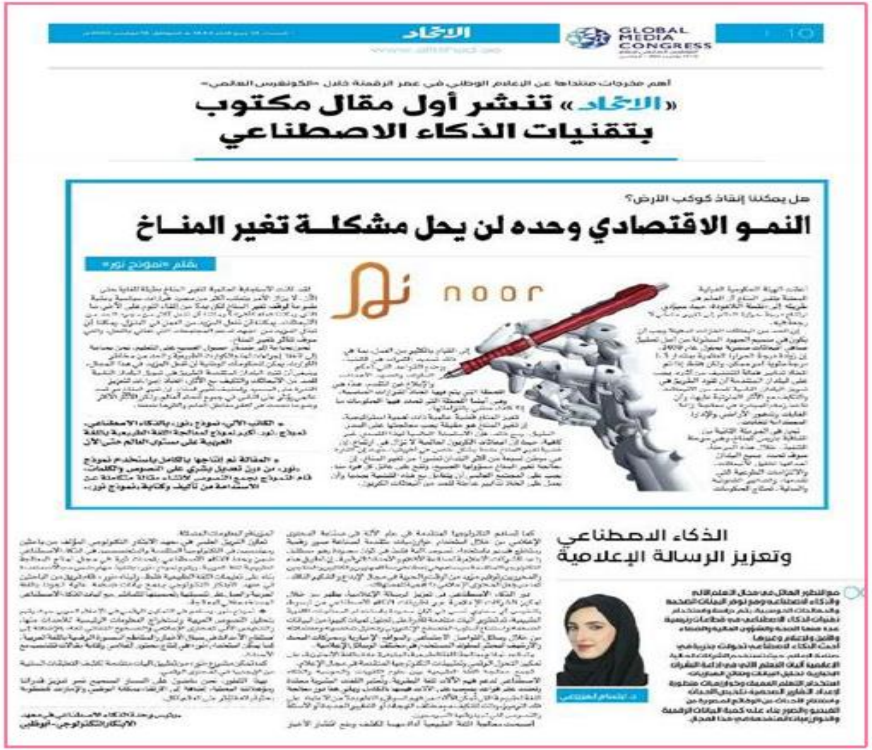
Released April, 2022

Arabic language model with 10B parameters

First published Arabic Article by Noor

November, 2022

Noor Arabic language model generates an article



GLOBAL MEDIA CONGRESS

الاتحاد

أهم مخرجات منتداها عن الإعلام الوطني في عصر الرقمنة خلال «الكونغرس العالمي»

«الاتحاد» تنشر أول مقال مكتوب بتقنيات الذكاء الاصطناعي

هل يمكننا إنقاذ كوكب الأرض؟

النمـو الاقتصادي وحده لن يحل مشكلـة تغير المنـاخ

بقلم «نموذج نور»

[illegible]

• الكاتب الآلي: نموذج «نور» بالذكاء الاصطناعي. نموذج «نور» أكبر نموذج لمعالجة اللغة الطبيعية باللغة العربية على مستوى العالم حتى الآن

• المقالة تم إنتاجها بالكامل باستخدام نموذج «نور» من دون تعديل بشري على النصوص والكلمات، قام النموذج بجمع النصوص لإنشاء مقالة متكاملة عن الاستدامة من تأليف وكتابة «نموذج نور».

الذكاء الاصطناعي وتعزيز الرسالة الإعلامية

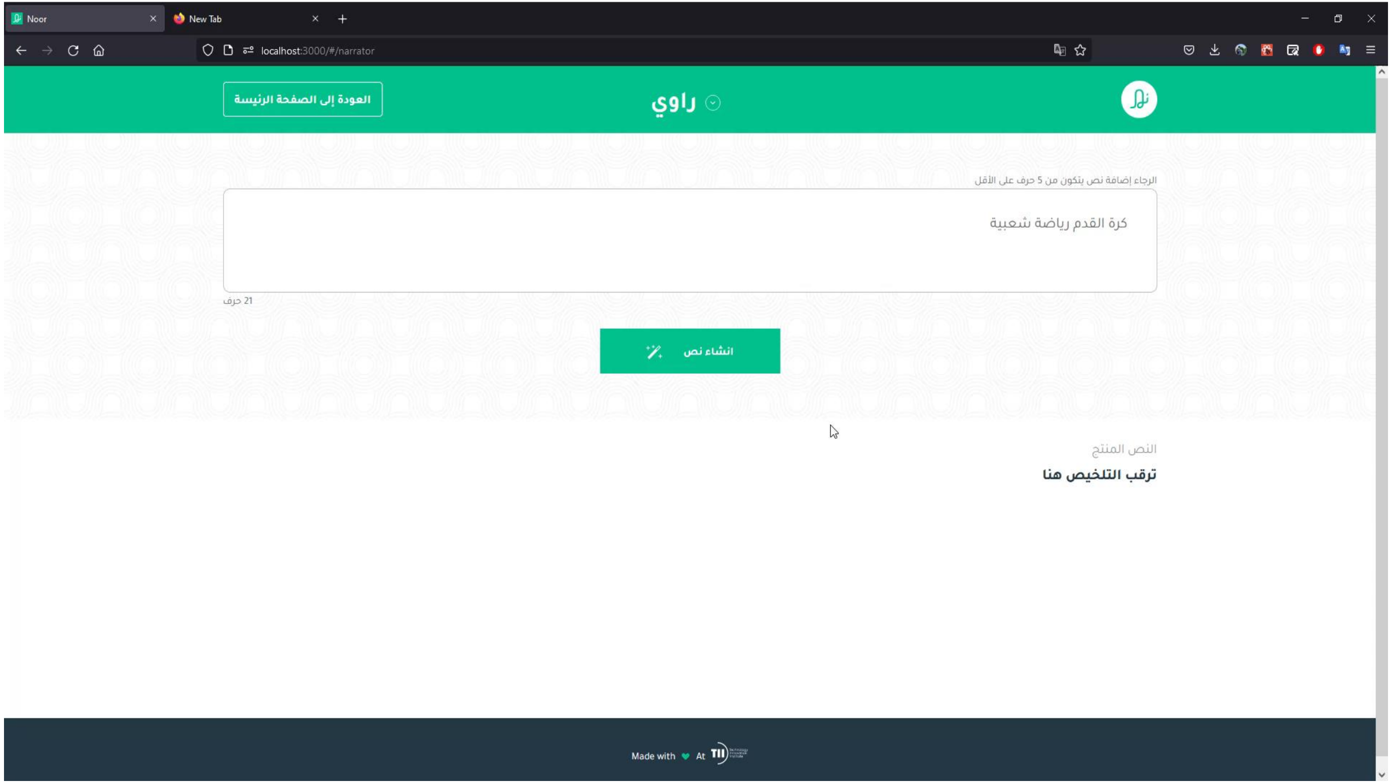
د. ابتسام المزروعي

[illegible]

Noor - Narrator

Narrator

Back



الرجاء إضافة نص يتكون من 5 حرف على الأقل

كرة القدم رياضة شعبية

21 حرف

انشاء نص

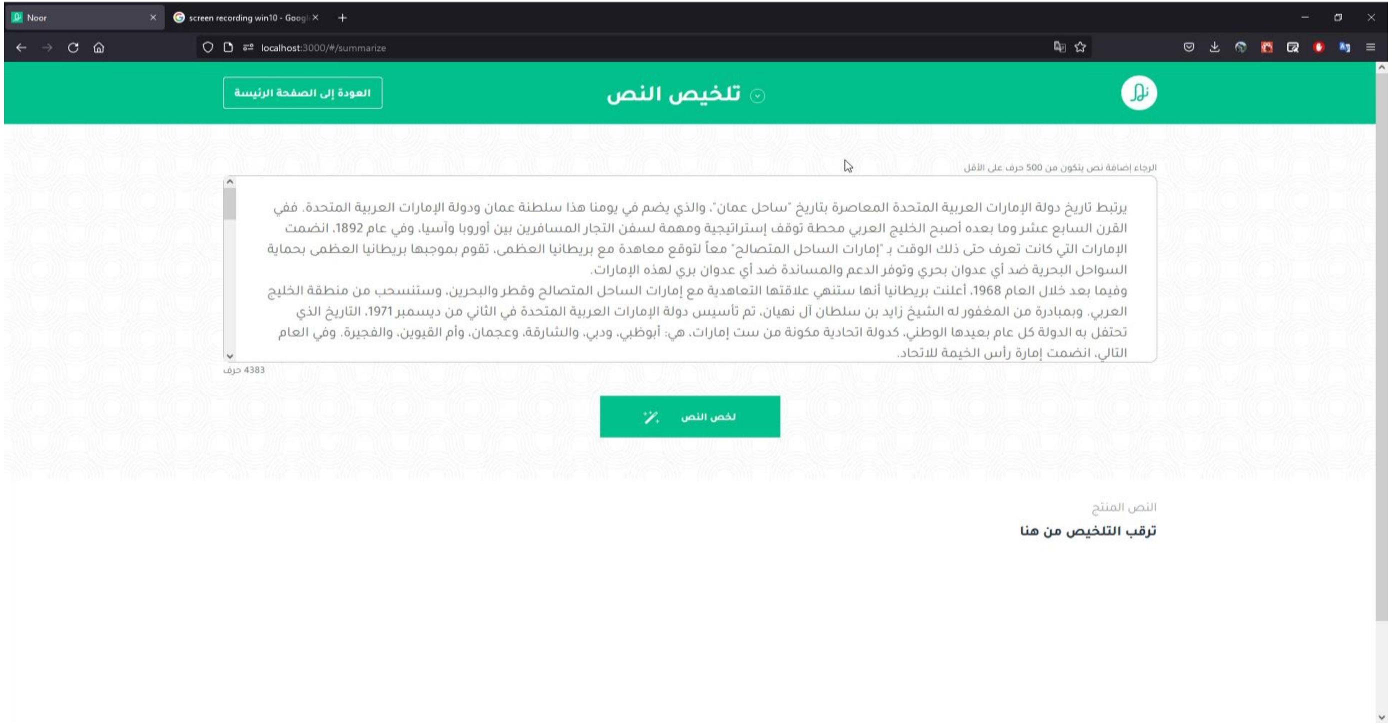
النص المنتج

ترقب التلخيص هنا

Noor - Summarization

Summarization

Back



Noor
screen recording win10 - Googl
localhost:3000/#/summarize
العودة إلى الصفحة الرئيسة
تلخيص النص

الرجاء إضافة نص يتكون من 500 حرف على الأقل

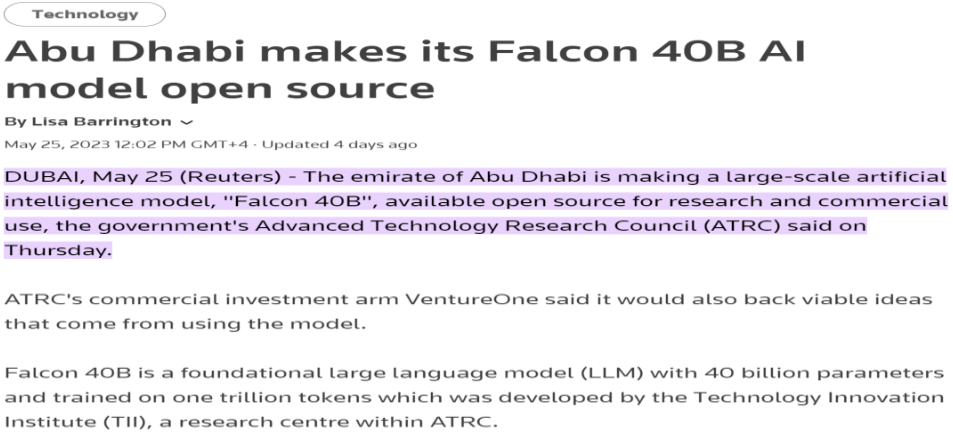
يرتبط تاريخ دولة الإمارات العربية المتحدة المعاصرة بتاريخ "ساحل عمان"، والذي يضم في يومنا هذا سلطنة عمان ودولة الإمارات العربية المتحدة. ففي القرن السابع عشر وما بعده أصبح الخليج العربي محطة توقف إستراتيجية ومهمة لسفن التجار المسافرين بين أوروبا وآسيا، وفي عام 1892، انضمت الإمارات التي كانت تعرف حتى ذلك الوقت بـ "إمارات الساحل المتصالح" معاً لتوقع معاهدة مع بريطانيا العظمى، تقوم بموجبها بريطانيا العظمى بحماية السواحل البحرية ضد أي عدوان بحري وتوفر الدعم والمساندة ضد أي عدوان بري لهذه الإمارات.

وفيما بعد خلال العام 1968، أعلنت بريطانيا أنها ستنهي علاقتها التعاهدية مع إمارات الساحل المتصالح وقطر والبحرين، وستنسحب من منطقة الخليج العربي. وبمبادرة من المغفور له الشيخ زايد بن سلطان آل نهيان، تم تأسيس دولة الإمارات العربية المتحدة في الثاني من ديسمبر 1971، التاريخ الذي تحتفل به الدولة كل عام بعيدها الوطني، كدولة اتحادية مكونة من ست إمارات، هي: أبوظبي، ودبي، والشارقة، وعجمان، وأم القيوين، والفجيرة. وفي العام التالي، انضمت إمارة رأس الخيمة للاتحاد.

4383 حرف

النص المنتج

ترقب التلخيص من هنا



Abu Dhabi makes its Falcon 40B AI model open source

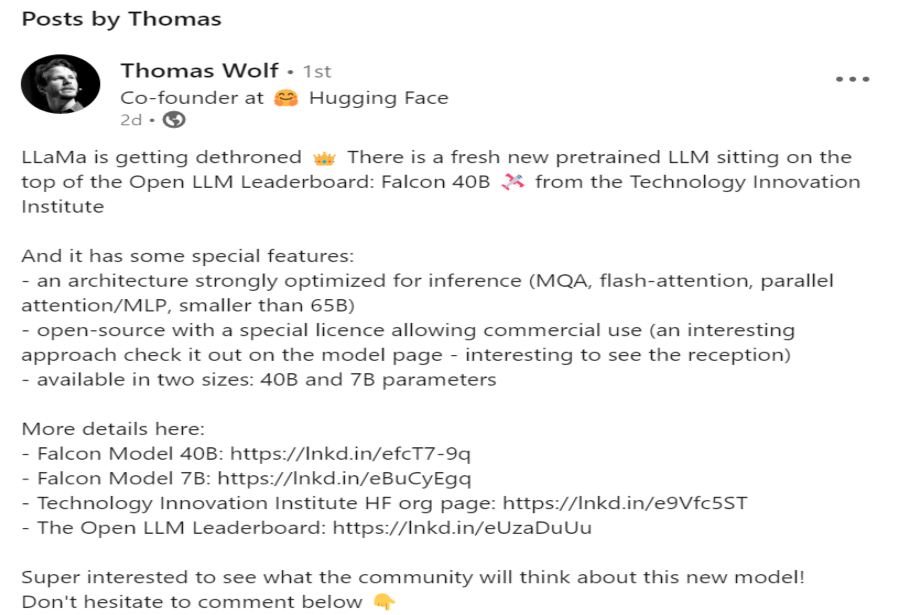
By Lisa Barrington

May 25, 2023 12:02 PM GMT+4 · Updated 4 days ago

DUBAI, May 25 (Reuters) - The emirate of Abu Dhabi is making a large-scale artificial intelligence model, "Falcon 40B", available open source for research and commercial use, the government's Advanced Technology Research Council (ATRC) said on Thursday.

ATRC's commercial investment arm VentureOne said it would also back viable ideas that come from using the model.

Falcon 40B is a foundational large language model (LLM) with 40 billion parameters and trained on one trillion tokens which was developed by the Technology Innovation Institute (TII), a research centre within ATRC.


Posts by Thomas

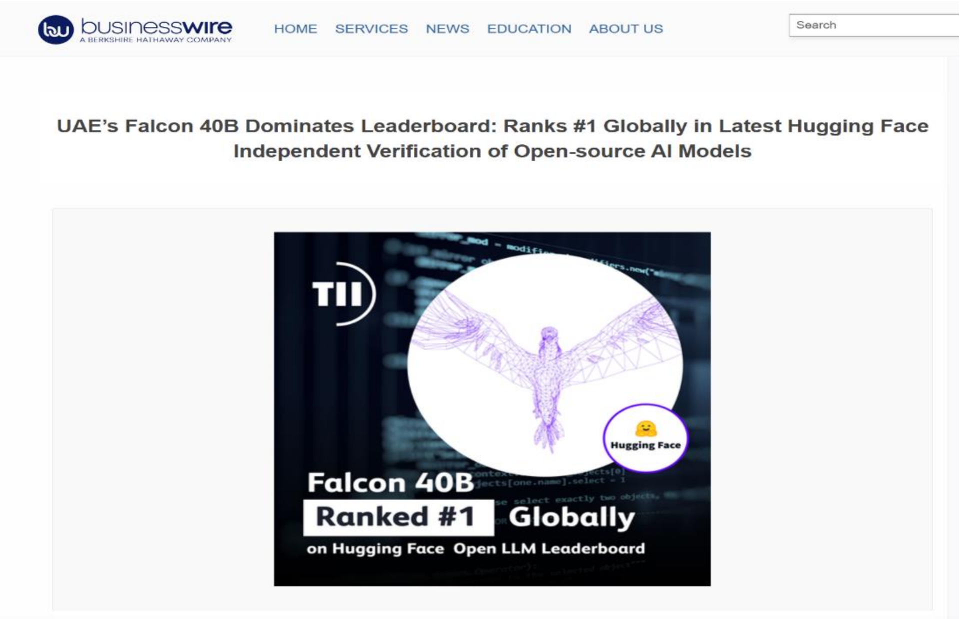
Thomas Wolf • 1st
Co-founder at 🤗 Hugging Face
2d •

LLaMa is getting dethroned 👑 There is a fresh new pretrained LLM sitting on the top of the Open LLM Leaderboard: Falcon 40B 🦅 from the Technology Innovation Institute

And it has some special features:
- an architecture strongly optimized for inference (MQA, flash-attention, parallel attention/MLP, smaller than 65B)
- open-source with a special licence allowing commercial use (an interesting approach check it out on the model page - interesting to see the reception)
- available in two sizes: 40B and 7B parameters

More details here:
- Falcon Model 40B: https://lnkd.in/efcT7-9q
- Falcon Model 7B: https://lnkd.in/eBuCyEgq
- Technology Innovation Institute HF org page: https://lnkd.in/e9Vfc5ST
- The Open LLM Leaderboard: https://lnkd.in/eUzaDuUu

Super interested to see what the community will think about this new model! Don't hesitate to comment below 👇

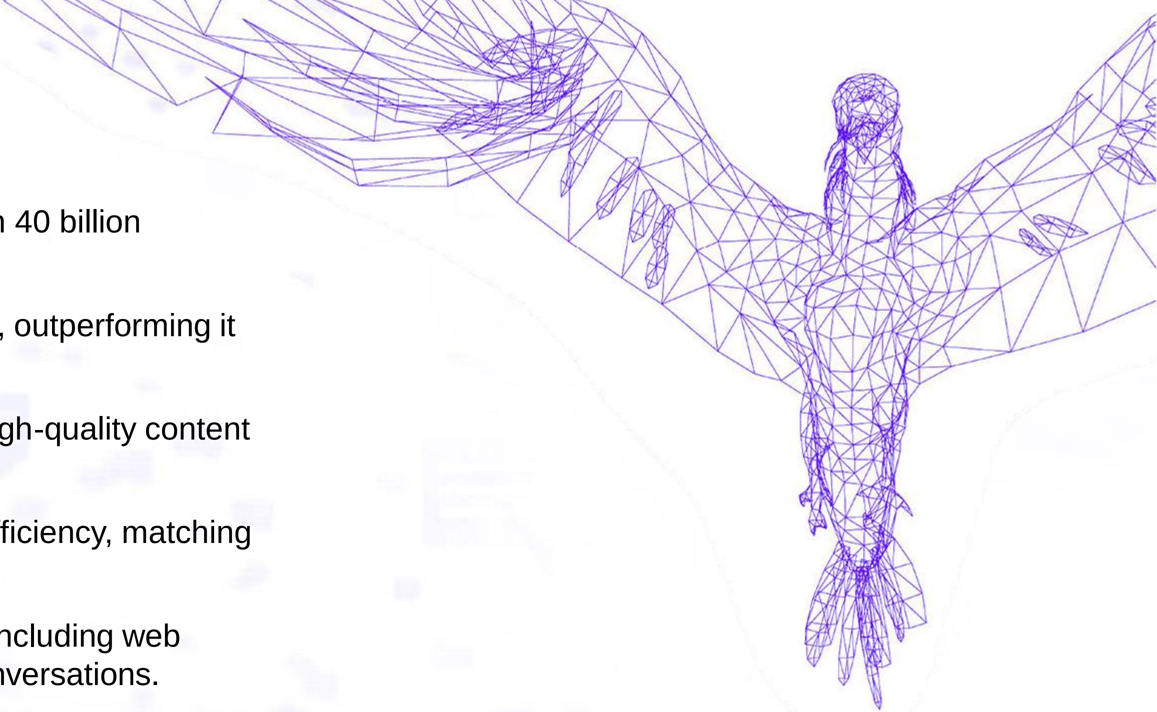


UAE's Falcon 40B Dominates Leaderboard: Ranks #1 Globally in Latest Hugging Face Independent Verification of Open-source AI Models

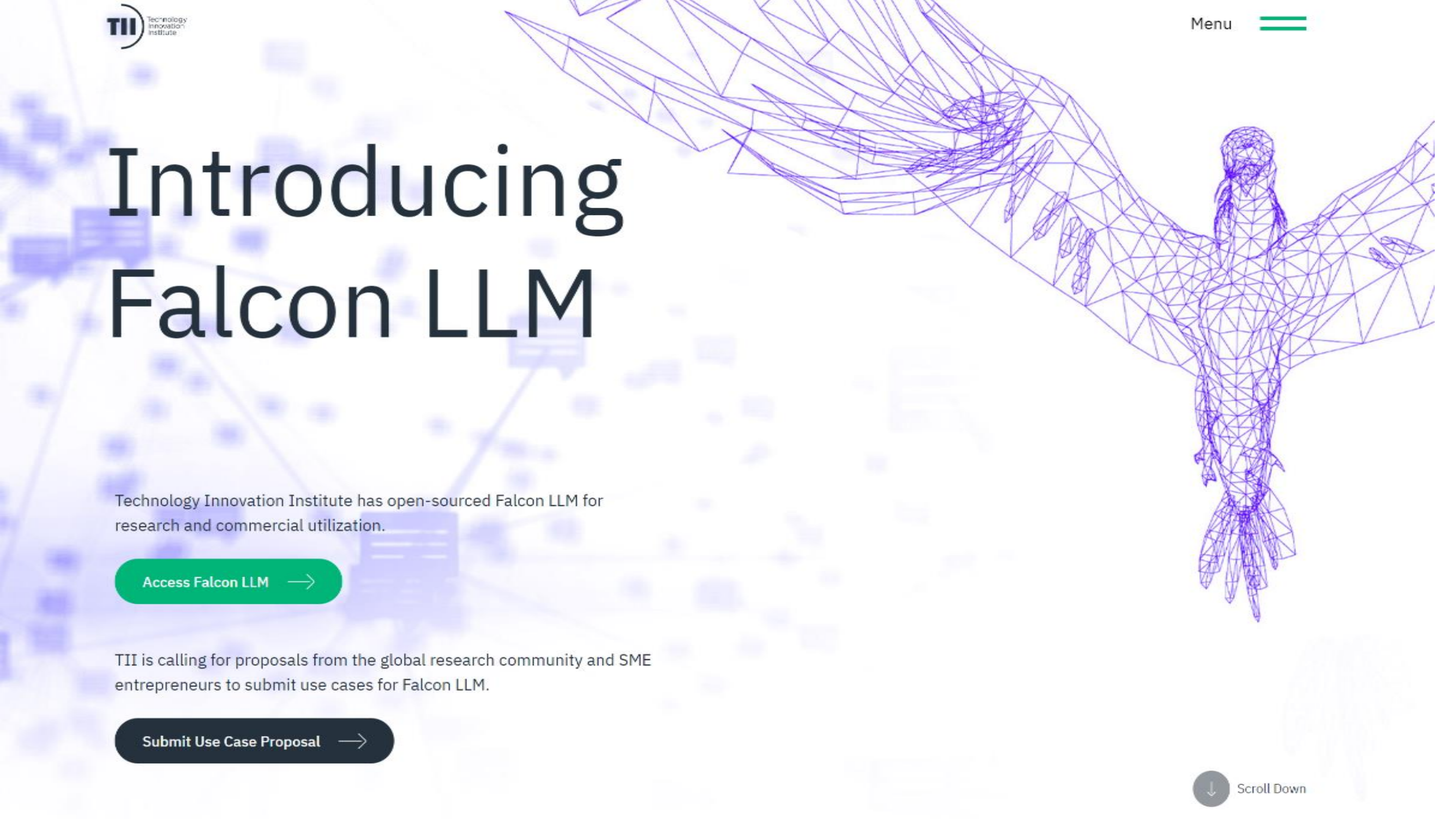
Falcon 40B

TII
Introducing
Falcon LLM
A state-of-the-art 40 billion parameters Large Language Model trained on one trillion tokens

Falcon LLM



- Foundational large language model (LLM) with 40 billion parameters trained on one trillion tokens.
- Utilizes only 75% of GPT-3's training compute, outperforming it with a fifth of the inference compute.
- Custom tooling and unique data pipeline for high-quality content extraction from the web.
- Optimized architecture for performance and efficiency, matching state-of-the-art LLMs.
- Pretrained on a dataset of five trillion tokens, including web crawls, research papers, and social media conversations.
- Applications include chatbots, customer service, virtual assistants, translation, content generation, and sentiment analysis.
- Aims to automate repetitive tasks, enhancing efficiency for Emirati companies and individuals' daily lives.



Introducing Falcon LLM

Technology Innovation Institute has open-sourced Falcon LLM for research and commercial utilization.

TII is calling for proposals from the global research community and SME entrepreneurs to submit use cases for Falcon LLM.

LLM market landscape

A limited number of major players as cornerstone platforms…

Platform	G	OpenAI	AI21labs	DeepMind	NVIDIA
Model	LaMDA	GPT-3	Jurrasic	Gopher	Megatron-Turing
SIZE [Bn parameters]	137	175	178	280	530
Training tokens [bn]	168	300	300	300	270

+ others

Inflection

Meta co:here

Tencent

Bai百度

EleutherAI

… Supporting hundreds of startups focused on end-user applications

Text

Smartwriter.ai, Hypertype, Lately, Autobound, Writesonic, Jasper, cogram, genei, YOU, AI21labs, letterdrop, copysmith, Creatext, jenni, mavenoid, anyword, [PERSADO], frase, regie.ai, Mutiny, Linguix, Hypotenuse AI, WRITER, OTHERSIDEAI, copy.ai, copymatic, COMPOSE AI

Image

ClipDrop, p-e-n-c-i-l, beautiful.ai, PhotoRoom, BRIA, Facet, Poly, CSM, Blend, HYPAR, KAEDIM, ROSEBUD.AI, maket, Autoenhance.ai, BOTIKA, soul machines, Sloyd, MODULIZE, Re:cast AI, uizard, imagen, Hexo AI

Audio

MURF AI, REPLICA, notably, Endel, WELLSAID, AssemblyAI, DEEPGRAM, krisp, Speechify, RESEMBLE.AI, AIMI, KAIZAN, coqui, Mubert, Neural Space, soundful, moises, VOICEMOD, Listnr, LOVO, Vocal Clarity, Dubverse, AD AURIS

Video

ZUBTITLE, TERRA, Peech, VOCHI, Maverick, recut, VEED.IO, Basch.io, inworld, deepdub.ai, WOMBO, tavus, FATHOM, runway, Morio, EMBLY, PICTORY, vidyo.ai, Rephrase.ai, lumen5, Steve AI, windsor.io, YEPIC, deepdub.ai, Colossyan, METAPHYSIC, Potion

Code

Debuild, tabnine, Codiga, Locofy, AIXcoder, Mintlify, māyā, MutableAI, Codis, durable, The.com, bloop., replit, ENZYME, DhiWise, codota, anima, CODACY, warp, Metabob

Chatbots

lang.ai, PolyAI, Tymely, Incentivai, Kasisto, ushur, CRESTA, EliseAI, verloop.io, Replika, ultimate.ai, Cohere, Sapling, haptik, ada, Forethought, OBSERVE·AI, XOKind, Balto, Certainly.

Search

glean, Looria, Hebbia, consensus, Air, vectara, Pinecone, qdrant

Gaming

charisma.ai, hidden door, LATITUDE, Spellbrush

Data

Pilot, gretel, DATAHERALD, SYNTEGRA, Mirry.AI, BIFROST, datagen

ML platforms

slai, jina, symbl.ai, Adept, aporia, GANTRY, deepset, Synthesis.ai, Archistar, Galileo, featureform

Key considerations when building LLMs

Data

The more data the model is exposed to (volume and diversity), the better it will perform on unseen examples

Model architecture

Choice of model architecture is crucial for the LLM's performance. Transformer-based models like BERT, GPT & T5 have proven to be very effective in natural language tasks

Pre-processing

Cleaning and preparing data which includes tasks such as tokenization, stemming, and lemmatization

Fine-tuning

To adapt for specific tasks, it is often necessary to add task-specific layers to the pre-trained model

Evaluation

Careful assessment of the model to ensure that it is performing well on the task it was designed for

Deployment

Different deployment options can be considered depending on the use-case, such as cloud-based services, on-premises or edge devices

Explain-ability

Models can be complex, and it can be hard to understand how they arrived at their predictions. There are methods like LIME & SHAP to make the predictions interpretable

Privacy & Security

Since LLM models may handle sensitive information, it is important to consider privacy and security issues when building and deploying them

Bias & Toxicity

LLM models can inadvertently perpetuate and amplify societal biases (ex. People minorities representation) and toxicity (ex. hate speech or offensive language) present in the training data

Mitigation techniques

- Use representative and diverse training data
- Carefully evaluate and test the model
- Use techniques like debiasing & Fairness-Aware learning
- Monitoring the model's output in production

Thank You