

# منتدى الدوحة للبيانات

من أجل الابتكار في التنمية المستدامة

23-22 أكتوبر 2024

**DOHA DATA FORUM**  
FOR INNOVATION IN SUSTAINABLE DEVELOPMENT

October 22-23, 2024



فانار  
Fanar

# Fanar: Arabic-Centric Large Language Model

Mohamed Eltabakh

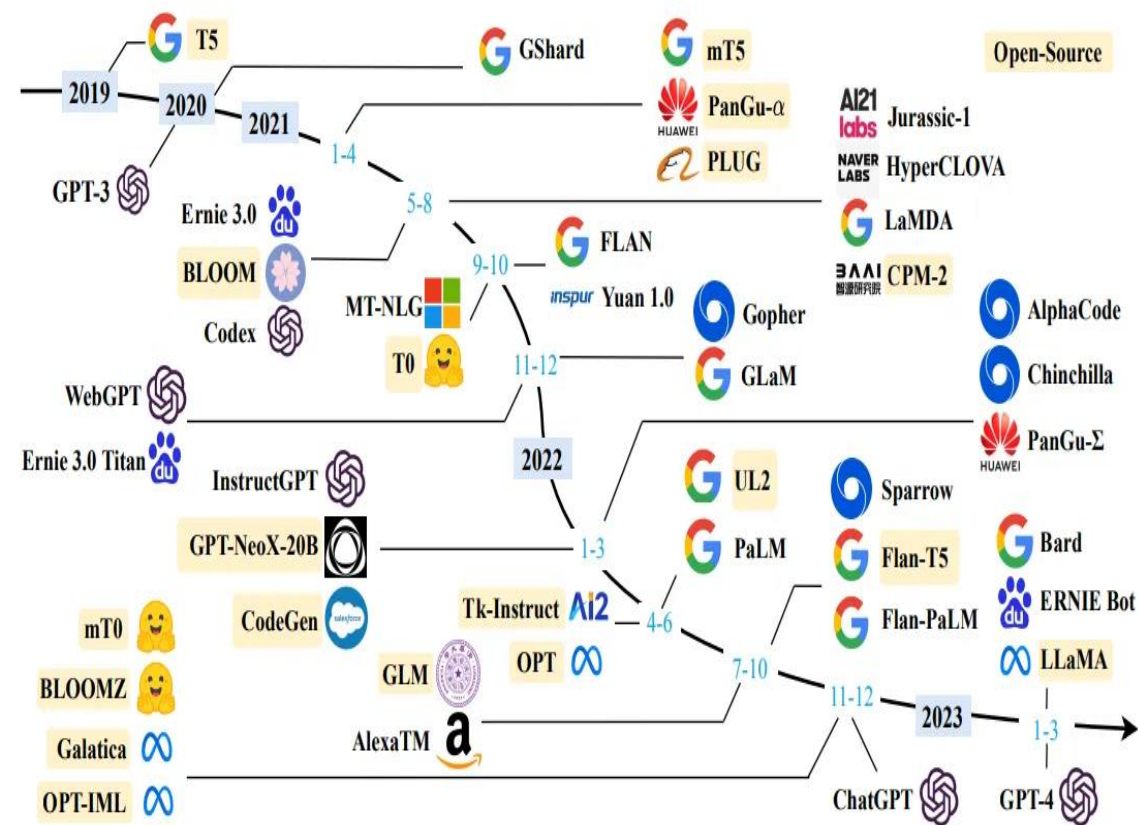
Principal Scientist, Qatar Computing Research Institute (QCRI)

منتدى الدوحة للبيانات  
من أجل الابتكار في التنمية المستدامة  
DOHA DATA FORUM  
FOR INNOVATION IN SUSTAINABLE DEVELOPMENT

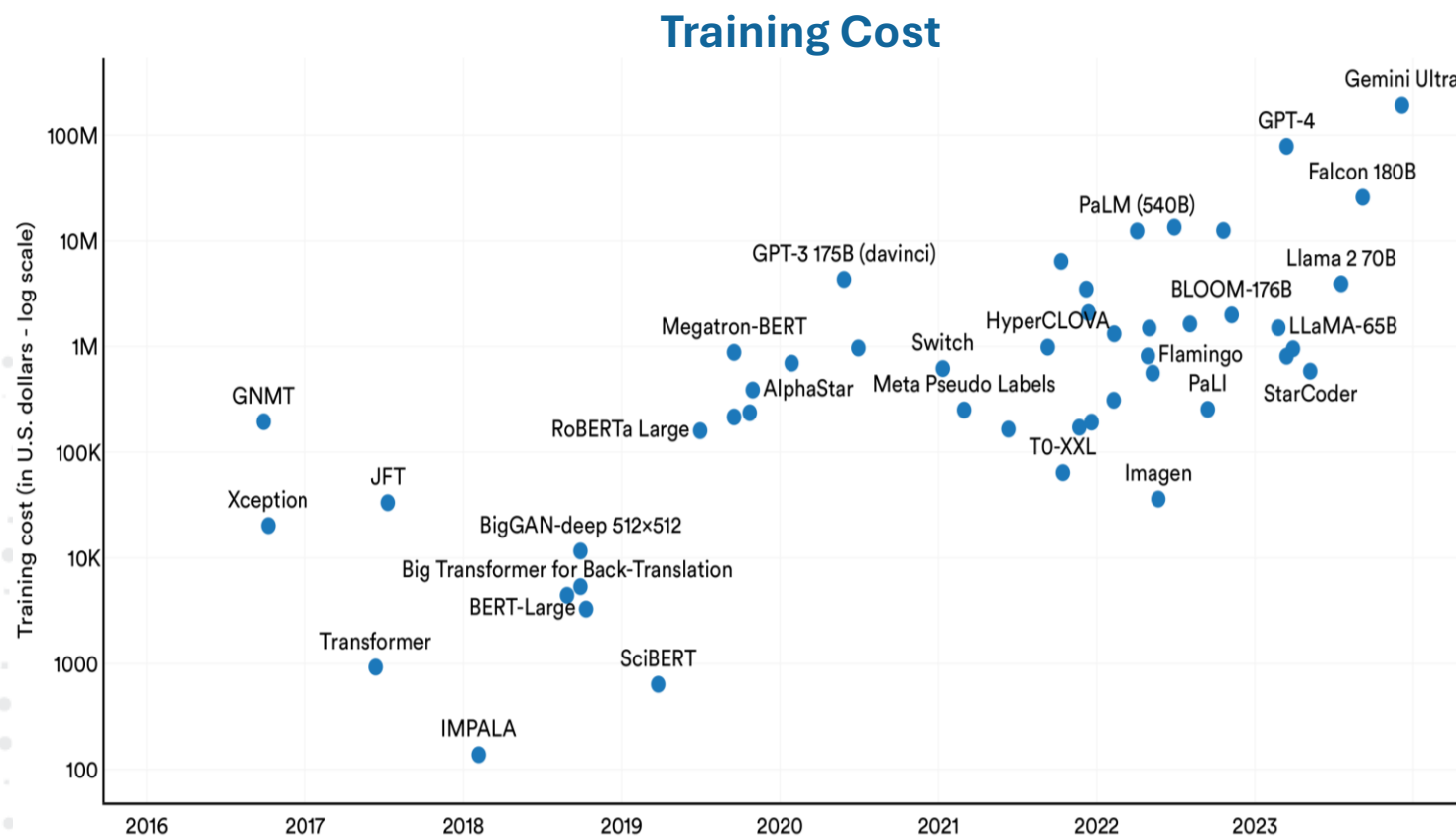


# Large Language Models (LLMs)

- LLMs are massive deep-learning models pre-trained on *huge amounts of data (the bigger the better)*
- Game-changer technology in so many aspects
- Several super-large & super-powerful models are built & more to come
- Profound impact on societies and applications



# LLM Pretraining from Scratch



Extremely expensive

Complex

Very prolonged process



LLM training

=

Space rocket launch

SpaceX's Falcon 9 rocket launches have been advertised at around **\$62 million** per launch

# What is Fanar (Release Date: Dec. 12<sup>th</sup>)

---



- **A family of highly capable Arabic-centric LLMs**
  - Trained on the largest amount of Arabic content
    - **Raw Arabic Data → over 1 trillion tokens → After cleansing → 400 Billion tokens**
- **Being developed fully from scratch in-house**
- **Modern Standard Arabic (MSA) + Multi-dialectal support**
- **Highly tuned and custom components**
  - Novel tokenizer, Arabic-oriented data mix, SFT and alignment [region, culture, religion]
- **Multi-modality support (e.g., audio, images, etc.)**

# Fanar- A National Project



## Project Sponsor

وزارة الاتصالات وتكنولوجيا المعلومات  
Ministry of Communications and Information Technology  
دولة قطر • State of Qatar



## Project inception, design, and development

QCRI  
معهد قطر لبحوث الحوسبة  
Qatar Computing Research Institute  
جامعة حمد بن خليفة  
HAMAD BIN KHALIFA UNIVERSITY



## Infrastructure Provider

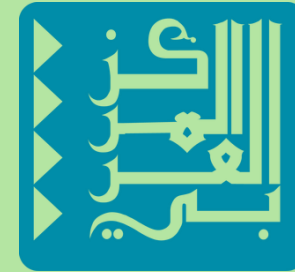


Google Cloud

## Cross-Partners Project Management

Liverx  
Innovative Services

## Project Partners



مكتبة قطر الوطنية  
Qatar National Library



جامعة قطر  
QATAR UNIVERSITY

وزارة الأوقاف والشؤون الإسلامية  
Ministry of Endowments and Islamic Affairs  
دولة قطر • State of Qatar



# How can LLMs (Fanar) Help in Data Statistics Domain

---



**Data Cleaning  
& Validation**



**Data Querying  
& Exploration**



TABLE (3)

المغادرون حسب مجموعات جنسيات الدول

الربع الأول ٢٠٢٢

DEPARTURES BY COUNTRY OF NATIONALITY GROUPS  
The First Quarter, 2022

Qatar قطر

جدو

## REGISTERED LIVE BIRTHS BY MONTH NATIONALITY, GENDER AND MUNICIPALITY 2016

TABLE (36)

جدول (٣٦)

Month Municipality	Gender	المجموع Total	ديسمبر Dec.	نوفمبر Nov.	اكتوبر Oct.	سبتمبر Sep.	اغسطس Aug.	يوليو Jul.	يونيو Jun.	مايو May.	ابريل Apr.	مارس Mar.	فبراير Feb.	يناير Jan.	النوع	الشهر البلدية
DOHA	M	5,347	402	469	499	442	437	426	485	473	425	447	385	457	ذكور	الدوحة
	F	5,108	435	430	480	409	437	392	454	421	414	413	401	422	إناث	
	T	10,455	837	899	979	851	874	818	939	894	839	860	786	879	مجموع	
AL RAYYAN	M	4,986	516	413	481	401	414	412	439	418	349	347	396	400	ذكور	الريان
	F	4,843	481	476	471	384	394	436	416	356	362	331	371	365	إناث	
	T	9,829	997	889	952	785	808	848	855	774	711	678	767	765	مجموع	
AL WAKRA	M	697	52	54	70	63	55	41	60	57	48	65	54	78	ذكور	الوكرة
	F	726	39	56	74	54	48	44	68	71	70	69	69	64	إناث	
	T	1,423	91	110	144	117	103	85	128	128	118	134	123	142	مجموع	
UMM SALAL	M	783	60	66	76	73	45	79	45	71	65	59	65	79	ذكور	ام صلال
	F	735	75	68	59	61	54	67	70	58	57	53	57	56	إناث	
	T	1,518	135	134	135	134	99	146	115	129	122	112	122	135	مجموع	
AL KHOR	M	513	43	41	39	46	30	39	34	58	49	28	50	56	ذكور	الخور
	F	508	36	35	48	24	36	42	61	48	45	37	51	45	إناث	
	T	1,021	79	76	87	70	66	81	95	106	94	65	101	101	مجموع	
AL SHAMAL	M	72	9	6	7	5	7	11	2	2	4	6	4	9	ذكور	الشمال
	F	53	4	5	7	3	6	5	1	5	1	5	6	5	إناث	
	T	125	13	11	14	8	13	16	3	7	5	11	10	14	مجموع	
AL DHAAYEN	M	277	27	22	29	11	23	21	17	18	22	33	31	23	ذكور	الظعنين
	F	282	19	27	27	20	18	22	21	23	25	25	27	28	إناث	
	T	559	46	49	56	31	41	43	38	41	47	58	58	51	مجموع	
AL SHEEHANIYA	M	800	59	85	72	55	49	71	52	72	75	69	64	77	ذكور	الشيحية
	F	835	58	72	73	67	71	62	74	74	61	71	71	81	إناث	
	T	1,635	117	157	145	122	120	133	126	146	136	140	135	158	مجموع	
OVERSEAS	M	116	4	9	10	11	5	22	3	10	7	15	9	11	ذكور	خارج قطر
	F	135	11	8	8	9	8	20	8	12	15	16	11	9	إناث	
	T	251	15	17	18	20	13	42	11	22	22	31	20	20	مجموع	
Total	M	13,591	1,172	1,165	1,283	1,107	1,065	1,122	1,137	1,179	1,044	1,069	1,058	1,190	ذكور	المجموع
	F	13,225	1,158	1,177	1,247	1,031	1,072	1,090	1,173	1,068	1,050	1,020	1,064	1,075	إناث	
	T	26,816	2,330	2,342	2,530	2,138	2,137	2,212	2,310	2,247	2,094	2,089	2,122	2,265	مجموع	



# Data Querying & Exploration Powered by LLMs

## PSA Discovery Portal

**Quick Search**  
Seamlessly search across 1000's of PSA data files

**View Tables**  
View retrieved tables relevant to your query

**Ask Questions**  
Express your query in simple natural language

I am interested in Femal Qatari population in years 2018 and 2019

Planning and Statistics Authority presents you with best regards, and with reference to your request, we are pleased to provide you with the required data via the links below.

- 1. Your information may be found at [Education\\_Chapter\\_4\\_2021\\_AE](#) on the following page: 22
- 2. Your information may be found at [Education\\_Chapter\\_4\\_2018\\_AE](#) on the following page: 22
- 3. Your information may be found at [Education\\_Chapter\\_4\\_2018\\_AE](#) on the following page: 4
- 4. Your information may be found at [Education\\_Chapter\\_4\\_2018\\_AE](#) on the following page: 23

For more information, please visit us at: [www.psa.gov.qa](http://www.psa.gov.qa) For further inquiries, you can contact us on the following e-mail: [pdr@psa.gov.qa](mailto:pdr@psa.gov.qa)

What is the ratio between Qatari males and females who are retired

4



POPULATION (15 YEARS AND ABOVE) BY RELATION TO LABOUR FORCE,  
NATIONALITY & GENDER  
2019

Nationality	gender	المجموع العام Grand Total	غير النشيطين اقتصادياً Economically Inactive						النشيطون اقتصادياً Economically Active				النوع	الجنسية
			المجموع Total	أخرى Other	متقاعد Retired	عاجز Disabled	متفرغ للدراسة Student	متفرغة لأعمال المنزل Housewife	المجموع Total	متنطل سبق له المعل Unemployed with previous employment	متنطل لم يسبق له العمل Seeking Work for first time	مشغلت Employed		
Qatari	Males	100,687	32,357	1,140	10,820	1,189	19,208	0	68,330	60	60	68,210	ذكور	قطريون
	Females	104,824	65,382	2,192	6,220	2,085	26,090	28,795	39,442	0	165	39,277	إناث	
	Total	205,511	97,739	3,332	17,040	3,274	45,298	28,795	107,772	60	225	107,487	المجموع	
Non-Qatari	Males	1,800,199	44,621	2,981	0	1,059	40,581	0	1,755,578	45	781	1,754,752	ذكور	غير قطريين
	Females	386,886	140,336	3,225	0	1,758	31,429	103,924	246,550	176	912	245,462	إناث	
	Total	2,187,085	184,957	6,206	0	2,817	72,010	103,924	2,002,128	221	1,693	2,000,214	المجموع	
Total	Males	1,900,886	76,978	4,121	10,820	2,248	59,789	0	1,823,908	105	841	1,822,962	ذكور	المجموع
	Females	491,710	205,718	5,417	6,220	3,843	57,519	132,719	285,992	176	1,077	284,739	إناث	
Total	Total	2,392,596	282,696	9,538	17,040	6,091	117,308	132,719	2,109,900	281	1,918	2,107,701	المجموع	

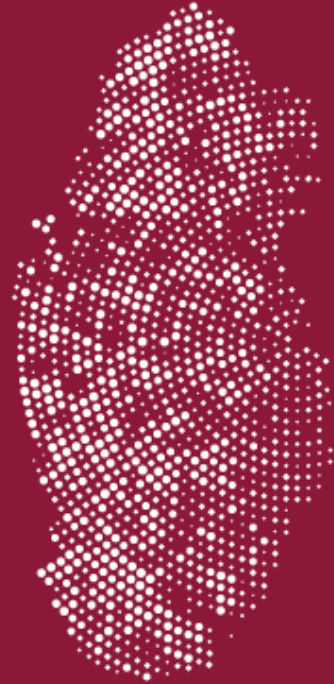
From the given table, the number of retired Qatari males in 2019 is 10,820 and the number of retired Qatari females in the same year is 6,220. This means the total number of retired Qatari is 17,040. The ratio of males to females is 63% to 37%.

# **In Summary –LLMs can play a key role**

---

- **Large Language Models (LLMs) can play a crucial role in the domain of data statistics**
- **Data cleaning and validation**
- **Data querying and exploration across different modalities**
- **Natural language interaction with the data**
- **Operating on data within files**

# Thank You



**منتدى الدوحة للبيانات**  
من أجل الابتكار في التنمية المستدامة

23-22 أكتوبر 2024

**DOHA DATA FORUM**  
FOR INNOVATION IN SUSTAINABLE DEVELOPMENT

October 22-23, 2024