FANAR

Arabic-Centric Large Language Models

www.fanar.qa

Dr. Mohamed Eltabakh
Qatar Computing Research Institute

Hamad Bin Khalifa University

Fanar- A National Project

Project Sponsor

وزارة الاتصــــــالات وتكنولوجيـــــا المعلومـــــات
Ministry of Communications and Information Technology
دولـــة قطــر • State of Qatar

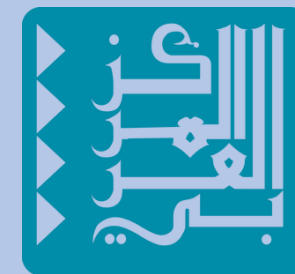**Project inception and development**
(30+ scientists and engineers)

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

Partners

Technology partners & infrastructure providers

Google Cloud

مكتبة قطر الوطنية
Qatar National Library

Cross-Partners
Project Management

Liver X
Innovative Services

جامعة قطر
QATAR UNIVERSITY

وزارة الأوقــــاف والشــــؤون الإسلاميـــــة
Ministry of Endowments and Islamic Affairs
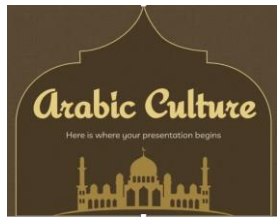دولـــة قطــر • State of Qatar

Motivation Behind Building Fanar

Content % on the Web

Preserve Arabic language and its dialects in the era of AI and LLMs

Cultural alignment & awareness relevant to 0.5B Arabs & 2.0B Muslims

Technology ownership & Digital sovereignty

Building in-house capacity

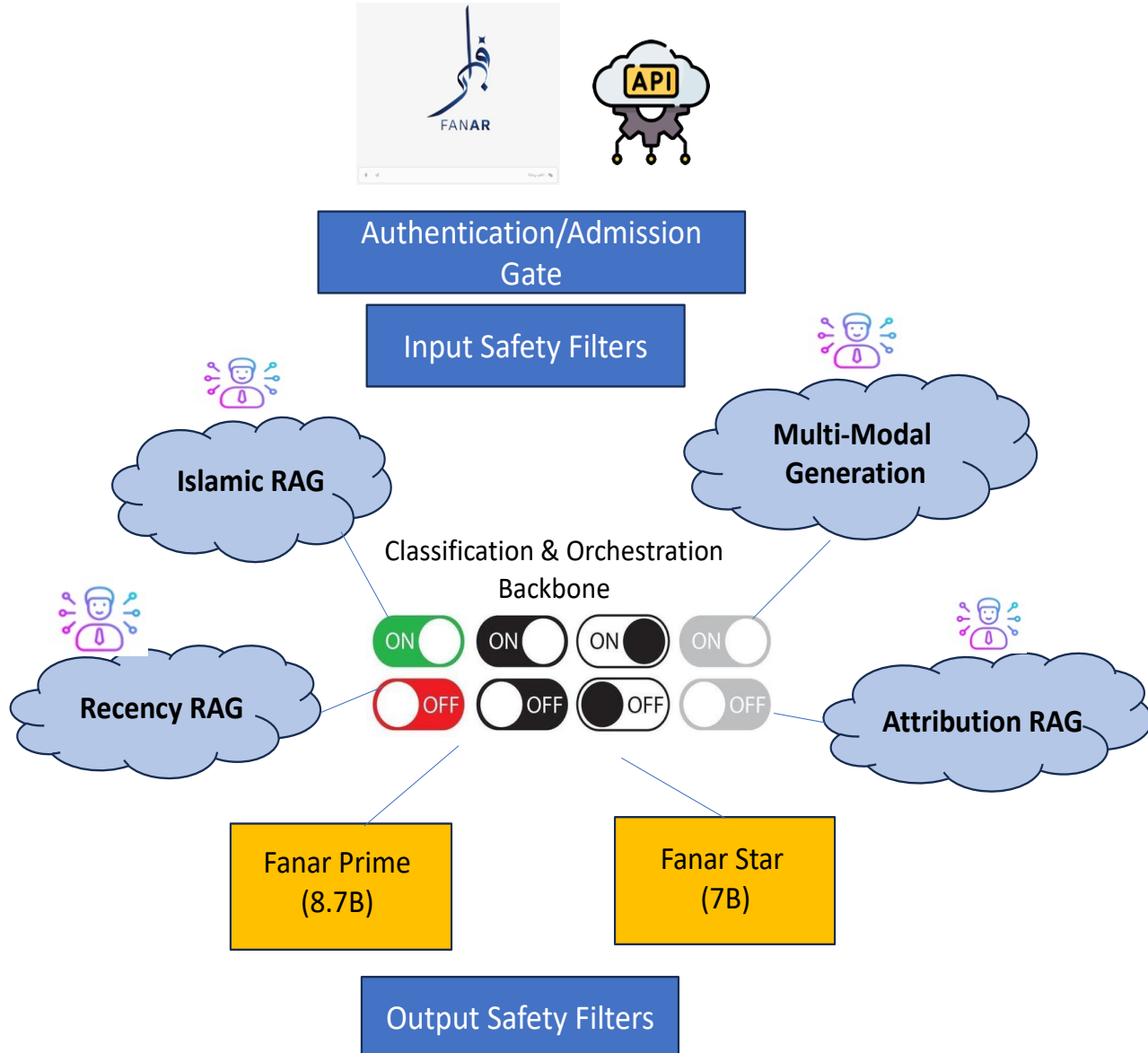| Rank	Language	16 May 2023	13 July 2024
1	English	55.5%	49.7%
2	Spanish	5.0%	5.9%
4	German	4.3%	5.4%
6	Japanese	3.7%	4.9%
5	French	4.4%	4.3%
3	Russian	4.9%	4.1%
7	Portuguese	2.4%	3.7%
9	Italian	1.9%	2.6%
11	Dutch	1.5%	2.1%
8	Turkish	2.3%	1.9%
12	Polish	1.4%	1.8%
10	Persian	1.8%	1.3%
13	Chinese	1.4%	1.2%
14	Vietnamese	1.3%	1.1%
15	Indonesian	0.7%	1.1%
16	Czech	0.7%	0.9%
17	Korean	0.7%	0.8%
18	Ukrainian	0.6%	0.6%
24	Hungarian	0.4%	0.6%
19	Arabic	0.7%	0.5%

Fanar Models

- **A family of highly capable Arabic-centric LLMs**
  - –Trained on the largest amount of Arabic content (Over 1.3T <u>cleaned</u> tokens) **[40% Arabic,   50% English,   10% code]**

- **Dual Models**
  - –**Fanar Star:** developed full-stack from scratch in-house **[7B parameters]**
  - –**Fanar Prime**: trained on Gemma 2 9B model **[8.7B parameters]**

- **Modern Standard Arabic (MSA) + Multi-dialectal support**

- **Highly tuned and custom components**

- **Multi-modality support (e.g., audio, images, etc.)**

Fanar Agentic Framework

- Connecting to Fanar through UI or APIs

- Different services connected through orchestration framework

- Services are activated based on prompt classification

- Fanar models are the brain of the system

Fanar Distinguishing Features (I)

اكتب قصة عن فتاة تدرس ومعلمتها تجلس بجانبها

كان هناك فتاة اسمها ليلى، طالبة مجتهدة تستيقظ كل صباح باكرا لتحضر دروسها بجد واج تبذل جهدا كبيرا في تعليم طلابها وتشجيعهم على تحقيق أهدافهم.

في أحد الأيام، وبينما كانت ليلى تجلس في مكانها المعتاد في الفصل، جلست معلمتها سارة بالحماس.

قالت سارة: "ليلى، كيف حالك اليوم؟"

أجابت ليلى: "الحمد لله، أنا بخير. كيف حالك أنت يا سارة؟"

ابتسمت سارة: "أنا دائما بخير، لكنني سعيدة لرؤية حماسك تجاه دراستك. ماذا عنك؟ هل و

Out-of-the-box model   *Fanar's Fine-tuned model*

كم مرة فازت هولندا بكأس العالم لكرة القدم

فازت هولندا بكأس العالم لكرة القدم مرة واحدة فقط، وذلك في عام 1988.

Fact-Checking Service

إسناد

بعد ____ قبل

فازت هولندا بكأس العالم لكرة القدم مرة واحدة فقط، وذلك في عام 1988.

ويعتبر منتخب هولندا لكرة القدم من أفضل منتخبات كرة القدم في العالم وسبق أن أحرز لقب بطولة أمم أوروبا عام 1988، كما أحرز المركز الثاني لبطولة كأس العالم لكرة القدم أعوام 1974 و1978و 2010 في مونديال جنوب أفريقيا حيث خسر النهائي أمام إسبانيا بعد أن قدم مباريات كبيرة وأخرج البرازيل من ربع النهائي وكان مرشحا لأحراز القب لولا سوء الطالع تحت قيادة المدرب فان مارفيك ونجوم من أمثال روبن وفان بيرسى وشنايدر وكاوت.

Who won Nobel Prize in Physics in 2024

John J. Hopfield and Geoffrey E. Hinton won the Nobel Prize in Physics in 2024.

[1] www.nytimes.com
[2] prl.aps.org
[3] www.eps.org
[4] www.chemistryworld.com

Recency Service

Fanar Distinguishing Features (II)

State-of-the-art in-house Speech-to-Text & Text-to-Speech

Multi-layers of precautions

Pre-training Phase

SFT/Alignment Phase

Islamic RAG

Post-Generation Verification

Grounded on top of 14 years of cutting edge research and expertise in QCRI

Fanar Benchmarking

		MMMLU(Ar)	ArabicMMLU		PIQA(AR)	OALL	ACVA	CulturalMCQ
		0-shot	0-shot	3-shot	0-shot	0-shot	5-shot	3-shot
Jais-family-6p7b	6.7B	25.34	29.12	34.81	65.18	48.20	60.13	34.00
OLMo-7B-0724-hf	7B	25.37	27.22	35.35	49.95	37.20	49.60	29.30
Fanar Star	7B	36.16	38.75	50.20	65.13	42.02	68.60	62.80
ALLaM-Base-2	7B	36.28*	44.45*	-	-	-	68.46*	-
Qwen2.5-7B	7B	51.77	60.15	65.08	59.68	48.66	80.37	65.70
ALLaM-Base-1	7B	34.42*	41.52*	-	-	-	66.18*	-
AceGPT-v2-8B	8B	41.71	45.96	58.55	63.17	43.58	78.36	67.50
Llama 3.1	8B	43.21	46.56	55.73	57.51	43.01	77.72	60.00
Fanar Prime	8.7B	**57.30**	**61.14**	**67.35**	**65.83**	**54.79**	**81.40**	**71.90**
Gemma-2-9b	9B	54.04	57.80	64.32	63.98	50.24	79.66	68.60
Jais-13B	13B	29.91	35.77	39.23	64.96	38.20	62.70	35.60
Jais-adapted-13b	13B	34.01	42.93	51.96	65.02	40.79	73.52	60.90

		Belebele(Ar)	Almieyar(Ar)	PIQA(Egy)	PIQA(Lev)	ArabicMMLU(Egy)	ArabicMMLU(Lev)
		3-shot	3-shot	0-shot	0-shot	0-shot	0-shot
Jais-family-6p7b	6.7B	34.54	32.17	60.23	58.38	28.50	29.46
OLMo-7B-0724-hf	7B	30.87	41.52	51.41	50.05	26.72	26.32
Fanar Star	7B	47.76	59.10	60.88	57.78	29.31	39.75
Qwen2.5-7B	7B	71.72	76.81	57.51	55.44	47.33	49.26
AceGPT-v2-8B	8B	60.61	66.83	61.48	56.75	54.53	53.91
Llama 3.1	8B	61.59	63.84	55.28	53.81	41.44	38.39
Fanar Prime	8.7B	**79.37**	**77.68**	61.97	57.78	**55.68**	**55.41**
Gemma-2-9b	9B	75.31	73.82	60.17	58.05	49.61	47.15
Jais-13B	13B	35.39	37.03	**62.73**	58.65	29.65	36.38
Jais-adapted-13b	13B	43.02	62.34	62.19	**59.25**	38.24	37.93

Thank You

www.fanar.qa