Fanar: Leading Arabic-Centric GenAI Platform
Fanar Hackathon @QCRI, 2025

Last Updated: 06/15/2025

A. Overview on Fanar

A.1 Fanar Overview and Architecture

Fanar is an innovative Arabic-centric Generative AI platform that bridges the Arabic language, regional culture, and Islamic values with the latest advancements in AI. Fanar is designed and developed entirely within the Qatar Computing Research Institute (QCRI) at Hamad Bin Khalifa University (HBKU), a member of Qatar Foundation for Education, Science, and Community Development (QF). It is sponsored by the Qatari government through the Ministry of Communications and Information Technology (MCIT). Fanar is the first regional Generative AI Platform created entirely by local talent ensuring the digital sovereignty of Qatar and Arab nations. Developed locally ensures full control over the data content and quality, model generation process, and safety guardrails, making it the ideal solution for sensitive, private, and sovereign applications.

Fanar platform encompasses several features including text-based Large Language Models (LLMs), speech and image generation systems, specialized Retrieval Augmented Generation (RAG) modules, and an attribution service to authenticate and correct facts in generated text (see Figure 1). At the heart of Fanar are *Fanar Star* and *Fanar Prime*, two highly capable Arabic LLMs that are best in the class on well-established benchmarks compared to similar sized models. Fanar Star is a 7B (7 billion) parameter model that was trained from scratch on nearly 1.2 trillion clean and deduplicated Arabic, English and Code tokens. Fanar Prime is an 9B parameter model continually trained on the Gemma2 9B base model on the same 1.2 trillion token set. The Fanar platform provides several other capabilities including a customized Islamic Retrieval Augmented Generation (RAG) system for adequately handling religious questions, and a Recency RAG for summarizing information about current or recent events that have occurred after the pre-training data cut-off date. The platform provides additional cognitive capabilities including in-house bilingual speech recognition that supports

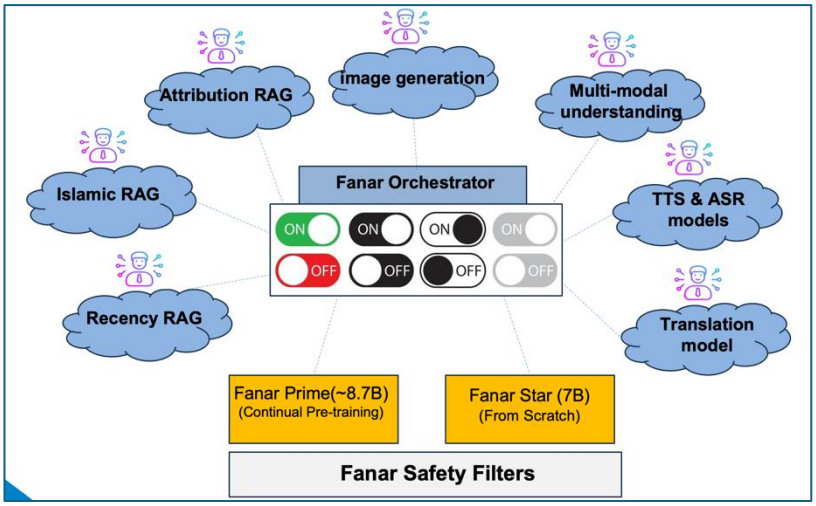

Figure 1: Fanar GenAI Platform.

multiple Arabic dialects, voice and image generation that is fine-tuned to better reflect regional characteristics.

Fanar Goals and Vision

- Revolutionizes Arabic-centric AI through culturally aligned solutions for over 400 million Arabic speakers.
- Establishes digital sovereignty and sets Qatar as a leader in culturally aligned AI research and development.
- Preserves and promotes Arabic culture by embedding cultural understanding and setting ethical standards for the development and use of AI models.
- Democratizes AI access through Fanar models and solutions, lowering barriers for AI adoption and innovation across sectors.
- Drives innovation and economic growth by enabling the development of new AI applications and industries tailored to the region.
- Addresses significant market needs by offering sophisticated Arabic AI technology currently lacking in the market.

A.2. Fanar Services

Fanar offers multiple services for applications. All services support bilingual interaction (in Arabic and English). Fanar is continuously under improvements and new features and capabilities are being added. The currently available services include:

- ***Fanar Text-to-Speech (TTS).*** Converting text to speech.
    - Input: Text (in Arabic or English), Output: Audio
- ***Fanar Automatic Speech Recognition (ASR).*** Converting speech to text
    - Input: Audio, Output: Text
- ***Fanar Translation (ShaheenLLM).*** Specialized engine for text translation
    - Input: Text (In Arabic or English), Output: Text (in English or Arabic)
- ***Fanar Image Generation.*** Generating images based on text description
    - Input: Text (in Arabic or English), Output: image
- ***Fanar Islamic RAG.*** Specialized module for Islamic content answering
    - Input: Text (In Arabic or English), Output: Text (in English or Arabic)
- ***Fanar Text Models.*** General-purpose LLMs supporting a broad range of tasks including, but not limited to: summarization, translation, sentiment analysis, entity and relationship extraction, code generation and analysis, reasoning tasks, linguistic tasks, creative writing, paraphrasing, and general knowledge Q & A.
- ***Fanar Image and video Understanding.*** Answering Q & A on a given image or video
    - Input: Image (or video) + Text (In Arabic or English), Output: Text (in English or Arabic)

A.3. Public Access to Fanar

Free-trial non-commercial public access to Fanar and its services is available via different channels. This public version of Fanar is deployed on a Google Cloud (GCP) infrastructure hosted in Europe (specifically Netherland) within the Qatari sovereign cloud managed by MCIT.

- Fanar Chat (https://chat.fanar.qa/)
- Fanar APIs (can be requested at https://fanar.qa/)
  - Designated API key: **fmFrMl3wHnB9SFnb8bzxNFpGCVE18Wcz**
- Fanar Text Model on Hugging Face
  - Fanar-1-9B (base) (https://huggingface.co/QCRI/Fanar-1-9B)
  - Fanar-1-9B-Instruct (https://huggingface.co/QCRI/Fanar-1-9B-Instruct)
- Fanar Mobile Apps
  - GooglePlay: https://play.google.com/store/apps/details?id=com.fanarmobile
  - AppleStore: https://apps.apple.com/in/app/fanar-%D9%81%D9%86%D8%A7%D8%B1/id6741857943

B. Fanar Hackathon @QCRI 2025

B.1. Hackathon Overview and Objectives

The Fanar Hackathon is a collaborative innovation event designed for QCRI interns to explore, experiment, and build practical solutions using the Fanar API platform. Spanning two weekends, the hackathon offers a focused environment for participants to work in teams, develop MVPs, and apply advanced AI techniques to real-world use cases. While the format is open-ended to encourage diverse ideas, participants are invited to consider themes such as **agentic AI**—systems that can act autonomously on behalf of users—and **multimodality**, which involves integrating and interpreting diverse data types such as text, images, and audio.

Objectives:

- To **encourage innovation** and experimentation using Fanar's multilingual, semantic, and context-aware APIs.
- To **develop MVPs** that demonstrate tangible, real-world applications of agentic and multimodal AI.
- To **gain feedback** from domain experts through evaluation and mentorship.
- To **highlight the versatility** and potential impact of the Fanar platform across different domains and user needs.

B.2. Hackathon Structure

The Fanar Hackathon will take place over two consecutive weekends—**June 20–21** and **June 27–28, 2025**—with teams comprising **2 to 3 members**. Participants will have access to the QCRI building each day from **10:00 AM to 12:00 midnight**, offering ample time for ideation, prototyping, and collaboration.

The **opening ceremony** will be held in the **Multipurpose Hall**, after which teams will have access to dedicated working spaces across different meeting rooms in the building. These rooms will be

available throughout the hackathon, and **mentors will be present on-site** each day **until 6:00 PM** to provide support and guidance across both technical and conceptual challenges.

To ensure a productive and comfortable environment, food will be provided daily to all participants (9:30 AM welcoming bites, 12:45 PM lunch, 3:00-7:00 PM light bites). The hackathon will conclude with a **presentation and awarding ceremony** on **Saturday, June 28**, from **5:00 PM to 9:00 PM**, also in the Multipurpose Hall. Each team will deliver a **5-minute pitch** (4 min presentation + demo + 1 min Q&A) to a panel of judges, showcasing their MVP and key learnings.

Prizes will be awarded as follows:

- **1st Place:** iPhones
- **2nd Place:** Apple Watches
- **3rd Place:** iPads

Registration link: https://forms.gle/NqVd6cFm7H2d1eSZ6

B.3. Hackathon Evaluation Criteria

Relevance and Impact (20%)	Does the project address a meaningful or original problem? Is it relevant to real-world use cases? Is there potential for positive societal, academic, or commercial impact?
Technical Innovation – Creativity and Originality (20%)	Does the project demonstrate creative use of technology? Are the agentic or multimodal aspects well-integrated (if applicable)? Does it go beyond a basic implementation?
Use of Fanar APIs (20%)	Are Fanar APIs integrated meaningfully into the solution? Does the team leverage Fanar diverse capabilities effectively (e.g., multilinguality, etc.)?
Functionality and Execution (20%)	Does the MVP work as intended? How complete is the prototype within the hackathon time constraints? Are there core features implemented and demonstrable?
Presentation and Communication (10%)	Was the problem, solution, and approach clearly communicated? Did the team present their work in a compelling and structured way?
Team Collaboration and Roles (5%)	Did the team demonstrate good collaboration and decision-making under time constraints? How well did the team work together to integrate their individual contributions into a cohesive solution?
User Experience and Design (5%)	Is the user interface intuitive and visually clear? Is the overall experience

	thoughtfully designed for end users?

C. Fanar Help

Participants are encouraged to join the dedicated **Slack channel**, where organizers and mentors will be available to respond to inquiries in real time. Whether it's a technical question, clarification about the schedule, or a request for resources, all support-related communication should be posted on Slack to ensure timely and organized responses. https://join.slack.com/t/qcri-fanarhackathon/shared_invite/zt-37o0lcs8x-crIm6Cbj2pmQjiOyRFac7Q