

# Dual LLM Configuration Walkthrough

I have implemented the ability to switch between **Ollama (Local)** and **Groq (Cloud)** for generating quiz questions, as requested.

## Setup & Usage

---

To start the project with your preferred AI provider, run the newly created startup script:

```
python start_project.py
```

### What happens next?

**Interactive Menu:** You (or your friend) will see a menu: `` Select AI Provider:

1. Ollama (Local - Free, Private)
2. Groq (Cloud - Fast, Requires API Key) ``

**Configuration:**

- If you select **Ollama**, the system configures itself to use your local Ollama instance (defaulting to `llama3` model).
- If you select **Groq**, it will check for a `GROQ_API_KEY`. If missing, it will ask you to enter it.

3. **Launch:** The Django server will start automatically with the chosen configuration.

## Manual Configuration (Advanced)

---

You can also manually control the provider by editing `backend/.env`:

```
# Options: 'ollama', 'groq', 'gemini', 'auto'  
LLM_PROVIDER=groq
```

## Verification

---

I have verified the implementation using automated tests to ensure: - Selecting `ollama` strictly tries to use the local Ollama API. - Selecting `groq` uses the Groq API. - The `auto` mode (default) still preserves the original fallback behavior (Ollama -> Groq -> Gemini).

Your friend can now easily choose their preferred way to run the AI!