

# Hướng dẫn sử dụng môi trường lập trình và nộp bài cho Kỳ thi AI Quốc gia

Ban Tổ Chức

## 1. Môi trường lập trình và thư viện hỗ trợ

Trong kỳ thi, thí sinh sẽ thực hiện lập trình trực tiếp trên nền tảng **Jupyter Notebook** do Ban Tổ Chức (BTC) cung cấp thông qua một liên kết web. Toàn bộ quá trình thực thi mã nguồn và sử dụng tài nguyên tính toán, đặc biệt là các tác vụ cần GPU, sẽ được thực hiện trực tiếp trên nền tảng notebook do BTC cung cấp. Máy tính được cấp cho thí sinh chỉ phục vụ mục đích soạn thảo mã và không được trang bị GPU, do đó không đảm bảo khả năng chạy các chương trình yêu cầu hiệu năng cao. Mỗi thí sinh sẽ được cấp một tài khoản riêng để đăng nhập trước giờ thi.

Để hỗ trợ quá trình viết mã và thử nghiệm cục bộ, **máy tính của thí sinh** được cài đặt sẵn các công cụ phát triển cơ bản, bao gồm:

- **Visual Studio Code** (phiên bản 1.89)
- **Vim** (phiên bản 9.1)
- **PyCharm Community Edition** (phiên bản 2024.1.2)

Ngôn ngữ lập trình sử dụng là **Python 3.10**, với các thư viện phổ biến đã được **cài đặt sẵn ở phiên bản ổn định mới nhất** tại thời điểm tổ chức kỳ thi. Thí sinh không được phép cài thêm thư viện mới hoặc tải mô hình từ Internet.

Danh sách các thư viện đã được cấu hình sẵn ở máy tính của thí sinh và trong môi trường Jupyter Notebook được trình bày trong Bảng 1 dưới đây.

**Lưu ý:**

- Thí sinh không được phép sử dụng **pip**, **conda** hay bất kỳ công cụ nào để cài thêm thư viện.
- Mọi mô hình và dữ liệu tải về từ Internet sẽ bị chặn trong môi trường thi.

Bảng 1: Danh sách thư viện đã được cài đặt sẵn trong môi trường thi

Nhóm thư viện	Tên thư viện
Học máy và xử lý NLP	torch, tensorflow, scikit-learn, xgboost, catboost, transformers, spacy, nltk, gensim, fasttext
Xử lý dữ liệu	pandas, numpy, scipy, csv, json, pickle
Xử lý ảnh	opencv-python, Pillow, torchvision, scikit-image
Trực quan hóa	matplotlib, seaborn, plotly, autoviz
Tiện ích	joblib, datasets, evaluate, os, sys, re, itertools, collections, time, pdb, pytorch-lightning, tensorboard, tqdm

## 1.2 Dữ liệu và mã nguồn mẫu

- Mỗi tác vụ đều đi kèm với **một tệp notebook mẫu**, trong đó đã bao gồm:
  - Mô tả tác vụ.
  - Mã nguồn của **mô hình mẫu tham khảo**.
- Bộ dữ liệu **tập huấn luyện (train)** và **tập kiểm tra công khai (public test)** đã được lưu sẵn trong môi trường notebook và có thể truy cập bất kỳ lúc nào.
- Tập dữ liệu **test riêng (private test)** sẽ chỉ được mở khóa trong **giờ cuối của cuộc thi**. Thí sinh cần mật khẩu (do BTC cung cấp) để truy cập vào tệp này. Việc truy cập trước giờ quy định hoặc sửa nội dung test đều bị coi là vi phạm quy chế.

## 2. Sử dụng mô hình ngôn ngữ lớn (LLM)

**Truy cập:** Thí sinh được cung cấp tài khoản truy cập vào LLM thông qua trình duyệt web trước khi bắt đầu thi. Mô hình sử dụng là **deepseek-r1-distill-qwen-32b**, hoạt động hoàn toàn offline.

**Thời gian sử dụng:** LLM chỉ được sử dụng trong **5 tiếng đầu** của kỳ thi (trong tổng thời gian thi 6 tiếng).

**Hạn chế:**

- Sau 5 tiếng đầu, quyền truy cập LLM sẽ tự động bị khóa.
- Thí sinh chỉ được dùng LLM để hỗ trợ tư duy và phát triển mô hình.

### 3. Nộp bài và đánh giá kết quả

#### Quy trình thi:

- **Giờ 0–5:** Làm bài, được phép dùng LLM.
- **Giờ thứ 6:** Không còn quyền truy cập LLM. BTC công bố password cho tập dữ liệu test (private). Thí sinh sử dụng mô hình của mình để tạo ra file kết quả.

#### Yêu cầu khi tạo file kết quả:

- Không được chỉnh sửa thủ công dữ liệu đầu vào/đầu ra.
- Quá trình chạy sẽ được ghi log toàn bộ để đảm bảo minh bạch.
- Mỗi bài, thí sinh cần nộp file kết quả là một file nén chấm **zip** bên trong chứa:
  1. Một file đầu ra: **.csv** hoặc **.pkl** (tùy theo yêu cầu đề thi).
  2. Một file mã nguồn dạng **.ipynb**. Lưu ý: File kết quả đầu ra phải đảm bảo được lấy từ việc chạy tuần tự từ file **.ipynb**. Các khối mã nguồn (code cell) được chạy lần lượt từ trên xuống dưới. Bài của thí sinh chỉ hợp lệ khi mã nguồn đã nộp sinh ra kết quả tương đồng với file đầu ra.

#### Cách nộp bài:

- Chỉ nộp một file kết quả duy nhất qua địa chỉ được BTC công bố cho mỗi bài.
- Mỗi bài thí sinh có tối đa 5 lần nộp.
- Việc nộp bài trễ sẽ không được chấp nhận.