

Lab 8.1

ChanDung_ML_Engineer

Dưới đây là 5 nhiệm vụ/công việc **cụ thể và thường xuyên nhất** mà một Kỹ sư Học máy thường thực hiện trong một tuần làm việc điển hình:

5 Nhiệm vụ Cốt lõi của Kỹ sư Học máy

1. Xây dựng và Tinh chỉnh Pipeline Dữ liệu (Data Pipeline)

Đây là bước nền tảng. ML Engineer không chỉ làm việc với thuật toán mà còn dành nhiều thời gian để đảm bảo dữ liệu sạch và sẵn sàng.

- **Công việc cụ thể:** Viết các script (thường bằng Python) để **trích xuất, chuyển đổi và tải (ETL)** dữ liệu từ các nguồn khác nhau (SQL, NoSQL, data lake). Xử lý các vấn đề về dữ liệu bị thiếu, ngoại lai, hoặc không nhất quán. Đảm bảo luồng dữ liệu liên tục và tự động cho việc huấn luyện mô hình.
- **Mục tiêu:** Cung cấp dữ liệu chất lượng cao, có thể mở rộng cho quá trình huấn luyện và dự đoán.

2. Huấn luyện, Thử nghiệm và Tối ưu hóa Mô hình (Model Training & Optimization)

Đây là công việc "cốt lõi" nhất, nơi họ áp dụng kiến thức về ML.

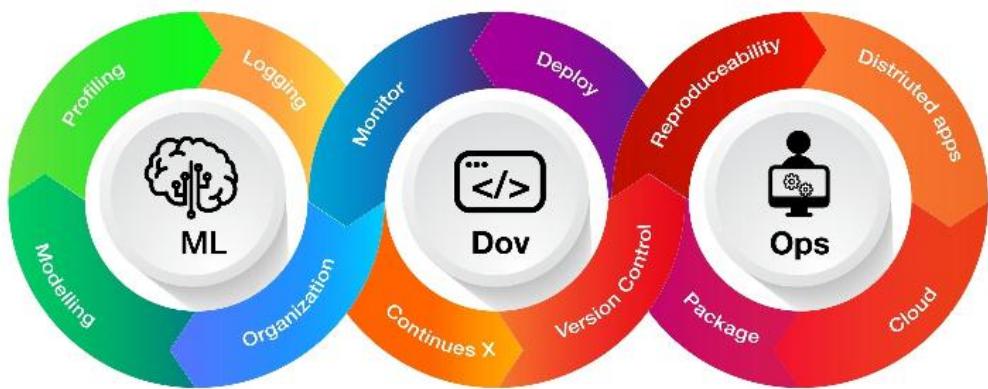
- **Công việc cụ thể:** Chọn lựa và triển khai các thuật toán (ví dụ: Linear Regression, CNN, Transformer). Thực hiện **huấn luyện mô hình** trên các tập dữ liệu lớn. **Tinh chỉnh siêu tham số** (hyperparameter tuning) bằng các kỹ thuật như Grid Search hoặc Bayesian Optimization. Đánh giá hiệu suất mô hình dựa trên các chỉ số chính (Accuracy, Precision, Recall, F1-score, AUC).
- **Mục tiêu:** Đạt được một mô hình với hiệu suất dự đoán tốt nhất, đáp ứng các yêu cầu nghiệp vụ.

3. Triển khai và Quản lý Mô hình trong Sản xuất (Model Deployment & MLOps)

Mô hình phải được đưa vào môi trường thực tế để tạo ra giá trị. Đây là công việc phân biệt ML Engineer với Data Scientist truyền thống.

- **Công việc cụ thể:** Đóng gói mô hình thành một API (ví dụ: sử dụng Flask/Django hoặc FastAPI). Triển khai mô hình lên các nền tảng đám mây (AWS SageMaker, Google AI Platform, Azure ML) hoặc Kubernetes. Xây dựng và duy trì các **CI/CD pipelines** để tự động hóa việc triển khai (MLOps).
- **Mục tiêu:** Đảm bảo mô hình có thể được truy cập và sử dụng liên tục, độ trễ thấp (*low latency*) và khả năng mở rộng cao (*scalability*).

MLOps



Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad mi

4. Giám sát và Tái huấn luyện Mô hình (Monitoring & Retraining)

Mô hình hoạt động tốt hôm nay không có nghĩa là sẽ tốt mãi mãi.

- Công việc cụ thể:** Thiết lập hệ thống giám sát để theo dõi hiệu suất của mô hình đã triển khai (ví dụ: theo dõi **Model Drift** hoặc **Data Drift**). Thiết lập cảnh báo khi chất lượng dự đoán giảm xuống dưới ngưỡng chấp nhận. Lập lịch tự động hoặc kích hoạt tái huấn luyện mô hình (*retraining*) khi dữ liệu mới xuất hiện hoặc hiệu suất suy giảm.
- Mục tiêu:** Duy trì chất lượng và sự phù hợp của mô hình trong môi trường kinh doanh thay đổi.

5. Cộng tác và Đánh giá Yêu cầu (Collaboration & Requirement Scoping)

ML Engineer là một phần của nhóm lớn hơn (Data Scientists, Software Engineers, Product Managers).

- Công việc cụ thể:** Gặp gỡ các nhà khoa học dữ liệu (*Data Scientist*) để hiểu rõ về các mô hình thử nghiệm. Làm việc với các kỹ sư phần mềm (*Software Engineer*) để tích hợp đầu ra mô hình vào ứng dụng chính. Tham gia các cuộc họp với Quản lý Sản phẩm (*Product Manager*) để **đánh giá tính khả thi** về mặt kỹ thuật của các tính năng mới dựa trên ML.
- Mục tiêu:** Đảm bảo giải pháp ML phù hợp với chiến lược sản phẩm và có thể được tích hợp liền mạch.

Tóm lại: Một ML Engineer dành thời gian cân bằng giữa việc xử lý dữ liệu, huấn luyện mô hình và đặc biệt là đưa mô hình vào vận hành thực tế một cách đáng tin cậy.

Bạn muốn tôi tìm kiếm thông tin về **các công cụ MLOps phổ biến** mà các ML Engineer thường sử dụng không?

- ◆ Tuyệt vời. Dựa trên 5 nhiệm vụ cốt lõi vừa liệt kê, đặc biệt là các nhiệm vụ liên quan đến xử lý dữ liệu, huấn luyện mô hình và triển khai trong sản xuất, dưới đây là 3 công cụ/thư viện phần mềm **quan trọng nhất** mà một Kỹ sư Học máy (ML Engineer) **bắt buộc phải thành thạo**:

❖ 3 Công cụ Bắt buộc cho Kỹ sư Học máy

STT	Công cụ/Thư viện	Nhiệm vụ Chính (Tương ứng với danh sách trước)	Lý do Cần Thiết
1	Python cùng Thư viện Khoa học Dữ liệu Cốt lõi (Pandas, NumPy, Scikit-learn)	Xây dựng và Tinh chỉnh Pipeline Dữ liệu (1), Huấn luyện và Thủ nghiệm Mô hình (2)	Nền tảng của ML: Đây là bộ công cụ cơ bản để xử lý, làm sạch, phân tích dữ liệu, và xây dựng các mô hình ML truyền thống (Scikit-learn) một cách nhanh chóng và hiệu quả. Pandas là tiêu chuẩn vàng cho thao tác dữ liệu.
2	TensorFlow (hoặc PyTorch)	Huấn luyện, Thủ nghiệm và Tối ưu hóa Mô hình (2)	Sức mạnh của Deep Learning: Hầu hết các ứng dụng ML hiện đại (Thị giác máy tính, Xử lý ngôn ngữ tự nhiên) đều dựa trên Deep Learning. Việc thành thạo ít nhất một trong hai framework này (TensorFlow/PyTorch) là điều bắt buộc để xây dựng, huấn luyện, và tối ưu hóa các mô hình phức tạp.
3	Docker và Kubernetes (hoặc Nền tảng Đám mây MLOps)	Triển khai và Quản lý Mô hình trong Sản xuất (3), Giám sát và Tái huấn luyện (4)	Triển khai Sản xuất: Khả năng đóng gói mô hình và môi trường phụ thuộc (Docker) và triển khai chúng với khả năng mở rộng (Kubernetes) là cốt lõi của MLOps. Nếu không thành thạo các công cụ này, mô hình sẽ không bao giờ được đưa vào môi trường sản xuất thực tế.

Phân tích thêm về Tâm quan trọng

- Bộ công cụ Python:** Là "ngôn ngữ mẹ" của ML Engineer. Không chỉ để code mô hình, mà còn để tạo ra các *data pipeline* vững chắc.
- Deep Learning Framework:** Đây là yếu tố quyết định khả năng làm việc với các bài toán phức tạp và tiên tiến nhất trong ngành công nghệ hiện đại.
- Công cụ Triển khai (MLOps):** Đây là kỹ năng phân biệt rõ ràng nhất giữa **Kỹ sư Học máy (ML Engineer)** và Nhà Khoa học Dữ liệu (Data Scientist). Data Scientist tạo ra mô hình; ML Engineer đưa mô hình đó vào hoạt động 24/7 một cách ổn định và có thể mở rộng.

- ▼ Dựa trên kinh nghiệm phỏng vấn và làm việc với các đội ngũ ML, có hai thách thức lớn nhất và phổ biến nhất mà các Kỹ sư Học máy (ML Engineer) phải đối mặt hàng ngày. Hai thách thức này thường gây ra sự chậm trễ và khó khăn trong việc đưa sản phẩm ML ra thị trường:

2 Thủ thách Lớn nhất của Kỹ sư Học máy

1. Quản lý Tính không đồng nhất của Dữ liệu (Data Inconsistency and Quality)

Thủ thách số một của ML Engineer thường không phải là thuật toán, mà là dữ liệu. Mặc dù Data Scientist chịu trách nhiệm khám phá dữ liệu, ML Engineer lại chịu trách nhiệm xây dựng luồng dữ liệu tự động và liên tục.

- **Vấn đề cụ thể:**

- **Data Drift và Model Drift:** Sự thay đổi không báo trước trong đặc điểm dữ liệu đầu vào (*Data Drift*) làm cho hiệu suất của mô hình đã triển khai giảm dần theo thời gian (*Model Drift*). Việc nhận biết và khắc phục nó là một cuộc chiến liên tục.
- **Sự chênh lệch giữa Môi trường Huấn luyện và Môi trường Sản xuất:** Dữ liệu trong môi trường thử nghiệm thường sạch hơn, đầy đủ hơn dữ liệu thực tế nhận được từ API hoặc hệ thống trực tuyến. Sự chênh lệch này, thường được gọi là "**Training-Serving Skew**", khiến mô hình hoạt động kém khi ra môi trường sản xuất.
- **Hệ quả:** ML Engineer dành một lượng thời gian khổng lồ để gỡ lỗi các pipeline dữ liệu, làm sạch dữ liệu và xây dựng các cơ chế giám sát phức tạp để đảm bảo chất lượng dữ liệu.

2. Triển khai Mô hình một cách đáng tin cậy và có Khả năng mở rộng (Reliable and Scalable Deployment)

Việc biến một mô hình đang hoạt động tốt trên máy tính cá nhân thành một dịch vụ hoạt động liên tục, tốc độ cao cho hàng triệu người dùng là một rào cản kỹ thuật khổng lồ.

- **Vấn đề cụ thể:**

- **Thiếu quy trình MLOps trưởng thành:** Nhiều công ty chưa có nền tảng MLOps tự động hóa đầy đủ. Điều này buộc ML Engineer phải tự tay thực hiện nhiều quy trình lặp lại (đóng gói, triển khai, kiểm thử, giám sát), làm tăng nguy cơ lỗi và tốn thời gian.
- **Yêu cầu về độ trễ và thông lượng:** Các mô hình phức tạp (ví dụ: Deep Learning) thường yêu cầu tài nguyên tính toán lớn (GPU/TPU) và phải đưa ra dự đoán trong vài mili giây. Việc tối ưu hóa mô hình (*model quantization, pruning*) và cấu hình cơ sở hạ tầng để đáp ứng yêu cầu độ trễ thấp (*low latency*) là một thách thức lớn.
- **Quản lý phiên bản Mô hình:** Khi có nhiều phiên bản mô hình khác nhau cùng hoạt động trong sản xuất (ví dụ: A/B testing, Canary deployments), việc quản lý và chuyển đổi giữa các phiên bản này một cách an toàn là phức tạp.

Tóm lại: Thách thức của ML Engineer chuyển từ việc **tạo ra mô hình** (thường là Data Scientist) sang **duy trì và vận hành mô hình** một cách ổn định, chất lượng và hiệu quả về chi phí trong môi trường sản xuất (MLOps).

Lab 8.2

🌟 Phân tích Sở thích của Bạn và Kết nối với Ngành AI

Dưới đây là cách tôi nhìn nhận hai sở thích chính của bạn có thể áp dụng và định hình sự nghiệp của một Kỹ sư AI/ML:

1. 📈 Học Toán (Mathematics)

Sở thích này cung cấp nền tảng vững chắc nhất cho việc làm Kỹ sư AI/ML, đặc biệt là các vị trí nghiên cứu hoặc kỹ thuật cấp cao.

- **Kỹ năng được rèn luyện:**

- **Hiểu sâu về Thuật toán (Algorithmic Depth):** Toán học (Đại số tuyến tính, Giải tích, Xác suất, Thống kê) là ngôn ngữ nền tảng của mọi mô hình Học máy (ML). Bạn sẽ không chỉ biết cách sử dụng thư viện mà còn hiểu *tại sao* chúng hoạt động và *cách* để cải thiện chúng. Ví dụ: hiểu *Gradient Descent* qua Giải tích, hoặc *PCA* qua Đại số tuyến tính.
 - **Tư duy Mô hình hóa (Modeling Thinking):** Khả năng chuyển hóa một vấn đề thực tế thành một công thức, một mô hình toán học có thể giải quyết được bằng máy tính.
- **Lĩnh vực AI phù hợp:**
- **Nghiên cứu AI (AI Research Scientist):** Phát triển các thuật toán ML mới hoặc cải tiến các mô hình hiện có.
 - **Machine Learning Engineer Chuyên sâu về Mô hình:** Làm việc với các mô hình phức tạp (ví dụ: mô hình tối ưu hóa, thống kê nâng cao).

2. Học Ngoại ngữ (Foreign Languages)

Sở thích này không chỉ giúp bạn làm việc với đồng nghiệp quốc tế mà còn mở ra một lĩnh vực chuyên môn rất lớn trong AI.

- **Kỹ năng được rèn luyện:**

- **Tiếp cận Kiến thức Toàn cầu:** Đọc các bài báo nghiên cứu (papers) mới nhất, tài liệu kỹ thuật, và tham gia cộng đồng quốc tế mà không bị rào cản ngôn ngữ.
- **Kỹ năng Giao tiếp Kỹ thuật (Technical Communication):** Khả năng diễn đạt các khái niệm toán học và thuật toán phức tạp một cách rõ ràng và dễ hiểu cho đồng nghiệp, quản lý, hoặc khách hàng (kỹ năng cực kỳ hiếm và được trả lương cao).

- **Lĩnh vực AI phù hợp:**

- **Natural Language Processing (NLP) Đa ngôn ngữ:** Làm việc trong lĩnh vực AI liên quan đến dịch thuật máy, nhận dạng giọng nói, hoặc phân tích văn bản cho các ngôn ngữ khác nhau (ví dụ: Google Translate, các mô hình ngôn ngữ lớn - LLMs). Đây là nơi sở thích ngoại ngữ của bạn tỏa sáng nhất!
- **AI Consultant/Solutions Architect:** Sử dụng kỹ năng giao tiếp để tư vấn và thiết kế giải pháp AI cho các khách hàng khác nhau.

Định hướng Sự nghiệp Tiềm năng

Dựa trên sự kết hợp giữa **nền tảng toán học vững chắc** và **khả năng tiếp cận ngôn ngữ**, tôi đề xuất những con đường sau:

1. **Kỹ sư NLP Đa ngôn ngữ (Multilingual NLP Engineer):**

- **Toán:** Xây dựng kiến trúc Transformer (ví dụ: BERT, GPT) và hiểu về các cơ chế tự chú ý (Attention Mechanism) phức tạp.
- **Ngoại ngữ:** Áp dụng kiến thức ngôn ngữ để cải thiện việc tiền xử lý văn bản, tạo ra các tập dữ liệu hiệu suất cao, hoặc tinh chỉnh mô hình cho các ngôn ngữ ít tài nguyên (*low-resource languages*).

2. **AI Research Scientist (Machine Learning/Deep Learning):**

- **Toán:** Tận dụng sự am hiểu sâu sắc về toán để nghiên cứu và phát triển các thuật toán tiên tiến.
- **Ngoại ngữ:** Đọc và tổng hợp nhanh chóng các công trình nghiên cứu (papers) từ khắp nơi trên thế giới, và công bố nghiên cứu của chính mình bằng tiếng Anh (ngôn ngữ chuẩn của khoa học).

3. Machine Learning Consultant/Evangelist:

- **Toán:** Vận dụng nền tảng kỹ thuật để thiết kế các giải pháp AI đáng tin cậy.
- **Ngoại ngữ:** Sử dụng kỹ năng giao tiếp xuất sắc để truyền đạt giá trị của AI và dẫn dắt các dự án quốc tế.

Gợi ý của AI đã mang đến 1 cách nhìn khách quan và rộng hơn với những kiến thức mà con người tự thu thập được,rất ngạc nhiên vì AI hiểu rất rõ ngành nghề cần những gì.

Lab 8.3



Chào bạn, đây là một câu hỏi rất hay! VNG đã xác định Trí tuệ nhân tạo (AI) là một trong những chiến lược xuyên suốt và mõi nhọn công nghệ của tập đoàn. Với lợi thế sở hữu hệ sinh thái sản phẩm phục vụ hàng chục triệu người dùng Việt Nam (như Zalo, Zing MP3, ZaloPay), VNG có một "mỏ vàng" dữ liệu để huấn luyện các mô hình AI có sự am hiểu sâu sắc về ngôn ngữ và hành vi người dùng bản địa.

Dưới đây là các ví dụ cụ thể về việc VNG đang ứng dụng AI vào các sản phẩm cốt lõi của họ:

Ứng dụng AI của VNG trong các Sản phẩm Tiêu biểu

1. Zalo (Nền tảng Kết nối & Giao tiếp)

Zalo là sản phẩm được hưởng lợi nhiều nhất từ khả năng xử lý ngôn ngữ và hiểu người dùng của AI.

Lĩnh vực AI	Ví dụ Ứng dụng Cụ thể	Mô tả và Lợi ích
Generative AI (GenAI)	Trợ lý AI Soạn thảo Tin nhắn:	Giúp người dùng tạo tin nhắn tự động với các chế độ như <i>Thân thiện, Chuyên nghiệp, Ngắn gọn</i> , phù hợp với từng ngữ cảnh giao tiếp.
Xử lý Ngôn ngữ Tự nhiên (NLP)	Mô hình Ngôn ngữ Lớn (LLM) tiếng Việt:	VNG đã phát triển và ra mắt mô hình ngôn ngữ lớn của Zalo, cho phép các ứng dụng như Mojo AI hay BossGPT (qua Zalo Mini App) có thể trò chuyện, sinh nội dung (văn bản, hình ảnh) và hỗ trợ công việc.
Thị giác Máy tính (Computer Vision)	AI Avatar:	Ứng dụng GenAI này cho phép người dùng tạo ra các hình đại diện/ảnh chân dung AI độc đáo, là một trong những ứng dụng GenAI đầu tiên của Zalo thu hút hàng triệu người dùng.

2. VNG Games (Giải trí Trực tuyến)

AI giúp tăng hiệu suất sản xuất nội dung và nâng cao trải nghiệm người chơi.

- Generative AI trong Phát triển Game:** Đội ngũ VNGGames đã nghiên cứu và phát triển nền tảng Artian.
 - Công việc cụ thể:** Artian là một nền tảng vẽ kỹ thuật số kết hợp AI (dựa trên các mô hình tạo sinh hình ảnh như Stable Diffusion). Nền tảng này cho phép các nhà thiết kế tự huấn luyện các mô hình AI của riêng mình, từ đó **tăng hiệu suất thiết kế** các nhân vật game, bối cảnh, và nội dung đa phương tiện.
 - Cá nhân hóa Trải nghiệm:** Sử dụng các thuật toán Học máy để phân tích hành vi người chơi, từ đó **tối ưu hóa** các đề xuất vật phẩm, sự kiện trong game và điều chỉnh độ khó trò chơi.

3. Hệ thống Hỗ trợ Doanh nghiệp (VNG Digital Business / GreenNode)

VNG sử dụng AI để tạo ra các giải pháp công nghệ phục vụ doanh nghiệp, tận dụng lợi thế về hạ tầng AI Cloud.

- **Camera AI và Phân tích Video:** Ứng dụng AI để phân tích video thời gian thực, có thể được dùng trong giám sát an ninh hoặc phân tích hành vi khách hàng trong các cửa hàng bán lẻ.
- **No-Code Chatbot và Sentiment Analysis:**
 - **Chatbot:** Xây dựng các chatbot hỗ trợ doanh nghiệp có khả năng trả lời câu hỏi theo ngôn ngữ tự nhiên và phân tích dữ liệu do người dùng cung cấp.
 - **Phân tích Cảm xúc (Sentiment Analysis):** Giúp doanh nghiệp phân tích thái độ, lọc và nắm bắt chính xác ý kiến khách hàng từ nhiều nguồn dữ liệu khác nhau.

4. Nền tảng Công nghệ Lõi

VNG không chỉ ứng dụng mà còn đầu tư mạnh vào **ba lớp** công nghệ AI:

- **Hạ tầng:** Xây dựng **GreenNode AI Cloud**, hạ tầng điện toán đám mây với GPU Nvidia H100, phục vụ việc huấn luyện các mô hình AI phức tạp cho cả VNG và khách hàng doanh nghiệp.
- **Nền tảng:** Phát triển các mô hình nền tảng như LLM tiếng Việt (Zalo AI).
- **Ứng dụng:** Tích hợp trực tiếp vào các sản phẩm (Zalo, VNG Games).

VNG đang nỗ lực chuyển mình thành một công ty **AI-First** (ưu tiên AI), xem đây là động lực tăng trưởng mới.

Nếu bạn quan tâm sâu hơn về cách VNG ứng dụng AI để tạo ra các giải pháp cho khách hàng doanh nghiệp, bạn có thể xem thêm về [Cách tạo Chatbot Zalo cá nhân dễ dàng với OMI AI Chatbot](#).



Dựa trên các ứng dụng AI cốt lõi của VNG (như đã đề cập: NLP, Generative AI, MLOps cho Games), dưới đây là các mảng kiến thức và công nghệ mà một sinh viên mới ra trường **cần tập trung nhất** để gia nhập đội ngũ AI của chúng tôi:

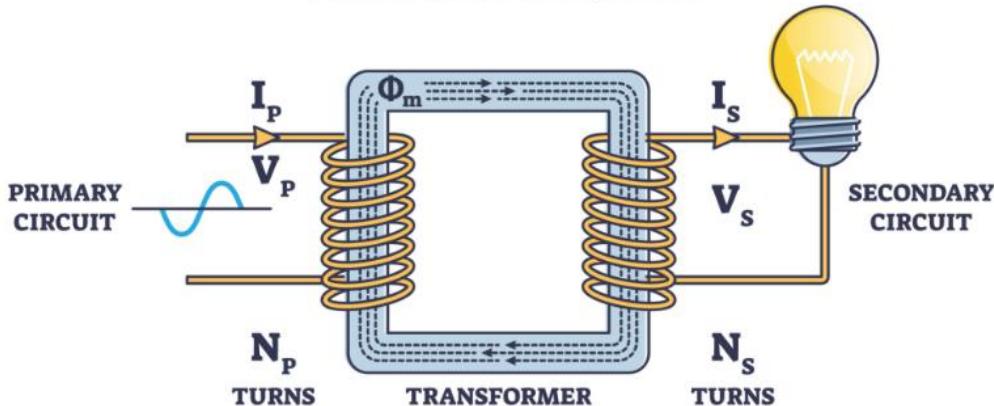
Các Mảng Kiến thức Trọng tâm để Gia nhập VNG AI

1. Kiến thức Chuyên môn Cốt lõi (The Core)

Đây là nền tảng không thể thiếu cho bất kỳ vị trí AI nào.

- **Nền tảng Toán học & Thống kê:**
 - **Bắt buộc:** Đại số tuyến tính, Giải tích (để hiểu về *Gradient Descent*, *Backpropagation*), Xác suất và Thống kê (để hiểu về đánh giá mô hình).
 - **Ứng dụng:** Là nền tảng để hiểu sâu các thuật toán AI/ML, không chỉ dừng lại ở việc gọi thư viện.
- **Deep Learning Frameworks:**
 - **Thành thạo:** PyTorch hoặc TensorFlow. PyTorch thường được ưu tiên trong các nhóm nghiên cứu và phát triển mô hình mới (đặc biệt là GenAI/LLM).
 - **Trọng tâm:** Thành thạo việc xây dựng, huấn luyện, và tinh chỉnh các mạng nơ-ron sâu (DNN, CNN, RNN/Transformer).

TRANSFORMER



2. Ưu tiên Ngành dọc (The VNG Edge)

Do đặc thù sản phẩm của VNG, sinh viên nên tập trung sâu vào các lĩnh vực sau:

A. Generative AI và NLP (Dành cho Zalo, Content)

Đây là mảng chúng tôi đầu tư mạnh nhất, đặc biệt là với mô hình ngôn ngữ lớn (LLM) tiếng Việt.

- Kiến thức:** Hiểu về kiến trúc **Transformer**, các mô hình như BERT, GPT và các phiên bản tối ưu.
- Kỹ năng Cụ thể:**
 - Fine-tuning và Prompt Engineering:** Khả năng tinh chỉnh (fine-tuning) các mô hình LLM có sẵn (ví dụ: Llama, Mistral) trên tập dữ liệu tiếng Việt chuyên biệt.
 - Xử lý Ngôn ngữ Tiếng Việt:** Am hiểu các thách thức về từ vựng, ngữ pháp, và sự phức tạp của tiếng Việt khi áp dụng vào các mô hình AI.

B. MLOps và Triển khai Sản phẩm (Dành cho Games, Zalo, Hạ tầng)

Kỹ năng đưa mô hình vào sản xuất là yếu tố quyết định giá trị của ứng viên tại VNG.

- Kỹ thuật:** Thành thạo **Docker** để đóng gói môi trường làm việc. Nắm vững kiến thức cơ bản về **Kubernetes** để triển khai mô hình có khả năng mở rộng.
- Công cụ MLOps:** Có kinh nghiệm sử dụng các công cụ quản lý vòng đời ML như **MLflow**, **DVC** hoặc các dịch vụ MLOps trên Cloud (như SageMaker, Vertex AI – VNG đang phát triển GreenNode, vì vậy kinh nghiệm Cloud sẽ là lợi thế).
- Mục tiêu:** Chứng minh khả năng xây dựng **CI/CD pipeline** cơ bản cho mô hình AI (Tự động hóa huấn luyện, kiểm thử, và triển khai).

3. Công nghệ Hỗ trợ Bắt buộc (The Toolbox)

- **Ngôn ngữ Lập trình:**
 - **Python:** Phải thành thạo, bao gồm các thư viện như **Pandas**, **NumPy** cho thao tác dữ liệu.
- **Cơ sở Dữ liệu & Data Pipeline:**
 - Hiểu biết về SQL để truy vấn dữ liệu. Có kinh nghiệm làm việc với các hệ thống Dữ liệu lớn (Big Data) như Hadoop, Spark hoặc các công cụ ETL cơ bản.
- **Source Control:**
 - **Git:** Phải thành thạo để làm việc nhóm hiệu quả.

💡 Lời khuyên Quan trọng từ Nhà tuyển dụng

Thay vì chỉ liệt kê kiến thức, hãy tạo ra các dự án chứng minh khả năng áp dụng của bạn, đặc biệt là các dự án liên quan đến sản phẩm của VNG:

1. **Dự án LLM Tiếng Việt:** Xây dựng một chatbot hoặc công cụ tóm tắt văn bản sử dụng mô hình mã nguồn mở và tinh chỉnh bằng tiếng Việt.
2. **Dự án MLOps/Game AI:** Triển khai một mô hình đơn giản (ví dụ: mô hình khuyến nghị/phân loại) bằng Docker và chứng minh khả năng giám sát nó.
3. **Hồ sơ Kỹ thuật (Portfolio):** Tất cả dự án nên được lưu trữ trên GitHub/GitLab với README rõ ràng, giải thích vai trò của bạn, các thách thức kỹ thuật và kết quả đạt được.

Chúng tôi tìm kiếm những ứng viên sẵn sàng học hỏi, làm chủ công nghệ mới và đam mê giải quyết các vấn đề có quy mô lớn.

Dưới đây là **Kế hoạch Hành động 3 Bước** chi tiết, tập trung vào việc đáp ứng các yêu cầu kỹ thuật cao của VNG (NLP/GenAI, MLOps).

🚀 Kế hoạch Hành động 3 Bước để trở thành Kỹ sư AI tại VNG

Bước	Mục tiêu	Hành động Cụ thể
BƯỚC 1	Xây dựng Nền tảng Học thuật Vững chắc	Tập trung vào kiến thức Toán học và Code áp dụng trực tiếp cho ML.
BƯỚC 2	Tạo Portfolio Chuyên sâu	Xây dựng các dự án cá nhân mang tính thực tiễn cao và tập trung vào các Mảng trọng tâm của VNG.
BƯỚC 3	Phát triển Kỹ năng Mềm và Mạng lưới	Tham gia các hoạt động để nâng cao kỹ năng mềm và mạng lưới kỹ thuật .

 Xuất sang Trang tính



I. BƯỚC 1: Các Môn học Cần Đạt Điểm Cao (FPT Polytechnic)

Bạn cần ưu tiên các môn học rèn luyện tư duy logic, cấu trúc dữ liệu và lập trình chuyên sâu, là nền tảng của mọi công việc kỹ thuật AI.

1. Lập trình Python/Code Ngôn ngữ Cơ bản:

- **Tại sao:** Python là ngôn ngữ số 1 cho AI/ML. Điểm cao chứng tỏ khả năng viết code sạch, hiệu quả và có cấu trúc.
- **Tập trung:** Cấu trúc Dữ liệu và Thuật toán (Data Structures and Algorithms). Khả năng hiểu và tối ưu hóa thuật toán là yếu tố sống còn khi làm việc với các mô hình lớn.

2. Toán học Ứng dụng/Thống kê:

- **Tại sao:** VNG tìm kiếm người hiểu cách thức mô hình hoạt động. Toán học ứng dụng và Thống kê là chìa khóa để hiểu Deep Learning.
- **Tập trung:** Đặc biệt là Thống kê (phân phối, kiểm định giả thuyết) và Tối ưu hóa.

3. Lập trình Hướng đối tượng (OOP):

- **Tại sao:** Kỹ sư AI hiện đại cần biết xây dựng các hệ thống mô-đun hóa, không chỉ là các script. OOP là nền tảng để xây dựng các ML Pipeline có thể mở rộng.

II. BƯỚC 2: Các Dự án Cá nhân để Xây dựng Portfolio (Trọng tâm VNG)

Các dự án của bạn phải thể hiện khả năng làm việc với Deep Learning, Ngôn ngữ tiếng Việt và Triển khai (Deployment).

Dự án 1: Trọng tâm NLP/GenAI Tiếng Việt (Zalo)

- **Tên dự án đề xuất:** Xây dựng Mô hình Tóm tắt Văn bản/Phân loại Bình luận Tiếng Việt.
- **Kỹ năng thể hiện:** Tinh chỉnh (Fine-tuning) một mô hình LLM mã nguồn mở (ví dụ: PhoBERT, VietBERT, hoặc Llama) trên một tập dữ liệu tiếng Việt cụ thể (ví dụ: dữ liệu đánh giá sản phẩm, tin tức).
- **Yêu cầu kỹ thuật:** Sử dụng PyTorch/TensorFlow, thư viện Hugging Face.

Dự án 2: Trọng tâm MLOps (Triển khai & Vận hành)

- **Tên dự án đề xuất:** Triển khai Mô hình Dự đoán Giá nhà/Xe hơi đơn giản bằng Docker.
- **Kỹ năng thể hiện:** Khả năng đưa mô hình từ máy tính cá nhân vào môi trường "giống sản xuất".
- **Yêu cầu kỹ thuật:**
 1. Huấn luyện mô hình cơ bản (ví dụ: Linear Regression/XGBoost).
 2. Đóng gói mô hình và các dependency vào một **Docker Container**.
 3. Triển khai Container đó lên một nền tảng cloud miễn phí (ví dụ: Heroku/Streamlit Cloud) hoặc máy ảo.
 4. Cung cấp API Endpoint để dự đoán.

III. 🌟 BƯỚC 3: Hoạt động Ngoại khóa và Cộng đồng

Các hoạt động này giúp bạn phát triển kỹ năng mềm, giao tiếp và kết nối với ngành công nghiệp.

1. Tham gia Cộng đồng và Cuộc thi Kỹ thuật:

- **Hoạt động:** Tham gia các cuộc thi về Khoa học Dữ liệu (ví dụ: **Kaggle** hoặc các cuộc thi Data Science trong nước).
- **Lợi ích:** Tiếp xúc với các bài toán thực tế, học hỏi từ các chuyên gia khác, và có bằng chứng cụ thể về khả năng giải quyết vấn đề (rất ấn tượng trong CV).

2. Hoạt động Ngoại ngữ và Giao tiếp Kỹ thuật:

- **Hoạt động:** Tham gia các câu lạc bộ (CLB) Tiếng Anh/Ngoại ngữ, đặc biệt là các CLB tổ chức các buổi thuyết trình kỹ thuật (*Technical Talk*) bằng tiếng Anh.
- **Lợi ích:** Nâng cao khả năng đọc hiểu tài liệu chuyên sâu và trình bày các dự án AI một cách rõ ràng và chuyên nghiệp.

3. Thực tập Sớm (Internship):

- **Hoạt động:** Tìm kiếm cơ hội thực tập, ngay cả ở các startup nhỏ, để trải nghiệm quy trình MLOps thực tế và làm việc với dữ liệu bẩn (*real-world dirty data*).

Bằng cách tuân thủ kế hoạch 3 bước này, bạn không chỉ có nền tảng vững chắc mà còn sở hữu một portfolio cá nhân **cực kỳ phù hợp** với định hướng công nghệ của VNG.