MODULE 1: GIỚI THIỆU VỀ EDA VÀ AI\ML

Ông Miguel giới thiệu khóa học này là học cách truy xuất dữ liệu, làm sạch, phân tích thăm dò (EDA), kỹ thuật đặc trưng, chuẩn bị dữ liệu để huấn luyện ML.

Course này yêu cầu Python, Jupyter Notebook, kiến thức cơ bản về toán (đại số tuyến tính, xác suất, thống kê).

Có 3 khhais niệm cơ bản của AI, ML,DL mà phải nhớ:

- AI (Artificial Intelligence): mọi hệ thống có thể cảm nhận, suy luận, hành động, thích ứng. Có thể dựa trên quy tắc, không nhất thiết phải “học” từ dữ liệu.

- ML (Machine Learning): **tập con của AI**. Máy có thể học khi tiếp xúc với nhiều dữ liệu hơn.

- DL (Deep Learning): **tập con của ML**, sử dụng mạng nơ-ron nhiều lớp, cải thiện đáng kể khả năng mô phỏng hành vi phức tạp.

Deep Learning đã khắc phục nhiều hạn chế của ML cổ điển tạo ra kỉ nguyên mưới

Điều chú ý : EDA là nền tảng trước khi làm ML, dữ liệu sạch quyết định chất lượng mô hình.

ML and DL   
ML thay vì con người viết sẵn quy tắc, máy tự học mẫu từ dữ liệu. Nhiều dữ liệu hơn mô hình học tốt hơn, nhưng sau một mức sẽ bị “diminishing returns” (lợi ích giảm dần). điển hình là phân loại email spam/không spam.   
**Features**: đặc trưng/dữ liệu đầu vào

**Target/Label**: cột muốn dự đoán

Supervised giám sát có nhãn (spam/not spam, loài hoa, giao dịch gian lận). Mục tiêu là dự đoán đúng nhãn.

Unsupervised ko giám sát không có nhãn (phân cụm khách hàng). Mục tiêu là khám phá cấu trúc dữ liệu, không có “đúng – sai” rõ ràng.  
DL dùng mạng nơ-ron sâu. Mạnh hơn ở chỗ tự động trích xuất đặc trưng thay vì phải “hand-craft” features. phù hợp dữ liệu lớn, bài toán phức tạp (nhận diện ảnh, giọng nói). Ko phù hợp nếu dữ liệu nhỏ, thay đổi theo thời gian .

So sánh để thấy rõ :

ML cổ điển con người phải chọn features thủ công còn DL mô hình tự học cách biểu diễn dữ liệu, từ pixel đến đặc trưng phức tạp.

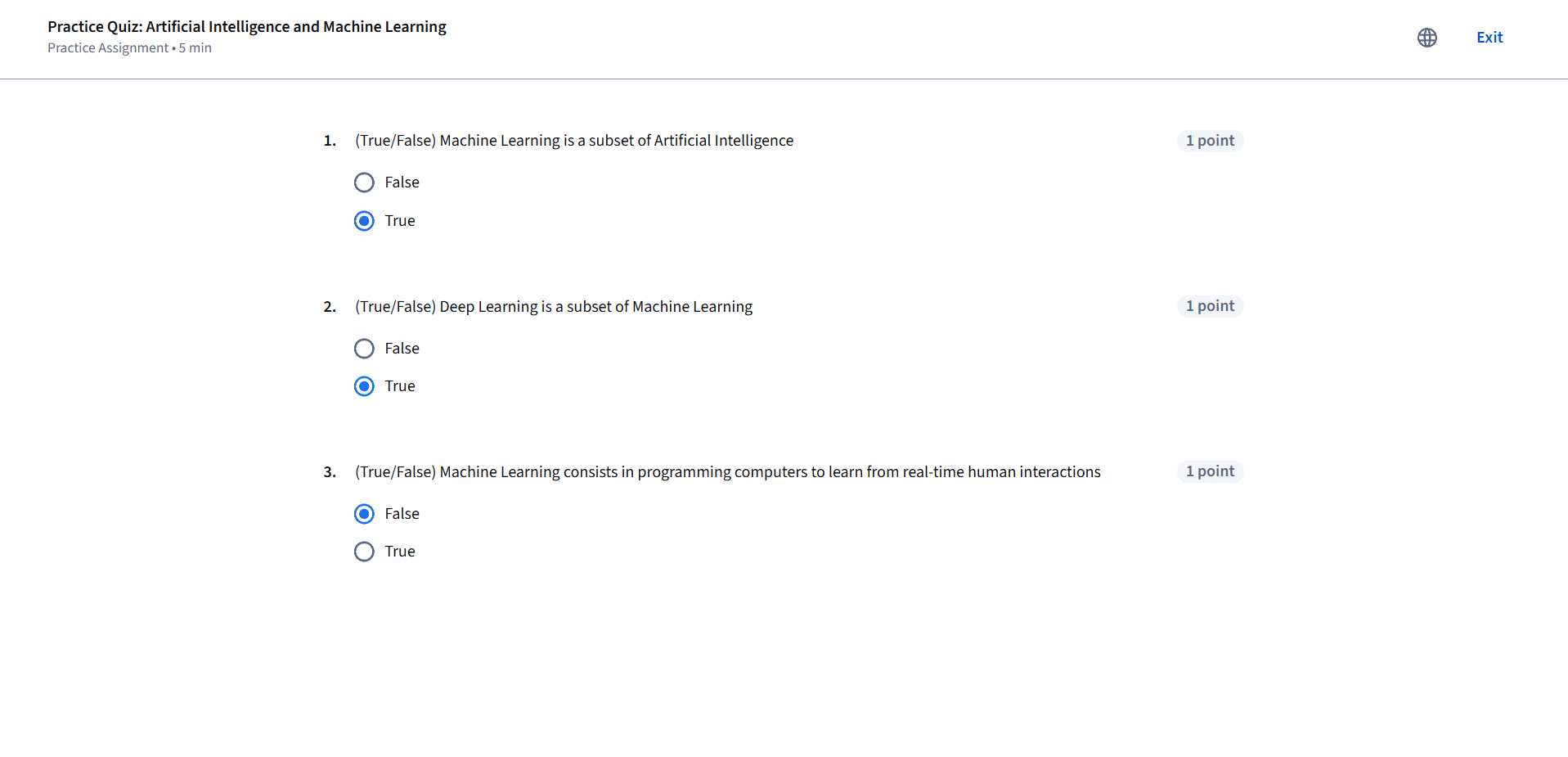
Lịch sử của AI : chủ yếu từ 1990s đến nay, AI đi từ “chỉ mới hữu ích” sang “bùng nổ thực sự” nhờ ba yếu tố:

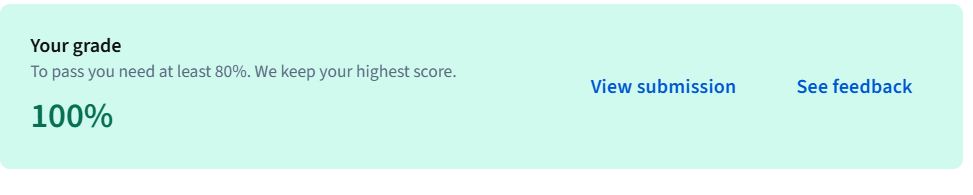
Thuật toán cải tiến (khắc phục gradient vanish/explode).

Dữ liệu lớn (ImageNet).

Sức mạnh tính toán (GPU, thư viện như TensorFlow/Keras).

Quiz 1 : Practice Quiz: Artificial Intelligence and Machine Learning





AI ở hiện tại bùng nổ mạnh nhất ở CPV ( thị giác ) và ngôn ngữ tự nhiên

AI đã lan tỏa khắp mọi ngành

AI trong cuộc sống có ở mọi nơi , như đi lại, MXH , siri alexa, CPV như quét khuôn mặt các thứ ,

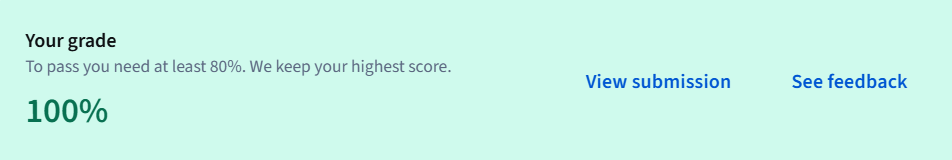
Làm quen với workflow ML

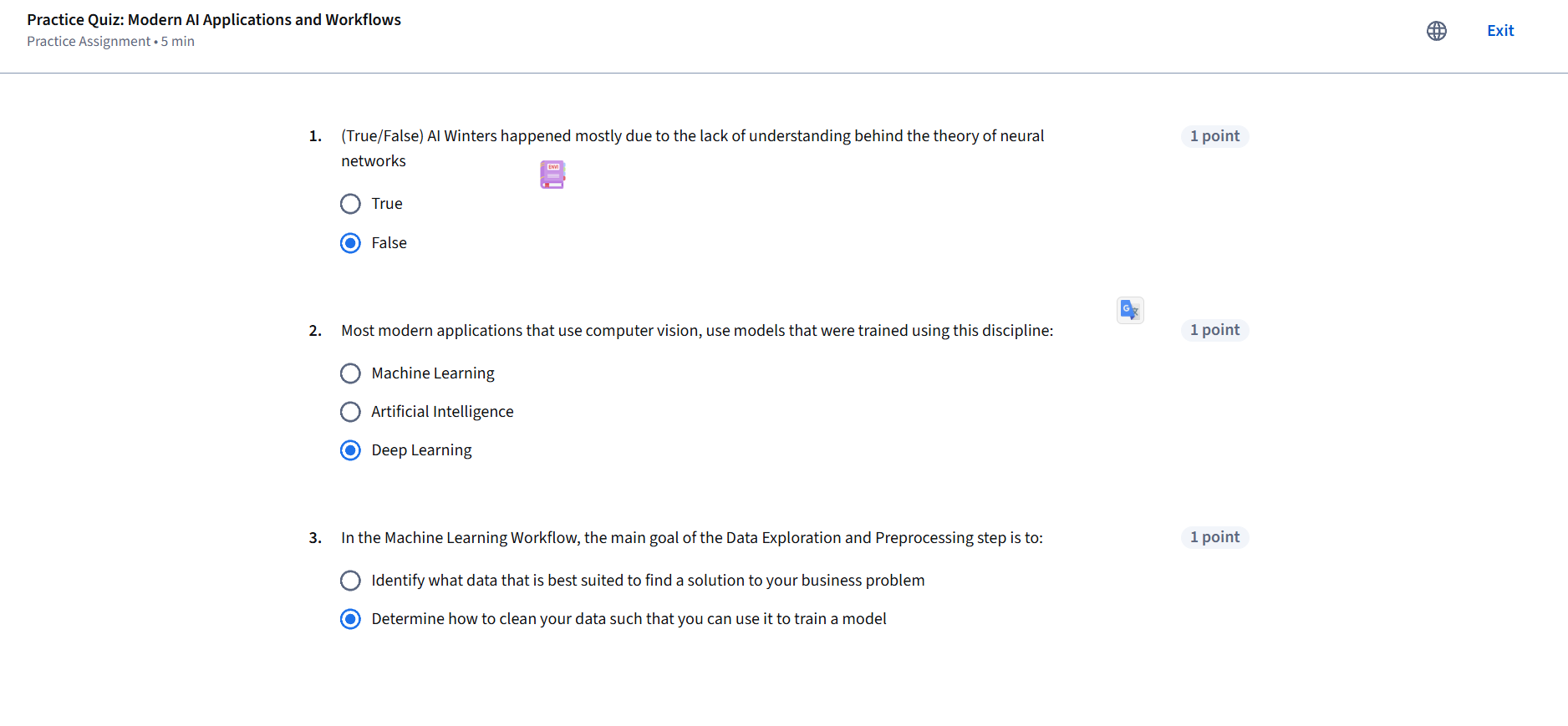
Quy trình làm việc ML điển hình

1. **Xác định vấn đề:** mình muốn giải quyết gì? Ví dụ: phân loại ảnh chó/mèo.
2. **Thu thập dữ liệu:** cần dataset đủ phong phú, nhiều góc chụp, ánh sáng, nhãn rõ ràng.
3. **Khám phá + tiền xử lý dữ liệu:** nhìn vào phân bố dữ liệu, heatmap, xử lý dữ liệu bẩn, chuyển pixel thành mảng số (cho deep learning).
4. **Xây dựng mô hình:** thử mô hình cơ bản trước, rồi tinh chỉnh.
5. **Xác thực/kiểm tra:** dùng test set chưa huấn luyện để đánh giá model.
6. **Ra quyết định & triển khai:** nếu đạt accuracy mong muốn thì đem vào production, bàn với stakeholder.

Target variable là biến mục tiêu là cái mình muốn dự đoán ở Iris là species

Quiz 2 : Practice Quiz: Modern AI Applications and Workflows





Graded Quiz: Module 1 - Modern AI and its Applications

