

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

1. Mục đích

“Hướng dẫn lựa chọn kỹ thuật AI” cung cấp một khuôn khổ tư duy có hệ thống, giúp các tổ chức và doanh nghiệp nhận diện đúng bản chất bài toán AI, đồng thời lựa chọn chính xác các kỹ thuật phù hợp với mục tiêu và điều kiện triển khai. Việc chuẩn hóa cách phân loại bài toán và kỹ thuật AI không chỉ giúp tối ưu nguồn lực đầu tư mà còn đảm bảo tính nhất quán, minh bạch và hiệu quả trong toàn bộ vòng đời phát triển và vận hành giải pháp AI.

2. Đối tượng áp dụng

Hướng dẫn áp dụng với các dự án phát triển sản phẩm ứng dụng học máy, trí tuệ nhân tạo tại các đơn vị trong Tập đoàn.

3. Định nghĩa, thuật ngữ và viết tắt

STT	Thuật ngữ & viết tắt	Giải thích
1	CNTT	Công nghệ thông tin
2	AI	Artificial Intelligence: Lĩnh vực khoa học máy tính nghiên cứu và phát triển các hệ thống có khả năng thực hiện những tác vụ đòi hỏi trí tuệ tương đương con người.
3	GPU	Graphics Processing Unit: Vi xử lý chuyên dụng tối ưu cho tính toán song song cường độ cao,

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Thuật ngữ & viết tắt	Giải thích
		thường được sử dụng trong huấn luyện mô hình DL.
4	TPU	Tensor Processing Unit: Vi xử lý do Google phát triển, tối ưu cho các phép toán tensor thường dùng trong huấn luyện và suy luận DL.
5	GAN	Generative Adversarial Network: Mô hình Generative AI gồm hai mạng (generator, discriminator) thi đua nhau để sinh dữ liệu có phân phối giống dữ liệu thật.
6	VAE	Variational Autoencoder: Mô hình Generative AI học phân phối tiềm ẩn để sinh dữ liệu mới qua cơ chế mã hóa–giải mã với chuẩn hóa xác suất.
7	GPT	Generative Pre-trained Transformer: Mô hình

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Thuật ngữ & viết tắt	Giải thích
		transformer tiên huấn luyện trên khối lượng lớn văn bản, có khả năng sinh ngôn ngữ tự nhiên.
8	Stable Diffusion	Kỹ thuật tạo ảnh từ nhiễu dựa trên quá trình khuếch tán ngược, kiểm soát chất lượng và phong cách.
9	Whisper	Mô hình nhận diện giọng nói của OpenAI, chuyển đổi âm thanh giọng nói thành văn bản.
10	MusicGen	Mô hình DL sinh nhạc tự động theo phong cách và cấu trúc đầu vào.
11	CNN	Convolutional Neural Network: Mạng nơ-ron tích chập, chuyên xử lý dữ liệu lưới (ảnh, video) bằng lớp tích chập để trích xuất đặc trưng.
12	RNN	Recurrent Neural Network: Mạng nơ-ron hồi tiếp, phù hợp xử lý dữ liệu tuần tự nhờ liên

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Thuật ngữ & viết tắt	Giải thích
		kết trạng thái qua thời gian.
13	LSTM	Long Short-Term Memory: Biến thể của RNN, giúp duy trì thông tin dài hạn qua cổng nhớ.
14	GRU	Gated Recurrent Unit: Biến thể RNN tích hợp cổng cập nhật và khôi phục, đơn giản hóa cấu trúc so với LSTM.
15	Transformer	Kiến trúc mạng dựa trên cơ chế self-attention, xử lý song song và đạt hiệu quả cao trong NLP.
16	ResNet	Residual Network: Mạng nơ-ron với kết nối tắt (skip connection), giảm thiểu mất mát thông tin khi mô hình sâu tầng.
17	EfficientNet	Kiến trúc CNN tự động cân chỉnh độ sâu, độ rộng và độ phân giải để tối ưu hiệu suất.

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Thuật ngữ & viết tắt	Giải thích
18	SVM	Support Vector Machine: thuật toán học có giám sát, phân loại bằng cách tìm siêu phẳng tối ưu phân tách các lớp dữ liệu.
19	KNN	k-Nearest Neighbors: thuật toán học máy có giám sát, phân loại dựa trên nhãn của k mẫu gần nhất trong không gian đặc trưng.
20	DBSCAN	Density-Based Spatial Clustering of Applications with Noise: thuật toán học máy không giám sát, phân nhóm dữ liệu dựa trên mật độ, tự động phát hiện điểm nhiễu.
21	XGBoost	Extreme Gradient Boosting: Thuật toán học máy có giám sát dựa trên kỹ thuật tăng cường (boosting), sử dụng nhiều cây quyết định liên tiếp

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Thuật ngữ & viết tắt	Giải thích
		để tối ưu độ chính xác. Thuật toán tập trung vào khả năng xử lý dữ liệu bị thiếu, kiểm soát tình trạng quá khớp (overfitting) tốt.
22	LightGBM	Light Gradient Boosting Machine: Thuật toán học máy có giám sát dựa trên kỹ thuật boosting, sử dụng nhiều cây quyết định liên tiếp để tối ưu độ chính xác. Thuật toán tập trung vào các kỹ thuật giúp tăng tốc độ huấn luyện, giảm bộ nhớ tiêu thụ.
23	K-means	Thuật toán phân cụm không giám sát, nhóm dữ liệu thành k cụm để tối thiểu hoá khoảng cách tới tâm cụm.
24	PCA	Principal Component Analysis: Phương pháp giảm chiều bằng cách

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Thuật ngữ & viết tắt	Giải thích
		chiều dữ liệu lên trục chính có phương sai lớn nhất.
25	t-SNE	t-Distributed Stochastic Neighbor Embedding: Phương pháp giảm chiều phi tuyến, giữ cấu trúc không gian cục bộ khi trực quan hóa dữ liệu cao chiều.
26	Isolation Forest	Thuật toán phát hiện bất thường, cô lập điểm dữ liệu khác biệt qua mô hình rừng cây ngẫu nhiên.

4. Các kỹ thuật AI

- Các kỹ thuật AI được chia thành 3 kỹ thuật chính: (1) Học máy (Machine Learning – ML); (2) Học sâu (Deep Learning – DL); (3) Trí tuệ nhân tạo tạo sinh (Generative AI – GenAI). Mỗi kỹ thuật sở hữu những ưu điểm và hạn chế đặc thù — từ khả năng giải thích rõ ràng, hiệu suất cao của ML, đến sức mạnh xử lý dữ liệu phi cấu trúc của DL và khả năng sáng tạo của GenAI. Cụ thể như sau:
- *Bảng 1. Các kỹ thuật AI*

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Kỹ thuật	Định nghĩa	Ưu điểm	Nhược điểm	Ứng dụng điển hình	Thuật toán điển hình
1	ML ¹	- Kỹ thuật AI sử dụng thuật toán thống kê để học từ dữ liệu quá khứ và đưa ra dự đoán mà không cần lập trình tường minh.	<ul style="list-style-type: none"> - Xử lý tốt dữ liệu có cấu trúc. - Một số thuật toán có khả năng giải thích được và tốt hơn học sâu. - Chi phí tính toán thấp hơn học sâu. - Mạnh với dữ liệu vừa và nhỏ. 	<ul style="list-style-type: none"> - Hiệu suất kém với dữ liệu phi cấu trúc (ảnh, video, âm thanh, văn bản). - Cần bước trích chọn đặc trưng thủ công. - Độ chính xác giới hạn với bài toán có mẫu phức tạp. - Giới hạn khả năng khai thác tương tác giữa các đặc trưng phức tạp (phi tuyến). 	<ul style="list-style-type: none"> - Dự báo nhu cầu, dự báo doanh thu. - Phân loại tín dụng, rủi ro. - Phát hiện gian lận. - Dự đoán khách hàng rời mạng. 	<ul style="list-style-type: none"> - Linear Regression, Logistic Regression. - Decision Tree, Random Forest. - Gradient Boosting, XGBoost, LightGBM. - SVM, KNN, - Naive Bayes. - K-means, DBSCAN
2	DL ²	Nhánh của ML sử dụng mạng nơ-ron sâu (Deep Neural Networks – DNN) có khả năng tự động học biểu diễn từ dữ liệu.	<ul style="list-style-type: none"> - Tự động trích xuất đặc trưng từ dữ liệu. - Xử lý mạnh dữ liệu phi cấu trúc (ảnh, âm thanh, văn bản, video). - Hỗ trợ học đa phương thức bao gồm ảnh, văn bản, dữ liệu bảng... - Tối ưu hóa tốt cho các bài toán có dữ liệu lớn, phức tạp. - Đạt độ chính xác rất 	<ul style="list-style-type: none"> - Yêu cầu dữ liệu rất lớn và tài nguyên phần cứng mạnh (GPU/TPU). - Tính đóng, khó giải thích. - Chi phí tính toán cao, độ trễ cao khi suy luận. - Dễ bị quá khớp(overfitting) nếu không có dữ liệu hoặc kỹ thuật chính quy hóa (regularization) tốt. - Quá trình huấn luyện phức tạp, nhạy với các siêu tham số, do đó cần tinh chỉnh kỹ lưỡng. 	<ul style="list-style-type: none"> - Nhận diện khuôn mặt, y tế chẩn đoán hình ảnh. - Phân tích văn bản, phân tích cảm xúc. - Dịch máy, nhận dạng tiếng nói. - Phát hiện đối tượng trong video. 	<ul style="list-style-type: none"> - CNN - RNN, LSTM, GRU - Transformer - ResNet, EfficientNet

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Kỹ thuật	Định nghĩa	Ưu điểm	Nhược điểm	Ứng dụng điển hình	Thuật toán điển hình
			cao trong thị giác máy tính, xử lý ngôn ngữ tự nhiên.			
3	GenAI	Nhánh của học sâu sử dụng các mô hình có khả năng sinh ra dữ liệu mới (văn bản, hình ảnh, âm thanh, video, code, nhạc).	<ul style="list-style-type: none"> - Sinh nội dung sáng tạo (văn bản, ảnh, video, nhạc, code). - Học biểu diễn phức tạp từ dữ liệu phi cấu trúc. - Tăng tốc quy trình sáng tạo, lập trình, nghiên cứu và phát triển (R&D). - Hỗ trợ xây dựng dữ liệu tổng hợp (synthetic data) để huấn luyện mô hình khác. - Kết hợp mạnh với tạo sinh dựa trên truy xuất tăng cường (RAG), hỗ trợ bài toán doanh nghiệp phức tạp. 	<ul style="list-style-type: none"> - Chi phí tính toán rất cao. - Đầu ra có thể chứa thông tin sai lệch, thiên kiến (bias), ảo giác (hallucination). - Độ tin cậy phụ thuộc mạnh vào prompt, chất lượng tinh chỉnh mô hình (fine tune). - Vấn đề về bản quyền, đạo đức và kiểm duyệt nội dung. - Khó kiểm soát chính xác logic nội tại của mô hình. 	<ul style="list-style-type: none"> - Sinh nội dung truyền thông - Trợ lý lập trình: sinh mã nguồn, sinh tài liệu phân tích. - Trợ lý ảo cho doanh nghiệp. - Tạo dữ liệu tổng hợp trong nghiên cứu y học, sản xuất. 	<ul style="list-style-type: none"> - GPT. - Stable Diffusion - Whisper. - MusicGen. - GAN, VAE

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

¹ Lưu ý: ML là khái niệm tổng quát bao gồm tất cả các phương pháp cho phép hệ thống học từ dữ liệu, bao gồm cả học sâu. Tuy nhiên trong tài liệu này, khi đề cập tới ML, ngầm hiểu là ML cổ điển, bao gồm lớp thuật toán thống kê, không bao gồm kiến trúc mạng nơ-ron sâu đặc trưng của học sâu.

² Lưu ý: DL là tập hợp các kỹ thuật học dựa trên kiến trúc mạng nơ-ron sâu, bao gồm cả mô hình dự đoán và mô hình sinh. Tuy nhiên trong tài liệu này, trí tuệ nhân tạo tạo sinh được tách riêng khỏi học sâu để nhấn mạnh nhóm mô hình có mục tiêu sinh dữ liệu mới (văn bản, hình ảnh, âm thanh...). Vì vậy, khi đề cập tới học sâu, ngầm hiểu là nhóm mô hình học sâu với mục tiêu dự đoán, nhận diện, không bao gồm nhóm mô hình sinh dữ liệu.

5. Bộ tiêu chí đánh giá mức độ phù hợp của các kỹ thuật AI

- Để lựa chọn kỹ thuật AI phù hợp với từng bài toán cụ thể trong doanh nghiệp, cần thiết lập một bộ tiêu chí đánh giá rõ ràng. Các tiêu chí này giúp tổ chức hiểu đúng bản chất bài toán, đặc điểm dữ liệu, cũng như yêu cầu triển khai thực tế, từ đó giới hạn và đề xuất phạm vi kỹ thuật có thể áp dụng.
- Để đánh giá mức độ phù hợp của các kỹ thuật AI với từng tiêu chí, tài liệu áp dụng thang đo 3 mức như sau:
 - Phù hợp (P): Kỹ thuật được thiết kế và hoạt động hiệu quả nhất cho tiêu chí này.
 - Cân nhắc (C): Kỹ thuật có thể đáp ứng tiêu chí trong một số điều kiện nhất định, hoặc khi được kết hợp với các kỹ thuật khác, nhưng không phải là lựa chọn hàng đầu.
 - Không phù hợp (K): Kỹ thuật không phù hợp hoặc hoạt động kém hiệu quả cho tiêu chí này.
- Bảng 2. Bảng đánh giá mức độ phù hợp của các kỹ thuật AI theo từng tiêu chí

STT	Tiêu chí	Thuộc tính đánh giá	Định nghĩa	Đánh giá chi tiết	ML	DL	Gen AI
1	Mục tiêu bài toán	Dự đoán	Dự báo biến liên tục (regression) hoặc phân loại nhãn (classification) dựa trên tập dữ liệu có nhãn.	- ML, DL phù hợp: bản chất của ML và DL là xây dựng mô hình học ánh xạ từ dữ liệu đầu vào tới một kết quả xác định. - GenAI cân nhắc: vì GenAI không được thiết kế cho bài toán dự đoán, tuy nhiên có thể sử dụng trong một số trường hợp đặc biệt như sinh kết quả dự đoán dựa trên ngữ cảnh, nhưng không phải là lựa chọn tối ưu cho các bài toán dự đoán hoặc phân loại nhãn rõ ràng.	P	P	C

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Tiêu chí	Thuộc tính đánh giá	Định nghĩa	Đánh giá chi tiết	ML	DL	Gen AI
2	Loại dữ liệu	Tạo sinh	Tạo mới dữ liệu phi cấu trúc như văn bản, hình ảnh, video, âm thanh hoặc mã nguồn.	<ul style="list-style-type: none"> - GenAI phù hợp: Đây là mục tiêu chính của GenAI - DL cân nhắc: Các mô hình DL truyền thống không linh hoạt và hiệu quả bằng các mô hình sinh hiện đại, tuy nhiên có thể cân nhắc trong trường hợp dữ liệu không quá lớn. - ML không phù hợp: Các thuật toán ML không có khả năng sinh dữ liệu mới. 	K	C	P
		Dữ liệu có cấu trúc	Dữ liệu dạng bảng, gồm hàng và cột với kiểu dữ liệu rõ ràng.	<ul style="list-style-type: none"> - ML phù hợp: Các thuật toán ML được tối ưu hóa để xử lý và tìm ra các mối quan hệ trong dữ liệu có cấu trúc - DL cân nhắc: DL có thể xử lý được có cấu trúc nhưng thường không mang lại kết quả vượt trội so với ML mà lại tốn kém tài nguyên hơn. Do đó DL không phải lựa chọn ưu tiên. - GenAI không phù hợp: GenAI không được thiết kế để làm việc với dữ liệu có cấu trúc. 	P	C	K
		Dữ liệu phi cấu trúc	Dữ liệu không có cấu trúc bảng định sẵn, bao gồm ảnh, văn bản tự do, âm thanh, video...	<ul style="list-style-type: none"> - ML không phù hợp: ML không xử lý trực tiếp dữ liệu phi cấu trúc, cần phải trích xuất các đặc trưng thủ công trước khi áp dụng mô hình, điều này rất tốn thời gian và không hiệu quả. - DL và GenAI phù hợp: DL và GenAI đều được thiết kế để xử lý dữ liệu phi cấu trúc thông qua các kiến trúc mạng nơ-ron sâu, có khả năng học các đặc trưng phức tạp một cách tự động. 	K	P	P
3	Nhãn dữ liệu	Có nhãn rõ ràng	Tập dữ liệu đã gán nhãn đầu ra (supervised).	<ul style="list-style-type: none"> - ML và DL phù hợp: Các kỹ thuật đều dựa trên dữ liệu có nhãn để học các mẫu và mối quan hệ. - GenAI cân nhắc: 	P	P	C

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Tiêu chí	Thuộc tính đánh giá	Định nghĩa	Đánh giá chi tiết	ML	DL	Gen AI
		Không nhãn	Tập dữ liệu chưa có nhãn, cần khám phá cấu trúc tiềm ẩn (unsupervised).	<ul style="list-style-type: none"> - GenAI phù hợp: Kỹ thuật này học và tạo ra cấu trúc từ dữ liệu thô, không cần nhãn. - ML và DL cân nhắc: Trong trường hợp các bài toán unsupervised, các kỹ thuật này có thể sử dụng để khám phá các cấu trúc từ dữ liệu. 	C	C	P
4	Số chiều dữ liệu	Số chiều dữ liệu nhỏ	Dữ liệu có số lượng đặc trưng (feature) nhỏ.	<ul style="list-style-type: none"> - ML phù hợp: ML hoạt động hiệu quả với dữ liệu có ít feature - DL và GenAI cân nhắc: Có thể cân nhắc sử dụng nhưng thường không cần thiết và có nguy cơ overfitting. 	P	C	C
		Số chiều dữ liệu lớn	Dữ liệu có số lượng feature rất lớn	<ul style="list-style-type: none"> - DL và GenAI phù hợp: Kiến trúc mạng nơ-ron sâu được thiết kế để xử lý và trích xuất đặc trưng từ dữ liệu có số chiều lớn một cách tự động. - ML không phù hợp: ML có hiệu suất kém và khó xử lý khi số lượng đặc trưng tăng lên. 	K	P	P
5	Số lượng dữ liệu	Khối lượng dữ liệu nhỏ	Tập dữ liệu nhỏ	<ul style="list-style-type: none"> - ML phù hợp: ML hoạt động tốt trên các tập dữ liệu nhỏ, yêu cầu ít tài nguyên tính toán và thường cho kết quả tốt. - DL và GenAI cân nhắc: DL và GenAI đòi hỏi lượng lớn dữ liệu để học các đặc trưng phức tạp. Với dữ liệu nhỏ, DL dễ bị overfitting và kém hiệu quả. Tuy nhiên có thể cân nhắc nếu sử dụng các mô hình đã được huấn luyện sẵn hoặc tinh chỉnh trên dữ liệu nhỏ. 	P	C	C

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Tiêu chí	Thuộc tính đánh giá	Định nghĩa	Đánh giá chi tiết	ML	DL	Gen AI
		Khối lượng dữ liệu lớn	Tập dữ liệu lớn	- DL và GenAI phù hợp: Các kỹ thuật này có khả năng khai thác các mối quan hệ phức tạp từ dữ liệu lớn. - ML cân nhắc: ML có thể hoạt động nhưng thường kém hiệu quả hơn, đặc biệt khi dữ liệu có nhiều đặc trưng phức tạp.	C	P	P
6	Yêu cầu khả năng giải thích	Yêu cầu khả năng giải thích rõ ràng, dễ kiểm tra	Yêu cầu khả năng giải thích rõ ràng, dễ kiểm tra nguyên nhân mô hình đưa ra quyết định cụ thể.	- ML phù hợp: Các thuật toán ML như Linear regression, Decision tree thường dễ giải thích, minh bạch và có thể kiểm tra các quyết định của mô hình. - DL và GenAI cân nhắc: Các kỹ thuật này dựa trên các mô hình mang tính chất hộp đen (black-box), rất khó để giải thích và truy vết nguyên nhân cụ thể. Tuy nhiên, có thể cân nhắc khi áp dụng kết hợp các kỹ thuật giải thích mô hình bổ sung hoặc đối với GenAI thì yêu cầu sinh phương pháp lý luận (tuy nhiên về bản chất mô hình vẫn không diễn giải được)	P	C	C

6. Các bài toán AI phổ biến

- Theo Viện tiêu chuẩn và Công nghệ Quốc gia của Hoa Kỳ (National Institute of Standards and Technology – NIST), bài toán AI được phân loại theo 14 bài toán phổ biến như sau:
- Bảng 3. Bảng phân loại các bài toán AI phổ biến

STT	Bài toán AI	Mô tả	Ứng dụng cụ thể
-----	-------------	-------	-----------------

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

1	Tạo nội dung	Tạo ra các sản phẩm, dữ liệu hoặc nội dung mới dựa trên các yêu cầu hoặc mục tiêu được xác định rõ ràng.	Tạo phụ đề tự động cho video quảng cáo; sinh mã lập trình từ mô tả; tạo hình ảnh từ nội dung văn bản
2	Tổng hợp nội dung	Kết hợp, phân tích, tóm tắt các dữ liệu hoặc thông tin riêng lẻ để tạo ra một nội dung hoàn chỉnh, có cấu trúc và ý nghĩa rõ ràng.	Tóm tắt báo cáo tài chính; chuyển đổi ghi chú y tế phi cấu trúc của bác sĩ thành hồ sơ bệnh án đầy đủ.
3	Ra quyết định	Hỗ trợ hoặc đề xuất quyết định tối ưu từ nhiều lựa chọn, dựa trên phân tích dữ liệu, dự đoán hoặc đánh giá rủi ro.	Hỗ trợ quyết định mua bán cổ phiếu; hệ thống đề xuất phê duyệt khoản vay tự động dựa trên lịch sử tín dụng.
4	Phát hiện	Nhận diện và xác định sự hiện diện của các vấn đề, sự kiện hoặc nguy cơ tiềm ẩn thông qua phân tích dữ liệu.	Phát hiện gian lận giao dịch ngân hàng; phát hiện tấn công mạng; nhận dạng sự cố kỹ thuật trong dây chuyền sản xuất.
5	Trợ lý ảo	Hỗ trợ người dùng bằng cách hiểu, phản hồi và thực hiện các yêu cầu hoặc tác vụ qua giao tiếp tự nhiên như một trợ lý cá nhân.	Siri, Alexa hỗ trợ đặt lịch họp; trợ lý ảo giúp trả lời câu hỏi khách hàng trực tuyến tự động.
6	Khám phá	Khai phá, nhận dạng hoặc tìm ra thông tin, sản phẩm hoặc quy trình hoàn toàn mới mà con người chưa biết tới hoặc chưa tiếp cận được trước đó.	Phát hiện phân tử mới trong dược phẩm; khám phá vật liệu pin tiên tiến; tối ưu công thức sản phẩm mới trong sản xuất.
7	Phân tích hình ảnh	Nhận diện và phân tích nội dung từ hình ảnh số để đưa ra thông tin hữu ích, hỗ trợ việc ra quyết định.	Hệ thống nhận dạng ung thư sớm từ hình ảnh chẩn đoán y tế; phân tích sản phẩm lỗi từ hình ảnh trên dây chuyền sản xuất.
8	Truy xuất thông tin	Hỗ trợ người dùng tìm kiếm, tiếp cận nhanh chóng và chính xác các thông tin cần thiết từ nguồn dữ liệu lớn, đa dạng.	Truy xuất tài liệu pháp lý theo từ khóa; tìm kiếm protein ổn định cho nghiên cứu thuốc nhanh hơn.
9	Giám sát	Theo dõi, giám sát liên tục trạng thái, chất lượng hoặc hiệu suất của hệ thống để phát hiện bất thường hoặc cải tiến hoạt động.	Giám sát tình trạng thiết bị công nghiệp theo thời gian thực; giám sát môi trường để phát hiện cháy rừng tự động.
10	Cải thiện hiệu suất	Tối ưu hóa, nâng cao chất lượng, độ chính xác hoặc hiệu suất vận hành của quy trình hoặc hệ thống hiện có.	Tối ưu thuật toán phân tích dữ liệu lớn; cải tiến độ chính xác mô hình dự báo thời tiết.

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

11	Cá nhân hóa	Điều chỉnh và cung cấp sản phẩm, dịch vụ, nội dung phù hợp nhất với từng cá nhân dựa trên hành vi, sở thích và đặc điểm riêng của họ.	Cá nhân hóa nội dung quảng cáo theo hành vi mua sắm trực tuyến; cung cấp nội dung giáo dục phù hợp theo năng lực học viên.
12	Dự báo	Dự đoán khả năng xảy ra của các sự kiện hoặc kết quả trong tương lai, dựa trên phân tích dữ liệu lịch sử và mô hình thống kê.	Dự báo doanh thu quý kế tiếp; dự báo khả năng hỏng hóc máy móc trong sản xuất; dự báo lưu lượng khách hàng để tối ưu dịch vụ.
13	Tự động hóa quy trình	Thực hiện các tác vụ thường xuyên, lặp lại hoặc dễ xảy ra sai sót nhằm giảm thời gian, tiết kiệm nguồn lực và nâng cao hiệu quả hoạt động.	Tự động hóa quản lý hóa đơn và thanh toán; xử lý, phân loại email nội bộ tự động.
14	Khuyến nghị	Đưa ra các gợi ý, lựa chọn hợp lý nhất cho người dùng nhằm hỗ trợ họ ra quyết định một cách hiệu quả.	Đề xuất sản phẩm phù hợp cho khách hàng mua sắm trực tuyến; gợi ý phản hồi nhanh trong hệ thống chăm sóc khách hàng.

7. Đánh giá mức độ phù hợp của các kỹ thuật AI theo từng bài toán phổ biến

Để đánh giá mức độ phù hợp của các kỹ thuật AI theo từng bài toán, thực hiện như sau:

- Thực hiện đánh giá các đặc điểm và yêu cầu của bài toán dựa trên bộ tiêu chí đã được xây dựng.
- Tham chiếu bảng đánh giá đã xây dựng, với mỗi tiêu chí xác định mức độ phù hợp của từng kỹ thuật AI:
 - Không phù hợp (K): Kết luận kỹ thuật không phù hợp.
 - Cân nhắc (C): Không cộng điểm.
 - Phù hợp (P): Cộng 1 điểm.
- Độ phù hợp tổng thể của một kỹ thuật AI được xác định bằng tổng điểm các tiêu chí (chỉ tính nếu không có tiêu chí nào bị đánh giá là không phù hợp).
- Kỹ thuật có tổng điểm cao nhất được ưu tiên lựa chọn (phù hợp), các kỹ thuật có tổng điểm thấp hơn cần được cân nhắc và đánh giá kỹ lưỡng, có thể được áp dụng trong các trường hợp đặc biệt hoặc kết hợp với các kỹ thuật khác để khắc phục các điểm yếu cho kỹ thuật chính. Trong trường hợp có nhiều kỹ thuật có tổng điểm bằng nhau, cần ưu tiên theo thứ tự quan trọng của tiêu chí từ mục tiêu bài toán, loại dữ liệu, nhãn dữ liệu, số chiều dữ liệu, số lượng dữ liệu, cho đến yêu cầu khả năng giải thích hoặc

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

cân nhắc kết hợp các kỹ thuật để tận dụng điểm mạnh của từng kỹ thuật.

- Căn cứ dựa trên đặc điểm của 14 bài toán AI phổ biến đã được phân loại, chiếu theo bộ tiêu chí đánh giá đã xây dựng để xác định mức độ phù hợp của các kỹ thuật AI cho từng bài toán. Cụ thể như sau:

- *Bảng 4. Bảng đánh giá mức độ phù hợp của các kỹ thuật AI theo từng bài toán phổ biến.*

STT	Bài toán	Chiều đánh giá		Mục tiêu bài toán	Loại dữ liệu	Nhãn dữ liệu	Số chiều dữ liệu	Số lượng dữ liệu	Yêu cầu khả năng giải thích	Tổng điểm	Kết luận
1	Tạo nội dung	Đặc điểm		Tạo sinh	Phi cấu trúc	Không nhãn	Lớn	Lớn	Không bắt buộc		
		Kỹ thuật	ML	K	K	C	K	C	P	0	K
			DL	C	P	C	P	P	P	4	C
			GenAI	P	P	P	P	P	P	6	P
2	Tổng hợp nội dung	Đặc điểm		Tạo sinh	Phi cấu trúc	Không nhãn	Lớn	Lớn	Không bắt buộc		
		Kỹ thuật	ML	K	K	C	K	C	P	0	K
			DL	C	P	C	P	P	P	4	C
			GenAI	P	P	P	P	P	P	6	P
3	Ra quyết định	Đặc điểm		Dự đoán	Có cấu trúc	Có nhãn	Nhỏ	Lớn	Có		
		Kỹ thuật	ML	P	P	P	P	C	P	5	P
			DL	P	C	P	P	P	C	4	C
			GenAI	C	K	C	C	P	C	0	K
4	Phát hiện	Đặc điểm		Dự đoán	Phi cấu trúc	Có nhãn	Nhỏ	Lớn	Có		
		Kỹ thuật	ML	P	K	P	P	C	P	0	K
			DL	P	P	P	P	P	C	5	P
			GenAI	C	P	C	C	P	C	2	C
5	Trợ lý ảo	Đặc điểm		Tạo sinh	Phi cấu trúc	Không nhãn	Lớn	Lớn	Có		
		Kỹ thuật	ML	K	K	C	K	C	P	0	K
			DL	C	P	C	P	P	C	3	C
			GenAI	P	P	P	P	P	C	5	P

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Bài toán	Chiều đánh giá		Mục tiêu bài toán	Loại dữ liệu	Nhãn dữ liệu	Số chiều dữ liệu	Số lượng dữ liệu	Yêu cầu khả năng giải thích	Tổng điểm	Kết luận
6	Khám phá	Đặc điểm		Tạo sinh	Phi cấu trúc	Không nhãn	Lớn	Lớn	Có		
		Kỹ thuật	ML	K	K	C	K	C	P	0	K
			DL	C	P	C	P	P	C	3	C
			GenAI	P	P	P	P	P	C	5	P
7	Phân tích hình ảnh	Đặc điểm		Dự đoán	Phi cấu trúc	Có nhãn	Lớn	Lớn	Không bắt buộc		
		Kỹ thuật	ML	P	K	P	K	C	P	0	K
			DL	P	P	P	P	P	P	6	P
			GenAI	C	P	C	P	P	P	4	C
8	Truy xuất thông tin	Đặc điểm		Dự đoán	Phi cấu trúc	Có nhãn	Lớn	Lớn	Có		
		Kỹ thuật	ML	P	K	P	K	C	P	0	K
			DL	P	P	P	P	P	C	5	P
			GenAI	C	P	C	P	P	C	3	C
9	Giám sát	Đặc điểm		Dự đoán	Có cấu trúc	Không nhãn	Nhỏ	Lớn	Có		
		Kỹ thuật	ML	P	P	C	P	C	P	4	P
			DL	P	C	C	P	P	C	3	C
			GenAI	C	K	P	C	P	C	0	K
10	Cải thiện hiệu suất	Đặc điểm		Dự đoán	Có cấu trúc	Có nhãn	Nhỏ	Nhỏ	Có		
		Kỹ thuật	ML	P	P	P	P	C	P	5	P
			DL	P	C	P	P	P	C	4	C
			GenAI	C	K	C	C	P	C	0	K
11	Cá nhân hóa	Đặc điểm		Dự đoán	Có cấu trúc	Có nhãn	Nhỏ	Lớn	Có		
		Kỹ thuật	ML	P	P	P	P	C	P	5	P
			DL	P	C	P	P	P	C	4	C
			GenAI	C	K	C	C	P	C	0	K

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

STT	Bài toán	Chiều đánh giá	Mục tiêu bài toán	Loại dữ liệu	Nhãn dữ liệu	Số chiều dữ liệu	Số lượng dữ liệu	Yêu cầu khả năng giải thích	Tổng điểm	Kết luận
12	Dự báo	Đặc điểm	Dự đoán	Có cấu trúc	Có nhãn	Nhỏ	Lớn	Có		
		Kỹ thuật	ML	P	P	P	C	P	5	P
			DL	P	C	P	P	C	4	C
			GenAI	C	K	C	P	C	0	K
13	Tự động hóa quy trình	Đặc điểm	Dự đoán	Có cấu trúc	Không nhãn	Nhỏ	Nhỏ	Không bắt buộc		
		Kỹ thuật	ML	P	C	P	P	P	5	P
			DL	P	C	P	C	P	3	C
			GenAI	C	K	C	C	P	0	K
14	Khuyến nghị	Đặc điểm	Dự đoán	Có cấu trúc	Có nhãn	Nhỏ	Lớn	Có		
		Kỹ thuật	ML	P	P	P	C	P	5	P
			DL	P	C	P	P	C	4	C
			GenAI	C	K	C	P	C	0	K

Đánh giá chung:

- ML: phù hợp với các bài toán yêu cầu mô hình đơn giản, có khả năng giải thích tốt và dữ liệu có cấu trúc, điển hình như ra quyết định, giám sát, cải thiện hiệu suất, cá nhân hóa, dự báo, tự động hóa quy trình và khuyến nghị. Đây là những trường hợp mà dữ liệu đã được gán nhãn đầy đủ, số chiều không cao và yêu cầu giải thích thường được đặt ra rõ ràng. Tuy nhiên, ML không phù hợp với các bài toán có đặc điểm sinh nội dung, dữ liệu phi cấu trúc và không nhãn như tạo nội dung, tổng hợp nội dung, trợ lý số, khám phá, phân tích hình ảnh hay truy xuất thông tin. ML cũng gặp hạn chế trong bài toán phát hiện, nơi dữ liệu phi cấu trúc và nhãn phức tạp khiến hiệu quả suy giảm.

- DL: phù hợp với phần lớn các bài toán bao gồm cả sinh nội dung (tạo nội dung, tổng hợp nội dung, trợ lý số, khám phá), lẫn các tác vụ dự đoán như ra quyết định, phát hiện, phân tích hình ảnh, truy xuất thông tin, cải thiện hiệu suất, cá nhân hóa, dự báo, tự động hóa quy trình và khuyến nghị. DL cần được cân nhắc trong các trường hợp yêu cầu giải thích rõ như giám sát, hoặc dữ liệu nhỏ như tự động hóa quy trình, khi mà DL có thể áp dụng được nhưng không tối ưu nếu thiếu lượng dữ liệu lớn hoặc tài nguyên tính toán phù hợp.

	VIETTEL AI RACE	TD328
	HƯỚNG DẪN LỰA CHỌN KỸ THUẬT AI	Lần ban hành: 1

- GenAI: cho thấy hiệu quả vượt trội trong các bài toán sinh nội dung và tương tác ngôn ngữ tự nhiên như tạo nội dung, tổng hợp nội dung, trợ lý số và khám phá – tất cả đều đạt điểm đánh giá tối đa. GenAI cũng có thể cân nhắc áp dụng ở các bài toán như phân tích hình ảnh, phát hiện và truy xuất thông tin, đặc biệt khi cần hỗ trợ trực quan hóa hoặc mở rộng khả năng phân tích dữ liệu phi cấu trúc. Tuy nhiên, GenAI hiện không phù hợp cho các bài toán truyền thống thiên về dự đoán định lượng và yêu cầu tính giải thích cao như ra quyết định, giám sát, cải thiện hiệu suất, cá nhân hóa, dự báo, tự động hóa quy trình và khuyến nghị, do đặc điểm sinh mô hình không rõ ràng, chi phí huấn luyện lớn và khó kiểm soát logic suy luận