

NGÔN NGỮ HỌC VÀ VIỆT NGỮ HỌC

THỬ NGHIỆM XÂY DỰNG CHATBOT HỖ TRỢ PHÁT HIỆN VÀ SỬA LỖI NGỮ PHÁP CHO NGƯỜI NƯỚC NGOÀI HỌC TIẾNG VIỆT ĐỖ HỒNG DƯƠNG*

TÓM TẮT: Bài viết trình bày một nghiên cứu thử nghiệm thiết kế Chatbot dựa trên nền tảng Custom GPTs của OpenAI nhằm phát hiện và giải thích lỗi ngữ pháp, cụ thể là lỗi trật tự từ trong tiếng Việt - một hiện tượng cú pháp phổ biến và phức tạp. Trên cơ sở khung lý thuyết ngôn ngữ về trật tự từ trong tiếng Việt và các loại lỗi thường gặp, cùng các kiến thức về trí tuệ nhân tạo tạo sinh, nghiên cứu đã đánh giá một số mô hình AI hiện có trong việc nhận diện và giải thích lỗi ngữ pháp tiếng Việt, đồng thời đề xuất quy trình phát triển Chatbot sử dụng tính năng Custom GPTs của OpenAI. Kết quả cho thấy Chatbot có tiềm năng trở thành công cụ hỗ trợ học tiếng Việt hiệu quả trong việc đề xuất sửa lỗi, nhưng việc nhận diện và giải thích lỗi ngữ pháp phức tạp vẫn cần được cải thiện thông qua huấn luyện chuyên sâu. Bài báo cũng đề xuất các hướng phát triển tiếp theo để tối ưu hóa Chatbot trong giáo dục ngôn ngữ.

TỪ KHOẢ: ngữ pháp tiếng Việt; lỗi trật tự từ; trí tuệ nhân tạo; Custom GPTs; Chatbot sửa lỗi ngữ pháp.

NHẬN BÀI: 25/09/2025.

BIÊN TẬP-CHỈNH SỬA-DUYỆT ĐĂNG: 16/12/2025

1. Đặt vấn đề

Ngữ pháp tiếng Việt là một thử thách đối với người nước ngoài học tiếng Việt, bởi tiếng Việt là một ngôn ngữ thiên ngữ dụng, đôi khi những câu mà người nước ngoài tưởng như sai ngữ pháp hoàn toàn (vì thiếu động từ, thiếu chủ ngữ, thiếu từ liên kết,...) lại không bị người bản ngữ cho là sai. Bởi vậy, việc nói đúng, viết đúng ngữ pháp theo cách của người Việt là một đích đến khá khó khăn đối với người nước ngoài học tiếng Việt.

Trong thời kỳ phát triển của trí tuệ nhân tạo, việc nhận diện lỗi và sửa lỗi dường như đã dễ dàng hơn rất nhiều, đặc biệt là với các ứng dụng AI tạo sinh. Nếu như trước đây, việc xử lý ngôn ngữ tự nhiên cần một quá trình xử lý dữ liệu đầu vào vô cùng vất vả (thu thập ngôn ngữ, thiết lập quy tắc xử lý, gán nhãn,...), thì bây giờ, AI có thể tự học, tự đào tạo để ra một sản phẩm ngôn ngữ hoàn chỉnh, không khác gì với ngôn ngữ của người bản xứ. Tuy nhiên, qua một quá trình quan sát, chúng tôi nhận thấy rằng, các ứng dụng AI dù rất thông minh và tạo sinh đạt trình độ siêu việt, nhưng lại chưa thực sự tốt trong việc nhận ra những lỗi ngữ pháp tiếng Việt. Một số ứng dụng dù có thể nhận ra câu có lỗi, nhưng không thể giải thích chính xác lỗi đó là gì để giúp cho người nước ngoài hiểu sâu hơn về ngữ pháp tiếng Việt, từ đó tránh mắc lỗi tương tự về sau. Bởi vậy, đề tài này ra đời, nhằm tìm một giải pháp để hỗ trợ người nước ngoài trong quá trình học ngữ pháp tiếng Việt thông qua ứng dụng AI.

Vì ngữ pháp tiếng Việt rất rộng, nên trong quy mô bài báo, chúng tôi sẽ thử nghiệm trên một hiện tượng ngữ pháp trước. Chúng tôi chọn trật tự từ trong tiếng Việt, bởi trật tự từ trong tiếng Việt mang tính tương đối tự do so với nhiều ngôn ngữ khác, song vẫn có những quy tắc ngữ pháp nhất định chi phối thứ tự các thành phần trong câu. Nghiên cứu sẽ tập trung xây dựng một Chatbot có khả năng phát hiện và giải thích một số dạng lỗi trật tự từ điển hình, từ đó đánh giá tính khả thi và tiềm năng phát triển của hướng tiếp cận này trong tương lai.

*TS; Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Hà Nội;
Email: duong1910@gmail.com

2. Kết quả và thảo luận

2.1. Trật tự từ trong tiếng Việt và những lỗi thường gặp của người nước ngoài

2.1.1. Vai trò của trật tự từ trong ngữ pháp

Trong ngôn ngữ học, trật tự từ đóng vai trò cốt lõi trong việc tạo lập cấu trúc và truyền tải ý nghĩa, đặc biệt trong tiếng Việt, một ngôn ngữ đơn lập phụ thuộc nhiều vào trật tự từ để biểu đạt quan hệ ngữ pháp và ngữ nghĩa.

Ngữ pháp cấu trúc (construction grammar - CxG), được phát triển bởi Adele Goldberg, xem mỗi cấu trúc ngôn ngữ là một cặp hình thức-ngĩa (form-meaning pairing) độc lập. Trong CxG, trật tự từ là một phần không thể tách rời của ý nghĩa cấu trúc. Theo Goldberg (1995), sự thay đổi trật tự từ tạo ra một cấu trúc mới với ý nghĩa riêng biệt, không phụ thuộc vào cấu trúc gốc. Điều này nhấn mạnh rằng trật tự từ không chỉ là vấn đề cú pháp mà còn quyết định ngữ nghĩa và ngữ dụng của câu.

Ngữ pháp tri nhận (cognitive grammar - CG), do Ronald Langacker khởi xướng xem ngôn ngữ như một biểu hiện của các quá trình nhận thức. Trật tự từ phản ánh cách con người khái niệm hóa và tổ chức thông tin trong tư duy. Trong tiếng Việt, trật tự từ xác định vai trò ngữ pháp và ngữ nghĩa của các thành phần trong câu. Ví dụ, câu “Tôi đọc sách” (I read a book) đặt “tôi” trước để làm nổi bật vai trò tác nhân (agent), trong khi “Sách tôi đọc” (The book I read) nhấn mạnh “sách” như chủ đề (topic). Theo Langacker (1987), trật tự từ ảnh hưởng đến cách người nghe diễn giải cảnh khái niệm (conceptual scene), làm cho nó trở thành yếu tố quan trọng trong việc truyền tải ý nghĩa.

Ngữ pháp chức năng (Functional Grammar), đặc biệt là Ngữ pháp chức năng hệ thống (Systemic Functional Grammar - SFG) của Michael Halliday tập trung vào cách ngôn ngữ được sử dụng để đạt được mục tiêu giao tiếp. SFG chia chức năng của ngôn ngữ thành ba siêu chức năng (metafunctions): Ý niệm (Ideational), Liên nhân (Interpersonal), Văn bản (Textual). Trong tiếng Việt, trật tự từ phục vụ cả ba siêu chức năng này. Ví dụ, trong câu “Cái này tôi thích lắm” việc đặt “cái này” ở đầu câu (textual function) làm nổi bật chủ đề, đồng thời thể hiện sự nhấn mạnh của người nói (interpersonal function) và xác định mối quan hệ giữa các thành phần trong câu (ideational function). Theo Halliday (1994), trật tự từ là công cụ để tổ chức diễn ngôn và đảm bảo giao tiếp hiệu quả, đặc biệt trong các ngôn ngữ nổi bật chủ đề như tiếng Việt.

2.2.2. Trật tự từ các thành phần câu và ngữ đoạn tiếng Việt và lỗi thường gặp của người nước ngoài

Tiếng Việt là một ngôn ngữ đơn lập, trong đó ý nghĩa ngữ pháp chủ yếu được thể hiện qua hai phương thức: trật tự từ và hư từ. Vì vậy, trật tự từ đóng vai trò then chốt trong việc truyền tải ý nghĩa và đảm bảo tính rõ ràng của câu. Một ví dụ điển hình cho vai trò này là trường hợp do Lê Văn Lý nêu ra với 05 từ: *nó, bảo, sao, không, đến*. Chỉ bằng cách hoán đổi trật tự của năm từ này, người ta có thể tạo ra hơn 30 câu khác nhau, chẳng hạn: *Nó bảo sao không đến, Nó bảo sao đến không, Nó bảo đến không sao, Nó bảo đến sao không, Bảo sao nó không đến, Bảo nó đến không sao...* Nguyễn Đức Dân (2013, tr.92) trích dẫn ví dụ này và kết luận: “Thay đổi trật tự từ là thay đổi nghĩa của câu, làm một câu đơn giản thành rắc rối, làm một câu đúng thành câu sai hoặc ngược lại”. Ông cũng nêu thêm một trường hợp sai lệch trong quảng cáo dầu xả trên truyền hình với câu: *Còn đâu 5 mùi khó chịu*. Cách dùng *còn đâu* khiến người nghe liên tưởng đến sự nuối tiếc, trong khi ý định ban đầu lại là khẳng định sự biến mất của các mùi khó chịu. Trật tự đúng phải là *Đâu còn (5 mùi khó chịu)*. Đây là một ví dụ tiêu biểu cho việc thay đổi trật tự từ có thể làm biến đổi sắc thái câu nói, dẫn đến cách hiểu sai lệch và những hệ quả đáng tiếc trong giao tiếp tiếng Việt.

(a) Trật tự các thành phần câu và lỗi thường gặp

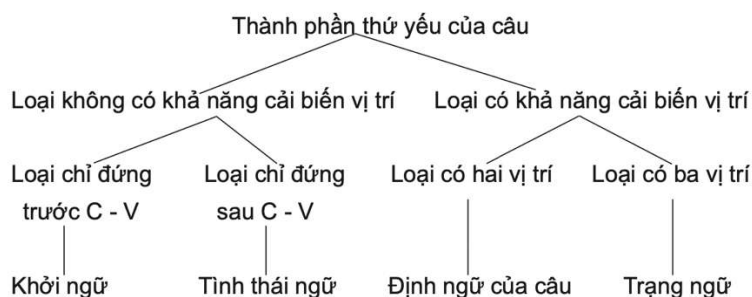
Theo lý thuyết loại hình học về trật tự từ, cấu trúc cơ bản của câu tiếng Việt là cấu trúc SVO với: S: Chủ ngữ, V: Động từ, O: Bổ ngữ (Tôi ăn cơm, Anh ấy đọc sách,...).

Theo lý thuyết về cú pháp tiếng Việt của Nguyễn Văn Hiệp (2009), câu tiếng Việt có thể có 7 thành phần: (P: Thành phần phụ, C: thành phần chính).

P1	P2	P3	C1	C2	C3	P4
<i>Trạng ngữ</i>	<i>Định ngữ câu</i>	<i>Khởi ngữ</i>	<i>Chủ ngữ</i>	<i>Vị ngữ</i>	<i>Bổ ngữ</i>	<i>Tình thái ngữ</i>
Hôm qua,	quả thật là	em gái em	nó	đã nấu	cả bữa cơm	đấy.
Ở đó,	có lẽ	những đứa trẻ này	chúng nó	sẽ thấy hơn	hạnh phúc	chẳng?

Về khả năng cải biến vị trí, trật tự của các thành phần chính là cố định, chỉ có một số ít trường hợp C2 đảo lên trước C1 nhằm mục đích nhấn mạnh (thường trong văn học, ví dụ: *Ngót rồi cơn mưa đầu mùa hạ*). C3 nếu đảo lên trước C1 sẽ trở thành P3 (xem Nguyễn Văn Hiệp, 2009).

Về các thành phần phụ, trật tự của P1, P2, P3 là không cố định, có thể đổi chỗ cho nhau và có thể đứng sau C1 (trước C2). Vị trí của P4 là cố định (cuối câu).



Sơ đồ về khả năng cải biến vị trí của các thành phần phụ (Nguyễn Văn Hiệp, 2009)

Ví dụ về khả năng cải biến vị trí của P1, P2, P3:

(+) [*Hôm qua*], [*quả thật là*] [*em gái em*] [*nó*] [*đã nấu*] [*cả bữa cơm*] [*đấy*].

(+) *Quả thật là hôm qua em gái em nó đã nấu cả bữa cơm đấy.*

(+) *Em gái em quả thật là hôm qua nó đã nấu cả bữa cơm đấy.*

(...)

Ngoài ra, P1 còn có khả năng đứng sau C3 (sau nòng cốt câu). Tức P1 có thể đứng ở cả 3 vị trí.

(+) *Quả thật là em gái em nó đã nấu cả bữa cơm hôm qua đấy.*

Tuy vậy, không phải trường hợp nào trạng ngữ cũng có thể đứng ở cả 3 vị trí. Ví dụ trong câu “*Ngày mai quán này mới mở cửa*”, tình hình vị trí của “*ngày mai*” như sau:

(+) *Ngày mai, quán này mới mở cửa.*; (+) *Quán này ngày mai mới mở cửa.*

(-) *Quán này mới mở cửa ngày mai.*

Trạng ngữ “*ngày mai*” không thể đứng ở cuối câu này vì “*mới*” cho biết một trạng thái sự việc xảy ra muộn ở một thời gian nào đó, trong ngữ cảnh này mốc thời gian cần được nhấn mạnh ở đầu hoặc giữa câu (trước “*mới*”). Nếu “*ngày mai*” đứng cuối câu thì cần bỏ “*mới*” đi cho hợp ngữ cảnh.

Tương tự, khi xuất hiện các phó từ biểu thị sự việc, hiện tượng xảy ra sớm, trạng ngữ cũng chỉ xuất hiện ở vị trí đầu hoặc giữa câu:

(+) *Bảy giờ mẹ đã về.*; (+) *Mẹ bảy giờ đã về.*; (-) *Mẹ đã về bảy giờ.*

Nếu “*đã*” không phải phó từ biểu thị sự việc xảy ra sớm (so với mốc thời gian) mà biểu thị thời quá khứ (*đã/ đang/ sẽ*) thì trạng ngữ cần được chuyển thành một giới ngữ (ví dụ *từ bảy giờ*)

Trong các trường hợp thông thường khác, trạng ngữ không đứng ở 1 trong 3 vị trí đã cho trước sẽ gây lỗi. Lỗi trật tự thành phần câu thường gặp là:

Lỗi vị trí C2, ví dụ: *Bắt đầu chúng tôi học từ 7 giờ.*

Lỗi vị trí P1, ví dụ: *Hôm qua tôi đã ở nhà thờ nhìn thấy bạn.*

Lỗi vị trí P2, ví dụ: *Nếu một người yêu thích một việc gì đó thì sẽ làm việc đó tốt hơn chắc chắn. / Tết trung thu ở Việt Nam là đặc biệt nhất theo em thấy.*

(b) *Trật tự các thành tố trong ngữ đoạn và lỗi thường gặp*

Trật tự các thành tố trong danh ngữ, Theo Nguyễn Tài Cẩn (1999, tr.203-246), mô hình cấu trúc của danh ngữ tiếng Việt như sau:

Bảng 1. Các thành phần trong cụm danh từ (theo Nguyễn Tài Cẩn, 1999)

PHẦN ĐẦU			PHẦN TRUNG TÂM		PHẦN CUỐI
Định tố chỉ toàn bộ	Định tố chỉ số lượng	Định tố “cái”	T1	T2	Định tố chỉ gồm một từ (có thể dùng kèm thành tố phụ của nó) (1) Định tố do mệnh đề đảm nhiệm (2)
-3	-2	-1	-0	+0	1
<i>Tất cả</i>	<i>những</i>	<i>cái</i>	<i>con</i>	<i>mèo</i>	<i>nhỏ này (1)</i>
<i>Cả</i>	<i>ba</i>	<i>cái</i>	<i>anh</i>	<i>sinh viên</i>	<i>mà hôm qua thi trượt (2)</i>

Hoàng Dũng và Nguyễn Thị Ly Kha (2004) cũng đề xuất mô hình danh ngữ với các thành tố phụ sau được diễn giải cụ thể:

Bảng 2. Các thành phần trong cụm danh từ (theo Hoàng Dũng và Nguyễn Thị Ly Kha, 2004)

CÁC THÀNH TỐ PHỤ TRƯỚC			TRUNG TÂM	CÁC THÀNH TỐ PHỤ SAU							
Lượng ngữ chỉ toàn thể	Lượng ngữ chỉ số lượng	Từ chỉ xuất	Từ chỉ đơn vị	Định ngữ hạn định		Định ngữ miêu tả			Định ngữ chỉ trở vị trí		
				1'		2'			3'		
-3	-2	-1	0	1'a	1'b	Định ngữ hàm ý phức số/ Định ngữ trang trí/ Định ngữ chỉ xuất			Định ngữ trực chỉ/ Định ngữ sở hữu/ Định ngữ chỉ vị trí/ Định ngữ là tiểu cú		
<i>tất cả</i>	<i>những</i>	<i>cái</i>	<i>bầy</i>	<i>cá</i>	<i>rô</i>	<i>đồng đúc</i>	<i>bé tí xíu</i>	<i>cuối cùng</i>	<i>ấy</i>	<i>của Đằm Sét</i>	<i>mà anh vừa thấy</i>
Đại từ	Số từ/ Phụ từ chỉ lượng	Quán từ	Danh từ đơn vị	Danh từ khối	Danh từ/ Vị từ, Số từ + danh từ	Vị từ, ngữ vị từ	Vị từ, ngữ vị từ	Vị từ/ Danh ngữ	Đại từ	Giới ngữ	Tiểu cú

Cách phân loại các thành tố phụ sau của Hoàng Dũng, Nguyễn Thị Ly Kha rất hữu ích cho các nghiên cứu ngôn ngữ học, nhưng với người nước ngoài học tiếng Việt, cách phân loại này khá phức tạp. Bởi vậy, nghiên cứu này sử dụng mô hình phân loại của Nguyễn Tài Cẩn để giải thích các lỗi sai trật tự trong danh ngữ.

Trật tự các thành tố trong danh ngữ là cố định, không linh hoạt như trong động ngữ. Các lỗi thường gặp là:

Lỗi vị trí -3, ví dụ: *Trong phòng ánh mắt tất cả đều nhìn về phía vị giáo sư mới đến.*

Lỗi vị trí -0, ví dụ: *Em có bút bi một chiếc.*

Lỗi vị trí +1, ví dụ: *Chị ấy có một quyển sách Trung Quốc lịch sử rất hay.*

Trật tự các thành tố trong ngữ vị từ:

Mô hình động ngữ trong tiếng Việt gồm các thành tố khái quát như sau:

Bảng 3. Các thành phần trong cụm ngữ vị từ

(trên cơ sở tham khảo Nguyễn Tài Cẩn, 1999; Nguyễn Lân Trung, 2011)

Phần phụ trước					Phần trung tâm		Phần phụ sau	
Phụ tố trước					Bổ tố trước	V tình thái	V miêu tả	Phụ tố sau
Tiếp diễn tương tự	Quan hệ thời gian	Phủ định	Tần suất	Mức độ		+	Ø	
Phụ tố cầu khiến								Danh ngữ
								Cú
						+ / Ø	+	Chế định (Bổ tố bắt buộc)
								Phi chế định (Bổ tố tự do)

Trong đó, phần trung tâm có thể có 1 vị từ (ví dụ: *Cô ấy sẽ đi ngay bây giờ/ Chị ấy rất xinh*) hoặc 2 vị từ (ví dụ: *Tôi muốn ăn phở/ Chị ấy béo ra*). Khi có 2 vị từ, người ta gọi là chuỗi vị từ (serial verb) với vai trò trung tâm thường được cho là từ thứ nhất.

Phần phụ trước gồm các nhóm:

(1) Nhóm chỉ sự tiếp diễn tương tự: *đều, cũng, vẫn, cứ,...*: thường đứng trước các nhóm khác; có khả năng kết hợp nội bộ (những kết hợp được chấp nhận: *cũng đều, đều vẫn, cũng cứ, cũng vẫn, vẫn cứ,...*, nhưng kết hợp không được chấp nhận: *cứ vẫn, vẫn cũng, cứ cũng,...*).

(2) Nhóm chỉ quan hệ thời gian: *từng, đã, sẽ, đang, mới, vừa,...*:

Vị trí sau nhóm (1): Ít có khả năng kết hợp nội bộ (những kết hợp được chấp nhận: *vừa mới, mới vừa, đã từng - từng đã*).

(3) Nhóm phủ định: *không, chưa, chẳng*; vị trí sau (1) và (2); không có khả năng kết hợp nội bộ.

(4) Nhóm chỉ tần suất: *hay, thường, luôn, hiếm khi, ít khi,...*; vị trí sau (1) và (2); ít có khả năng kết hợp nội bộ (kết hợp được chấp nhận: *thường hay, thường ít khi*, ví dụ: *Anh ấy thường hay đi chơi/ Anh ấy thường ít khi đi chơi*).

(5) Nhóm chỉ cầu khiến: *hãy, đừng, chớ*; vị trí sau (1); có khả năng kết hợp nội bộ (*hãy đừng*).

(6) Nhóm chỉ mức độ: *rất, hơi, khi, quá*; đứng sau (1) và (2); không có khả năng kết hợp nội bộ.

Ví dụ các kết hợp sau (kết hợp giữa các nhóm) có thể chấp nhận:

(+) *Tất cả sinh viên ai cũng đều đã từng gặp khó khăn trong học tập.*

(+) *Tôi muốn nhắc các em sinh viên ở đây, nam hay nữ, ai cũng không nên đi làm thêm ban đêm.*

(+) *Em nói mãi rồi mà anh ấy cũng vẫn cứ đang chưa đồng ý.*

Một số kết hợp có thể đảo vị trí trong ngữ vị từ: Ví dụ: - *ung dung bước/ bước ung dung, rả rích rơi/ rơi rả rích, buồn bã khóc / khóc buồn bã, chăm chỉ học tập/ học tập chăm chỉ,...*

Các lỗi trật tự ngữ vị từ thường gặp là (thường ở 4 nhóm đầu tiên):

- Lỗi trật tự nhóm (1): *Anh ấy đang vẫn làm.*; - Lỗi trật tự nhóm (2): *Chị ấy vừa đã đến.*
 - Lỗi trật tự nhóm (3): *Chị ấy không đang làm việc*; - Lỗi trật tự nhóm (4): *Em bé luôn cứ cãi mẹ.*
- Lỗi trật tự phần phụ sau của ngữ vị từ thường gặp là:
- Lỗi trật tự nhóm từ chỉ khả năng: *Vì kĩ túc xá ồn ào quá, em không được đọc sách.*
 - Lỗi trật tự nhóm từ chỉ sự phối hợp hoạt động: *Trong kì nghỉ lễ Tết, tôi đã với chị Huệ đi Nha Trang và Đà Nẵng du lịch.*
 - Lỗi trật tự nhóm từ chỉ tiếp diễn của hành động: *Chúng em tiếp học tiếng Việt.*
 - Lỗi trật tự nhóm từ chỉ kế tiếp nhận hành động: *Cô quản lí kĩ túc xá tốt bụng cho chúng em chuẩn bị nhiều thứ.*
 - Lỗi trật tự nhóm từ chỉ nơi chốn và mục đích: *Năm ngoái, tôi đã từng đến làm việc công ty này rồi.*

2.2. Ứng dụng Chatbot trong sửa lỗi ngữ pháp tiếng Việt cho người nước ngoài

2.2.1. Khái quát về Chatbot

Chatbot là từ viết tắt của *chatting robot*, trong tiếng Việt có thể dịch là trợ lí ảo hay ứng dụng đối thoại tự động. Trong bài nghiên cứu này, chúng tôi giữ nguyên cách gọi Chatbot trong tiếng Anh để thống nhất về mặt thuật ngữ. Chatbot là các hệ thống phần mềm được thiết kế để mô phỏng hội thoại với con người thông qua văn bản hoặc giọng nói, dựa trên các công nghệ xử lí ngôn ngữ tự nhiên (NLP) và trí tuệ nhân tạo (AI). Trong lĩnh vực giáo dục ngôn ngữ, Chatbot đóng vai trò quan trọng trong việc hỗ trợ học tập, cung cấp phản hồi tức thời và cá nhân hóa trải nghiệm người dùng. Dựa trên cơ chế hoạt động và mức độ phức tạp, Chatbot có thể được phân loại thành ba nhóm chính như sau:

Nhóm 1. Chatbot dựa trên luật (Rule-based Chatbots)

Loại Chatbot này hoạt động trên các quy tắc được lập trình sẵn, sử dụng các mẫu câu (patterns) hoặc từ khóa để phản hồi. Hệ thống sẽ so khớp đầu vào của người dùng với các quy tắc được định nghĩa trước để tạo câu trả lời tương ứng. Loại Chatbot này có ưu điểm: 1/Dễ phát triển và triển khai, không yêu cầu dữ liệu huấn luyện lớn; 2/Phù hợp cho các kịch bản đơn giản, như trả lời câu hỏi cố định hoặc hướng dẫn cơ bản. Nhưng bên cạnh đó cũng tồn tại các hạn chế như: 1/Thiếu linh hoạt, không xử lí được các câu hỏi ngoài phạm vi quy tắc; 2/Khả năng tương tác tự nhiên hạn chế, đặc biệt với các ngôn ngữ phức tạp như tiếng Việt. Khi ứng dụng trong giáo dục, loại Chatbot này làm tốt việc hỗ trợ trả lời câu hỏi trắc nghiệm hoặc cung cấp từ vựng cơ bản.

Nhóm 2. Chatbot dựa trên trí tuệ nhân tạo (AI-based Chatbots)

Đây là loại Chatbot sử dụng các mô hình học máy (machine learning) hoặc học sâu (deep learning) để hiểu và phản hồi dựa trên ngữ cảnh. Các mô hình như BERT, GPT hoặc PhoBERT (dành riêng cho tiếng Việt) được huấn luyện trên tập dữ liệu lớn để xử lí ngôn ngữ tự nhiên. Ưu điểm của Chatbot này là: 1/Khả năng hiểu ngữ cảnh và phản hồi tự nhiên, linh hoạt; 2/Có thể học hỏi và cải thiện qua thời gian nếu được huấn luyện thêm; 3/Phù hợp với các tác vụ phức tạp như phát hiện lỗi ngữ pháp hoặc giải thích ngữ pháp theo ngữ cảnh. Tuy nhiên, hạn chế của AI-based Chatbot là yêu cầu tài nguyên tính toán lớn và dữ liệu huấn luyện chất lượng cao, cũng như quá trình phát triển phức tạp, đòi hỏi chuyên môn sâu về AI và NLP.

Chatbot trí tuệ nhân tạo có thể ứng dụng tốt trong việc hỗ trợ sửa lỗi ngữ pháp, giải đáp thắc mắc ngữ pháp, hoặc cung cấp bài tập tương tác.

Nhóm 3. Chatbot kết hợp (hybrid Chatbots)

Chatbot kết hợp giữa quy tắc và AI, sử dụng quy tắc cho các kịch bản đơn giản và AI cho các tình huống phức tạp. Hệ thống này tận dụng ưu điểm của cả hai loại Chatbot để tăng hiệu quả. Hệ thống kết hợp sẽ giúp cân bằng giữa tính đơn giản và khả năng xử lí ngữ cảnh phức tạp, cũng như dễ dàng tùy chỉnh cho các mục đích cụ thể, như giáo dục ngôn ngữ. Tuy nhiên, hạn chế khi sử dụng Chatbot kết hợp là cần thiết kế cẩn thận để đảm bảo sự chuyển đổi mượt mà giữa quy tắc và AI, chi phí phát

triển có thể cao hơn Chatbot dựa trên luật. Trong giáo dục, loại Chatbot này có thể hỗ trợ giảng dạy kết hợp giữa bài tập cố định và phản hồi cá nhân hóa.

Các nền tảng phát triển Chatbot: Để xây dựng Chatbot, có thể sử dụng các nền tảng như Dialogflow, Rasa, Microsoft Bot Framework hoặc các mô hình ngôn ngữ lớn như GPT của OpenAI. Mỗi nền tảng có đặc điểm riêng:

Dialogflow: Phù hợp cho Chatbot dựa trên luật và kết hợp, dễ tích hợp với các ứng dụng.

Ras: Mở mã nguồn, linh hoạt, phù hợp cho các dự án yêu cầu tùy chỉnh cao.

Custom GPTs (OpenAI): Dựa trên mô hình GPT, cho phép tạo Chatbot AI nhanh chóng với giao diện thân thiện, không yêu cầu lập trình phức tạp.

Mô hình chuyên biệt như PhoBERT: Tối ưu cho tiếng Việt nhưng đòi hỏi kỹ năng lập trình và tài nguyên tính toán lớn.

2.2.2. Lựa chọn Chatbot hỗ trợ nhận diện và sửa lỗi ngữ pháp cho người nước ngoài học tiếng Việt

Để có cơ sở lựa chọn Chatbot cho mục tiêu nghiên cứu, chúng tôi tiến hành so sánh năng lực phát hiện lỗi ngữ pháp của một số ứng dụng AI thông dụng bằng cách đặt các câu hỏi giống nhau cho các ứng dụng AI để nhận kết quả trả lời. Chúng tôi đã đặt ra 20 câu hỏi (tương ứng với 20 câu chứa lỗi ngữ pháp), nhận về 100 câu trả lời (của 5 ứng dụng AI), từ đó đánh giá sơ bộ khả năng nhận diện, phân tích và sửa lỗi của từng AI. Các câu được đưa ra phân tích gồm:

1. Bàn gỗ cái kia là của ông nội làm cho tôi lâu rồi.
2. Chị ấy làm được những bài tập rất khó tất cả nên được cao điểm.
3. Em thường ít đi phương tiện công cộng giao thông.
4. Em thường thích thêm gia vị một chút.
5. Tất cả này hai trường đều có sinh viên học tiếng Việt.
6. Mỗi ngày chúng em khoảng có bốn đến năm tiết học.
7. Chị ấy không đang ăn sáng (trả lời cho câu hỏi: Chị ấy có đang ăn sáng không?)
8. Em đi chơi thường buổi tối.
9. Em cố gắng ôn tập để được thi điểm cao.
10. Cuộc hành trình nữa bắt đầu lại.
11. Chúng em một chút hiểu hôm qua về bài cô dạy.
12. Tôi từng đã với gia đình năm ngoái đi du lịch Thái Lan.
13. Em học bây giờ ở Hà Nội thành phố.
14. Em sẽ được điểm cao chắc chắn trong kì thi này.
15. Làm ơn chị chỉ đường đến trường Nhân văn cho tôi.
16. Có một dịp tốt đi chơi em với anh ấy.
17. Cuối tháng này bạn ấy sẽ thi hình như.
18. Cô giáo cho chúng em làm buổi sáng nay bài kiểm tra.
19. Bắt đầu Tết của Lào từ ngày 14/4.
20. Anh ấy đều hàng tuần lên thư viện mượn sách.

Các ứng dụng AI được chọn là các ứng dụng đang được dùng rất phổ biến hiện nay, gồm: Chatgpt, Gemini, Grok, Claude, Deepseek.

Dưới đây là 2 ví dụ phản hồi của các ứng dụng:

Câu 1: Câu này có lỗi không? Nếu có, hãy sửa cho tôi: “Bàn gỗ cái kia là của ông nội làm cho tôi lâu rồi.”.

Bảng 4. Phản hồi của các ứng dụng AI cho câu hỏi 1

	Nhận diện lỗi	Giải thích	Đề xuất sửa
Chatgpt	Câu sai về trật tự từ trong danh ngữ và cách diễn đạt chưa tự nhiên	Trong tiếng Việt, trật tự từ trong danh ngữ thường là: [Số lượng] + [Định ngữ] + Danh từ + (Từ chỉ định).	“Cái bàn gỗ kia là ông nội làm cho tôi từ lâu rồi.” (Tự nhiên hơn và đúng)

		"Bàn gỗ cái kia" là một tổ hợp sai thứ tự. Ngoài ra, phần "là của ông nội làm cho tôi lâu rồi" chưa thật mạch lạc. Cấu trúc "là của A làm" không thật sự tự nhiên.	cú pháp.) Hoặc: <i>"Cái bàn gỗ kia là do ông nội làm cho tôi từ lâu rồi."</i> (Trang trọng, rõ ràng hơn.)
Gemini	Câu không có lỗi ngữ pháp hay chính tả nào. Chúng rất tự nhiên và thể hiện cảm xúc tốt.	- "Bàn gỗ cái kia" thành "Chiếc bàn gỗ kia": Từ "chiếc" nghe tự nhiên và thanh thoát hơn khi đi kèm với đồ vật. "Cái" cũng đúng nhưng ít trang trọng hơn. - "làm cho tôi lâu rồi" thành "làm cho tôi đã lâu rồi": Thêm "đã" nhấn mạnh khoảng thời gian từ lúc sự việc diễn ra, tạo cảm giác về kỉ niệm xa xưa.	<i>"Chiếc bàn gỗ kia là do ông nội làm cho tôi đã lâu rồi."</i>
Grok	Câu không có lỗi ngữ pháp nghiêm trọng, nhưng có thể cải thiện để câu văn mượt mà và tự nhiên hơn	- Cụm "của ông nội làm cho tôi" hơi lủng củng, dễ gây hiểu nhầm. Cách dùng "của" ở đây không sai, nhưng không phải cách diễn đạt tự nhiên nhất trong tiếng Việt. - "Cái kia" cũng hơi thừa, vì trong văn viết hoặc ngữ cảnh rõ ràng, người đọc thường hiểu được bạn đang nói về cái bàn nào. - "Lâu rồi" ở cuối câu ổn, nhưng có thể thay bằng cụm khác để tăng tính mạch lạc."	Kết quả sau khi sửa (gợi ý): <i>"Cái bàn gỗ kia do ông nội làm cho tôi từ lâu."</i> Hoặc <i>"Cái bàn gỗ ấy là ông nội làm cho tôi từ lâu."</i> Nếu bạn muốn giữ giọng điệu thân mật, có thể dùng: <i>"Cái bàn gỗ ấy ông nội làm cho tôi từ lâu rồi. Khi nhìn thấy nó, tôi nhớ ông lắm."</i>
Claude	Câu có một chút vấn đề về trật tự từ và cách diễn đạt.	- "Bàn gỗ cái kia" → "Cái bàn gỗ kia" (trật tự từ tự nhiên hơn) - "là của ông nội làm" → "là do ông nội làm" hoặc bỏ "là của" (tránh dư thừa) - "lâu rồi" → "từ lâu rồi" (hoàn chỉnh hơn về mặt ngữ pháp)	Câu đã sửa: <i>"Cái bàn gỗ kia là do ông nội làm cho tôi từ lâu rồi."</i> Hoặc có thể sửa thành: <i>"Cái bàn gỗ kia ông nội làm cho tôi từ lâu rồi."</i>
Deepseek	(Lần 1: chỉ nêu cách sửa, không nêu lỗi) (Lần 2 sau khi thêm câu hỏi "Câu trên đúng hay sai ngữ pháp): Câu không sai về ngữ pháp hình thức, nhưng	- "bàn gỗ cái kia" không phải cách nói tự nhiên trong tiếng Việt. Thường người ta sẽ nói "cái bàn gỗ kia" hoặc "bàn gỗ kia". - Cụm "làm cho tôi lâu rồi" nghe hơi lủng củng, không rõ ràng bằng các cách diễn đạt khác như "đã lâu rồi" hoặc "làm cho tôi từ lâu rồi".	<i>"Cái bàn gỗ kia là của ông nội làm cho tôi từ lâu rồi."</i> Hoặc một cách khác: <i>"Cái bàn gỗ kia là do ông nội làm cho tôi đã lâu rồi."</i>

	về mặt cú pháp và phong cách, nó không tự nhiên và có thể gây khó hiểu.	
--	---	--

Câu 15: Câu này có sai không: “Làm ơn chị chỉ đường đến trường Nhân văn cho tôi.”.

Bảng 5. Phản hồi của các ứng dụng AI cho câu hỏi 15

	Nhận diện lỗi	Giải thích	Đề xuất sửa
Chatgpt	Đây là một câu nói phổ biến trong văn nói đời thường, người Việt vẫn thường dùng theo cách này. Tuy nhiên, nếu xét theo ngữ pháp chuẩn và văn phong mạch lạc, thì câu này chưa trọn vẹn và hơi lủng củng về cấu trúc.	- "Làm ơn chị chỉ...": Đây là hai thành phần rời nhau ("làm ơn" và "chị chỉ...") không được kết nối mạch lạc. Trong văn nói, người ta bỏ qua tiểu tiết này, nhưng trong văn viết hay giao tiếp lịch sự, nên có liên kết rõ hơn. - "cho tôi" ở cuối câu là thừa - vì việc “chỉ đường” vốn đã hướng đến người hỏi. Dùng “giúp tôi” thì nghe lịch sự hơn.	Phiên bản 1 (tự nhiên, đúng văn nói lịch sự): “Chị làm ơn chỉ đường đến trường Nhân văn giúp tôi với.” Phiên bản 2 (ngắn gọn, trang trọng hơn): “Làm ơn chỉ giúp tôi đường đến trường Nhân văn.” Phiên bản 3 (cực kì lịch sự, dùng trong tình huống trang trọng): “Phiền chị chỉ giúp tôi đường đến trường Nhân văn được không ạ?”
Gemini	Câu không sai ngữ pháp	Việc dùng "chị" ở đây nghe hơi cụt và giống như một lời yêu cầu trực tiếp quá, thiếu đi sự trang trọng hoặc mềm mại cần thiết khi nhờ vả người lạ. Trong giao tiếp tiếng Việt, chúng ta thường thêm các từ xưng hô vào đầu hoặc cuối câu để thể hiện sự lịch sự. - "Làm ơn" đã thể hiện sự lịch sự rồi, nhưng kết hợp với cách dùng từ xưng hô chưa thật tự nhiên khiến câu nghe chưa được trọn vẹn.	- “Chị ơi, làm ơn chỉ đường đến trường Nhân văn giúp em/ tôi với ạ.”. - “Làm ơn chỉ đường đến trường Nhân văn giúp tôi/em với ạ.”. - “Xin lỗi chị, chị có thể chỉ giúp em/tôi đường đến trường Nhân văn không ạ?”.
Grok	Câu không hoàn toàn sai nhưng có vấn đề nhỏ về ngữ pháp và ngữ cảnh	- Từ "chị" trong câu này có thể không phù hợp nếu người được hỏi không phải là nữ hoặc không rõ giới tính, tuổi tác. Trong tiếng Việt, việc sử dụng đại từ nhân xưng như "chị" cần dựa trên ngữ cảnh và mối quan hệ. Nếu bạn đang nói với một người lạ, đặc biệt khi không rõ giới tính hoặc tuổi, dùng "chị" có thể gây thiếu tự nhiên hoặc không lịch sự. - Cụm "làm ơn chị chỉ đường" có	- "Làm ơn chỉ đường đến trường Nhân văn cho tôi." - “Xin vui lòng chỉ đường đến trường Nhân văn giúp tôi.”. - "Chị ơi, làm ơn chỉ đường đến trường Nhân văn cho tôi.”. Hoặc: "Chị làm ơn chỉ

		trật tự từ hơi ngược so với cách nói thông thường. Trong tiếng Việt, "làm ơn" thường đứng đầu câu hoặc sau động từ để diễn đạt sự lịch sự. - "Trường Nhân văn" có thể cần bổ sung thông tin nếu bạn muốn cụ thể hơn (ví dụ: Trường Đại học Khoa học Xã hội và Nhân văn), tùy thuộc vào ngữ cảnh.	<i>giúp tôi đường đến trường Nhân văn."</i>
Claude	Câu này không sai về ngữ pháp nhưng có thể cải thiện để tự nhiên hơn:	- Thêm "với" ở cuối câu sẽ làm câu lịch sự và tự nhiên hơn trong giao tiếp hàng ngày - Hoặc có thể đổi vị trí "cho tôi" để câu văn trôi chảy hơn	Các cách sửa tự nhiên hơn: - " <i>Làm ơn chị chỉ đường đến trường Nhân văn cho tôi với.</i> " - " <i>Chị làm ơn chỉ đường đến trường Nhân văn cho tôi với.</i> " - " <i>Làm ơn chị chỉ cho tôi đường đến trường Nhân văn.</i> "
Deepseek	Câu này đúng về cấu trúc và ngữ pháp, nhưng có thể chỉnh sửa để nghe tự nhiên và lịch sự hơn một chút.	X	- Thêm từ "xin" để thể hiện sự lịch sự: " <i>Làm ơn, chị xin chỉ đường đến trường Nhân văn cho tôi.</i> " Hoặc đơn giản hơn: " <i>Làm ơn, chị chỉ đường đến trường Nhân văn giúp tôi.</i> "

Để đánh giá kết quả trả lời của các ứng dụng, chúng tôi dựa trên 3 nội dung trả lời: nhận diện lỗi, giải thích lỗi và sửa lỗi với các tiêu chí như sau:

- *Nhận diện lỗi* được coi là đúng khi ứng dụng gọi được đúng tên lỗi (như: lỗi sai trật tự từ/ sắp xếp không đúng thành phần/ sai ngữ pháp); không đúng khi chỉ nói chung chung như: câu chưa tự nhiên, chưa mượt mà,... (vì "chưa tự nhiên" hay "chưa mượt mà" vẫn có nghĩa là "đúng nhưng chưa hay, ít phổ biến"). Ví dụ trong 2 bảng phản hồi nêu trên, chỉ có Chatgpt và Claude nhận diện đúng lỗi câu 1, và không ứng dụng nào nhận diện đúng lỗi của câu 15.

- *Giải thích* được coi là đúng khi phân loại đúng lỗi, phân tích đúng cơ chế sắp xếp trong câu tiếng Việt. Trong 2 bảng phản hồi nêu trên, chỉ có Chatgpt giải thích đúng lỗi sai trật tự cụm danh từ ở câu 1, còn lại các giải thích khác đều không đúng.

- *Sửa lỗi* được coi là đúng khi câu đề xuất sửa được người bản ngữ chấp nhận. Trong 2 bảng phản hồi nêu trên, tất cả các câu ứng dụng Ai đề xuất đều được chấp nhận trong văn phong tiếng Việt.

Kết quả trả lời của các ứng dụng được tổng hợp như bảng dưới đây.

Bảng 6. Tổng hợp kết quả trả lời đúng của các ứng dụng

Nội dung	Nhận diện lỗi	Giải thích lỗi	Sửa lỗi
----------	---------------	----------------	---------

Ứng dụng			
Chatgpt	75%	45%	100%
Gemini	65%	35%	100%
Claude	70%	35%	100%
Grok	40%	20%	100%
Deepseek	70%	40%	100%

Dựa vào bảng kết quả trả lời, có thể đánh giá sơ bộ như sau:

1/Nhận diện lỗi: Chatgpt đang tỏ ra hiệu quả hơn cả với 75% nhận diện đúng lỗi, tiếp theo là Claude và Deepseek với 70%, Gemini 65% lần. Grok, một ứng dụng được đánh giá là có khả năng phân tích và viết nghiên cứu rất tốt, lại có vẻ gặp khó khăn nhất khi số lần nhận diện đúng chỉ là 40%. Câu khó nhận diện lỗi nhất (100% ứng dụng không nhận diện đúng lỗi) là câu 7 (*Chị ấy không đang ăn sáng*), câu 10 (*Một cuộc hành trình nữa bắt đầu lại*), và 15 (*Làm ơn chị cho tôi hỏi đường đến trường Nhân văn*). Đây đều là các câu có lỗi trật tự từ liên quan đến ngữ vị từ (cụ thể là động ngữ).

2/Giải thích: Trừ Deepseek trong lần đầu hỏi không đưa ra lời giải thích, các ứng dụng khác đều đưa ra các giải thích chấp nhận được đối với lỗi trong câu (nếu nhận diện đúng). Tuy nhiên, không phải giải thích nào cũng cụ thể và phân loại đúng lỗi. Ví dụ trong câu hỏi 1, chỉ Chatgpt đưa ra được giải thích đúng về danh ngữ, còn lại đều chỉ chung chung “nên thêm vào cho tự nhiên hơn/ mượt mà hơn”. Vì vậy trong phần giải thích lỗi, mức độ trả lời đúng của các ứng dụng khá thấp, lần lượt là Chatgpt (45%), Deepseek (40%), Gemini và Claude (35%) và cuối cùng vẫn là Grok (20%).

3/Đề xuất sửa: Đây là phần được các ứng dụng làm tốt nhất. Tất cả các ứng dụng đều đề xuất các cách sửa đúng và rất tự nhiên với người bản ngữ, đạt 100%. Ngay cả Grok dù chỉ đạt trọng số trả lời đúng rất thấp ở phần nhận diện lỗi và giải thích lỗi thì đến phần sửa lỗi cũng vẫn sửa đúng toàn bộ. Điều này chứng tỏ AI tạo sinh đang rất mạnh trong việc tạo ra các sản phẩm ngôn ngữ tự nhiên và giống ngôn ngữ bản xứ. Các ứng dụng AI không hề gặp bất cứ khó khăn gì trong việc sản sinh câu. Vấn đề của các ứng dụng AI chỉ là ở phần nhận diện và phân tích, giải thích lỗi.

Trong 20 câu hỏi, chúng tôi thiết kế 5 câu hỏi về lỗi ngữ danh từ, 5 câu hỏi lỗi ngữ vị từ, 5 câu hỏi lỗi thành phần câu, 5 câu hỏi kết hợp các loại lỗi (lỗi thành phần câu và lỗi đoạn ngữ). Các ứng dụng xử lý lỗi thành phần câu tốt nhất, sau đó đến lỗi danh ngữ, và gặp khó khăn nhất với lỗi động ngữ (đặc biệt các ứng dụng gặp khó khăn với các lỗi đảo trật tự phụ từ trong động ngữ). Đây là một gợi ý để chúng tôi chọn lọc kiến thức đào tạo Chatbot kĩ, sâu, để Chatbot hiểu được nguyên tắc vận hành của tiếng Việt.

Từ kết quả thử nghiệm trên, chúng tôi thấy rằng, chưa có ứng dụng trí tuệ nhân tạo nào đồng thời làm tốt được cả 3 việc: 1/Nhận diện lỗi; 2/Phân loại lỗi (giải thích lỗi) và 3/Đề xuất cách sửa. Trong đó, các ứng dụng vẫn gặp khó khăn trong việc hiểu sâu về các kiến thức tiếng Việt để phân tích và giải thích cho người dùng. Điều đó đặt ra yêu cầu cần phải tạo một Chatbot được đào tạo kĩ càng về các kiến thức tiếng Việt nói chung và các kĩ thuật nhận diện lỗi ngữ pháp, phân loại lỗi (giải thích lí do mắc lỗi) và sửa lỗi nói riêng.

Cũng dựa vào kết quả nghiên cứu, chúng tôi thấy rằng nền tảng Chatgpt dường như đang đem lại hiệu quả tốt nhất trong việc phân tích lỗi. Vì vậy chúng tôi quyết định chọn sử dụng chức năng Custom GPTs của OpenAI để phát triển Chatbot.

Quyết định này còn dựa trên các tiêu chí về sự đơn giản trong thao tác đối với những người dạy không chuyên về công nghệ, không có kiến thức lập trình chuyên sâu. Custom GPTs cung cấp công cụ tạo Chatbot thông qua giao diện kéo-thả và nhập văn bản, giúp giảm thiểu rào cản kĩ thuật. Người dạy, dù không có kinh nghiệm lập trình, vẫn có thể dễ dàng cấu hình Chatbot trong thời gian ngắn.

So với việc phát triển Chatbot từ đầu bằng các framework như Rasa hoặc huấn luyện mô hình như PhoBERT, Custom GPTs cho phép tạo nguyên mẫu (prototype) chỉ trong vài giờ. Điều này phù hợp với các dự án nghiên cứu cần thử nghiệm và đánh giá nhanh. Dựa trên mô hình GPT-4 hoặc GPT-4o, Custom GPTs có thể hiểu và phản hồi bằng tiếng Việt, đồng thời cung cấp giải thích bằng tiếng Anh, đáp ứng nhu cầu của người nước ngoài học tiếng Việt. Mặc dù không tối ưu cho tiếng Việt như các mô hình chuyên biệt, GPT vẫn đủ khả năng phát hiện các lỗi ngữ pháp cơ bản như sai trật tự từ, thiếu từ chức năng hoặc sai loại từ. Ngoài ra, Custom GPTs hỗ trợ tích hợp thêm dữ liệu hoặc điều chỉnh hướng dẫn khi đề tài phát triển, cho phép cải thiện Chatbot dựa trên phản hồi từ người dùng thực tế. Quan trọng hơn, Custom GPTs không yêu cầu cơ sở hạ tầng phức tạp vì hoạt động trên nền tảng đám mây của OpenAI, loại bỏ nhu cầu thiết lập máy chủ hoặc quản lý tài nguyên tính toán. Người dạy chỉ cần tài khoản ChatGPT Plus để bắt đầu.

Tuy nhiên, Custom GPTs cũng có những hạn chế nhất định như: Mô hình GPT không được huấn luyện chuyên sâu cho tiếng Việt, có thể gặp khó khăn với các lỗi ngữ pháp phức tạp hoặc ngữ cảnh đặc thù. Để khắc phục, cần cung cấp tập dữ liệu mẫu (corpus) chi tiết, bao gồm các câu sai, câu đúng và giải thích, để tăng cường khả năng phản hồi của Chatbot. Ngoài ra, Chatbot chỉ hoạt động trong hệ sinh thái của OpenAI và yêu cầu chi phí thuê bao (đối với người lập). Tuy nhiên, trong giai đoạn nghiên cứu, chi phí này được xem là hợp lý so với lợi ích mang lại.

2.2.3. Quy trình phát triển Chatbot với Custom GPTs

Quy trình phát triển Chatbot được tiến hành qua bốn giai đoạn chính: 1/Chuẩn bị; 2/Xây dựng Chatbot; 3/Thử nghiệm - tinh chỉnh; 4/Triển khai. Mỗi bước được thiết kế nhằm đảm bảo Chatbot có thể phát hiện và giải thích chính xác các lỗi ngữ pháp về trật tự từ của người học tiếng Việt như ngoại ngữ.

Giai đoạn 1: Chuẩn bị

Trước hết, nhóm nghiên cứu xác định rõ yêu cầu chức năng của Chatbot, bao gồm: 1/Khả năng phát hiện lỗi trật tự từ trong câu tiếng Việt; 2/Giải thích lỗi bằng ngôn ngữ dễ hiểu, ưu tiên tiếng Việt hoặc tiếng Anh; 3/Đề xuất cách sửa lỗi (câu đúng); 4/Cung cấp ví dụ minh họa hoặc bài tập ngữ pháp theo yêu cầu.

Song song với đó là quá trình thu thập và xây dựng nguồn dữ liệu, bao gồm khung lý thuyết về trật tự từ, bảng tập hợp lỗi phổ biến từ thực tế giảng dạy và các tài liệu nghiên cứu liên quan (bài báo, luận văn, luận án). Để sử dụng nền tảng Custom GPTs, nhóm nghiên cứu đăng ký tài khoản ChatGPT Plus có trả phí.

Giai đoạn 2: Xây dựng Chatbot

Sau khi truy cập công cụ GPT Builder trên nền tảng ChatGPT, nhóm tiến hành cấu hình Chatbot với tên gọi "VietGrammarBot". Trong phần mô tả và hướng dẫn (Instructions), Chatbot được định nghĩa là trợ lý học tập chuyên hỗ trợ người nước ngoài sửa lỗi ngữ pháp tiếng Việt, với nhiệm vụ chính là phát hiện lỗi trật tự từ, giải thích nguyên nhân, đề xuất phương án sửa, và khuyến khích học viên thông qua giao tiếp thân thiện. Hệ thống cũng được cấu hình ngôn ngữ đầu vào là ngôn ngữ của người dùng (câu lệnh) và tiếng Việt (câu cần kiểm tra), và đầu ra là tiếng Việt, tiếng Anh hoặc ngôn ngữ của người dùng tùy ngữ cảnh.

Phân khởi động hội thoại (conversation starters) được xây dựng dưới dạng các câu hỏi thường gặp của người học như: *Câu này có đúng ngữ pháp không: "Tôi không đang ăn"?*, *Bạn có thể sửa lỗi trong câu: "Cô ấy được vượt qua kì thi" không?*; *Câu "Ngày mai tôi đã đi học" có vấn đề gì không?*; *Giúp mình kiểm tra câu: "Những cái áo đẹp này tôi mới mua hôm qua."* Các câu khởi đầu được thiết kế ngắn gọn, rõ ràng, phù hợp với trình độ trung cấp và có thể dùng song ngữ để tạo thuận lợi cho người học nước ngoài.

Để tăng cường năng lực chuyên biệt, nhóm nghiên cứu đã tải lên vào mục "Knowledge" các tài liệu sau để huấn luyện và hỗ trợ Chatbot phản hồi chính xác các yêu cầu của người dùng:

- Khung lý thuyết về cú pháp tiếng Việt (chúng tôi sử dụng tập *Quy tắc cú pháp tiếng Việt* - Đề tài của nhóm tác giả Đào Minh Thu, Đào Thị Minh Ngọc, Nguyễn Mai Vân, Lê Kim Ngân, Lê Thanh

Hương, Nguyễn Phương Thái, Đỗ Bá Lâm). Đây là tập quy tắc cú pháp được thiết lập dựa trên các lý thuyết ngữ pháp hiện đại, được ứng dụng hiệu quả để gắn nhãn và phân tích câu tự động.

- Lý thuyết về trật tự từ tiếng Việt và các lỗi thường gặp của người nước ngoài (như phần cơ sở lý thuyết chúng tôi đã xác lập).

- Bảng tập hợp lỗi phổ biến từ thực tế giảng dạy (500 lỗi thường gặp và được tiếp tục cập nhật.).

- Các tài liệu nghiên cứu liên quan (các bài báo, luận văn, luận án đã công bố).

Giai đoạn 3: Thử nghiệm và tinh chỉnh

Trong giai đoạn này, các câu tiếng Việt đúng và sai được đưa vào để kiểm tra phản hồi của Chatbot. Có nhiều câu Chatbot xử lý tốt hơn hẳn so với trước khi được huấn luyện, ví dụ: khi được hỏi “*Câu này có lỗi không: “Em ấy không đang làm việc”*”, Chatgpt bình thường dù nhận diện được lỗi trật tự từ (*kết hợp không đúng*), nhưng giải thích sai (*phủ định của thì hiện tại tiếp diễn thường dùng “không + động từ + đang” hoặc “không đang + động từ”*).

ChatGPT ▾

Câu này có lỗi không
Em ấy không đang làm việc

Câu “Em ấy không đang làm việc” có lỗi về ngữ pháp trong tiếng Việt tự nhiên. Câu này kết hợp không đúng giữa trạng thái phủ định (“không”) và thì hiện tại tiếp diễn (“đang + động từ”).

Giải thích:

Trong tiếng Việt, phủ định của thì hiện tại tiếp diễn thường dùng “không + động từ + đang” hoặc “không đang + động từ”, nhưng cấu trúc “không đang” không tự nhiên và ít được dùng.

Cách nói đúng và tự nhiên hơn:

- “Em ấy không làm việc.” → đơn giản, chỉ trạng thái không làm việc.
- “Em ấy đang không làm việc.” → nhấn mạnh là ngay hiện tại thì em ấy không làm việc, nhưng câu này vẫn hơi gượng.
- “Em ấy không làm việc lúc này.” → rõ ràng, tự nhiên hơn.
- “Hiện tại, em ấy không làm.” ↓ → cũng là cách diễn đạt hợp lý.

Còn Chatbot VietGrammarBot đưa ra phản hồi đúng ở cả 3 nội dung (nhận diện, giải thích, sửa lỗi), với cách giải thích rõ ràng, ngắn gọn kèm ví dụ bổ sung:

VietGrammarBot 4.0 ▾

Đây là lỗi trật tự phụ từ trước động từ — một lỗi phổ biến ở người học tiếng Việt.

Trong cấu trúc động ngữ, nếu cả “đang” (thời gian) và “không” (phủ định) cùng xuất hiện, “đang” phải đứng trước để biểu thị mốc thời gian phủ định.

3. Câu sửa đúng:

- ✓ “Em ấy đang không làm việc.”

4. Ví dụ đúng để tham khảo:

- “Tôi đang không nghĩ gì cả.”
- “Họ đang không sử dụng máy tính.”

Những trường hợp VietGrammarBot phản hồi thiếu chính xác được sử dụng để tinh chỉnh hướng dẫn và bổ sung dữ liệu vào kho kiến thức của Chatbot. Quá trình này cho thấy các trường hợp sai trật tự thành phần câu và trật tự danh ngữ đã được nâng cấp đáng kể so với trước khi được huấn luyện. Tuy nhiên, động ngữ vẫn là một thách thức khi quá trình tinh chỉnh đôi khi phải làm đi làm lại để Chatbot nhận diện đúng được lỗi tiếp theo. Ví dụ: Để kiểm tra Chatbot về lỗi trật tự thành phần phụ sau của động ngữ, chúng tôi đưa vào câu “*Con người có ý chí quyết tâm thì mới được vượt qua khó khăn*”. Chatbot không phát hiện ra lỗi ở trật tự của từ “*được*” mà cho rằng câu này thiếu chủ ngữ (trước *được*). Vì vậy, chúng tôi thêm vào quy tắc dùng từ “*được*” trong phần hướng dẫn (instructions) của Chatbot. Sau 2 lần tinh chỉnh, chúng tôi đặt tiếp câu hỏi với câu “*Bên ngoài ồn ào làm tôi không được đọc sách*”, Chatbot đã trả lời đúng lỗi trật tự từ.

Giai đoạn 4: Triển khai và đánh giá

Sau khi hoàn tất kiểm thử, dự kiến là một khối lượng việc khổng lồ, Chatbot sẽ được chia sẻ rộng rãi trong cộng đồng người học tiếng Việt. Với người có chuyên môn kỹ thuật, Chatbot có thể được tích hợp vào website hoặc ứng dụng thông qua API. Nhóm nghiên cứu cũng chủ động thu thập phản hồi từ người dùng để tiếp tục cải thiện hiệu quả hoạt động của hệ thống. Ở thời điểm hiện tại, Chatbot vừa được sử dụng trong phạm vi nhóm nghiên cứu để tiếp tục quá trình thử nghiệm và tinh chỉnh, vừa được chia sẻ trong giới chuyên môn để nhận được góp ý của người dùng¹.

Để mở rộng tiềm năng ứng dụng, nhóm nghiên cứu cũng đề xuất tích hợp các bài tập tương tác như điền từ, sắp xếp câu; cá nhân hóa nội dung theo trình độ người học; và tạo cơ chế kết nối với giáo viên trong các trường hợp Chatbot không xử lý được yêu cầu quá phức tạp.

3. Kết luận

Nghiên cứu này đã trình bày một thử nghiệm bước đầu trong việc ứng dụng công nghệ trí tuệ nhân tạo - cụ thể là nền tảng ChatGPT với tính năng Custom GPTs - nhằm xây dựng Chatbot VietGrammarBot hỗ trợ người học tiếng Việt như một ngoại ngữ nhận diện và sửa lỗi ngữ pháp, tập trung vào hiện tượng sai trật tự từ. Với sự kết hợp giữa khung lý thuyết ngôn ngữ học chức năng và phương pháp tiếp cận thực hành, nhóm nghiên cứu đã bước đầu triển khai một quy trình thiết kế, thử nghiệm và tinh chỉnh Chatbot trên nền tảng dữ liệu thực tiễn và phản hồi từ người dùng mục tiêu.

Kết quả cho thấy Chatbot có tiềm năng trở thành công cụ hỗ trợ học tập ngôn ngữ hiệu quả, đặc biệt trong việc cá nhân hóa phản hồi và tạo điều kiện cho người học thực hành ngôn ngữ qua tương tác tự nhiên. Tuy nhiên, nghiên cứu cũng nhận thức rõ các giới hạn hiện tại như phạm vi xử lý lỗi còn hẹp, dữ liệu huấn luyện chưa đủ đa dạng, và chưa có đánh giá định lượng trên diện rộng. Đây là những vấn đề cần được tiếp tục mở rộng và khảo sát trong các nghiên cứu kế tiếp để có thể tạo nguồn dữ liệu đầu vào huấn luyện Chatbot được chuyên nghiệp hơn.

Về mặt lý luận, nghiên cứu đóng góp vào xu hướng tích hợp ngôn ngữ học với công nghệ AI trong giảng dạy tiếng Việt như ngoại ngữ, đồng thời gợi mở khả năng phát triển các công cụ học tập dựa trên hội thoại thông minh và ngữ liệu thực hành. Về mặt ứng dụng, sản phẩm Chatbot có thể được triển khai trong môi trường lớp học, nền tảng tự học (như Duolingo, Anki), hoặc tích hợp vào hệ thống LMS, góp phần nâng cao hiệu quả giảng dạy ngữ pháp tiếng Việt theo hướng giao tiếp và lấy người học làm trung tâm. Dự kiến trong thời gian tới, Chatbot sẽ được đưa vào hỗ trợ sinh viên học tiếng Việt tại Khoa Ngôn ngữ học, Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia Hà Nội.

Người dùng có thể truy cập chatgpt.com, vào mục Gpts, gõ VietGrammarBot để tìm và sử dụng thử Chatbot. Hiện nay Chatbot vẫn đang trong giai đoạn thử nghiệm.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Nguyễn Tài Cẩn (1999), *Ngữ pháp tiếng Việt*. Nxb Đại học Quốc gia Hà Nội.
2. Nguyễn Linh Chi (2009), *Lỗi ngôn ngữ của người nước ngoài học tiếng Việt (trên tư liệu lỗi từ vựng - ngữ pháp của người Anh, Mỹ)*. Luận án Tiến sĩ Ngôn ngữ học.

3. Nguyễn Đức Dân (2013), *Từ câu sai đến câu hay*. Nxb Trẻ.
4. Hoàng Dũng, Nguyễn Thị Ly Kha (2004), “Về các thành tố phụ sau trung tâm trong danh ngữ tiếng Việt”. Tạp chí *Ngôn ngữ* số 4. tr.24-34.
5. Cao Xuân Hạo (2007), *Tiếng Việt: Mấy vấn đề về ngữ âm, ngữ pháp, ngữ nghĩa*, Nxb Giáo dục.
6. Nguyễn Văn Hiệp (2009), *Cú pháp tiếng Việt*. Nxb Giáo dục.
7. Vũ Hoàng Phương Loan (2015), *Khảo sát lỗi ngữ pháp trong sử dụng tiếng Việt của sinh viên Trung Quốc tại một số trường Đại học ở Hà Nội*, Luận văn thạc sĩ Ngôn ngữ học, Trường ĐHKHXHNV Hà Nội.
8. Bùi Thu Loan (2015), *Khảo sát lỗi sử dụng từ Hán Việt của sinh viên Trung Quốc học tiếng Việt (trên cơ sở các bài viết của sinh viên Trung Quốc ở Việt Nam)*, Luận văn thạc sĩ Ngôn ngữ học, Trường ĐHKHXHNV Hà Nội.
9. Đặng Lê Thủy Tiên (2018), *Khảo sát lỗi trật tự từ trong câu tiếng Việt của sinh viên Lào học tiếng Việt (tại Trường Đại học Quảng Bình)*. Luận văn thạc sĩ Ngôn ngữ học, Trường ĐHKHXHNV Hà Nội.
10. Đào Minh Thu, Đào Thị Minh Ngọc, Nguyễn Mai Vân, Lê Kim Ngân, Lê Thanh Hương, Nguyễn Phương Thái, Đỗ Bá Lâm (2010), *Tập quy tắc cú pháp tiếng Việt*. Đề tài KC.01.01.05/06-10. https://www.jaist.ac.jp/~bao/VLSP-text/Mar2009/SP85_baocaokythuat2009thang3.pdf.
11. Nguyễn Lâm Trung (2011), “Về động ngữ tiếng Việt”. Tạp chí *Khoa học ĐHQGHN, Ngoại ngữ* (27). tr. 225-231.

Tiếng Anh

12. Dik, S. C. (1997), *The Theory of Functional Grammar* (Vol. 1–2). Mouton de Gruyter.
13. Evans, V., & Green, M. (2006,) *Cognitive Linguistics: An Introduction*. Lawrence Erlbaum Associates.
14. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014), *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge.
15. Langacker, R. W. (2008), *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
16. Saussure, F. de. (1983), *Course in General Linguistics* (Trans. Roy Harris). Duckworth. (Original work published 1916).

A pilot study on developing a grammar-correction Chatbot for Vietnamese as a foreign language learners

Abstract: This paper presents a pilot study on the design of a Chatbot using OpenAI's Custom GPTs platform to detect and explain grammatical errors, with a focus on word order errors in Vietnamese-a syntactic phenomenon that is both common and complex. Grounded in theoretical frameworks on Vietnamese word order and common learner errors, as well as current advances in generative artificial intelligence, the study evaluates the performance of existing AI models in identifying and explaining Vietnamese grammatical issues. As part of the broader field of AI-assisted language learning, the research proposes a structured development process for a grammar-focused Chatbot leveraging the Custom GPTs feature. Findings suggest that while the Chatbot shows strong potential as a language learning support tool-particularly in error correction-its ability to handle complex syntactic structures still requires improvement through more targeted training. The paper concludes with recommendations for further development to optimize the Chatbot for educational use in Vietnamese language instruction.

Key words: Vietnamese grammar; word order errors; AI; Custom GPTs; grammar correction Chatbot.