

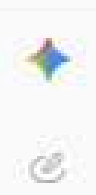


Đánh giá mức độ ổn định của các mô hình ngôn ngữ lớn với các phương ngữ Tiếng Việt

Group 3

Members: Quang Lý, Nhật Anh, Thanh Tú, Ngọc Ánh, Hà Phương

Mentors: Đức Hoàng, Quang Minh, Hải Nam, Thành Luân, Tổ Trinh, Cao Cường



Làm quen với Gemini, trợ lý AI cá nhân của bạn

+

Hỏi Gemini

Flash-Lite ▾

🎤



Đánh giá mức độ ổn định của các mô hình ngôn ngữ lớn với các phương ngữ Tiếng Việt

Group 3

Members: Quang Lý, Nhật Anh, Thanh Tú, Ngọc Ánh, Hà Phương

Mentors: Đức Hoàng, Quang Minh, Hải Nam, Thành Luân, Tổ Trinh, Cao Cường





1. Introduction

2. Benchmarks

3. Result analysis

4. Mitigation

Group 3



1. Introduction



Group 3



1. Introduction

Motivation

Dataset

Example

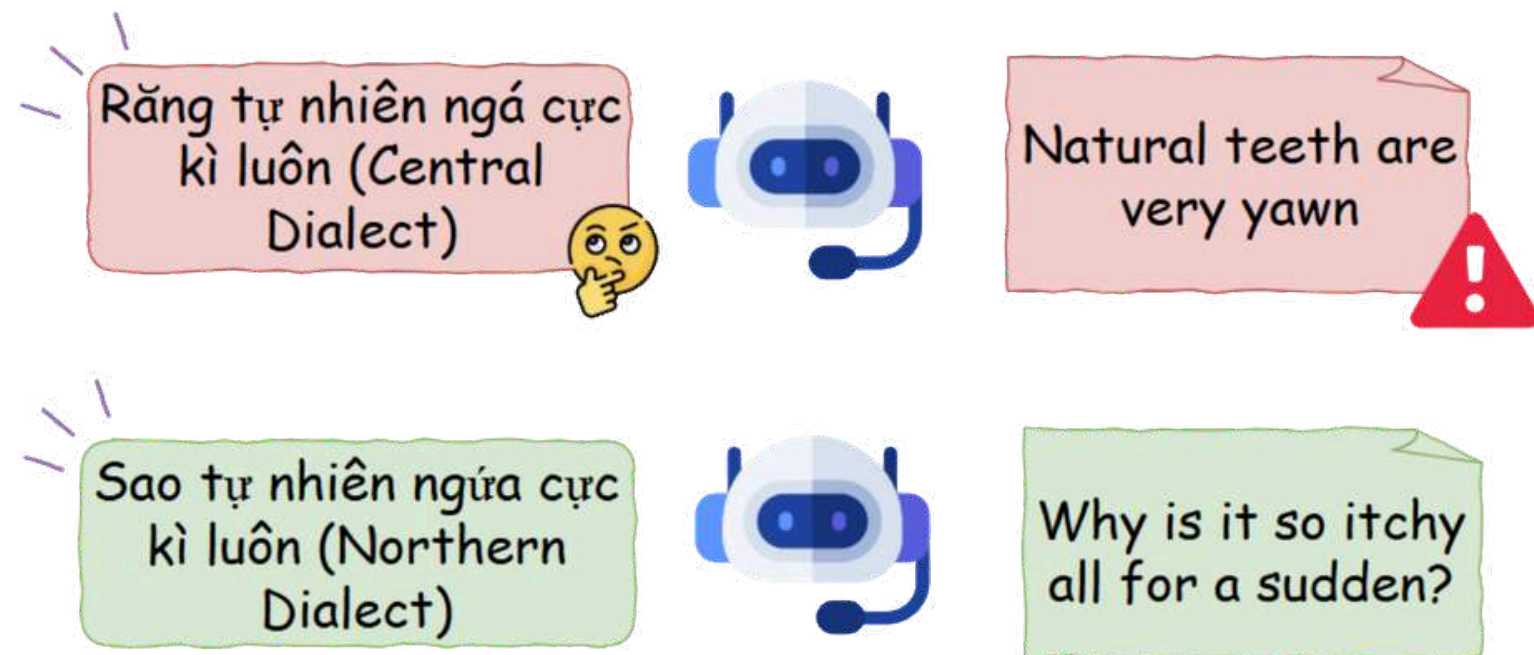
Vocab

Group 3

Motivation



MÔ HÌNH NGÔN NGỮ GẶP KHÓ KHĂN VỚI PHƯƠNG NGỮ VIỆT NAM



- ✓ Cùng ý nghĩa
- ✓ Cùng bài toán
- ✗ Khác biệt về phương ngữ
- ✗ Khác biệt về output mô hình



Trong ví dụ này, câu nói phương ngữ **miền Trung khác** với biến thể phương ngữ **miền Bắc** ở hai từ "**răng**" và "**sao**", "**ngá**" và "**ngứa**".

Mục tiêu nghiên cứu:

"Đánh giá tính bền vững của các mô hình ngôn ngữ lớn với các biến thể phương ngữ"



1. Introduction

Motivation

Dataset

Example

Vocab

Group 3



Dataset

Tên Gọi

Phạm vi địa lý

Đặc điểm chính

Nếu muốn làm model



Em phải hiểu dataset

Group 3

Mã

Tên Gọi

Phạm vi địa lý

Đặc điểm chính

facebook.com/seas.cvn

seas.cvn@gmail.com

<https://seas-cvn.com/>

Dataset



Mã	Tên Gọi	Phạm vi địa lý	Đặc điểm chính
PNB	Phương ngữ Bắc	Các tỉnh miền Bắc	Chuẩn thanh điệu, phân biệt rõ các dấu hỏi/ngã.
PNT1	Phương ngữ Trung 1	Thanh Hóa	Khu vực chuyển tiếp, lai giữa âm Bắc và Trung.
PNT2	Phương ngữ Trung 2	Hà Tĩnh, Nghệ An	Âm vực nặng, hệ thống từ vựng địa phương đậm nét.
PNT3	Phương ngữ Trung 3	Quảng Bình, Quảng Trị, Huế	Hệ thống thanh điệu đặc biệt, khác biệt lớn.
PNT4	Phương ngữ Trung 4	Các tỉnh Trung Bộ còn lại	Giao thoa giữa Phương ngữ Trung và Nam.
PNN	Phương ngữ Nam	Các tỉnh miền Nam	Biến âm phụ âm cuối, âm sắc nhẹ nhàng hơn.



1. Introduction

Motivation

Dataset

Example

Vocab

Group 3

Example



Dialect	Input
Standard	Bọn tôi có nhiều niềm vui khi chụp album này nên không mệt nhọc gì.
PNB	Bọn tôi có nhiều niềm vui khi chụp album này nên không mệt nhọc gì.
PNT1	Chúng tôi có nhiều niềm vui khi chụp album này nên không mệt mỏi chi.
PNT2	Nhà choa có nhiều niềm vui khi chụp album ni nên nỏ mệt mỏi chi.
PNT3	Choa có nhiều niềm vui khi chụp album ni bớ nỏ dộc chi.
PNT4	Chúng tôi có nhiều niềm vui khi chụp album ni nên không mệt mỏi chi.
PNN	Tụi tui có nhiều niềm vui khi chụp album này nên không oải gì.



1. Introduction

Motivation

Dataset

Example

Vocab

Group 3

Vocab



#	STANDARD (Chuẩn)	PNB (Phương ngữ Bắc)	PNN (Phương ngữ Nam)	PNT1 (Phương ngữ Trung 1)	PNT2 (Phương ngữ Trung 2)	PNT3 (Phương ngữ Trung 3)	PNT4 (Phương ngữ Trung 4)
1	gì	gì	sao	chi	chi	chi	chi
2	nào	gì	sao	mô	mô	mô	mô
3	vào	vào	vô	vô	vô	vô	vô
4	như thế nào	gì	ra sao	răng	a răng	chi	ra ren
5	không	không	hông	nỏ	nỏ	nỏ	không
6	thế nào	thế nào	sao	răng	răng	răng	ren
7	đã	từng	đã	đà	đạ	đạ	đã
8	làm	đã khiến	mắn	mắn	mắn	mắn	làm
9	và	và	với	ví	với	với	với
10	các	các	mấy	các	mấy	mấy	mấy
11	bố	bố	ba	bố	bố	bọ	ba
12	vì	vì	tại	vì	tại	vì	tại/do
13	mẹ	mẹ	má	miệ	mệ	mạ	má
14	những	các	mấy	mấy	những	mấy	mấy
15	đâu	gì	gì	đi mô	mô	mô	mô
16	đến	đến	tới	tới	đến	tới	tới
17	sao	thế nào	sao	răng	răng	răng	răng
18	cậu	nó	nó	nó	lạo	hắn	thằng nó
19	khi	khi	lúc	khi	cấy đoạn	chầu	lúc
20	này	này	này	ni	ni	ni	ni
21	cha	bố	ba	bố	bố	bọ	ba
22	thế	thế	vậy	rứa	rứa	rứa	rứa
23	hỏi	hỏi	hỏi	hỏi	hỏi	hỏi	hỏi

Group 3



1. Introduction

Motivation

Dataset

Example

Vocab

Group 3



1. Introduction

Group 3



1. Introduction

2. Benchmarks

3. Result analysis

4. Mitigation

Group 3



2. Benchmarks

Group 3



2. Benchmarks

Group 3



2. Benchmarks

Yêu cầu Cách đánh giá Metrics

Group 3



Yêu cầu

Giả sử $x^{\text{chuẩn}}$ là một đầu vào tiếng Việt chuẩn và $x^{\text{phương ngữ}}$ là biến thể phương ngữ của nó từ phương ngữ $d \in \{\text{Bắc, Trung, Nam, hỗn hợp}\}$. Giả sử hai đầu bảo toàn cùng một ý nghĩa:

$$\text{Sem}(x^{\text{chuẩn}}) \approx \text{Sem}(x^{\text{phương ngữ}})$$

Đối với một mô hình f và một tác vụ xuôi dòng T , tính mạnh mẽ yêu cầu:

$$f_T(x^{\text{chuẩn}}) \approx f_T(x^{\text{phương ngữ}})$$



Yêu cả

Giả sử x^{chu}
phương ngữ d

Đối với một m



ĩ của nó từ
ĩa:



2. Benchmarks

Yêu cầu

Cách đánh giá

Metrics

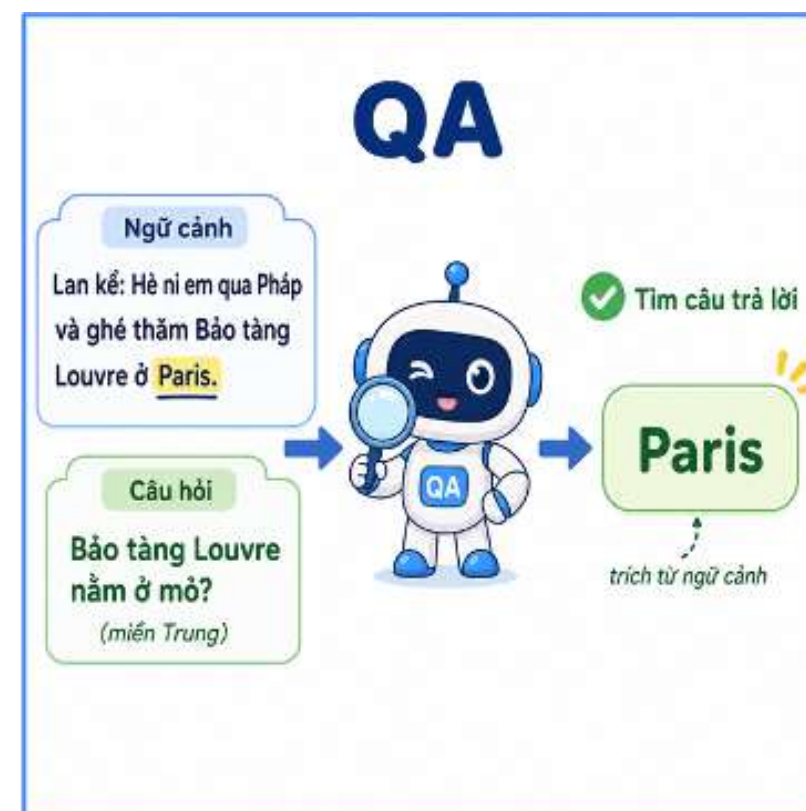
Group 3

Cách đánh giá



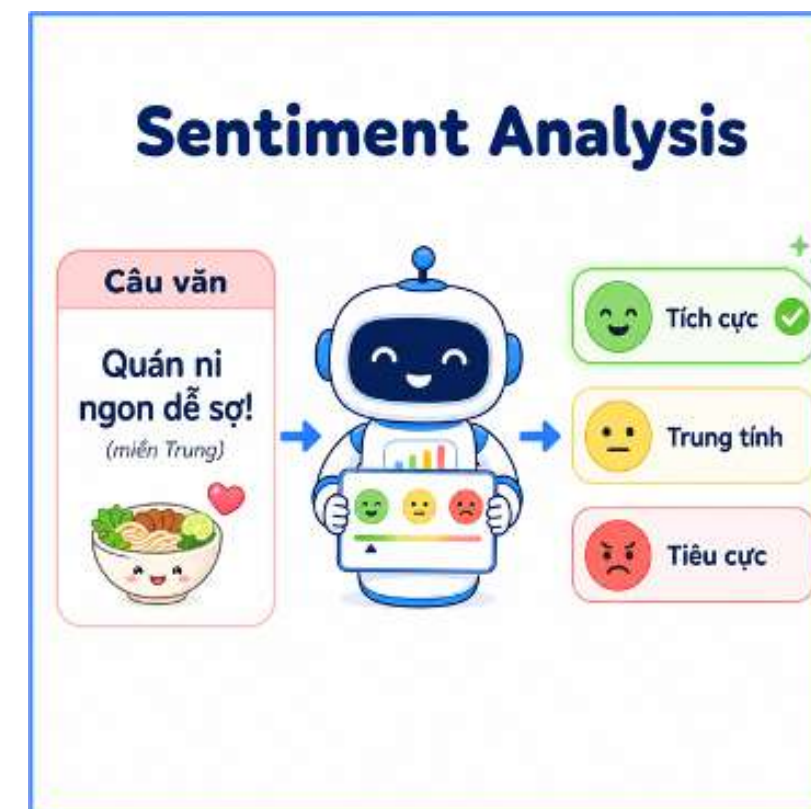
Suy luận ngôn ngữ

Trắc nghiệm



Trả lời câu hỏi

Phân tích cảm xúc





2. Benchmarks

Yêu cầu

Cách đánh giá

Metrics

Group 3



Metrics

Accuracy

NLI, MCQA, Sentiment Analysis

$$\text{Accuracy} = \frac{\text{Tổng số câu đúng}}{\text{Tổng số câu}}$$



Token-F1

Binary Classification

QA

$$F1 = \frac{\text{Tổng số câu AI đúng}}{\text{Tổng số câu}}$$



2. Benchmarks

Yêu cầu

Cách đánh giá

Metrics

Group 3



1. Introduction

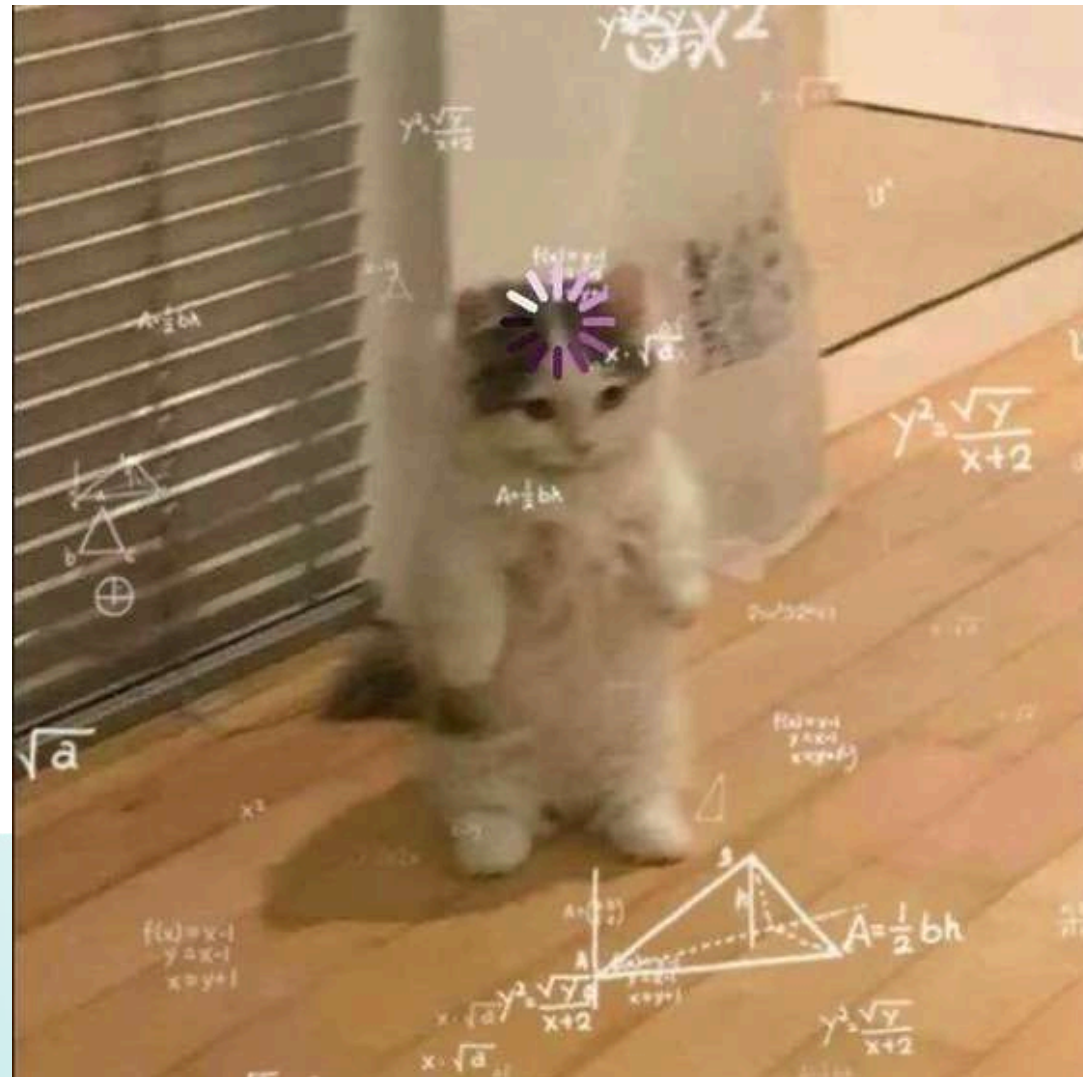
2. Benchmarks

3. Result analysis

4. Mitigation

Group 3

3. Result analysis



Group 3



3. Result analysis

Điểm số trung bình

Đánh giá độ ổn định

Sự giảm độ chính xác



Group 3

Điểm số trung bình



• NLI

Mô hình	Chuẩn	Bắc	Nam	Trung 1	Trung 2	Trung 3	Trung 4
Qwen2.5-	63	63	62	60	63	63	59
Qwen2.5-	63	65	60	60	61	60	63
Qwen2.5-	55	56	56	55	55	51	55
Qwen2.5-	33	33	33	33	33	33	33
Gemma-	53	52	57	51	49	47	52
Mistral-	49	51	46	51	47	43	51
SeaLLMs-	47	48	44	43	44	42	46
Vistral-	19	21	17	20	18	18	18

• Sentiment

Mô hình	Chuẩn	Bắc	Nam	Trung 1	Trung 2	Trung 3	Trung 4
Qwen2.5-	49	51	47	46	47	43	48
Qwen2.5-	38	41	40	38	35	34	44
Qwen2.5-	51	48	51	51	37	34	46
Qwen2.5-	24	23	25	24	22	20	22
Gemma-	39	41	39	40	41	38	40
Mistral-	36	39	37	36	33	29	38
SeaLLMs-	38	39	40	40	36	34	37
Vistral-	32	32	34	34	28	29	32

• MCQA

Mô hình	Chuẩn	Bắc	Nam	Trung 1	Trung 2	Trung 3	Trung 4
Qwen2.5	95	95	95	94	94	94	96
Qwen2.5	92	91	89	88	86	86	88
Qwen2.5	92	91	89	87	87	84	89
Qwen2.5	66	63	63	56	60	62	62
Gemma-	90	91	90	91	89	89	92
Mistral-	73	75	69	72	70	68	72
SeaLLMs	91	91	88	87	88	89	89
Vistral-	88	88	88	86	85	87	89

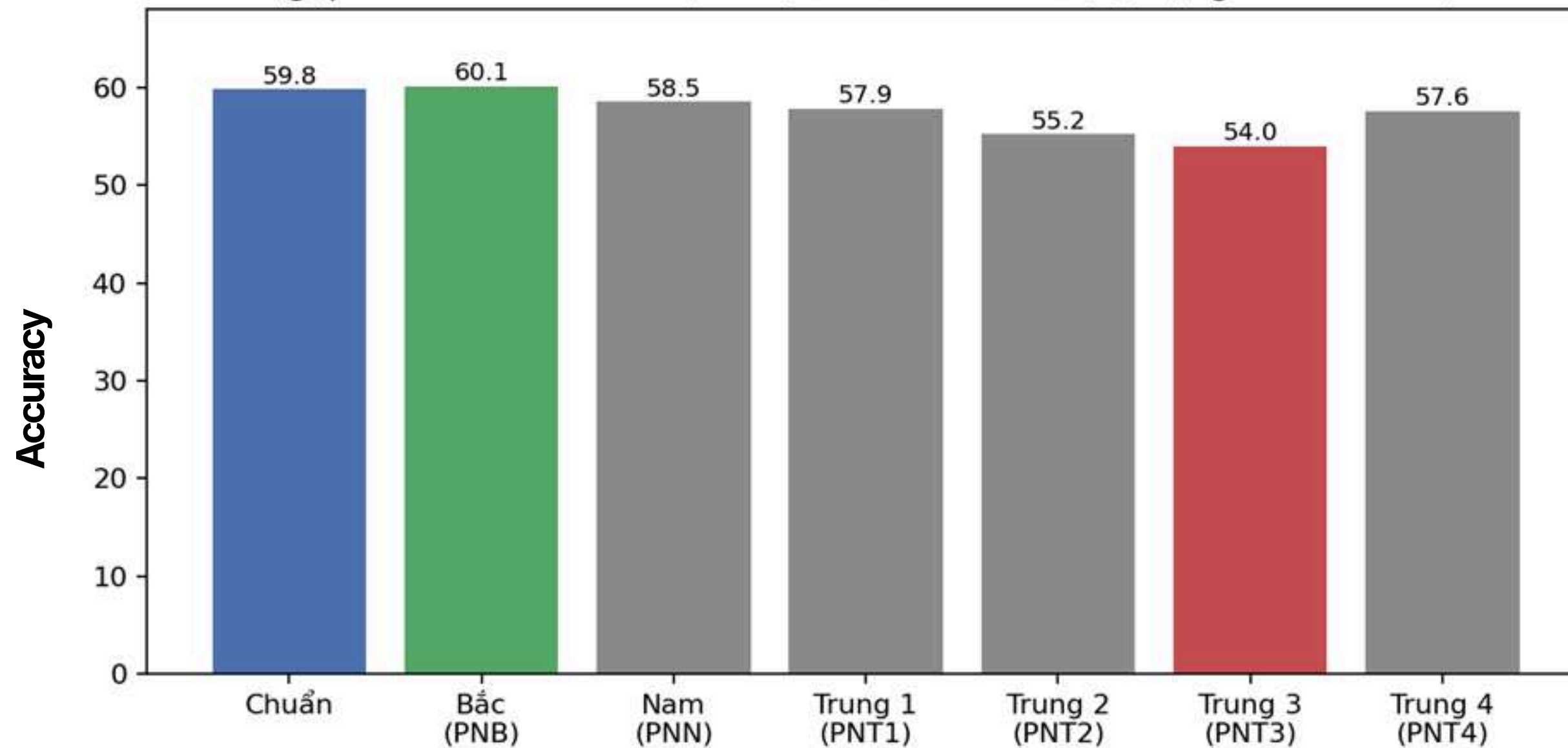
• QA

Mô hình	Chuẩn	Bắc	Nam	Trung 1	Trung 2	Trung 3	Trung 4
Qwen2.5	82	80.6	81.5	83.6	81.7	78.6	81.3
Qwen2.5	73.9	73	70.7	70.6	68.7	68.2	68.2
Qwen2.5	67.4	67.1	64.7	66.3	59.7	58.2	59.8
Qwen2.5	40.6	40.1	37.8	32.6	18.3	16.6	30.8
Gemma-	78.4	78.4	79.2	74.2	73.1	73.1	74.5
Mistral-	55.2	57.2	46.1	49	32.1	33.9	38.2
SeaLLMs	75.1	74.4	71.8	68.4	65.2	62.7	67.7
Vistral-	65.9	65.5	62.4	64	60.9	58.7	32.5

Điểm số trung bình



Điểm trung bình theo phương ngữ
(gộp 8 mô hình x 4 tác vụ MCQA/NLI/Sentiment/QA, trọng số đều nhau)



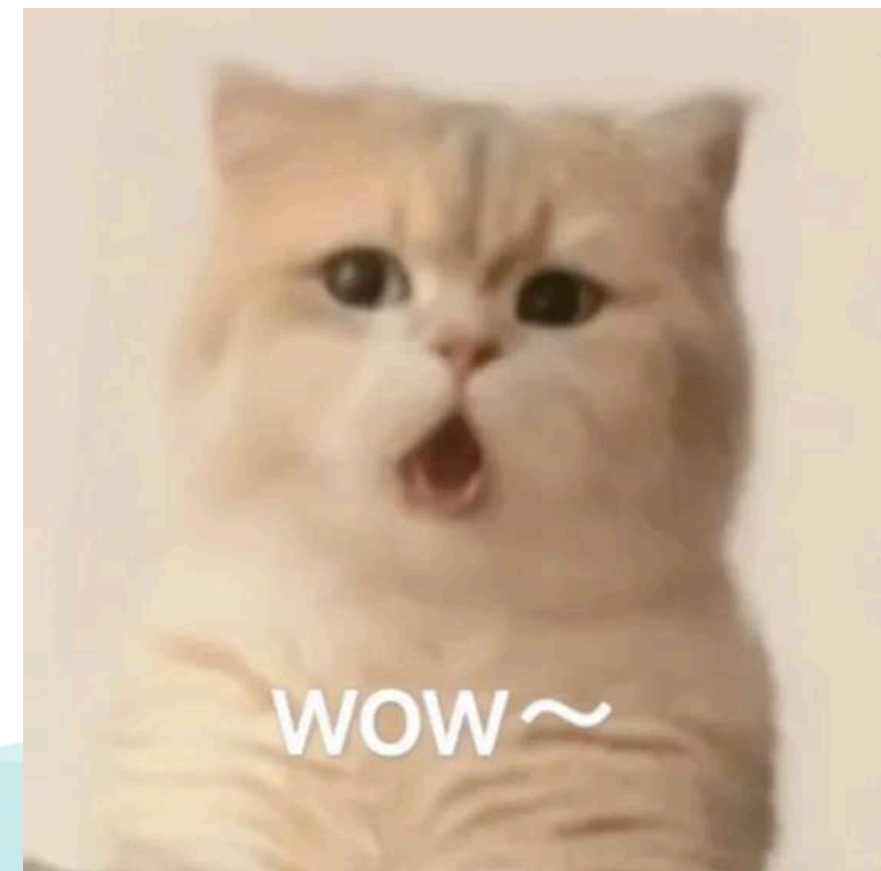
Thấp nhất



Điểm số trung bình

Điểm số của PNT3 và PNT4 gần như thấp nhất ở mọi task

➔ Phương ngữ có ảnh hưởng rất mạnh



Group 3



3. Result analysis

Điểm số trung bình

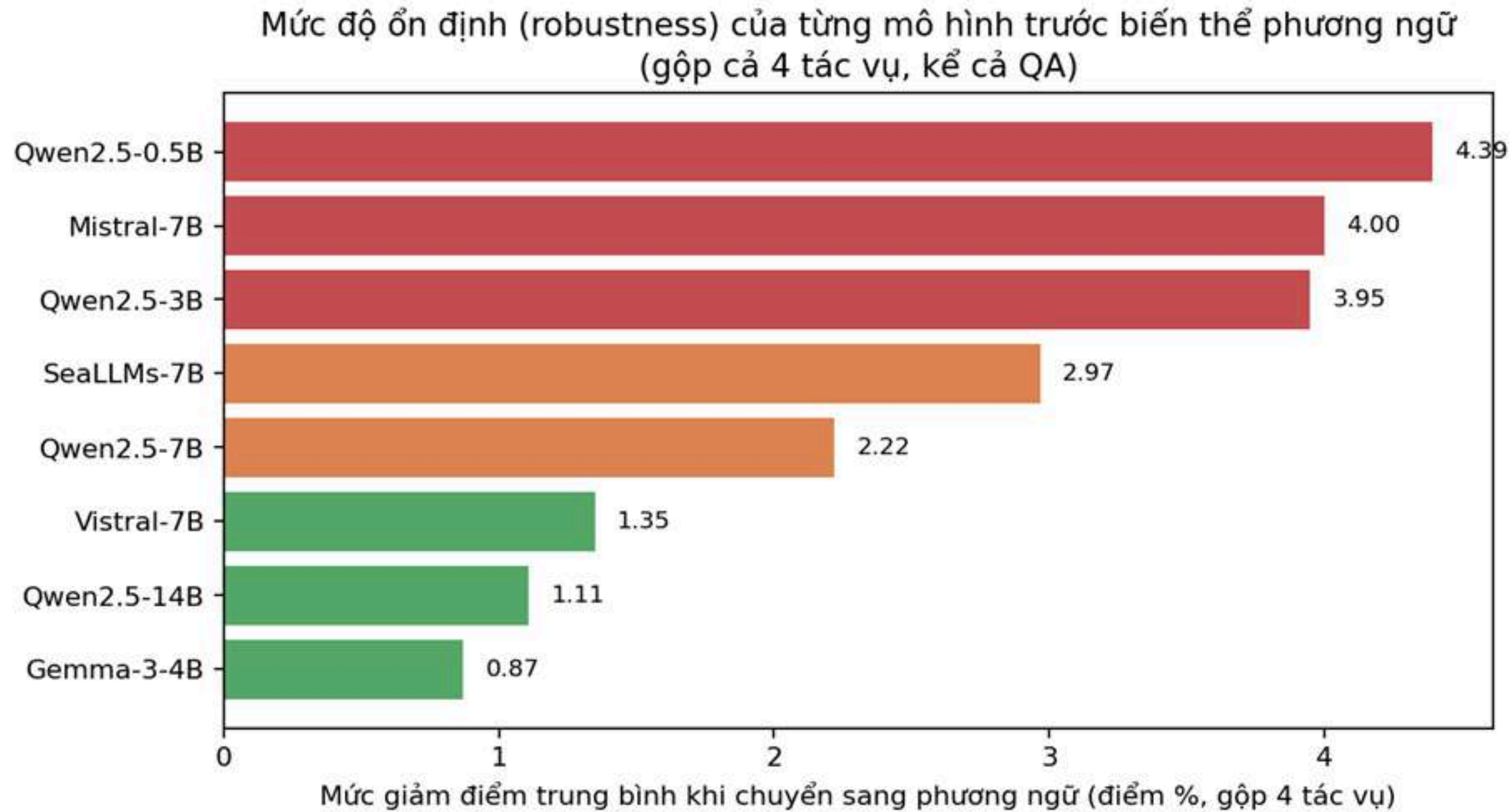
Đánh giá độ ổn định

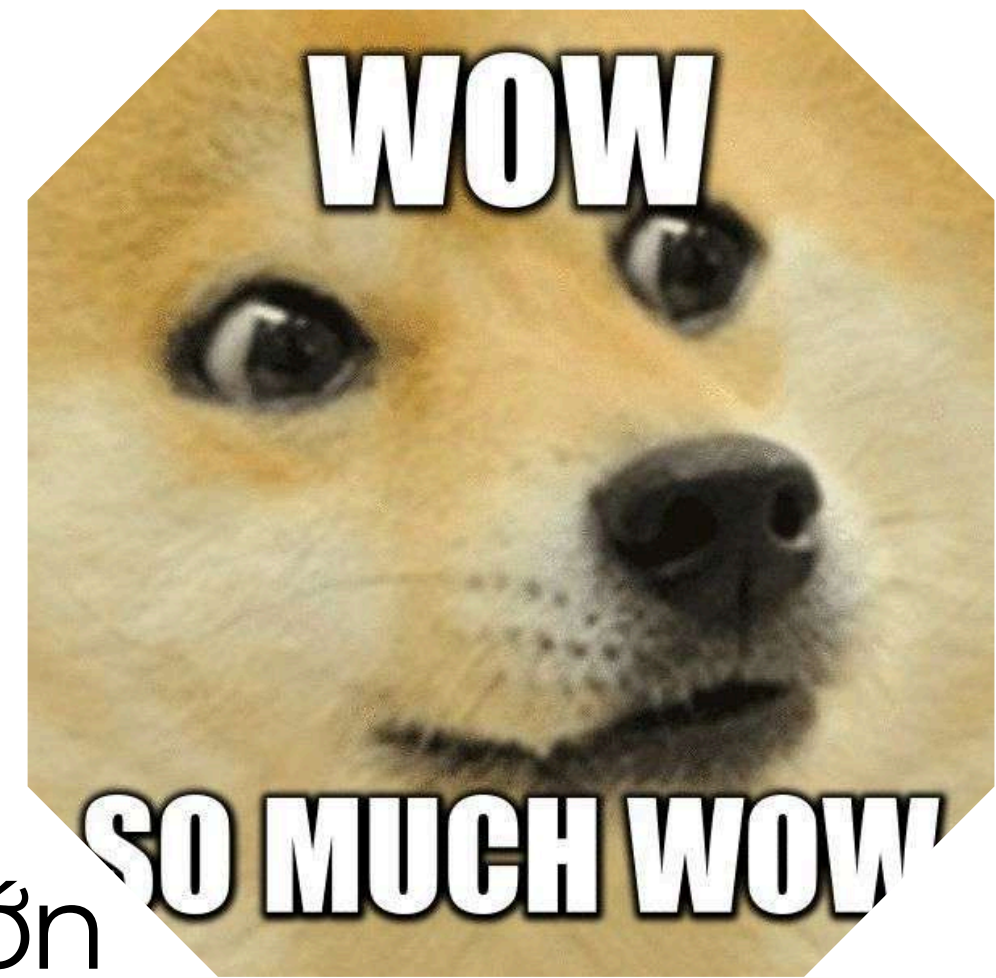


Sự giảm độ chính xác

Group 3

Đánh giá độ ổn định





Đánh giá độ ổn định

Các model có sự chênh lệch độ ổn định rất lớn

➔ Model có ảnh hưởng đến số liệu khảo sát

Group 3



3. Phân tích kết

Điểm số trung bình

Đánh giá độ ổn định

Sự giảm độ chính xác

Group 3

Sụt giảm độ chính xác



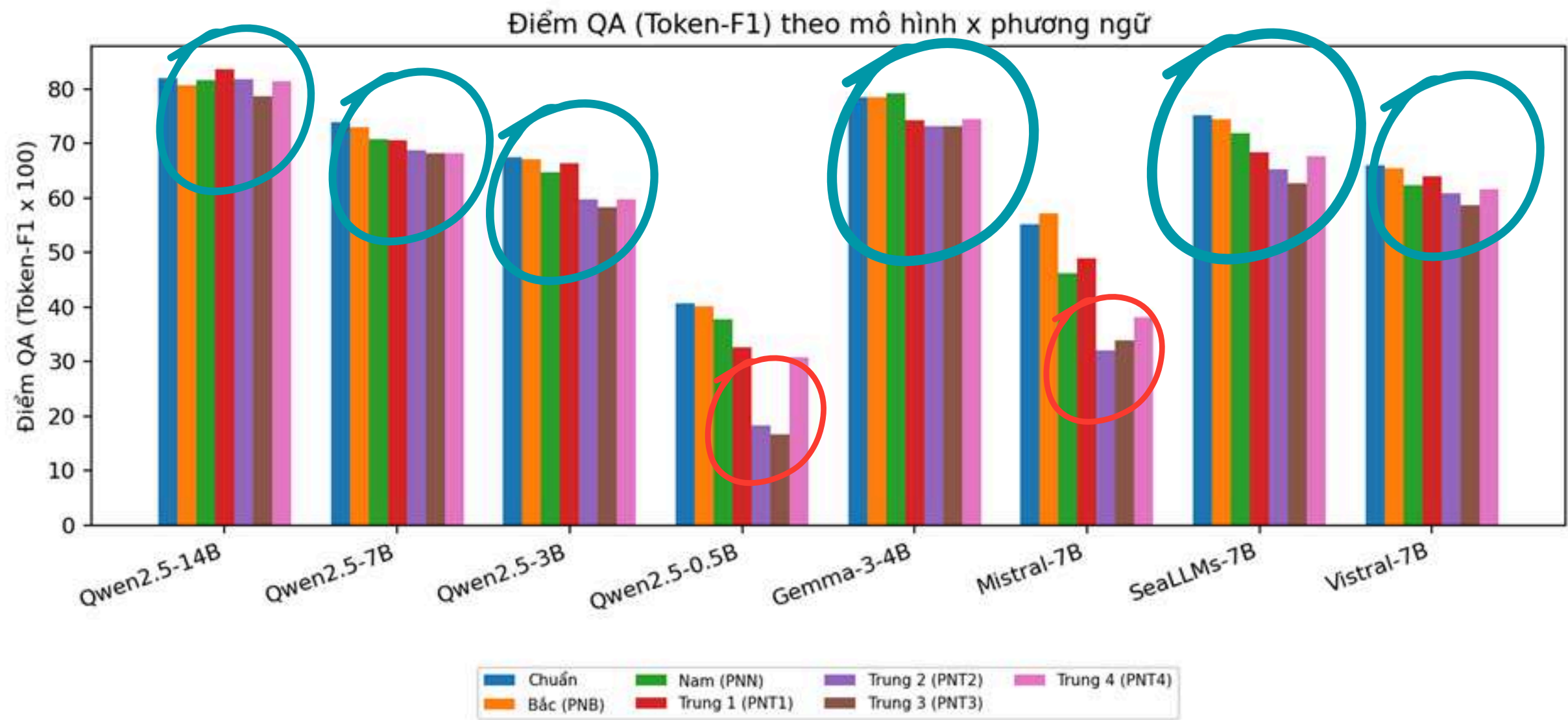
Tác vụ	PNB	PNN	PNT1	PNT2	PNT3	PNT4
MCQA	0.90%	3.80%	4.90%	4.90%	5.90%	3.50%
NLI	1.50%	4.80%	4.00%	4.80%	7.90%	3.00%
Sentiment	2.00%	4.50%	5.00%	9.60%	11.20%	5.60%
QA	2.80%	10.40%	10.80%	26.40%	26.10%	20.40%

Harmful Flip Rate

Kết quả cho thấy ở các task khó như QA thì phương ngữ ảnh hưởng rất mạnh

Độ chính xác giảm mạnh

Sụt giảm độ chính xác



Phương ngữ chỉ ảnh hưởng mạnh lên các model yếu, để tăng độ chính xác và ổn định thì có thể sử dụng model mạnh nhưng sẽ cần nhiều tài nguyên hơn vì vậy phải tìm cách tối ưu.



3. Result analysis

Điểm số trung bình

Đánh giá độ ổn định

Sự giảm độ chính xác

Group 3

41



1. Introduction

2. Benchmarks

3. Result analysis

4. Cải thiện

Group 3



4. Cải thiện

Group 3



4. Cải thiện

Ý tưởng chính

Cấu trúc mBART

Phân tích

Group 3



Ý tưởng chính

- Dialect - aware Prompting:
Ex: “Dưới đây là một câu hỏi MCQA có sử dụng phương ngữ miền trung, ví dụ rằng, rứa, nớ,.. Hãy trả lời nó một cách thật kĩ...”
- Seq2Seq:
Train mô hình dịch từ phương ngữ sang ngôn ngữ chuẩn

Group 3



4. Cải thiện

Ý tưởng chính

Cấu trúc mBART

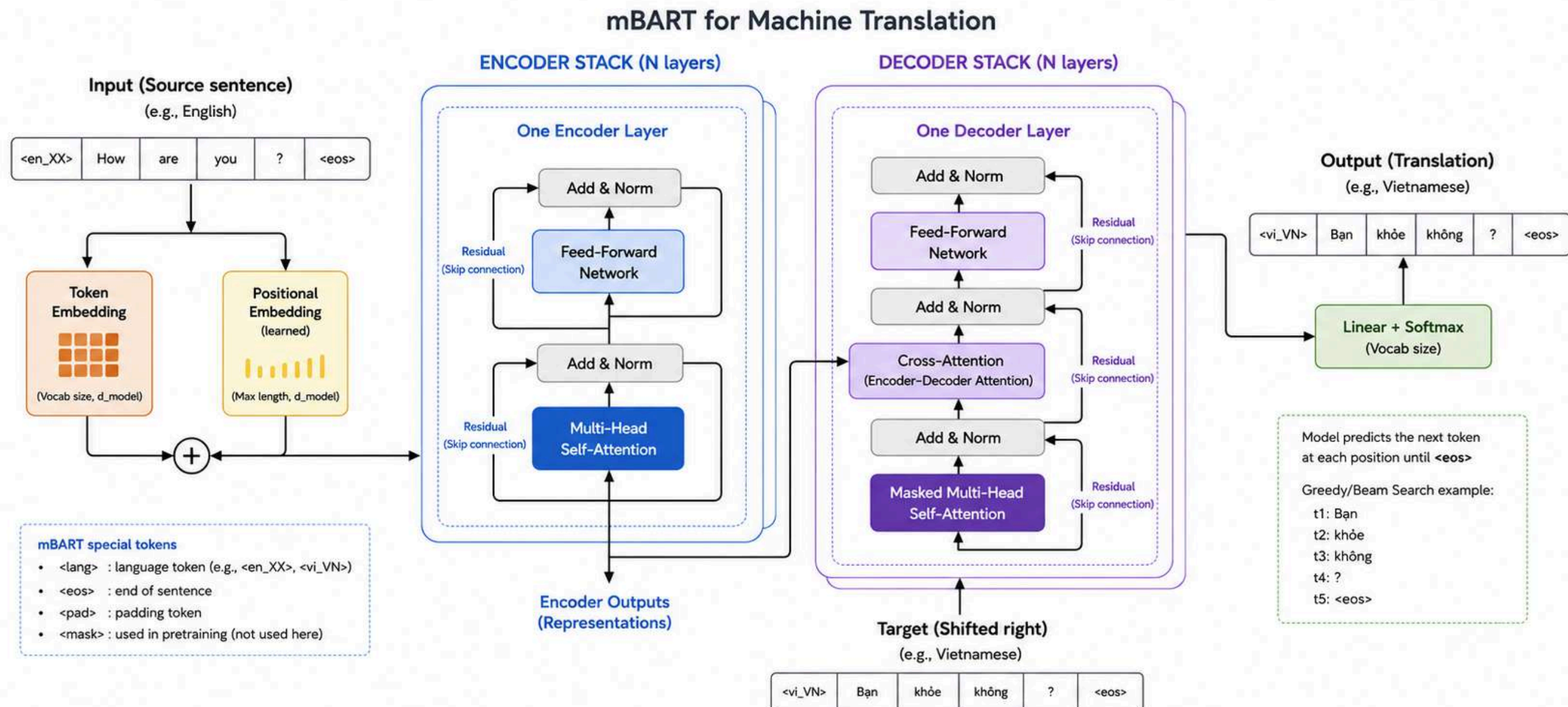
Phân tích

Group 3

Cấu trúc mBART



- Nền tảng của Transformer
- Đóng vai trò là một bộ lọc tiền xử lý chuẩn hóa phương ngữ khi chuyển phương ngữ sang ngôn ngữ chuẩn trước khi đưa vào LLM
- Giúp kết nối các từ tương ứng(attention)





4. Cải thiện

Ý tưởng chính

Cấu trúc mBART

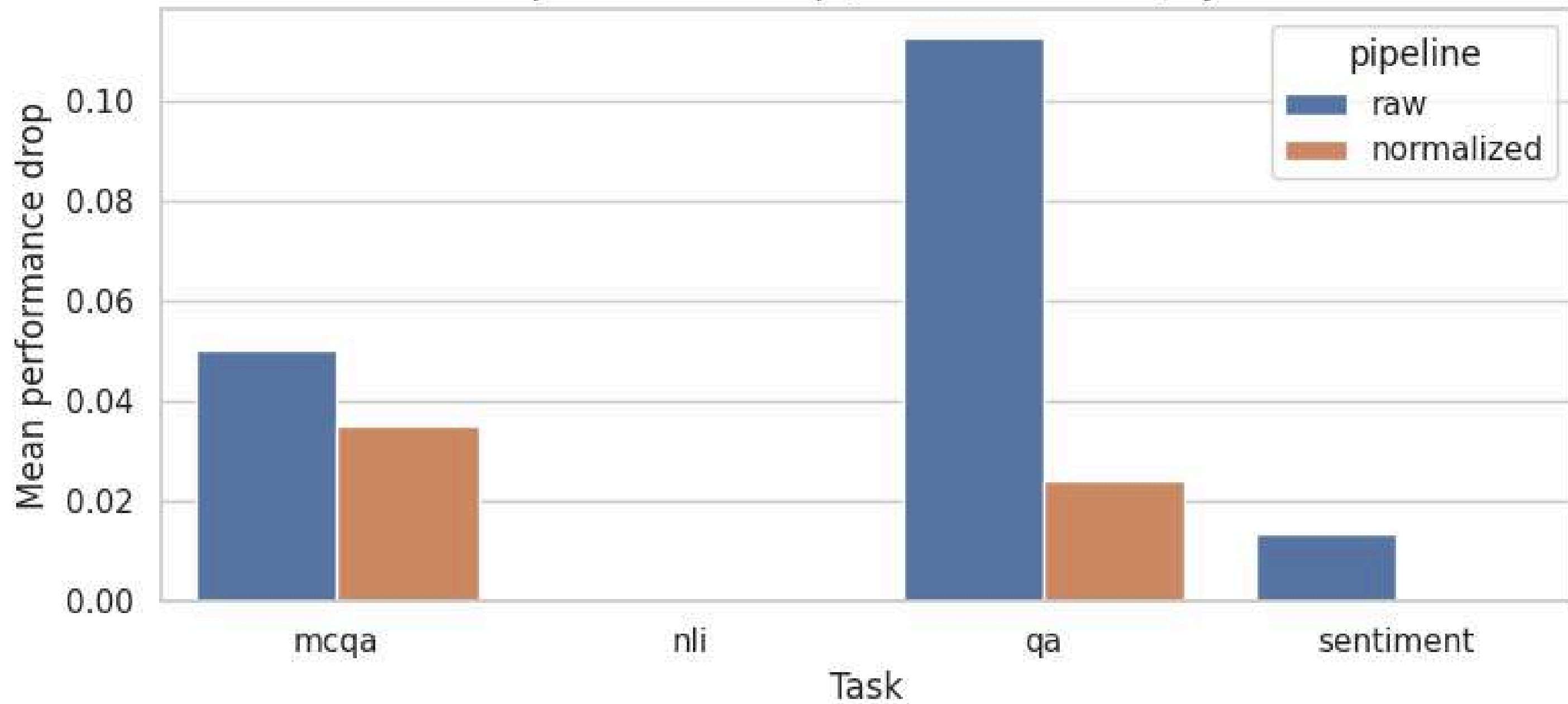
Phân tích

Group 3

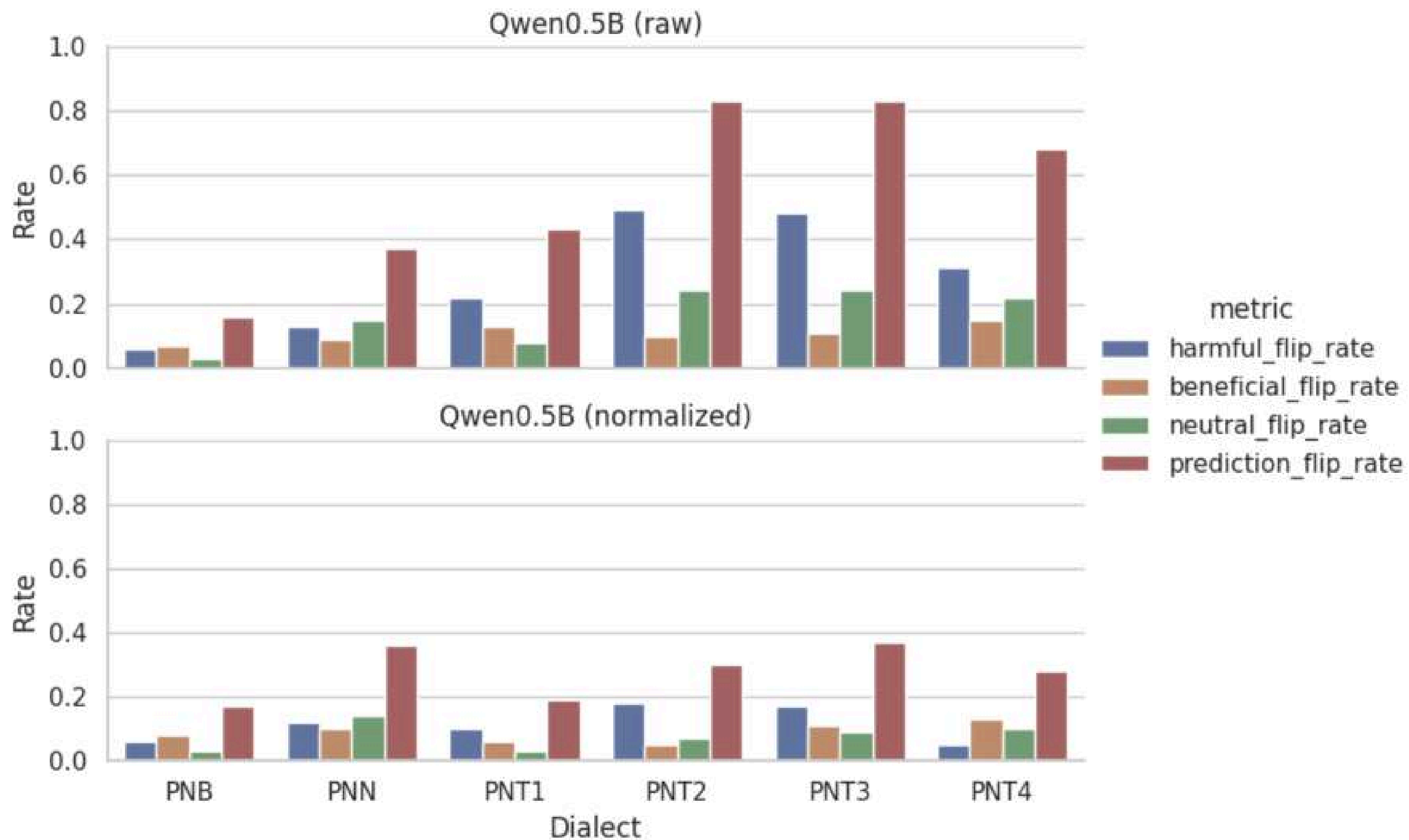
Phân tích



Mean performance drop (standard - dialect) by task



Flip rates by dialect - task: qa





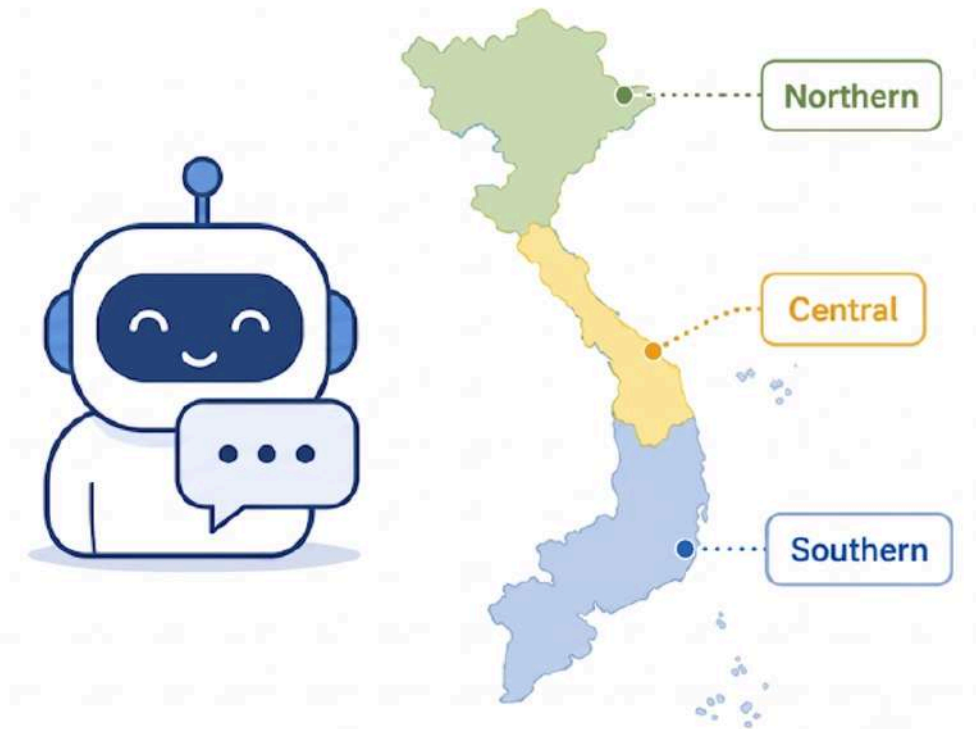


Tóm lại

Suy cho cùng, đích đến cao nhất của mọi công nghệ ngôn ngữ không phải là tạo ra một cỗ máy biết tuốt, mà là tạo ra một "người bạn" đủ tinh tế để hiểu chúng ta là ai, đến từ đâu, và trân trọng sự nguyên bản của mỗi cá nhân.



QR to our website demo





That's the end