

Phân tích ChatGPT
dưới góc độ Ngôn ngữ học:
KHẢ NĂNG và **HẠN CHẾ**



GV: Đinh Điền
ddien@hcmus.edu.vn
<https://www.clc.hcmus.edu.vn>
6/2023

KHẢ NĂNG và HẠN CHẾ



- **KHẢ NĂNG:** cần biết
 - ✓ ChatGPT&AI làm được gì?
 - ✓ Làm như thế nào?
 - Ứng dụng vào việc nào trong NCKH?

- **HẠN CHẾ:** cần biết
 - ✓ ChatGPT&AI chưa làm được gì?
 - ✓ Tại sao ?
 - Biết cách làm chủ ChatGPT&AI trong NCKH

ỨNG DỤNG ChatGPT & AI trong NGHIÊN CỨU KHOA HỌC

1

Giới thiệu về ChatGPT & AI

2

Khả năng của ChatGPT & AI

3

Hạn chế của ChatGPT & AI

4

Lưu ý khi ứng dụng ChatGPT&AI

I. ChatGPT là gì?



- ChatGPT là một “người máy đối thoại” (chatbot)
- Đối thoại với con người bằng ngôn ngữ con người
- ❖ Sản phẩm của cty OpenAI (30/11/2022)
- ✓ Ảnh hưởng sâu rộng đến mọi người
- ✓ >> so với các thành quả AI trước đây (chơi cờ)
- ✓ Vì nó liên quan đến Ngôn ngữ con người
- ❑ Ngôn ngữ: Phương tiện giao tiếp (truyền thông tin) quan trọng nhất của loài người.
- Nghiên cứu khoa học: Chịu tác động lớn
- Vì thông tin khoa học: Dạng ngôn ngữ (con người)
- “(thông) TIN”: 信 = 亻 (nhân) + 言 (ngôn)
- Làm chủ THÔNG TIN => làm chủ Thế giới (Lance Winslow)

Thử hỏi ChatGPT?

DD

Hãy giới thiệu về Hội Sinh viên VN và
khoa học của sinh viên



Hội Sinh viên Việt Nam là một tổ chức quan trọng, tập hợp các sinh viên từ đại học và cao đẳng trên toàn quốc. Mục tiêu của hội là đại diện cho quyền thúc đẩy sự phát triển của giáo dục và nghiên cứu khoa học, và tạo cơ hội tham gia vào hoạt động xã hội.



DD

Các hoạt động nghiên cứu khoa học của sinh viên, dưới sự hỗ trợ của Hội Sinh viên Việt Nam, thường bao gồm:

- Dự án nghiên cứu:** Sinh viên thường được khuyến khích và hỗ trợ để tham gia vào các dự án nghiên cứu, cả cá nhân và nhóm, trong và ngoài trường học. Điều này giúp họ xây dựng và phát triển kỹ năng nghiên cứu của mình.
- Hội thảo và hội nghị:** Hội Sinh viên Việt Nam thường tổ chức và/hoặc tài trợ cho các hội thảo và hội nghị, nơi sinh viên có thể trình bày kết quả nghiên cứu của mình, học hỏi từ các nhà nghiên cứu hàng đầu và mở rộng mạng lưới xã hội học thuật của họ.
- Cuộc thi nghiên cứu khoa học:** Các cuộc thi nghiên cứu sinh viên là cơ hội tuyệt vời để sinh viên trình bày công trình nghiên cứu của mình và có cơ hội đạt được sự công nhận và giải thưởng.
- Chương trình học bổng nghiên cứu:** Một số Hội Sinh viên Việt Nam có thể cung cấp hoặc hỗ trợ các chương trình học bổng dành riêng cho nghiên cứu khoa học, nhằm khuyến khích và hỗ trợ sinh viên trong việc thực hiện công trình nghiên cứu của mình.

ChatBot: Mô hình Dự đoán ~ MH Ngôn ngữ

Tôi sẽ đến tham dự buổi hội thảo lúc 8 giờ

Từ hiện tại

Xử lý Ngôn ngữ tự nhiên (NLP)

Từ kế tiếp

Câu hỏi

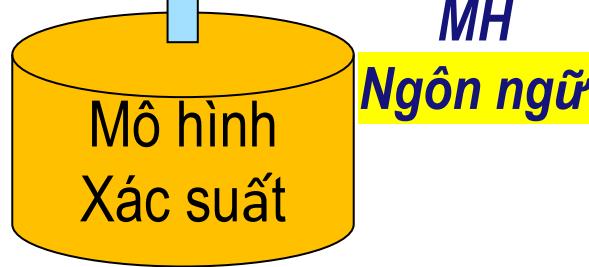
Đầu vào

Đầu ra

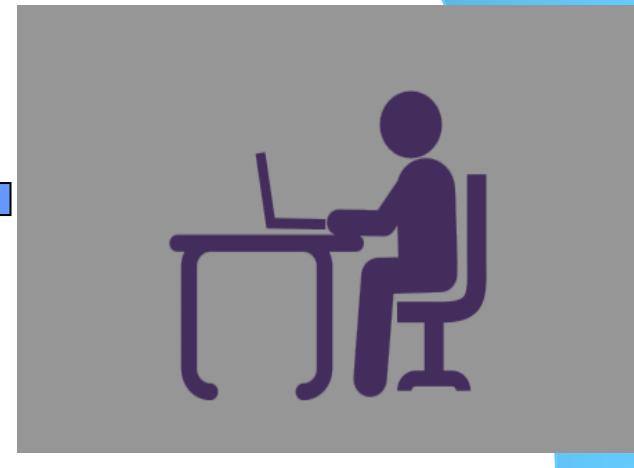
Câu trả lời
Đáp ứng

Yêu cầu

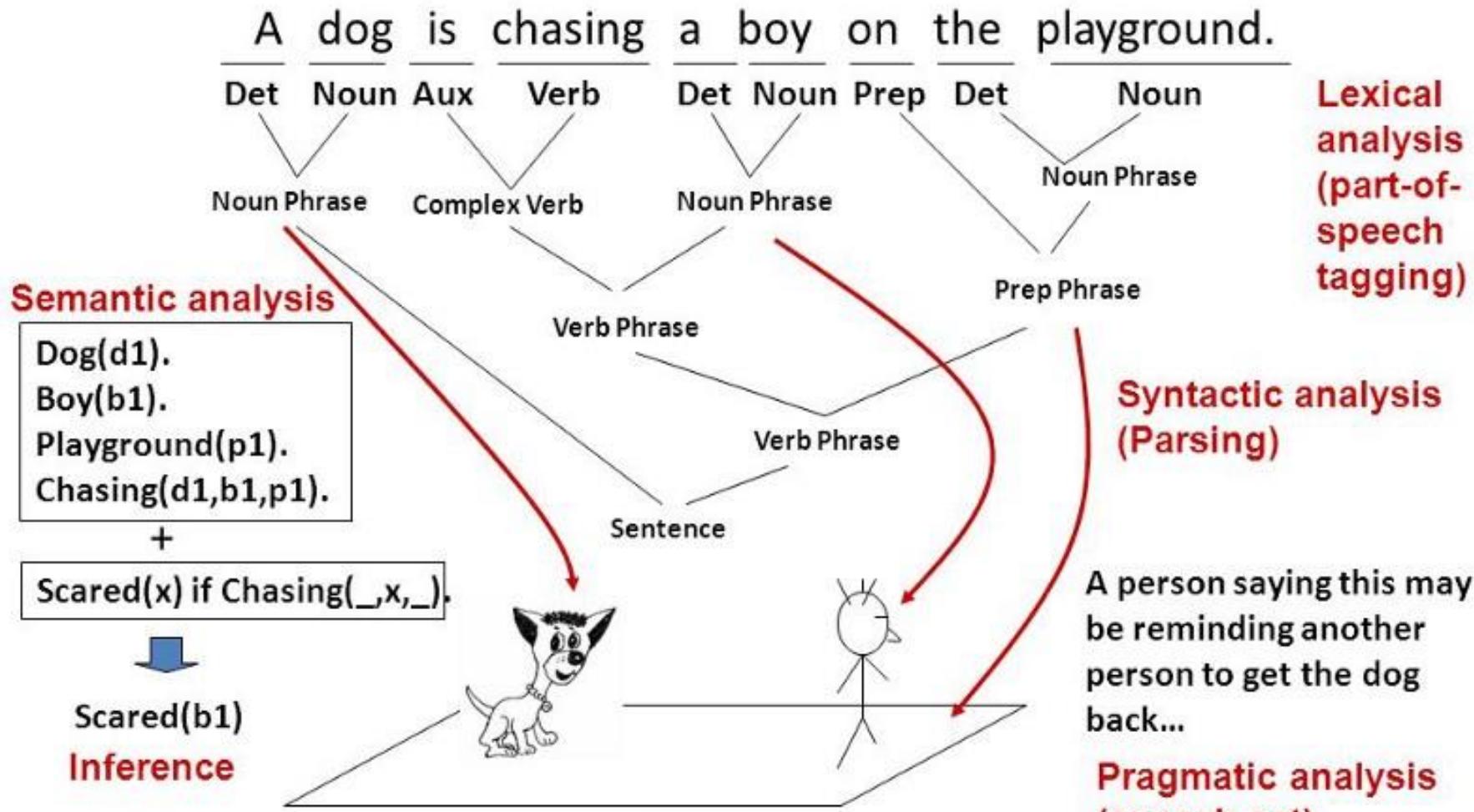
Dự đoán



Huấn luyện bằng Học Máy
(Machine Learning)



An example of NLP



Cơ chế hoạt động chatGPT

- ChatGPT: huấn luyện trên Mô hình ngôn ngữ lớn
- Kho ngôn ngữ: 300 tỷ từ với 175 tỷ tham số
- Học máy GPT (Generative Pre-trained Transformer)
- Do OpenAI đề xuất (2018)
- Từ mô hình Học sâu (Deep Learning) Transformer do Google Brain đưa ra (2017)
- Dựa trên mô hình Seq2Seq (LVQuoc et al., 2014)
- MHNN lớn: không chỉ dự đoán từ kế tiếp, mà
 - ✓ Câu (sentence) kế tiếp
 - ✓ Đoạn văn (paragraph) kế tiếp
 - ✓ Toàn văn bản (text)

Công nghệ học tăng cường với sự phản hồi của con người

- ✓ Reinforcement Learning with Human Feedback (RLHF)
- ❑ Con người:
 - ✓ Dạy máy cách trả lời đúng
 - ✓ Phản hồi kết quả của máy
 - ✓ Đánh giá xếp hạng các câu trả lời của máy
 - ❖ Giúp máy tối ưu hóa câu trả lời.
 - ✓ Đầu vào (prompt) tốt: cụ thể, chi tiết, chính xác
 - Đầu ra tốt
- ❑ Con người: cần học cách “sử dụng” ChatGPT một cách hiệu quả

II. Khả năng của ChatGPT



❑ Hai nhóm tác vụ Ngôn ngữ:

1. Ngôn ngữ máy/lập trình

(Machine/Programming languages)

2. Ngôn ngữ con người/tự nhiên

(Human/Natural languages)

Ngôn ngữ lập trình

1. Sinh mã (code generation): không cần đến lập trình viên

DU

Hãy viết một đoạn code python về kiểm tra số chẵn lẻ



+

☰



Dưới đây là một đoạn mã Python để kiểm tra xem một số có phải là số chẵn hay số lẻ:



```
python                                         Copy code

def kiem_tra_chan_le(number):
    if number % 2 == 0:
        return "Số {} là số chẵn.".format(number)
    else:
        return "Số {} là số lẻ.".format(number)

# Kiểm tra số chẵn/lẻ
so = int(input("Nhập một số nguyên: "))
print(kiem_tra_chan_le(so))
```

Trong mã này, chúng ta định nghĩa một hàm `kiem_tra_chan_le` để kiểm tra số chẵn/lẻ.

Nếu số đó chia hết cho 2 (khi chia lấy dư bằng 0), chúng ta cho rằng số đó là số chẵn. Ngược lại, nếu số đó không chia hết cho 2, chúng ta cho rằng số đó là số lẻ.

Ngôn ngữ lập trình

1. Sinh mã (code generation): Tự viết chương trình
2. Giải thích mã (code explanation): Giải thích bằng tiếng Anh/tiếng Việt/... đoạn mã nào, làm gì, ...
3. Chuyển mã (code conversion): Chuyển tự động một chương trình viết bằng mã ngôn ngữ lập trình A sang mã ngôn ngữ lập trình B. Chẳng hạn chuyển từ C sang Python.
4. Sửa lỗi (bug fixing): Tìm lỗi logic của chương trình và sửa lỗi đó.

II. Ngôn ngữ con người: Chức năng cơ bản

1. Đổi thoại
2. Tìm kiếm thông tin
3. Tạo văn bản
4. Dịch văn bản
5. Tóm tắt văn bản
6. Phân loại văn bản
7. Sửa lỗi văn bản
8. Xử lý dữ liệu
9. Siêu văn bản

- Văn bản = **Ngôn ngữ**:
- Hiểu ngôn ngữ không?
- Tiếng Anh/tiếng Việt?
- Hiểu ngữ nghĩa sâu?

1. Đối thoại (text conversation): hỏi – đáp (chat)

DD

Có phải David Luan, người làm ở OpenAI, là người Việt phải không?



Xin lỗi, như một mô hình ngôn ngữ trí tuệ nhân tạo, tôi không có quyền truy cập vào thông tin cá nhân về những người đã làm việc tại OpenAI hoặc ở bất kỳ đâu khác, bao gồm quốc tịch hoặc dân tộc của họ. Điều này cũng giúp bảo vệ sự riêng tư của cá nhân và không đưa ra giả định về danh tính của họ mà không có sự đồng ý của họ. Do đó, tôi không thể xác nhận liệu David Luan có phải là người Việt hay không.

1. Đổi thoại: Gợi ý ý tưởng nghiên cứu

DU

Tôi đang có một bộ ngữ liệu nhiều văn bản cùng với tên tác giả của từng văn bản đó, bạn có thể gợi ý ý tưởng nghiên cứu giúp tôi được không?



AI

Tất nhiên! Dựa trên thông tin mới về bộ dữ liệu của bạn, dưới đây là một số ý tưởng nghiên cứu có thể bạn quan tâm:



1. Phân loại tác giả: Bạn có thể xây dựng một mô hình học máy để phân loại tác giả dựa trên nội dung của các văn bản. Mục tiêu là xác định những đặc trưng riêng biệt trong phong cách viết của từng tác giả và sử dụng chúng để nhận dạng tác giả của các văn bản không được biết trước.
2. Phân tích phong cách viết: Bạn có thể nghiên cứu về phân tích phong cách viết của các tác giả trong bộ dữ liệu. Điều này bao gồm việc tìm hiểu các đặc điểm ngôn ngữ, cấu trúc câu, sử dụng từ vựng, hoặc phương pháp sắp xếp ý trong văn bản. Bằng cách phân tích những yếu tố này, bạn có thể tìm ra sự đa dạng và đặc trưng của phong cách viết của từng tác giả.



1. Đối thoại: Tư vấn hướng nghiên cứu

DU

Ngữ liệu văn bản đã được phân loại theo độ khó có giúp ích được gì cho các nghiên cứu về mô hình ngôn ngữ lớn không?



Ngữ liệu văn bản đã được phân loại theo độ khó có thể mang lại nhiều lợi ích cho các nghiên cứu về mô hình ngôn ngữ lớn, bao gồm:



1. Xây dựng mô hình ngôn ngữ: Dữ liệu về độ khó của các văn bản có thể được sử dụng để xây dựng và cải thiện các mô hình ngôn ngữ lớn. Đối với mô hình ngôn ngữ, việc đánh giá độ khó của một từ, câu hoặc văn bản có thể cung cấp thông tin hữu ích để điều chỉnh quá trình sinh và dự đoán ngôn ngữ.

2. Đánh giá hiệu suất mô hình: Dữ liệu văn bản được phân loại theo độ khó có thể được sử dụng để đánh giá hiệu suất của các mô hình ngôn ngữ. Bằng cách sử dụng các văn bản có độ khó khác nhau, ta có thể đánh giá khả năng của mô hình trong việc xử lý và hiểu các văn bản phức tạp hơn.

3. Đánh giá và so sánh mô hình: Các tập dữ liệu văn bản phân loại theo độ khó có thể được sử dụng để đánh giá và so sánh hiệu suất của các mô hình ngôn ngữ khác nhau. Bằng cách sử dụng các văn bản có độ khó đã biết trước, ta có thể so sánh khả năng dự đoán và sinh ngôn ngữ của các mô hình khác nhau và đo lường sự khác biệt giữa chúng.



1. Đổi thoại: hỏi-đáp thủ tục hành chánh

DU

Tôi là một công dân ở Quận 5, tôi có công việc và muốn chuyển sang nhà khác cũng ở quận 5. Tôi nên làm những thủ tục gì để có thể dời đồng hồ nước sang nhà mới

TH

TH



Để chuyển đồng hồ nước từ nhà cũ sang nhà mới trong cùng một khu vực, bạn cần thực hiện các bước sau:

TH

1. Liên hệ với Công ty Cấp nước địa phương: Đầu tiên, hãy liên hệ với Công ty Cấp nước địa phương, đây là tổ chức quản lý và cung cấp nước cho khu vực của bạn. Hỏi về quy trình chuyển đồng hồ nước và yêu cầu hướng dẫn cụ thể.

2. Đơn chuyển đồng hồ nước: Công ty Cấp nước sẽ yêu cầu bạn điền đơn chuyển đồng hồ nước. Đơn này thường có sẵn tại văn phòng công ty hoặc có thể tải xuống từ trang web của công ty. Đơn đầy đủ thông tin cần thiết trong đơn, bao gồm thông tin cá nhân, địa chỉ cũ và địa chỉ mới.

1. Đối thoại: tiếp nhận khiếu nại của khách hàng

DU

Tôi là một công dân thuộc Quận Thủ Đức, trong quá trình sử dụng và tiêu thụ nước ở thời gian qua thì một tháng gia đình tôi chỉ tiêu thụ xấp xỉ 50 khối nước. Trong tháng trước thì gia đình tôi nhận được phiếu đóng tiền nhưng mà số nước tiêu thụ lên đến 80 khối nước trong khi lượng nước sử dụng vẫn như các tháng trước. Tôi cần làm gì để khiếu nại về khối lượng nước cao bất thường này?



ĐC

Để khiếu nại về khối lượng nước cao bất thường mà bạn nhận được trên phiếu đóng tiền, bạn có thể thực hiện các bước sau:

CN

1. Kiểm tra lại chỉ số đồng hồ nước: Đầu tiên, hãy kiểm tra lại chỉ số đồng hồ nước của gia đình bạn để xem liệu nó có chính xác hay không. Đảm bảo rằng chỉ số được ghi lại đúng và không có sự nhầm lẫn.
2. Liên hệ với Công ty Cấp nước địa phương: Liên hệ trực tiếp với Công ty Cấp nước địa phương để báo cáo về tình huống và khiếu nại về lượng nước cao bất thường trên phiếu đóng tiền. Hỏi về quy trình khiếu nại và yêu cầu hướng dẫn cụ thể.
3. Làm đơn khiếu nại: Công ty Cấp nước có thể yêu cầu bạn điền đơn khiếu nại để gửi cho họ. Trong đơn này, bạn cần cung cấp thông tin chi tiết về tình huống, bao gồm chỉ số nước đúng của bạn và lý do tại sao bạn cho rằng lượng nước được ghi trên phiếu đóng tiền không chính xác.

1. Đổi thoại: gợi ý

DU

ChatGPT có thể gợi ý cho tôi mục tiêu tổng quát và kế hoạch cho năm 2023 cho một công ty cấp nước với công suất là 1967616 m³/ngày, giảm tỷ lệ nước thải toàn còn 17.09%

TH

TB



Dựa trên thông tin bạn đã cung cấp về công ty cấp nước có công suất là 1.967.616 m³/ngày và tỷ lệ nước thải hiện tại là 17.09%, dưới đây là một số gợi ý về mục tiêu tổng quát và kế hoạch cho năm 2023:



Mục tiêu tổng quát:

- Giảm tỷ lệ nước thải: Mục tiêu chính là giảm tỷ lệ nước thải xuống mức thấp hơn so với hiện tại. Một tỷ lệ nước thải thấp hơn sẽ giúp tăng hiệu suất hoạt động và tiết kiệm nguồn nước.
- Tăng cường hiệu suất hoạt động: Mục tiêu là nâng cao hiệu suất hoạt động của hệ thống cấp nước để đảm bảo cung cấp nước hiệu quả, đáng tin cậy và bền vững cho khách hàng.
- Nâng cao chất lượng nước: Mục tiêu là đảm bảo chất lượng nước cung cấp đáp ứng các tiêu chuẩn chất lượng nước quốc gia và quốc tế. Điều này đảm bảo sự an toàn và sức



1. Đôi thoại: giải đáp thủ tục

DU

Trong quá trình làm thủ tục cấp giấy phép hoạt động in, nhưng tôi có làm mất CCCD thì có cách nào để xử lý trong trường hợp này không? (tôi đang cần gấp)



Nếu bạn đã làm mất CCCD trong quá trình làm thủ tục cấp giấy phép hoạt động in, bạn có thể thực hiện các bước sau để xử lý:

1. Báo mất CCCD: Bạn cần thông báo về việc mất CCCD cho cơ quan cấp giấy phép hoạt động in và yêu cầu hướng dẫn thủ tục cấp lại CCCD mới.
2. Tìm kiếm lại CCCD: Nếu bạn vẫn còn thời gian, hãy kiểm tra lại các đồ vật của bạn hoặc đến các cơ quan công quyền gần nhất để tìm kiếm lại CCCD. Nếu tìm thấy, bạn cần thông báo cho cơ quan cấp giấy phép hoạt động in để hủy đơn yêu cầu cấp lại CCCD.
3. Điều chỉnh thủ tục: Nếu bạn không tìm thấy CCCD, cơ quan cấp giấy phép hoạt động in sẽ hướng dẫn bạn các thủ tục cấp lại CCCD mới. Trong quá trình đợi, bạn có thể yêu cầu cơ quan cấp giấy phép hoạt động in điều chỉnh thời gian xử lý đơn hoặc cung cấp giấy chứng nhận tạm thời để đáp ứng nhu cầu gấp của bạn.

Lưu ý: Bạn cần báo mất CCCD và thực hiện các thủ tục cấp lại CCCD mới ngay khi phát hiện mất để tránh những rủi ro về an ninh thông tin, giao dịch tài chính và các vấn đề pháp lý khác.

2. Tìm kiếm và tổng hợp thông tin: Google <> Bing

Thế nào là độ khó văn bản (readability)



All Images Videos News Books More Tools

About 14,600 results (0.45 seconds)

Độ khó của văn bản là chỉ số xác định văn bản dễ hay khó đọc ở mức nào. Độ khó của văn bản đóng vai trò vô cùng quan trọng trong việc soạn thảo, phát hành và lựa chọn sách, đặc biệt là trong lĩnh vực giáo dục.



hueuni.edu.vn

<https://jos.hueuni.edu.vn> › article › download PDF



MÔ HÌNH SO SÁNH CÁC VĂN BẢN TIẾNG VIỆT THEO ĐỘ KHÓ

- ✓ Searching for: độ khó văn bản là gì?
- ✓ Generating answers for you...

Độ khó của văn bản là độ đo xác định một văn bản là dễ hay khó đọc thế nào. Nó là một hướng nghiên cứu nhằm xác định sự phù hợp giữa văn bản và người đọc ¹ ². Theo Alan Bailin và Ann Grafstein, độ khó của văn bản là độ đo xác định một văn bản là dễ hay khó đọc thế nào ¹. Độ khó của văn bản được xác định dựa trên các yếu tố như số lượng từ ngữ khó hiểu, cấu trúc câu phức tạp, từ vựng chuyên ngành ³.

Learn more: [1. bing.com](https://bing.com) [2. clc.hcmus.edu.vn](https://clc.hcmus.edu.vn) [3. researchgate.net](https://researchgate.net) [4. vns.edu.vn](https://vns.edu.vn)

1 of 20



Làm thế nào để xác định độ khó của một văn bản?

Các yếu tố nào ảnh hưởng đến độ khó của một văn bản?

2. Tìm kiếm thông tin: tài liệu nghiên cứu

DD

Tôi đang nghiên cứu về độ khó của văn bản (readability), hãy tìm giúp tôi một số tài liệu nghiên cứu mới để tôi tham khảo.

số tài liệu quan trọng bạn nên tìm đọc khi nghiên cứu về độ khó của văn bản (readability):

1. "**Readability: A view from cognitive psychology**" của Klare, G. (1984). Đây là một nghiên cứu quan trọng về readability từ góc độ tâm lý học nhận thức, cung cấp cho bạn kiến thức cơ bản về cách mà con người tiếp thu thông tin từ văn bản.
2. "**The Flesch Reading Ease Formula**" của Rudolf Flesch (1948). Đây là công thức đánh giá độ khó của văn bản đã được sử dụng rộng rãi, dựa trên số lượng từ trong mỗi câu và số lượng âm tiết trong mỗi từ.
3. "**Readability Revisited: The New Dale-Chall Readability Formula**" của Dale và Chall (1995). Đây là một công thức đánh giá độ khó của văn bản khác, dựa trên số lượng từ không nằm trong danh sách 3000 từ dễ nhất.
4. "**Coh-Metrix: Analysis of text on cohesion and language**" của Graesser, McNamara, Louwerse, và Cai (2004). Coh-Metrix là một công cụ phân tích văn bản tính toán các chỉ số liên quan đến độ khó của văn bản như sự liên kết văn bản và ngữ ngữ liệu.
5. "**Exploring the readability of web pages for people with intellectual disabilities**" của Williams and Merten (2011). Đây là một nghiên cứu về độ khó của văn bản trong môi trường trực tuyến, đặc biệt đối với những người có khuyết tật trí tuệ.

Hãy tìm đọc những tài liệu này trên các cơ sở dữ liệu học thuật như Google Scholar, JSTOR, hoặc PubMed.

2. Tìm kiếm thông tin: tìm kiếm tài liệu (ChatSonic)

A

Hãy kiểm giúp tôi vài tài liệu nghiên cứu về độ khó của văn bản

CS

Sure, I'd be happy to help! Here are some relevant search results on the difficulty level of texts:

1. According to researchers, the two main factors that influence the difficulty level of a text are its meaning and grammatical structure. [1]
2. The Flesch formula is a tool that can be used to objectively assess the difficulty level of a text. It takes into account factors like sentence length and syllable count to determine how easy or difficult a text is to read. [1]
3. There is a growing interest in using AI to optimize search engines, create copies, and generate unique content. One tool that is available for this purpose is Smodin, which uses a recreation method to rewrite or recreate text in over 100 languages. [2]
4. In academic research, the use of other people's work or ideas without proper citation is known as plagiarism. This is a serious offense that can result in consequences like failing a class or losing one's job. [3]

Do these help?

References:

- [1] Độ khó của văn bản và việc kiểm tra ngôn ngữ
- [2] Viết lại, viết lại hoặc viết lại văn bản | đổi từ miễn phí
- [3] Lỗi đạo văn trong nghiên cứu khoa học và cách phòng tránh

2. Tìm kiếm thông tin: tìm kiếm tài liệu (ChatSonic)

A

Please suggest to me some research papers on the text readability along with links for viewing or downloading

CS

Sure, here are some research papers on text readability along with their links to view or download:

1. "A Neural Approach to Automated Readability Assessment" by X. Liu et al. [1] - <https://www.aclweb.org/anthology/D16-1085.pdf>
2. "Assessing Text Readability: A Review of Approaches and Tools" by S. Sharoff [2] - <https://www.aclweb.org/anthology/W13-1001.pdf>
3. "Factors Affecting Text Readability: A Review of Empirical Studies" by M. Al-Sabbagh [3] - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7421355/>
4. "Readability Assessment of Web Pages: A Review" by V. Chakravarthy et al. [4] - <https://www.sciencedirect.com/science/article/pii/S1877050915010197>

I hope these resources are helpful for your research on text readability.

References:

- [1] [Supervised and Unsupervised Neural Approaches to Text ...](#)
- [2] [Computational Assessment of Text Readability: A Survey of ...](#)
- [3] [A Unified Framework for Predicting Text Quality - UPenn CIS](#)

2. Sinh văn bản (text generation): viết dàn bài

DU

Bạn hãy gợi ý một dàn ý chi tiết cho bài báo khoa học chuyên ngành Khoa học máy tính cho những công việc trong nghiên cứu mà tôi đã thực hiện sau đây:

TI



Đề tài: Độ khó văn bản tiếng Việt trong lĩnh vực Khoa học máy tính: Nghiên cứu, phân tích và đánh giá

TH



I. Giới thiệu

- Giới thiệu về lĩnh vực Khoa học máy tính và vai trò của độ khó văn bản trong quá trình nghiên cứu và xử lý ngôn ngữ tự nhiên.
- Nhấn mạnh sự thiếu hụt nghiên cứu đối với độ khó văn bản tiếng Việt trong lĩnh vực này.

II. Khái quát về độ khó văn bản

- Định nghĩa độ khó văn bản và vai trò của nó trong việc hiểu và xử lý ngôn ngữ.
- Các khía cạnh quan trọng liên quan đến độ khó văn bản, bao gồm độ phức tạp ngữ pháp, từ vựng, cấu trúc câu, ngữ nghĩa, v.v.
- Đánh giá tầm quan trọng của việc đo lường và đánh giá độ khó văn bản trong ngôn ngữ học và Khoa học máy tính.

III. Các phương pháp tiếp cận nghiên cứu độ khó văn bản

- Tổng quan về các phương pháp tiếp cận phổ biến để nghiên cứu và đo lường độ khó văn bản.

vào các hoạt động phức tạp và mang lại giá trị gia tăng, nó cũng có thể dẫn đến việc thay thế một số công việc, đặc biệt là những công việc chỉ yêu cầu kỹ năng ngôn ngữ và giao tiếp cơ bản. Do đó, chúng ta cần cung cấp đào tạo và hỗ trợ đầy đủ cho nhân viên của chúng ta để thích nghi với công nghệ mới và phát triển kỹ năng và năng lực mới.

Tóm lại, việc áp dụng ChatGPT trong hệ thống dịch vụ công trực tuyến của chúng ta đồng nghĩa với cơ hội và thách thức đối với chúng ta làm cán bộ quản lý nhà nước. Bằng cách sử dụng công nghệ này, chúng ta có thể cải thiện hiệu quả, chất lượng và trải nghiệm người dùng của các dịch vụ công của chúng ta, và nâng cao hình ảnh và uy tín của thành phố của chúng ta. Tuy nhiên, chúng ta cũng cần ý thức được các rủi ro và giới hạn của công nghệ, và áp dụng các biện pháp tích cực để đảm bảo chất lượng, độ tin cậy và việc sử dụng đạo đức của công nghệ này. Với kế hoạch và thực thi cẩn thận, tôi tin rằng chúng ta có thể thành công trong việc tích hợp ChatGPT vào hệ thống dịch vụ công của chúng ta và tạo ra một tương lai tốt đẹp cho thành phố và công dân của chúng ta.

Cảm ơn quý vị đã chú ý.



In conclusion, the adoption of ChatGPT in our online public service system represents both opportunities and challenges for us as civil servants. By leveraging this technology, we can improve the efficiency, quality and user experience of our public services, and enhance the image and reputation of our city. However, we also need to be aware of the risks and limitations of the technology, and take proactive measures to ensure its quality, reliability and ethical use. With careful planning and execution, I believe that we can successfully integrate ChatGPT into our public service system and create a better future for our city and its citizens.

Thank you for your attention.

Another challenge is the potential impact on human resources. While ChatGPT can automate many routine tasks and free up staff to focus on more complex and value-added activities, it may also lead to the displacement of some jobs, especially those that require only basic language and communication skills. Therefore, we need to provide adequate training and support to our staff to adapt to the new technology and develop new skills and competencies.

3. Sinh câu hỏi: tự sinh các câu hỏi thường gặp

DA

Các câu hỏi thường gặp trong dịch vụ công



Dưới đây là một số câu hỏi thường gặp khi sử dụng dịch vụ công:



1. Làm thế nào để tìm hiểu về các dịch vụ công cần thiết cho tôi?
2. Làm thế nào để nộp đơn xin cấp giấy tờ như hộ chiếu, giấy khai sinh...?
3. Tôi có thể làm thủ tục trực tuyến hay không?
4. Làm thế nào để tra cứu thông tin về hồ sơ của mình?
5. Làm thế nào để kiểm tra tiến độ xử lý hồ sơ của tôi?
6. Tôi phải trả phí bao nhiêu để sử dụng dịch vụ công?
7. Làm thế nào để khiếu nại hoặc phản ánh nếu tôi gặp sự cố trong quá trình sử dụng dịch vụ công?

3. Sinh câu hỏi: tự sinh câu trả lời

DA

Hãy hướng dẫn trả lời các câu hỏi sau:



1. Làm thế nào để tìm hiểu về các dịch vụ công cần thiết cho tôi?
2. Làm thế nào để nộp đơn xin cấp giấy tờ như hộ chiếu, giấy khai sinh...?
3. Tôi có thể làm thủ tục trực tuyến hay không?
4. Làm thế nào để tra cứu thông tin về hồ sơ của mình?
5. Làm thế nào để kiểm tra tiến độ xử lý hồ sơ của tôi?



CB



1. Để tìm hiểu về các dịch vụ công cần thiết cho bạn, bạn có thể truy cập trang web của cơ quan cung cấp dịch vụ công hoặc ứng dụng di động của họ. Thông thường, các cơ quan cung cấp dịch vụ công đều có danh mục các dịch vụ công mà họ cung cấp, đồng thời cũng cung cấp thông tin về các yêu cầu, thủ tục và giấy tờ cần thiết cho mỗi dịch vụ. Bạn cũng có thể liên hệ với cơ quan cung cấp dịch vụ công để được hướng dẫn chi tiết.
2. Để nộp đơn xin cấp giấy tờ như hộ chiếu, giấy khai sinh, bạn có thể truy cập trang web của cơ quan cung cấp dịch vụ công để tìm hiểu thông tin chi tiết về thủ tục và giấy tờ



3. Sinh văn bản: viết công trình nghiên cứu

DU

Hãy viết 1 đoạn tổng quan nghiên cứu về chuyển tự chữ Nôm, có liệt kê các ý về ứng dụng trong nghiên cứu lịch sử, địa lý, y học cổ truyền, chủ quyền biển đảo



TH

PC

SV

TH

PC

Nghiên cứu về chuyển tự chữ Nôm là một lĩnh vực quan trọng trong việc khôi phục và giải mã các tài liệu cổ truyền bằng ngôn ngữ chữ Nôm, một hệ thống chữ viết được sử dụng rộng rãi trong lịch sử Việt Nam. Chuyển tự chữ Nôm đang thu hút sự quan tâm của nhiều nhà nghiên cứu và nhà khoa học với các ứng dụng đa dạng trong lĩnh vực lịch sử, địa lý, y học cổ truyền và chủ quyền biển đảo. Dưới đây là các ý chính về ứng dụng của chuyển tự chữ Nôm trong các lĩnh vực này:

1. Nghiên cứu lịch sử: Chuyển tự chữ Nôm giúp nhà nghiên cứu khôi phục và giải mã các tài liệu cổ truyền viết bằng chữ Nôm, đóng góp vào việc tái hiện và khám phá lịch sử Việt Nam. Việc chuyển tự chữ Nôm giúp nghiên cứu các văn bản, di chỉ, sử liệu lịch sử, giúp hiểu rõ hơn về văn hóa, xã hội, sự phát triển của dân tộc qua thời gian.
2. Nghiên cứu địa lý: Chuyển tự chữ Nôm cung cấp cơ sở dữ liệu quan trọng cho nghiên cứu địa lý. Các tài liệu viết bằng chữ Nôm thường chứa thông tin quý giá về địa danh, mô tả địa lý, hệ thống địa lí, biểu đồ và bản đồ lịch sử. Chuyển tự chữ Nôm giúp tiếp cận và phân tích những thông tin này để tìm hiểu về quá trình hình thành và biến đổi địa lý của đất nước.

3. Sinh văn bản: soạn giáo án

Vd: Soạn giáo án cho bài “Làng” của Kim Lân



Chào mừng các em đến với buổi học về bài văn "Làng" của tác giả Kim Lân. Bài văn này được viết với mục đích giới thiệu về cuộc sống trong một làng quê, nơi mà tác giả đã dựa dẫm suốt thời thơ ấu của mình.



1. Giới thiệu về tác giả:

- Kim Lân là một nhà văn người Việt Nam, sinh ra và lớn lên trong một làng quê.
- Nhiều tác phẩm của ông đã được chú ý về cuộc sống trong làng quê và giá trị của nền văn hoá dân gian.

2. Nội dung bài văn:

- Bài văn mô tả cuộc sống trong làng quê, với những truyền thống văn hóa và giá trị gia đình được truyền lại từ nhiều thế hệ.
- Tác giả còn nhấn mạnh sự quan tâm và sự tôn trọng của các dân làng đối với mối quan hệ gia đình và bạn bè.
- Bài văn còn gợi lên sự trải nghiệm của tác giả trong quá trình lớn lên trong một môi trường những nền văn hóa và giá trị gia đình quan trọng.

3. Phân tích từng đoạn

3. Sinh văn bản: lập dàn bài

DU

Hãy lập dàn bài cho một bài báo về chủ đề thiên văn



I. Đầu bài



- A. Mở đầu: giới thiệu chủ đề thiên văn, nhắc đến những nỗ lực nghiên cứu của cộng đồng nhà khoa học.
- B. Đề cập đến những phát hiện mới nhất về thiên văn.

II. Phần chính

- A. Giới thiệu về thiên văn hệ mặt trời của chúng ta.
- B. Nhắc đến các hành tinh trong hệ mặt trời và sự quan tâm của nhà khoa học đối với chúng.
- C. Mô tả những phương tiện và kỹ thuật được sử dụng để nghiên cứu thiên văn.
- D. Trình bày những kết quả và phát hiện mới nhất về thiên văn.

III. Kết luận

- A. Tóm lại những điều đã được nói trong bài báo.
- B. Đề xuất những định hướng cho nghiên cứu tiếp theo về thiên văn.
- C. Kết thúc bài báo bằng một câu nhận xét về sứ mệnh của nghiên cứu về thiên văn.

3. Sinh văn bản: viết bài báo

DU

Hãy viết một bài báo dựa vào dàn bài trên



"Khám phá Thiên Văn: Những Phát Hiện Mới Nhất"



Trong suốt nhiều năm qua, cộng đồng nhà khoa học đã tiếp tục nỗ lực để khám phá và hiểu hơn về thiên văn. Những phát hiện mới nhất về thiên văn đã cho chúng ta một cái nhìn mới về vũ trụ của chúng ta.

Thiên văn hệ mặt trời của chúng ta gồm nhiều hành tinh độc đáo và nổi bật. Nhà khoa học đang quan tâm đến những hành tinh này, nhằm tìm hiểu hơn về sự sống và dạng tồn tại của chúng. Phương tiện và kỹ thuật hiện đại như teleskop và sức mạnh tính toán đã giúp cho việc nghiên cứu về thiên văn trở nên dễ dàng hơn bao giờ hết.

Mới đây, nhà khoa học đã phát hiện ra một hành tinh mới tại hệ mặt trời của chúng ta. Hành tinh này có kích thước gần giống với Trái Đất và có thể có dạng sống. Điều này đã tạo ra nhiều câu hỏi và thú vị cho nhà khoa học về tương lai của chúng ta và sự tồn tại của sinh

3. Sinh văn bản: soạn bài tập

NG

Đọc đoạn văn sau và đưa ra bài quiz phù hợp:

Hoạt động chào đời hàng năm của rùa sông Nam Mỹ khổng lồ (*Podocnemis expansa*) ở phía tây lưu vực Amazon được coi là tập hợp rùa con lớn nhất trên hành tinh. Ước tính có khoảng 80.000 con rùa cái trưởng thành tập trung tại đây từ cuối tháng 9 để làm tổ và đẻ trứng. Hàng trăm nghìn rùa con bắt đầu nở từ giữa tháng 12 sang tháng 1.



Bài quiz:



1. Tại sao rùa sông Nam Mỹ khổng lồ tập trung tại phía tây lưu vực Amazon?
 - A. Để tổ và đẻ trứng
 - B. Để sống và tìm kiếm thức ăn
 - C. Để gặp gỡ với nhau
2. Khi nào rùa con bắt đầu nở tại phía tây lưu vực Amazon?
 - A. Từ cuối tháng 9
 - B. Từ giữa tháng 12 sang tháng 1
 - C. Từ đầu tháng 1
3. Bao nhiêu con rùa cái trưởng thành tập trung tại phía tây lưu vực Amazon?
 - A. Khoảng 80.000 con
 - B. Hàng trăm nghìn con
 - C. Chỉ vài trăm con.

3. Sinh văn bản: soạn bài tập

NG

Hãy viết một câu tiếng Anh sử dụng các từ reach, rich, least, list, seat, sit, leave, live.



"The wealthy man reached the top of the list of richest people in the world and chose to sit in the most comfortable seat, leaving the least luxurious option for those who live less extravagantly."



NG

Bạn có thể viết một đoạn văn Tiếng Anh sử dụng nhiều phrasal verb nhất có thể được không



Sure! Here's a paragraph that makes use of multiple phrasal verbs:



After finishing her shift at the cafe, Sarah headed home to put her feet up and relax. She had been running around all day, trying to keep up with the busy rush of customers coming in and out. As she settled into the couch, she flipped through the channels, settling on her favorite show. Just as she was getting into it, her phone rang, and she reluctantly picked it up, only to find her friend asking if she wanted to come out for a drink. Despite her initial hesitation, she finally agreed, knowing that she could use a night out to unwind and let off some steam after a long week of work.

3. Sinh văn bản: soạn đề thi

Vd: Soạn đề thi cuối kỳ môn Ngữ văn lớp 12:



Dưới đây là một đề thi cuối kỳ Ngữ Văn lớp 12:



I. Viết lại câu (2 điểm):

1. Chị tôi có thể chạy nhanh hơn anh ta.
-> Anh ta có thể chạy chậm hơn chị tôi.
2. Cây này cao hơn cây kia.
-> Cây kia thấp hơn cây này.

II. Trả lời câu hỏi (3 điểm):

1. Hãy giới thiệu một vài tác phẩm của nhà văn Nguyễn Tuân.
2. Tại sao bạn thích bài văn "Làng" của Kim Lân?

III. Viết một đoạn văn miêu tả về cuộc sống trong làng quê (5 điểm):

IV. Phân tích bài văn "Làng" của Kim Lân (10 điểm):

- Nội dung bài văn
- Những cảm xúc và suy nghĩ của tác giả khi viết bài văn
- Những giá trị văn hoá và gia đình được truyền lại trong bài văn
- Những lời khuyên hay học được từ bài văn

Tổng cộng: 20 điểm.

Lưu ý: Đề thi trên chỉ là một ví dụ và có thể sẽ có sự thay đổi tùy theo yêu cầu và tiêu chuẩn của mỗi trường học.

4. Dịch văn bản (text translation): dịch đa ngữ

DD

Hãy dịch câu sau sang tiếng Anh, Pháp, Hoa, Nhật, Hàn, Đức, Nga, Latin và Quốc tế ngữ: "Chúng tôi đang học một ngôn ngữ"



4. Dịch văn bản: dịch đa ngữ

DD

Hãy dịch câu sau sang tiếng Anh, Pháp, Hoa, Nhật, Hàn, Đức, Nga, Latin và Esperanto: "Chúng tôi đang học một ngôn ngữ"



- Tiếng Anh: "We are learning a language"
- Tiếng Pháp: "Nous apprenons une langue"
- Tiếng Hoa: "我们正在学习一门语言" (Wǒmen zhèngzài xuéxí yī mén yǔyán)
- Tiếng Nhật: "私たちは言語を学んでいます" (Watashitachi wa gengo o manande imasu)
- Tiếng Hàn: "우리는 언어를 배우고 있습니다" (Urineun eoneoreul baeugo issseubnida)
- Tiếng Đức: "Wir lernen eine Sprache"
- Tiếng Nga: "Мы изучаем язык" (My izuchayem yazyk)
- Tiếng Latin: "Nos discimus linguam"
- Esperanto: "Ni lernas lingvon"



5. Tóm tắt văn bản: Tóm tắt một chương sách

DU

Summary of the content of chapter 6 of the book "Deep Learning with Python"



Chapter 6 of the book "Deep Learning with Python" focuses on the use of convolutional neural networks (CNNs) for computer vision tasks. The chapter starts with an overview of how CNNs work, including the use of convolutional layers and pooling layers to extract features from images. The author then provides a detailed explanation of how to build a CNN model, including the use of data augmentation techniques to increase the size of the training set and prevent overfitting.



The chapter also covers the use of pre-trained CNN models, such as the VGG16 and VGG19 models, for transfer learning. Transfer learning involves taking a pre-trained model and using it as a starting point for a new model, allowing for faster and more accurate training on a smaller dataset.

The author also discusses the use of data generators and callbacks in Keras, a popular deep learning library. Data generators allow for efficient loading and preprocessing of large image datasets, while callbacks can be used to monitor the training process and adjust the learning rate or stop training early if necessary.

The chapter concludes with a discussion of visualizing the features learned by a CNN, including the use of activation maximization to generate images that maximally activate a particular feature. Overall, chapter 6 provides a comprehensive introduction to using CNNs for computer vision tasks, with practical examples and tips for improving model performance.

5. Tóm tắt văn bản: Tóm tắt một bài báo khoa học

DD

Hãy tóm tắt bài báo "Development of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel" - Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975).



Finished browsing



Bài báo "Development of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel" của tác giả Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975) tập trung vào việc điều chỉnh lại ba công thức đọc hiểu để phù hợp hơn với việc sử dụng trong Hải quân. Đồng thời, những công thức này được kiểm tra đôi với sự hiểu biết của họ thông qua 18 đoạn văn được lấy từ Sách hướng dẫn Đào tạo Hạng. Điểm số từ bài kiểm tra đọc và đoạn văn tài liệu đào tạo cho phép tính toán cấp độ học của các đoạn văn. Cấp độ đọc theo tỷ lệ này dựa trên nhân viên Hải quân đang đọc tài liệu đào tạo Hải quân và hiểu nó¹.

5. Tóm tắt văn bản: Tóm tắt cả một tập tin dự án

DD

Hãy tóm tắt nội dung tập tin tại địa chỉ sau

https://www.dropbox.com/s/ky4kd3s8cbhsxfo/SmartCity_District_10.pdf?dl=0

Tài liệu PDF bạn cung cấp là một báo cáo về việc xây dựng Đô thị thông minh tại Quận 10. Dưới đây là một số điểm chính:

1. Mục tiêu của việc xây dựng đô thị thông minh là hiện đại hóa hành chính, nâng cao năng lực quản lý, và phục vụ công dân, doanh nghiệp tốt hơn. Điều này sẽ cho phép tiếp nhận và xử lý khôi lượng thông tin lớn, hỗ trợ lãnh đạo ra quyết định chính xác và kịp thời, và cho phép dự báo dài hạn hơn, toàn diện hơn, độ chính xác cao hơn (Trang 10).
2. Việc xây dựng các hệ thống thông tin thông minh cùng với các giải pháp đồng bộ sẽ cung cấp môi trường phát triển ngày càng tiện ích, ưu việt cho doanh nghiệp và người dân. Hệ thống hạ tầng giao thông, đô thị của địa phương được xây dựng, hiện đại hóa và các ứng dụng công nghệ mới trong lĩnh vực y tế, giáo dục, giao thông, văn hóa, doanh nghiệp... sẽ góp phần giảm thiểu công sức con người, hướng đến phát triển kinh tế và xã hội (Trang 10).
3. Lộ trình triển khai đề án được phân thành 3 giai đoạn từ 2020 đến 2025, bao gồm việc xây dựng trung tâm Xử lý hình ảnh, giám sát an ninh trật tự, giao thông công cộng thông minh qua hệ thống camera, xây dựng Trung tâm điều hành đô thị thông minh của Quận 10, xây dựng cơ sở dữ liệu hộ tịch, và nâng cấp các ứng dụng công nghệ thông tin nhằm tăng cường kết nối giữa các cơ quan nhà nước (Trang 11-12).
4. Các cơ quan liên quan như Ban Điều hành Đề án, Văn phòng Hội đồng nhân dân và Ủy ban nhân dân quận, Phòng Tư pháp, và Công an quận sẽ chịu trách nhiệm trong việc thực hiện và quản lý Đề án (Trang 12).

6. Phân loại văn bản (text classification): theo độ khó

DD

Hãy phân loại các văn bản sau theo độ khó (text readability):

Văn bản 1:

Summer in India is hot. And it is getting even hotter.

People are suffering from the heat. Many people die. Schools must close. Many countries, including India, are buying more air conditioners (ACs) to cool down. ACs are good for health and work. But they are bad for the natural world. They use bad chemicals. ACs with bad chemicals can make the world even hotter.

Rich countries have strict rules to make ACs better for the natural world. But poor countries cannot afford those ACs. It is a difficult problem to solve.



Dựa trên từ vựng, cấu trúc câu và chi tiết được cung cấp, ta có thể phân loại độ khó đọc của các văn bản như sau:

1. **Văn bản 1:** Dễ - Văn bản này sử dụng ngôn ngữ đơn giản, câu ngắn và dễ hiểu. Điều này thích hợp cho người mới học tiếng Anh hoặc trẻ em.
2. **Văn bản 3:** Trung bình - Văn bản này sử dụng từ vựng và cấu trúc câu phức tạp hơn so với Văn bản 1, nhưng không quá khó so với Văn bản 2. Có sự cân nhắc giữa thông tin chi tiết và độ phức tạp của ngôn ngữ.
3. **Văn bản 2:** Khó - Văn bản này chứa nhiều thông tin chi tiết, sử dụng từ vựng và cấu trúc câu phức tạp. Nó đòi hỏi độc giả phải có kiến thức về tiếng Anh và các vấn đề liên quan đến môi trường để hiểu đúng.

6. Phân loại: theo cảm xúc (<https://duongdaynong.tphcm.gov.vn/trang-chu#>)

DU

Hãy phân loại các đánh giá sau là tích cực hay tiêu cực:

1. Cán bộ trực ban giao tiếp rất khó chịu, giọng điệu hằn học. Gần 14h chiều ngày thứ ba nhưng bàn trực ban không có ai, 3-4 người dân phải ngồi đợi.
2. Tại địa chỉ số 215/36C, bãi xe Nam, Nguyễn Xí, Phường 13, Quận Bình Thạnh, nhà lấn chiếm hết hẻm làm bãi giữ xe, xây nhà và trổ cửa trái phép lấn ra sông Vầm Thuật.
3. Việc đốt (nhóm) bếp than vào mỗi sáng sớm từ 3 đến 4 giờ sáng của hộ gia đình 144/10 đường Lê Văn Thọ, phường 11, quận Gò Vấp rất độc hại. Gây ảnh hưởng âm thầm đến sức khỏe của rất nhiều người sinh sống xung quanh.
4. Về việc gần tết, các cán bộ không hề lơ là trong công việc và luôn có mặt ở cơ quan trong giờ làm việc hành chánh để hỗ trợ người dân.
5. 133 Nguyễn Tiểu La, Phường 5 Quận 10, nhà cho thuê để các bao xi măng, bao cát, đá lấn chiếm lòng lề đường mấy năm nay. Người dân có liên hệ UBND Phường nhưng không được giải quyết triệt để.



1. Tiêu cực: Cán bộ trực ban không được hỗ trợ và cách hành xử của họ rất khó chịu.
2. Tiêu cực: Ngôi nhà lấn chiếm hẻm, làm bãi giữ xe, xây nhà và trổ cửa trái phép, lấn ra sông Vầm Thuật.
3. Tiêu cực: Hộ gia đình đốt (nhóm) bếp than vào mỗi sáng sớm từ 3 đến 4 giờ sáng, gây ảnh hưởng đến sức khỏe của nhiều người.
4. Tích cực: Các cán bộ không lơ là trong công việc, luôn có mặt ở cơ quan và hỗ trợ người dân trong giờ làm việc hành chánh.
5. Tiêu cực: Nhà cho thuê bị lấn chiếm lòng lề đường mấy năm nay bởi các bao xi măng, bao cát, đá và việc này chưa được giải quyết triệt để bởi UBND Phường.

7. Sửa lỗi văn bản: kiểm tra chính tả, ngữ pháp, viết lại,..

Neural machine translation (NMT) is the deep learning based approach for machine translation, which yields a state-of-the-art translation performance in scenarios which large-scale parallel corpora are available. Although the high-quality and domain-specific translation is crucial in the real world, but domain-specific corpora are usually scarce or nonexistent, and thus vanilla NMT performs poorly in such scenarios. Domain adaptations that leverages both out-of-domain parallel corpora as well as monolingual corpora for in-domain translation, very important for domain-specific translation. In this research aims to improve performance of domain adaptation models by exploring the use of composite index methods. Specifically, the research focuses in optimizing nearest neighbor search for natural language processing tasks in domain adaptation.

Neural machine translation (NMT) is a deep learning-based approach to machine translation, which yields state-of-the-art translation performance in scenarios where large-scale parallel corpora are available. Although high-quality and domain-specific translation is crucial in the real world, domain-specific corpora are often scarce or nonexistent, leading to poor performance of vanilla NMT in such scenarios. Domain adaptation, which leverages both out-of-domain parallel corpora and in-domain monolingual corpora, is essential for domain-specific translation. This research aims to improve the performance of domain adaptation models by exploring the use of composite index methods. Specifically, the research focuses on optimizing nearest neighbor search for natural language processing tasks in domain adaptation.

8. Phân tích dữ liệu: phân tích số liệu

Tôi đang nghiên cứu sự tương quan giữa độ khó văn bản. Bạn có thể giúp tôi phân tích và giải thích mối quan hệ này không?

Độ khó của văn bản | Độ dài trung bình câu theo từ | Tỉ lệ câu đơn | Trung bình độ cao cây cú pháp | Tỉ lệ từ đa tiết | Tỉ lệ từ đa nghĩa

2 | 823.1 | 0.868501529 | 5.220183486 | 230.9844904 | 1032.050222

1 | 1050 | 0.9 | 5.4 | 6.057142857 | 39.47619048

2 | 2494.118 | 0.529411765 | 8.176470588 | 31.27102804 | 145.8551402

1 | 1190.476 | 0.761904762 | 6.476190476 | 18.96825397 | 87.24206349

1 | 1511.111 | 1 | 6.777777778 | 18.43065693 | 47.81021898

1 | 1608.333 | 0.583333333 | 7.583333333 | 28.96315789 | 63.12631579

2 | 1952.941 | 0.764705882 | 7.647058824 | 52.42168675 | 123.3222892

Để phân tích mối quan hệ giữa độ khó văn bản và các yếu tố khác, chúng ta có thể xem xét các chỉ số trong bảng dữ liệu mà bạn cung cấp. Dưới đây là một phân tích tương quan sơ bộ:

- Độ dài trung bình câu theo từ: Độ dài trung bình của câu theo số từ trong văn bản. Dựa trên dữ liệu, không có một mối tương quan rõ ràng giữa độ dài câu và độ khó văn bản. Các mẫu có cùng mức độ khó (độ khó 1, 2 hoặc 3) có thể có độ dài câu khác nhau.
 - Tỉ lệ câu đơn: Tỉ lệ câu đơn là tỉ lệ của các câu chỉ có một động từ chính. Dữ liệu không cho thấy một mối quan hệ rõ ràng giữa tỉ lệ câu đơn và độ khó văn bản.
 - Trung bình độ cao cây cú pháp: Đây là trung bình số lượng cây cú pháp trong mỗi câu. Không có mối quan hệ rõ ràng giữa trung bình độ cao cây cú pháp và độ khó văn bản trong dữ liệu của bạn.
 - Tỉ lệ từ đa tiết: Tỉ lệ từ đa tiết đo lường tần suất của các từ có nhiều hơn một tiết âm. Dữ liệu cho thấy mối quan hệ tương đối giữa tỉ lệ từ đa tiết và độ khó văn bản. Mẫu với độ khó 2 có tỉ lệ từ đa tiết cao hơn so với mẫu với độ khó 1. Tuy nhiên, độ khó 3 không có tỉ lệ từ đa tiết cao như độ khó 2.
 - Tỉ lệ từ đa nghĩa: Tỉ lệ từ đa nghĩa đo lường tần suất của các từ có nhiều hơn một nghĩa. Dữ liệu cho thấy mối quan hệ tương đối giữa tỉ lệ từ đa nghĩa và độ khó văn bản. Mẫu với độ khó 2 có tỉ lệ từ đa nghĩa cao hơn so với mẫu với độ khó 1, và độ khó 3 có tỉ lệ từ đa nghĩa cao hơn so với cả độ khó 1 và 2.
- Tóm lại, dữ liệu của bạn cho thấy rằng tỉ lệ từ đa tiết và từ đa nghĩa có mối quan hệ với độ khó văn bản. Tuy nhiên, các chỉ số khác như độ dài câu, tỉ lệ câu đơn và trung bình độ cao cây cú pháp không có mối tương quan rõ ràng với độ khó văn bản. Cần thêm nghiên cứu để xác định các yếu tố khác có thể ảnh hưởng đến độ khó văn bản một cách chính xác hơn.

8. Phân tích dữ liệu: phân tích số liệu (offline)

SWC-009423MT

PHIẾU KẾT QUẢ KIỂM NGHIỆM



28/03/2023

II. KẾT QUẢ THỬ NGHIỆM

TT	Tên chỉ tiêu	Đơn vị	Phương pháp thử	Giới hạn phát hiện	Quy chuẩn 01- 1:2018/BY T	Kết quả thử nghiệm		
						SWC- 009423MT/ 0537	SWC- 009423MT/ 0538	SWC- 009423MT/ 0539
1	Độ màu	Pt-Co	Hach method 8025	-	≤ 15	0	0	0
2	Mùi	-	SMEWW 2130B:2017	-	Không có mùi lạ	Không có mùi	Không có mùi	Không có mùi
3	Độ đục (*)	NTU	SMEWW 2130B:2017	0.035 NTU	≤ 2	<0.25	<0.25	<0.25
4	pH (*)		TCVN					
5	Coliform							
6	E. Coli							
7	Clo dư							

Dựa vào bảng báo cáo chất lượng nước mà bạn đã cung cấp, dưới đây là đánh giá chất lượng nước cho từng chỉ tiêu:

- Độ màu (Pt-Co): Phương pháp thử Hach. Giới hạn phát hiện theo SWC-01 là 15. Kết quả thử nghiệm của bạn là 0, 0 và 0, thể hiện rằng mẫu nước không có màu.
- Mùi: Phương pháp thử SMEWW. Kết quả cho thấy không có mùi lạ (Không có) cho cả ba mẫu nước.
- Độ đục (NTU): Phương pháp thử SMEWW. Giới hạn phát hiện theo SWC-0537 là 0.035 NTU. Kết quả thử nghiệm cho thấy độ đục của nước trong giới hạn cho phép (<0.25 NTU) cho cả ba mẫu nước.
- Độ pH: Phương pháp thử TCVN 6492:2011 và TCVN 6187-1:2009. Giới hạn tối đa theo TCVN 6492:2011 là 6.0 - 8.5. Kết quả thử nghiệm cho thấy độ pH của nước là 7.2 cho cả ba mẫu nước, nằm trong khoảng cho phép.
- Coliform CFU/100mL: Phương pháp thử TCVN 6187-1:2009 và ISO 9308-1:2014. Kết quả thử nghiệm cho thấy mẫu nước không có Coliform (KPH) cho cả ba mẫu.

8. Phân tích dữ liệu: phân tích số liệu (online)

DD

Hãy phân tích số liệu sản xuất công nghiệp trên địa bàn quận Tân Phú trong năm 2022



Finished browsing



Dựa vào các số liệu tìm được, sản xuất công nghiệp tại quận Tân Phú trong năm 2022 đã có sự phục hồi và phát triển, tuy nhiên cũng gặp một số khó khăn từ tháng 10 năm 2022 với xu hướng tăng trưởng chậm lại. Chỉ số sản xuất công nghiệp (IIP) ước cả năm 2022 tăng 7,8% so với cùng kỳ năm trước, chủ yếu do ngành công nghiệp chế biến, chế tạo [\[9thsource\]](#).

Từ tháng 2 đến tháng 9 năm 2022, sản xuất công nghiệp đã có một chu kỳ tăng trưởng với IIP toàn ngành công nghiệp tám tháng liên tục tăng cao. Tuy nhiên, từ tháng 10/2022, sản xuất công nghiệp có xu hướng tăng trưởng chậm lại, với IIP toàn ngành công nghiệp ước cho quý IV/2022 chỉ tăng 3,0% [\[10thsource\]](#).

9. Siêu văn bản: hình ảnh, âm thanh, video

DD

Dựa vào bài giới thiệu tóm tắt về hoạt động nghiên cứu khoa học của Hội Sinh viên Việt Nam, hãy xây dựng kịch bản cho video cỡ 90 giây



Dưới đây là một kịch bản cho video cỡ 90 giây giới thiệu về hoạt động nghiên cứu khoa học của Hội Sinh viên Việt Nam:

[Màn hình mở đầu với logo Hội Sinh viên Việt Nam, tiếng nhạc dịu dàng bắt đầu]

Giọng đọc (Voice-over): "Hội Sinh viên Việt Nam, nơi khơi nguồn sức trẻ, tài năng và đam mê khoa học."

[Hiện hình ảnh các sinh viên đang nghiên cứu trong phòng thí nghiệm, thư viện, và trên máy tính]

Giọng đọc (Voice-over): "Chúng tôi tạo ra một môi trường tạo điều kiện cho tất cả sinh viên Việt Nam nghiên cứu, khám phá và phát triển."

[Chuyển sang hình ảnh các sinh viên đang thảo luận, cùng nhau giải quyết các vấn đề]

Giọng đọc (Voice-over): "Từ các nhóm thảo luận sôi động, các dự án khoa học thực tế đến việc tạo lập mạng lưới kết nối cùng các nhà nghiên cứu hàng đầu."

Video clip: sản phẩm của ChatGPT và AI

VIDEO MADE WITH
PIC
Nhân Dân tv



Đ/c. Nguyễn Minh Triết
Bí thư T.U Đoàn
Chủ tịch Hội Sinh viên Việt Nam

Ngôn ngữ con người: Chức năng => Ứng dụng

- 1. Dịch vụ khách hàng
- 2. Giáo dục
- 3. **Nghiên cứu KH**
- 4. Sáng tạo nội dung
- 5. Trợ lý cá nhân
- 6. Quản lý Nhân sự
- 7. Chăm sóc sức khỏe
- 8. Giải trí – Trò chơi
- 9. Thương mại điện tử
- 10. Tài chính & Ngân hàng
- 11. Dịch thuật ngôn ngữ
- 12. Quy định & Pháp luật

Vô vàn Ứng dụng

Ứng dụng AI trong từ điển học: độ khó văn bản



Ứng dụng AI trong tính toán độ khó của văn bản (Text readability)

- ❖ Từ điển Oxford Anh-Việt:
- ✓ Chỉ sử dụng 3.000 từ phổ biến nhất trong tiếng Anh
- ✓ Dùng cấu trúc câu đơn giản

phil·an·throp·ist /fɪ'lænθrəpɪst/ noun a rich person who helps the poor and those in need, especially by giving money •nhà từ thiện, mạnh thường quân

- ❖ Trong một từ điển tiếng Việt:
 - Định nghĩa từ “đường” (sugar): “một hợp chất kết tinh có vị ngọt” (“hợp chất” = compound, “kết tinh” = crystallize).
 - SGK lớp 1: “gà qué (35.216)”, “con ngóe (23.670)”

Từ điển tàn số

Rank	Word	f
1	the	1.3712
2	of	1.7254
3	be	1.7322
4	and	1.8024
5	a	1.8101
6	to	1.8224
7	in	1.9179
8	have	2.1294
9	it	2.1578
10	that	2.1992
11	he	2.2107
12	you	2.2921
13	on	2.3084
14	with	2.3312
15	for	2.3710
16	his	2.4079
17	as	2.4162
18	at	2.4193
19	do	2.4428
20	not	2.5276

号	字	f
1	的	1.4198
2	是	1.7300
3	不	1.7832
4	我	1.8229
5	一	1.8300
6	有	1.8756
7	大	1.9585
8	在	2.0009
9	人	2.0315
10	了	2.0571
11	中	2.1139
12	到	2.1177
13	资	2.1878
14	要	2.2207
15	以	2.2375
16	可	2.2384
17	这	2.2413
18	个	2.2650
19	你	2.2737
20	会	2.2836

Sđt	Từ	f
1	và	1.8199
2	của	1.8223
3	có	1.9560
4	các	1.9591
5	là	1.9682
6	một	1.9861
7	được	2.0125
8	không	2.0317
9	trong	2.0433
10	cho	2.0504
11	đã	2.0826
12	những	2.1467
13	với	2.1483
14	người	2.1599
15	ở	2.2107
16	để	2.2750
17	này	2.2809
18	đến	2.3032
19	vào	2.3090
20	tôi	2.3140

Ứng dụng AI trong dịch tự động chữ Nôm

Chuyển tự từ chữ Nôm sang chữ Quốc ngữ

- “năm” (5): 卯 = 南 (sound /nán/ “nam”) + 五 (meaning: *ngũ*)
 - “năm” (year): 卯 = 南 (sound /nán/) + 年 (meaning: *niên*)
 - “中” => “đúng, trong, trung, trúng, truồng”
 - 穢 = “trăm/trặn” ;



Hàn Thuyên



Alexandre de Rhodes

Dịch tự động văn bản Hán-Nôm sang chữ Quốc ngữ

Truyện Kiều

慕辭沖塲得些

Trăm năm trong cõi người ta

字才字命窖羅怙饒

Chữ tài chữ mệnh khéo là ghét nhau

浪辭嘉靖朝明

Rằng: Năm Gia Tinh triều Minh

羣方滂朗台京凭傍

Bốn phương phảng lặng hai kinh vững vàng

固茹員外戶王

Có nhà viên ngoại họ Vương

家資擬拱常常塲中

Gia tư nghỉ^{đỏ} cũng thường thường bậc trung



Nguyen Du
(1766-1820)

- *nghĩ* (think)
- *nghỉ* (he)



Nghiệm thu hệ thống tự động chuyển văn bản chữ Nôm sang chữ Quốc ngữ

Ý KIẾN BẠN ĐỌC(2)

① 14/02/2023 - 19:33

Chia sẻ























































































































































































































































































































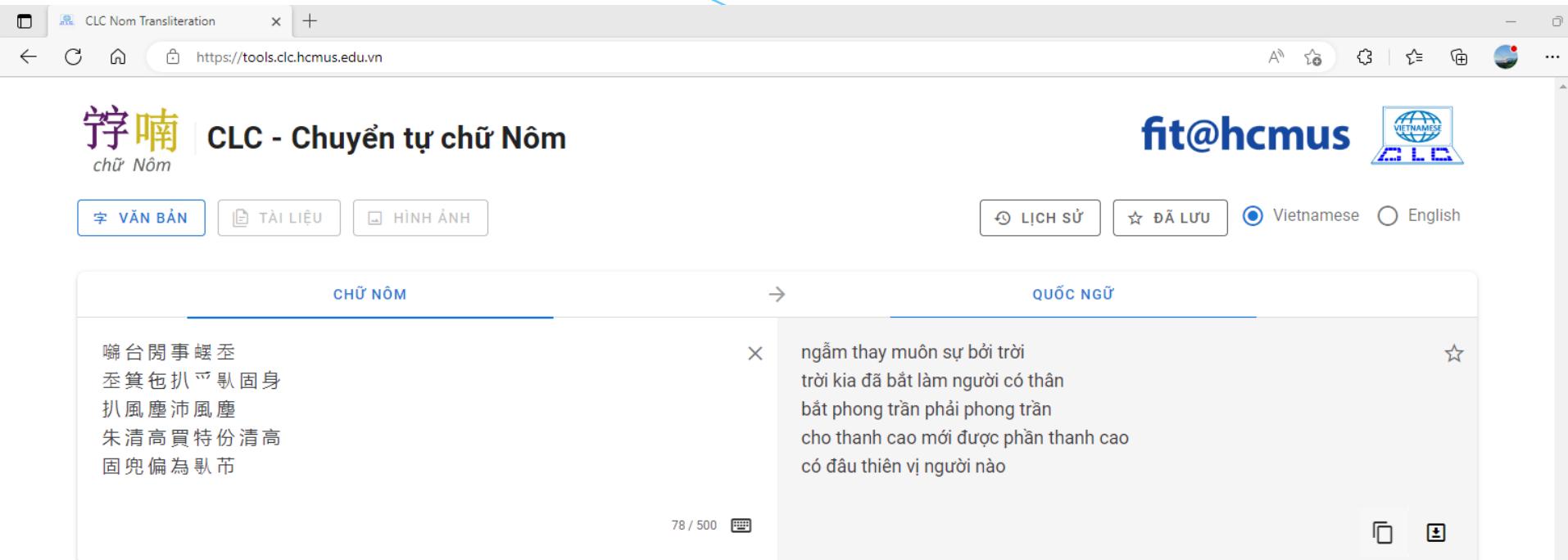









Website: <https://tools.clc.hcmus.edu.vn>



Công trình này được thực hiện bởi một số giảng viên Khoa Công nghệ Thông tin và Trung tâm Ngôn ngữ học Tính toán thuộc Trường ĐH Khoa học Tự nhiên - ĐHQG - HCM trong khuôn khổ đề tài nghiên cứu khoa học của Sở Khoa học & Công nghệ TpHCM



Đang thử nghiệm: Dịch chữ Nôm từ hình ảnh

The screenshot shows a web browser window with the title "NOM OCR BETA VERSION - HCM". The address bar displays "localhost:8000". The main content area contains two rows of Chinese characters arranged in a grid. Below each row is a Vietnamese translation.

Row 1:

爛	花	雲	枚	頭	沒	浪	稿	選	候	解	解
強	咲	祐	骨	懸	殊	蔚	蒼	之	戈	脫	脫
色	王	莊	格	企	狃	員	若	彼	沒	撲	撲
稍	說	重	雪	次	翟	外	精	齊	局	局	些
漫	端	恪	情	卒	次	戶	朝	斯	漬	局	些
麻	莊	鴻	神	娥	王	明	烟	豐	機	局	些

越南翻译:

- 森辭脫撲馴些 - trâm nâm trong cõi người ta
- 笄才笄命害罵恬饒 - chữ tài chữ mệnh khéo là ghét nhau
- 廸戈沒局漬機 - trái qua một cuộc bê dâu
- 仍調體貸厄疚悉 - những điều trông thấy đã đau đớn lòng
- 遇之彼齊斯豐 - lạ gì bi sắc tư phong
- 至擯涓貝鵝紅打惺 - trời xanh quen với má hồng đánh ghen

Row 2:

揚	運	困	沒	翠	王	眾	風	至	仍	笄	笄
皮	輸	賤	駟	翫	兒	資	方	擇	調	才	命
才	浩	若	沒	哭	𠂇	模	傍	消	離	笄	害
色	還	憚	駢	姊	笄	模	固	貝	覺	命	罵
更	雪	譏	辯	姪	絞	綱	錄	鵝	厄	打	恬
分	談	惺	辯	𡇠	絞	常	群	紅	方	惺	饒
欣	半	妾	近	院	𡇠	又	傳	打	疽	打	饒
	脣	囊	近	翠	雲	中	史	惺	疽	惺	饒

越南翻译:

- 廸戈沒局漬機 - trái qua một cuộc bê dâu
- 仍調體貸厄疚悉 - những điều trông thấy đã đau đớn lòng
- 遇之彼齊斯豐 - lạ gì bi sắc tư phong
- 至擯涓貝鵝紅打惺 - trời xanh quen với má hồng đánh ghen

Bottom controls:

- Choose File
- Upload and Send



Công trình này được thực hiện bởi một số giảng viên Khoa Công nghệ Thông tin và Trung tâm Ngôn ngữ học Tính toán thuộc Trường ĐH Khoa học Tự nhiên - ĐHQG - HCM trong khuôn khổ đề tài nghiên cứu khoa học của Sở Khoa học & Công nghệ TpHCM



KÝ HIỆU CHỮ BRAILLE VIỆT NGỮ

Dịch tự động văn bản chữ sáng (printed) sang chữ nổi (braille) cho người khiếm thị

Trung tâm dữ liệu đa ngữ

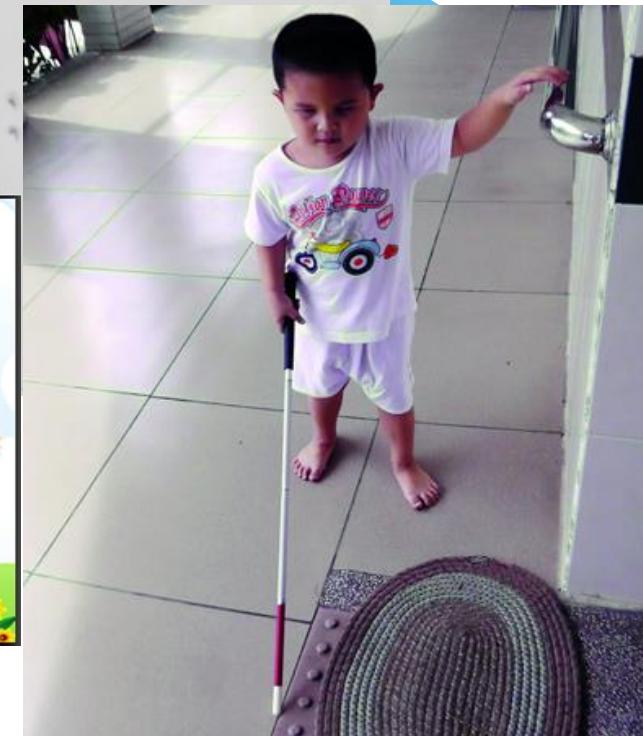
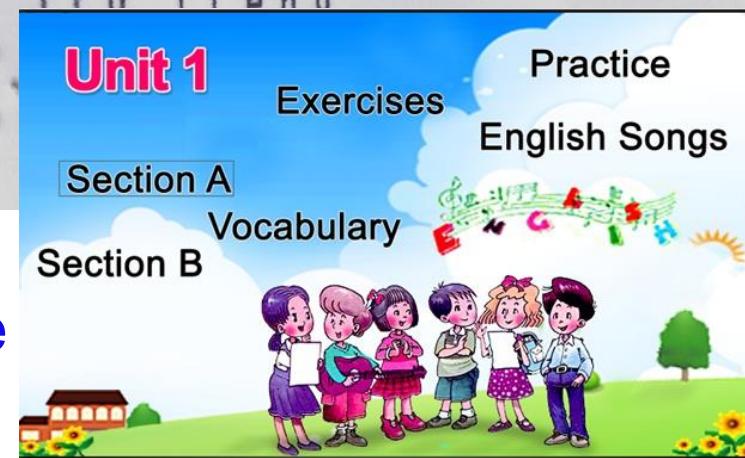
Kim Tự Điện (KMDC) chúng

tôi chuyên sản xuất các

Phần mềm có hỗ trợ tiếng

Việt cho người

Phần mềm học
ngoại ngữ cho trẻ
em khiếm thị



Download: www.clc.hcmus.edu.vn

Cách đặt câu hỏi/yêu cầu (Prompt Engineering)

- ❑ Chất lượng/Cách thức câu hỏi/yêu cầu sẽ ảnh hưởng rất lớn đến chất lượng câu trả lời/kết quả.
- ❑ Yêu cầu từ tổng quát đến chi tiết (top-down), từ dàn ý (outline) đến các mục nhỏ hơn => Nội dung cụ thể.
- ❑ Quá trình “chia - để - trị” => có khả năng xử lý văn bản rất lớn (vd. viết cả cuốn sách; tóm tắt cả một nghị định dài, ...).
- Áp dụng P.E. trong tất cả các chức năng của ChatGPT:
- ✓ Vd: Translation <> Google Dịch thuật:
 - ✓ Ngữ cảnh
 - ✓ Miền và chủ đề
 - ✓ Phong cách/Định dạng/Cấu trúc/Chiến lược
 - ✓ Đối tượng (reader)
 - ✓ Trình độ giáo dục/Độ phức tạp/Độ khó văn bản (readability)

3. Hạn chế của ChatGPT

- Với ChatGPT 3.5: Dữ liệu chỉ được cập nhật đến tháng 9/2021.
- Nhưng những phiên bản sau: Cho phép truy cập Internet
- Trả lời sai: nhất là những vấn đề mang tính đặc thù (đất nước, dân tộc, văn hóa, lịch sử, chính trị, pháp luật, ngôn ngữ, ...) vì không thiếu dữ liệu huấn luyện (vd: chữ Nôm).
- Trả lời sai nhưng rất thuyết phục: “ảo giác” (hallucination)
- Trả lời thiên vị: Chủng tộc, giới tính, tôn giáo, ...
- Hiện chỉ là văn bản (text): Chưa có tiếng nói (voice)
- Phiên bản ChatGPT 4.0: Hiểu hình ảnh, “thông minh” hơn.

Không chính xác: lịch sử

NG

Quang Trung và Nguyễn Huệ là một người



Không, Quang Trung và Nguyễn Huệ là hai nhà lãnh đạo của Việt Nam trong lịch sử, hoạt động trong thời kỳ khác nhau. Quang Trung là vị vua thứ hai của nước Việt trong thời kỳ Lý, còn Nguyễn Huệ là một trong những nhà lãnh đạo chính trị và quân sự trong thời kỳ Tây Sơn.



DD

Hải đội nào được cử ra Hoàng Sa sớm nhất?



Trong lịch sử, Hải đội 146 của Việt Nam là đơn vị được cử ra Hoàng Sa sớm nhất được biết đến. Hải đội 146 được thành lập vào năm 1956 và được giao nhiệm vụ thám hiểm, bảo vệ chủ quyền trên các đảo Hoàng Sa và Trường Sa của Việt Nam. Tuy nhiên, trước đó cũng có nhiều đơn vị thuộc Việt Nam đã thăm dò, khám phá, đánh dấu chủ quyền tại khu vực này từ rất lâu trước đó.

~~Không chính xác: ngữ nghĩa từ Hán-Việt~~

DD

Hãy giải thích nghĩa của chữ "giai" trong thành ngữ "bách niên giai lão"



Trong thành ngữ "bách niên giai lão", chữ "giai" có nghĩa là "già nua", "đã già". "Bách niên" nghĩa là "trăm tuổi", "lão" có nghĩa là "người già". Vì vậy, "bách niên giai lão" được dịch là "trăm tuổi đã già nua" hoặc "trăm tuổi đời đã già".

Thành ngữ này thường được dùng để miêu tả một người đã sống đến tuổi rất cao, đánh giá đó là sự kiện đáng kính và đáng quý. Nó thể hiện sự tôn trọng và ngưỡng mộ đối với người già, đồng thời cũng nhắc nhở về sự quan trọng của sức khỏe, thời gian và đời sống. Ngoài ra, thành ngữ này còn có ý nghĩa đề cao lòng kiên trì, nghị lực, sức mạnh và tinh thần của con người, vì để sống đến độ tuổi trăm tuổi, cần phải có những phẩm chất đó.

giai (階): *giai cấp, giai đoạn, ...*

"bậc thềm"

giai (佳): *giai điệu, giai thoại, giai nhân, ...*

"hay/đẹp"

giai (偕): *giai lão, ...*

"cùng nhau"

~~Không chính xác: cả ngũ nghĩa từ thông dụng~~

Model: GPT-4

Talk-t

DD

Hãy giải thích ý nghĩa của chữ "trắc" trong từ "bất trắc"



Trong tiếng Việt, từ "bất trắc" thường được dùng để mô tả một tình huống không ổn định, khó lường, rủi ro hoặc gặp phải những khó khăn và vấn đề. Từ này bao gồm hai chữ "bất" và "trắc".

Chữ "trắc" trong từ "bất trắc" có nghĩa là **không ổn định, không chắc chắn**. Khi kết hợp với chữ "bất", nó nhấn mạnh ý nghĩa của sự không ổn định, rủi ro hoặc tình trạng khó lường. Ví dụ, bạn có thể dùng từ "bất trắc" để mô tả tình hình kinh tế, chính trị, thời tiết, hoặc một mối quan hệ nào đó.

Tóm lại, ý nghĩa của chữ "trắc" trong từ "bất trắc" là sự không ổn định, không chắc chắn, phản ánh một tình huống không tốt hoặc khó khăn.

Trắc: đo/lường: **trắc nghiệm, quan trắc, sinh trắc, ...**
“Sông sâu khó đo, lòng người không lường được”

Chưa thể hiểu: nghĩa hàm ý, ẩn dụ, hoán dụ

2. "Thiều quang chín chục đã ngoài sáu mươi": "Thiều quang" có nghĩa là ánh sáng mờ nhạt, chỉ thời gian gần hoàng hôn. Số "chín chục" ở đây ám chỉ tuổi thọ trung bình của con người, còn "sáu mươi" là thời gian mà ánh sáng bắt đầu mờ dần, tượng trưng cho tuổi già. Như vậy, câu thơ này có thể hiểu là thời gian cuộc đời đã đi qua nhanh chóng, giống như cách mà ánh sáng của ngày chuyển mình vào hoàng hôn.

Tổng kết lại, hai câu thơ này mang ý nghĩa về sự vụt qua của thời gian, sự thay đổi từ mùa xuân sang hoàng hôn tượng trưng cho quá trình từ thanh xuân đi vào tuổi già. Đồng thời, chúng cũng nêu lên một cảm giác nuối tiếc về những thời gian đã qua nhưng vẫn phải đón nhận và chấp nhận sự thay đổi của cuộc sống.

*Ngày xuân con én đưa thoi => dệt nên Tấm lụa mùa Xuân (90 ngày)
60 ngày = đã qua 2 tháng => Thanh minh trong tiết Tháng Ba*

Không thể hiểu: Nghĩa hình tượng

2. "Phận liễu sao mà nảy nét ngang."

- "Phận liễu" là số phận tầm thường, không có gì nổi bật, giống như cây liễu nhỏ bé và mong manh.
- "Sao mà" là "làm sao" hoặc "làm cách nào".
- "Nảy nét ngang" là phát triển và đạt được sự kiên cường, mạnh mẽ trong cuộc sống.

Câu thơ này thể hiện sự trăn trở của tác giả về số phận của mình, liệu với một số phận tầm thường như cây liễu, có thể kiên cường và vượt qua khó khăn, thử thách trong cuộc sống hay không.

Nhìn chung, hai câu thơ này diễn đạt nỗi lòng của tác giả về sự trăn trở, tiếc nuối về số phận và duyên nợ trong cuộc đời. Tác giả nhấn mạnh rằng mỗi con người phải đối mặt với những khó khăn và thử thách trong cuộc sống, dù số phận có tốt hay không. Đồng thời, tác giả cũng đặt câu hỏi liệu có thể vượt qua những thử thách đó để trở nên kiên cường và mạnh mẽ hơn.

Đặc thù tiếng Việt: không hiểu chữ Nôm

NG

GPT có thể dịch giúp tôi câu "慕辭韙塲𠂇些" không



Rất vui làm việc cho bạn! Câu "慕辭韙塲𠂇些" trong chữ Nôm có thể dịch như sau: "Hãy cho tôi một vài."

Truyện Kiều: “*Trăm năm trong cõi người ta*”

Kết quả: “*Hãy cho tôi một vài.*” !?

Tiếng Anh: chính xác hơn

DD

Xin cho tôi biết nguồn gốc của từ "breakfast"



Từ "breakfast" được hình thành từ hai từ "break" và "fast". Trong tiếng Anh cổ, "to break" có nghĩa là "phá vỡ, làm gián đoạn", và "fast" có nghĩa là "đói khát" trong thời gian dài, thường là do tôn giáo hay y học. Vì vậy, "breakfast" có nghĩa là bữa ăn đầu tiên trong ngày, giúp phá vỡ trạng thái đói khát sau một đêm ngủ.

break: phá vỡ **fast:** ăn chay/nhịn ăn (do đêm: ngủ)

breakfast: bữa đầu tiên trong ngày mà phá chay

Tiếng Anh: hiểu rất tốt

DD

Xin cho biết nguồn gốc của từ "disaster" trong tiếng Anh?



Từ "disaster" trong tiếng Anh có nguồn gốc từ tiếng Ý, từ "disastro". "Disastro" được ghép từ hai phần: tiền tố "dis-" mang nghĩa "không tốt" hay "xấu" và "**astro**" nghĩa là "sao" (trong ngữ cảnh thiên văn học). Trong tiếng Ý, "disastro" được sử dụng để chỉ một sự kiện xấu xảy ra do ảnh hưởng của các vì sao, vì người xưa thường tin rằng các thiên thể có ảnh hưởng đến cuộc sống trên Trái đất.

astro/aster: **astronomy, astronaut, asteroid**
disappear: biến mất => **disaster:** tai họa

Từ "disastro" sau đó được chuyển sang tiếng Pháp dưới dạng "désastre" và cuối cùng trở thành "disaster" trong tiếng Anh vào khoảng thế kỷ 16. Ngày nay, từ "disaster" được sử dụng để chỉ một sự kiện đột ngột và khủng khiếp gây hậu quả nghiêm trọng về mặt nhân vật, kinh tế, môi trường hoặc xã hội.

Tiếng Anh: kiến thức cập nhật

DD

Xin cho biết ý nghĩa của thương hiệu NVidia?



NVidia là một công ty đa quốc gia của Mỹ chuyên về thiết kế và sản xuất chip đồ họa. NVidia được biết đến rộng rãi với các sản phẩm của mình trong lĩnh vực đồ họa máy tính, đặc biệt là chip đồ họa cho máy tính xách tay và máy tính để bàn. Họ cũng có mặt trong các lĩnh vực khác như trí tuệ nhân tạo, xe tự hành, và máy chủ.

Thương hiệu NVidia được xem như một biểu tượng của công nghệ tiên tiến và chất lượng cao trong ngành công nghệ thông tin. Nó tượng trưng cho sự đổi mới liên tục, độ tin cậy và hiệu suất mạnh mẽ. Công ty đã tạo ra nhiều công nghệ đột phá trong lĩnh vực đồ họa máy tính và trí tuệ nhân tạo, và thường xuyên đẩy giới hạn của những gì có thể thực hiện với công nghệ.

Tên "NVidia" bắt nguồn từ Latinh "*invidia*", có nghĩa là "*ghen tị*". Nó cũng là một sự chơi chữ với từ "*video*", một chỉ dẫn đến trọng tâm của công ty: đồ họa và video.

Tiếng Anh: hiểu rất sâu rộng

DD

Xin cho tôi biết nguồn gốc của từ "Milky Way"



The creation of the Milky Way in Greek and Roman Mythology!

A large, dark background image of a star-filled sky, representing the Milky Way galaxy. A single meteor streaks across the upper left.

Thần Zeus: Vua các vị thần
Hera: Vợ thần Zeus
Hercule: Con rơi của thần Zeus

Tiếng Anh: câu quá khó => cũng dịch sai

Câu lắt léo (nhập nhằng):

“*Xem lại album, mình thấy ông già đi nhanh quá*”

DA

Hãy dịch câu sau đây sang tiếng Anh: Xem lại album, mình thấy ông già đi nhanh quá



Review the album, I see that the old man walks too fast.

SAI: “*Review the album, I see that the old man walks too fast*”
(ông già / đi nhanh!)

ĐÚNG: “*Review the album, I see that the grandfather gets old too fast*”
(ông / già đi nhanh)

Sự phát triển của ChatGPT & các sản phẩm AI tương tự

- Hạn chế về ngữ liệu của ChatGPT: sẽ được giải quyết
- Hạn chế về tiếng nói: khá đơn giản (ChatGPT4, 10^{14})
- Hiểu rõ hơn về ngữ nghĩa (semantics) văn bản: thông qua ngữ âm (phonetics)
- Phát triển: thêm phần xử lý ảnh (image/video): để nhận diện khuôn mặt, biểu cảm của con người (khả thi)
- Kết nối với người máy (robot): Trợ lý “thật” đắc lực!
- Tuy không thể chính xác tuyệt đối: nhưng rất hữu dụng.
- Phạm vi ứng dụng: theo vào trí tưởng tượng con người.

Các biến thể của ChatGPT: VoiceGPT

The screenshot shows the VoiceGPT website (voicegpt.us/chatgpt) running in a browser. The main header reads "VoiceGPT" with a microphone icon. Below the header, there are tabs for "VoiceGPT", "VoiceNews", "VoiceDALL-e", "Feed", and "Top Inviters". A user profile "DT Dương...hị An" is visible on the right.

The central content area is titled "VoiceGPT" and "Auto-voice mode" (which is turned on). It features three sections: "Examples", "Capabilities", and "Limitations".

- Examples:**
 - "Make Eminem-style jokes about Max Payne."
 - "Make a song about a programmer and someone who isn't a programmer."
 - "Make a list of the best..."
- Capabilities:**
 - Remembers what user said earlier in the conversation
 - Allows user to provide follow-up corrections
 - Trained to decline inappropriate requests
- Limitations:**
 - May occasionally generate incorrect information
 - May occasionally produce harmful instructions or biased content
 - Limited knowledge of world and..."

On the left and right sides of the main content area, there are two large, semi-transparent promotional banners. The left banner is for "Tesse Technology" and the right banner is for "VOICELYAI". Both banners include download links for Google Play and the App Store.

At the bottom of the page, there is a footer message: "VoiceGPT 5 Mar Version. Built on top of ChatGPT & DALL-E. Our goal is to make AI system more natural and safe to interact with. Your feedback will help us improve." and "Made with ❤ by TESSE".

Các sản phẩm AI khác: Bing, Bard, Copilot,...

- Voice-GPT, Visual-GPT, Auto-GPT, ...: biến thể ChatGPT
- Otter, Notta: Ghi biên bản cuộc họp: tóm tắt, slides, email,...
- Copilot: Kết nối ChatGPT với MS Word,...
- Bing: Tích hợp ChatGPT khi tìm kiếm thông tin
- Bard: Sản phẩm Chatbot.AI của Google
- DALL-E: Tạo ảnh từ văn bản
- Mid-Jurnet: Tạo video từ kịch bản
- D-ID: Tạo video người nói từ hình ảnh tĩnh
- Designrr: Tạo sách điện tử từ Blog, Podcast, video, ChatGPT
- ChatSonic: Hỗ trợ nghiên cứu khoa học.

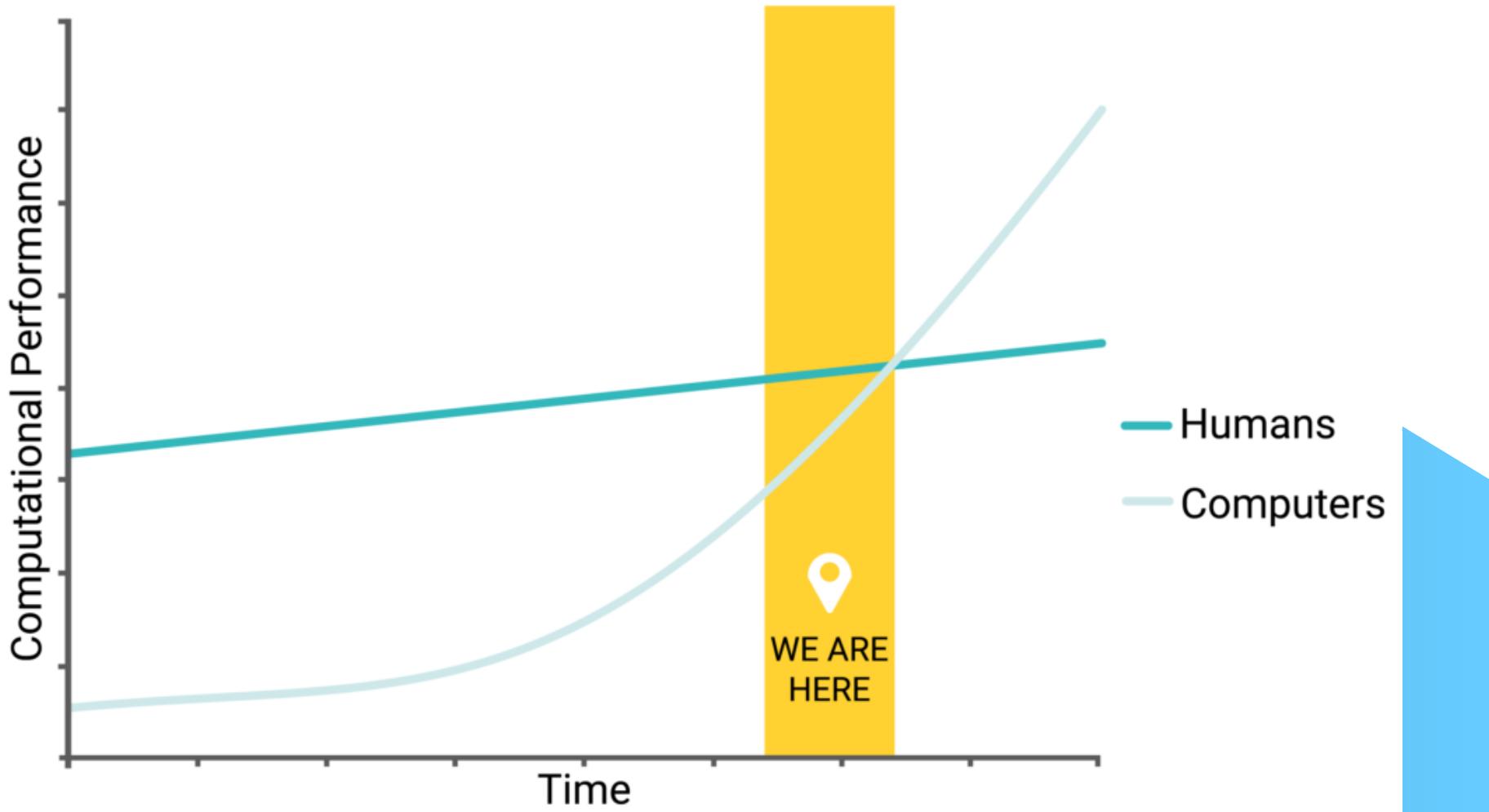
ChatGPT & Con người

- ❑ Câu trả lời của ChatGPT: **có tác vụ không chính xác**
 - ✓ Vì đâu ra phụ thuộc vào: Tác vụ, cách hỏi, lịch sử hỏi, ...
 - Thách thức trong KIỂM TRA/GIÁM SÁT thông tin
- ❑ ChatGPT hay các sản phẩm tương tự (Bard, Erniebot, Bing, Auto-GPT, Voice-GPT, Visual-GPT, Local-GPT,...)
 - ✓ So với Con người, Máy:
 - ❑ Nhanh hơn, **biết nhiều hơn** (do nó học từ hàng trăm triệu sách/tài liệu, bằng **hàng trăm ngôn ngữ**), siêu bộ nhớ.
 - ❑ Tìm kiếm Google: chỉ tìm trong một ngôn ngữ (truy vấn)
 - ChatGPT: đa ngôn ngữ, khó truy nguồn; không biết độ tin cậy.

Cách đặt câu hỏi/yêu cầu (Prompt Engineering)

- ❑ Chất lượng/Cách thức câu hỏi/yêu cầu sẽ ảnh hưởng rất lớn đến chất lượng câu trả lời/kết quả.
- ❑ Yêu cầu từ tổng quát đến chi tiết (top-down), từ dàn ý (outline) đến các mục nhỏ hơn => Nội dung cụ thể.
- ❑ Quá trình “chia - để - trị” => có khả năng xử lý văn bản rất lớn (vd. viết cả cuốn sách; tóm tắt cả một nghị định dài, ...).
- Áp dụng P.E. trong tất cả các chức năng của ChatGPT:
- ✓ Vd: Translation <> Google Dịch thuật:
 - ✓ Ngữ cảnh
 - ✓ Miền và chủ đề
 - ✓ Phong cách/Định dạng/Cấu trúc/Chiến lược
 - ✓ Đối tượng (reader)
 - ✓ Trình độ giáo dục/Độ phức tạp/Độ khó văn bản (readability)

Tốc độ phát triển: Máy & Con người



AI đã hơn con người!

tuoi tre
online

Podcast

YouTube

C



VIDEO

THỜI SỰ

THẾ GIỚI

PHÁP LUẬT

KINH DOANH

CÔNG NGHỆ

XE

DU LỊCH

NHỊP SỐNG TRẺ

VĂN HÓA

GIẢI TRÍ

TI



aa



1



Công nghệ

26/05/2023 21:42 GMT+7

'Đọc' thấu suy nghĩ của con người, AI đã làm được



GIA MINH

▶ Nghe đọc bài 2:13 | 1x | 🔍

Tuổi Trẻ trên Google News

Lần đầu tiên, các nhà nghiên cứu sử dụng GPT1, tiền thân của AI ChatGPT, để dịch hình ảnh MRI thành văn bản nhằm hiểu được suy nghĩ của con người.



AI phát triển không kiểm soát được?

OpenAI kêu gọi kìm hãm trí tuệ nhân tạo

Câu chuyện của tác giả Ngọc Diệp • 8 giờ trước

TP - Đội ngũ đằng sau hệ thống ChatGPT cho biết, cần có một cơ quan giám sát để chống lại rủi ro của những AI “siêu thông minh”.

Trong một ghi chú ngắn đăng trên trang web của công ty, những người đồng sáng lập, ông Greg Brockman và ông Ilya Sutskever cùng giám đốc điều hành, ông Sam Altman, đã nhấn mạnh sự cần thiết của một cơ quan quản lý quốc tế chuyên “kiểm tra hệ thống, giám sát việc tuân thủ các tiêu chuẩn an toàn và đặt ra các hạn chế về mức độ triển khai và mức độ bảo mật” nhằm giảm các “rủi ro” mà các [hệ thống AI](#) có thể gây ra.

“Có thể hình dung rằng trong vòng 10 năm tới, các hệ thống AI sẽ vượt qua trình độ kỹ năng chuyên gia trong hầu hết các lĩnh vực và thực hiện nhiều hoạt động hiệu quả bằng những tập đoàn lớn nhất hiện nay. Xét về mặt tích cực và tiêu cực, siêu trí tuệ sẽ mạnh hơn các công nghệ khác mà nhân loại từng đổi mới trong quá khứ. Chúng ta có thể có một tương lai thịnh vượng hơn nhiều; nhưng chúng ta phải quản lý rủi ro để đạt được điều đó. Với khả năng xảy ra rủi ro hiện hữu, chúng ta cần phải hành động”, họ viết

4. An toàn thông tin

Công nghệ

29/03/2023 20:20 GMT+7

Nguy cơ 'ghê gớm' gì khiến Elon Musk kêu gọi tạm dừng phát triển trí tuệ nhân tạo? 4



NGHI VŨ



Nghe đọc bài 3:12

1x



v

Tuổi Trẻ trên Google News



Trước sự phát triển nhanh như vũ bão của công nghệ trí tuệ nhân tạo (AI), Elon Musk cùng nhiều chuyên gia trong ngành đã kêu gọi 'giảm tốc' để xem xét thêm về các vấn đề đạo đức và pháp lý.

Theo Hãng tin Reuters, tỉ phú Elon Musk cùng một nhóm gồm các chuyên gia về trí tuệ nhân tạo (AI) và các giám đốc công nghệ cấp cao đã cùng ký tên vào một lá thư kêu gọi tạm dừng việc phát triển các hệ thống mạnh hơn bản nâng cấp GPT-4 mới nhất của ChatGPT, do lo ngại các nguy cơ có thể xảy ra cho xã hội và nhân loại.

Nội dung lá thư yêu cầu tạm dừng phát triển trí tuệ nhân tạo cho đến khi có các "giao thức an toàn" chung, được các chuyên gia độc lập thẩm định và giám sát.

Những rủi ro tiềm ẩn:

- TIN TẮC:
 1. Ngụy tạo thông tin có chủ đích
 2. Đưa ra những ý tưởng vô đạo đức
- RÒ RỈ THÔNG TIN:
 1. Chia sẻ thông tin với ChatGPT: cá nhân, tài khoản
 2. Dự đoán hành động dựa vào thông tin theo dõi được.
- LỪA ĐẢO QUA MẠNG:
 1. Giả danh người thân: tiếng nói, hình ảnh, video
 2. Lừa đảo tài chính: chuyển tiền, đánh cắp mật khẩu,...

Công nghệ giả mạo (deepfake):



Công nghệ giả mạo (deepfake):

Vụ Tổng thống Putin bị giả mạo - hé lộ vũ khí mới trong xung đột Nga-Ukraine

QUANG HIẾU 06:20 | 08/06/2023 ★★★★★

Vụ việc Tổng thống Nga Vladimir Putin bị giả mạo hình ảnh, giọng nói bằng công nghệ deepfake mới đây đã dấy lên hồi chuông cảnh báo về loại “vũ khí” mới trong xung đột ở Ukraine.



Video giả, ảnh hưởng thật

Trong bài phát biểu, giọng nói của Tổng thống Nga Vladimir Putin được tạo ra bởi công nghệ deepfake cho biết tình trạng khẩn cấp đã được ban bố ở các vùng Belgorod, Voronezh và Rostov của Nga do các cuộc tấn công sắp tới từ lực lượng Ukraine. Ông Putin giả mạo cũng yêu cầu cư dân trong các khu vực sơ tán khỏi nhà của họ để tìm nơi trú ẩn sâu hơn bên trong nước Nga.

Deepfake là công nghệ sử dụng trí tuệ nhân tạo (AI) để lấy hình ảnh, giọng nói của một người ghép vào video của người khác. Công nghệ deepfake sẽ thu thập hình ảnh khuôn mặt của một đối tượng, sau đó thay thế khuôn mặt này vào mặt của một người khác trong video. Đối với các tập tin âm thanh, deepfake sử dụng bản ghi âm giọng nói của một người thực để huấn luyện máy tính nói chuyện giống hệt người ấy.

Lưu ý:

- Đảm bảo tính bảo mật và riêng tư của dữ liệu người dùng khi sử dụng ChatGPT:
 1. Tránh cung cấp thông tin riêng tư
 2. Không cung cấp thông tin tài khoản ngân hàng cho một ứng dụng chưa được kiểm chứng
 3. Kiểm tra thông tin trước khi tiến hành các bước tiếp theo.

Văn bản của ChatGPT & của Người: khác nhau không?

GPTZero

Publish on Jan 3, 2023

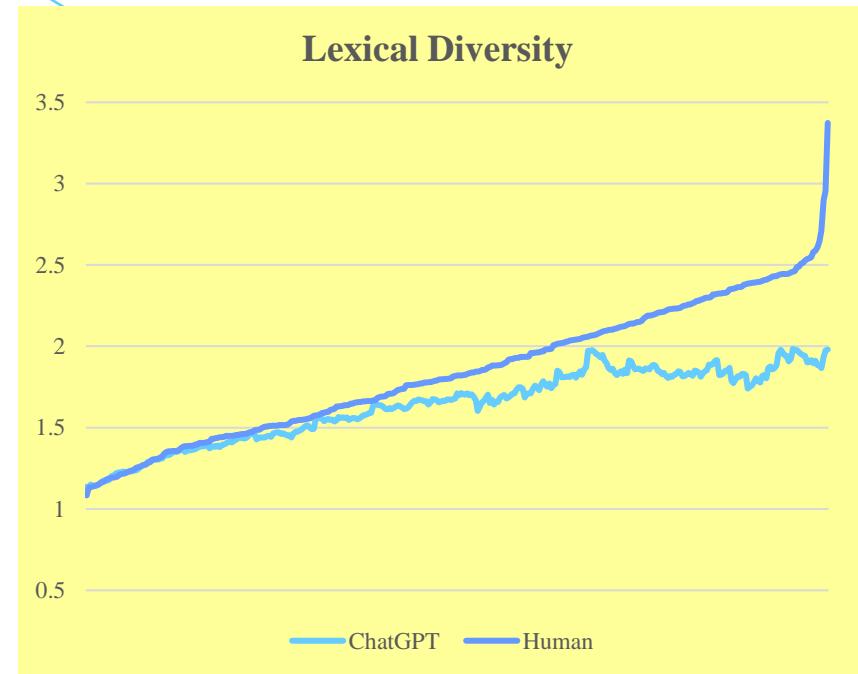
Created by Edward Tian
(Princeton University)

Main idea: calculate two properties of text which are -- “perplexity” and “burstiness”

Web tool: [GPTZero](#)

At least 250 characters

Source code: [Code](#)



❖ Phát hiện tự động văn bản ChatGPT viết được không?

DetectGPT

Publish on Jan 31, 2023

Released by
Stanford University

Main idea: use a language model to detect its own generations.

Web tool: [DetectGPT](#)
At least 40 words.

Source code: [Code](#)

OpenAI
Classifier

Publish on Jan 31, 2023

Released by OpenAI

Main idea: create a classifier to distinguish between text written by a human and text written by AIs.

Web tool: [OpenAI Classifier](#)
At least 1000 characters.

Source: Not Found

Our classifier is not fully reliable.

Hành động

- ✓ Hiểu rõ hơn về chatGPT để tận dụng **CƠ HỘI** ứng dụng.
- ✓ Phải học hỏi nhiều hơn để làm chủ được nó:
- ✓ Lập trường chính trị ổn định, chuyên môn vững chắc
- ✓ Biết nhận ra những sai sót của nó
- ✓ Sẵn sàng đối phó với những **THÁCH THỨC** của nó
- Xa hơn: tự nghiên cứu xây dựng sản phẩm tương tự cho các bài toán đặc thù của VN: hỏi đáp pháp luật VN, truy vấn lịch sử - địa lý VN, y học dân tộc, ... (từ kho tàng chữ Nôm).
- ❖ Khai thác ChatGPT & các AI khác để phục vụ hoạt động **nghiên cứu khoa học**.

Ngành CNTT chúng ta

- ✓ Máy: xử lý ngôn ngữ máy (NN hình thức) dễ hơn ngôn ngữ người
- ✓ Cần tìm hiểu hoạt động bên trong của ChatGPT:
- ✓ Tích hợp API vào các hệ phần mềm hiện nay (Vscode, Office, ...)
- ✓ Tham khảo các mã nguồn (phiên bản GPT trước)
- ✓ Nghiên cứu các công trình khoa học do OpenAI công bố

Publications
(openai.com)

Efficient Training of Language Models to
July 28, 2022

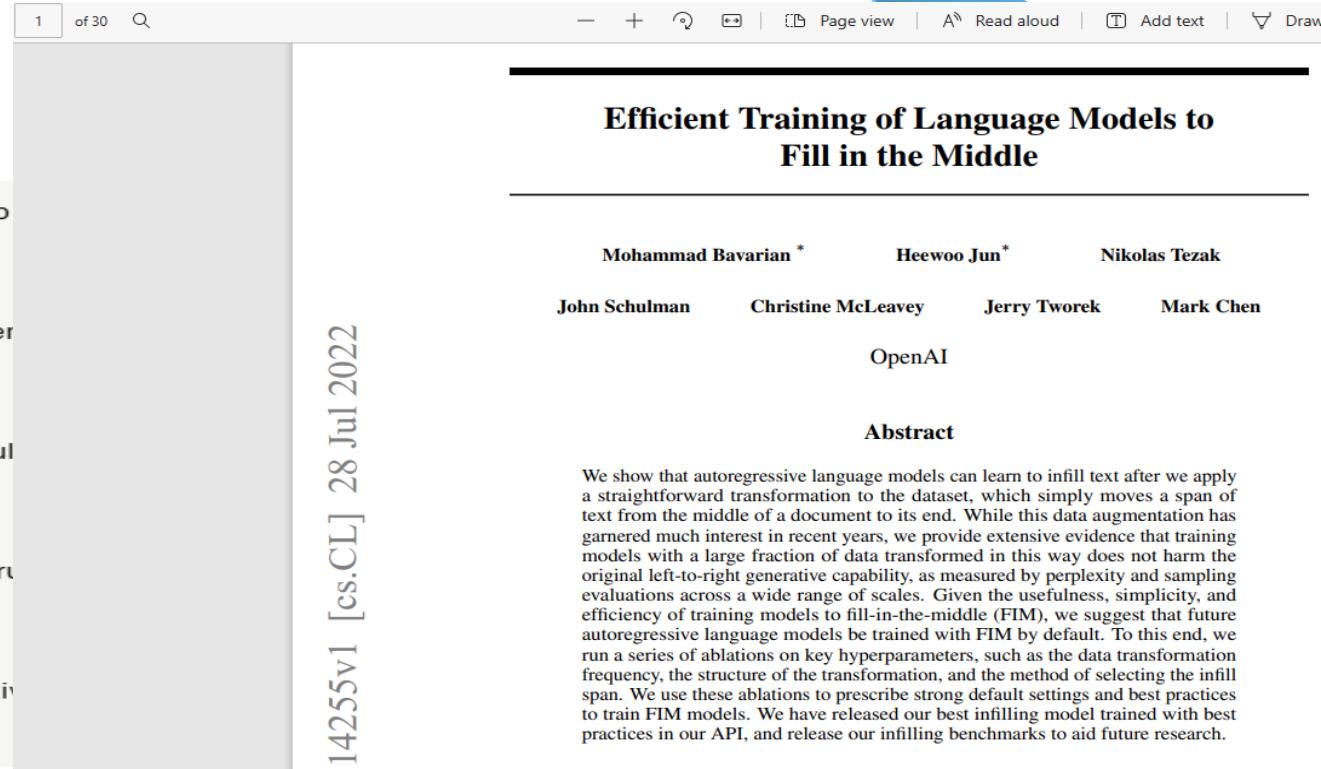
Hierarchical Text-Conditional Image Gen
April 13, 2022 — Read blog post

Formal Mathematics Statement Curricul
February 3, 2022 — Read blog post

Training language models to follow instru
January 27, 2022 — Read blog post

Text and Code Embeddings by Contrastiv
January 24, 2022 — Read blog post

14255v1 [cs.CL] 28 Jul 2022



The screenshot shows a list of publications on the OpenAI website. The first item is the paper 'Efficient Training of Language Models to Fill in the Middle'. The page includes a navigation bar with 'Page view', 'Read aloud', and 'Add text' options.

Efficient Training of Language Models to Fill in the Middle

Mohammad Bavarian * Heewoo Jun * Nikolas Tezak

John Schulman Christine McLeavey Jerry Tworek Mark Chen

OpenAI

Abstract

We show that autoregressive language models can learn to infill text after we apply a straightforward transformation to the dataset, which simply moves a span of text from the middle of a document to its end. While this data augmentation has garnered much interest in recent years, we provide extensive evidence that training models with a large fraction of data transformed in this way does not harm the original left-to-right generative capability, as measured by perplexity and sampling evaluations across a wide range of scales. Given the usefulness, simplicity, and efficiency of training models to fill-in-the-middle (FIM), we suggest that future autoregressive language models be trained with FIM by default. To this end, we run a series of ablations on key hyperparameters, such as the data transformation frequency, the structure of the transformation, and the method of selecting the infill span. We use these ablations to prescribe strong default settings and best practices to train FIM models. We have released our best infilling model trained with best practices in our API, and release our infilling benchmarks to aid future research.

Dữ liệu: là quan trọng nhất

≡ VNEXPRESS | Góc nhìn >

Góc nhìn



Đặng Hùng Võ

Nguyên Thứ trưởng Bộ
Tài Nguyên Môi trường

AI cũng khóc

Thứ năm, 30/5/2019, 00:56 (GMT+7) | 86 |

Trên thế giới mươi năm trước, người ta đã hoàn thành cuộc cách mạng công nghiệp thứ ba với việc gắn mọi hàng hóa, khái niệm, quá trình với tính từ "điện tử", viết tắt bằng tiền tố "E" ở đầu từ như: thương mại điện tử (E-commerce), chính quyền điện tử (E-government), xã hội điện tử (E-society) hay công dân điện tử (E-citizen).

Ngày nay, con người bàn về cuộc cách mạng công nghiệp thứ tư với việc gắn tính từ "thông minh" vào trước nhiều thứ như: cuộc sống thông minh (smart life), thành phố thông minh (smart city), quản lý thông minh (smart administration) hay xã hội thông minh (smart society). "Thông minh" được hiểu là trí tuệ nhân tạo, giúp con người đưa ra lựa chọn tối ưu khi mọi thông tin chính xác được kết nối đầy đủ qua mạng Internet.

Dữ liệu: Dầu mỏ mới

THE NEW NEW WORLD

How Cheap Labor Drives China's A.I. Ambitions



Workers at the headquarters of Ruijin Technology Company in Jiaxian, in central China's Henan Province. They identify objects in images to help artificial intelligence make sense of the world. Yan Cong for The New York Times

Data is the new oil, it has been said for years now. If data is the new oil, then China is already the largest producer with its factories packed with laborers working hard to annotate images and data for machine learning (*Analytics India Magazine*).

Kết luận:

Số hóa > Công nghệ > AI

Thứ năm, 23/3/2023, 07:00 (GMT+7)

Bill Gates: 'AI là đột phá mang tính cách mạng'

Tỷ phú Bill Gates đánh giá những bước tiến lớn của AI hiện nay tương đương việc tạo bộ xử lý, PC, Internet và điện thoại di động.



Tầm nhìn nghề nghiệp

Theo tỷ phú Mỹ, AI có thể được sử dụng như một công cụ để cải thiện năng suất công việc cho con người, giúp giảm sự bất bình đẳng toàn cầu tại nơi làm việc, cũng như các vấn đề về chăm sóc sức khỏe và giáo dục.

Những lo ngại

Bên cạnh lợi ích, Gates cho rằng sự phát triển của AI đặt ra những câu hỏi khó về lực lượng lao động, hệ thống pháp luật, quyền riêng tư, sự thiên vị. "Thế giới cần đảm bảo tất cả đều được hưởng lợi từ trí tuệ nhân tạo thay vì chỉ một nhóm nhất định. Các chính phủ và tổ chức từ thiện cần đóng vai trò quan trọng trong việc đảm bảo nó làm giảm sự bất bình đẳng. Đó cũng là ưu tiên của tôi liên quan đến AI", tỷ phú cho biết.

Nên dùng/cấm?

Page 10A The Daily Item — Sumter, S.C. Saturday, April 5, 1986



AP photo

Elementary school teachers picket against use of calculators in grade school
The teachers feel if students use calculators too early, they won't learn math concepts

Math teachers protest against calculator use

By JILL LAWRENCE

"My older kids don't pay any attention to an answer being absurd," he said. "Teachers are shy."

5/4/1986:

Các giáo viên Toán tuần hành đòi cấm học sinh tiểu học sử dụng máy tính tay (calculator)

Vì lo ngại:

Học sinh sẽ không chịu học cách tính các phép tính

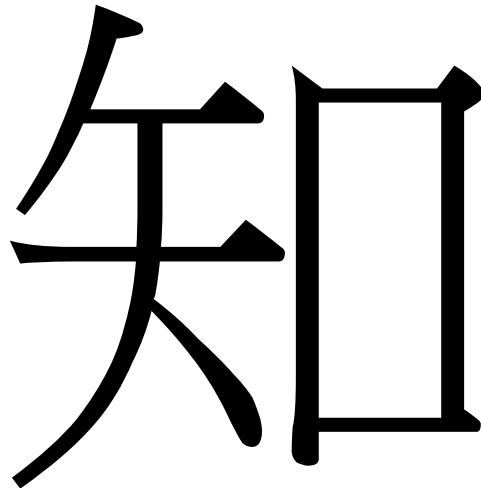
Tâm thế chúng ta?

■ ChatGPT:

- ✓ Công cụ trợ giúp rất đắc lực cho việc tìm kiếm và xử lý (phân loại, tóm tắt, tổng hợp,...) thông tin khoa học.
- ✓ Công cụ trợ giúp viết công trình khoa học
- Tìm hiểu **CƠ HỘI** ứng dụng vào **nghiên cứu khoa học**
- ❑ **Đối phó THÁCH THỨC** mới
- ✓ Khó truy nguồn thông tin và đánh giá độ tin cậy.
- ✓ **Khả năng rò rỉ thông tin**
- **THAY ĐỔI** cách thức QUẢN LÝ/GIÁM SÁT thông tin
- **TÂM THẾ** chúng ta?
- Có **LO** không?

TRI THỨC là
SỨC MẠNH

THỈ
(mũi tên
→
sâu sắc)



KHẨU
(mouth: lời nói)

Người có TRI thức thì lời nói
sâu sắc như mũi tên

TRI
(knowledge)

Sinh viên Việt Nam => TRÍ tuệ:
dùng TRI thức của mình: NCKH
=> để VIẾT lên được điều tốt đẹp cho đời.

TRI
(*knowledge*)

VIẾT
(*make*)



TRÍ
(*wisdom*)

Nhất nam *viết* hữu, thập nữ *viết* vô
一男曰有, 十女曰无

Lời kết:

Nhân (Human) 忧 (worry) 优 (good)

人 + 尤 = 优

Nhân hữu ưu sâu ưu điểm đại

人 有 尤 愁 优 点 大

LO

HỌC HỎI

CƠ HỘI

Go (đi) + Slow (chậm) + After (sau) => Cách mạng TTNT?

"tụt HẬU" (lag)

HÀNH(go)
後
MỊCH(slow)
TÙY(after)

UNS: companion (bạn đồng hành)

pan: panis (Latin) = bread (bánh mì)

le pain (Fr), パン (Jp), 빵 (Ko), ...

com: to share (chia sẻ)

companion: chia sẻ mẫu bánh mì cuối cùng

XIN CẢM ƠN