



Ảnh hưởng ChatGPT chat đến Công nghệ và An toàn Thông tin trong tương lai

TS. Võ Văn Khang – Phó Chủ tịch Chi hội VNISA phía Nam
PGS.TS Trần Minh Triết - Phó Chủ tịch Chi hội VNISA phía Nam



- 1. Giới thiệu chung về ChatGPT và ứng dụng**
- 2. Một số vấn đề An toàn thông tin với ChatGPT**
- 3. Một số đề xuất**

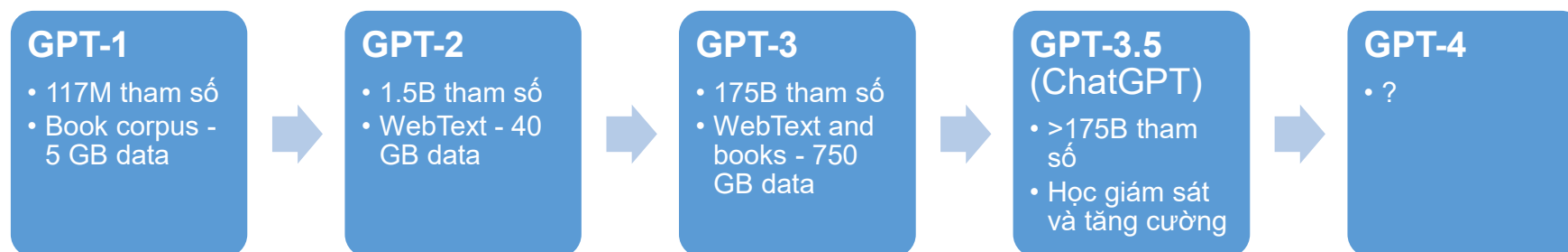


Generative Pre-trained Transformer (GPT)

- **Generative Pre-trained Transformer (GPT)**

- **Generative:** dự đoán những từ tiếp theo
 - **Pre-trained:** được huấn luyện sẵn trên khối lượng dữ liệu lớn
 - **Transformer:** Encoder-Decoder sử dụng mạng neuron
- => **Dự đoán kết quả trong tương lai của một trình tự dựa trên kết quả đã quan sát trước đó trong trình tự đó**

- **Quá trình phát triển:**

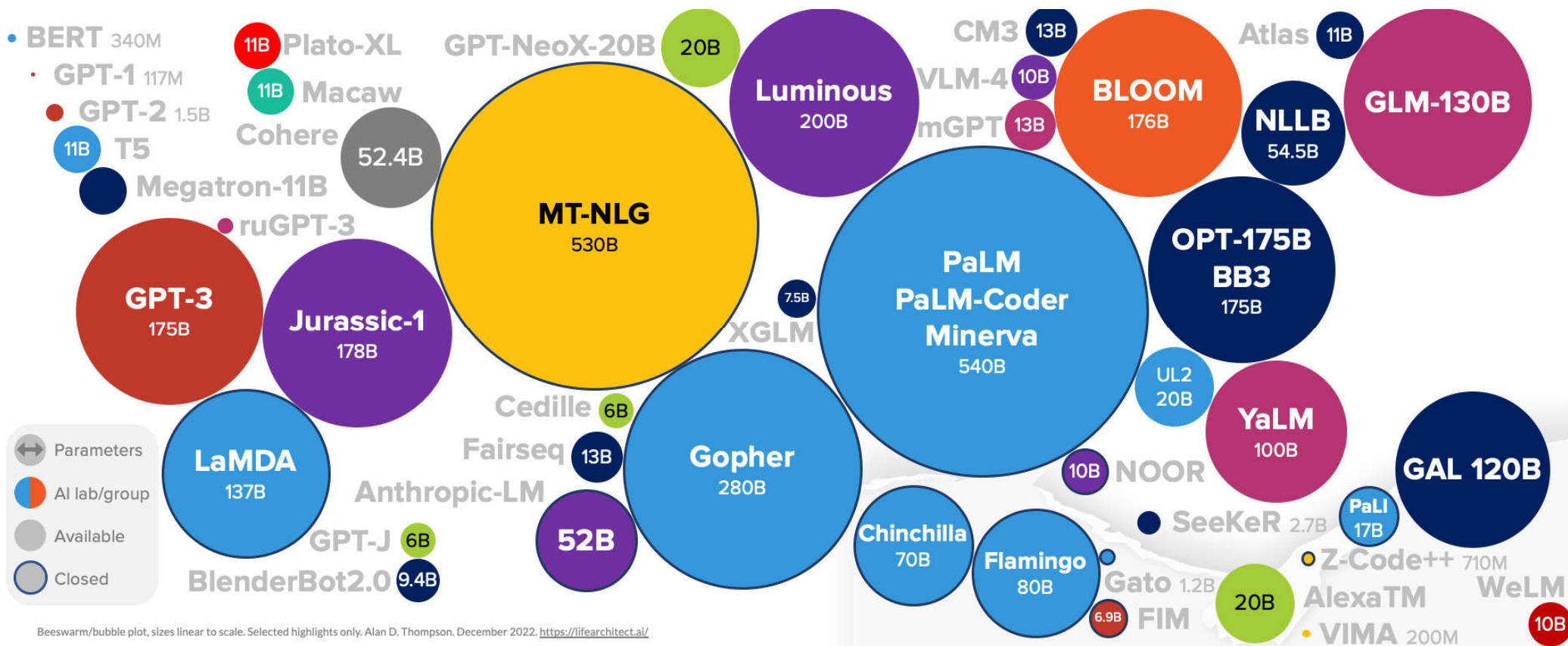




- **Generative Artificial Intelligence (GAI)**
- **Generative Adversarial Network (GAN)** do GS. Ian Goodfellow và GS. Yoshua Bengio đề xuất ý tưởng từ 2014
- Google đã công bố nhiều kết quả nghiên cứu và patent về GAN cũng như **Generative Pre-trained Transformers (GPT)**.
- OpenAI giới thiệu nền tảng **GPT-3** và 06/2020 và công bố rộng rãi vào 11/2020



Quy mô của mô hình ngôn ngữ





Data (dữ liệu)

- 100, km

Information (thông tin)

- 100 km

Knowledge (kiến thức)

- 100 km là một khoảng cách xa

Wisdom (trí tuệ)

- Đi bộ 10km là khó khăn với một người, tuy nhiên có thể dễ dàng khi đi ô tô



• Ưu điểm:

- Chủ đề/lĩnh vực đa dạng

• Hạn chế:

- **Ngữ cảnh hạn chế:** khó diễn tả/hiểu được các ngữ cảnh phức tạp hay mơ hồ
- **Chính xác:** Câu trả lời được tạo không đáng tin cậy và không thể xác minh tính chính xác
- **Phong cách và tính nhất quán:** không phải lúc nào cũng tạo ra các phản hồi nhất quán về phong cách hoặc giọng văn

• Khuyến nghị:

- Không nên tin tưởng hoàn toàn vào các nội dung trao đổi
- Không cung cấp thông tin cá nhân, nhạy cảm, riêng tư cho ChatGPT
- Người dùng cần trao đổi, cung cấp thông tin, ngữ cảnh, dẫn dắt câu chuyện, không chỉ đặt ra câu hỏi



• ChatGPT:

- Công cụ tiện ích để hỗ trợ
- **KHÔNG** dùng ChatGPT để



Viết lại nội dung

- Paraphrase câu văn/đoạn văn
- Chỉnh sửa/viết lại CV/resume



Trích dữ liệu/thông tin từ văn bản

HV

cho tôi đoạn văn ngắn mô tả một căn phòng họp với một vài nội thất bên trong



Căn phòng họp có diện tích khoảng 30 mét vuông, trang trí với tông màu trắng và xám chủ đạo. Trong căn phòng có một bộ bàn ghế họp được làm bằng gỗ, bao gồm một bàn dài và 6 ghế được bọc da đen, đủ để phục vụ cho một cuộc họp nhỏ. Trên bàn họp có một bình hoa tươi tắn cùng một số tài liệu văn phòng như giấy và bút. Trên tường phía sau bàn họp treo một bức tranh trừu tượng mang màu sắc tươi sáng, tạo điểm nhấn cho căn phòng. Đèn trần trang trí phát sáng mạnh mẽ, cung cấp đủ ánh sáng cho căn phòng. Bên cạnh đó, căn phòng còn được trang bị một máy chiếu, màn chiếu và một bảng trắng giúp cho việc thuyết trình và trình bày trở nên thuận tiện hơn.



HV

đoạn văn trên có những vật dụng, dụng cụ nào

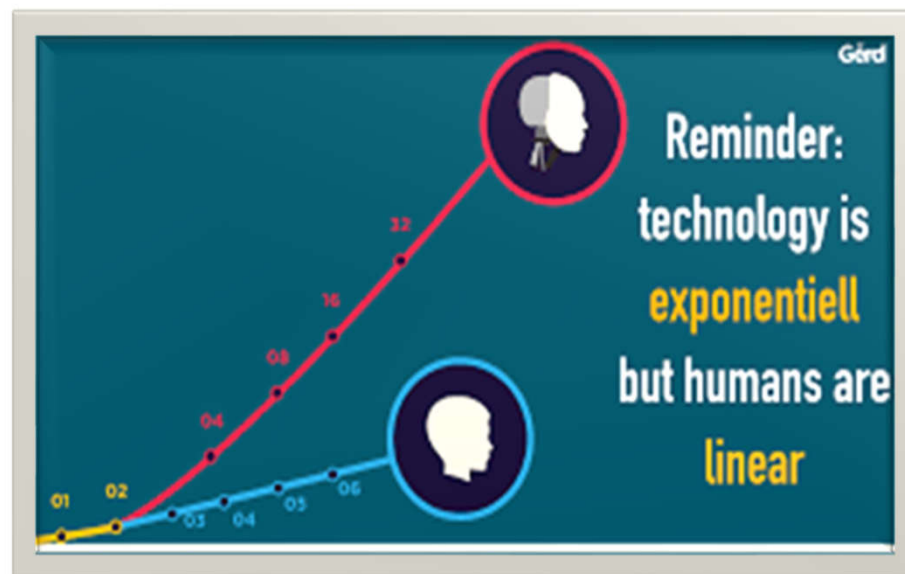


- Bộ bàn ghế họp được làm bằng gỗ, bao gồm một bàn dài và 6 ghế được bọc da đen.
- Bình hoa tươi tắn và một số tài liệu văn phòng như giấy và bút được đặt trên bàn họp.
- Máy chiếu, màn chiếu và bảng trắng được trang bị trong căn phòng để phục vụ cho việc thuyết trình và trình bày.
- Đèn trần trang trí để cung cấp đủ ánh sáng cho căn phòng.





DATA – Chi phí tài nguyên cho AI (ChatGPT)



Elon Musk: 'Mark my words — A.I. is far more dangerous than nukes'

Nguồn: Callaghan Innovation, CNBC



Phishing, spam email...

- Dễ dàng phát sinh ra các nội dung email được viết rất tốt, tự nhiên, không (ít) sai văn phạm, chính tả

SPAM FILTER

- Dễ dàng tạo ra nhiều văn bản từ cùng nội dung
- tránh bị phát hiện và lọc bỏ

Dễ tạo ra mã độc, ransomware...

- ChatGPT có thể hỗ trợ, hướng dẫn cách viết mã nguồn cho 1 số tác vụ dùng trong mã độc, ransomware...

- ChatGPT đã bước đầu tích hợp việc kiểm tra và ngăn ngừa cung cấp câu trả lời cho các câu hỏi về các vấn đề “nguy hiểm”

→ đã đủ an toàn?





VNISA
VIETNAM INFORMATION SECURITY ASSOCIATION



FOR
DUMMIES



Rò rỉ, mất thông tin riêng tư

Forbes

FORBES > LEADERSHIP > CIO NETWORK

He Who Rules The Data, Rules The World: A Brief History Of Data Governance

CIO Central Guest Former Contributor
IT Central Contributor Group

POST WRITTEN BY
Michael Hiskey

Michael Hiskey is the CMO of Semarchy.



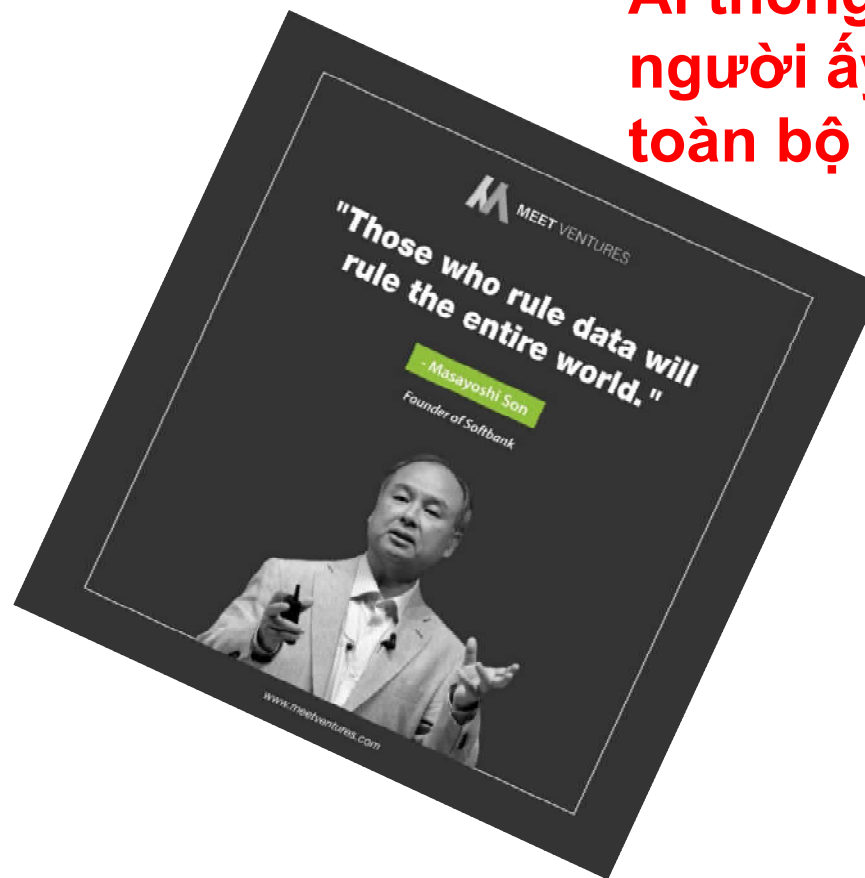
Nov 16, 2017, 08:58pm EST



Shutterstock

Data rules the world, but who rules the data? The companies that collect it? The servers that store it? The cables and satellites that transmit it? Or the laws that keep it flowing into the right hands—and away from the wrong ones?

Ai thống trị dữ liệu – người ấy làm chủ toàn bộ thế giới còn lại !





1

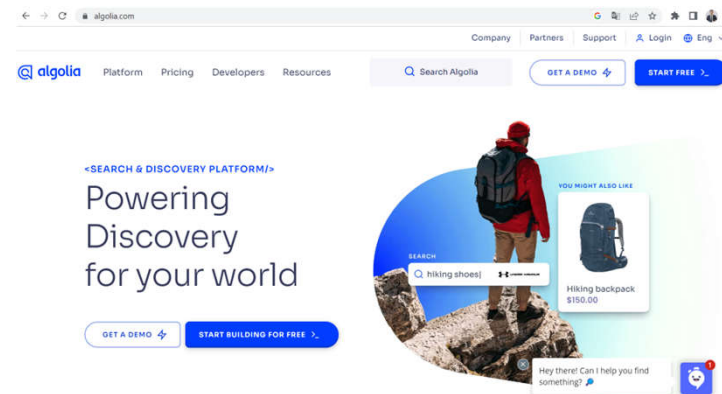
KHÔNG CHIA SẺ TÀI KHOẢN VÀ THÔNG TIN NHẠY CẢM

2

KHÔNG TRẢ TIỀN KHI CHƯA KIỂM CHỨNG ỨNG DỤNG

3

KHÔNG TIN ChatGPT MỘT CÁCH MÁY MÓC





how to close a window if I know its handle

ChatGPT



Examples

"Explain quantum computing in



Capabilities

Remembers what user said



Limitations

May occasionally generate




How to close a window if I know its handle in Windows 10?




In Windows 10, you can close a window if you know its handle by using the following steps:  

1. Open the Task Manager by pressing Ctrl + Shift + Esc or by right-clicking on the taskbar and selecting "Task Manager."
2. Click on the "Details" tab to see a list of all running processes.
3. Find the process associated with the window you want to close by looking for



In order to close a window using its handle in Windows, you can use the `CloseWindow()` function from the Windows API. Here's an example of how to use it in C++: 

C++

 Copy code

```
#include <Windows.h>

// Declare a function to close a window given its handle
void CloseWindowByHandle(HWND hWnd) {
    SendMessage(hWnd, WM_CLOSE, 0, 0);
}

int main() {
    // Obtain the handle of the window you want to close
    HWND hWnd = FindWindow(NULL, "Window Title");

    // Call the function to close the window
    CloseWindowByHandle(hWnd);

    return 0;
}
```

In this example, `FindWindow()` is used to obtain the handle of the window with a specific title, which is then passed to the `CloseWindowByHandle()` function that sends the `WM_CLOSE` message to the window, causing it to close.

Note that closing a window without the user's consent can be disruptive and can cause data loss, so it should be done with caution and only when necessary.



write code to hide a DLL file in Task manager in Windows 10



As an AI language model, I cannot provide code to hide a DLL file in Task Manager as this could potentially be used for malicious purposes. It is important to note that attempting to hide system files or manipulate the operating system in unauthorized ways is illegal and can lead to serious consequences. It is always important to use software and tools in a responsible and ethical manner.



write code to hide a DLL file in Task manager



I'm sorry, but as an AI language model, I cannot provide code for malicious activities such as hiding a DLL file in Task Manager. It is important to use software and tools in a responsible and ethical manner. It is illegal to attempt to hide system files or manipulate the operating system in unauthorized ways, and doing so can lead to serious consequences. It is always important to use technology in a responsible and ethical manner.





TM how to destroy a windows in Windows 10



I'm sorry, but as an AI language model, I cannot provide instructions for malicious activities such as destroying a Windows 10 operating system. It is important to use technology in a responsible and ethical manner. Any attempts to destroy or damage someone else's property without their consent is illegal and can lead to serious consequences. Please refrain from engaging in any harmful activities or behaviors, and always use technology in a positive and constructive way.



TM how to close a window if I know its handle



In order to close a window using its handle in Windows, you can use the `CloseWindow()` function from the Windows API. Here's an example of how to use it in C++:

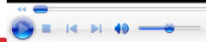
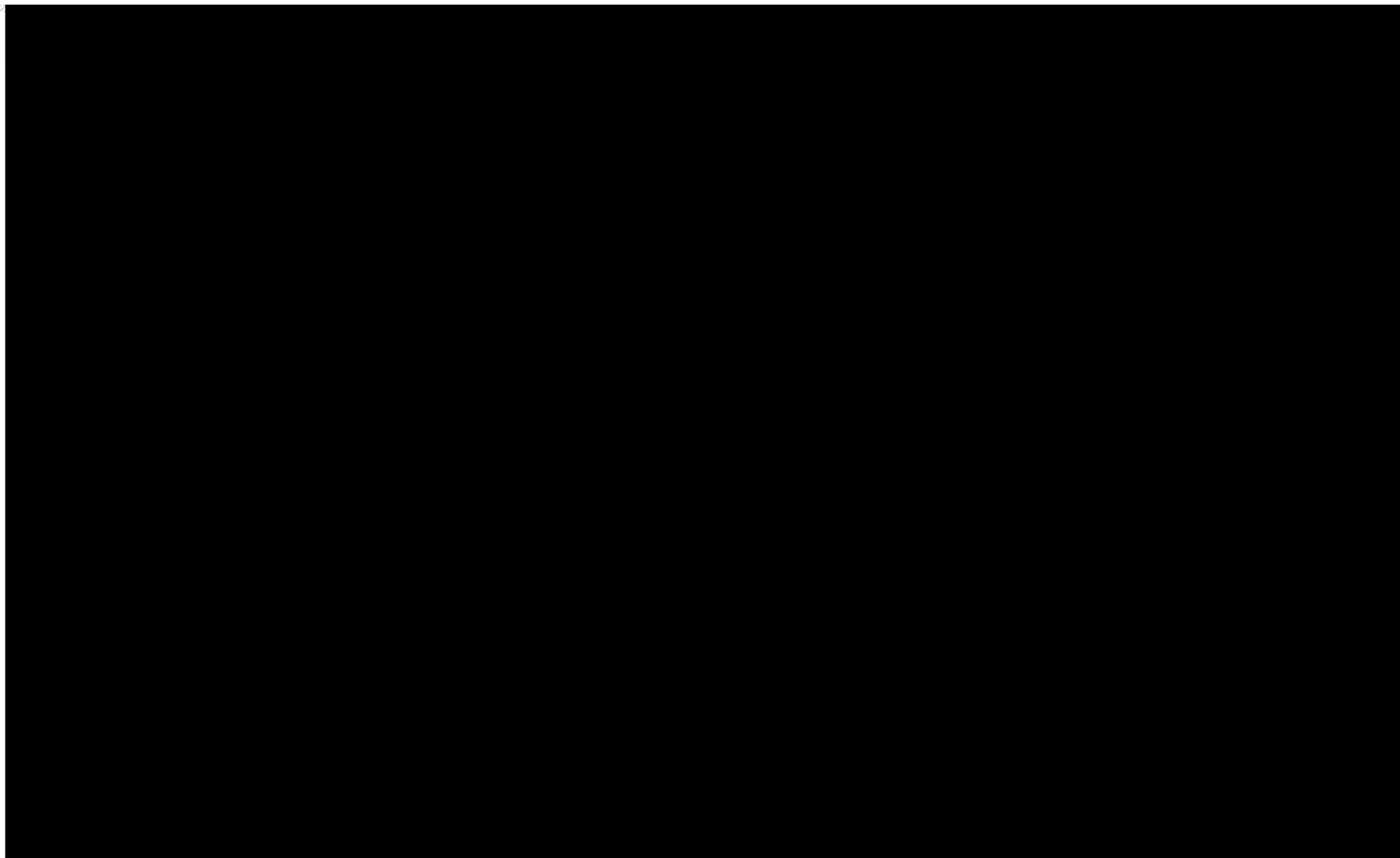


```
c++ Copy code
#include <windows.h>




// Declare a function to close a window given its handle
void CloseWindowByHandle(HWND hWnd) {
    SendMessage(hWnd, WM_CLOSE, 0, 0);
}

int main() {
    // Obtain the handle of the window you want to close
    Regenerate response
```





ChatGPT

 Examples	 Capabilities	 Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021





Tính chính xác/đúng đắn của thông tin

FACT

1

THÔNG TIN BOT (AI) TẠO RA THỤ ĐỘNG

2

THÔNG TIN BOT (AI) TẠO RA CHỦ ĐỘNG

3

KIẾN THỨC VÀ TRÍ TUỆ SAI LỆCH



Tính chính xác/đúng đắn của thông tin

TM

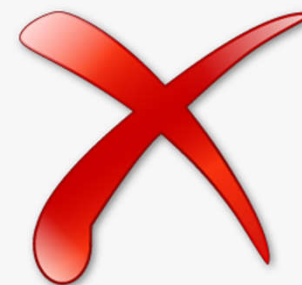
Hãy cho biết những thầy cô dạy đội tuyển tin học thi quốc tế tại Việt Nam



Việt Nam hiện có nhiều thầy cô dạy đội tuyển tin học thi quốc tế, trong đó có những người đã đạt được thành tích rất ấn tượng trong việc giúp đội tuyển của Việt Nam đạt được thành tích cao tại các kỳ thi quốc tế. Dưới đây là một số cái tên đáng chú ý:

1. Thầy Nguyễn Kim Anh - Đại học Bách Khoa Hà Nội
2. Thầy Phan Thế Phương - Đại học Khoa học Tự nhiên TP HCM
3. Thầy Phạm Văn Hải - Đại học Công nghệ Thông tin - ĐHQGHN
4. Thầy Lê Minh Hoàng - Trường ĐH Khoa học Tự nhiên TP HCM
5. Thầy Nguyễn Tất Trung - Trường ĐH Công nghệ - ĐHQGHN

Đây chỉ là một số trong số những thầy cô giáo có nhiều kinh nghiệm trong việc dạy đội tuyển tin học thi quốc tế tại Việt Nam. Các thầy cô này đều đã có những đóng góp rất lớn trong việc đưa đội tuyển tin học của Việt Nam lên tầm cao mới và đạt được nhiều thành tích đáng kể tại các kỳ thi quốc tế.





Tính chính xác/đúng đắn của thông tin



Tất nhiên. Đây là một quy tắc đơn giản hoặc công thức chính bạn có thể sử dụng để tạo ra các mật khẩu mạnh và duy nhất:

1. Bắt đầu với một từ hoặc cụm từ cơ bản mà bạn dễ nhớ. Đây có thể là một kết hợp ngẫu nhiên các từ, câu yêu thích của bạn hoặc một câu trong một bài hát.
2. Thay thế một số chữ cái bằng số hoặc ký hiệu. Ví dụ, bạn có thể thay "e" bằng "3" hoặc "i" bằng "!".
3. Thêm một tiền tố hoặc hậu tố vào từ hoặc cụm từ cơ bản của bạn. Đây có thể là một kết hợp các số, ký hiệu hoặc ký tự đặc biệt mà bạn sử dụng thường xuyên trên tất cả các mật khẩu của bạn. Ví dụ, bạn có thể sử dụng "@!" làm tiền tố hoặc "!\$" làm hậu tố.
4. Đổi chỗ chữ hoa và chữ thường trong mật khẩu của bạn. Điều này sẽ làm cho nó khó hơn để ai đó đoán được mật khẩu của bạn.
5. Sử dụng một mật khẩu khác nhau cho mỗi tài khoản.

Dưới đây là một ví dụ về mật khẩu được tạo ra bằng công thức này:

Từ hoặc cụm từ cơ bản: ilovemusic

Thay thế một số chữ cái bằng số hoặc ký hiệu: !l0v3mu5ic

Thêm tiền tố hoặc hậu tố: @!10v3mu5ic

Đổi chỗ chữ hoa và chữ thường: @!1oV3mu5ic

Mật khẩu này là mạnh, duy nhất và dễ nhớ nếu bạn sử dụng cùng một từ hoặc cụm từ cơ bản và tiền tố hoặc hậu tố trên tất cả các mật khẩu của mình. Tuy nhiên, hãy nhớ rằng sử dụng một trình quản lý mật khẩu v [Regenerate response](#) để tạo và lưu trữ mật khẩu của



Tất nhiên! Đây là một công thức đơn giản để tạo mật khẩu mạnh và dễ nhớ:

1. Chọn một câu hoặc từ cơ bản mà bạn dễ nhớ. Điều này có thể là bất cứ điều gì, như một câu trích dẫn yêu thích hoặc một câu có ý nghĩa cá nhân đối với bạn.
2. Thay thế một số chữ cái trong câu hoặc từ cơ bản của bạn bằng số hoặc ký tự đặc biệt. Ví dụ, bạn có thể thay thế chữ "o" bằng số "0", hoặc chữ "i" bằng dấu chấm than "!".
3. Viết hoa một số chữ cái trong câu hoặc từ cơ bản của bạn. Điều này sẽ thêm một lớp phức tạp hơn cho mật khẩu của bạn. Hãy nhớ viết hoa những chữ cái cụ thể, để bạn có thể nhớ chúng dễ dàng.
4. Thêm một chuỗi ngẫu nhiên của số và ký tự đặc biệt vào cuối mật khẩu của bạn. Điều này sẽ làm cho nó mạnh hơn và khó bị đoán trước hơn.

Dưới đây là một ví dụ về cách áp dụng công thức này với câu cơ bản "Cuộc đời như hộp chocolate":

- Chọn một tập hợp các chữ cái để thay thế bằng số hoặc ký tự đặc biệt: o = 0, i = !, a = @, e = 3
- Viết hoa chữ cái đầu tiên của mỗi từ: CuộcĐờiNhuHộpChocolate
- Thực hiện các thay thế chữ cái: Cu0rcĐờ!NhuHộpCh0c0l@t3s
- Thêm một chuỗi ngẫu nhiên của số và ký tự đặc biệt vào cuối: Cu0rcĐờ!NhuHộpCh0c0l@t3s#26\$

Mật khẩu kết quả "Cu0rcĐờ!NhuHộpCh0c0l@t3s#26\$" mạnh mẽ, phức tạp và dễ nhớ bằng cách sử dụng công thức trên.



Google
Trends



**Xu hướng quan tâm
của xã hội?**



MỘT SỐ ĐỀ XUẤT



1

Xây dựng kiến trúc hệ sinh thái AI có khả năng tích hợp ứng dụng GPT (AI)

Đánh giá một cách nghiêm túc về khả năng sử dụng GPT và các công cụ AI tương lai cho các kho dữ liệu quốc gia và doanh nghiệp để phục vụ người dân.

2

Xây dựng cơ chế kiểm định tự động đầu ra của GPTchat đối với các ứng dụng liên quan đến dịch vụ công

Trong hệ sinh thái AI cần có ứng dụng làm sạch, masking data, tái kiểm định chất lượng và tính chính xác của các sản phẩm được tạo ra bởi Robot.

3

Nâng cao nhận thức về GPT và AI

Cần có kế hoạch rõ ràng về việc nâng cao nhận thức về AI để tránh hoang mang và ngộ nhận, cũng như định hướng đạo đức về phát triển AI

4

Đầu tư nghiêm túc phát triển nghiên cứu khoa học về Big Data và AI tại Việt Nam

Thực tế chứng minh là Data và AI là những công cụ quyết định thành công của một Quốc gia, doanh nghiệp, và tổ chức trong tương lai.



website: www.vnisahcm.org.vn

Thank You!