Generative Al growing pains, security value and implications

Michael Melore, CISSP

IBM Senior Cybersecurity Advisor Chair SecRT (Security <Leaders> Round Table) FBI InfraGard Committeemen NASCIO Privacy and Cyber Security Committees CISA / ATARC Cyber Defense Task Force

@MichaelMelore mmelore@us.ibm.com

Champion of No Cost Trusted Advisory Programs

- Architect led collaborative workshops
 Interactive no products pitched education
- XForce Cyber Range Exercises Exclusive Sessions available at cost
- Solution and Threat Briefings
- IBM Center for Government Cybersecurity Understand cyber threats with the IBM Center for Government Cybersecurity https://www.ibm.com/security/services/us-federal-cybersecurity-center
- Onsite collaborative exercises/knowledge sharing/tech days Can be jointly delivered with State Officials

IBM Sponsored External Committees

www.IJIS.org IJIS Institute Cybersecurity Task Force

www.NASCIO.org National Assoc State CIOs

Privacy and Cybersecurity Working Groups

www.ATARC.org Advanced Technology Academic Research Center

Government Grants, Higher Ed, and Cyber Defense Working Groups California California Cybersecurity Task Force- Workforce Development-Education Subcommittee

- XForce Exchange Threat Intelligence Forum https://exchange.xforce.ibmcloud.com
- IBM Security Learning Academy
 Hundreds of no cost Tutorials/Hand on Labs
 https://www.securitylearningacademy.com
- IBM's New Collar / Apprentice programs
 https://www.ibm.com/impact/feature/apprenticeship
- IBM's Skills Build (Resources for Learners, Educators and Organizations)
 https://skillsbuild.org
- XForce Webinars on Threat Landscapes
- IBM Security Solution User Groups
- Regional/Nationwide Security Leaders Round Tables
 www.SECRT.us

Grass roots, non sponsored, no cost Security Executive Round Tables Nationwide, initially founded by Mike Melore and a dozen CISOs, now at 21 active Nationwide chapters, 1500+ opt in security leaders as members.



The marketing moment offered by ChatGPT is incredible. We've seen these moments before a company called Netscape brought the web browser to everybody's attention."

What ChatGPT has done is help make AI real to many people who kind of were aware of it but didn't maybe quite see what the power of AI would be."

> IBM CEO, Arvind Krishna Excerpts from a Financial Times interview February 16, 2023





Nutrition Fa

Serving Size Servings Per Container

Amount Per Serving	9
Calories	Calories from
	9
Total Fat	g
Saturated Fat	g
Cholesterol	g
Sodium	g
Total Carbohydra	ate g
Dietary Fiber	g
Sugar	g
Protein	g
Vitamin A %	Vitamir
Calcium %	• Iron

*Percent Daily Values are based on a 2,00 Your daily values may be higher or lower or your calorie needs.

ВВС

News

Home

Reel

Sport

ChatGPT: New AI chatbot has everyone talking to it

() 7 December 2022

By Chris Vallance

Technology reporter

A new chatbot has passed <mark>one million users in less than a week</mark>, the project behind it says.

ChatGPT was publicly released on Wednesday by OpenAI, an artificial

list 15 major events in the history of AI in short sentences

Sure, here are 15 major events in the history of AI in short sentences:

- 1. 1956: The term "Al" is coined.
- 2. 1959: First Al program plays checkers.
- 3. 1966: ELIZA chatbot simulates human conversation.
- 4. 1974: MYCIN expert system for medical diagnosis developed.
- 5. 1981: The first Al winter begins.
- 6. 1997: IBM's Deep Blue defeats Garry Kasparov in chess.
- 7. 2005: Stanford's Stanley wins the DARPA Grand Challenge.
- 8. 2011: IBM's Watson wins Jeopardy against human champions.
- 9. 2012: Google's neural network learns to recognize cats.
- 10. 2015: AlphaGo defeats Lee Sedol in the game of Go.
- 11. 2016: Google's AlphaGo Zero beats AlphaGo with no human data.
- 12. 2017: OpenAI's DOTA 2 AI defeats top professional players.
- 13. 2018: AI-powered voice assistants become widespread.
- 14. 2019: GPT-2 language model produces highly realistic text.
- 15. 2020: Al aids in COVID-19 vaccine research and development.

ChatGPT has initiated a massive surge of interest in AI

A Google Trends comparison of the search terms "ai", "chatgpt", and "covid-19" from April, 2019 to April, 2023

"Interesting analysis evocative of two truths: 1. The first report is usually wrong 2. It is exceedingly difficult to unseat a first impression

Heather McMahon Artemist Advisory Group, LLC

Clearing up the Unclear

•**Misinformation** is false, but not created or shared with the intention of causing harm.

•**Disinformation** is deliberately created to mislead, harm, or manipulate a person, social group, organization, or country.

•Malinformation is based on fact, but used out of context to mislead, harm, or manipulate.

Internet search and online advertising is at the forefront of AI investment

Bloomberg

Microsoft Invests \$10 Billion in ChatGPT Maker OpenAl

By Dina Bass

January 23, 2023 at 9:06 a.m. EST Updated on January 23, 2023 at 5:03 p.m. EST

Microsoft Corp. is investing \$10 billion in OpenAI, whose artificial intelligence tool ChatGPT has lit up the internet since its

The **A** Register[®]

Of course Facebook will monetize an ad-generating AI

╃ <u>Katyanna Quach</u>

8

Wed 5 Apr 2023 // 22:45 UTC

Meta is building a generative AI system capable of churning out online ads to order and will charge businesses to use it, according to chief technology officer Andrew Bosworth.

Kari Paul and agencies Tue 7 Feb 2023 23.41 GMT

Microsoft to power Bing with AI as race with Google heats up

Company to work with OpenAI to improve search and Edge web browser as rival unveils ChatGPT competitor

Microsoft is revamping its search products with more artificial intelligence, using technology behind the wildly popular ChatGPT, as tech companies race to take advantage of increasingly powerful AI tools.

The company detailed its plans at a special event on Tuesday, saying it would work with OpenAI, the startup behind the <u>ChatGPT</u> tool, to upgrade its Bing search engine and Edge web browser and enhance the information available.

$\mathsf{E} \mathsf{A} \mathsf{R} \mathsf{T} \mathsf{H} \cdot \mathsf{O} \mathsf{R} \mathsf{G}$

ENERGY

The Environmental Impact of ChatGPT: A Call for Sustainable Practices In AI Development

Robert Pearl, M.D. Contributor 0 GLOBAL COMMONS APR 28TH 2023 4 MINS Markets Tech Media Success Perspectives Apr 24, 2023, 04:15am EDT EARTH.ORG IS POWERED BY OVER 150 CONTRIBUTING WRITERS The New Hork Times ChatGPT, a large language model developed by OpenAI, has garnered widespread atter Google shares lose \$100 billion after remarkable natural language processing capabilities. However, as with any large languag company's AI chatbot makes an error **Disinformation Researchers Raise** training and developing the AI system requires a tremendous amount of energy, resultin during demo significant environmental costs that are often overlooked. In this article, we take a look a Alarms About A.I. Chatbots already now regarding the environmental impact of ChatGPT. By Catherine Thorbecke, CNN Updated 9:41 AM EST, Thu February 9, 202 MARKETS BUSINESS INVESTING TECH POLITI Researchers used ChatGPT to produce clean, convincing text that (CNN) — Google's much-hyped new AI chatbot tool Bard, which has yet to be released to the public, In generative AI legal Wild West, the repeated conspiracy theories and misleading narratives. being called out for an inaccurate response it produced in a demo this week courtroom battles are just getting started By Tiffany Hsu and Stuart A. Thompson Ellen Sheng BELLENSHENG PUBLISHED MON, APR 3 2023-10-56 AM EDT | UPDATED MON, APR 3 2023-11-29 AM EDT INDEPENDENT Published Feb. 8, 2023 Updated Feb. 13, 2023 KEY As companies including Microsoft, Google and OpenAI launch generative POINTS Al to the general public, lawsuits are piling up from creative industries Soon after ChatGPT debuted last year, researchers tested what the LIFESTYLE TRAVE about copyrighted work co-opted or used by AI. CULTURE artificial intelligence chatbot would write after it was asked Getty Images, the photo licensing company, filed a lawsuit against Stable Tech questions peppered with conspiracy theories and false narratives. TIME \equiv ChatGPT built with help of underpaid, The New AI-Powered Bing Is Threatening Users. exploited Kenyan employees, report alleges That's No Laughing Matter Kenyan workers were tasked with labelling content from 'darkest recesses of the internet', TIME reports BY BILLY PERRIGO 🔰 Vishwam Sankaran • Monday 23 January 2023 10:45 FEBRUARY 17, 2023 10:58 AM EST OpenAI's chatbot ChatGPT was reportedly built using vital contributions hortly after Microsoft released its new AI-powered search tool, Bing, to a C from outsourced, underpaid Kenyan labourers. select group of users in early February, a 23 year-old student from The chatbot was built with help from a Kenya-based data labeling team Germany decided to test its limits. who earned less than \$2 per hour, according to an investigation by TIME.

Forbes

ChatGPT's Use In Medicine Raises Questions Of Security, Privacy, Bias

Generative AI: Code generation

> .env		1	and the second s
> .git		2 - name: "Deploy Web Application Stack"	
> .vscode		3 hosts: all	
> webapp		4 become: true	
ansible.cfg		5 gather_facts: false	
≣ inventory		6 tasks:	
≣ logs.out		7 - name: "Install nginx and node;s 12 Packages on RHEL"	
≝ main.yaml	M	8 dnf:	
webapp_messa	ge	9 name:	
S webapp.comig ≤ webapp.service	1	10 - nginx	
	1	11 – nodejs	
	1	12 state: present	
	1	13 enablerepo: epel	
	1	<pre>uwhen: ansible_os_family == "RedHat"</pre>	
	1	15	
	1	16 – name: "Install NodeJS Dependencies"	
	1	17 npm:	
	1	18 path: /opt/app	
	1	19 name:	
		20 – express	
		- body-parser	
		22 – cookie-parser	
		23 – debug	
		– morgan	
		25 - serve-favicon	

Briefly outline the top 5 risks in using ChatGPT

Using ChatGPT, like any AI system, comes with certain risks. Here are the top five risks to consider:

Misinformation and Inaccurate Information: ChatGPT can generate plausible-sounding but false information, potentially leading to the spread of misinformation or incorrect advice.

Bias and Discrimination: ChatGPT may inadvertently produce biased or discriminatory content based on the data it was trained on, which can perpetuate stereotypes and offend users.

Privacy Concerns: If not properly configured and monitored, ChatGPT can generate text that unintentionally leaks sensitive information or violates privacy rights.

Malicious Use: There is a risk that ChatGPT can be used for malicious purposes, such as generating spam, phishing emails, or harmful content.

Security Vulnerabilities: Like all software, ChatGPT may have security vulnerabilities that could be exploited by malicious actors to gain unauthorized access or disrupt its functioning.

What is a good security policy for ChatGPT?

A good security policy for ChatGPT should be comprehensive, clear, and tailored to your specific use case and organization. Below is a template that outlines key components of a security policy for ChatGPT. Remember to customize it according to your organization's needs, industry regulations, and risk tolerance.

[Organization Name] ChatGPT Security Policy

1. Purpose and Scope

Define the purpose of this security policy, including the protection of data, resources, and the responsible use of ChatGPT within your organization.

Specify the scope, outlining which systems, users, and data this policy applies to.

2. Access Control and Authentication

Describe user authentication methods required to access ChatGPT.

Specify who is authorized to use ChatGPT and under what conditions.

Outline procedures for onboarding and offboarding users.

3. Data Privacy

Address the handling of sensitive or personal data when using ChatGPT.

Define data classification levels and how each should be treated. Specify data encryption, storage, and retention policies.

Generative AI: Image generation

DALL-E

Midjourney

Award-winning painting generated by Midjourney

Stable Diffusion

Foundation model risks

IBM's point of view on foundation model opportunities, risks, and mitigations outlines three categories of risk to help clarify potential risks and mitigation mechanisms.

	Input Risks associated with the content provided to foundation models	Output Risks associated with the content generated by foundation models	Other challenges Risks associated with how foundation models are used		
Traditional Risks known from earlier forms of Al	Data laws Privacy Robustness	Fairness	Transparency		
Amplified Known risks intensified by foundation models	Fairness Intellectual property Privacy Transparency	Explainability Misuse	Accountability Environment Human agency Human dignity Impact on jobs Legal uncertainty		
New Emerging risks intrinsic to the generative capabilities of foundation models	Intellectual property Value alignment Privacy Robustness	Fairness Harmful code generation Intellectual property Misuse Privacy Traceability Value alignment	Diversity and inclusion Impact on education Intellectual property		

Cost of a Data Breach Report

553 breaches studied 16 countries/regions

17 industries

18th year

Cost of a Data Breach 2023 Report

US average cost of a data breach (global \$4.45M)

82%

Breaches were based on Cloud data

\$1.4M

Complex Security Environments

\$181

Cost per stolen PII customer record

\$1N

Cost increase where remote work was a factor in the breach

\$850 thousand

Cost savings due to security AI and automation

20%

Top Attack Vector: Phishing Second: Compromised Credentials

- Discover and protect data across cloud environments
- Top Mitigating Factors
- Security AI and Automation
- Adopt an attacker's perspective of your organization's environment

Time to identify and contain

US average: 218 days (\$1.02M savings under 200 Days)

170 Identify

48 Contain

\$1.49 million

Average cost savings with incident response teams and IR testing vs. low levels in IR teams & testing

Factors that may increase or reduce the cost of a data breach

	DevSecOps approach	-723,001	
	Employee training	-617,494	
	Threat intelligence	-500,535	
	Security information and event management (SIEM)	-481,352	
Measured	Incident response (IR) plan and testing	-479,416	
	Security orchestration, automation and response (SOAR) tools	-478,655	
in USD	Insurance protection	-468,589	
	AI, machine learning-driven insights	-456,211	
	Board-level oversight	-424,600	
	Encryption	-401,289	
	IR team	-393,829	
	Endpoint detection and response tools	-387,571	
	Offensive security testing	-383,473	
	Proactive threat hunting	-382,822	
	Attack surface management (ASM) tools	-368,520	
	Identity and access management (IAM)	-322,761	
	CISO appointed	-263,513	
	Managed security service provider (MSSP)	-234,560	
	Data security and protection software	-167,754	
	Noncompliance with regulations		465,083
	Migration to the cloud		466,429
	IoT or OT environment impacted		466,429
	Third-party involvement		476,462
	Supply chain breach		477,602
	Remote workforce		482,116
	Security skills shortage		562,844
	Security system complexity		579,627
	-1,00	0.000 -800.000 -600.000 -400.000 -200.000	0 200,000 400,000 600,000 800,000

Total cost and frequency of data breaches by initial attack vector

Advanced Analytics

DETECT

Cognitive

Threat Hunting

The reality of expanding attack surfaces.

of assets are unknown or unmanaged due to rapid transformation. 1

76%

Organizations have been compromised by an unknown or unmanaged asset. ₂

50%

By 2026, non-patchable attack surfaces will grow to account for more than half of an enterprise's total exposure. 3 How Attack Surface Management Works

Build Resilience with Offensive Security

01. Discover

What does my attack surface look like to an adversary?

02. Prioritize

What makes my organization tempting to an adversary?

03. Authorize

What assets should be in scope for continuous validation?

04. Validate

How can I test my program free from assumptions?

05. Customize & Repeat

Can an adversary reach my objective?

ASM & CAASM

Complementary but Different Solutions

Attack Surface Management (ASM)

Discover Shadow IT & Validate Risk of Internet Facing Assets

Cyber Asset Attack Surface Management (CAASM)

Aggregates and Consolidates Asset Data from Known Sources

Tells you what you don't know.

Organizes what you already know.

XDR - Extended Detection and Response

Phishing Kit Hosted on Coronavirus-Themed Domain

Last updated: Apr 6, 2020, 9:32 PM

Top relevant threats

Pri

The to p sev Up and sur thre ind

Fa Ca Ea

relevant infeats									
oritize threats for your organization e X-Force Threat Score helps you prioritize threats based on rerity and risk. date your organization profile d connect a data source to face relevant threats, refine eat scores, and scan for threat icators.	on 74%	 Threat severity Indicator risk My organization profile My environment 	Early (Net Can Earl X-Forr 21	warning flix Squatting npaign - X-Force y Warning :e Threat Score % an now	Early warni Google Campa Early W X-Force Th 21%	ng Squatting ign - X-Force /arning reat Score			
^{y warning} cebook Squatting mpaign - X-Force rly Warning	Early warning WellsFargo S Campaign - Early Warnin	Squatting X-Force ng	Ea D C E	abuse@namecheap.com Filters Q Find filters		73 results for "abuse@nan	necheap.com"		
			Fil	ters	73	Threat activity		Utility Relief Abuse from Threat Actors Last updated: Apr 28, 2020, 12:20 AM	
			Ň		(21) (4)	Threat activity		OneTone Vulnerability Leads to JavaScript Cookie Hija Last updated: Apr 20, 2020, 11:42 PM	king:
				 Malware analysis Threat group profile 	(2) (3)	Threat activity		GitHub Users Targeted In A Sawfish Phishing Campaig Last updated: Apr 20, 2020, 10:45 PM	n
			~	✓ Indicator report □ Vulnerability	(24)	Threat activity		Malwarebytes Brand Abused In Malvertising Campaign Last updated: Apr 10, 2020, 3:59 PM	
				🗆 Signature	(18)				

🗆 URL

Threat activity

Public Service Announcement Nationwide Credit Reporting Agencies

Innovis

Information that empowers

experian

Trans

Lockdown Your Credit Reporting

• <u>https://www.equifax.com/</u>

<u>https://www.experian.com/</u>

<u>https://www.transunion.com/</u>

• https://Innovis.com/

Generative Al growing pains, security value and implications

Michael Melore, CISSP

IBM Senior Cybersecurity Advisor Chair SecRT (Security <Leaders> Round Table) FBI InfraGard Committeemen NASCIO Privacy and Cyber Security Committees CISA / ATARC Cyber Defense Task Force

@MichaelMelore mmelore@us.ibm.com