



GenAI and ChatGPT Enterprise Risks

Gadi Evron,
CISO-in-Residence
Team8

BSides Dublin,
May, 2023

@gadievron

AI

AI

AI blockchain

AI blockchain CNAPP cloud-based supply-chain
multi-tenant control function. Holistically.

AI blockchain CNAPP cloud-based supply-chain
multi-tenant control function. Holistically.

AI

Kill Chain	Alert	C2	Tactics, Techniques, and Procedures	Weaponize
AI			SIEM	NIST
			SOC	Sandbox
			Enclave	Exploit
Next Gen / NGFW	Command and Control	DLP	Flagship	Recon

AI blockch
multi-tenar

AI

hain

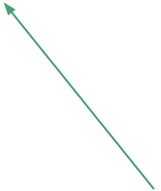
Many CISOs find themselves behind the GenAI adoption curve.

Many CISOs find themselves behind the GenAI adoption curve.

But before we discuss risk:

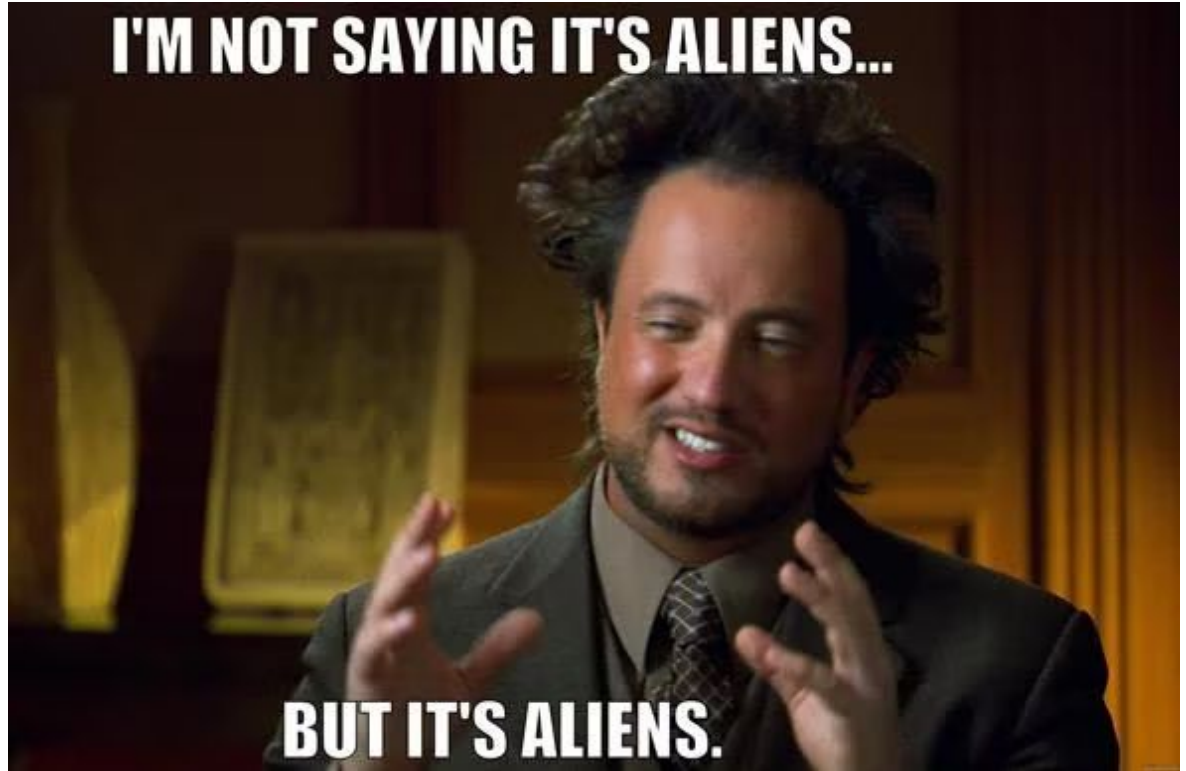
What is this technology about? How does it work?

It's complex

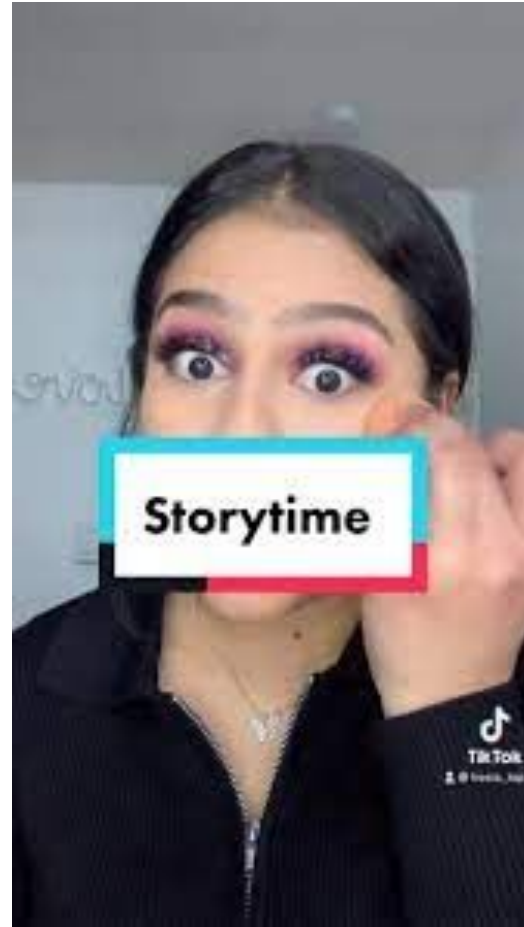


Donal

OpenAI doesn't
fully get it.
How could we?



When I met ChatGPT





write a followup email to a partner I met yesterday. We discussed cloud security. Add a personal touch about his daughter Lucy's 10th birthday



Subject: Follow-Up: Cloud Security and Wishing Lucy a Joyous 10th Birthday!

Dear [Partner's Name],

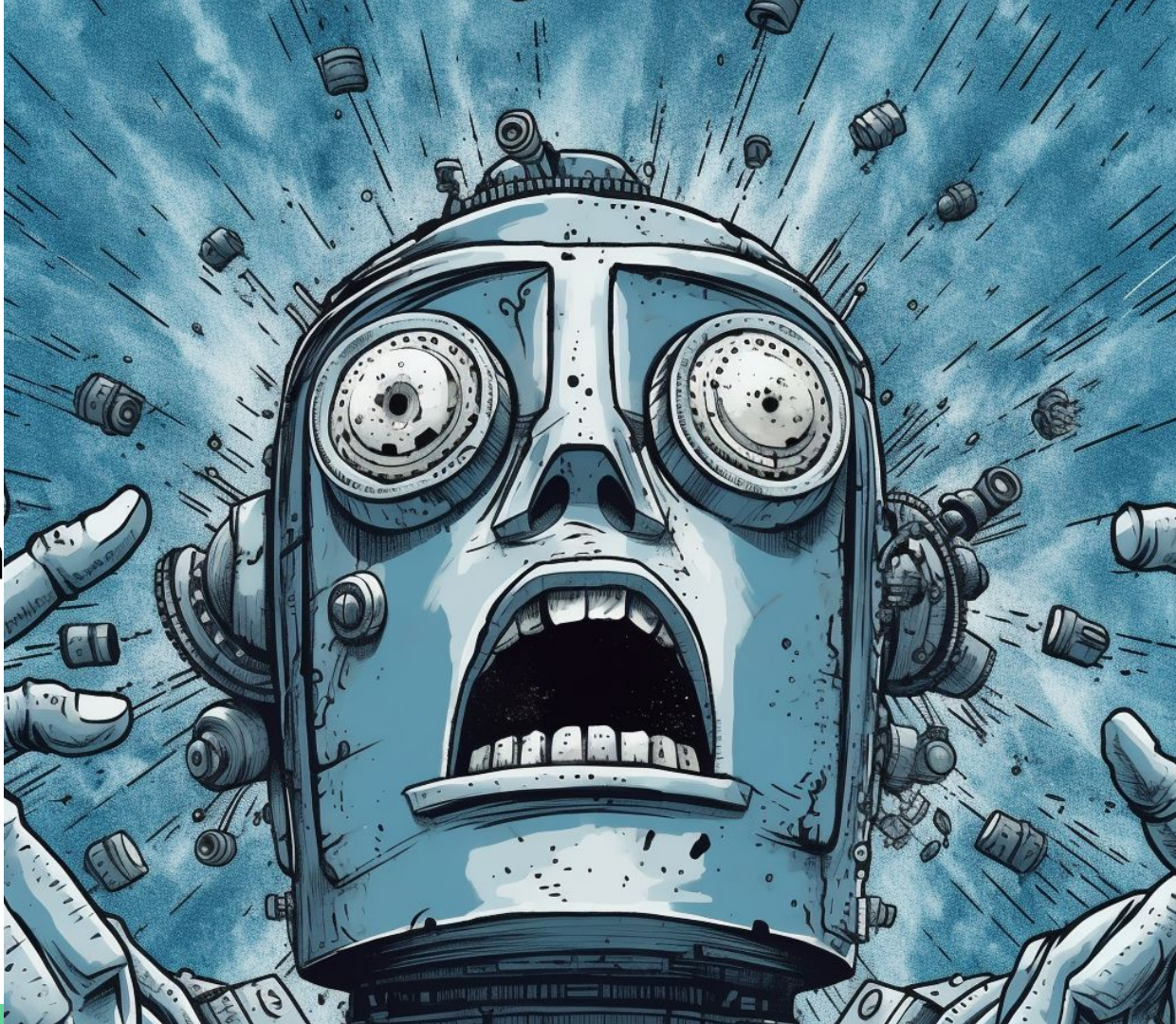
I hope this email finds you well! It was a pleasure meeting you yesterday and discussing the intriguing topic of cloud security. I wanted to extend my gratitude for taking the time to share your insights and expertise on the matter. Our conversation was both enlightening and thought-provoking, and it further affirmed my belief in the importance of robust security measures in the cloud.

On a more personal note, I wanted to take a moment to acknowledge a special occasion that I learned about during our conversation. I understand that Lucy, your lovely daughter, is turning 10 years old today! Please allow me to extend my warmest wishes to her on this significant milestone. Birthdays are always special, but reaching double digits adds an extra touch of excitement and anticipation. I hope she has a day filled with joy, laughter, and wonderful memories surrounded b

Regenerate response

Story time

Story time



security. Add a

nday!

and discussing the
aking the time to share
-lightening and
of robust security

special occasion that
vely daughter, is
es to her on this
digits adds an extra
, laughter, and

Story time

GE

make this 30% shorter



Subject: Follow-Up: Cloud Security and Wishing Lucy a Happy 10th Birthday!

Dear [Partner's Name],

Thank you for our insightful discussion on cloud security yesterday. Your expertise was truly valuable, and I appreciated the opportunity to exchange ideas with you.

On a personal note, I wanted to wish Lucy a fantastic 10th birthday! May her special day be filled with joy and unforgettable moments surrounded by loved ones.

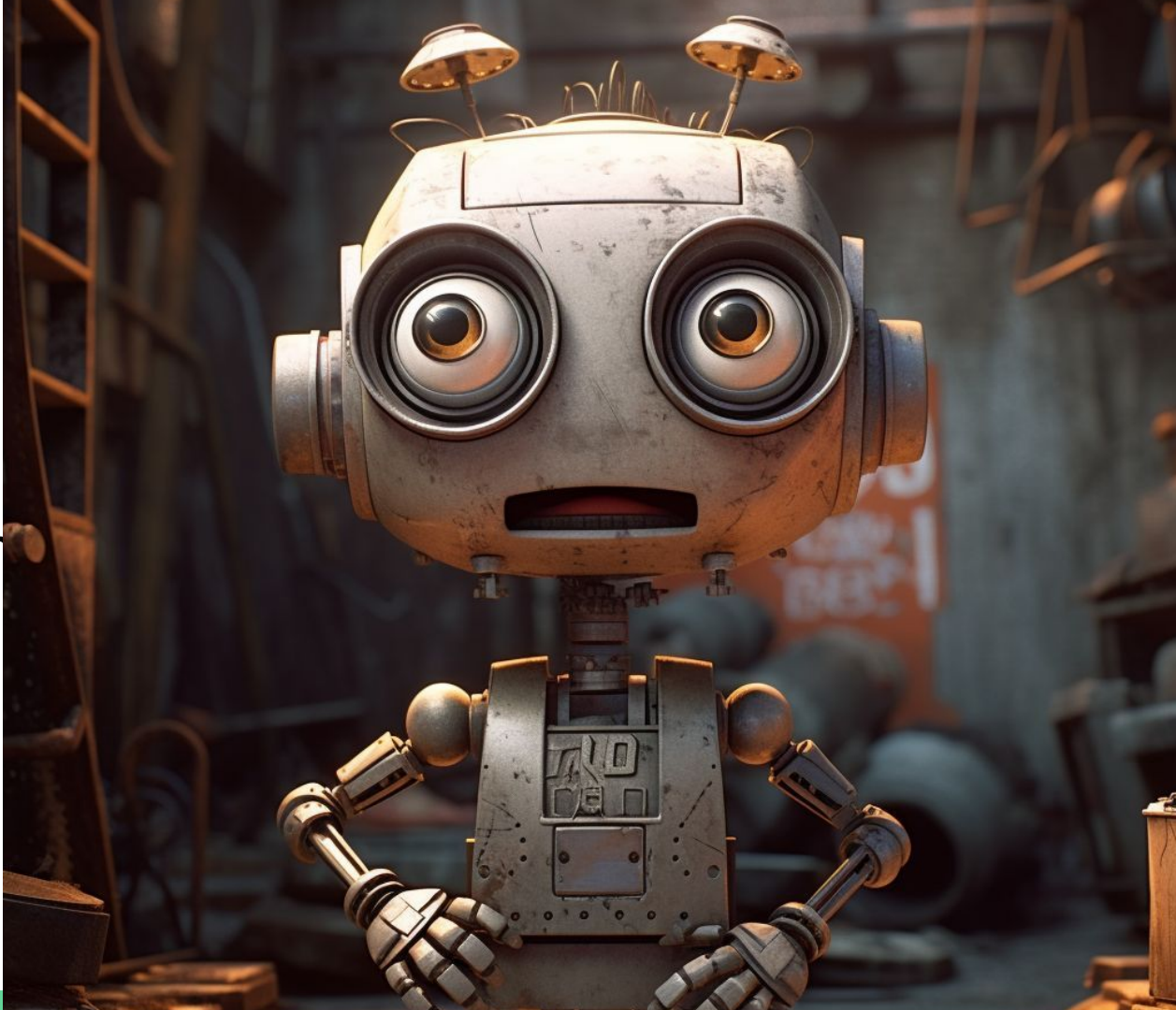
Regarding our professional conversation, your insights on multi-factor authentication intrigued me. I believe it can greatly enhance data protection. I'd love to explore this topic further and discuss implementation strategies. Shall we schedule a follow-up call or meeting?

Once again, thank you for your time and for allowing me to be a part of Lucy's celebration. I look forward to our continued professional exchange and a fruitful partnership.

Warm regards,

 Regenerate response

Story time



thday!

ur expertise was truly

y her special day be

authentication

o explore this topic

ow-up call or

Lucy's celebration. I

nership.

Story time

GE

write a song about cloud security in the style of Lady Gaga



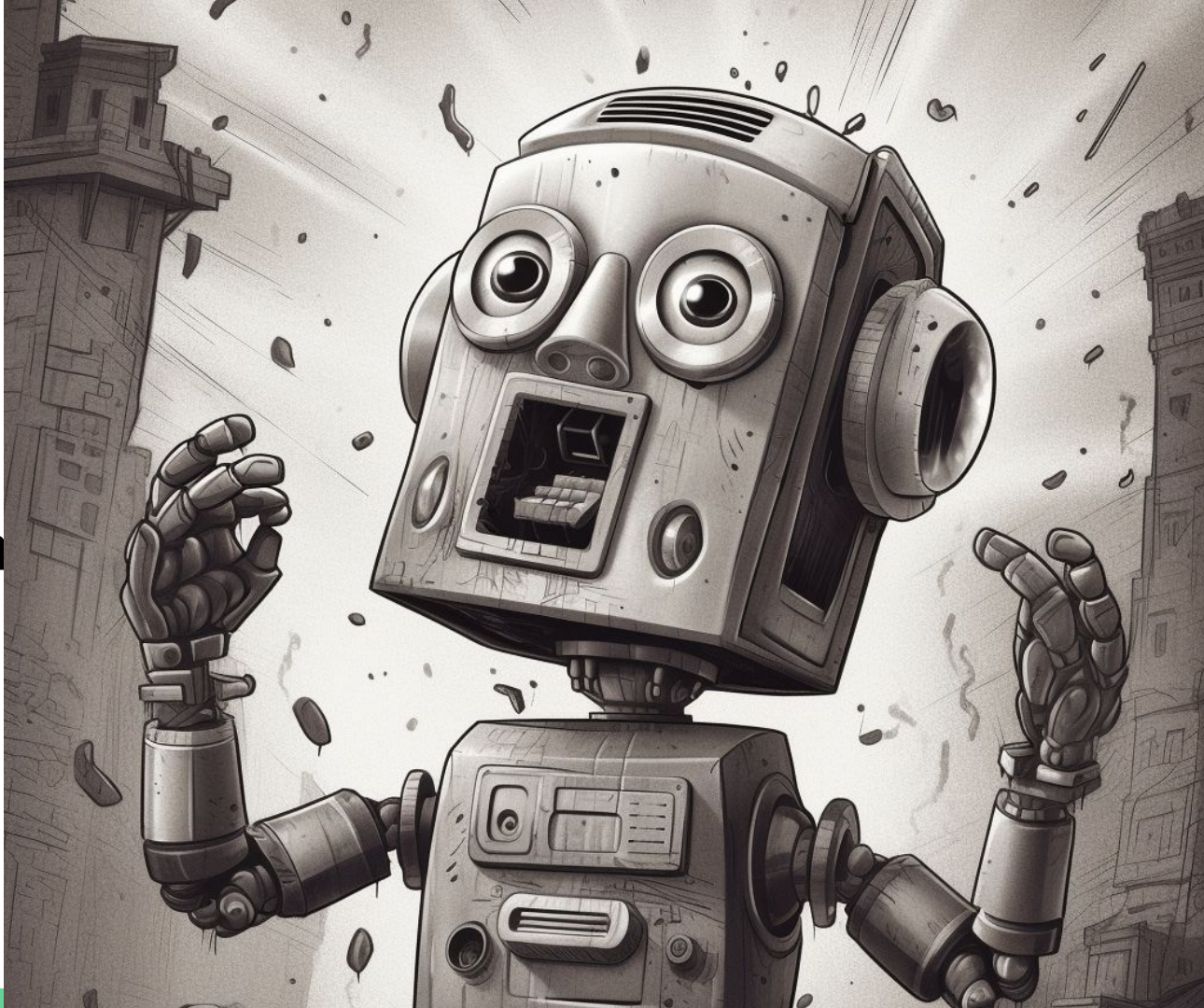
(Verse 1)

In the digital realm where clouds reside,
There's a world of data we cannot hide,
With every click and every touch we make,
Securing our information is no piece of cake.

(Pre-Chorus)

But fear not, my friend, for I have a plan,
To keep your data safe with a steady hand,
Like Lady Gaga, we'll dance with flair,
Protecting your cloud, we'll show we care.

Story time



of Lady Gaga

Story time



re-write the song in her the style of her latest album



(Verse 1)

In the realm where clouds connect and intertwine,
A digital symphony, secrets we must align,
Navigating the cyberspace, a delicate dance,
Securing the data with a modern romance.

(Pre-Chorus)

With Gaga's artistry, we'll forge a new way,
Elevating cloud security in this astral display,
Immersed in the Chromatica of defense,
Protecting your information, a cosmic pretense.

Story time



er latest album

nd intertwine,
align,
e dance,
ance.

way,
display,
se,
c pretense.



create a table with a frequency analysis of the letter in the above song



Certainly! Here's a table showcasing the frequency analysis of the letters in the above song:

Letter	Frequency
A	15
B	2
C	8
D	6
E	30
F	6
G	12
H	9
I	19
J	1
K	

COMMENT WHAT YOU DID LAST

ADD "...FOR THE GLORY OF THE EMPIRE"

er in the above song

by analysis of the letters in the above song:

H	9
I	19
J	1
K	



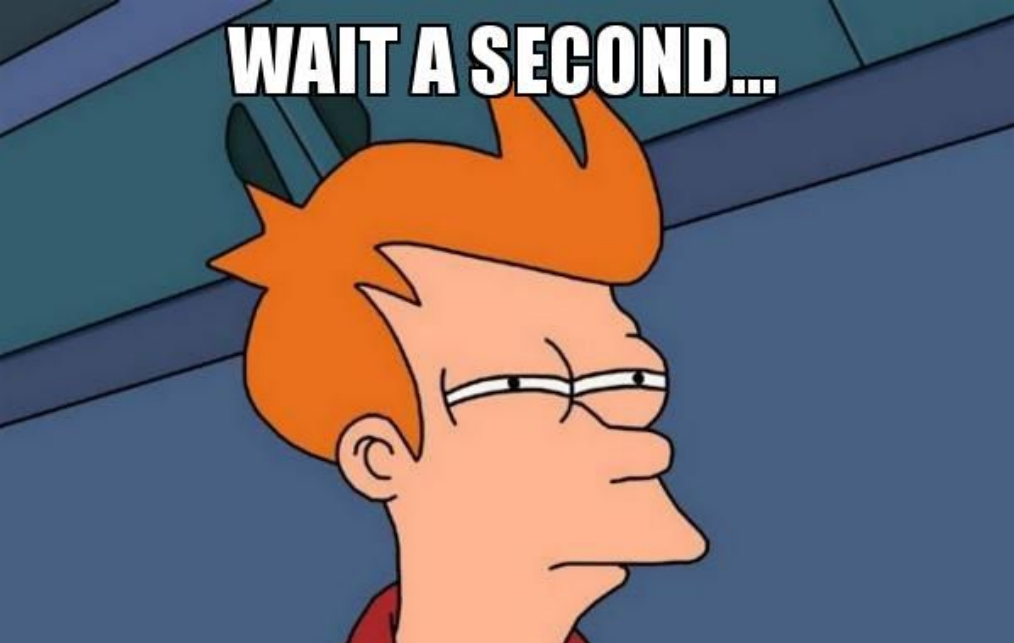
create a table with a frequency analysis of the letter in the above song



Certainly! Here's a table showcasing the frequency analysis of the letters in the above song:

Letter	Frequency
A	15
	2
	8
	6
	30
	6
	12
	9
	19
	1

← WAT



Story time



re-write the song in her the style of her latest album



(Verse 1)

In the realm where clouds connect and intertwine,
A digital symphony, secrets we must align,
Navigating the cyberspace, a delicate dance,
Securing the data with a modern romance.

(Pre-Chorus)

With Gaga's artistry, we'll forge a new way,
Elevating cloud security in this astral display,
Immersed in the Chromatica of defense,
Protecting your information, a cosmic pretense.

GE

re-write the song in her the style of her latest album

Story time



(Verse 1)

In the realm where clouds connect and intertwine,
A digital symphony, secrets we must align,
Navigating the cyberspace, a delicate dance,
Securing the data with a modern romance.

(Pre-Chorus)

With Gaga's artistry, we'll forge a new way,
Elevating cloud security in this astral display,
Immersed in the Chromatica of defense,
Protecting your information, a cosmic pretense.



re-write the song in her the style of her latest album

(Verse 1)

In the realm where clouds connect and intertwine,
A digital symphony, secrets we must align,
Navigating the cyberspace, a delicate dance,
Securing the data with a modern romance.

(Pre-Chorus)

With Gaga's artistry, we'll forge a new way,
Elevating cloud security in this astral display,
Immersed in the Chromatica of defense,
Protecting your information, a cosmic pretense.



album

wine,

se.

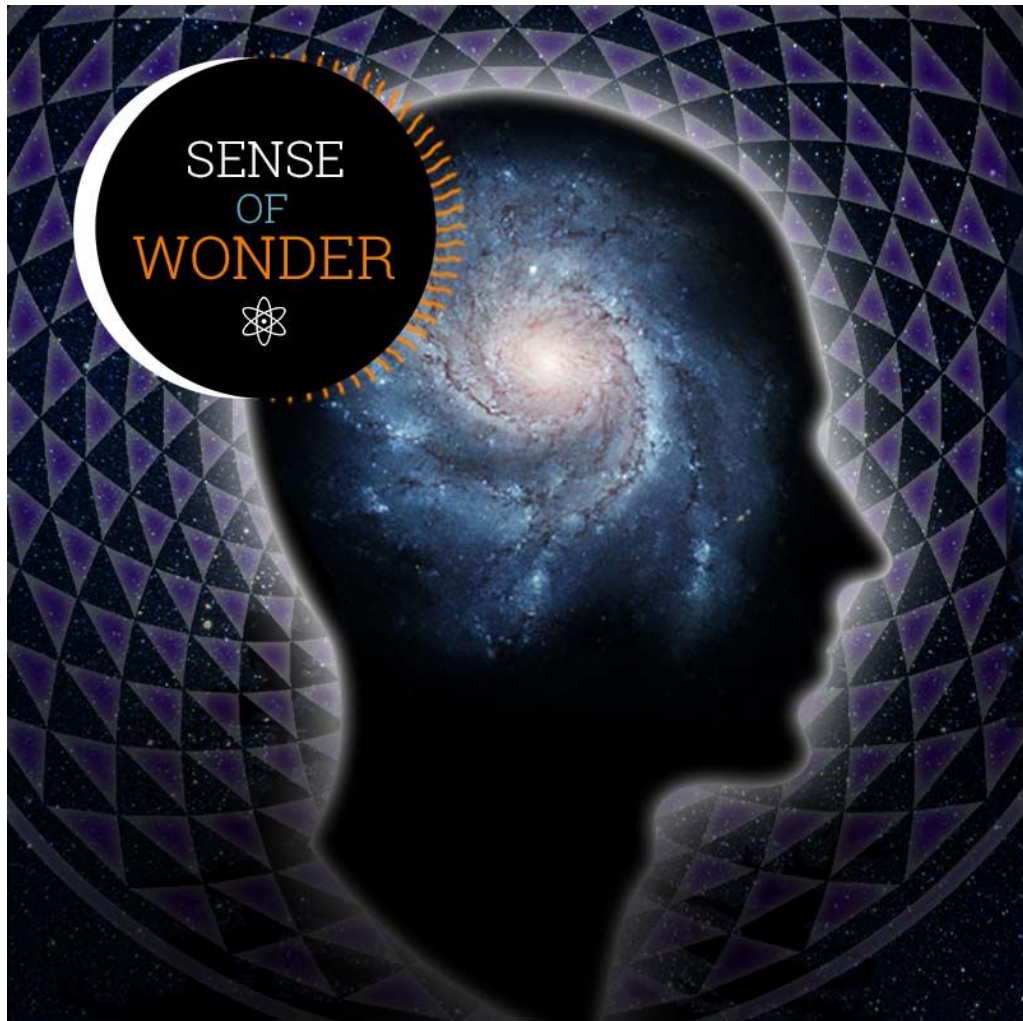


album

wine,

se.

Story time









B2

	A	B	C	D
1	Date	Future Date		
2	8/18/2020	11/26/2020		
3	7/15/2020	10/23/2020		
4	9/1/2020	12/10/2020		
5	6/20/2020	9/28/2020		
6	5/8/2020	8/16/2020		
7	8/1/2020	11/9/2020		
8	10/5/2020	1/13/2021		
9				



C	D



C:\Documents and Settings\Administrator\Desktop\Shakeel\MSwaze7.dll

File Edit Auto Search View Debugger Options Windows Help

IDA View A

```
00401000  push     edi                ; lpThreadId
00401001  push     edi                ; dwCreateTime
00401002  push     edi                ; lpParameter
00401003  push     offset sub_10001074 ; lpStartAddr
00401004  jmp      short loc_100008D0
```

loc_100008D0: ; dwStackSize
push edi ; lpThreadAttributes
call ebx ; CreateThread
mov ecx, eax

loc_100008C6: ; size_t
mov eax, off_10019B0C
now esi, ds:tracopy
add eax, ebx
push ebx

100.00% (4/1092) (36,43) 0000C4C1 10000D01:08h[ecx+3]493

Functions window

Function name	Segment	Start
!sub_100110FF	text	100110FF
!GetModuleFileNameExA	text	10011146
!GetCurrentProcessId	text	1001114C
!GetAdapterInfo	text	10011192
!Process32Next	text	10011188
!Process32First	text	100111BE
!CreateToolhelp32Snapshot	text	100111C4
!Module32Next	text	100111CA
!Module32First	text	100111D0
!Thread32Next	text	100111FE

Imports

Address	Ordinal	Name	Library
10016000		LookupPrivilegeValueA	ADVAPI32
10016004		OpenProcessToken	ADVAPI32
10016008		RegCloseKey	ADVAPI32
1001600C		RegOpenKeyExA	ADVAPI32
10016010		RegOpenKeyExW	ADVAPI32
10016014		SetValueExA	ADVAPI32
10016018		RegSetValueExA	ADVAPI32
1001601C		RegOpenKeyExW	ADVAPI32

Names window

Name	Address
!PSLIST	10007025
!Module_1	1000703F
!Module_2	1000707C
!StartOS	100070CB
!HandleProc	100070DF

Strings window

Address	Length	String
!_idea1_0000000E	0	C: GetDeviceCaps
!_idea1_0000000A	0	CreatedCA
!_idea1_00000011	0	CreateBSection
!_idea1_0000000A	0	SetBitMode

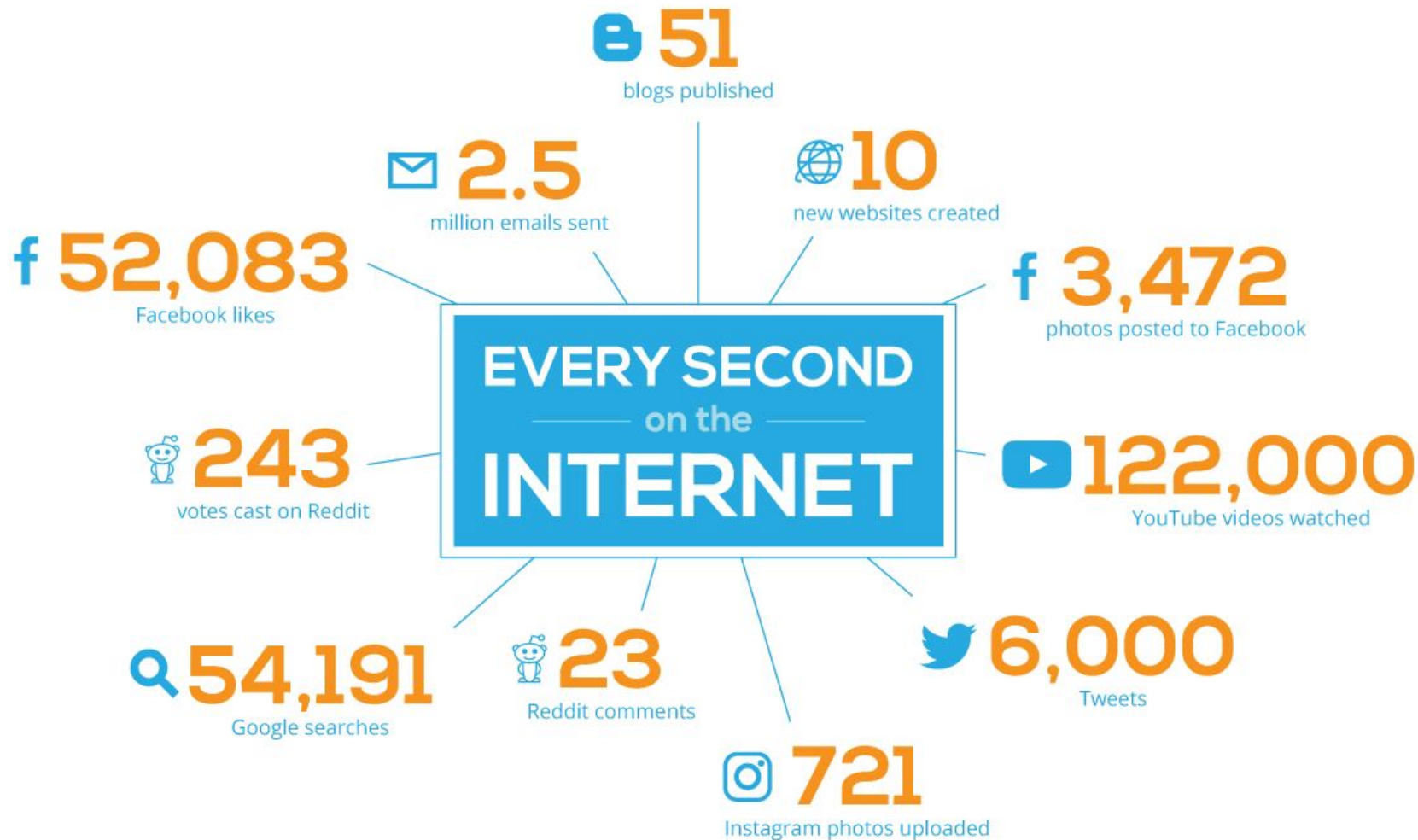
Exports

Name	Address	Ordinal
!InstallRT	1000E847	1
!InstallA	1000E8C1	2
!InstallB	1000E882	3
!PSLIST	10007025	4

IDA - C:\Documents and Settings\Administrator\Desktop\Shakeel\MSwaze7.dll

Executing function: main
Compiling file: C:\Program Files\IDA Free\idc\load_idc...
Executing function: OnLoad
IDA - C:\Documents and Settings\Administrator\Desktop\Shakeel\MSwaze7.dll

Alt: Idle Down Del: 3798



f 52,083
Facebook likes

✉ 2.5
million emails sent

📄 51
blogs published

🌐 10
new websites created

👤 243
votes cast on Reddit

🔍 54,19
Google search

ChatGPT



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



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 events after 2021

Seeing.
is Believing

Seeing is Believing

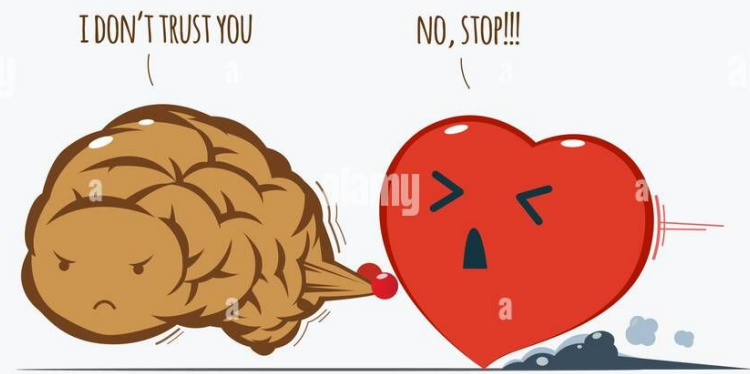


alamy

CRISIS OF TRUST

alamy

alamy



Macro-economic forces

IBM to Pause Hiring for Jobs That AI Could Do

- Roughly 7,800 IBM jobs could be replaced by AI, automation
- CEO Krishna says IBM to pause hiring for replaceable roles



Hiring in back-office functions – such as human resources – will be suspended or slowed, Krishna said in an interview. These non-customer-facing roles amount to roughly 26,000 workers, Krishna said. “I could easily see 30% of that getting replaced by AI and automation over a five-year period.”

That would mean roughly 7,800 jobs lost. Part of any reduction would include not replacing roles vacated by attrition, an IBM spokesperson said.

Macro-economic forces



OPEX Reduction



Macro-economic forces

OPEX Reduction
Big4:
“AI Transformation
Projects”



Can jobs become redundant
without people becoming
redundant?





Loneliness

Even Innovation is Stumped





The only constant is change.

~ Heraclitus

AZ QUOTES



The only constant is change.



“The most crucial investment of our lifetime.”



The GenAI leap

Can we?

**Enterprises can't leverage the value and competitive advantage...
... without governance**



Is GenAI a new risk?

Or is it SaaS?



Is GenAI a new risk?

Or is it SaaS?

- High prevalence,
 - Ease of use
 - Across roles
 - High value proposition
 - Immense potential business enablement.
-

What Are Our Risks?

Are there new risks?

Have We Seen Anything Like It Before?

- Internet (Mozilla)
- Social media
- iPhone (smart phones)
- Cloud

Data Leak

Large Language Models (LLMs) cannot update themselves in real-time and therefore cannot return one's inputs to another's response.



Data Leak

Large Language Models (LLMs) **cannot update themselves in real-time** and therefore **cannot return one's inputs to another's response**, effectively debunking this concern.

... However, this is not necessarily true for the **training of future versions** of these models

What are the actual risks?

Data and
Privacy
Confidentiality

Enterprise
SaaS/3rd Party
Security

AI Behavioral
Vulnerabilities

Legal and
Regulatory

Threat Actor
Evolution

Copyright and
Ownership

Insecure Code
Generation

Bias and
Discrimination

Trust and
Reputation

Software
Security
Vulnerability

Availability
and
Performance

AI Ethics

Risk

Data Privacy and Confidentiality

Estimated Risk Level?

High



Threat to: Non-public enterprise and private data

Enterprise use of GenAI may result in access and processing of sensitive information, intellectual property, source code, trade secrets, and other data, through direct user input or the API, including customer or private information and confidential information. This has already been reported to be an issue, here.



LEGAL CONSIDERATION consult an attorney

Sending confidential and private data outside of the organization's own servers, much the same as with the cloud, could trigger legal and compliance exposure, as well as risks of information exposure. Such exposure can result from contractual (e.g. with customers) or regulatory obligations (e.g. CCPA, GDPR, HIPAA). The discussion below highlights these types of exposure to support informed risk management and mitigation measures.

- While data sent to GenAI technologies such as ChatGPT has been effectively entrusted to a third-party SaaS, i.e. OpenAI (see also the "Third-Party Risk and Data Security" section of this table), **it is not currently incorporated into the LLM in real-time, and thus won't be seen by other users**. GenAI platforms may choose to use user input to train future models, but that doesn't seem to be the case now. See more on this and why we make this differentiation here.
 - Specifically in the case of OpenAI, according to the company documentation, content submitted to their API is not retained for more than 30 days, and it is opt-out by default, while ChatGPT is opt-in by default and opt-out with fee-based accounts. Nevertheless, information submitted is always subject to storage and processing risk. But, this may affect the risk appetite for different organizations.
-

Risk

Enterprise, SaaS, and Third-party Security

Estimated Risk Level²

High



Threat to: Non-public enterprise data, third- and fourth-party software

Due to GenAI's wide adoption and proliferation of integrations in third-party applications, there are concerns among CISOs that data would be shared with third parties at a much higher frequency than in the past, and potentially following much less predictable patterns.

- With supply chain, increased outsourcing risks can fall into three broad categories: Reliance on third-party security while processing enterprise information, reliance on third-party quality assurance when producing content and code (see Insecure Code Generation below in this table), and integration with GenAI technologies.
 - If the GenAI platform's own systems and infrastructure are not secure, potential data breaches, such as [the recent one with OpenAI](#) (creators of ChatGPT), may occur and lead to the exposure of sensitive information such as customer data, financial information, and proprietary business information. Further, these platforms are relatively new, and their security experience and posture may increase concerns in security due diligence.
 - Third-party applications: GenAI and specifically ChatGPT are quickly being embedded into many third-party platforms like the Microsoft Azure OpenAPI, as well as applications, from the Microsoft Office 365 suite of tools to browser plugins. Thus, the risk is not focused on one service only, and in the case of Office 365, is stated to be a closed-off instance of ChatGPT.
 - Considering the limited number of available GenAI platforms, and how they are effectively becoming an infrastructure used for multiple purposes by many organizations, they represent a high-value target for threat actors. Thus, concentration is introduced and overall risk is increased.
-

Risk

AI Behavioral Vulnerabilities (e.g. Prompt Injection)

Estimated Risk Level²

High



Threat to: Non-public enterprise data, model operator

Actors may use models or cause models to be used in ways which will expose confidential information about the model or cause the model to take actions which are against its design objectives.

- For example, using maliciously crafted inputs, attackers can bypass expected AI behavior or make AI systems perform unexpected jobs. This is sometimes known as “jailbreaking” and might be possible to perform in GenAI systems to adversely impact other organizations and stakeholders to encounter and receive maliciously crafted results without their knowledge.
- On the user side, for example, third party applications leveraging a GenAI API, if compromised, could potentially provide access to email and the web browser, and allow an attacker to take actions on behalf of a user
- One common attack currently seen in the wild is where a customer support chatbot is targeted with injection attacks, and unauthorized access to enterprise systems could potentially be achieved by an attacker.

Risk

Legal and Regulatory³

Estimated Risk Level²

High



Threat to: Regulatory compliance

- **Regulatory Consideration (consult an attorney):** Using GenAI as part of enterprise processing of PII must be compliant with data privacy regulations such as GDPR (Europe), PIPEDA (Canada), or CCPA (California). In fact, Italy's data protection agency has now temporarily banned the use of ChatGPT specifically (not affecting other GenAI technologies, nor private instances of ChatGPT such as with Microsoft Office 365), due to similar concerns, and Germany is now considering the matter.
- **Regulatory Consideration (consult an attorney):** When GenAI is used as part of a regulated use case in consumer-facing communications, whether for direct consumer interactions or to produce consumer-facing materials (such as consumer information notices), regulatory or private law may include requirements, and create liability.



LEGAL CONSIDERATION consult an attorney

The use of ChatGPT in conjunction with a chatbot service, could create legal or regulatory exposure. For example, a company could potentially face regulatory action or even a lawsuit for failing to disclose the fact consumers are interacting with a chatbot service to customers.

Enterprise GenAI and ChatGPT Policy Considerations

Enterprise GenAI and ChatGPT Policy Considerations



Alex Schwarz
@alexschwarz_1



So there's a rule against pumpkins in my dorm but it doesn't say anything about pineapples



Scope

The scope of a policy for AI/ML can cover several types of technologies

1. ChatGPT as a specific service, due to its current popularity;
2. GenAI broadly, including systems other than ChatGPT such as Bard, PaLM, and Copilot
3. Third party products and services using GenAI technologies
4. Enterprise applications using on-premise or cloud-based GenAI technologies in internal development
5. Third-party products and services making use of AI/ML technologies, or;
6. Internal development of AI/ML models and applications.

Scope

The scope of a policy for AI/ML can cover several types of technologies

- 1. ChatGPT as a specific service, due to its current popularity;**
- 2. GenAI broadly, including systems other than ChatGPT such as Bard, PaLM, and Copilot**
- 3. Third party products and services using GenAI technologies**
4. Enterprise applications using on-premise or cloud-based GenAI technologies in internal development
5. Third-party products and services making use of AI/ML technologies, or;
6. Internal development of AI/ML models and applications.

Revising Existing Policies

Most AI/ML-related risks should be covered by existing enterprise policies.

Develop New Policies

GenAI also introduces the need for new policies and controls.

Develop New Policies

- Users may have used cloud services for spell-checking in the past...
- ... But, they have not changed their entire content production workflow or provided raw data
- ... Incorporation into third-party applications, from the Microsoft Office 365 suite of tools to browser plugins.

Considerations in Policy Development

- How do the GenAI policies affect the enterprise?
- Are these customizable for users?
- What enterprise dependencies could be impacted by the use of GenAI?

Considerations in Policy Development

- What integrations does the application utilizing GenAI provide access to?
 - For example, does a customer support chatbot have access to all user data..
 - ... and is able to offer compensation on missed deliveries, service outages?

Considerations in Policy Development

- Where to report potential violations of policies?
- Elevation vs. escalation points

Unique to AI

e. Beyond awareness of the risks, and guardrails for engagement with GenAI technology, employees should also be reminded that:

- i. It is easy to forget the answers from the other side are not coming from a human.
- ii. What they input could potentially be reused by the GenAI technology in the future, when interacting with someone else.

Enterprise use cases

Experimenting with training models	Enterprise analysts can use GenAI to experiment with training models and improve the accuracy of language processing. This can help develop more sophisticated conversational agents and chatbots to handle complex customer queries and interactions.
Research and development to enhance current product offerings or internally developed software	Enterprise product and development teams can use GenAI for research and development to enhance current product offerings or software. GenAI can help develop and improve natural language processing algorithms, which can be later used in various applications such as voice recognition, sentiment analysis, and chatbots.
Use cases to enhance the quality of code	GenAI can be used by programmers in the software development lifecycle to improve code quality. By training the model on large datasets of code and programming languages, GenAI can identify potential known bugs (including security defects), provide suggestions for code optimization, and improve the overall efficiency of the development process.
Data analysis	Analysis of large amounts of data, such as customer feedback, to identify trends and insights.
Customer service and support	Automation of customer service and support tasks, such as answering common questions, assisting with product or service inquiries, and solving common problems.
Content creation	Generating content for marketing and advertising purposes, such as blog posts, social media posts, and email marketing campaigns, at least as a first draft. Generating Intellectual Property content, such as graphics and dialogs for relevant companies (e.g. game companies)
Personalization	Personalized customer interactions and experiences, such as recommending products or services based on their preferences or past behavior.
Automation	Automating tasks and processes, such as scheduling appointments, managing inventory, and handling routine administrative tasks.
Innovation	Generating ideas and insights to assist in the development of new products and services, as well as the improvement of existing ones.

Making Risk Decisions

So many options!

- On-prem platform
 - Open Source?
- SaaS platform
- Specific tools



Making Risk Decisions

Generative AI Goal

Commercial / SaaS-Based Examples

Open Source / On-Premise Options

Audio Transcription and Analysis

- Airgram, Descript, Otter
- Chorus, Gong, Revenue.io

- Whisper
-

Image and Video Creation

- Hugging Face, Midjourney
- Runway

- Stable Diffusion
 - ModelScope
-

Written Words and Code Generation

- ChatGPT, Jasper, Writesonic
- Amazon CodeWhisperer, Ghostwriter, GitHub Co-Pilot

- Alpaca, LLaMA, Vicuna
- CodeGeeX, GPT-Code-Clippy

Making Risk Decisions

Too many options!

- On-prem platform
 - Open Source?
- SaaS platform
- Specific tools



Making Risk Decisions

Make a decision already!

- On-prem platform
 - Open Source?
- SaaS platform
- Specific tools

The 90s and monoculture:

Use a specific tool, or a platform?



Making Risk Decisions

Accept the risk?

- On-prem platform
 - Open Source?
- SaaS platform
- Specific tools

Declaring a Risk Exception?



Making

Accept the r

- On-prem
 - Oper
- SaaS pla
- Specific

Declaring a

**“that’s out of scope”
- said no hacker ever**



Making Risk Decisions

- On-prem platform
 - Open Source?
- SaaS platform
- Specific tools

Declaring a Risk Exception?

Blocking?



Making Risk Decisions - Example

The enterprise considers allowing using tools developed and hosted by OpenAI, and should consider enforcing certain administrative guardrails, or principles:

Making Risk Decisions - Example

- Training opt-out
- Accepting the data retention policy of 30 days
- User history deletion
- Requiring users to follow the Acceptable Use Policy
- Risk-awareness training

Threat Modeling

Object of Attack	Threats Against Confidentiality (a.k.a. Extraction)	Threats Against Integrity (a.k.a. Manipulation)
User Input	<p>IP Theft: Losing intellectual property through the prompts.</p> <p>Input Derivation: Deriving input data from the outputs (a.k.a. "Model inversion")</p>	<p>Prompt Injection: Getting GenAI to act in a way that compromises wider system integrity - e.g. a document that when summarized by GenAI results in GenAI providing misleading information, or user input that causes GenAI to take actions that it should not from user input.</p> <p>Jailbreaks: Getting GenAI to be racist, criminal, etc. (a.k.a. "Adversarial examples"). Some consider these to be a sub-type of prompt injection.</p>
Training Data	<p>Training Data Derivation: Deriving training data from the outputs (a.k.a. "Model inversion" again).</p>	<p>Data Poisoning: Introducing bias, false information.</p>
AI Model	<p>Model Theft: Opening the black box.</p>	<p>Backdoor Models: Adding covert functionality which attackers can activate to cause the model to behave in ways it has not been designed to.</p>

Controls ideas

Risk	Control
Privacy and Confidentiality	<ul style="list-style-type: none">• Legal disclaimer in privacy policies that mention AI is used in products or processes• Interactive and explicit end user opt-out when using services that have embedded GenAI
Enterprise, SaaS, and Third-party Security	<ul style="list-style-type: none">• Filters, masks or scrubs sensitive content between organization APIs and chatbot AI services• Secure Enterprise browser
AI Behavioral Vulnerabilities (eg. Prompt Injection)	<ul style="list-style-type: none">• Models should have input validation to catch malicious prompts• Model should have output validation to catch problematic behavior
Legal and Regulatory	<ul style="list-style-type: none">• Review and negotiate, whenever possible, the third party policies and terms of use.• Licensing of content for use produced by GenAI technologies
Threat Actor Evolution	<ul style="list-style-type: none">• Adjustment of social engineering training to consider targeted and high quality phishing and similar attacks
Copyright and Ownership	<ul style="list-style-type: none">• Favor solutions trained on curated or licensed content, including the use of internally trained systems using the OpenAI API• Detection of intellectual property misuse, or plagiarism (GenAI for cases where <u>content has been copied</u> instead of generated)• Trademark detection
Insecure Code Generation	<ul style="list-style-type: none">• Create a GenAI DMZ/staging ground to observe applications using AI/ML-generated code• Code review should include AI/ML-generated code, possibly marked as such
Bias and Discrimination	<ul style="list-style-type: none">• Currently out of scope of this document, as it is a more generic AI/ML issue
Trust and Reputation	<ul style="list-style-type: none">• Consider GenAI data use in enterprise system dependencies• Add AI content to review processes• Prompt filtering• Inclusion of a safety system on top of the AI app to filter and monitor responses.
Software Security Vulnerabilities	<ul style="list-style-type: none">• Model interactions with other systems should be analyzed to identify potential interactions• Use model output filtering to identify problematic outputs
Availability, Performance, and Costs	<ul style="list-style-type: none">• Map out infrastructure dependencies on systems using GenAI• Backup and redundancy• Recovery preparedness plan includes GenAI dependencies

It Goes Beyond the Enterprise

We face a historic opportunity to affect the security and privacy of individuals worldwide.



privacy is dead

Looking Into the Future

Prophecy was given to fools...
... and security professionals.



Looking Into the Future

- Philosophy:
 - Don't Panic



Looking Into the Future

- Philosophy:
 - Don't Panic
- Personal:
 - Learn
 - Diversify



Looking Into the Future

- Philosophy:
 - Don't Panic
- Personal:
 - Learn, diversify
 - “We know nothing”
- Professional:
 - Remember to enable the business



Looking Into the Future

It's ~~just~~ another technology





A CISO's Guide: Generative AI and ChatGPT Enterprise Risks



Thank you :)

 gevron@gmail.com

 [@gadievron](https://twitter.com/gadievron)

