

A group of four hikers with backpacks are seen from behind, walking along a dirt trail on a grassy mountain slope. The sun is low on the horizon, creating a warm, golden glow and long shadows. In the background, there are rolling hills and a large, snow-capped mountain peak under a clear blue sky. The overall mood is adventurous and serene.

verinext

TRUST IN WHAT'S NEXT

Session Abstract

Every day, we network-enable more aspects of our professional and personal lives. We make conscious decisions to adopt technologies in the pursuit of convenience or an apparent need to feel connected. Often the decisions are thrust upon us by employers, service providers or even those we've come to rely on in the physical world. Each step towards deeper and total connectivity comes with consequences; risks associated with our personal privacy, the security of our business and even our safety. In this session, we will explore the changes we've already allowed into our lives, where things are leading, and portents for personal, business and public security.

Session Abstract

- ❑ People want Automation. Automation enables & allows Freedom. Manual is bad
- ❑ Automation is giving rise to Large Language Models (LLM) & Artificial Intelligence (AI)
 - Automation is both the ingredients and often the output of AI
- ❑ The Cyber Threat Environment is already driven by AI. Our defense capability, visibility & speed is lagging.
- ❑ Ironically, now we usher in an era of:
 - Discovering our AI Systems & Understanding Vulnerabilities to / from
 - Protecting our own AI Systems
 - Defending from adversarial AI Systems

Where do we start?

verinext

Historical Perspective





is Times Square station). It's very heavy (500lbs), it has facial-recognition 8K cameras, HD microphones and a panic button. The city has leased it for only \$9/hr! Ohh... and If you tamper with it, you're going to Jail.



NYC is going robot patrols

New York's mounted unit has shrunk considerably over the last decade: it now has 79 police officers and 60 horses, down from the 130 officers and 125 horses it had before, Deputy Commissioner Paul J. Browne said.

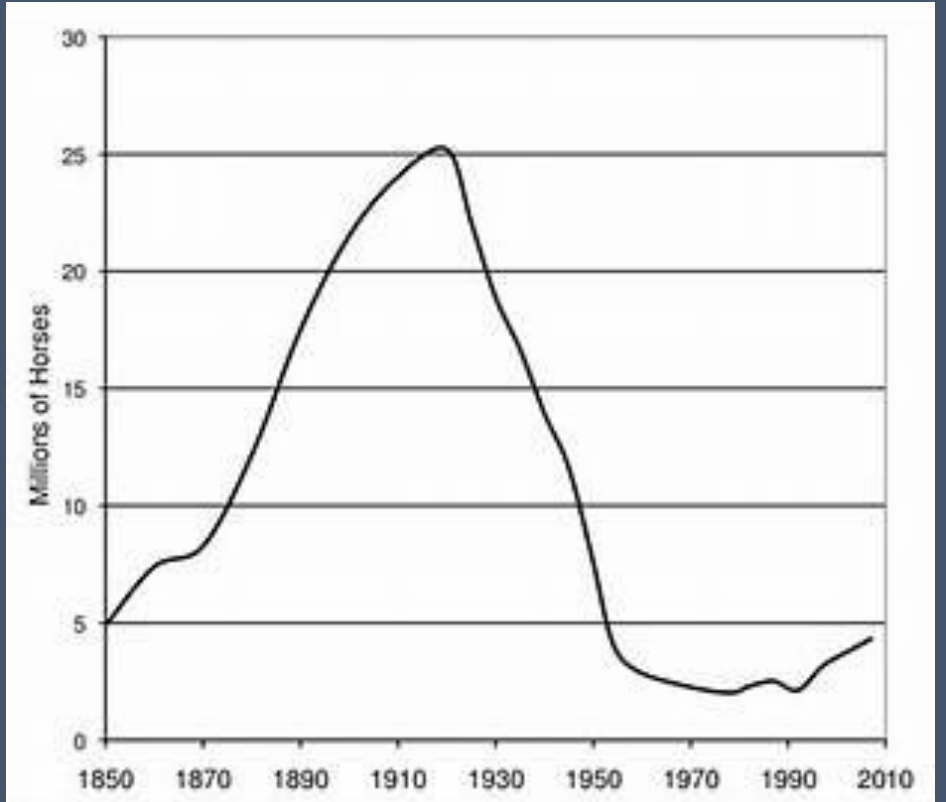
So, Two Horses Were Chatting...

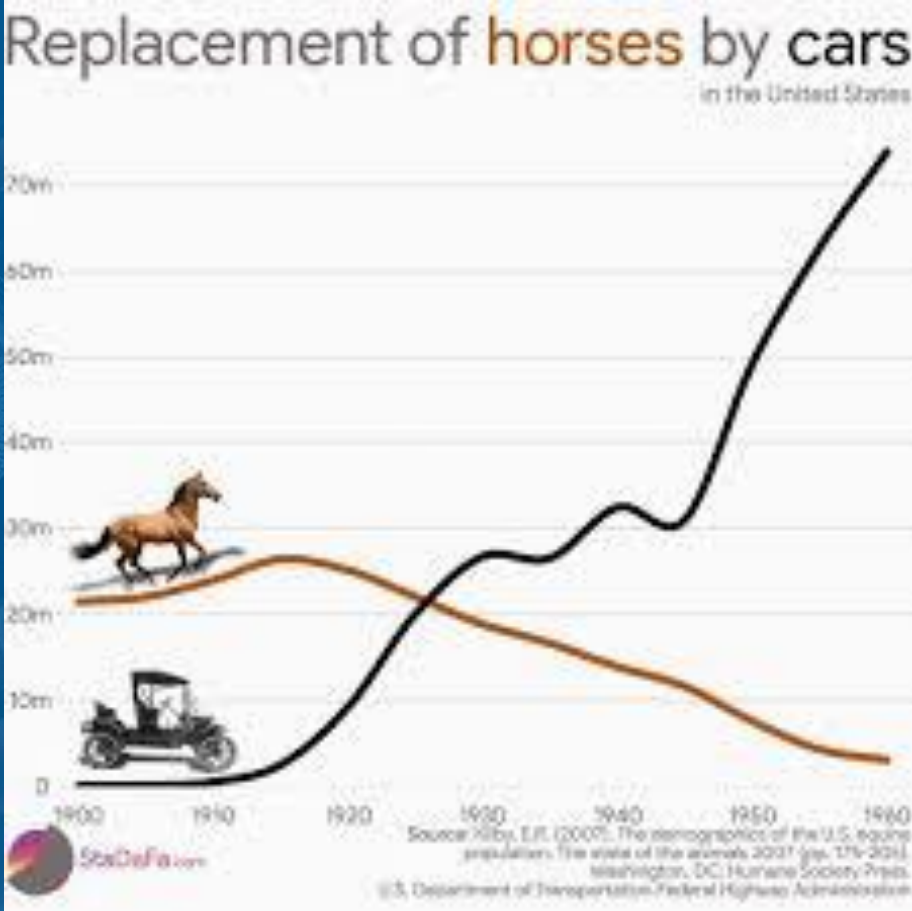
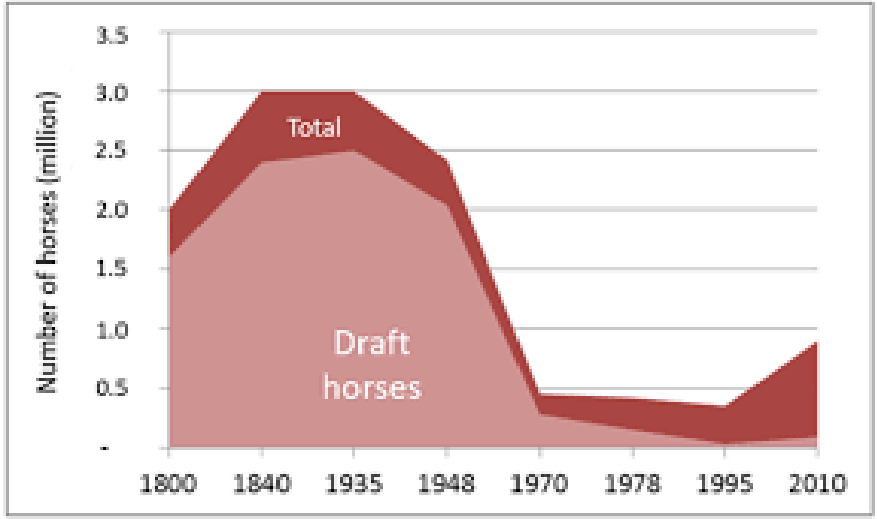


26m

5m

US Horse Population



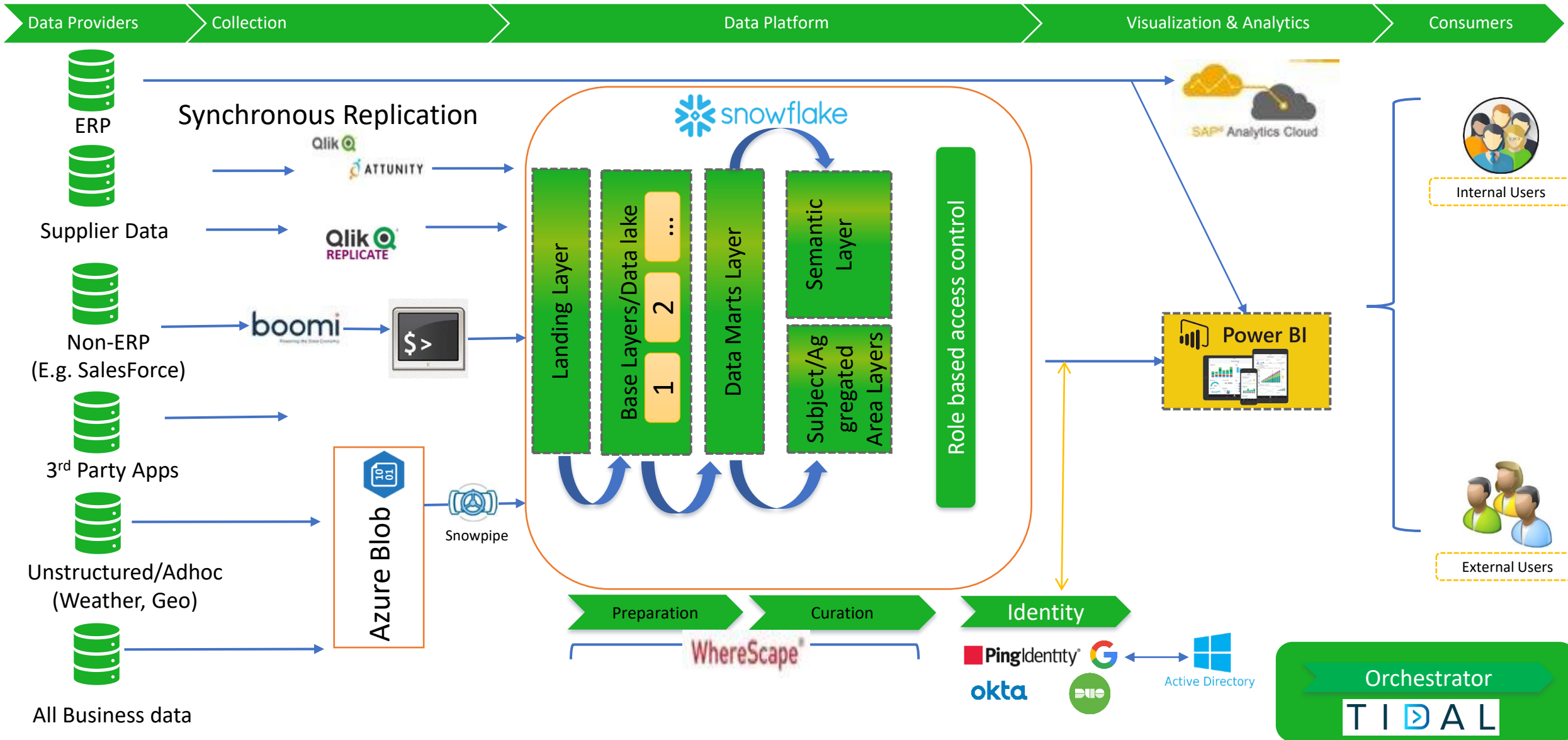


Horse Population & Models

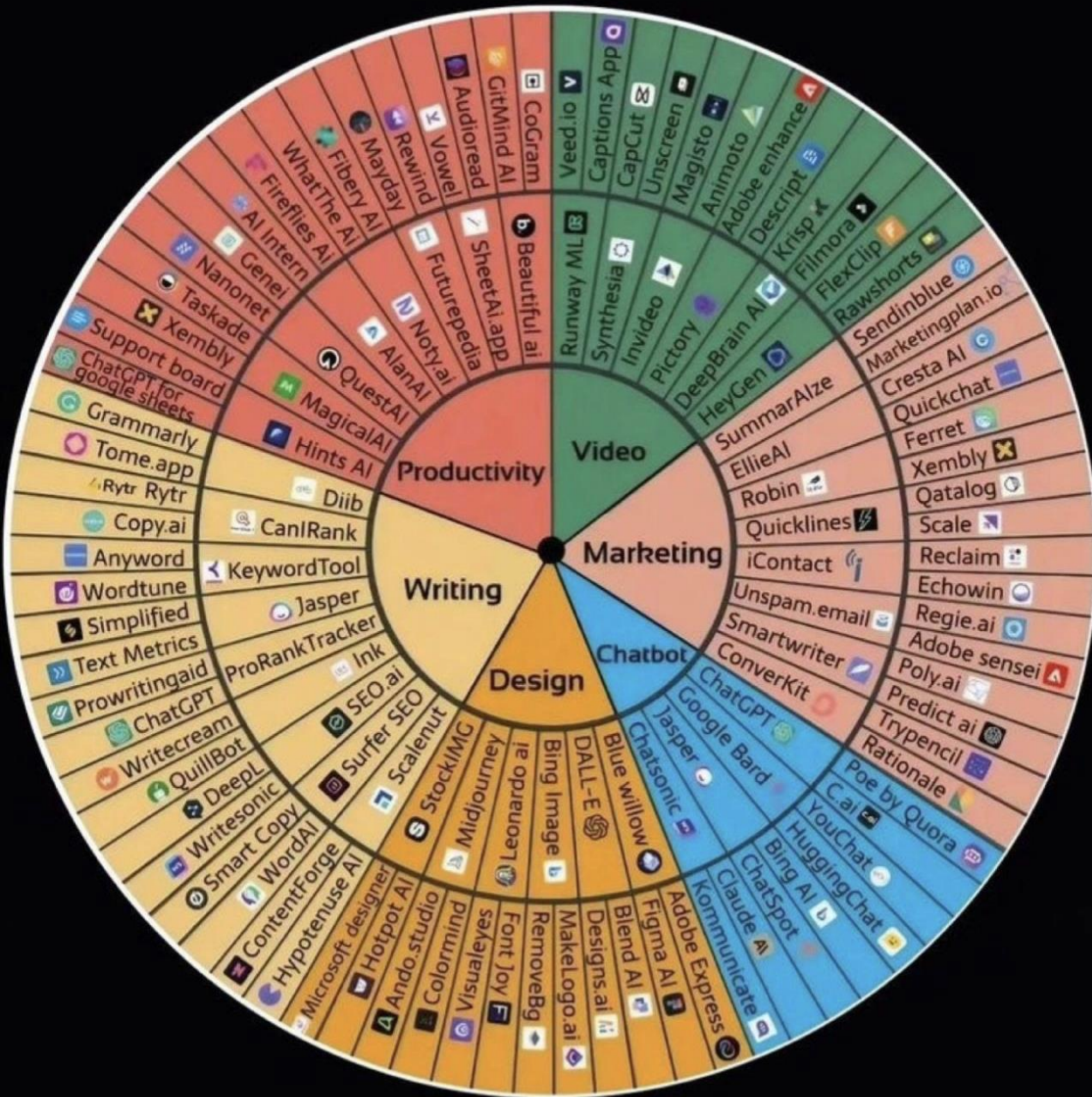
Large Language Model in Architectural Concept



Example Typical Architecture, Vendor & Dataflow Diagram



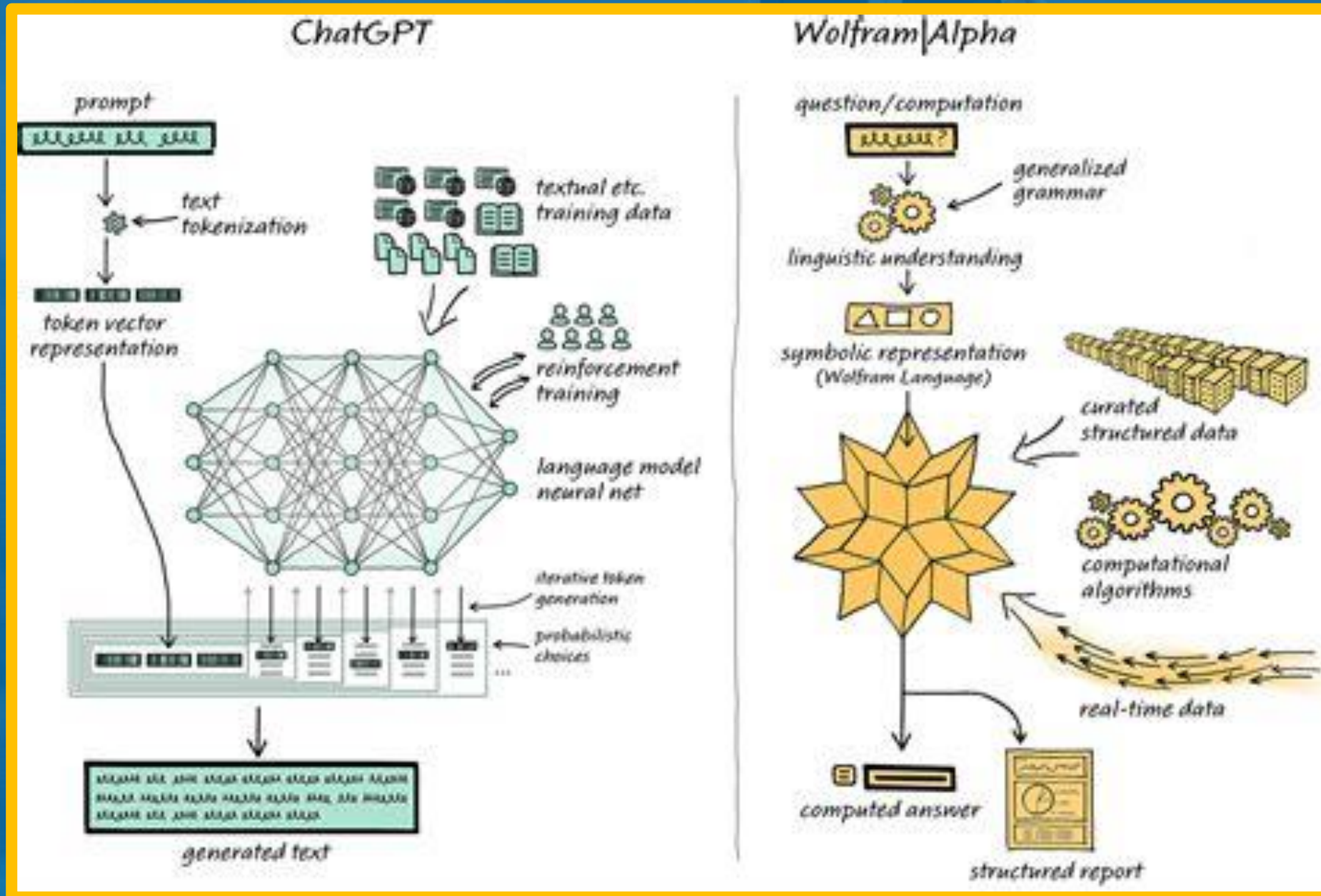
120 MIND-BLOWING AI TOOLS



As we know AI
Tools are
Exploding in
Availability

In earnest, CyberSecurity of / from
these tools has just begun

How to think about CyberSecurity of Large Language Models (LLMs)



Threats To / From Large Language Models (LLMs)

OWASP Top 10 for LLM

LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

Training data poisoning refers to manipulating the data or fine-tuning process to introduce vulnerabilities, backdoors or biases that could compromise the model's security, effectiveness or ethical behavior.

LLM04

Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre-trained models, and plugins add vulnerabilities.

LLM06

Sensitive Information Disclosure

LLM's may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. Implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control due to lack of application control. Attackers can exploit these vulnerabilities, resulting in severe consequences like remote code execution.

LLM08

Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10

Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

How to think about CyberSecurity of Large Language Models (LLMs)

Threats TO the Feed

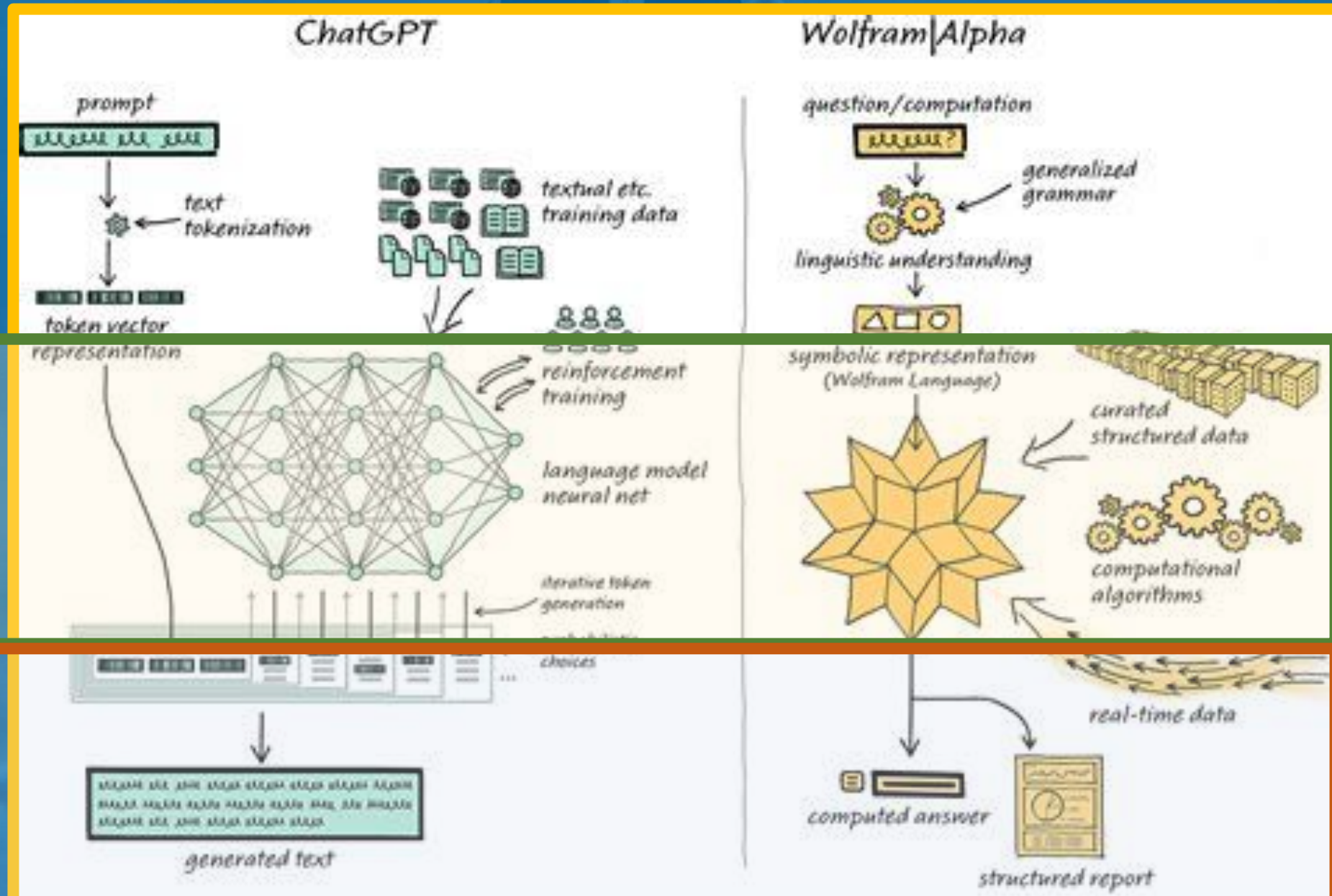
- Integrity:
- MiM & Injections
- Compliance

Threats to the Model

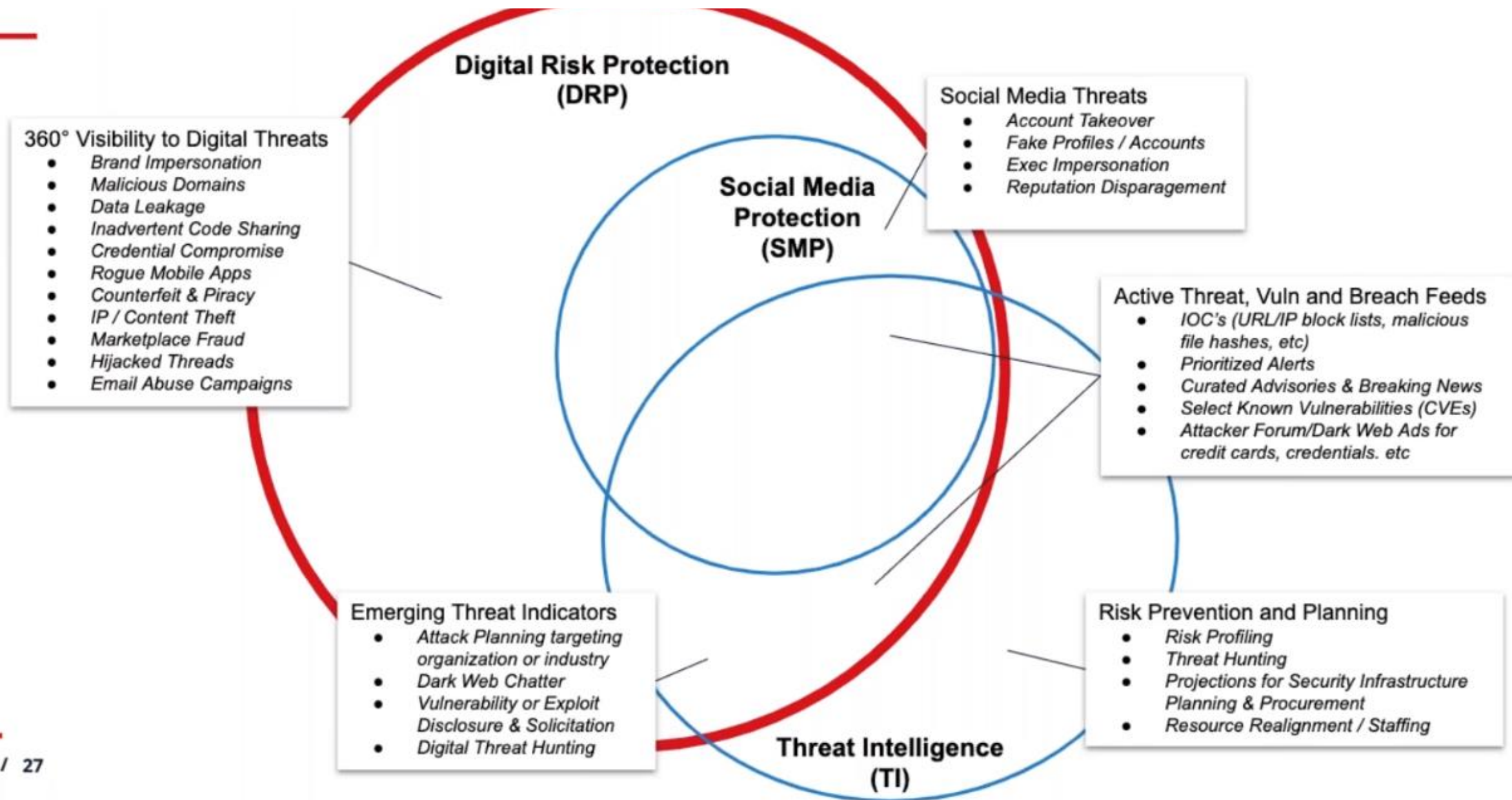
- Availability
- API & Bots / DDoS
- Ransomware / Malware

Threats to the Results

- False / Fake Information
- Privacy / Law Breach
- Used as a Threat Platform

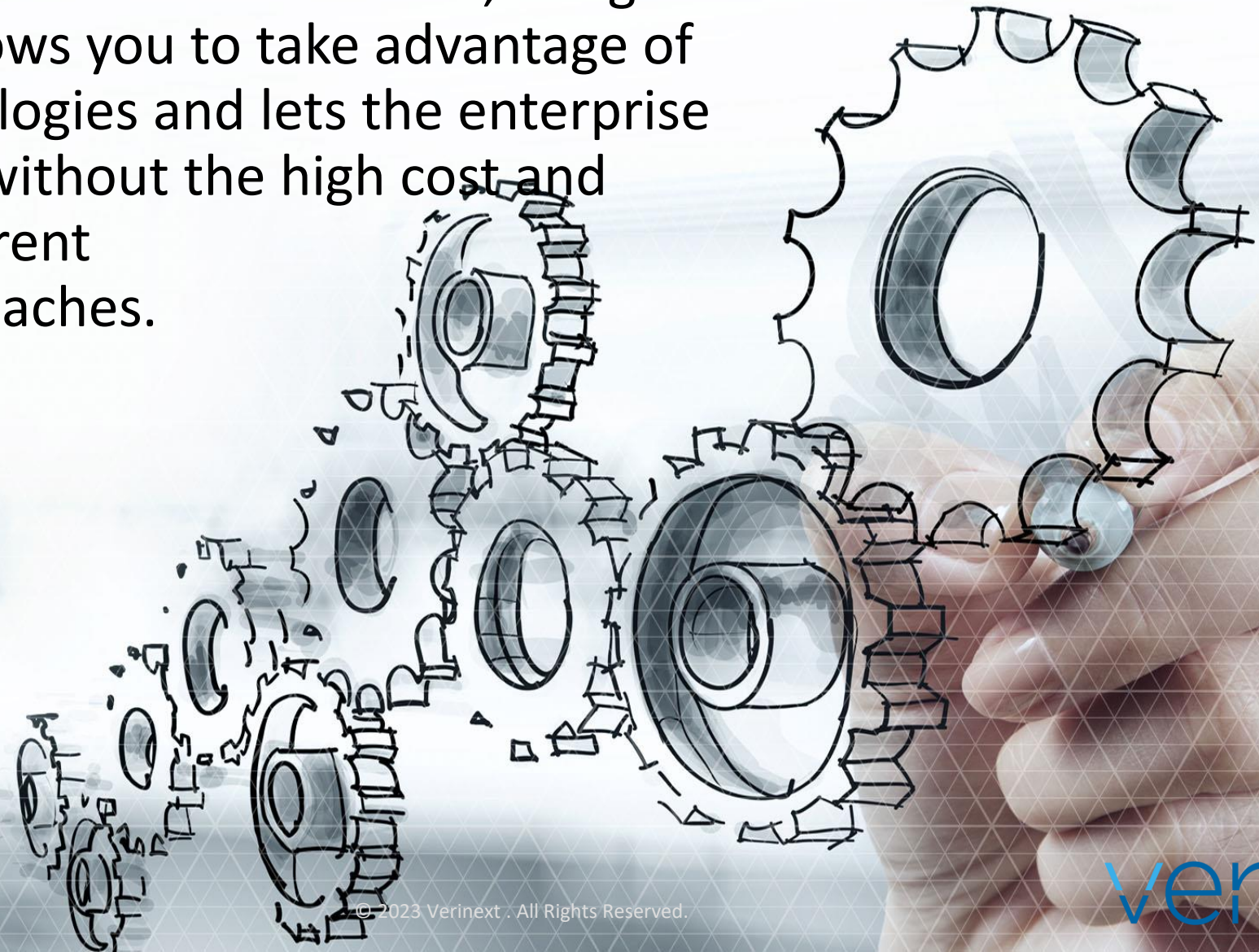


AI Security Requires External Concepts

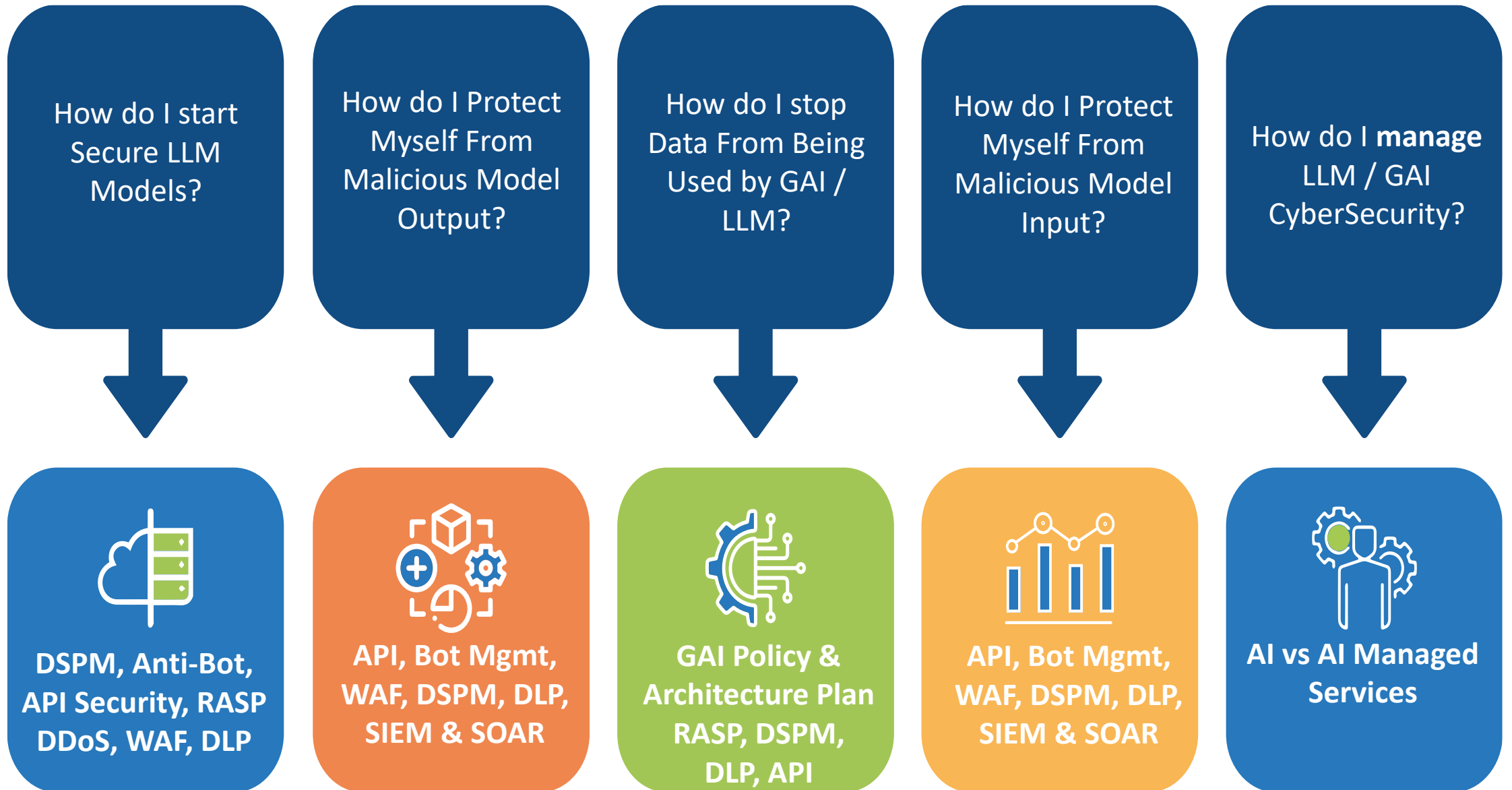


Innovation doesn't require 'rip and replace'

We have developed a business-driven, integrative approach that allows you to take advantage of innovative technologies and lets the enterprise embrace change without the high cost and business risk inherent with lock-in approaches.



How to Start thinking of LLM & GAI Cybersecurity



High Level Components of a AI Security Program

Generative AI (GAI) / LLM Policy



All Good Things
CyberSecurity start
with Policy desires

Defined Data Architecture



Junk In / Junk Out

You Must Know &
Discover Your Assets

GAI / LLM Security Testing



Ensuring your
defenses are
working as
intended

Prepare for AI: Curation Tools & Enable Culture of Security



Defensive mechanisms
are layered in order to
protect valuable data
and information.

Step 2: Data Architecture Strategy

Define Guiding Principles

Replication Strategy/plan

Data Profiling

Define Application (e.g. Snowflake) Architecture

Technology Stack

Data Governance and Management

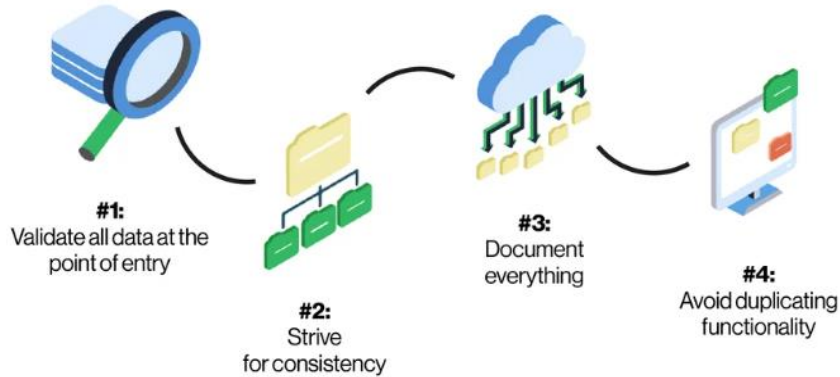
Data Security

Data Archive Strategy

Data Mart Approach

Data Cleanup and Alternative Data Warehouse Decommissioning

Step 4: Define LLM / GAI Guiding Principles, Tools, & Curation



1. *Centralized data management: a centralized view of the business across functions*
2. *Restrict data movement and bring over only necessary attributes to data warehouse*
3. *Ensure Data security and access control*

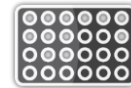
4. *Provide role based access to the data*
5. *Data Curation: manage the lifecycle of the data from collection, cleansing, transforming, cataloging to archiving*
6. *Establish a common vocabulary*
7. *Eliminate the copy of the data and increase the data sharing*
8. *Archive historical data*
9. *Build easy to use data structure for visual layer*

Stay Focused. Be Prepared.

Build your protection strategy. Develop an incident response plan.
Remember this is mostly about integrity



Look to Automation. It has become necessary to fight automated threats with automation technology.



Cover the Blind Spot. Choose a solution with the widest coverage to protect from multi-vector attacks.



Simplify with Services. Fully managed services will provide the resources and expertise needed to combat today's attacks



Single point of contact is crucial when under attack - it will help to divert internet traffic and deploy mitigation solutions.

Questions?

verinext



verinext

TRUST IN WHAT'S NEXT

Security = Confidentiality, Integrity & Availability:

Zero trust focuses on Confidentiality & Integrity

Zero Trust is required at the Perimeter, In Transit, At Rest, SaaS & in Trusted Supply-Chain Providers

Legacy systems may not adapt to zero trust

All-in-one zero-trust products don't exist

Best of breed approach to zero-trust creates gaps

Zero trust requires ongoing administration, maintenance

Trust is consolidated and, like password managers, at risk

Solutions for multi-cloud, service-provider, hybrid and powerful scale are still lacking

ZT focuses on human interactions & IoT, but software & automation is still developing

Balancing Zero Trust Theory with Real-world Functionality