Security and AI – Empowering organizations to stay ahead of evolving threats

John Velisaris
IBM Security
jvelisaris@us.ibm.com

Overview

- The (re)rise of AI
- Securing AI
- Attackers and AI
- Defenders and AI

The speed, scope, and scale of generative AI impact is unprecedented

Massive early adoption

Up to

80%

of enterprises are working with or planning to leverage foundation models and adopt generative AI[1]

Broad-reaching and deep impact

Generative AI could raise global GDP by

7%

within 10 years[2]

Critical focus of AI activity and investment

Generative AI expected to represent

30%

of overall market by 2025[3]

Artificial Intelligence (AI)

Human intelligence exhibited by machines

Learning, reasoning, perceiving, and problem solving.

Machine Learning (ML)

Systems that learn from historical data

Discover patterns and generate corresponding outputs

Deep Learning (DL)

ML technique that mimics human brain function

Enable complex applications, like image and speech recognition.

Foundation Model

Generative AI systems

Generate sequences of related data elements (for example, like a sentence).

1950's

1980's

2010's

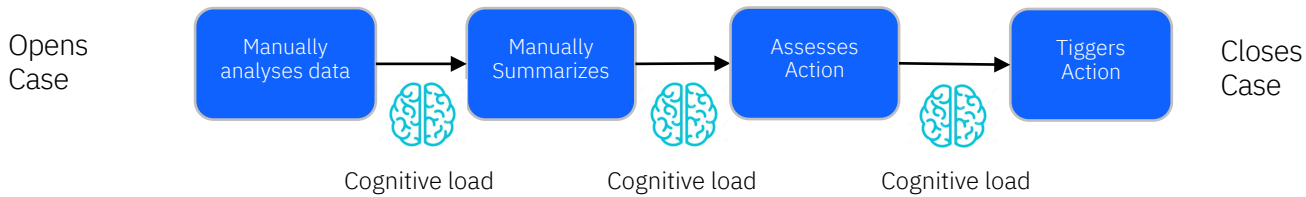
2020's

.....

Efficiency with Generative AI

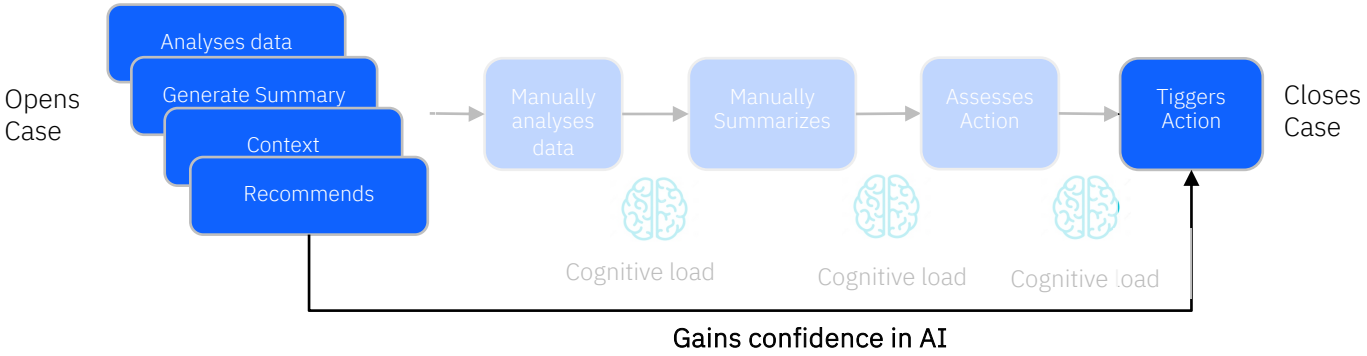
- Time saving
- Contextual insights
- Ease of navigation
- Recommended actions
- Dynamic Updates
- Collaborative effort
- Learning/adaptability
- Increased Accuracy

Human Workflow



AI Workflow

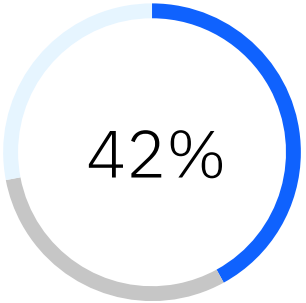
AI + Human Workflow



There’s a broad spectrum of concerns with AI

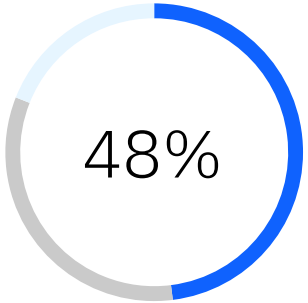
80% of surveyed business leaders have major concerns[1]

Trust



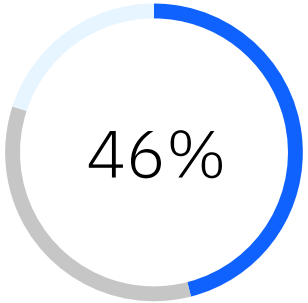
believe generative AI cannot be **trusted**

Explainability



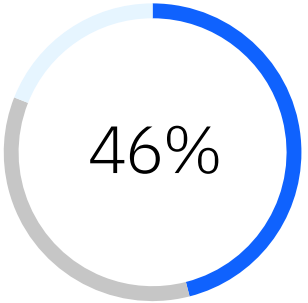
believe decisions made by generative AI are not sufficiently **explainable**

Ethics



concerned about the safety and **ethical** aspects of generative AI

Bias



believe that generative AI will propagate established **biases**

Agree Neutral Disagree

Business is adopting AI

AI for Business


Security


Talent


Marketing


Automation


Finance


Regulations

Security for AI


Trust AI models, data, vendors


Model, data, prompts access controls


Model, Data Infrastructure protection


Privacy Controls and management


Employee education


Secure design and engineering


AI Security


Threat monitoring and response

Adversarial AI


Theft


Phishing


Social engineering


Malware


Fakes


Poison

So are attackers

Attacker’s Use of AI in Security

AI Powered Attacks

Generate: DeepHack tool learned SQL injection

Automate: Generate targeted phishing attacks on Twitter

Refine: Neural network powered password crackers

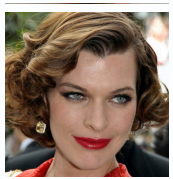
Evade: Generative adversarial networks learn novel steganographic channels

Attacking AI

Poison: Microsoft Tay chatbot poisoning via Twitter (and Watson Urban Dictionary “poisoning”)

Evade: Real-world attacks on computer vision for facial recognition biometrics and autonomous vehicles

Harden: Genetic algorithms and reinforcement learning (OpenAI Gym) to evade malware detectors



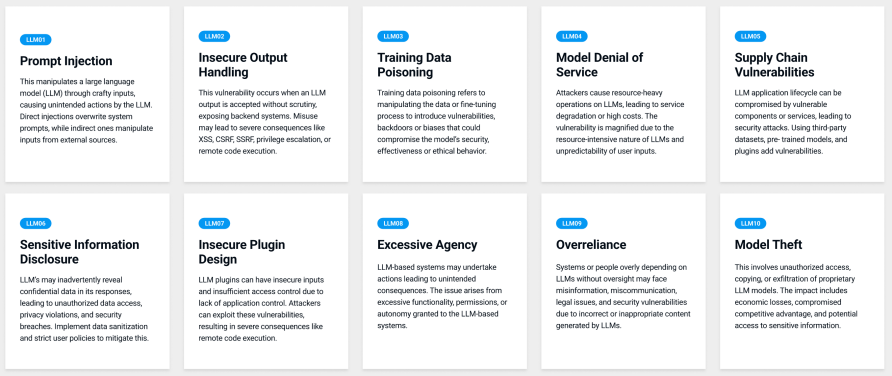
Theft of AI

Theft: Stealing machine learning models via public APIs

Transferability: Practical black-box attacks learn surrogate models for transfer attacks

Privacy: Model inversion attacks steal training data

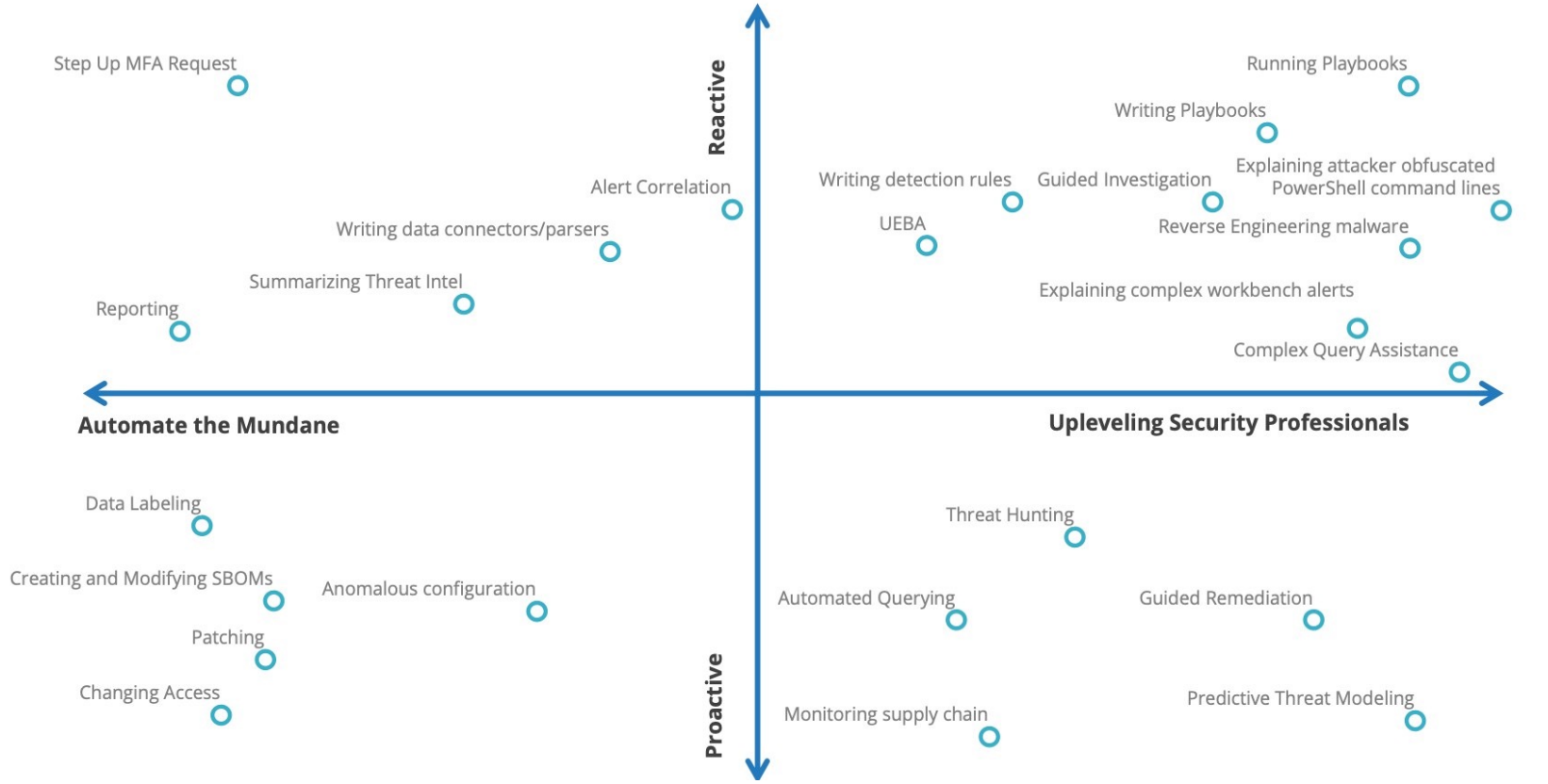
Securing AI



1. Leverage trusted AI by evaluating vendor policies and practices.
2. Enable secure access to users, models and data.
3. Safeguard AI models, data, and infrastructure from adversarial attacks.
4. Implement data privacy protection in the training, testing & operations phases.
5. Conduct threat modeling and secure coding practices into the AI dev lifecycle.
6. Perform threat detection & response for AI applications and infrastructure.
7. Assess and decide AI maturity through the IBM AI framework.

OWASP Top 10 for Large Language Model Applications

AI for Security Use Cases



© 2023 IDC

AI + Human for security operations

Identify

Automatically scan your attack surface for hidden assets, vulnerable systems and exploitable misconfigurations

Protect

Take automated action like your analysts would, through ML-powered protection

Detect

Assess the risk of threats in real-time using AI models to recognize and categorize deviations

Triage

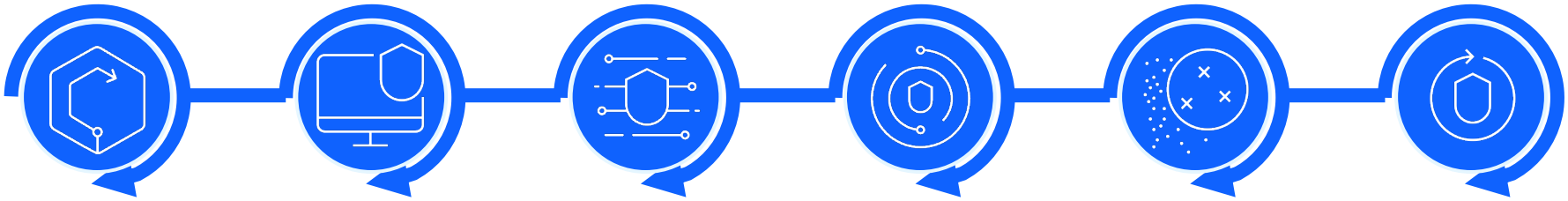
React faster to urgent incidents using alert severity scoring powered by ML

Investigate

Automatically investigate cases that warrant it, with data mining, risk assessment, and timeline generation

Respond

Dynamically create playbooks in incident response that adapt to threat context



Exposure Management

EDR

SIEM

SOC Workflow

Investigation and Threat Hunting

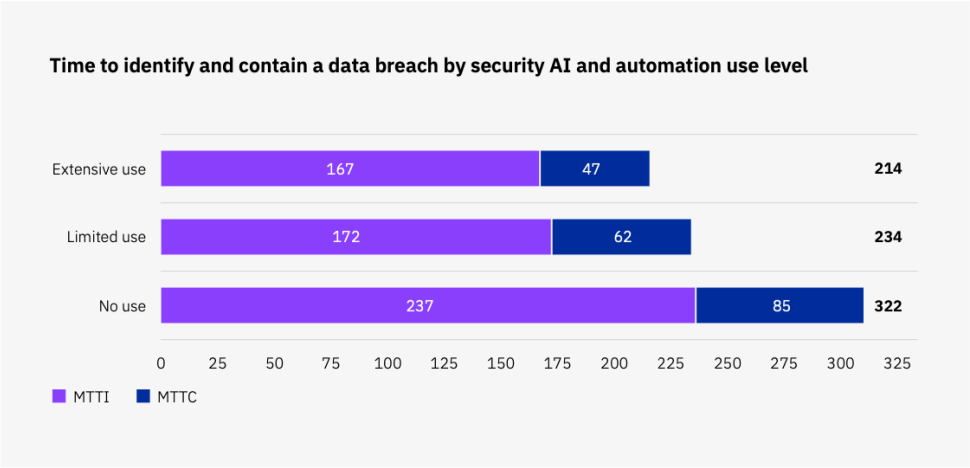
Response

Traditional use of machine learning

Increasing use of Generative AI

AI for Security

Cost of a Data Beach Report 2023



108 days

Organizations with extensive use of security AI and automation identified and contained a data breach 108 days faster than organizations with no use.

In Summary

AI in Security brings speed and efficiency so we can…

Proactively Protect, Accurately Detect and Respond Faster

… with lower costs & complexity

But this must be….

…Built on a strong foundation of security and trust..

Thank you

Follow us on:

ibm.com/security

securityintelligence.com

ibm.com/security/community

xforce.ibmcloud.com

@ibmsecurity

youtube.com/ibmsecurity