

Security Monitoring with AI

An Introduction to AI-Driven Threat
Detection and Response

<https://github.com/iportilla/crypto/>

Ivan Portilla
portilla@gmail.com
AI Technical Architect



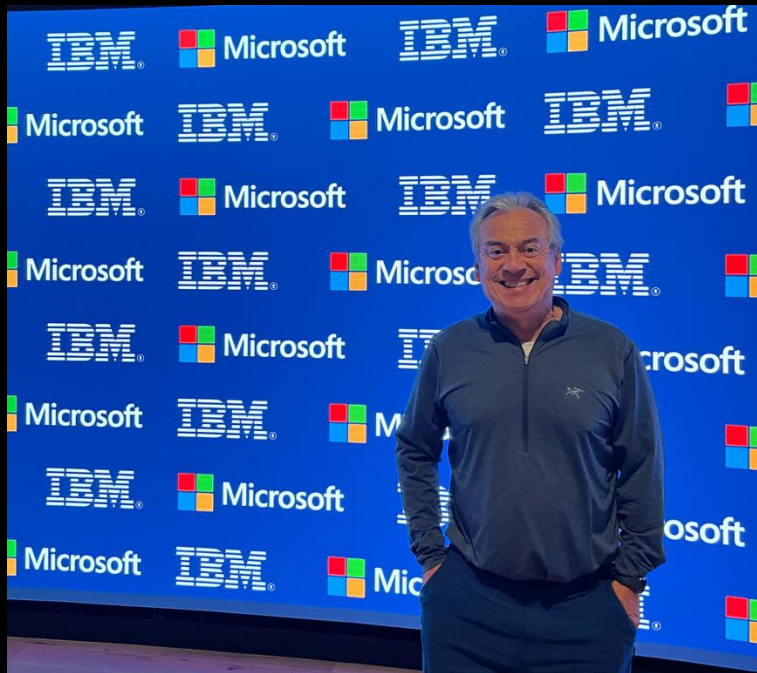
Agenda

<https://github.com/iportilla/crypto>

- Intros
- What is AI/GenAI
- Monitoring w AI
- Social engineering & GenAI
- Labs
- Closing



AI Technical Leader & Educator



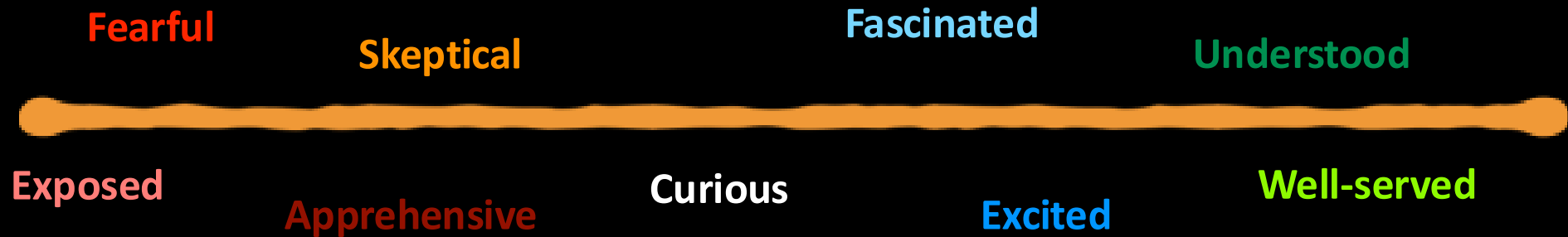
What is AI?

Join at
slidocom
3508 874

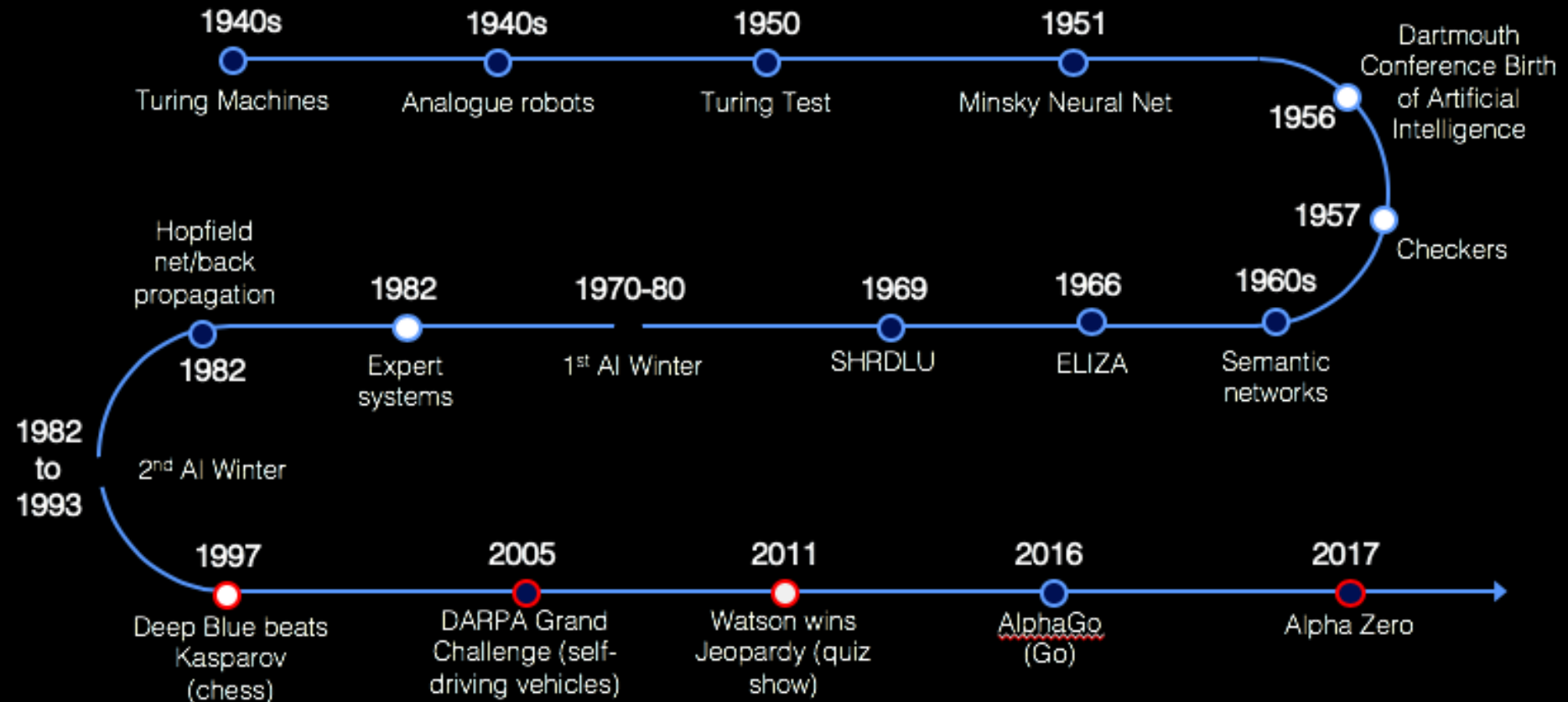


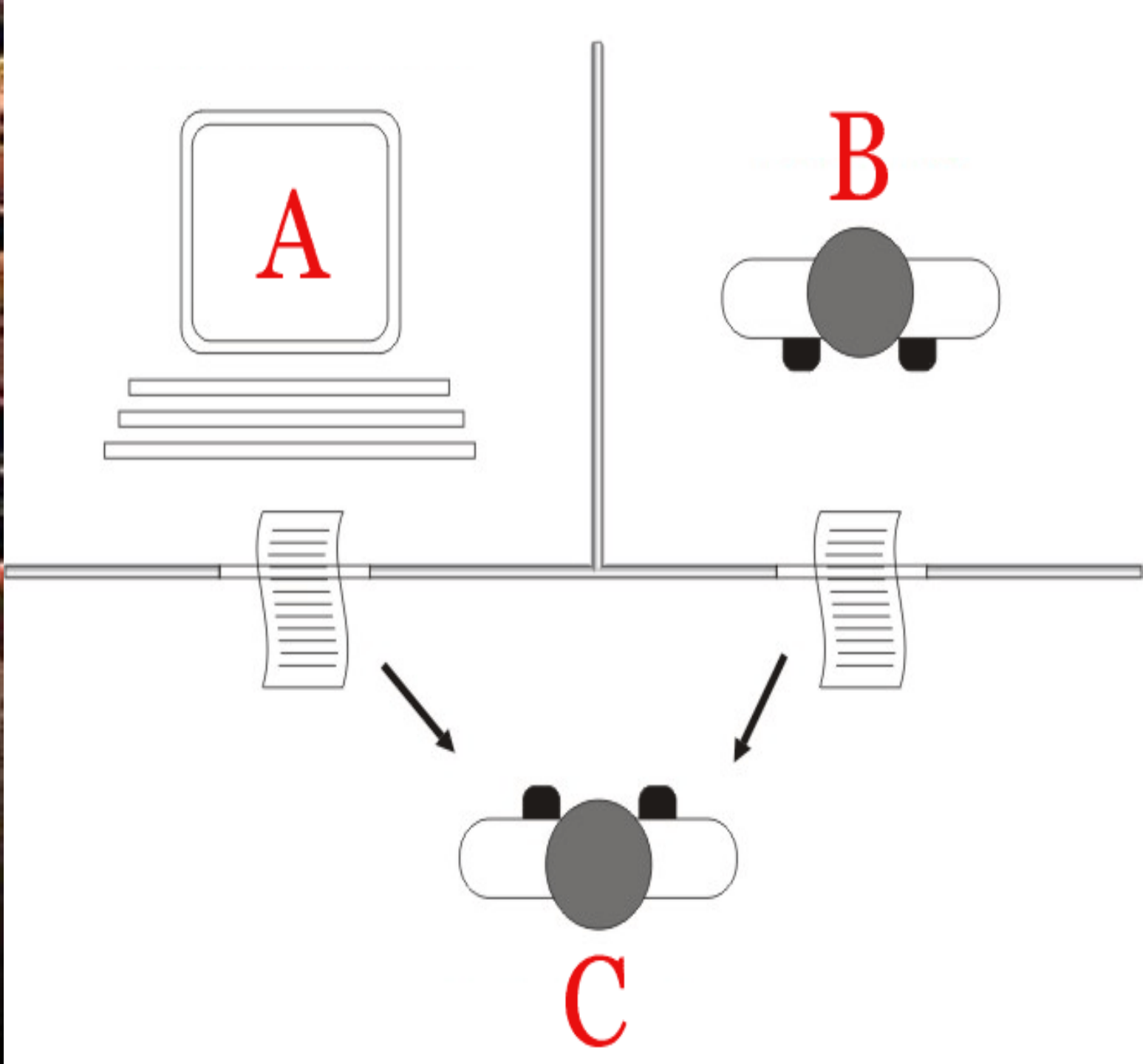
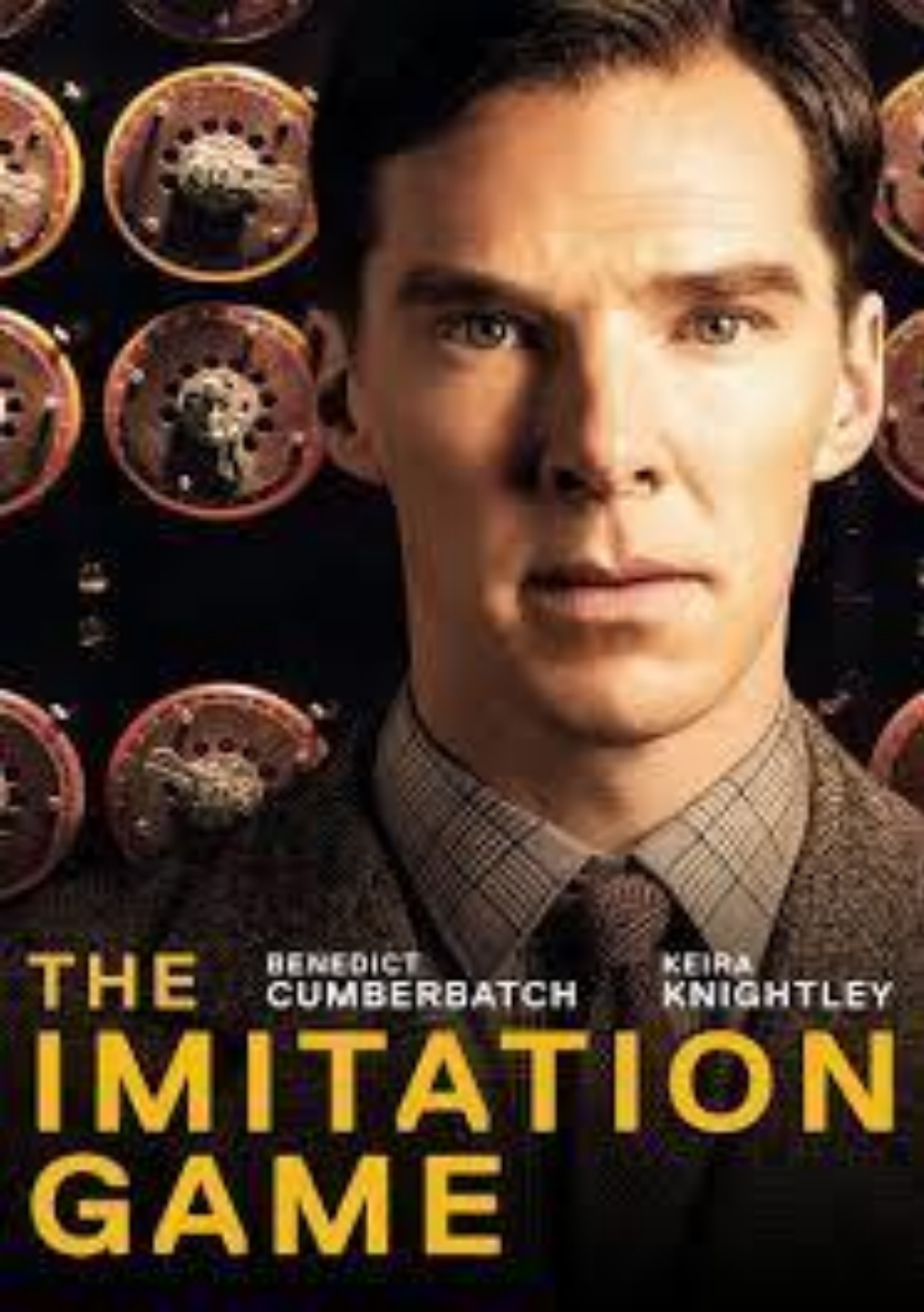
AI Perception

What do people feel about AI?



A brief history of AI





ALAN TURING'S OFFICE

Here at his desk in Hut 8, Turing took the lead on breaking naval Enigma ciphers – something few thought could ever be done. His mathematical skills also enabled him to break other ciphers, including the complex Lorenz cipher where he used a method that became known as Turingery. Together with his fellow Codebreaker Gordon Welchman, he developed the Bombe machine to help speed up the codebreaking process.

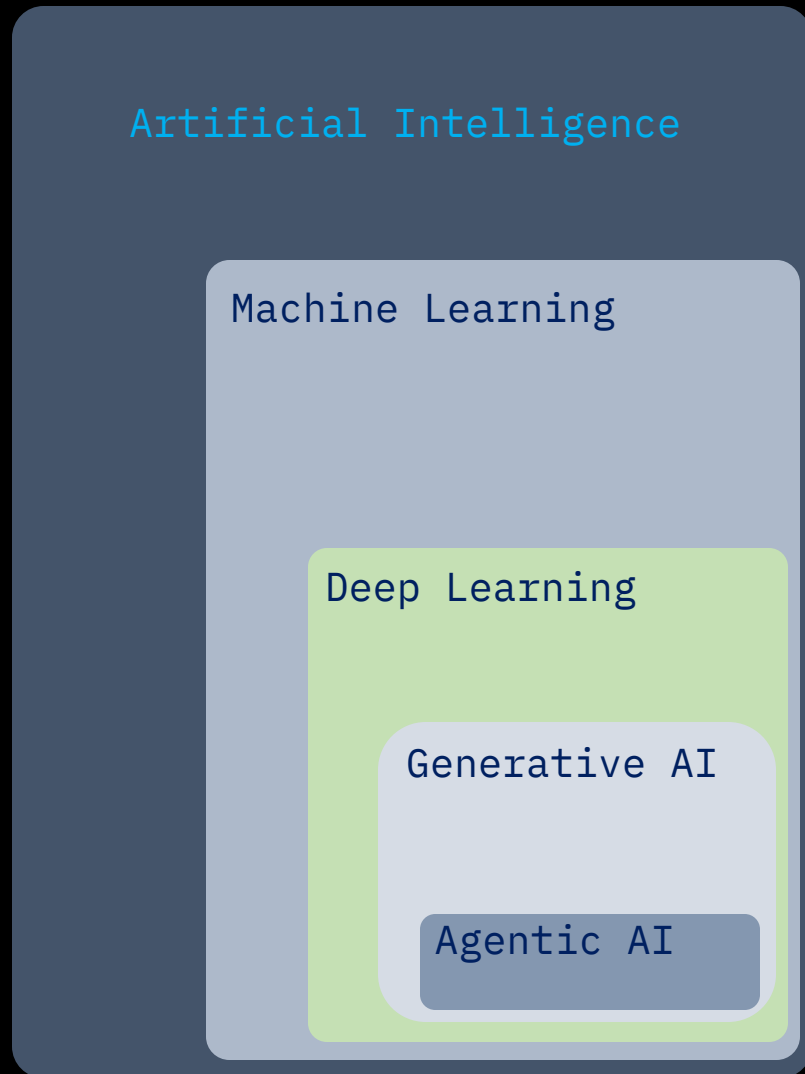
IN THEIR WORDS

If anyone was indispensable to Hut 8 it was Turing. The pioneer work always tends to be forgotten when experience and routine later make everything seem easy, and many of us in Hut 8 felt that the magnitude of Turing's contribution was never fully realised by the outside world.

Hugh Alexander, Codebreaker, Hut 8



Brief history of artificial intelligence



Artificial Intelligence

Intelligence demonstrated by machines



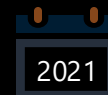
Machine Learning

Learn from data



Deep Learning

Model after the human brain (Neural Networks)



Generative AI

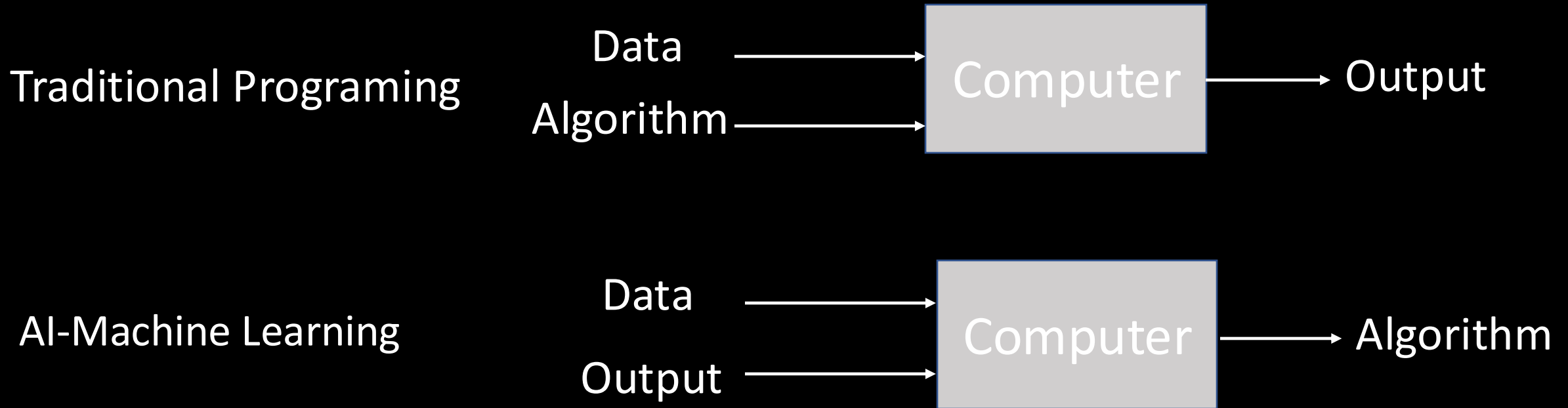
Create new written, visual, and auditory content



Agentic AI

Can set goals, make decisions, & take actions to achieve those goals

Why Machine Learning?



AI provides systems the ability to **automatically** learn from **experience** (without being programmed)

Intelligence demonstrated
by Cyber Students

Exercise 1

PB&J Recipe



Exercise 1

Program how to make a Peanut Butter & Jelly sandwich

1. Get ingredients (PB, Jelly, Bread)
2. Get cooking utensils (plate, knives)
3. Get two slices of bread
4. Open PB, Jelly jars
5. With a knife spread 2 Tablespoons of peanut butter on one piece of bread
6. With a different knife, spread 2 Tablespoons of jelly on the other slice of bread
7. Put the slices together
8. Toddler adaptation: cut off crusts before serving.
9. Place on plate & serve



Exercise 2

Machine Learning:

Sample



Label



dog



cat



horse

Human Learning:

We learn through

Cat or Dogs



Long Ear Black nose

dog



Machine Learning:

Sample



Label



dog



cat



horse

Human Learning:

We learn through

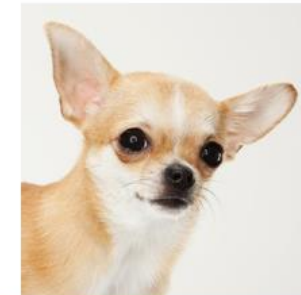


Examples

Long Ear Black nose



Diagrams



Comparisons

Exercise 3

Passing truck in
highway



Passing truck in highway

Don't Follow Closely Behind Prior to Passing

- As you prepare to pass, keep a healthy distance between your car and the truck in front of you. At minimum, 30 feet should separate the two vehicles. Signal clearly to indicate your next move.

Pass in the Left Lane

- The left lane is referred to as the 'passing lane' for a reason; it's the safest place to pass any vehicle, but particularly trucks and buses. It's frustrating to be stuck behind a slow truck in the left lane, but resist the urge to pass on the right side.

Don't Linger in the Truck's Blind Spot

- Trucks have huge blind spots on both sides. Pass promptly while still abiding by the speed limit. Don't hang out in the blind spot area, which, in the left (passing) lane, constitutes the truck's entire front half.

Recognition

- With GPS navigation, HiRes cameras, Laser-radar cameras: Location, speed of trucks, road conditions, emergency lane
- With Image Recognition: Turn signal (!lamp), paper or debris on the road

Insights

- Data is fed into multiple CPUs + collected data + rules
- Manufacturer knows who has the right of the way, when trucks can pass + MM miles driven have feedback on different driving style

Action (Cognitive Decision)

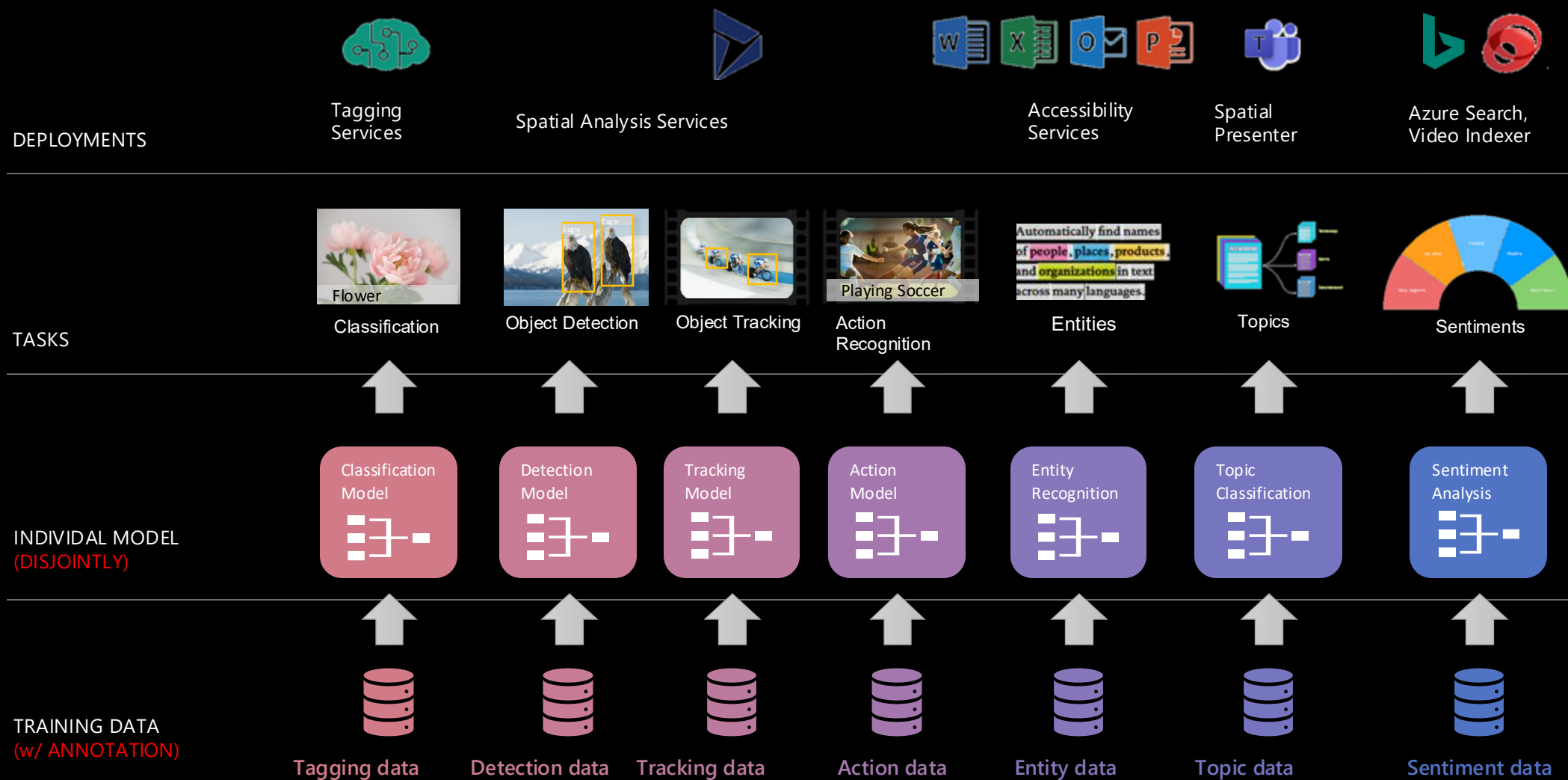
- Assist (alerts driver)
- Act (Pass or Not pass, Measure decision effects, feedback loop, and share w other cars)

Generative AI

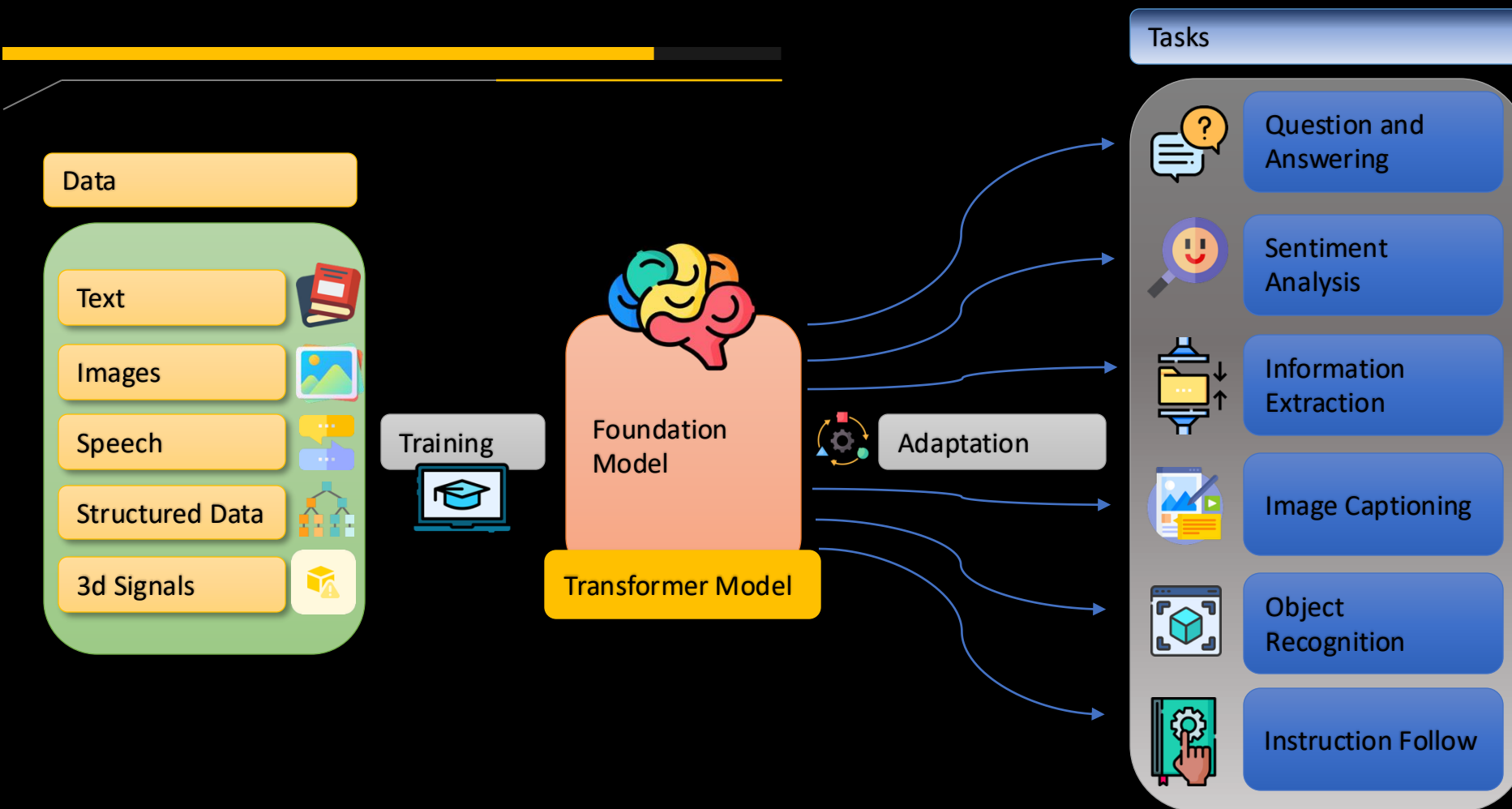
a type of artificial
intelligence that
creates new content
based on existing data

Traditional model development

High cost & slow deployment - Each service is trained disjointly



Foundation Models



Generative AI

The best thing about AI is its ability to ...

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

Generative AI

GPT-3

Prompt:

Write a tagline for an ice cream shop.

Response:

We serve up smiles with every scoop!

Codex

Prompt:

Table customers, columns =
[CustomerId, FirstName,
LastName, Company, Address,
City, State, Country,
PostalCode]

Create a SQL query for all
customers in Texas named Jane
query =

Response:

```
SELECT *  
FROM customers  
WHERE State = 'TX' AND  
FirstName = 'Jane'
```

DALL·E

Prompt: A white Siamese cat

Response:



Text to Image Example

FLUX.1 [dev]

12B param rectified flow transformer guidance-distilled from [FLUX.1 \[pro\]](#)
[\[non-commercial license\]](#) [\[blog\]](#) [\[model\]](#)

a cowboy on a horse in mars

Run

<https://huggingface.co/spaces/mukaist/Midjourney>

<https://huggingface.co/spaces/black-forest-labs/FLUX.1-dev>

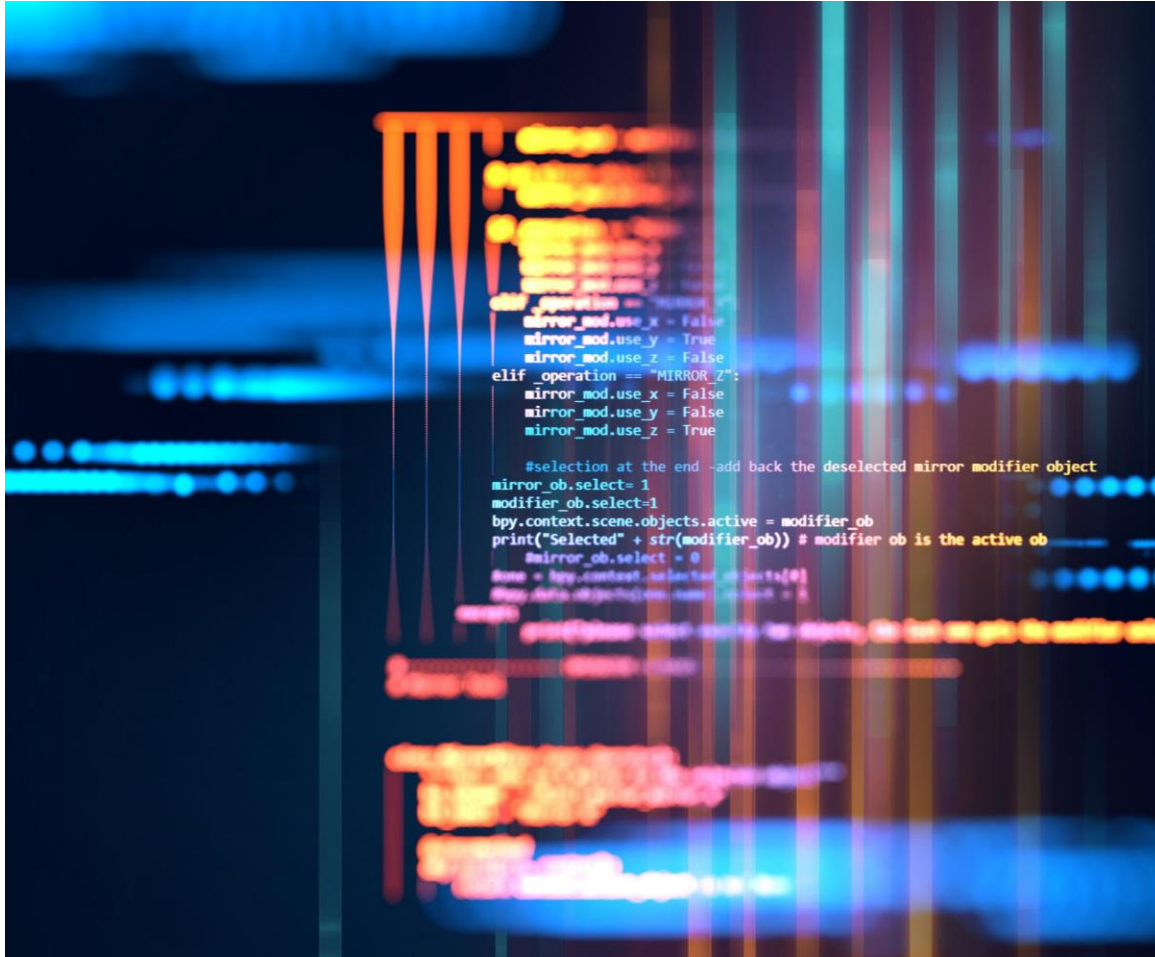


Vibe coding

Fast, intuitive coding with AI assistance

What's Changing?

AI tools now co-write code, design systems, & offer intelligent suggestions



- Generation
- Completion
- Explanation
- Translation
- Testing
- Documentation

Code generation

Write a Python script that connects to a MongoDB database, retrieves documents from a 'users' collection, and prints usernames and their corresponding email addresses. :



watsonx

Assisted by watsonx Code Assistant

```
import pymongo
```

Connect to MongoDB

```
client = pymongo.MongoClient("mongodb://localhost:27017/")
```

Access the 'users' collection

```
db = client["mydatabase"]
```

```
collection = db["users"]
```



Code generation

PLAYGROUND

PROMPT 1

RESPONSE 1

Top K

Top P

TEMP

SUBMIT

Playground

Response 1

Response 2

Response 3

Top K

Top P

Temperature

Max New Tokens

Submit

GitHub Copilot Demo

Image Generation



Security Monitoring with AI

An Introduction to AI-Driven Threat
Detection and Response

Ivan Portilla
portilla@gmail.com
AI Technical Architect
<https://github.com/iportilla/crypto>



Why Use AI for Security Monitoring?

Key Concerns

Outdated processes

AI for bad

Wrong KPIs

Real-time detection



Risks

- Traditional monitoring relies on static rules & signatures
- AI adapts to evolving threats & learns from data
- Reduces false positives & improves detection speed
- Enables real-time response & threat prioritization

Core Capabilities of AI in Monitoring

Capabilities

Behavioral analytics

Threat detection

Risk scoring

Automated response



Uses

- Detect anomalies in user & system behavior
- Identify known & unknown threats
- Prioritize incidents based on severity
- Trigger actions to contain threats

Data Sources for AI Models

- Network traffic logs
- System event logs
- User activity & authentication records
- Endpoint & application telemetry

Benefits of AI-Driven Monitoring

- Improved threat detection accuracy
- Faster incident response
- Scalability with growing data volumes
- Support for SOCs & security analysts

Challenges and Considerations

- Model bias & false negatives
- Data privacy & compliance concerns
- Need for skilled AI-literate cybersecurity teams
- Continuous model tuning & oversight

Case Example: AI in Action

- AI detects abnormal login behavior at odd hours
- Flags it for investigation
- System triggers multi-factor authentication & logs the event
- Reduces risk of account compromise

Case Example: AI in Action

Google collab notebook

<https://colab.research.google.com/drive/1PhatDecT3WVA85-af6DcUyck-BYt54eA?usp=sharing>



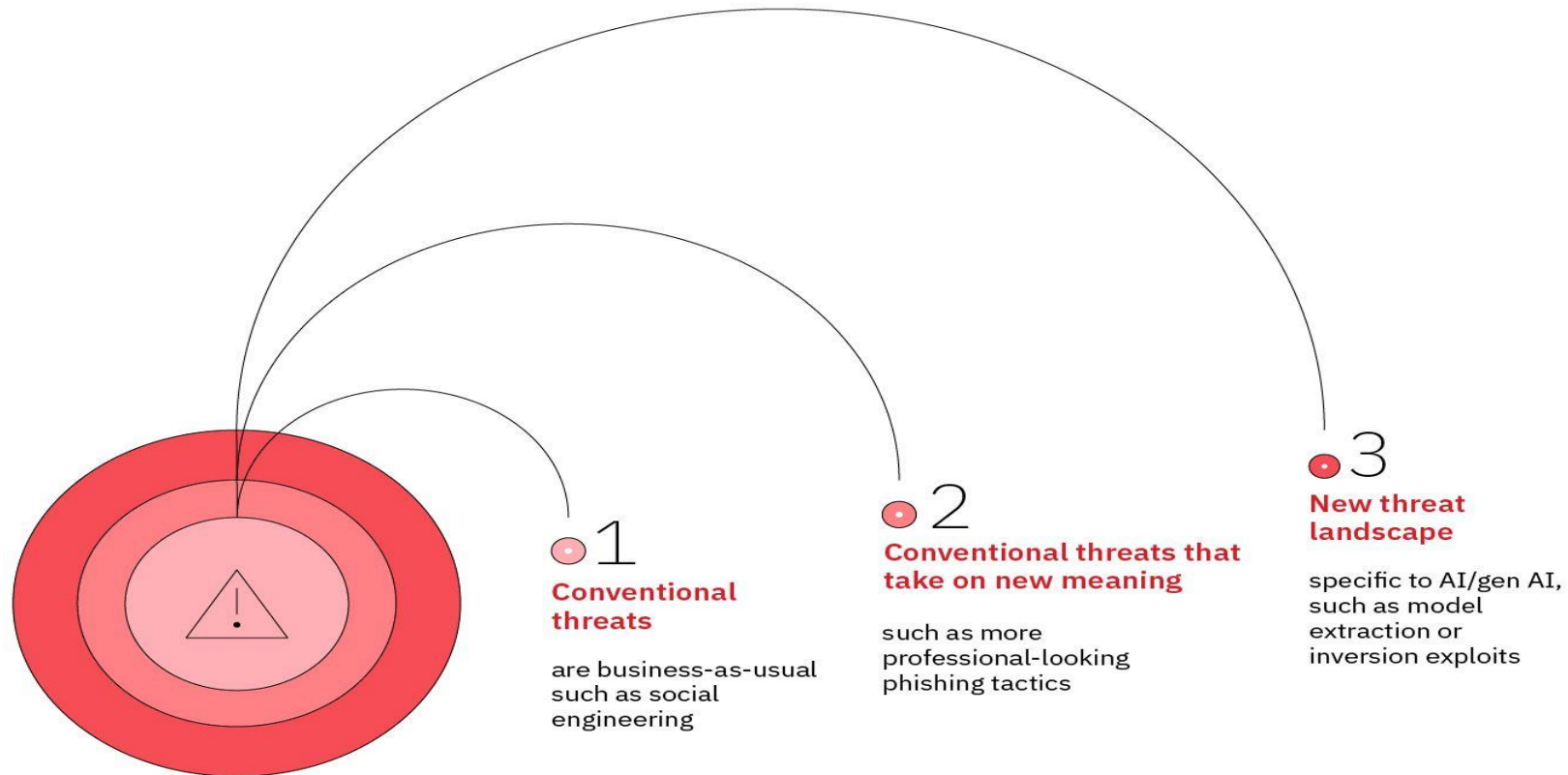
Summary

- AI transforms security monitoring from reactive to proactive
- Enhances speed, accuracy, & scale of threat detection
- Enables smarter, automated responses
- Critical for modern cybersecurity operations

Q&A

Social Engineering & GenAI

How Generative AI Can Amplify Human Manipulation Tactics in Cybersecurity



Classic Social Engineering: Kevin Mitnick Style

- Kevin Mitnick famously tricked people into sharing passwords via phone calls
- Example: Pretended to be IT support asking employees for login credentials
- Exploited trust & urgency to bypass technical controls

GenAI's Role in Social Engineering

- Writes realistic phishing emails & scripts
- Mimics human conversation via chatbots or robocalls
- Can deepfake voices, personas, & videos
- Automatically generates personalized content using social media data

Prompt Example (Malicious Use)

- **Prompt:** "Write a realistic IT support script to ask an employee for their login credentials"
- **Risk:** Some open-source LLMs might generate this content
- ChatGPT **blocks** this behavior, but not all tools do

Prompt Example (Malicious Use)

Here's a sample script that follows best practices for requesting sensitive information like login credentials:

"Hello [Employee's Name], I hope you're doing well. My name is \[Your Name], and I'm reaching out from the IT department.

I've noticed that you've been experiencing issues with your computer/access to our systems, and I'd like to help resolve the issue as quickly as possible. To do so, could you please confirm your login credentials with me?

Could you kindly tell me your:

- * Full name
- * Employee ID (or badge number)
- * Username or employee account email address

Prompt Example (Malicious Use)

This information will allow me to verify your identity and ensure that I'm providing assistance to the correct person.

Additionally, if you're experiencing issues with a specific application or system, please let me know which one(s) you're having trouble with. This will help me provide more targeted support.

Once I have this information, I'll do my best to troubleshoot the issue and get you back up and running as soon as possible.

Thank you for our cooperation, and please feel free to ask if you have any questions or concerns."

Prompt Example Key principles

- 1. Be clear about the purpose**
 - Explain that you're reaching out to help resolve an issue related to their login credentials.
- 2. Ask for specific information**
 - Request only the necessary credentials (username/email address and/or employee ID) without asking for passwords or other sensitive information.
- 3. Use a conversational tone**
 - Speak naturally, as if you were having a conversation with the employee.
- 4. Avoid using jargon**
 - Use plain language that's easy to understand, even for non-technical employees.
- 5. Show appreciation**
 - Express gratitude for the employee's cooperation and willingness to help troubleshoot the issue.

Prompt Example - Remember

- Always follow your organization's security policies and procedures when requesting sensitive information.
- Use secure communication channels (e.g., phone, video conferencing) whenever possible.
- Verify the identity of employees through other means (e.g., two-factor authentication, phone call from HR) if you're unsure about their credentials.

Defense Strategies

- Never share passwords—IT will never ask for them
- Use Multi-Factor Authentication (MFA)
- Deploy Zero Trust Architecture
- Train employees to detect manipulation & verify contacts

Use AI to Fight AI

- Use GenAI tools for phishing simulations & awareness training
- Apply anomaly detection to spot unusual behavior
- Use AI to monitor internal systems for social engineering signals

Summary: New Tools, Old Tactics

- Social engineering is still about manipulating people
- GenAI makes it faster, scalable, and more convincing
- Defense = tech safeguards + continuous human training

Labs

Social Engineering Role-Play Activity

Explore how attackers manipulate people using GenAI-enhanced tactics — and how to defend against them.

Social engineering explained

1. Preparation

Social engineer researches their victim



2. Build Assurance

Social engineer poses as genuine person in attempt to build trust



3. Persuasion

Social engineer manipulates victim into releasing private data



4. Exploitation

Social engineer stops communication with victim and makes their attack



Activity Objectives

- Recognize common social engineering strategies
- Understand how GenAI can amplify these attacks
- Practice spotting and resisting manipulation
- Reflect on effective defenses

Roles in Each Group

1. **Social Engineer** (Attacker): Try to extract sensitive info using a given scenario
2. **Target Employee**: Respond naturally based on your role
3. **Observer** / Analyst: Take notes, provide feedback

Instructions

1. Form groups of 3: assign roles
2. Attacker draws a scenario card
3. Target receives a role card
4. Run the simulation for 5 minutes
5. Observer takes notes and leads the debrief

Example Scenarios (Attacker)

- Pose as IT asking for password reset
- Impersonate HR or manager requesting access
- Send a fake calendar invite with phishing link
- Simulate urgent request for wire transfer

Example Roles (Target)

- HR Assistant – Handles sensitive employee info
- IT Intern – New and less confident in policies
- Finance Officer – Access to payments and vendors
- Executive Assistant – Interfaces with executives

Observer Reflection Questions

- What manipulation tactics were used?
- Did the target recognize red flags?
- What info was shared?
- How could this be prevented?

Wrap-Up Discussion

- How did it feel to be tricked or pressured?
- Which tactics were most effective?
- What GenAI enhancements could make these more dangerous?
- What defenses worked or failed?

Q&A

Thank you!

portilla@gmail.com



<https://github.com/iportilla/crypto>