Navigating the Ethical Implications of AI in Cybersecurity: Balancing Innovation and Integrity

Dr Angel Jimenez-Aranda

25th October 2024

University of Salford MANCHESTER

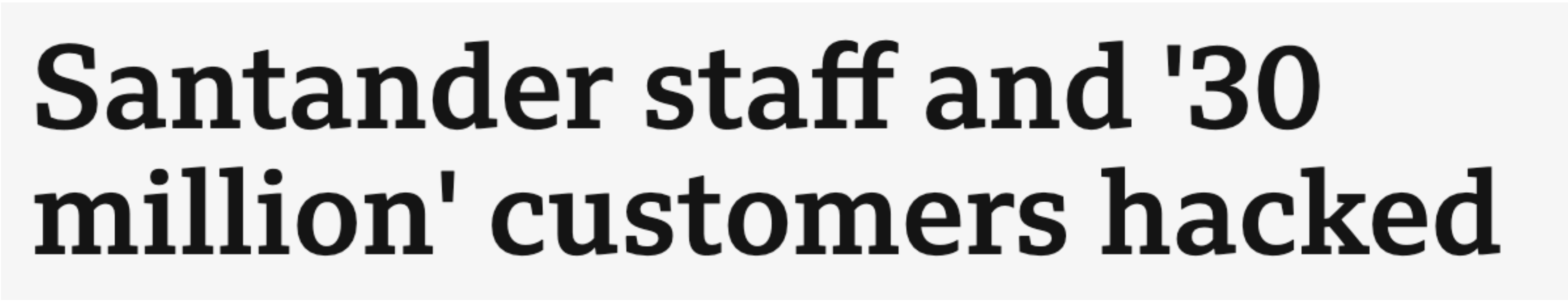
What is Cybersecurity?

*Deploying **people, policies, processes and technologies***
to protect organisations, their critical systems and sensitive information
from digital attacks.

Data Breaches in the News

2024

Ticketmaster Hack: Personal Data of 560 Million Customers Potentially Compromised


Santander staff and '30 million' customers hacked

Dell API abused to steal 49 million customer records in data breach

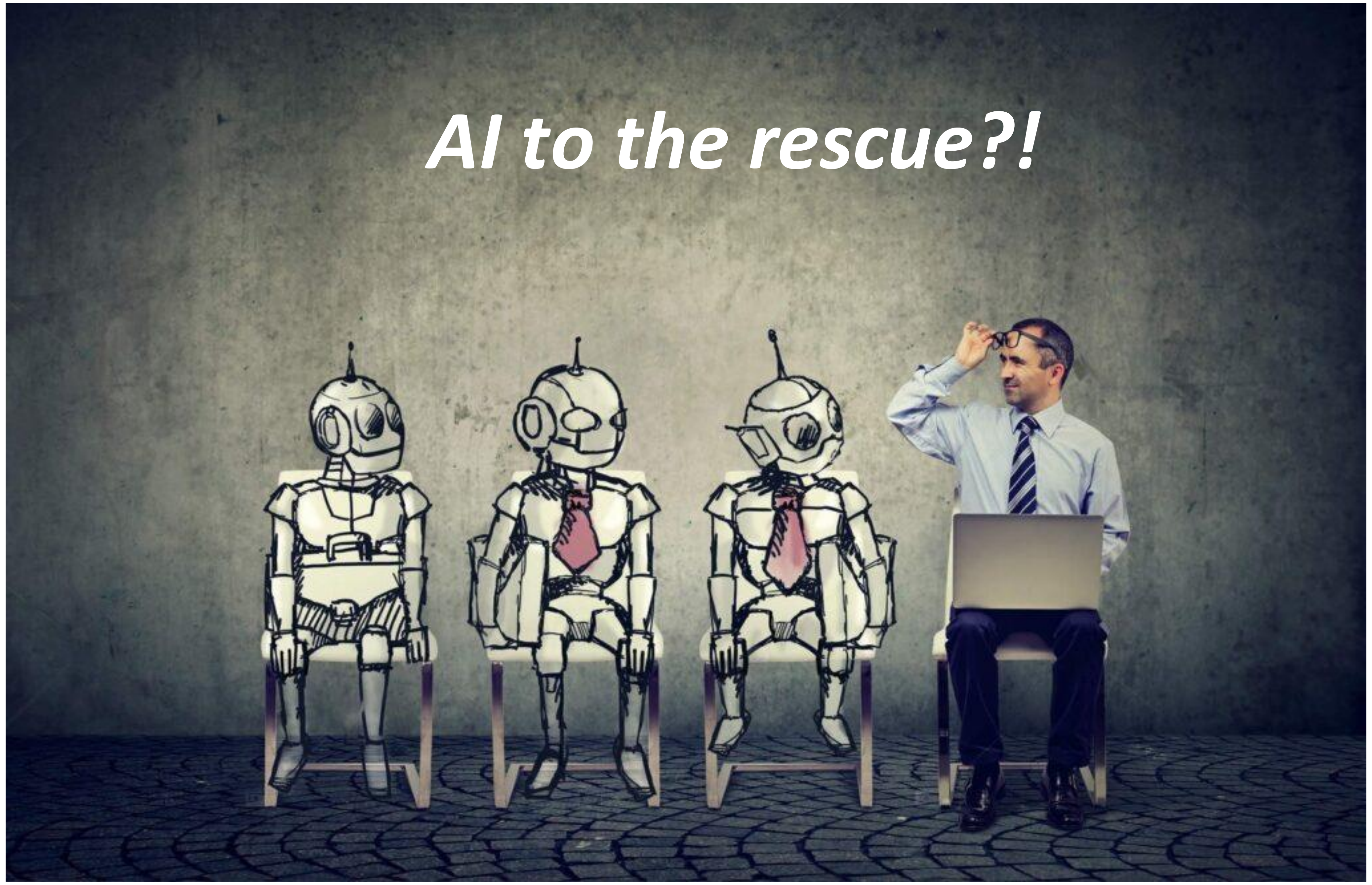
AT&T acknowledges data leak that hit 73 million current and former users

NEWS 14 MAR 2024

French Employment Agency Data Breach Could Affect 43 Million People

UK confirms Ministry of Defence payroll data exposed in data breach

The Role of AI in Cybersecurity



AI to the rescue?!

The Role of AI in Cybersecurity

AI's Impact on the Cybersecurity Landscape:

New capabilities

Speed and efficiency improvements

… for both defenders and attackers

The Role of AI in Cybersecurity

University of Salford MANCHESTER

REACTIVE MANAGEMENT

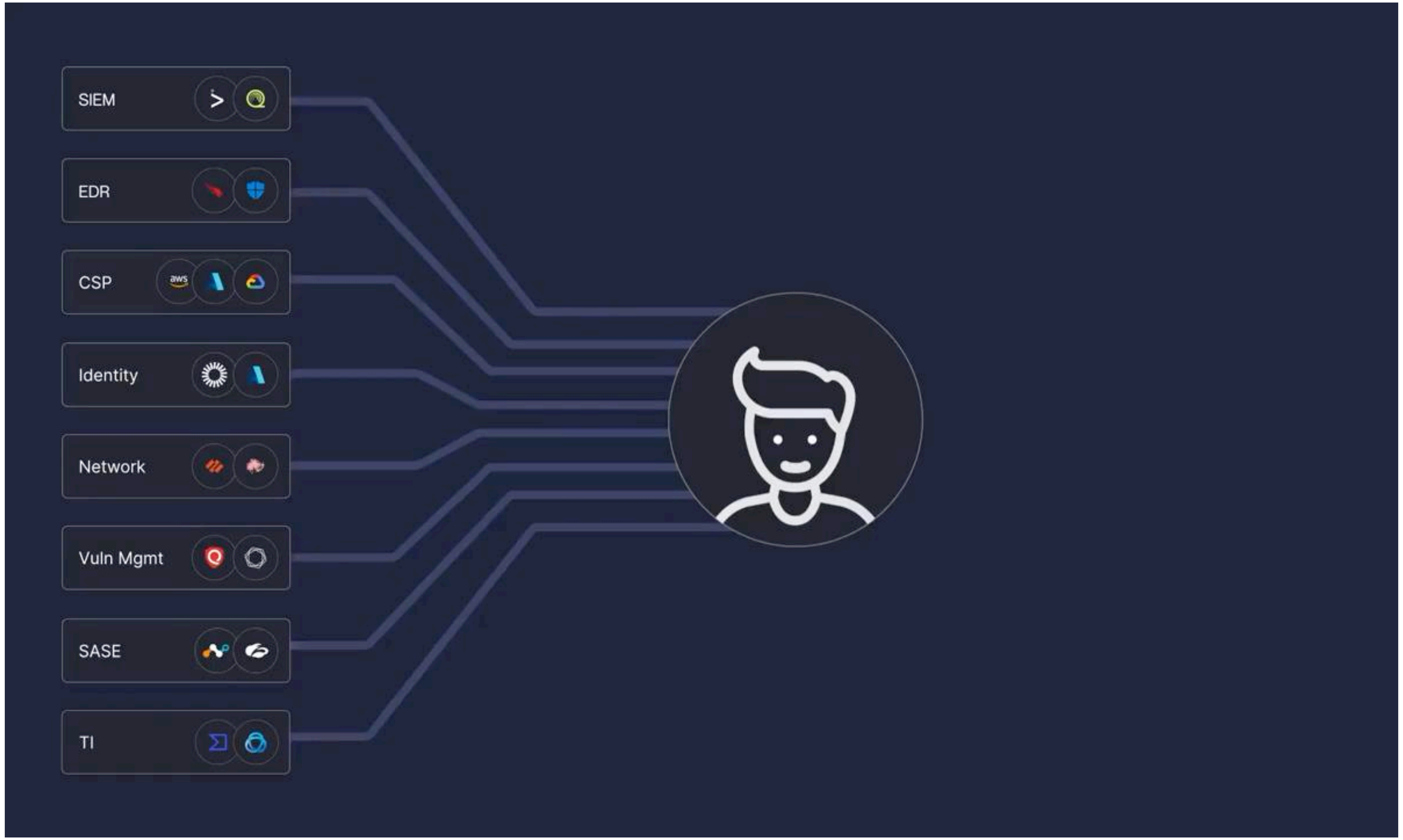
- Threat detection and prevention
- Automation in incident response

PROACTIVE MANAGEMENT

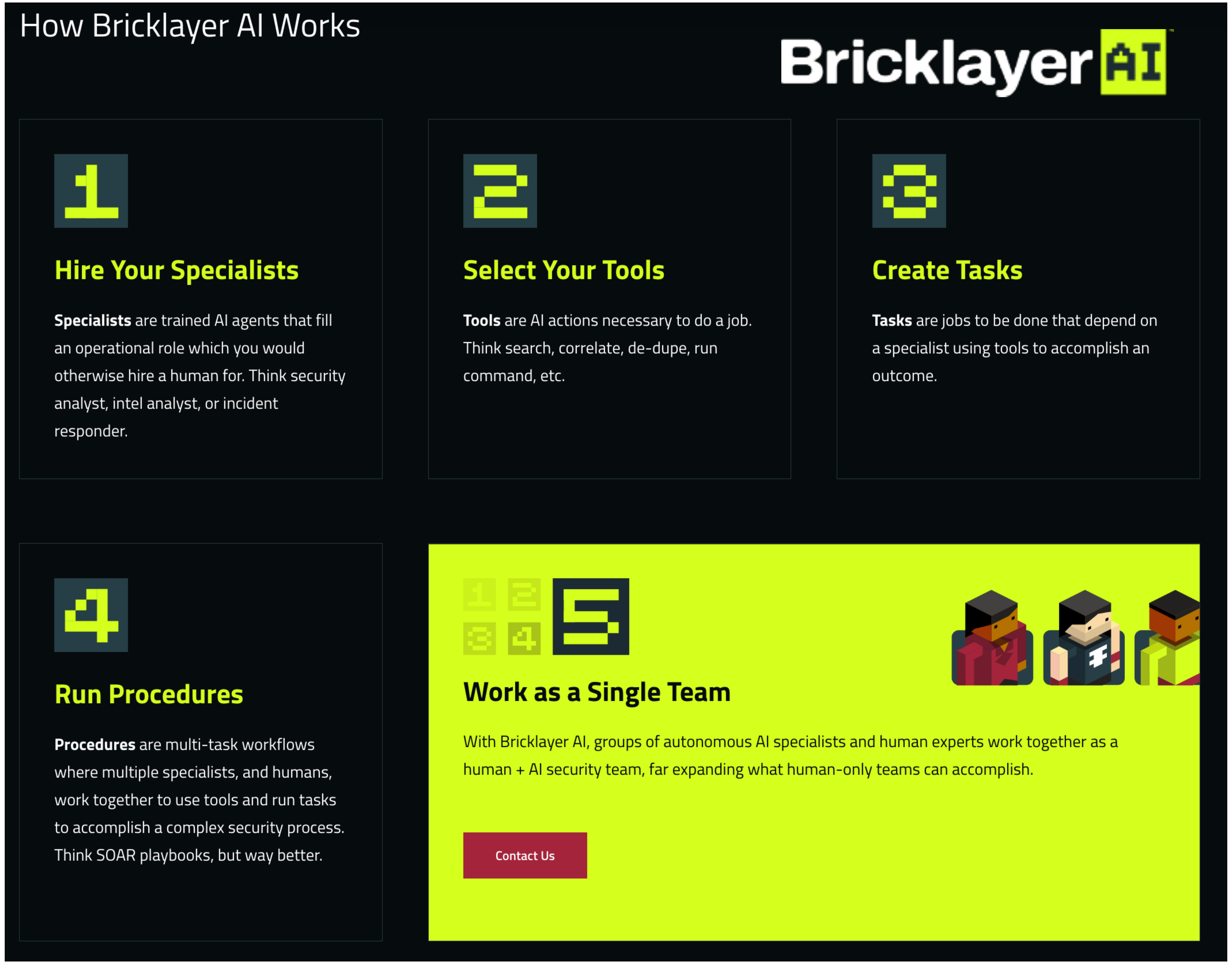
- Vulnerability scanning
- Predictive analytics

The rise of AI in cybersecurity

Dropzone AI



The rise of AI in cybersecurity



How Bricklayer AI Works

1

Hire Your Specialists

Specialists are trained AI agents that fill an operational role which you would otherwise hire a human for. Think security analyst, intel analyst, or incident responder.

2

Select Your Tools

Tools are AI actions necessary to do a job. Think search, correlate, de-dupe, run command, etc.

3

Create Tasks

Tasks are jobs to be done that depend on a specialist using tools to accomplish an outcome.

4

Run Procedures

Procedures are multi-task workflows where multiple specialists, and humans, work together to use tools and run tasks to accomplish a complex security process. Think SOAR playbooks, but way better.

5

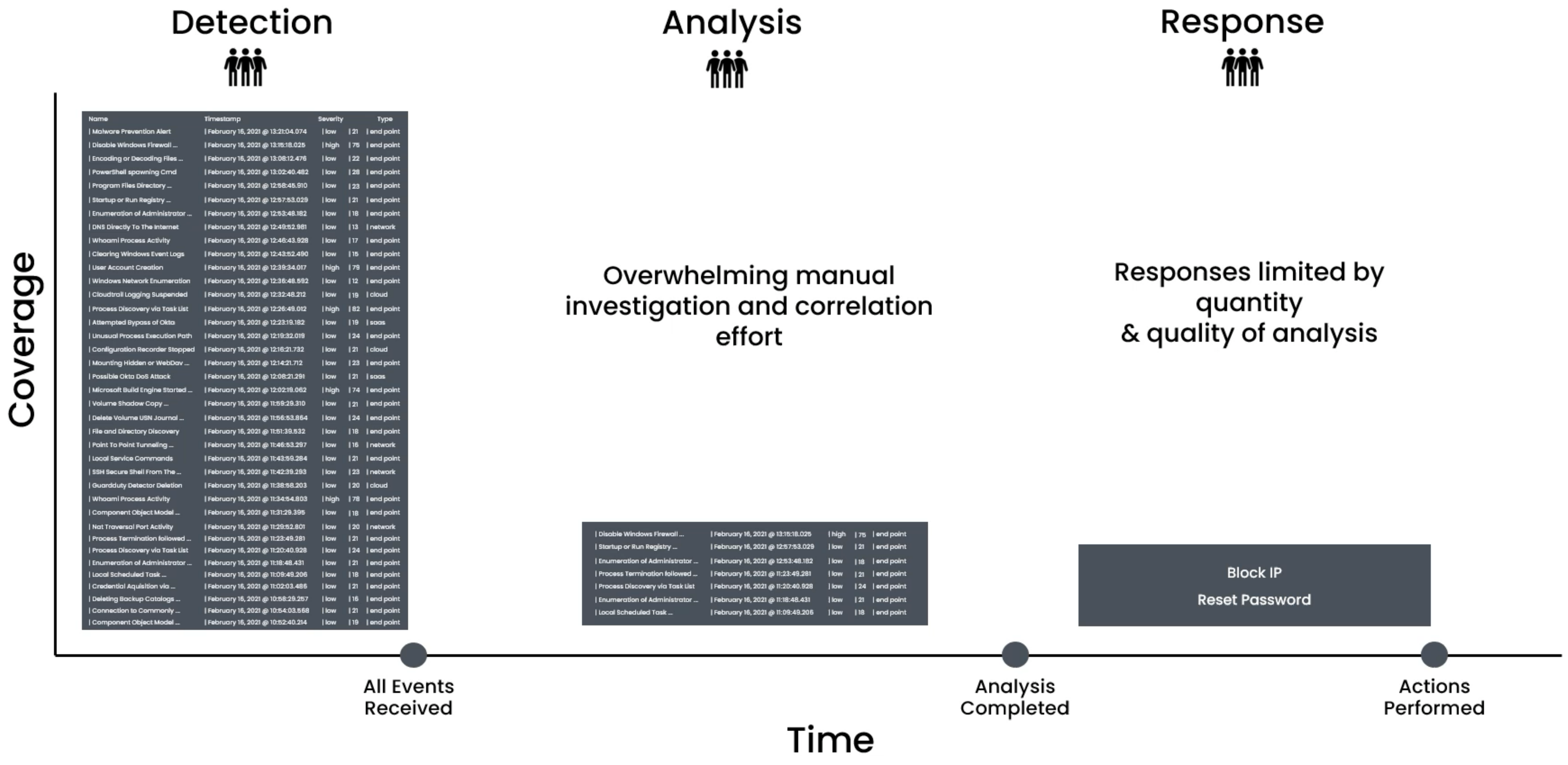
Work as a Single Team

With Bricklayer AI, groups of autonomous AI specialists and human experts work together as a human + AI security team, far expanding what human-only teams can accomplish.

The rise of AI in cybersecurity

DTonomy

Security Analysis



```
Name                               Timestamp                              Severity        Type
| Malware Prevention Alert          | February 16, 2021 @ 13:21:04.074    | low    | 21   | end point
| Disable Windows Firewall ...      | February 16, 2021 @ 13:15:18.025    | high   | 75   | end point
| Encoding or Decoding Files ...    | February 16, 2021 @ 13:08:12.476    | low    | 22   | end point
| PowerShell spawning Cmd           | February 16, 2021 @ 13:02:40.482    | low    | 28   | end point
| Program Files Directory ...       | February 16, 2021 @ 12:58:45.910    | low    | 23   | end point
| Startup or Run Registry ...       | February 16, 2021 @ 12:57:53.029    | low    | 21   | end point
| Enumeration of Administrator ...  | February 16, 2021 @ 12:53:48.182    | low    | 18   | end point
| DNS Directly To The Internet      | February 16, 2021 @ 12:49:52.981    | low    | 13   | network
| Whoami Process Activity           | February 16, 2021 @ 12:46:43.928    | low    | 17   | end point
| Clearing Windows Event Logs       | February 16, 2021 @ 12:43:52.490    | low    | 15   | end point
| User Account Creation             | February 16, 2021 @ 12:39:34.017    | high   | 79   | end point
| Windows Network Enumeration       | February 16, 2021 @ 12:36:48.592    | low    | 12   | end point
| Cloudtrail Logging Suspended      | February 16, 2021 @ 12:32:48.212    | low    | 19   | cloud
| Process Discovery via Task List   | February 16, 2021 @ 12:26:49.012    | high   | 82   | end point
| Attempted Bypass of Okta          | February 16, 2021 @ 12:23:19.182    | low    | 19   | saas
| Unusual Process Execution Path    | February 16, 2021 @ 12:19:32.019    | low    | 24   | end point
| Configuration Recorder Stopped    | February 16, 2021 @ 12:16:21.732    | low    | 21   | cloud
| Mounting Hidden or WebDav ...     | February 16, 2021 @ 12:14:21.712    | low    | 23   | end point
| Possible Okta DoS Attack          | February 16, 2021 @ 12:08:21.291    | low    | 21   | saas
| Microsoft Build Engine Started ...| February 16, 2021 @ 12:02:19.062    | high   | 74   | end point
| Volume Shadow Copy ...            | February 16, 2021 @ 11:59:29.310    | low    | 21   | end point
| Delete Volume USN Journal ...     | February 16, 2021 @ 11:56:53.864    | low    | 24   | end point
| File and Directory Discovery      | February 16, 2021 @ 11:51:39.532    | low    | 18   | end point
| Point To Point Tunneling ...      | February 16, 2021 @ 11:46:53.297    | low    | 16   | network
| Local Service Commands            | February 16, 2021 @ 11:43:59.284    | low    | 21   | end point
| SSH Secure Shell From The ...     | February 16, 2021 @ 11:42:39.293    | low    | 23   | network
| Guardduty Detector Deletion       | February 16, 2021 @ 11:38:58.203    | low    | 20   | cloud
| Whoami Process Activity           | February 16, 2021 @ 11:34:54.803    | high   | 78   | end point
| Component Object Model ...        | February 16, 2021 @ 11:31:29.395    | low    | 18   | end point
| Nat Traversal Port Activity       | February 16, 2021 @ 11:29:52.801    | low    | 20   | network
| Process Termination followed ...  | February 16, 2021 @ 11:23:49.281    | low    | 21   | end point
| Process Discovery via Task List   | February 16, 2021 @ 11:20:40.928    | low    | 24   | end point
| Enumeration of Administrator ...  | February 16, 2021 @ 11:18:48.431    | low    | 21   | end point
| Local Scheduled Task ...          | February 16, 2021 @ 11:09:49.206    | low    | 18   | end point
| Credential Aquisition via ...     | February 16, 2021 @ 11:02:03.486    | low    | 21   | end point
| Deleting Backup Catalogs ...      | February 16, 2021 @ 10:58:29.257    | low    | 16   | end point
| Connection to Commonly ...        | February 16, 2021 @ 10:54:03.568    | low    | 21   | end point
| Component Object Model ...        | February 16, 2021 @ 10:52:40.214    | low    | 19   | end point
```

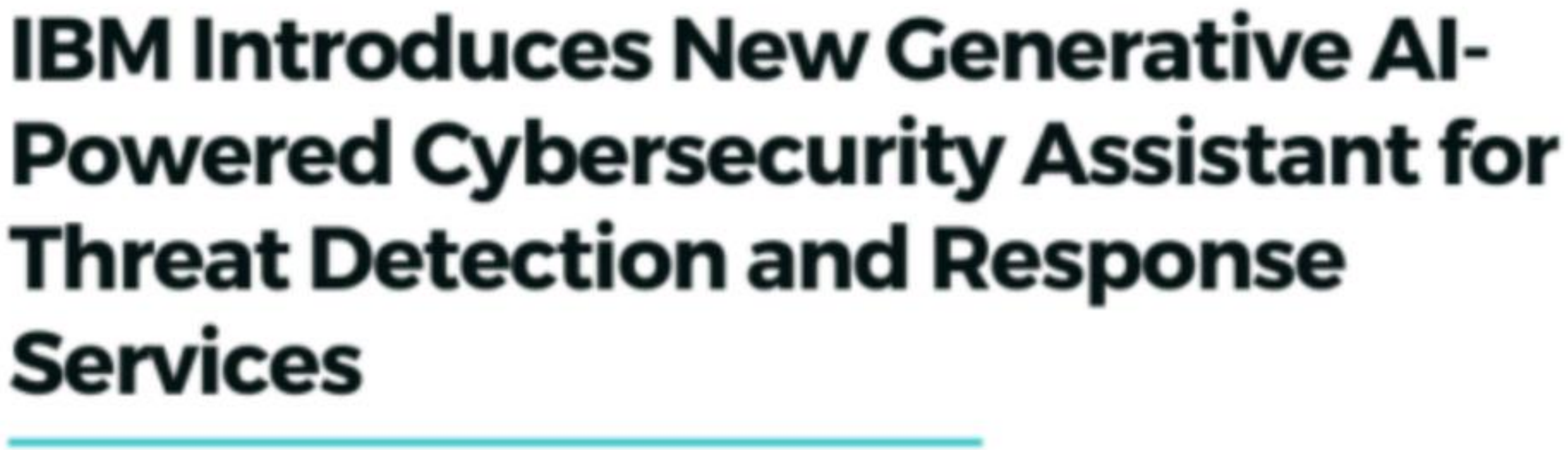
Overwhelming manual investigation and correlation effort

```
| Disable Windows Firewall ...      | February 16, 2021 @ 13:15:18.025    | high   | 75   | end point
| Startup or Run Registry ...       | February 16, 2021 @ 12:57:53.029    | low    | 21   | end point
| Enumeration of Administrator ...  | February 16, 2021 @ 12:53:48.182    | low    | 18   | end point
| Process Termination followed ...  | February 16, 2021 @ 11:23:49.281    | low    | 21   | end point
| Process Discovery via Task List   | February 16, 2021 @ 11:20:40.928    | low    | 24   | end point
| Enumeration of Administrator ...  | February 16, 2021 @ 11:18:48.431    | low    | 21   | end point
| Local Scheduled Task ...          | February 16, 2021 @ 11:09:49.206    | low    | 18   | end point
```

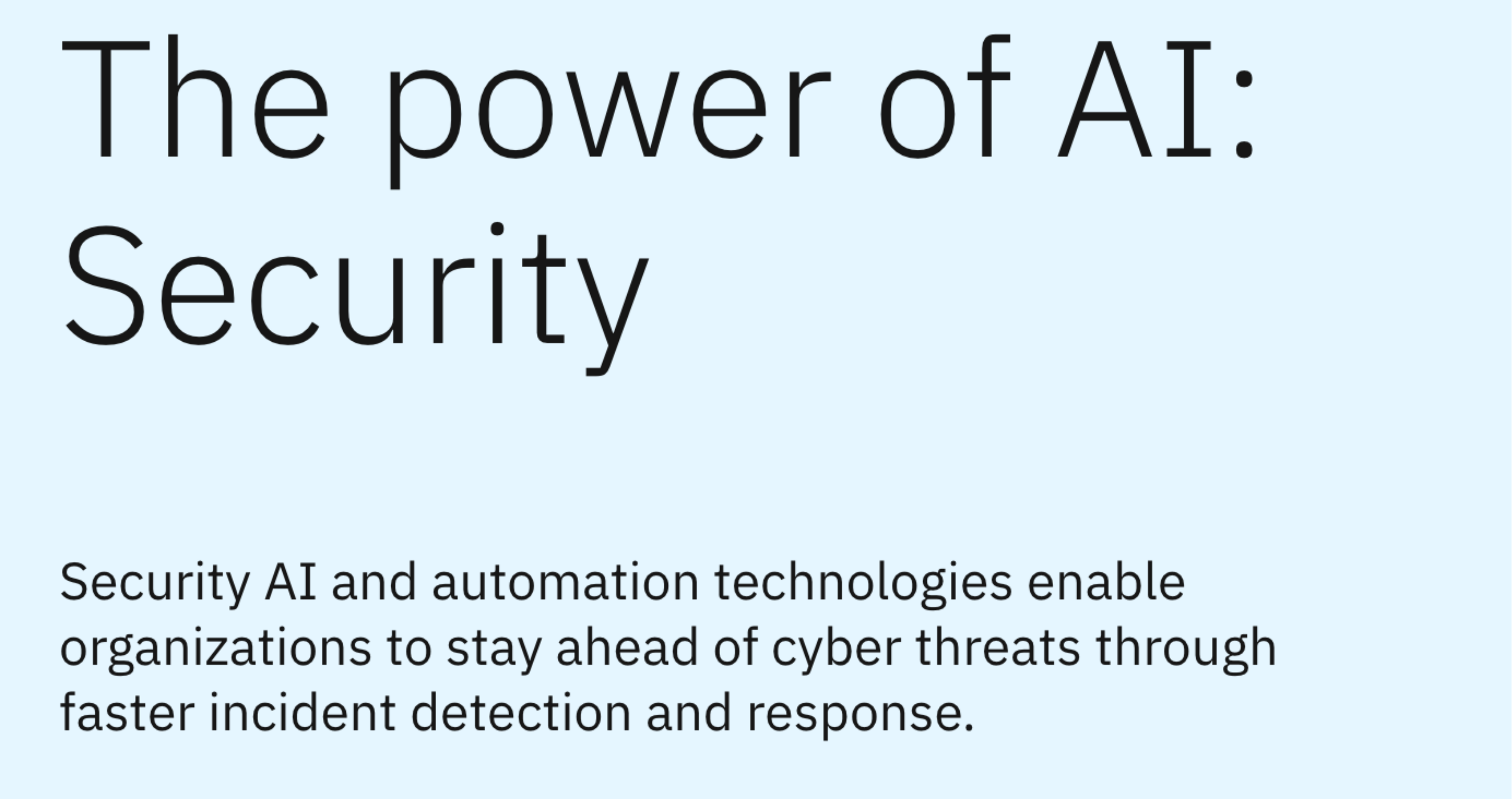
Responses limited by quantity
& quality of analysis

Block IP
Reset Password

The rise of AI in cybersecurity



IBM Introduces New Generative AI-Powered Cybersecurity Assistant for Threat Detection and Response Services



The power of AI: Security

Security AI and automation technologies enable organizations to stay ahead of cyber threats through faster incident detection and response.

Ethical Considerations in AI for Cybersecurity

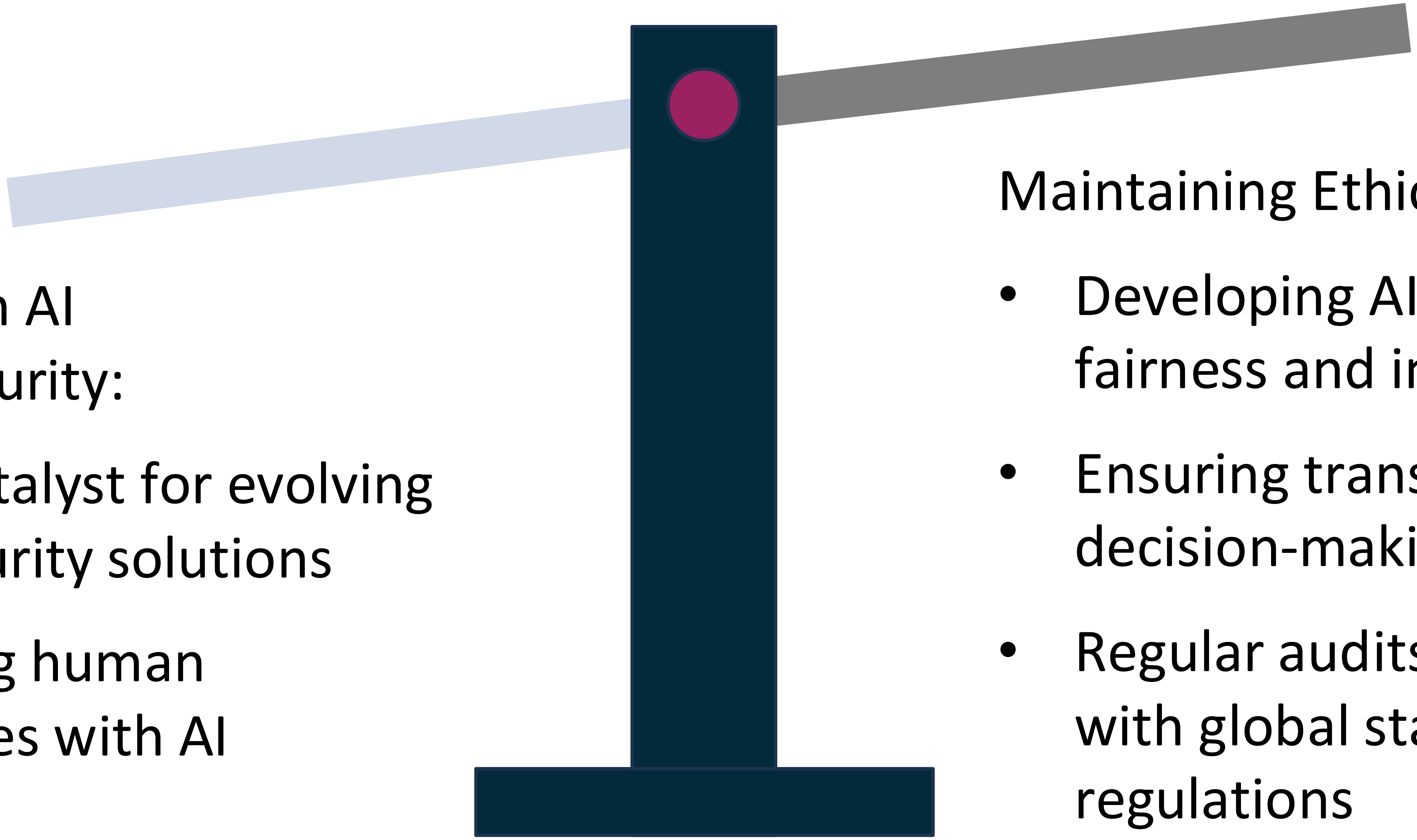
Bias in Algorithms

Privacy Concerns

Accountability and Transparency

Autonomy and Human Oversight

Balancing Innovation with Integrity



Innovation in AI
for Cybersecurity:

- AI as a catalyst for evolving cybersecurity solutions
- Enhancing human capabilities with AI

Maintaining Ethical Standards:

- Developing AI systems with fairness and inclusivity in mind
- Ensuring transparency in AI decision-making
- Regular audits and compliance with global standards and regulations

Legal and Regulatory Frameworks

Examples of Regulations and Guidelines:

- GDPR (General Data Protection Regulation)
- AI Act (EU) and NIST AI Risk Management Framework (US)
- Cybersecurity-specific regulations

Impact of Compliance on AI Development:

- Encouraging ethical AI practices
- Challenges of adhering to legal standards in fast-paced innovation

Practical Strategies for Ethical AI in Cybersecurity

Key Best Practices:

Ethical AI Frameworks

Human in the Loop (HITL) Systems

Cross-functional Collaboration

Transparent AI Audits

Challenges and Future Directions

Challenges:

- Balancing innovation with ethical concerns
- Rapidly evolving cyber threats and the pace of AI development
- Ensuring global cooperation on AI ethics in cybersecurity

Future Trends:

- Growing focus on explainable AI (XAI)
- AI-driven advancements in proactive cybersecurity
- Ethical AI innovation hubs and collaborative global efforts

Conclusion

- The growing adoption of AI in cybersecurity continues to transform the field.
- Along with its benefits, AI introduces new ethical challenges and risks that must be carefully managed.
- Responsible use of AI in cybersecurity is essential to ensure fairness, transparency, and accountability.
- Achieving the right balance between technological innovation and ethical responsibility is crucial for sustainable progress in the industry.

Q
&
A

University of Salford MANCHESTER

salford.ac.uk

Dr Angel Jimenez-Aranda
a.jimenez-aranda@salford.ac.uk