AI Security - What is it?

Navigating the Intersection of Security, Compliance, and AI in Enterprise

May 2024

Intersection of Security and AI

Today's Objectives

- Defining AI
- Understanding the intersection between AI and Cybersecurity
- Provide a glimpse into how and where we are seeing AI being used

Introduction

Adam Durbin
Chief Technology Officer
Mantel Group

Natalie Rouse
CEP | Mantel Group
AI Forum Executive Council

Mantel Group Overview

Mantel Group pursues technologies that change the way clients do business in the real world, offering end-to-end solutions across four capability pillars:

Digital, Cloud, Data and Cybersecurity.

Highlights

- Leading independent digital transformation services company in Australia & New Zealand
- 850+ experts
- Operating at over 150 clients
- Consistently winning Best Workplaces awards

We are top tier partners

Best Workplaces™ Great Place To Work. AUSTRALIA 2021

Best Workplaces™ Great Place To Work. AUSTRALIA 2022

Best Workplaces™ for Women Great Place To Work. AUSTRALIA 2022

Best Workplaces™ in Technology Great Place To Work. AUSTRALIA 2023

Best Workplaces™ Great Place To Work. AUSTRALIA 2023

Great Place To Work. Certified APR 2023–APR 2024 NZ

Before we start - what is AI?

A simple definition

> “Artificial Intelligence (AI) at its most simple, is a sub-field of computer science with the goal of creating programs that can **perform tasks generally performed by humans**.
>
> These tasks can be considered intelligent, and include visual and audio perception, learning and adapting, reasoning, pattern recognition and decision-making.
>
> **‘AI’ is used as an umbrella term** to describe a collection of related techniques and technologies including machine learning, predictive analytics, natural language processing and robotics.”

Why Security for AI is no longer an option

Some simple examples of how AI is impacting security

It's Here

Individuals and organisations are being targeted today leveraging AI enabled attacks and AI systems being compromised.

2 simple examples:

- Phishing
- Deep Fakes[1]

It's New

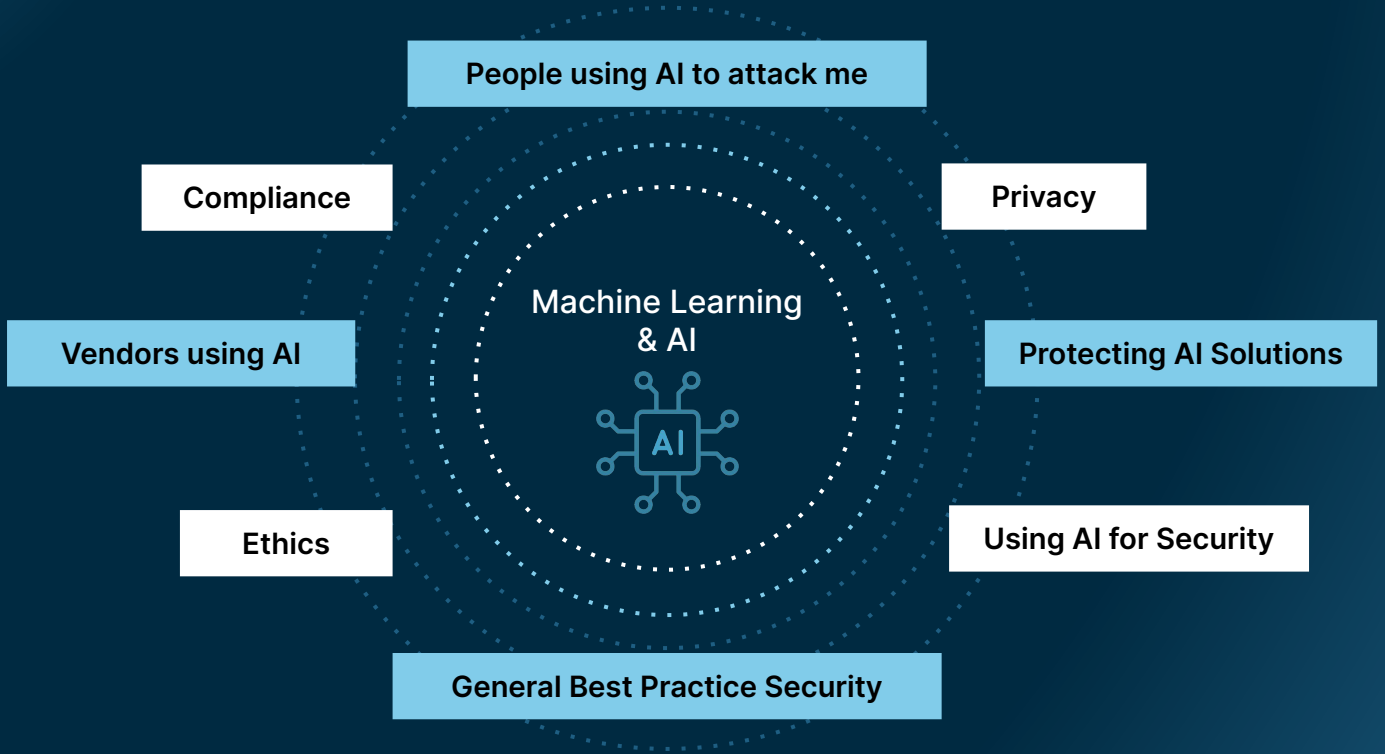
25 out of 28 companies (~90%) interviewed had not implemented any strategies to account for Adversarial AI attacks.

Security has been a challenge for new technologies in the past - this is not new.

Source[1]: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

Simply breaking down AI Security

Terms like Adversarial AI and AI Security are often thrown around however where are the key intersections we see between Security and AI.



Trustworthy AI Principles

5 Artificial Intelligence (AI) Ethics Principles designed to ensure AI is safe, secure and reliable.

Human Oversight & Accountability

AI stakeholders should retain an appropriate level of human oversight of AI systems and their outputs

Wellbeing

AI stakeholders should utilise AI systems in service of the wellbeing of New Zealand's people and environment

Transparency

The operation and impacts of an AI system should be transparent, traceable, auditable and generally explainable to a degree appropriate to its use and potential risk profile

Reliability, Security & Privacy

AI stakeholders must ensure AI systems and related data are reliable, accurate and secure and the privacy of individuals is protected throughout the AI system's life cycle

Fairness & Justice

AI systems must respect applicable laws, human rights, Māori Data Sovereignty, democratic values, and principles of equality & fairness.

Protecting the use of AI/ML

Understanding some of the potential attacks against ML models

During Development

Data Access

- System/Data Access

Poisoning

- Indirect/direct Poisoning
- Data Injection
- Data Manipulation
- Logic Corruption

During Run-Time

Evasion

- Single-step (Gradient)
- Iterative (Gradient)
- Gradient-free

Oracle

- Extraction
- Inversion
- Membership Inference

"A TAXONOMY AND TERMINOLOGY OF ADVERSARIAL MACHINE LEARNING"

So why use AI?

From increased productivity to reimagined customer experience

Internal → External

Low Risk

Banking

‘Security Architecture as Code’ Framework

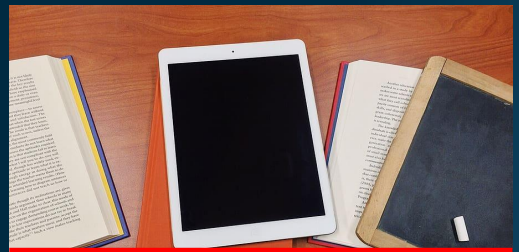
- Compliant, de-risked and standardised output
- Speed to delivery
- Unlocked $m’s business value

Medium Risk

Energy

Internal IT Support interface

- Knowledge base integrated
- Business productivity enabler
- Accuracy over 90%



High Risk

Education

Customised Services chatbot

- Knowledge corpus powered
- Unique ‘adapter first’ approach to training
- Enhanced student experience

Mantel group