



Spring 2024

GIPAW Conference

AI Risks and AI Risk Mitigation





As we went
over in the
**“Using AI to be
more
productive”**, AI
can be a great
tool...

but...

Artificial Intelligence Risks

- 1. Cyberattacks and Data Breaches:** AI systems rely on large amounts of data, including sensitive personal information, which can be targeted by cyber attackers. Proper data security measures are crucial to prevent breaches.
- 2. Lack of AI Governance:** Many local governments lack formal AI governance programs, increasing risks of ethical concerns, privacy violations, and public distrust.
- 3. Absence of Clear Guidelines:** The lack of comprehensive AI strategies and guidelines can lead to misuse, biased algorithms, and discriminatory outcomes.
- 4. Algorithmic Bias and Discrimination:** AI models can perpetuate biases present in their training data, resulting in unfair or discriminatory decisions impacting citizens.

Artificial Intelligence Risks

- 5. Amplification of Human Biases:** AI systems trained on biased data can amplify societal biases and discrimination against certain groups.
- 6. Lack of Transparency and Accountability:** The opaque nature of AI decision-making processes raises concerns over transparency and accountability, eroding public trust.
- 7. Job Displacement:** Automation through AI may lead to job losses in certain sectors, requiring transition plans and retraining programs.

Artificial Intelligence Risks

- 8. Public Records Compliance.** AI creates challenges for local government to comply with WI public records statutes. That require full disclosure of government records with few exemptions and impose penalties for noncompliance.
- 9. Public Trust.** Unless the use of AI can degrade public trust. A tool that works well in the private sector may not work well in government without full disclosure and understanding.
- 10. Lack of Knowledge.** Elected officials, decision makers and the public may not fully understand the tools and associated risks.
- 11. Too much trust.** People trust the work product / output of AI too much.

AI Cyber Risk – Dark Web Tools

- **DarkBERT:** A GPT-based malware that uses the entirety of the Dark Web as its knowledge base. This adversarial tool can analyze new pieces of Dark Web content, written in its own dialects and heavily-coded messages, and extracting useful information from it.
- **WormGPT:** Unlike ChatGPT, which has built-in protections against misuse, WormGPT is designed with “no ethical boundaries or limitations,” making it a potent tool for hackers.
- **FraudGPT:** Tool to assist with fraud.

There are others



AI Cyber Risk – Deepfake

Deepfakes are fake video and/or audio simulations created using AI and deep learning techniques.

They substitute a person's likeness with a "digital fake" in order to mislead, perpetrate fraud or carry out other malicious functions.

- They swap in a synthetic/manipulated version of a person's face, voice, etc.
- **The purpose is to deceive by making it appear someone said or did something they didn't**
- Common malicious uses include non-consensual porn, defamation, fraud, and disinformation campaigns
- Deepfake technology is becoming more accessible but also more advanced and harder to detect and has been successfully used in theft.




AI Cyber Risk – Deepfake Example #1

Binance: The Chief Communications Officer of Binance, a blockchain ecosystem, was deepfaked. The fraudsters held 20-minute “investment” Zoom calls, trying to convince the company’s clients to turn over their Bitcoin for scammy investments.

The clients were sent links to faked LinkedIn and Telegram profiles claiming to be the officer, inviting them to various meetings to talk about different listing opportunities.

The criminals then used a convincing-looking holograph of the officer in Zoom calls to try and scam several representatives of legitimate cryptocurrency projects.



AI Cyber Risk – Deepfake Example #2

Hong Kong: A multinational company's Hong Kong office suffered a significant financial loss of HK\$200 million (US\$25.6 million) due to a sophisticated scam involving deepfake technology.

The scam featured a digitally recreated version of the company's chief financial officer, along with other employees, who appeared in a video conference call instructing an employee to transfer funds.

The scammers were able to convincingly replicate the appearances and voices of targeted individuals using publicly available video and audio footage.



AI Cyber Risk – Deepfake Other Examples

#3 CEO Fraud : A scammer used AI-powered voice technology to impersonate a German CEO. The UK CEO, believing they were interacting with their German counterpart, followed instructions that led to financial loss.

#4 CEO Fraud: In one case, a US-based business lost nearly \$400k when the payments team received an email from the CEO asking for payments to be set up for new beneficiaries. In another case, a global commodity trading platform provider lost £920,000 (\$984k) when an employee received an email from the CEO requesting a new payment.

Deepfake scams can be used in elections, personal and corporate scams.

CONTROL vs NO CONTROL

Things you CAN mitigate

- Cyber Attacks
- Governance
- Policy
- Training
- Transparency / Accountability
- Bank Fraud

Things you can't directly mitigate

- Job Displacement
- Bias
- Lack of Guidelines/Law
- Deepfake

Mitigation 1a: AI Training & Education

Train staff, decision makers, contractors and elected officials.

- Appropriate Use and Misuse of AI tools
- Personal responsibility and decision making when using AI generated content – including identifying AI hallucinations
- Identifying AI hallucinations
- Understanding of “approved” AI tools

Mitigation 1b: AI Training & Education

Train staff, decision makers, contractors and elected officials.

- Public Records Compliance
- Ethical AI Governance – public oversight
- Knowing and using mitigations for financial deepfakes
- Contractor, vendor and partner management as it relates to AI including clarification of accountability and responsibility

Mitigation 2: Government Workforce Planning

Sooner or later the use of AI will impact government operations, likely impacting: call centers, repetitive processes, administrative functions and recurring/repetitive well-defined activities.

- Be transparent, communicate changes in advance
- **Review and audit AI work product**
- Prepare staff and consider education and retraining programs for impacted private sector and or public sector workers

Mitigation 3: Cyber Security

The bad guys are using AI – you will need to if you aren't already.

Consider tools such as: DarkTrace, CrowdStrike Falcon, Vectra AI, Symantec Endpoint, IBM Watson for Security, Palo Alto Network Cortex XDR, CheckPoint Infinity, and Fortinet FortiAI

Mitigation 4: Policy & Guidelines

Recommend, Propose or Establish

- AI Use Policy. Draft a policy and submit it.
 - Appropriate use
 - Employee Responsibility
- AI Misuse Policy
- Policy Protections for Deepfake Financial Fraud (two signature)
- County Board and/or County Exec (elected official oversight)
- Required training for everyone
- Policy AI Public Records Compliance

Questions and Answers

august@nevermanconsulting.com
414-380-9701

Neverman
Consulting
Trustworthy Solutions for Complex Problems