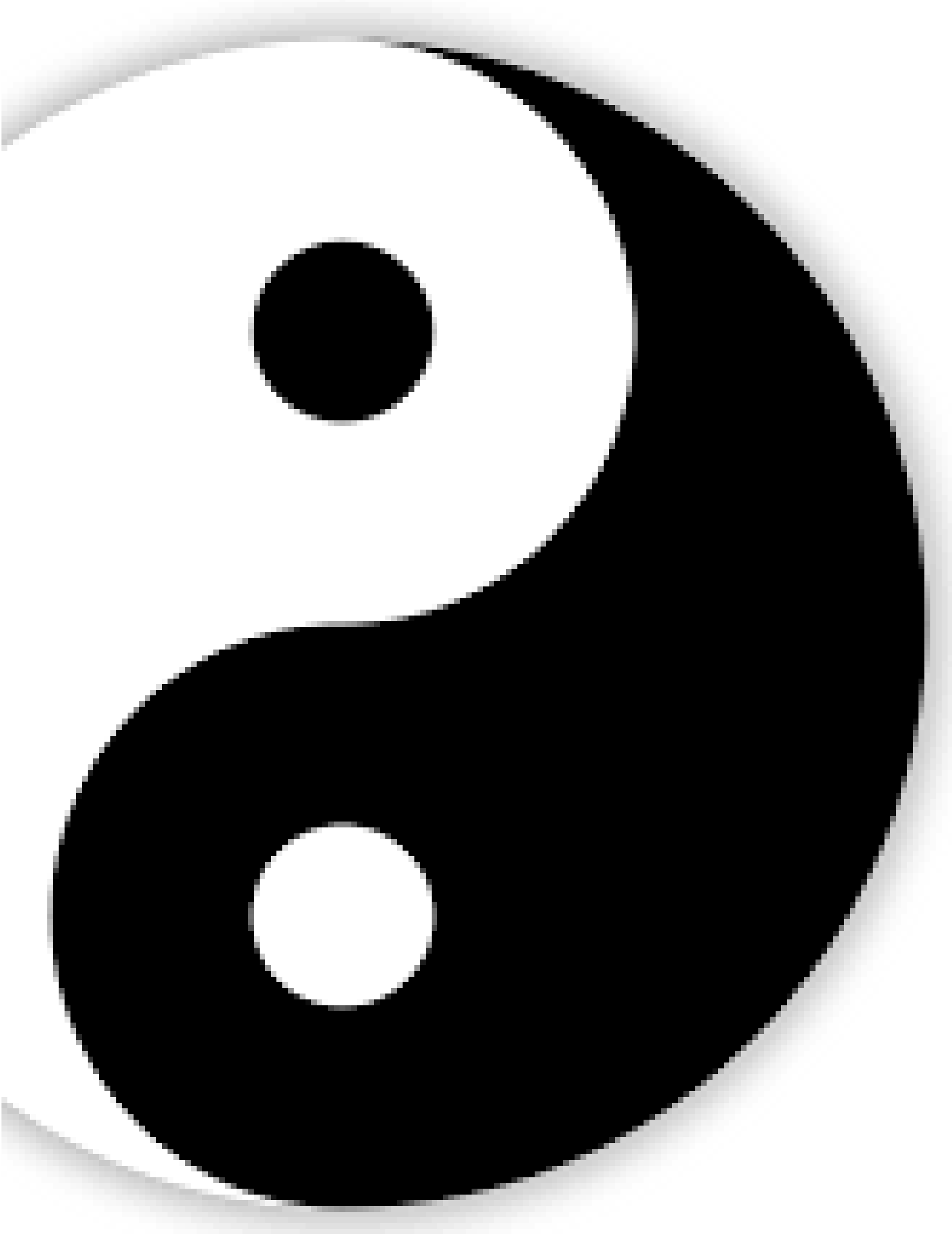
# AI: Through the lens of ChatGPT & Deep Fakes

***The Yin and the Yang***

Presenter: Deb Wells

24 May 2023



# Overview

- Introduction
- What is ChatGPT?
- What are Deep Fakes?
- Exercises
- The Elephant in the Room

# Who Am I?

- Senior Manager, Cybersecurity Engineering, BECU
- Adjunct Lecturer at Central Washington University
- Spent 21 years in the US Air Force
  - Computers and Communications
- Wife, Mom, and animal lover!

“You have to talk about ‘The Terminator’ if you’re talking about artificial intelligence. I actually think that that’s way off. I don’t think that an artificially intelligent system that has superhuman intelligence will be violent. I do think that it will disrupt our culture.”

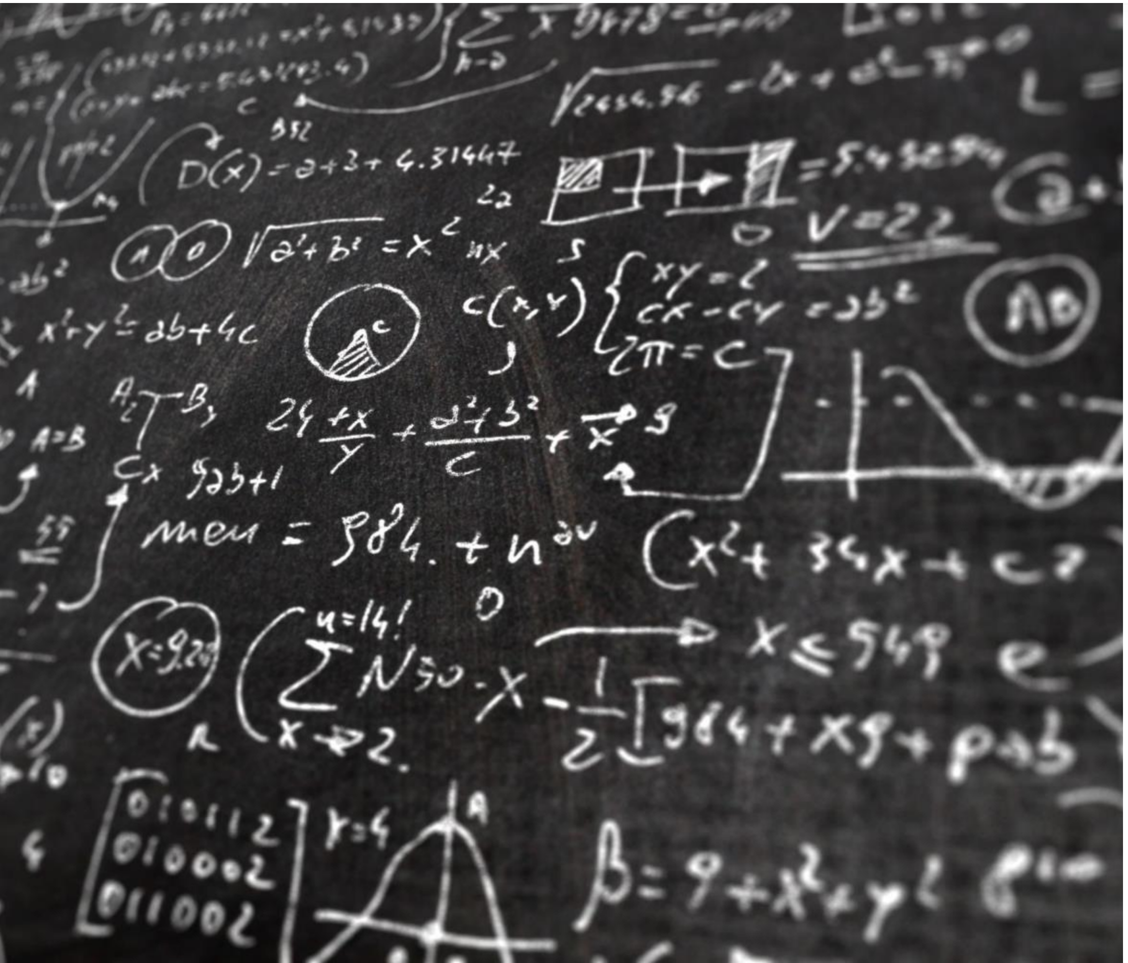
—Gray Scott

# Chat GPT

## What is ChatGPT?

- Yin
- Yang

# What is ChatGPT?



GPT = Generative Pre-Trained Transformer

Developed by OpenAI

Uses deep learning algorithms to generate human-like text based on a given prompt or input

Generates responses to what you want

Pretrained by humans

Transforms input into meaningful output

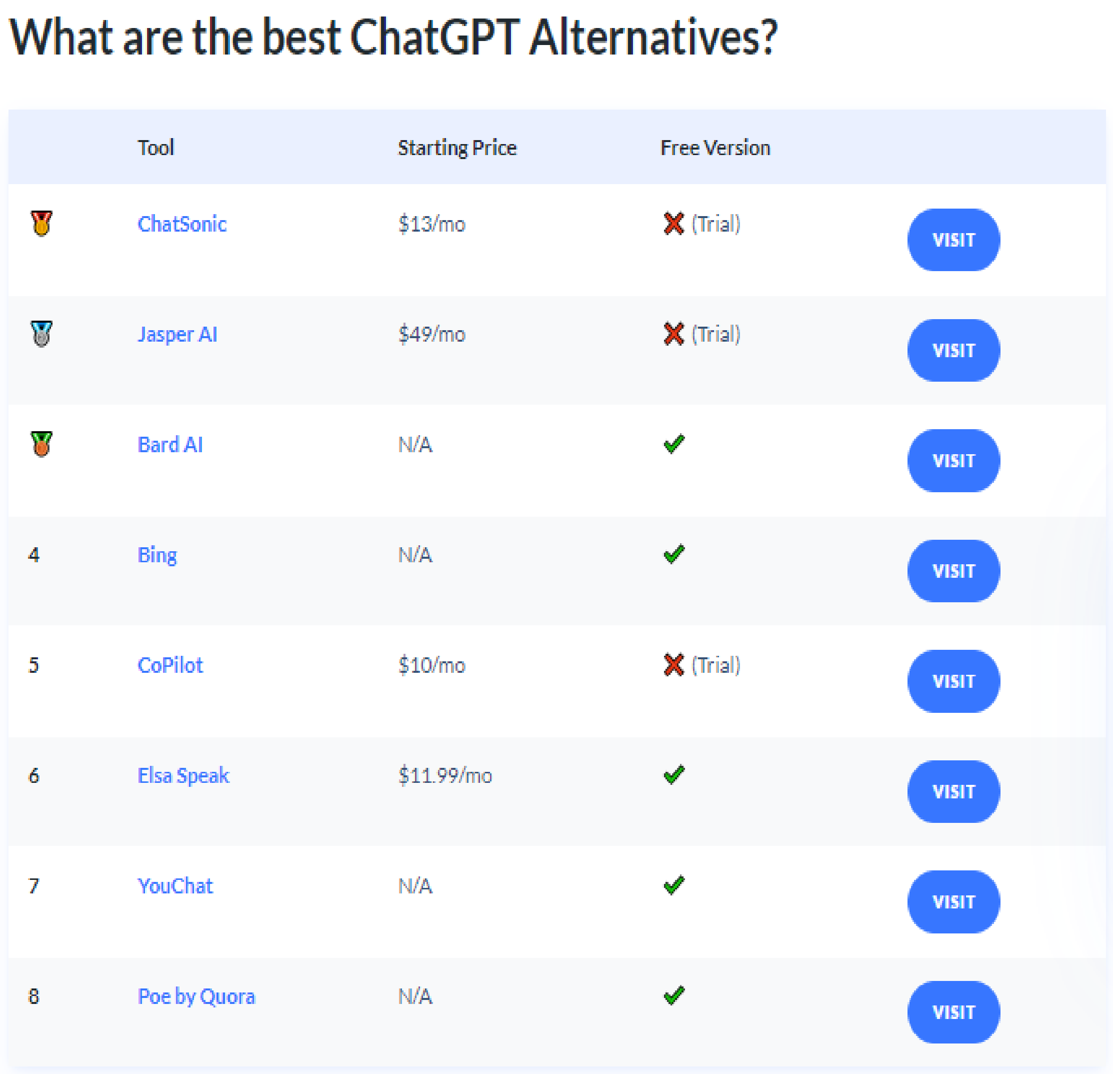
Understands human language and context

Tailored responses based on previous conversations

Retrieves vast amounts of information

# Who Makes this Software

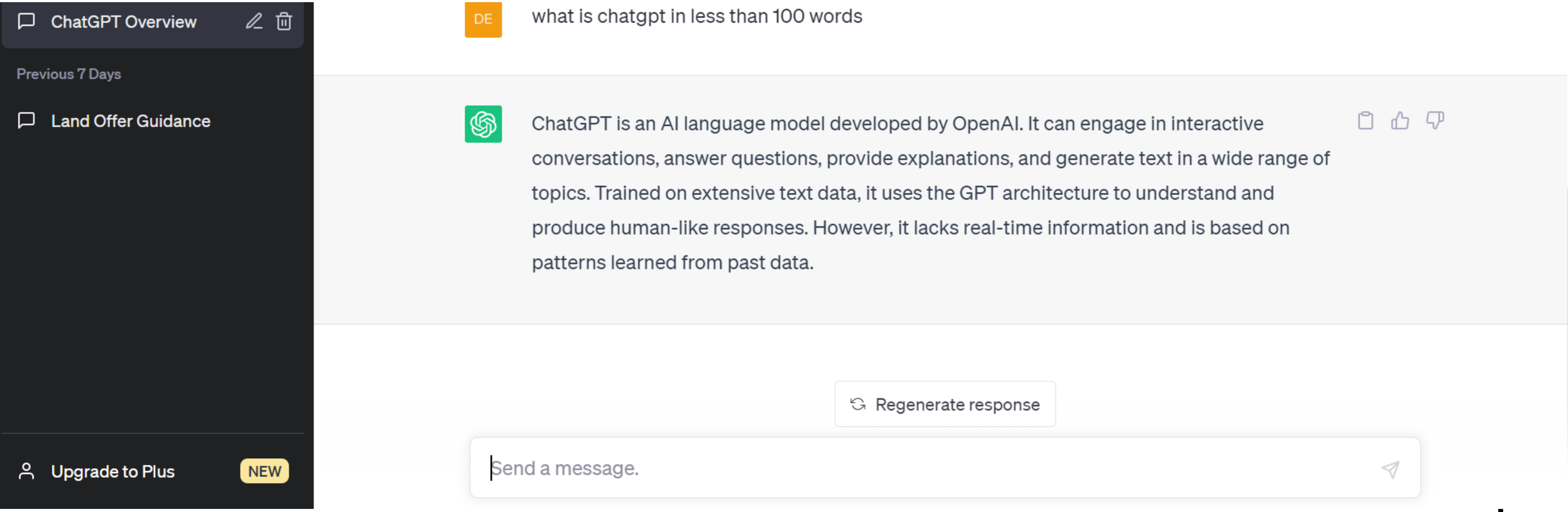
OpenAI – Sam Altman is the “Father of the ChatGPT”



## What are the best ChatGPT Alternatives?

| | Tool | Starting Price | Free Version | |
|---|---|---|---|---|
| 🥇 | ChatSonic | $13/mo | ✗ (Trial) | VISIT |
| 🥈 | Jasper AI | $49/mo | ✗ (Trial) | VISIT |
| 🥉 | Bard AI | N/A | ✓ | VISIT |
| 4 | Bing | N/A | ✓ | VISIT |
| 5 | CoPilot | $10/mo | ✗ (Trial) | VISIT |
| 6 | Elsa Speak | $11.99/mo | ✓ | VISIT |
| 7 | YouChat | N/A | ✓ | VISIT |
| 8 | Poe by Quora | N/A | ✓ | VISIT |

# SAMPLE from CHATGPT



what is chatgpt in less than 100 words

ChatGPT is an AI language model developed by OpenAI. It can engage in interactive conversations, answer questions, provide explanations, and generate text in a wide range of topics. Trained on extensive text data, it uses the GPT architecture to understand and produce human-like responses. However, it lacks real-time information and is based on patterns learned from past data.

# The Yang…positive, bright, the good

Medical break throughs –

Useful for generating high-quality human-like text

Automating customer service and engagement

Powering chatbots and virtual assistants

Language translation

Advancing research and development in natural language processing

More automated, more efficient redlining of contracts

Applying pre-set language and parameters to large batches of contracts

# The Yin…the negative, dark, the bad

Fake news can come out of GPT

Polarization – even more so than today!

Putting people out of work

Super Intelligent AI taking control

Bias, discrimination

National Security Concerns

# Real World Examples

Massey Students - Students caught cheating with ChatGPT offered amnesty for confession

Furman University – another case of cheating using ChatGPT

*“Hallucinations”* - Refers to the generation of outputs that may sound plausible but are either factually incorrect or unrelated to the given context

# How Can I Tell?

AI Content Detector - https://copyleaks.com/ai-content-detector

AI Writing Check

Originality.AI

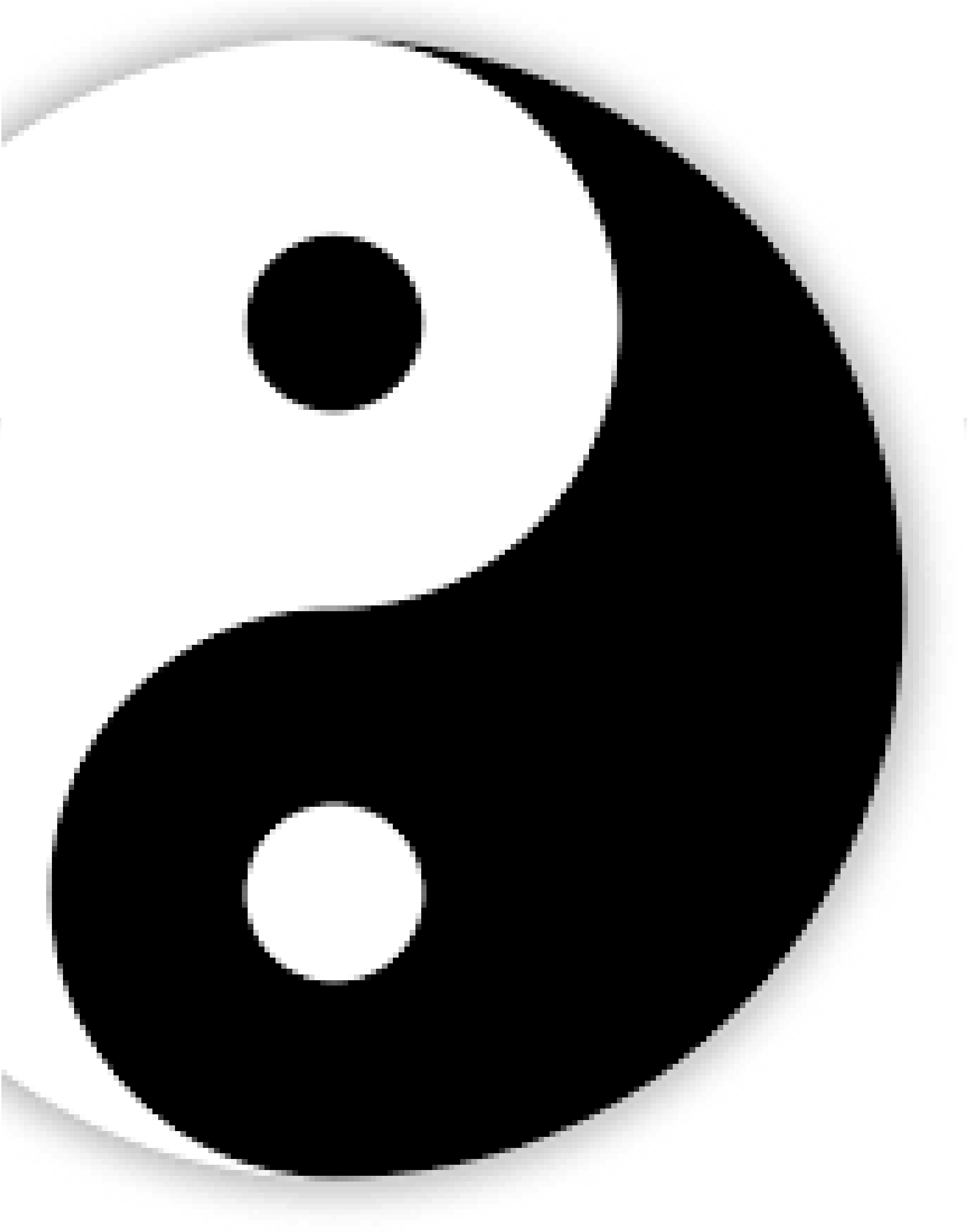
Writer

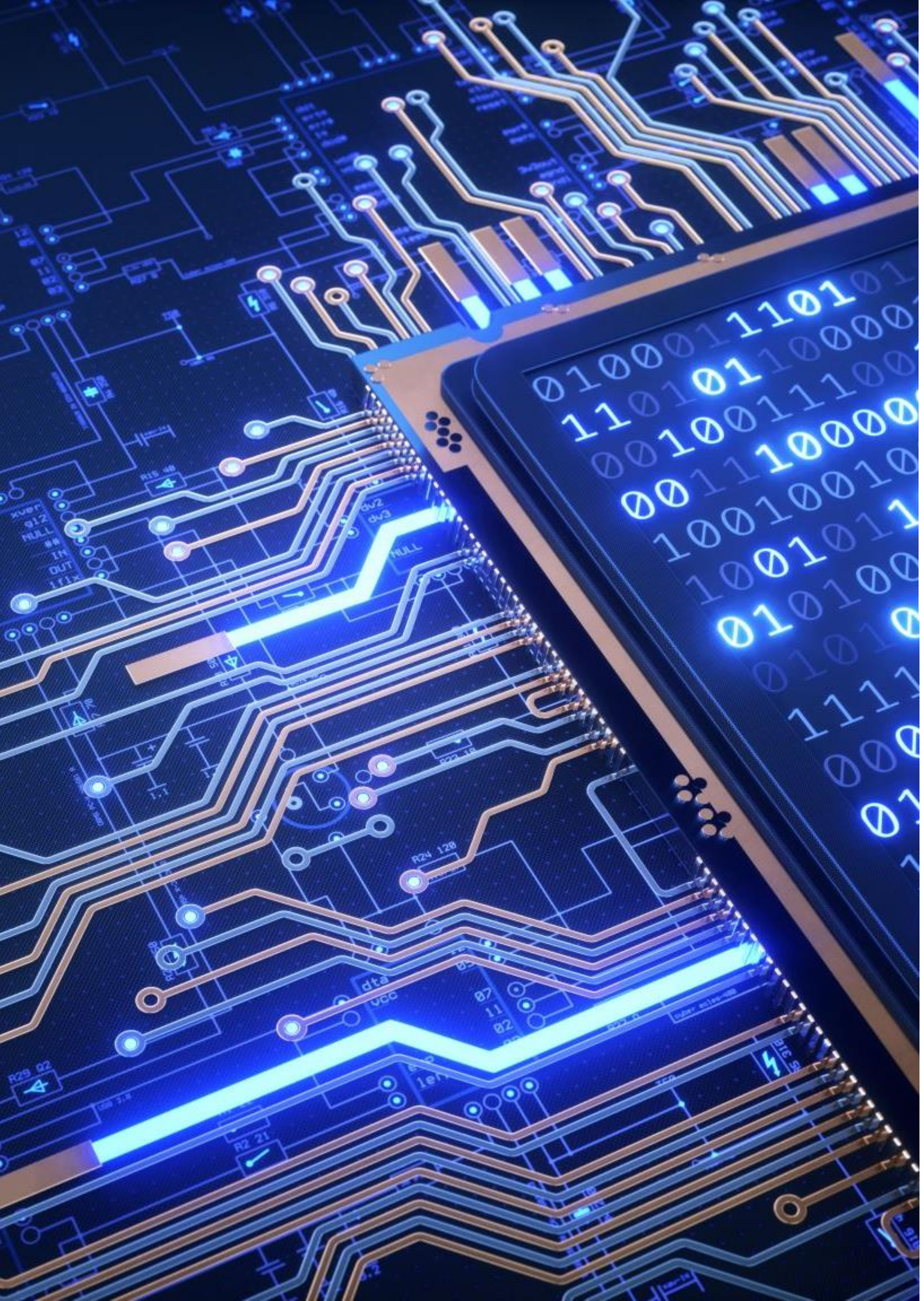
Copyleaks AI Content Detector

# Deepfakes

## What are Deepfakes?

- Yin
- Yang





# Overview – Deepfakes

Introduction to Deepfakes

Technology surrounding Deepfakes

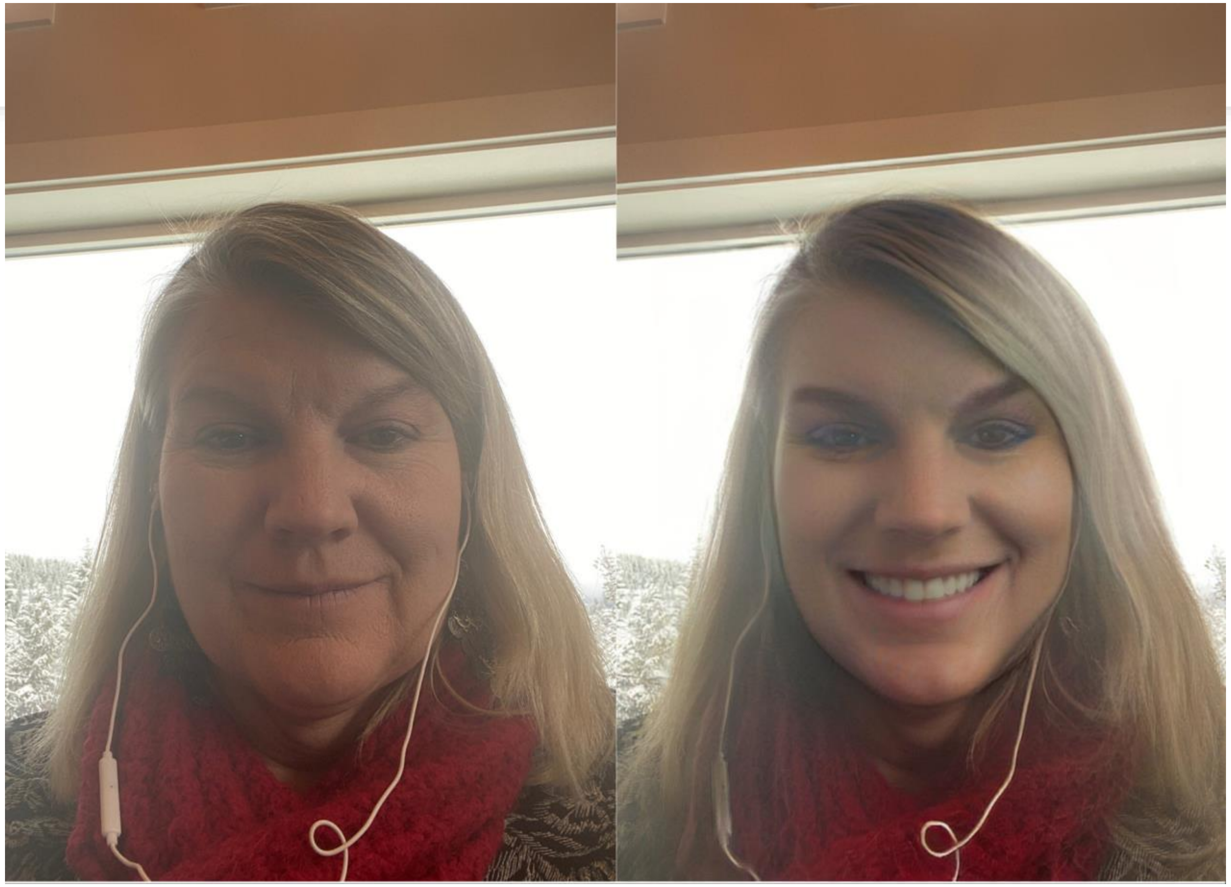
Deepfake Categories and Examples

Cybersecurity Concerns

Discerning deepfakes

Governance and Laws Around Deepfakes

# *“Seeing is believing”*…or is it?



This was done on FaceApp

# Deepfakes - Defined

Conjunction – Deep (meaning AI or ML) + Fake (not real)

- Output = Term ***Deepfake***

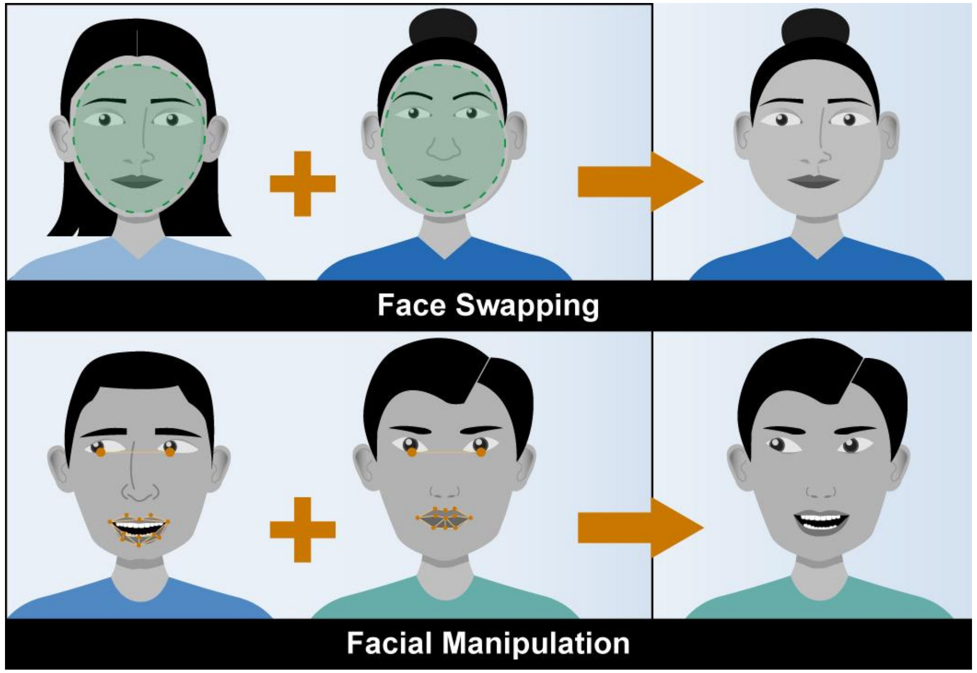
Artificial images and sounds put together with machine-learning algorithms

- Can create people who do not exist
- Can impose on real people actions and words they did not really say

Started in late 2017 – Reddit user began uploading videos of celebrities onto the body of porn star

Synthetic Media

# Deepfakes



Source: GAO. | GAO-20-379SP

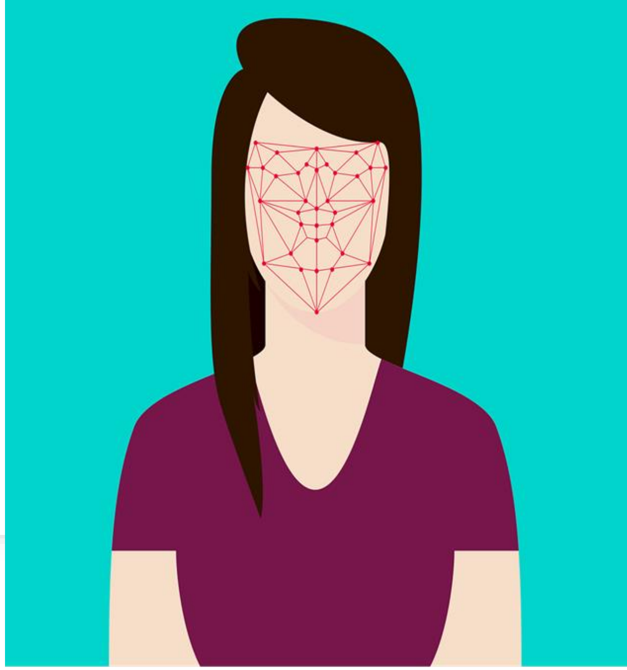
Traditionally, the better the quality of the deepfake, more images required to make video/audio look and sound better

- Takes tens of minutes of videos and hundreds of photos
- Hence the reason by political and celebrities are the main target – today
- Known as passive information

**Samsung is perfecting deepfake s/w to allow them with the use of only 1 photo!

500 images for a perfect deepfake

# Deepfakes are made of…



Generative Adversarial Network (GAN)

- Used for face generation
- It produces faces that otherwise do not exist

GAN uses two separate neural networks — or a set of algorithms designed to recognize patterns

- First, network generates the image
- Second, learns how to distinguish fake from real image

Output = an algorithm that trains itself using the information generated above to generate fake photos of a real person

TOOLBOX™

# TYPES OF DEEPFAKE FRAUDS

Textual Deepfakes

Deepfake Video

Deepfake Audio

Deepfakes on Social Media

Real-time or Live Deepfakes

# Deepfakes also are made of…

Artificial intelligence (AI) algorithm known as encoder/decoder

- Used in face-swapping or face-replacement technology
- First, you run thousands of face shots of two people through the encoder to find similarities between the two images
- Then, a second AI algorithm, or decoder, retrieves the face images and swaps them
  - End result -- a person’s real face can be superimposed on another person’s body

# Who makes such software?

- Open-source Python software
  - Faceswap and DeepFaceLab
    - Faceswap is free, open-source, multi-platform software
    - Runs on Windows, macOS, and Linux
    - DeepFaceLab is an open-source that also enables face-swapping.
  - FakeApp was developed in 2018
  - FaceApp easily downloaded and used – remember my opening picture!!
  - Zao
  - Reface
  - And many more…..

# The Yang…positive, bright, the good

- **Commercial Uses**

  CereProc uses digital voices for people who have lost the use of their voice

  Significant cost & time savings for artificial videos (multiple languages)

- **Creative Uses**

  Nicholas Cage and face-swapping images

  Holocaust survivors talking to an audience using holograms, authors reading their own books

  Fun, entertainment

# The Yin…the negative, dark, the bad

- Porn/Revenge Porn

  Invasion of sexual privacy

- Political campaigns

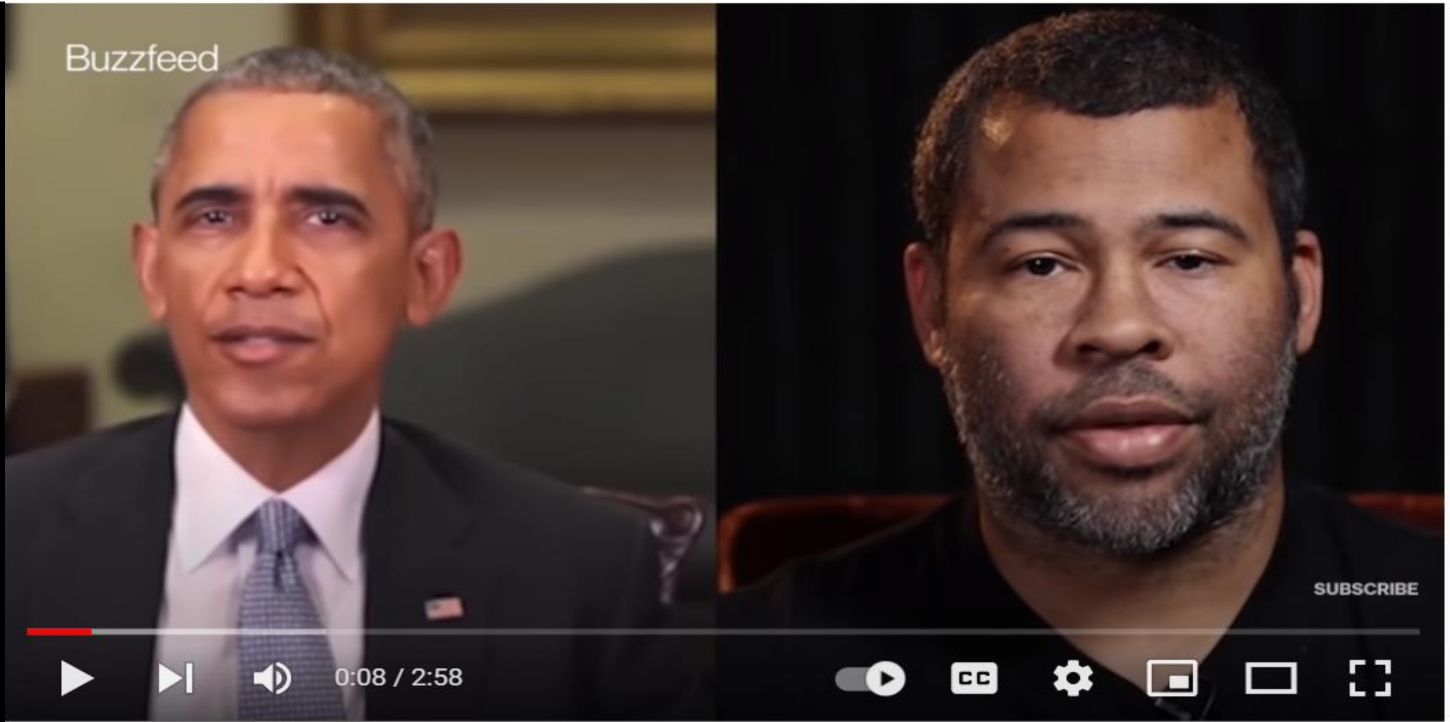
  Launching info warfare campaigns: Example -- Gabon, East Africa

- Potential to create chaos and manipulate public perception makes it a threat, both at individual and societal level

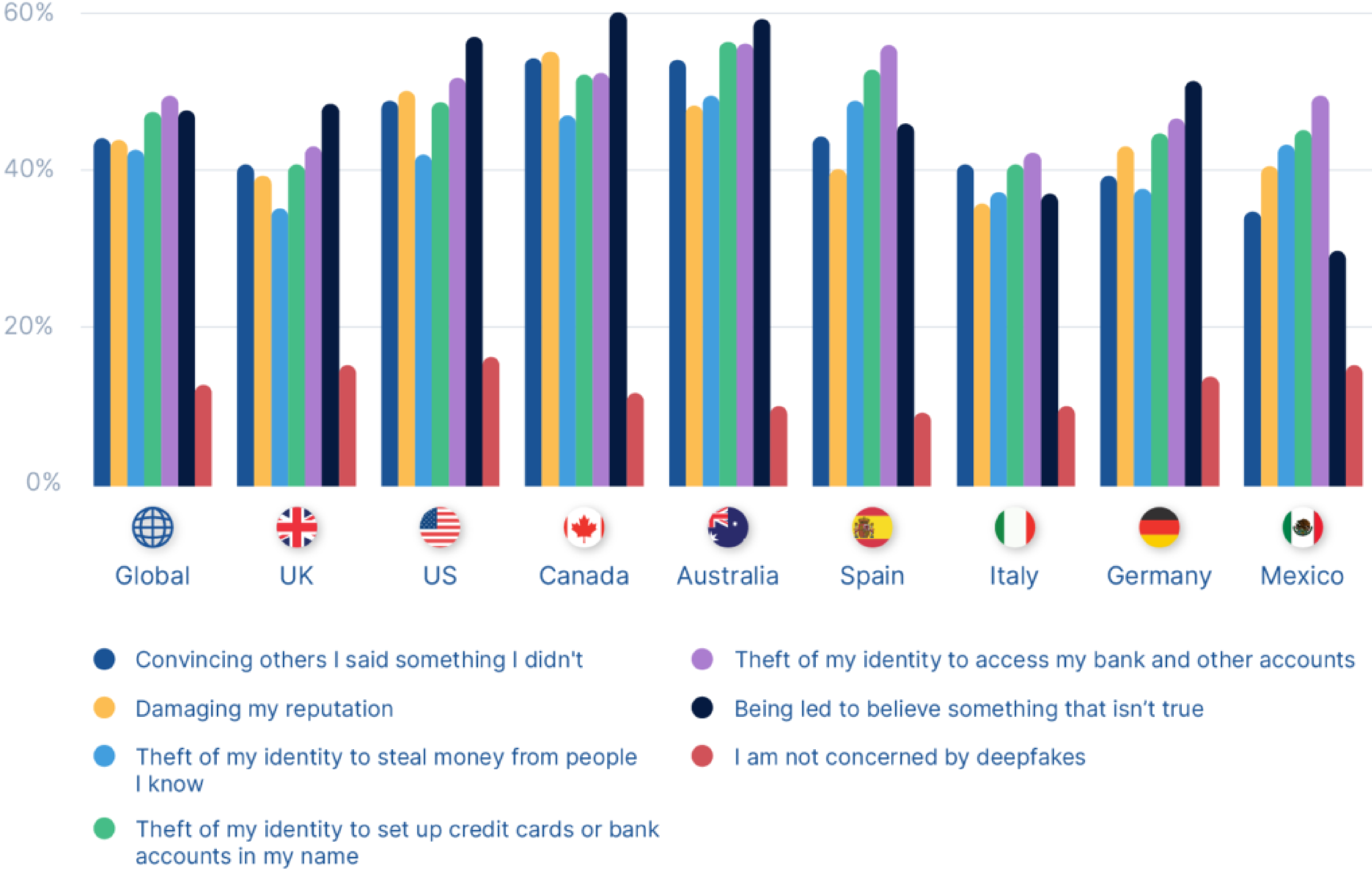
# Examples of Deepfakes

If you watch a Buzzfeed video from 17 April 2018, you will see a video of President Obama making some very outlandish and brazen statements…is it really him?

Scarlett Johansson’s face was transposed on a porn star back in 2017



## Which of the following worries you most about how deepfakes could be used against you? Please select all that apply.



# How can I tell?

Look into their eyes…unnatural eye movement

Unnatural facial expressions.

Awkward-looking facial movement or their body does not look right

Unnatural coloring

Hair and teeth that look fake

Blurring or misalignment

Inconsistent noise or audio

Images that look unnatural when slowed down

Hash discrepancies

Reverse image searches

# Countering Deepfakes

Web browser extensions that help identify a deepfake

Filtering software – Deeptrace provides this type of protection

Social Media Rules

Using soft biometrics to detect

Deepfake Detection Challenges – like Bug Bounties

Research Technologies – using watermarks and blockchain to detect a deepfake

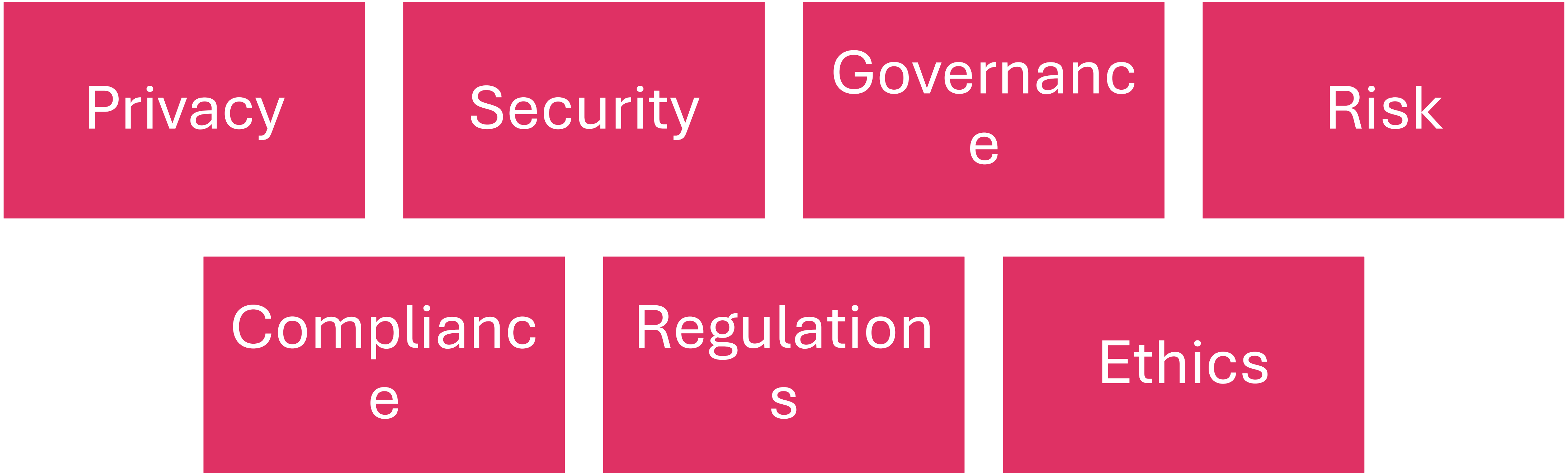
Corporate best practices

Laws and governance

# Break for some fun...

# The Elephant in the room…

# What does this mean for Privacy, Ethics, Security, GRC,  and Regulations?



Privacy

Security

Governance

Risk

Compliance

Regulations

Ethics

**GRC = Governance, Risk, and Compliance**

# Privacy…some yang moments

Minimize the risk of privacy breaches by encrypting personal data

Reduces human error – have another set of “eyes” looking at the work; or having the AI build the documents, frameworks, and a human then oversees the work

Detecting potential cybersecurity incidents – especially since many cybersecurity tools are now using AI/ML to help make efficacy even better!

# Privacy…some yin moments

Data leaks

Intellectual property stolen

Personal Information – at a larger magnitude than even today

**Begs the question…**

*“what ethical or regulatory obligations do organizations have to disclose to customers when AI is doing the work instead of a human?”*

# Some Harvard Professors thoughts on AI/ML

*“Virtually every big company now has multiple AI systems and counts the deployment of AI as integral to their strategy,” said Joseph Fuller, professor of management practice at Harvard Business School,*

*“But we are discovering that many of the algorithms that decide who should get parole, for example, or who should be presented with employment opportunities or housing … replicate and embed the biases that already exist in our society.”*

*“Companies have to think seriously about the ethical dimensions of what they’re doing and we, as democratic citizens, have to educate ourselves about tech and its social and ethical implications — not only to decide what the regulations should be, but also to decide what role we want big tech and social media to play in our lives,”*

# Security…some yang moments

Automating compliance processes for speedy decision-making – in many cases, the AI embedded tool will make some rapid decisions for the security professional

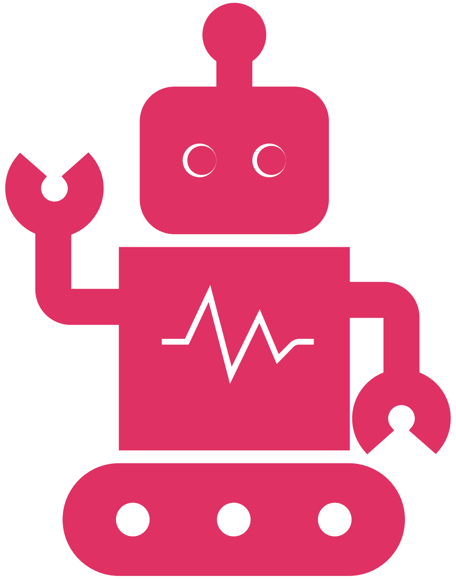
-- F5 Shape

Ensuring faster and secure transactions

-- Can learn the fastest, most reliable paths and push traffic that way

Easy monitoring of regulatory change management – can scan through the network to see what is happening and if there are any holes or vulnerabilities

-- PCI DSS or HIPAA



# Security...some yin moments

Phishing scams

Data breaches

Hoaxes

Pornography

Reputation smearing

Election manipulation

Social engineering

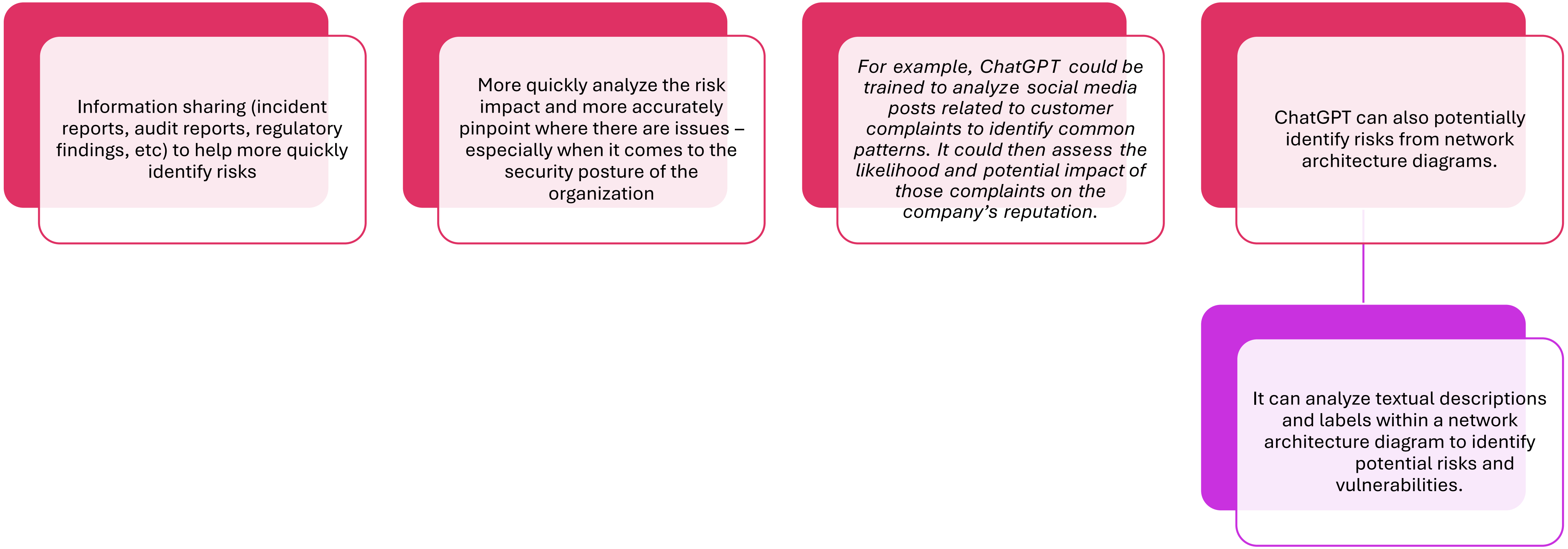
Automated disinformation attacks

Identity theft

Financial fraud

Blackmail

# Risk – some yang moments



Information sharing (incident reports, audit reports, regulatory findings, etc) to help more quickly identify risks

More quickly analyze the risk impact and more accurately pinpoint where there are issues – especially when it comes to the security posture of the organization

*For example, ChatGPT could be trained to analyze social media posts related to customer complaints to identify common patterns. It could then assess the likelihood and potential impact of those complaints on the company's reputation.*

ChatGPT can also potentially identify risks from network architecture diagrams.

It can analyze textual descriptions and labels within a network architecture diagram to identify potential risks and vulnerabilities.

# Risk…some yin moments

- Generative AI holds an intrinsic capacity for potential inaccuracies
- Information misuse – take risks on misinformation…not good…
- May retain inherent bias
- More data in, more data out, more data messes

# Governance…some yang moments

Governance is very underdeveloped in AI

- Speed and scale of adoption of AI threatens to outpace the regulatory responses to address the concerns raised

Responsibility and liability for harms resulting from the use of AI applications remain ambiguous under many legal frameworks

Automation of routine and manual tasks are expected to displace millions of jobs that will not be evenly distributed within and across countries

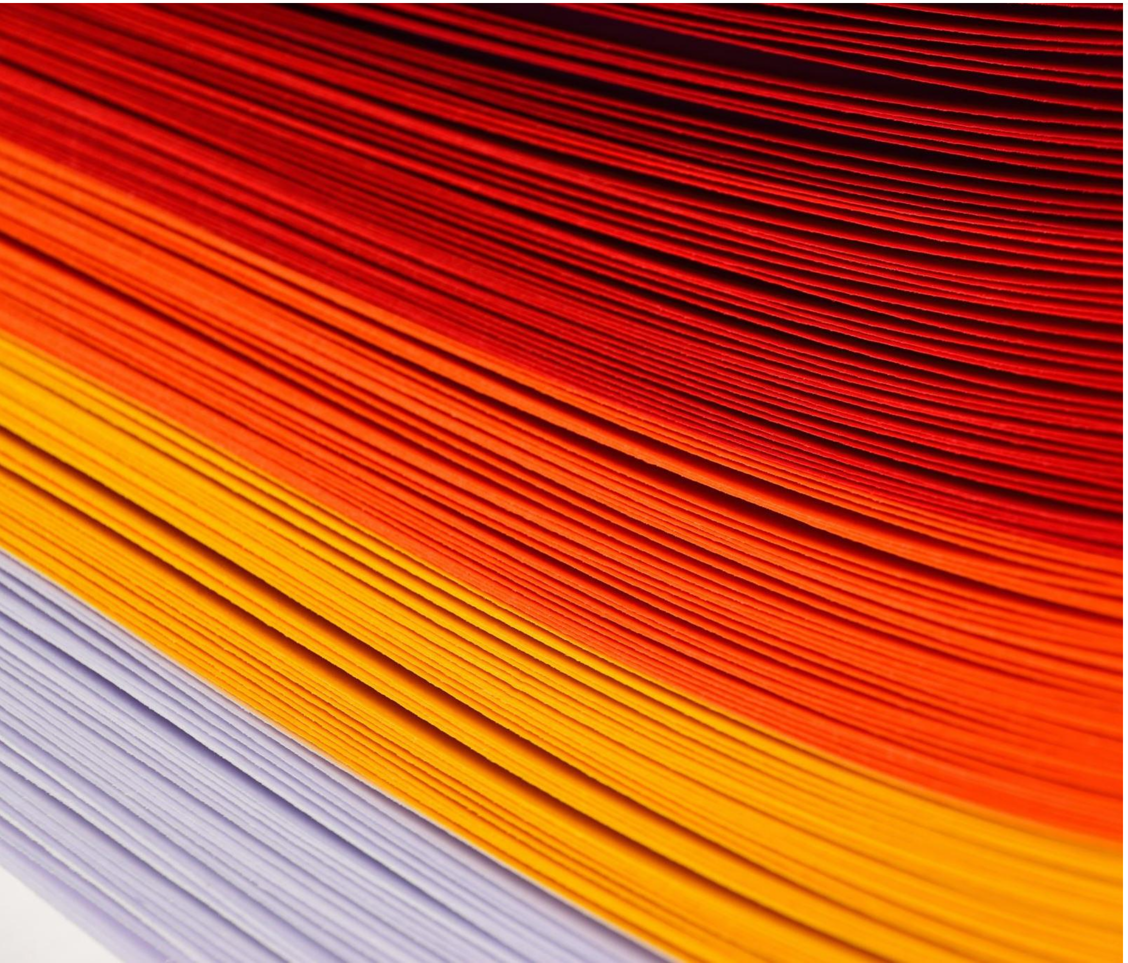
# Governance…some yin moments

Begs many questions on protecting the IP from being plagiarized, stolen or used without permission by nefarious AI tools

Governing bodies will need to implement protections to ensure their organization is not inadvertently infringing on the IP of others

Legal teams and business leaders will need to work together on putting together the right amount of governance for the organization – especially in light that there really is not too much out there yet…

# Regulation…some yang moments



Crafting regulations and governing documents can be mundane and very tedious – typically put off until the last minute or not until something bad happens…

GRC analysts can use ChatGPT to generate draft policy or procedure documents by supplying basic information and guidelines.

Uses natural language processing capabilities to create a coherent, well-structured document that meets the company's GRC management requirements

# Regulation…some yin moments

Currently, the laws that govern AI use and ethics implications are limited

Europe has shown to lead on this front, with the impending AI Act

--aims to implement certain protections and limitations around the use of advanced technology and reinforce privacy rights within those applications.

US’  “AI Bill of Rights” is in the works

# AI Bill of Rights - Proposed

Human alternatives, consideration, and fallback

Notice and explanation

Data Privacy

Algorithmic Discrimination Protections

Safe and Effective Systems

# Governance & Regulations

National Defense Authorization Act of 2020

Privacy Act of 1974

Copyright Laws and Intellectual Property

Fair Use Doctrine

Communications Decency Act

State Laws – Virginia, Texas, and California

Photoshop law in Israel

Cyberstalking Law

General Data Protection Regulation (GDPR) – EU

AI Law - EU

# Artificial Intelligence Act

Artificial Intelligence Act being implemented in the EU

Breaks the risk down into three categories:

- Applications and systems that create an **unacceptable risk**, such as government-run social scoring of the type used in China, are banned.
- **High-risk applications**, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements.
- Applications not explicitly banned or listed as high-risk are largely left unregulated.

# Compliance – some yang moments

**Can be helpful in ensuring the systems, processes, and other areas of the business are compliant**

Think of vulnerability scanning; quickly finding changes in the network

**Support Third-Party Assessment Program**

Provides guidance on assessment criteria and assist third-party assessors in understanding specific assessment requirements

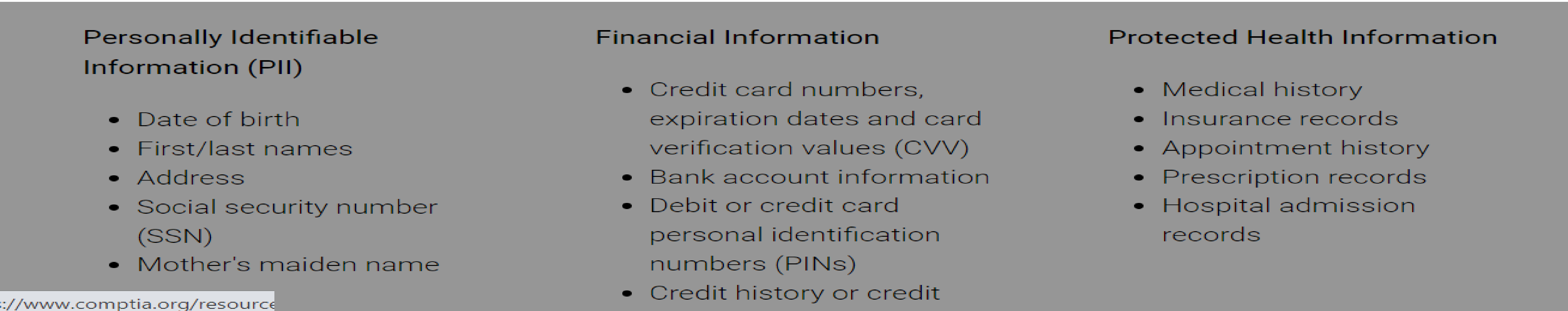
Provides information on industry best practices for security and compliance, including frameworks such as SOC 2, Payment Card Industry Data Security Standard (PCI DSS) and Health Insurance Portability and Accountability Act (HIPAA)

Provides ongoing education and training for third-party assessors

# Compliance – some yin moments

Compliance information that is not accurate and can cause issues with regulators or other officials in industry

Sensitive data can be leaked and cause compliance gaps



**Personally Identifiable Information (PII)**

- Date of birth
- First/last names
- Address
- Social security number (SSN)
- Mother's maiden name

**Financial Information**

- Credit card numbers, expiration dates and card verification values (CVV)
- Bank account information
- Debit or credit card personal identification numbers (PINs)
- Credit history or credit

**Protected Health Information**

- Medical history
- Insurance records
- Appointment history
- Prescription records
- Hospital admission records

://www.comptia.org/resource

# Parting thoughts…

There is always a balance between good and evil when it comes to technological advancements

It is not all bad and…

Implement oversight and supervision of AI tools

Identify use cases already in flight, as certain areas of the business may already be leveraging AI technology that needs to be validated.

Be prepared for regulatory transparency and compliance, because local and global legislation will require it at some point in time

Be prepared to prevent, detect, mitigate and defend against various personal and commercial claims of harm.

# Questions?

Deborah.wells@cwu.edu

# References

https://www.youtube.com/watch?v=WAiqNav2cRE

Https://www.youtube.com/watch?v=yAgQWnD31nE

https://www.youtube.com/watch?v=L_Guz73e6fw

https://medium.com/predict/chatgpt-and-its-impact-on-criminal-forensics-and-criminals-1e4f0a57754d

https://www.law.com/legaltechnews/2023/03/16/with-chat-gpt-will-legal-and-compliance-become-more-intelligent-more-complicated-or-both/?slreturn=20230421145950

https://securityintelligence.com/posts/using-chatgpt-as-an-enabler-for-risk-and-compliance/

http://csunplugged.mines.edu/activity-AI.html https://www.iproov.com/blog/deepfakes-statistics-solutions-biometric-protection#:~:text=Summary%3A,the%20deepfake%20threat%20is%20growing.

https://www.purevpn.com/blog/deepfake-apps/

https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

https://artificialintelligenceact.eu/