

This presentation may contain simulated phishing attacks.

The trade names/trademarks of third parties used in this presentation are solely for illustrative and educational purposes.

The marks are property of their respective owners, and the use or display of the marks does not imply any affiliation with, endorsement by, or association of any kind between such third parties and KnowBe4.

Cybercriminals don't care about this and use them anyway to trick you....

This presentation, and the following written materials, contain KnowBe4's proprietary and confidential information and is not to be published, duplicated, or distributed to any third party without KnowBe4's prior written consent. Certain information in this presentation may contain "forward-looking statements" under applicable securities laws. Such statements in this presentation often contain words such as "expect," "anticipate," "intend," "plan," "believe," "will," "estimate," "forecast," "target," or "range" and are merely speculative. Attendees are cautioned not to place undue reliance on such forward-looking statements to reach conclusions or make any investment decisions. Information in this presentation speaks only as of the date that it was prepared and may become incomplete or out of date; KnowBe4 makes no commitment to update such information. This presentation is for educational purposes only and should not be relied upon for any other use.

KnowBe4

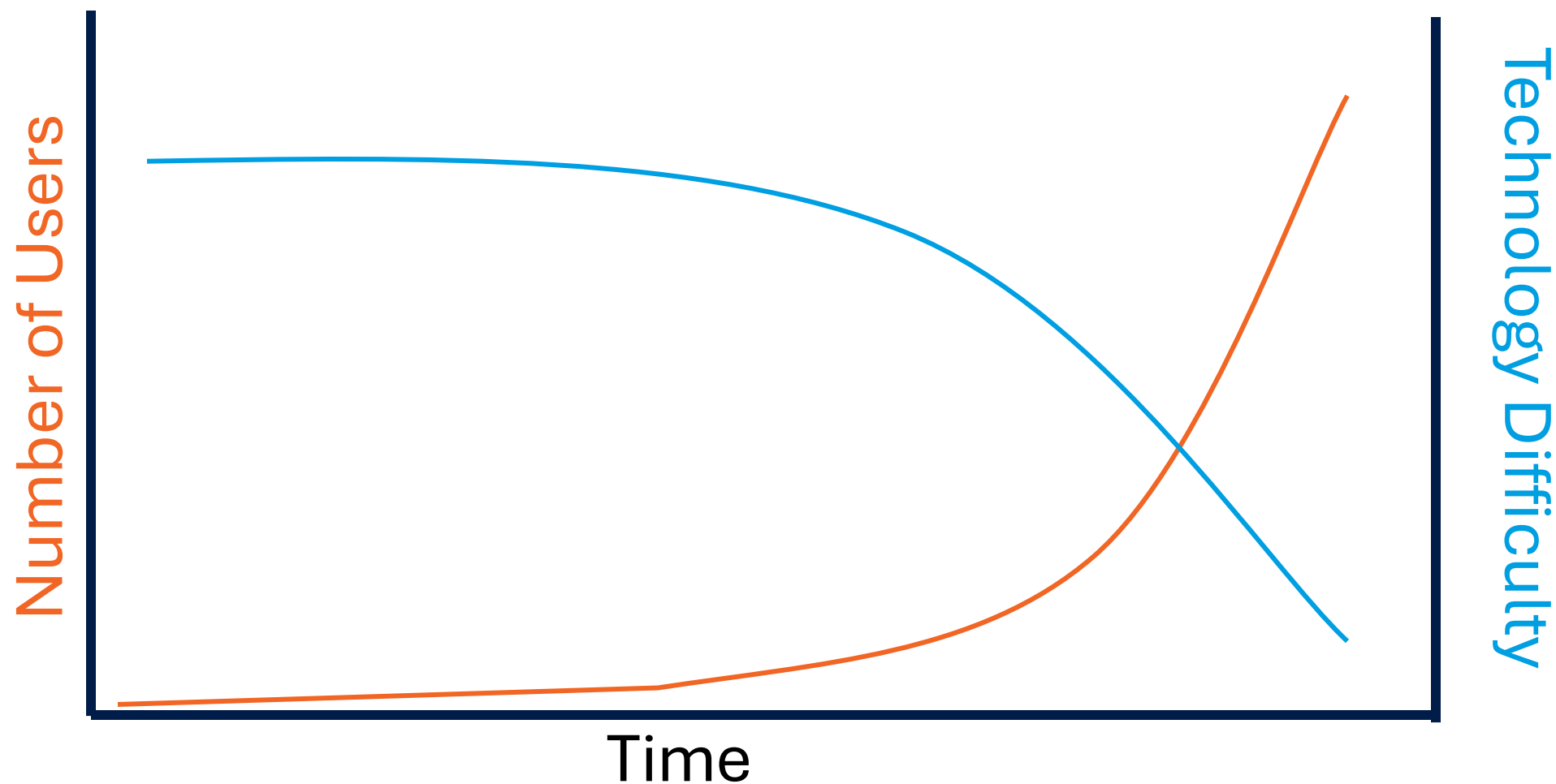
AI vs AI

Transforming Cybersecurity Through
Proactive Technologies

James R. McQuiggan, CISSP, SACP
Security Awareness Advocate



AI, Generative AI and AI Agents... Oh My!



Cybercriminals are leveraging all
aspects of AI to **attack**
organizations.

James R. McQuiggan, CISSP, SACP, OSC

Security Awareness Advocate, KnowBe4 Inc.

Producer, Security Masterminds Podcast

Professor, Cyber Threat Intelligence, Full Sail

President, ISC2 Central Florida Chapter

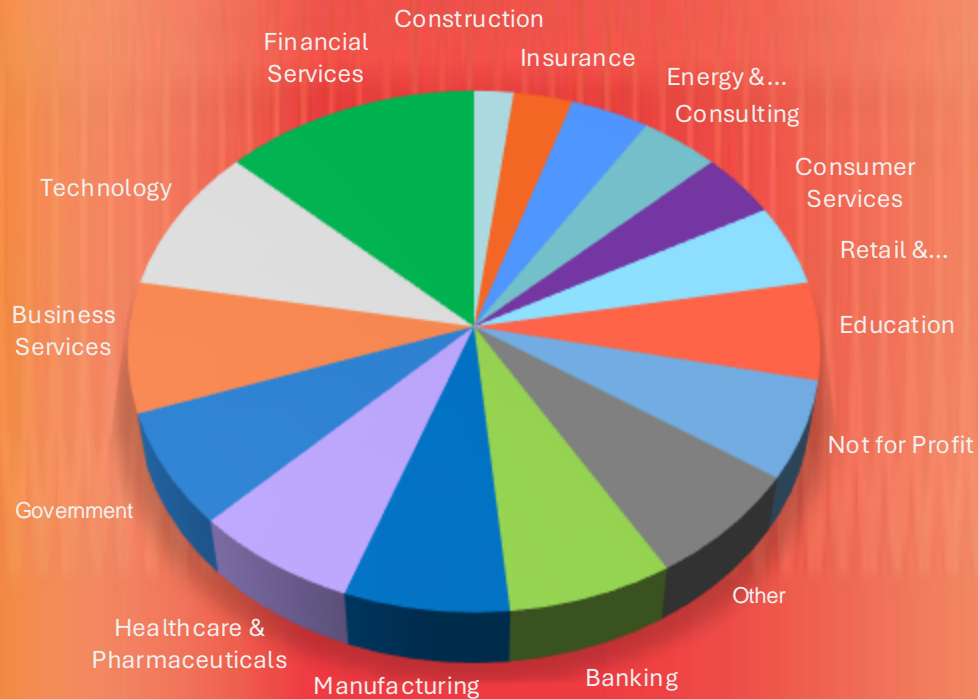
ISC2 North American Advisory Council

Cyber Security Awareness Lead, Siemens

Product Security Officer, Siemens Gamesa



Over
70,000
Customers



About KnowBe4

- The world's largest integrated Security Awareness Training and Simulated Phishing platform
- We help tens of thousands of organizations manage the ongoing problem of social engineering
- CEO & employees are industry veterans in IT Security
- Global Sales, Courseware Development, Customer Success, and Technical Support teams worldwide
- Offices in the USA, UK, Netherlands, India, Germany, South Africa, United Arab Emirates, Singapore, Japan, Australia, and Brazil

Our mission

To help organizations manage the ongoing problem of social engineering

We do this by

Empower your workforce to make smarter security decisions every day.

Outcomes for the next 97 minutes... (and 257 slides)

AI is available
to all, and
cyber
criminals are
using for their
needs

AI Agents are
leveraging Gen
AI for
efficiency

How can we
use AI Agents
to protect
users and
organizations?



AI Attack Vectors

Examining the various attack vectors used by cybercriminals

KnowBe4

© 2024 KnowBe4, Inc. All rights reserved. 10



AI vs AI

Real World Examples of how AI is used to defend and protect organizations

KnowBe4

© 2024 KnowBe4, Inc. All rights reserved. 11



AI vs AI Protections

Looking at how AI can protect and defend our organizations.

KnowBe4

© 2024 KnowBe4, Inc. All rights reserved. 12

KnowBe4

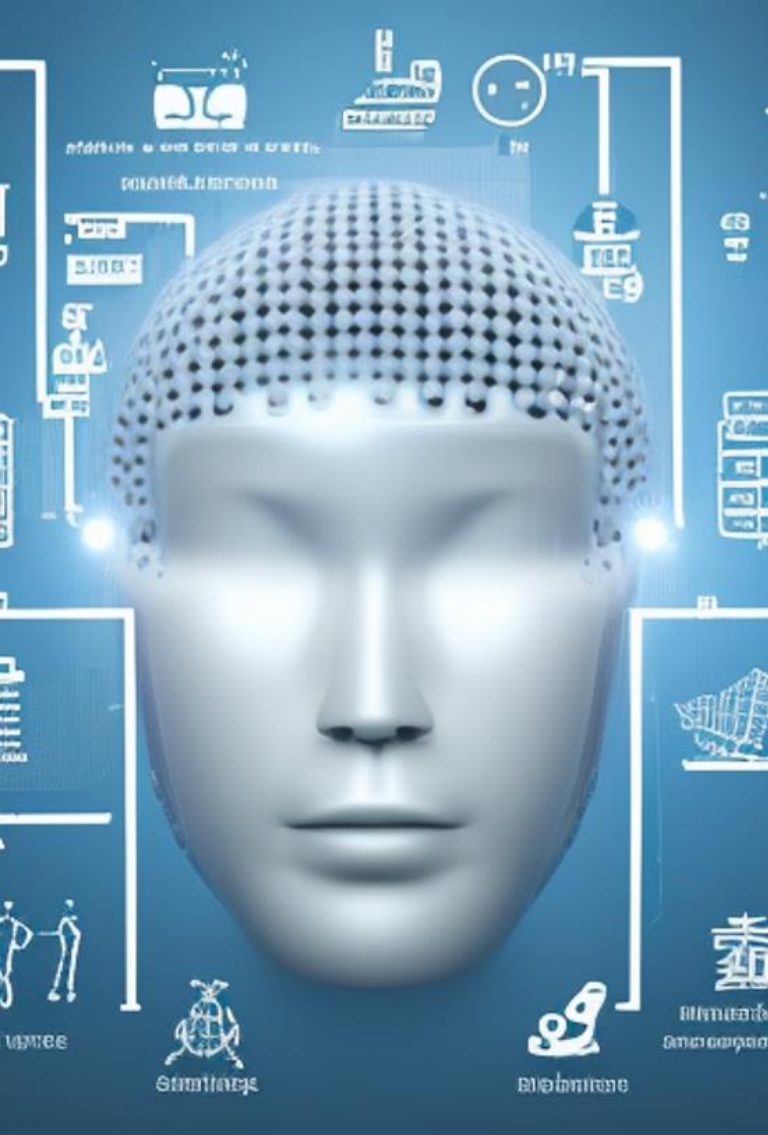
Wrap-Up / Q&A



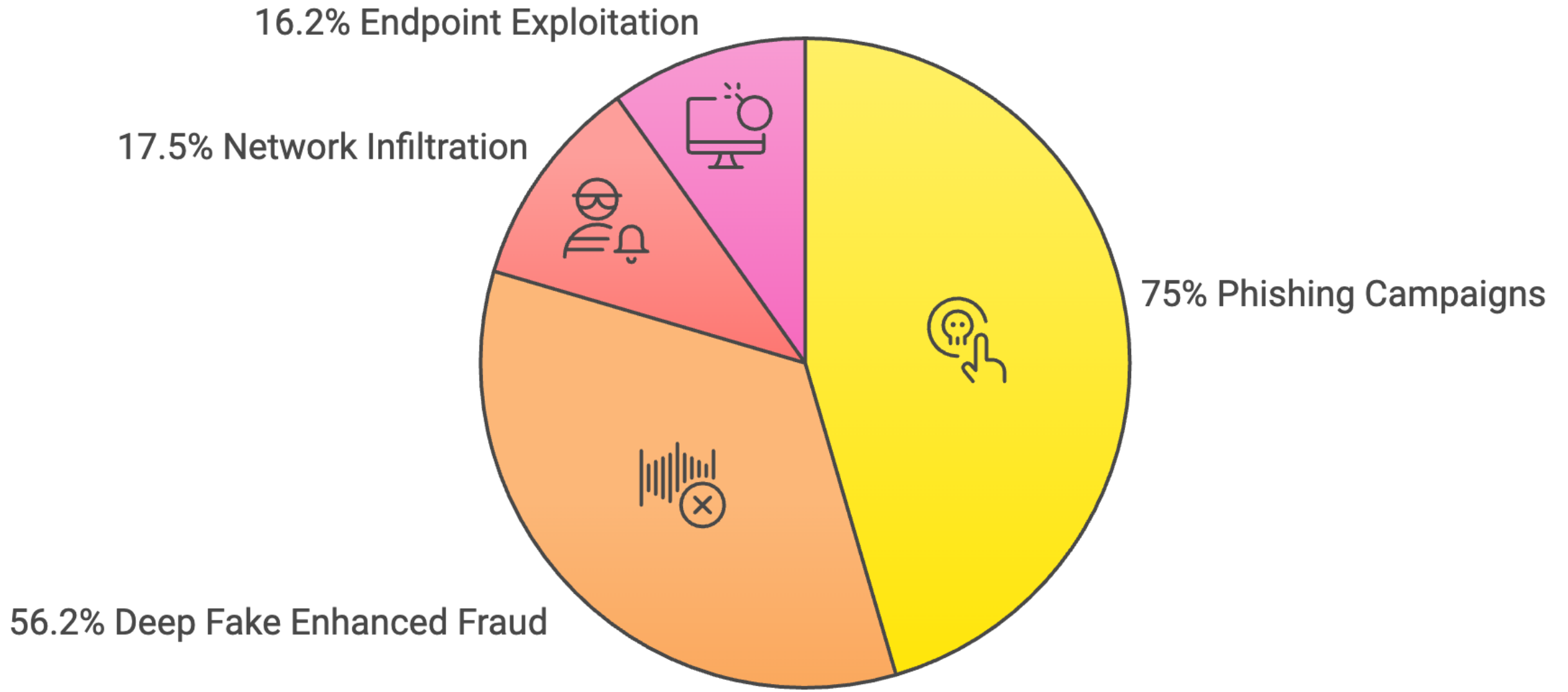
© 2024 KnowBe4, Inc. All rights reserved. 13

AI Attack Vectors

Examining the various attack vectors used by cybercriminals

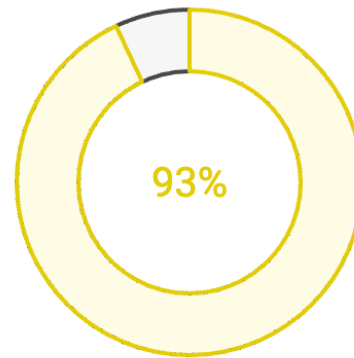


CISO Survey – AI Threats – Team 8

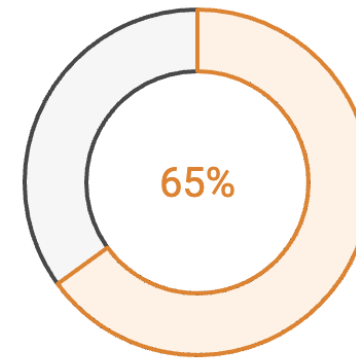




Perceptions of AI Threats in Cybersecurity



Daily AI Attacks



Offensive AI Norm

AI is Supercharging Phishing Attacks & Social Engineering

Spear Phishing Personalization

Tailoring phishing attacks to individuals using AI insights.

Phishing Site Automation

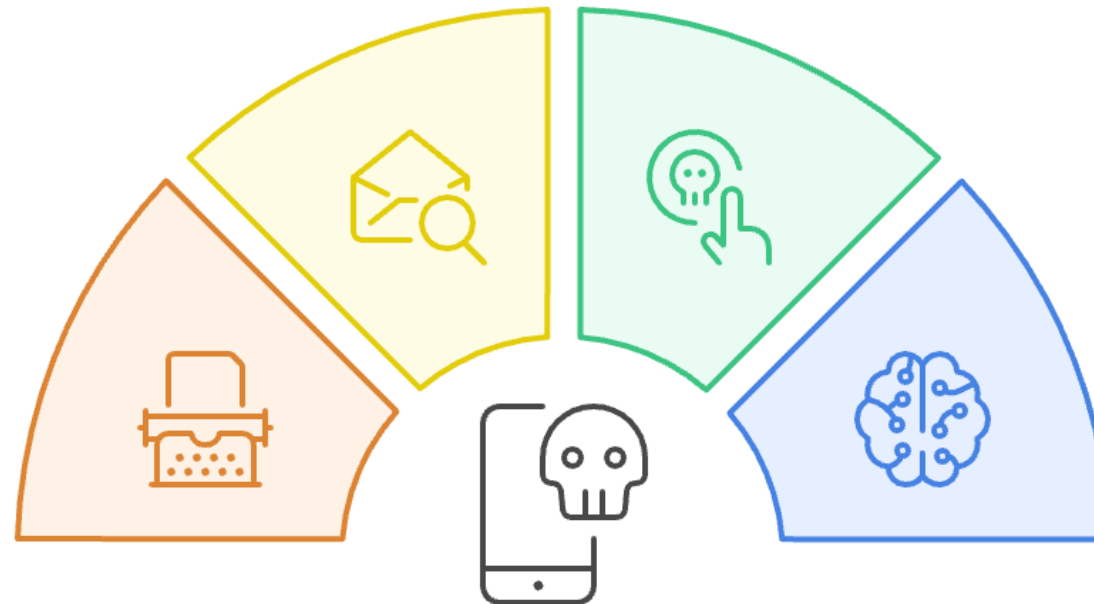
Automating the creation of fraudulent websites to deceive users.

Automated Content Generation

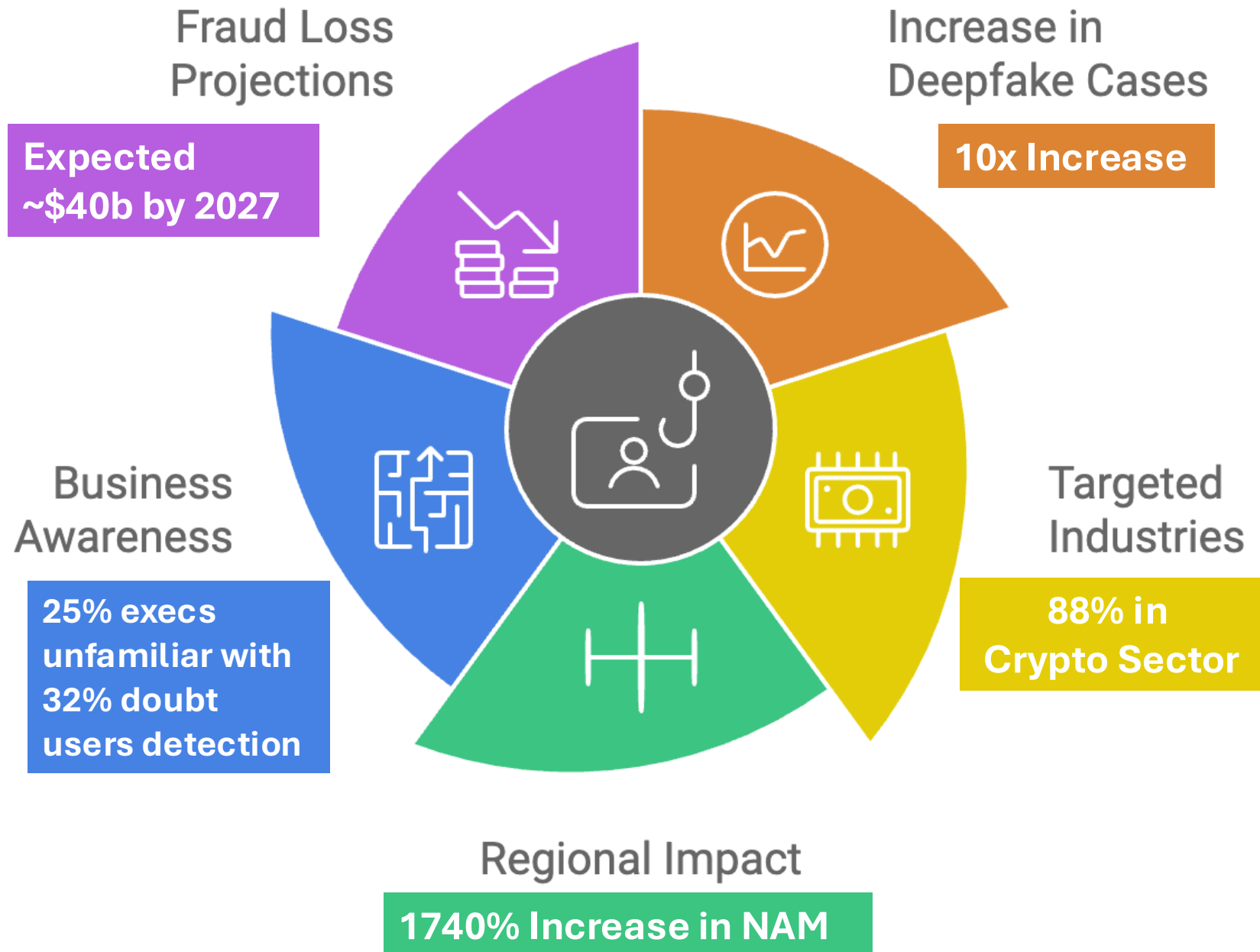
The use of AI to create deceptive content at scale.

Deepfake Technology

Creating realistic but fake videos to mislead viewers.



Deepfake Scams



HONG KONG

Multinational loses HK\$200 million to deepfake video conference scam, Hong Kong police say

Police received a report of the incident on January 29, at which point some HK\$200 million (US\$26 million) had already been lost via 15 transfers.



NEWS SERVICE by AFP

12:09, 5 FEBRUARY 2024



Why you can trust Hong Kong Free Press



Listen to this article



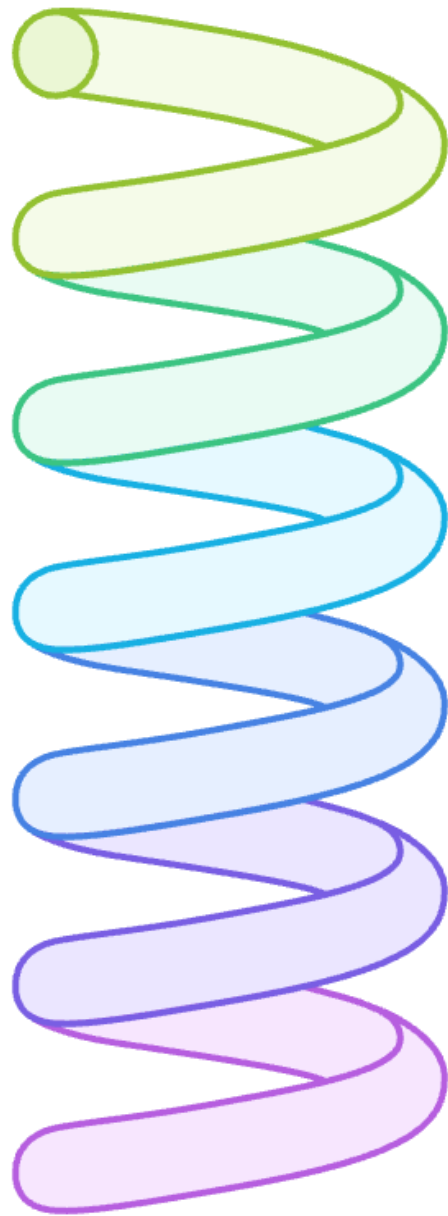
Scammers tricked a multinational firm out of some US\$26 million by impersonating senior executives using deepfake technology, Hong Kong police said Sunday, in one of the first cases of its kind in the city.

HKFP

Support the HKFP team
as a monthly Patron.



AI & Phishing



Phishing attacks increase by 28%



44% of attacks originate from compromised accounts



45% of phishing emails contain malicious hyperlinks



AI integrated into phishing toolkits



75% of kits offer AI features



82% of kits include deepfake capabilities

Source: <https://www.scworld.com/news/phishing-attacks-armed-with-ai-capabilities-are-on-the-rise>

AI-Created Deepfakes Used In Attempt Theft

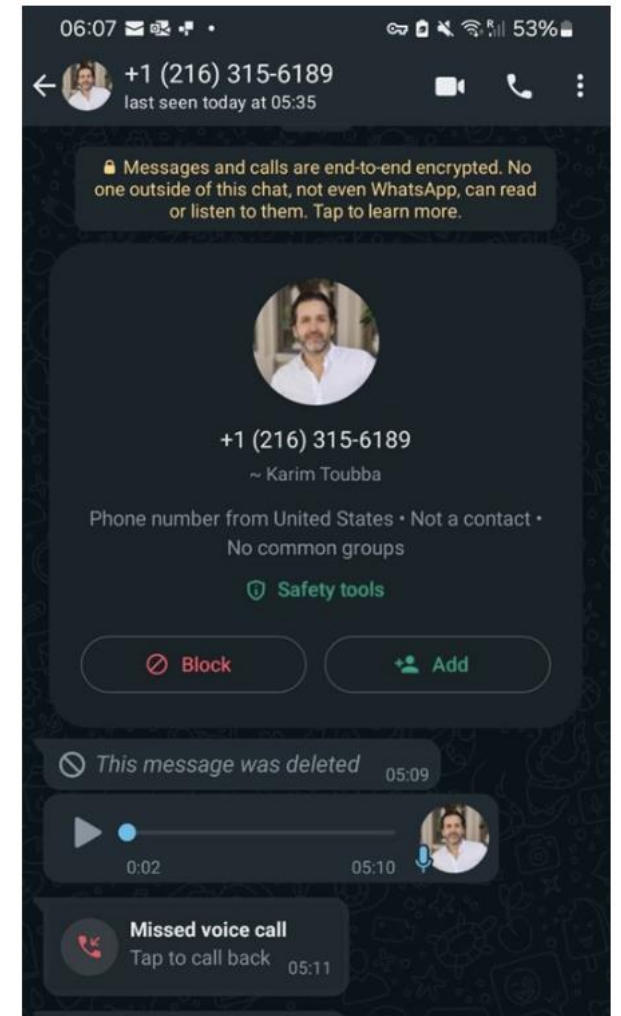
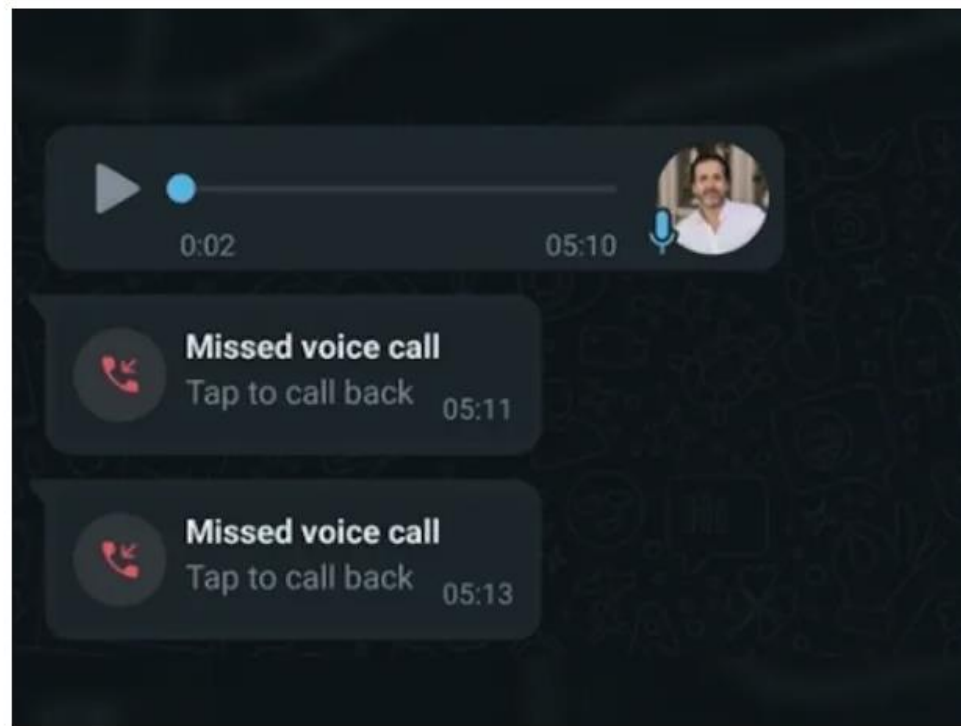
Audio Deepfake Attacks: Widespread And 'Only Going To Get Worse'

BY KYLE ALSPACH ▶

OCTOBER 3, 2024, 11:23 AM EDT

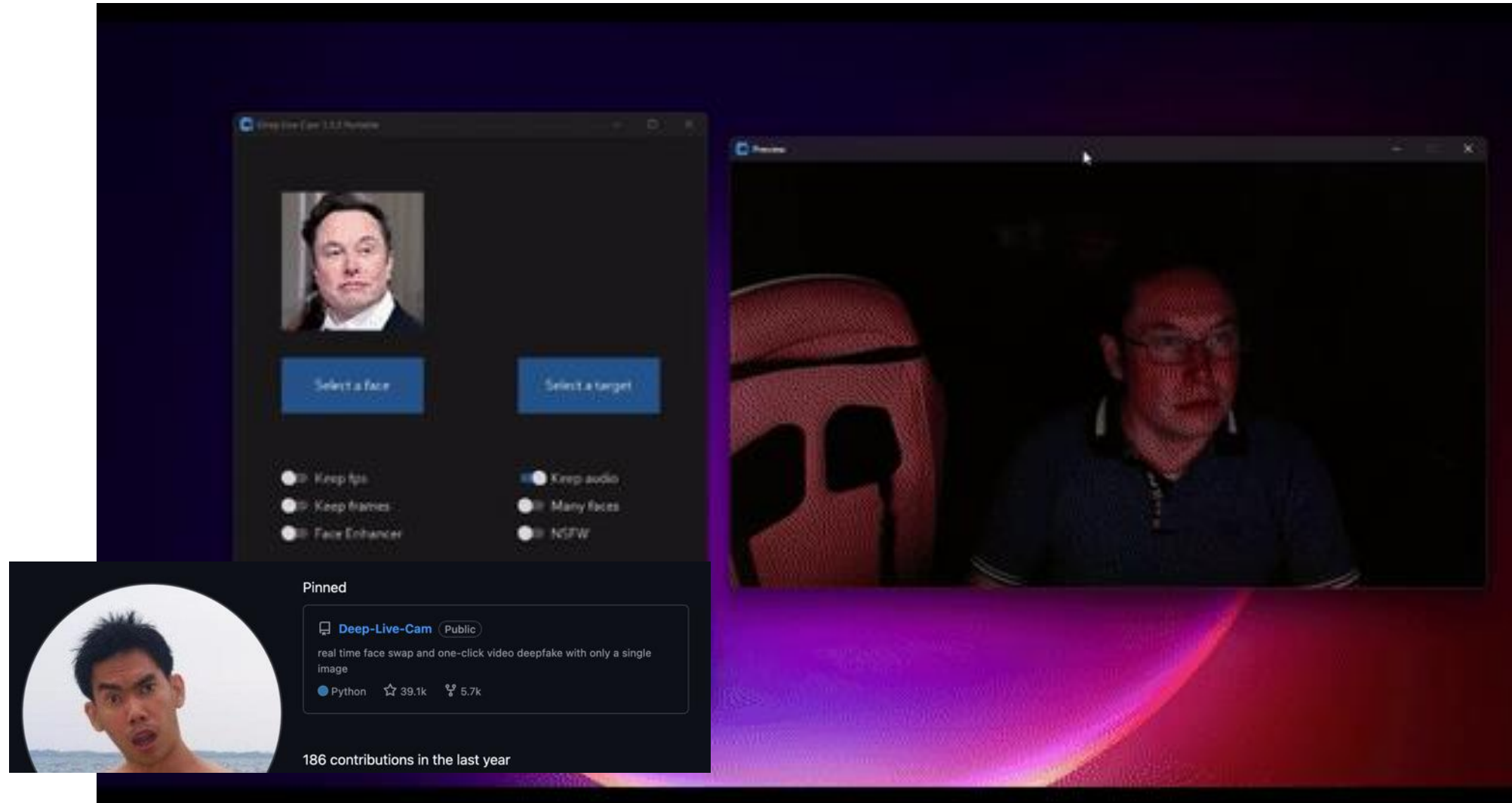
A cybersecurity researcher tells CRN that his own family was recently targeted with a convincing voice-clone scam.

Fake
Calls



<https://blog.lastpass.com/posts/2024/04/attempted-audio-deepfake-call-targets-lastpass-employee>


Deepfake & Webcams - LIVE



Dark Web Activity

14 June

Black Market © 239 edited 12:35



Black Market © Plus Plan

🔥 New Released

- ✅ The most advanced deep fake video impersonation application using the latest DeepFake AI technology.
- ✅ Supported on Windows machine with GPU and minimum 8GB RAM.
- ✅ Simply upload any person photo and let the DeepFake AI make it live with enhanced 3D dimensions following your text scripts expressions, movements and voice for the high resolution video generation.
- ✅ Best for generating your own fake / clone video statement and conference telling about anything based on your text scripts with your own preferred voice cloning module.
- ✅ The new era of video spoofing, love scamming and false statement spreading.
- ✅ Unlimited high resolution deep fake video generations.

Bundle Package Fee:
Lifetime = 🇺🇸 USD160 / 🌐 USDT160

00:00 LIVE

Avatar AI
VideoCallSpoofer




Black Market © Plus Plan

🔥 Hot Selling

- ✅ The latest AI technology of video call spoofer tool.
- ✅ Simply upload any person photo and let the Avatar AI make it live following your expressions and movements for the video call session.
- ✅ Supported for most of video call applications and platforms (Whatsapp, Telegram, Google Meet, Zoom, Microsoft Teams and many more)
- ✅ Supported for all Windows / Mac PC & Laptop machines.
- ✅ Supported for all iOS / Android smartphones.
- ✅ Remote installation and setup services included.

Bundle Package Fee:
Lifetime = 🇺🇸 USD200 / 🌐 USDT200

Black Market © 1,1K edited 09:28



Black Market © Premium Plan

🔥 Hot Selling

- ✅ The most advanced deep fake video impersonation tool using well known DeepFake AI model.
- ✅ Simply upload any person photo and let the DeepFake AI make it live following your expressions, movements and voice for the high resolution video generation.
- ✅ Best for generating your own fake video statement and conference telling about anything that you want.
- ✅ The new era of video spoofing, love scamming and false statement spreading.
- ✅ Unlimited high resolution deep fake video generation.

Subscription Fee:
1 month = 🇺🇸 USD60 / 🌐 USDT60
3 months = 🇺🇸 USD150 / 🌐 USDT150
6 months = 🇺🇸 USD250 / 🌐 USDT250
Lifetime = 🇺🇸 USD400 / 🌐 USDT400

Source: <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/surging-hype-an-update-on-the-rising-abuse-of-genai>



AI Agents

Real World Examples of how AI is used to defend and protect organizations

2023 was the year for GEN AI
2024 was the year for Gen AI +
2025 is the year for AI AGENTS

What the internet was to AI,
Social Media is to AI Agents



AI Agents

AI Agent Automation

Autonomous decision-making to achieve goals

Every day at sunset write a recap email on the cyber stories of the day.

Rule-Based Automation

Following a set list of steps, including language model calls

Every day at sunset write a social media post about AI Agents and Human Risk Management.

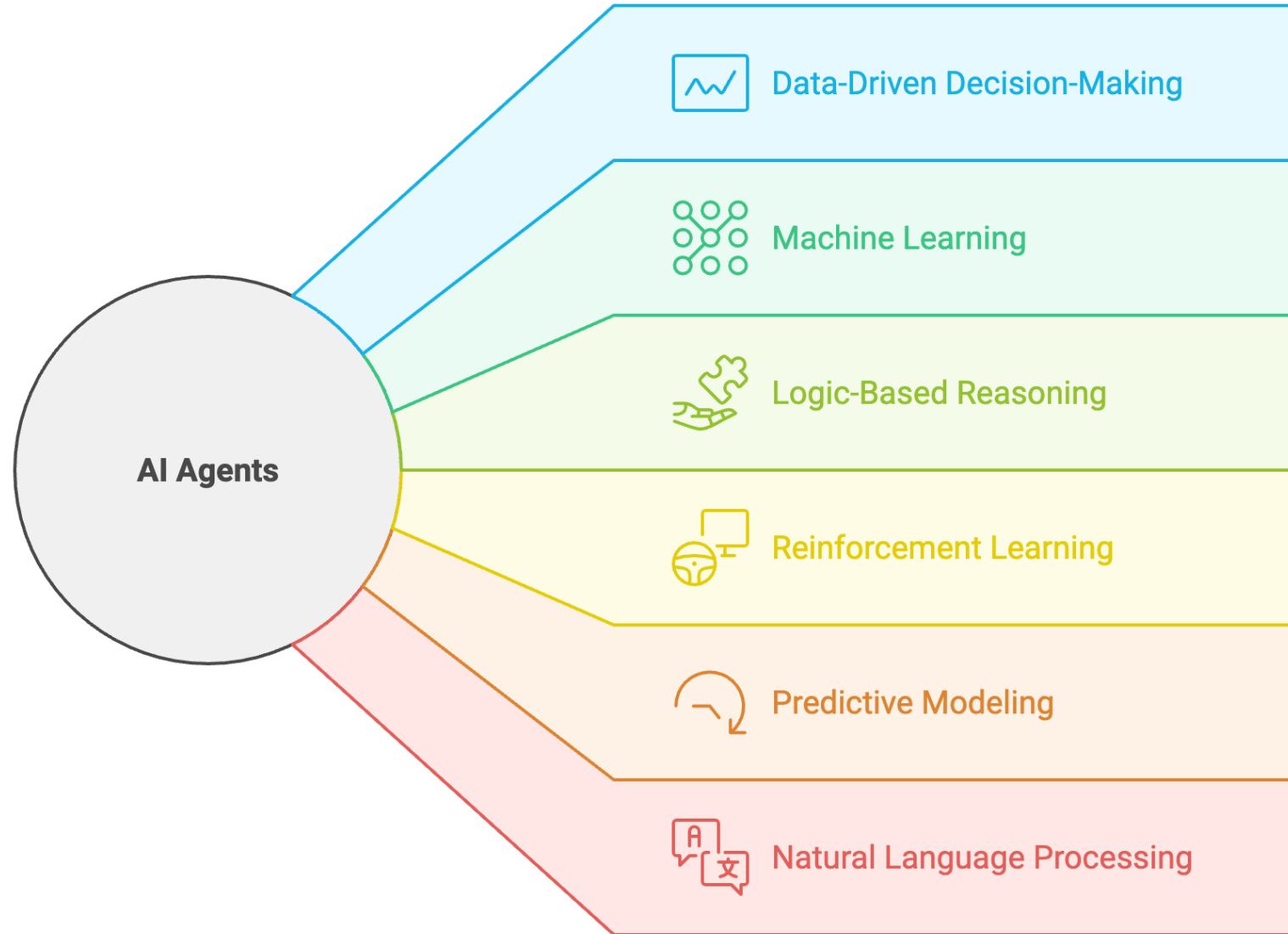
Basic Automation

Simple conditional actions like "If X happens, do Y"

Every day at sunset turn on the lights



AI Agents



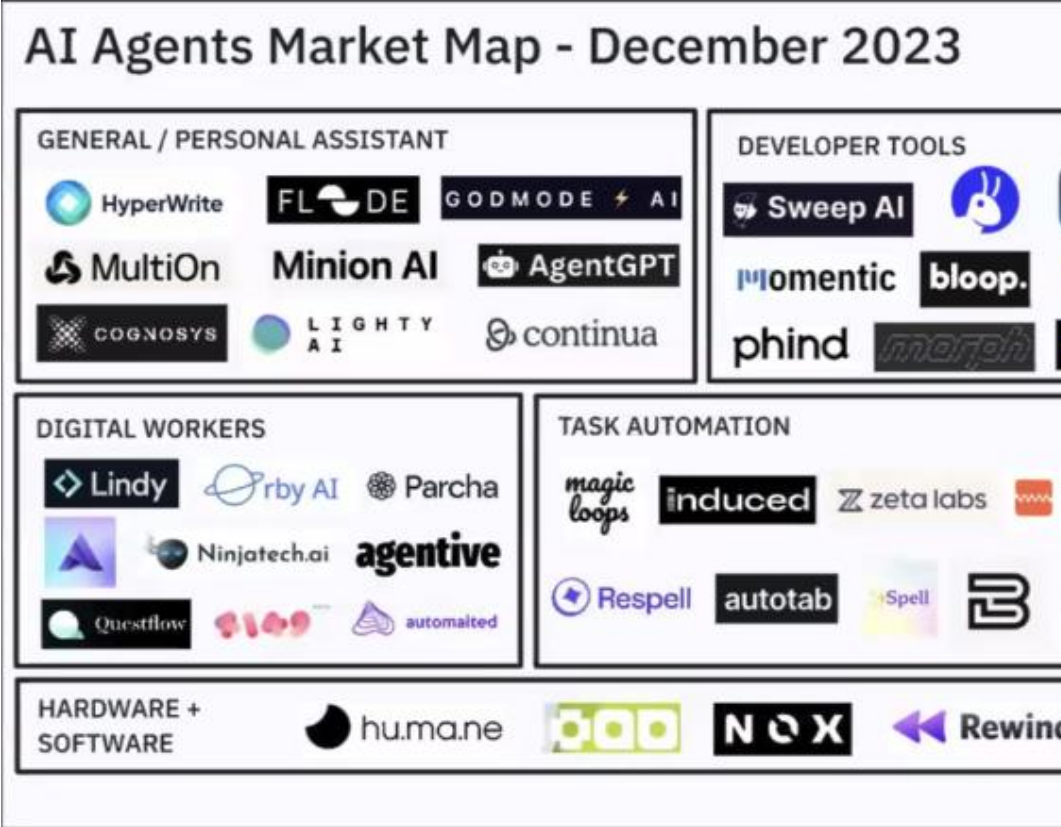
AI Agent Examples

Remember: some people will call the left one an AI agent too, but you can quickly see why these two columns would be treated differently.

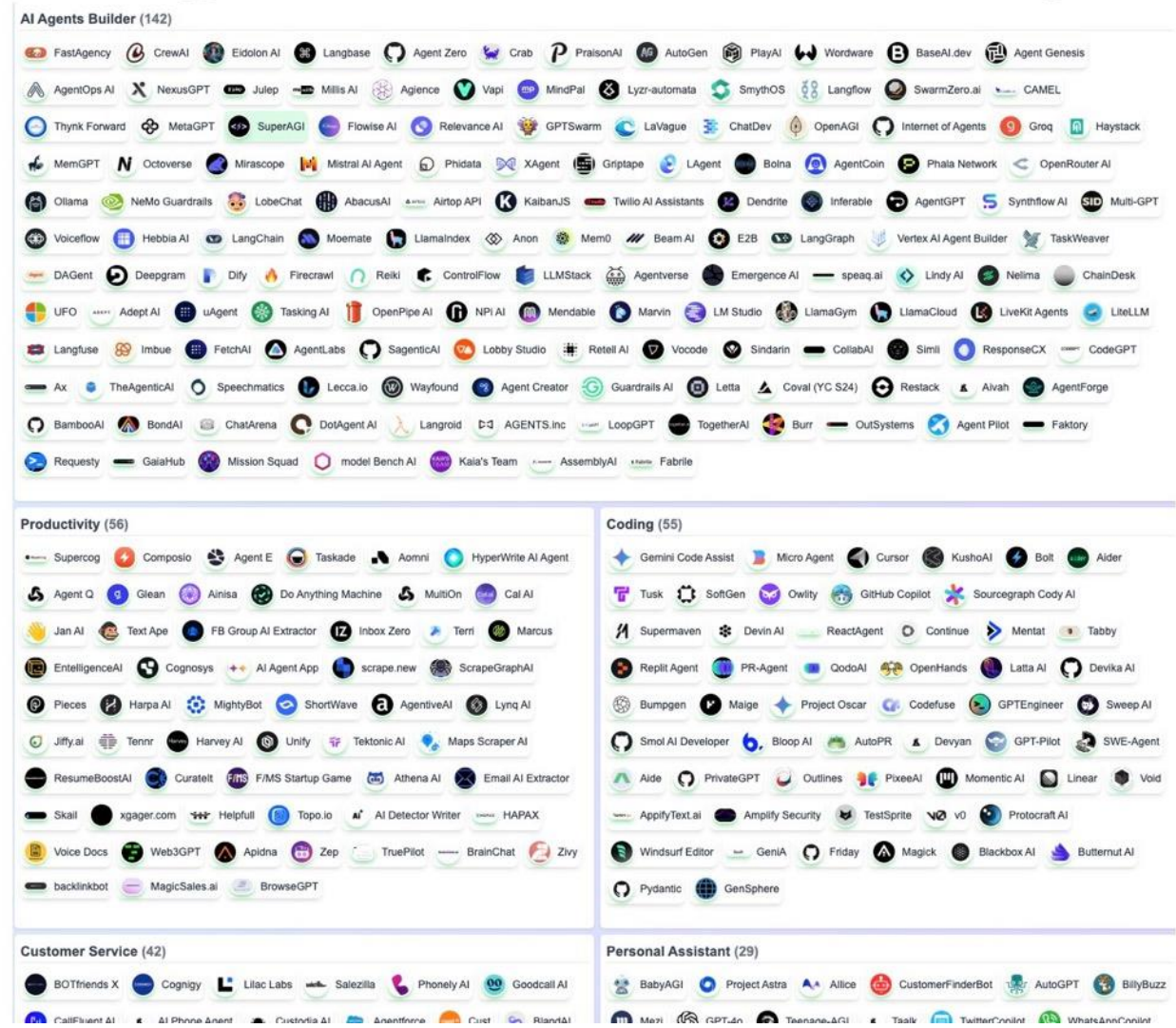
AI AUTOMATIONS	AI AGENTS
Follows a recipe	Writes the recipe
You define every step	Defines its own path
Executes instructions	Creates strategies to achieve goals
Coffee maker	Barista
Zapier and Make.com with Claude 3.5 Sonnet	Crew AI, Claude Computer Use
Valuable	Valuable
Reliable order/progression of steps	May have lower standardization
Task-specific	Multi-tasking/general purpose

Source: Introduction to AI Agents, Allie K Miller (Maven)

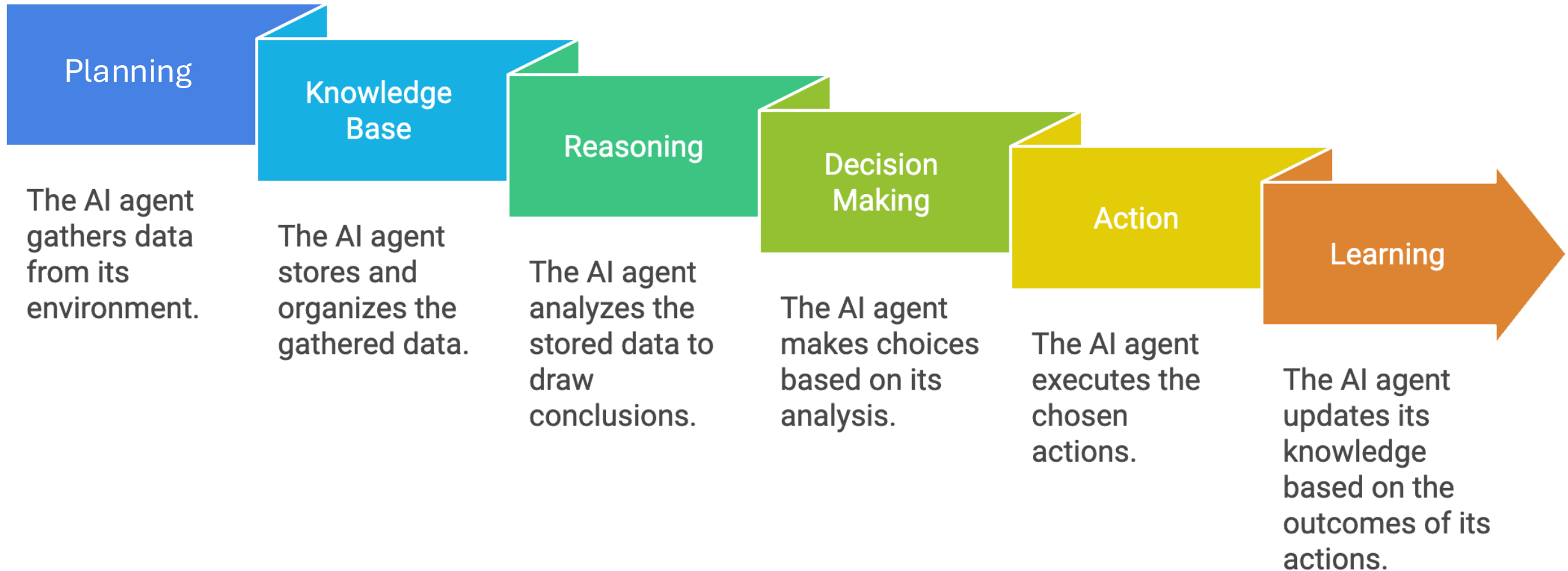
AI Agent – 1 year



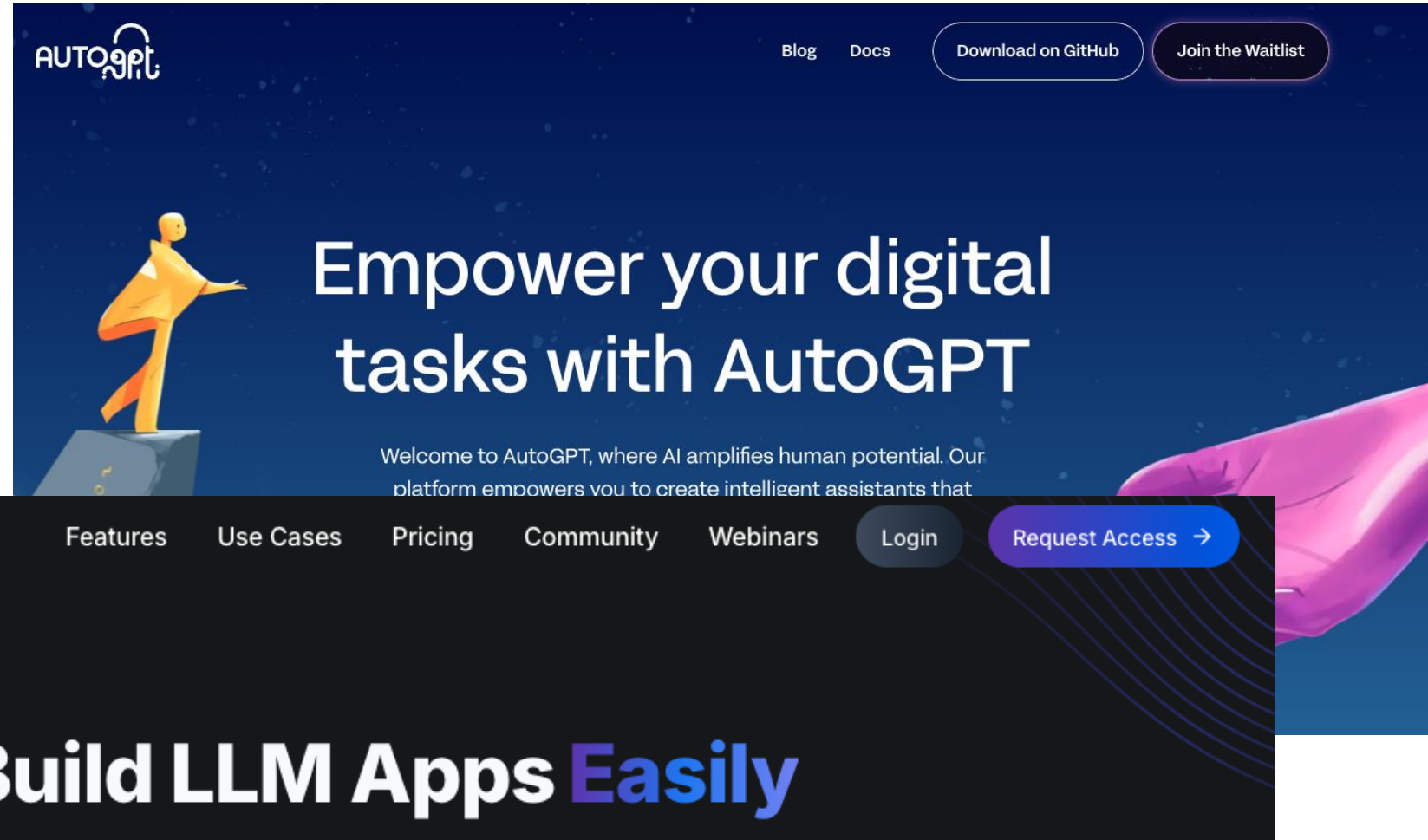
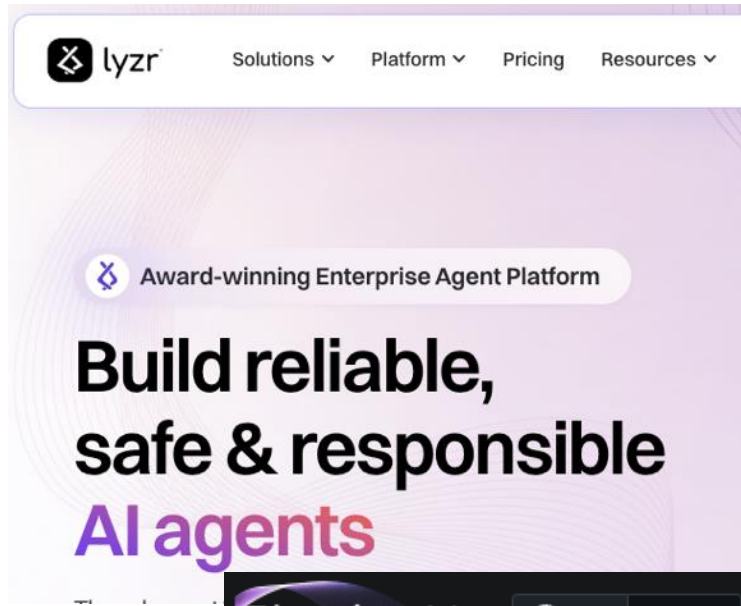
AI Agents Market Landscape



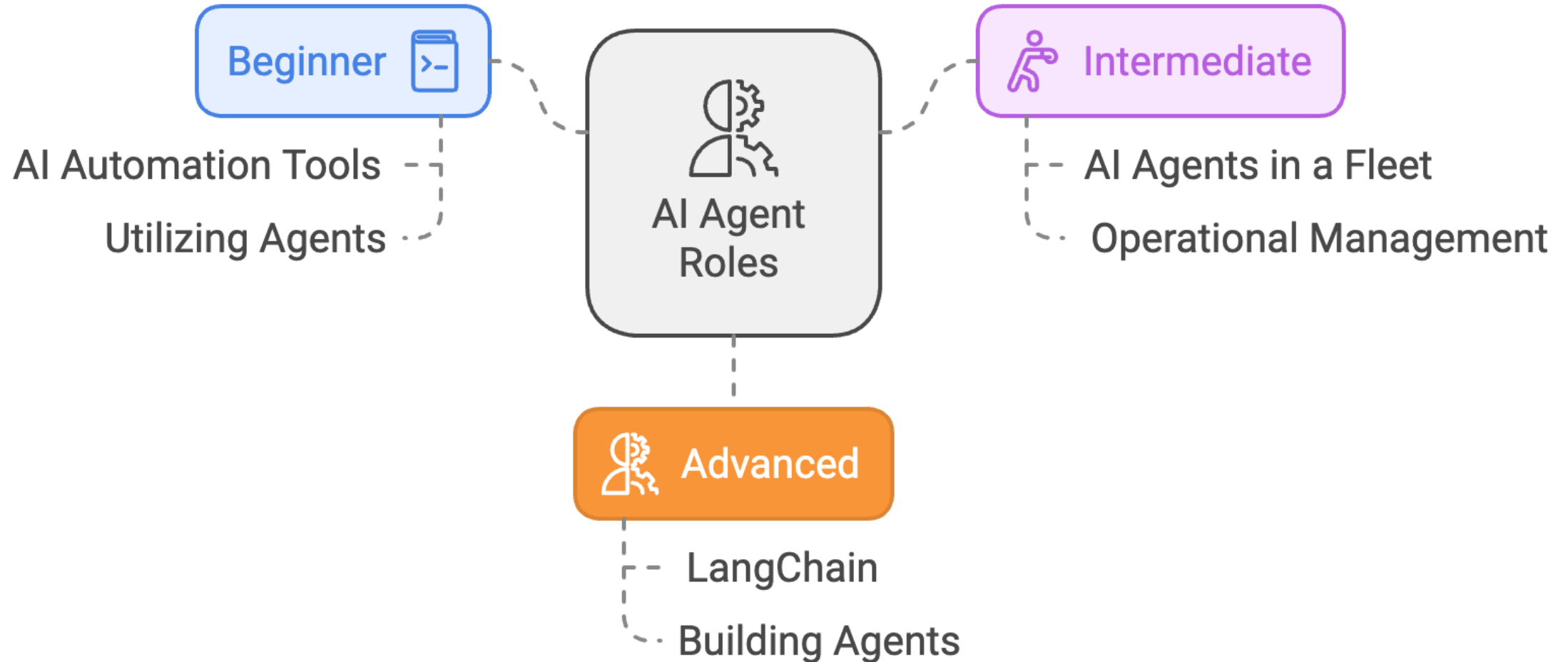
AI Agent Breakdown



AI Agent Development



AI Agent Roles





AI vs AI

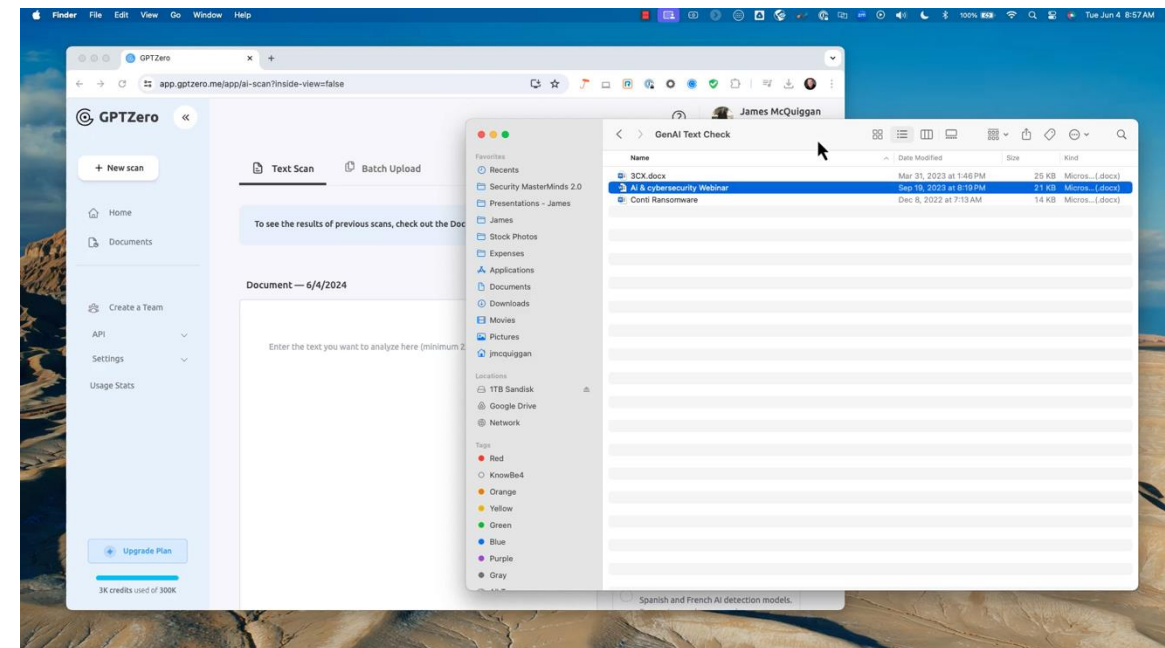
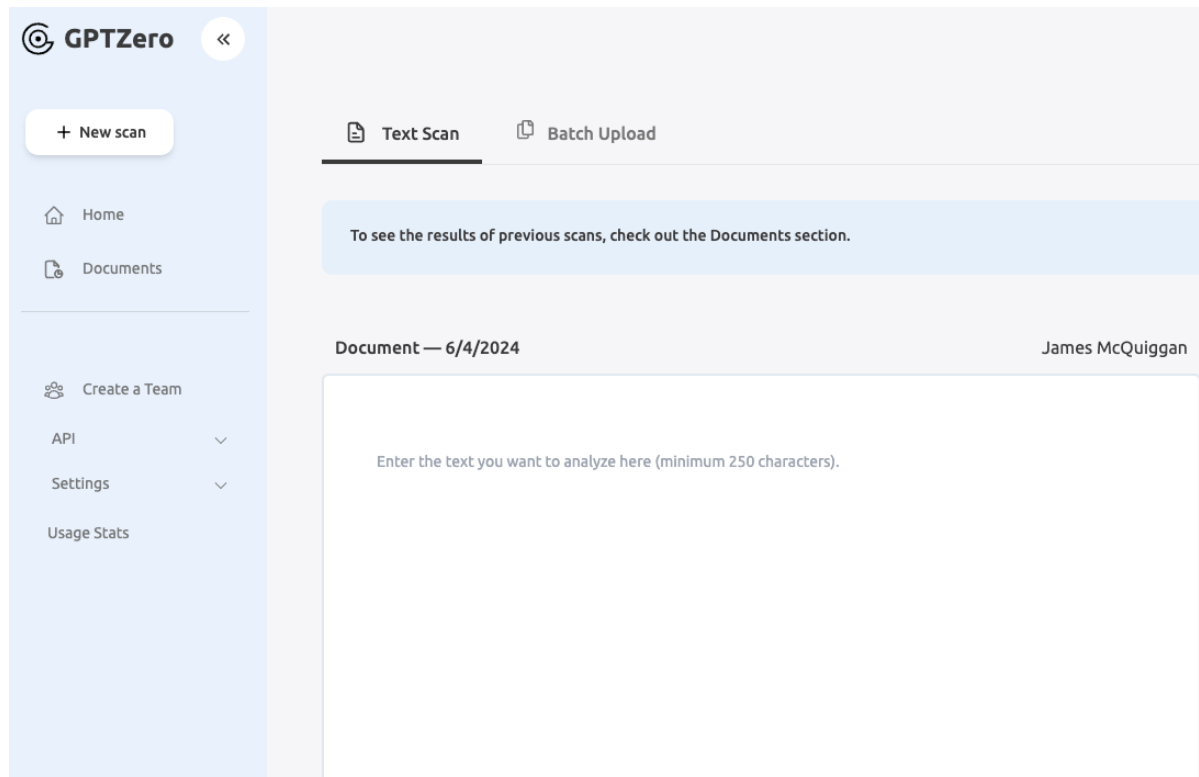
Real World Examples of how AI is used to defend and protect organizations

FlipSide – Used Against the Scammers

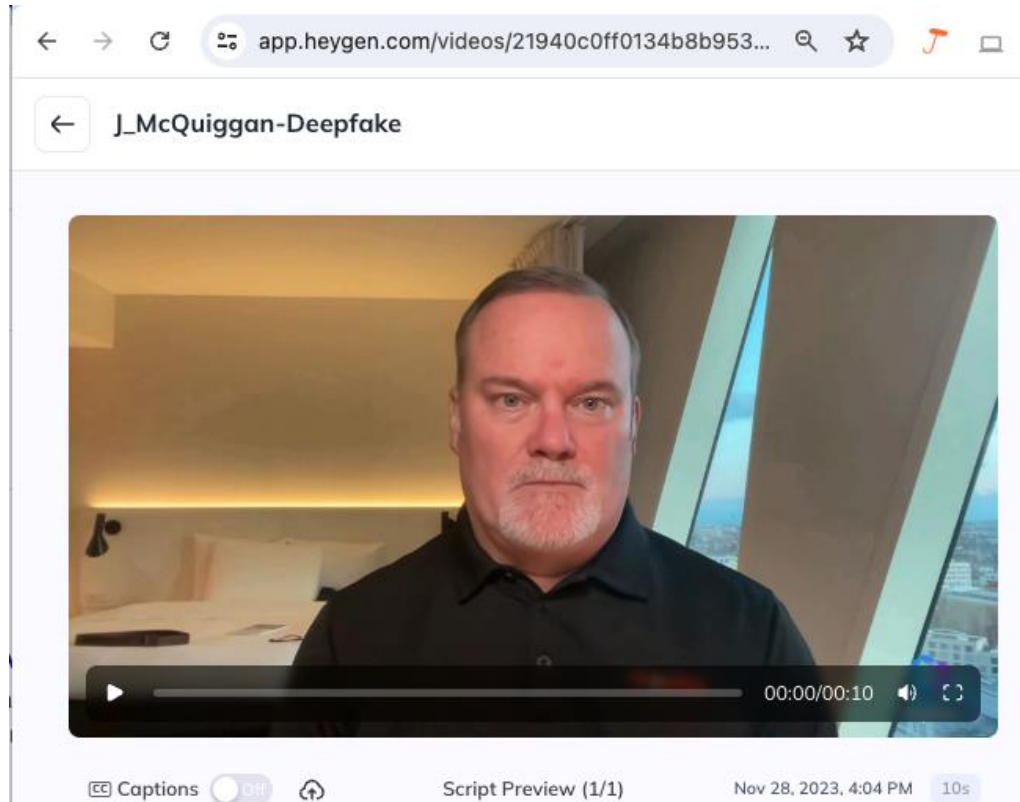


GPTZero – GenAI Synthetic Text Detection

- <https://app.gptzero.me/app/ai-scan>





Modality – Text to Video (HeyGen Demo)



Now I took this one step further and created this audio and video which I recorded from a hotel room using HeyGen and all I did was type in a script of what I wanted to say.

James HeyGen Video - Deepware

 **DEEPPFAKE DETECTED**



Name:	James -BSidesCPH.mp4	User	202
Size:	5.5 MB	Source	

Deepware aims to give an opinion about the scanned video and is not responsible for the result. As Deepware Scanner is still in beta, the results should not be treated as an absolute truth or evidence.



Model Results

- Avatarify:** DEEPPFAKE DETECTED(94%)
- Deepware:** NO DEEPPFAKE DETECTED(0%)
- Seferbekov:** NO DEEPPFAKE DETECTED(31%)
- Ensemble:** NO DEEPPFAKE DETECTED(4%)

Video

- Duration:** 9 sec
- Resolution:** 1920 x 1080
- Frame Rate:** 25 fps
- Codec:** h264

Audio

- Duration:** 9 sec
- Channel:** stereo
- Sample Rate:** 48 khz
- Codec:** aac

- O**
- n:** 9 sec
- I:** stereo
- Rate:** 48 khz
- aac**

✓

NO DEEPPFAKE DETECTED

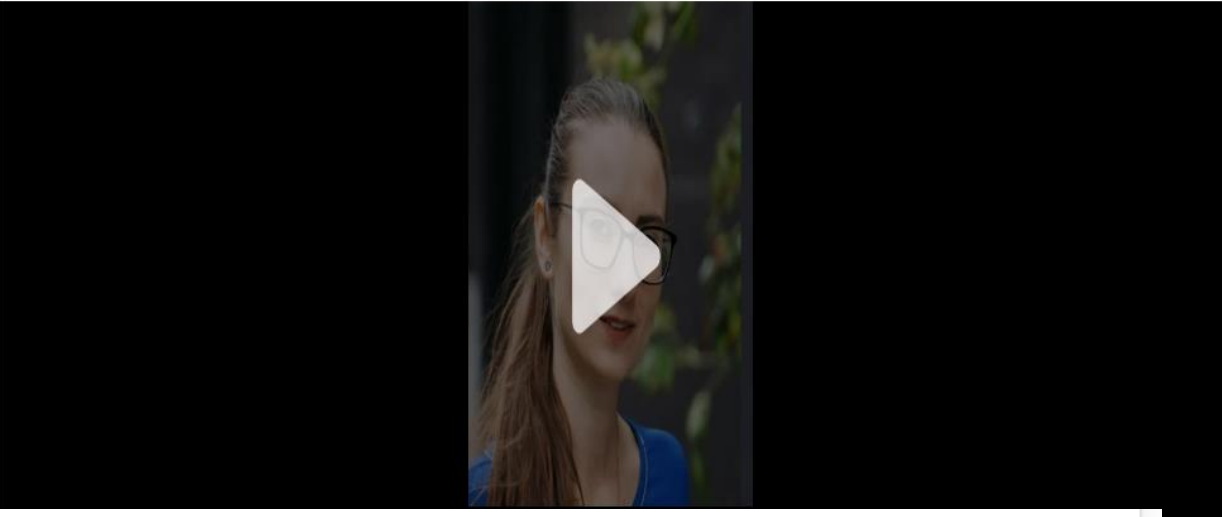
Name:

af66f232-b4ac-4a35-a2dd-d6c93655ba68.mp

Size:

9.5 MB

Deepware aims to give an opinion about the scanned video and is not responsible for the result. As Deepware Scanner is still in beta, the results should not be treated as an absolute truth or evidence.



Model Results

- Avatarify: NO DEEPPFAKE DETECTED(43%)
- Deepware: NO DEEPPFAKE DETECTED(0%)
- Seferbekov: NO DEEPPFAKE DETECTED(1%)
- Ensemble: NO DEEPPFAKE DETECTED(0%)

Video

- Duration: 37 sec
- Resolution: 512 x 512
- Frame Rate: 30 fps
- Codec: h264

Audio


- Duration: 37 sec
- Channel: mono
- Sample Rate: 48 khz
- Codec: aac

Andrada's Audio File

detect.resemble.ai/results/0b7e6bac1708987c39e00b3d2805fd0c

Resemble Detect

Detect deepfake audio from any source with our powerful AI Model. [Try it out yourself →](#)



A horizontal audio waveform visualization showing amplitude over time. The waveform is rendered in a light orange color against a dark background. It consists of a series of peaks and valleys, indicating the frequency and volume of the audio signal.

Result: **Fake**

<https://detect.resemble.ai/results/0b7e6bac1708987c39e00b3d2805fd0c>



AI vs AI Protections

Looking at how AI can protect and defend our organizations.

AI in Cybersecurity

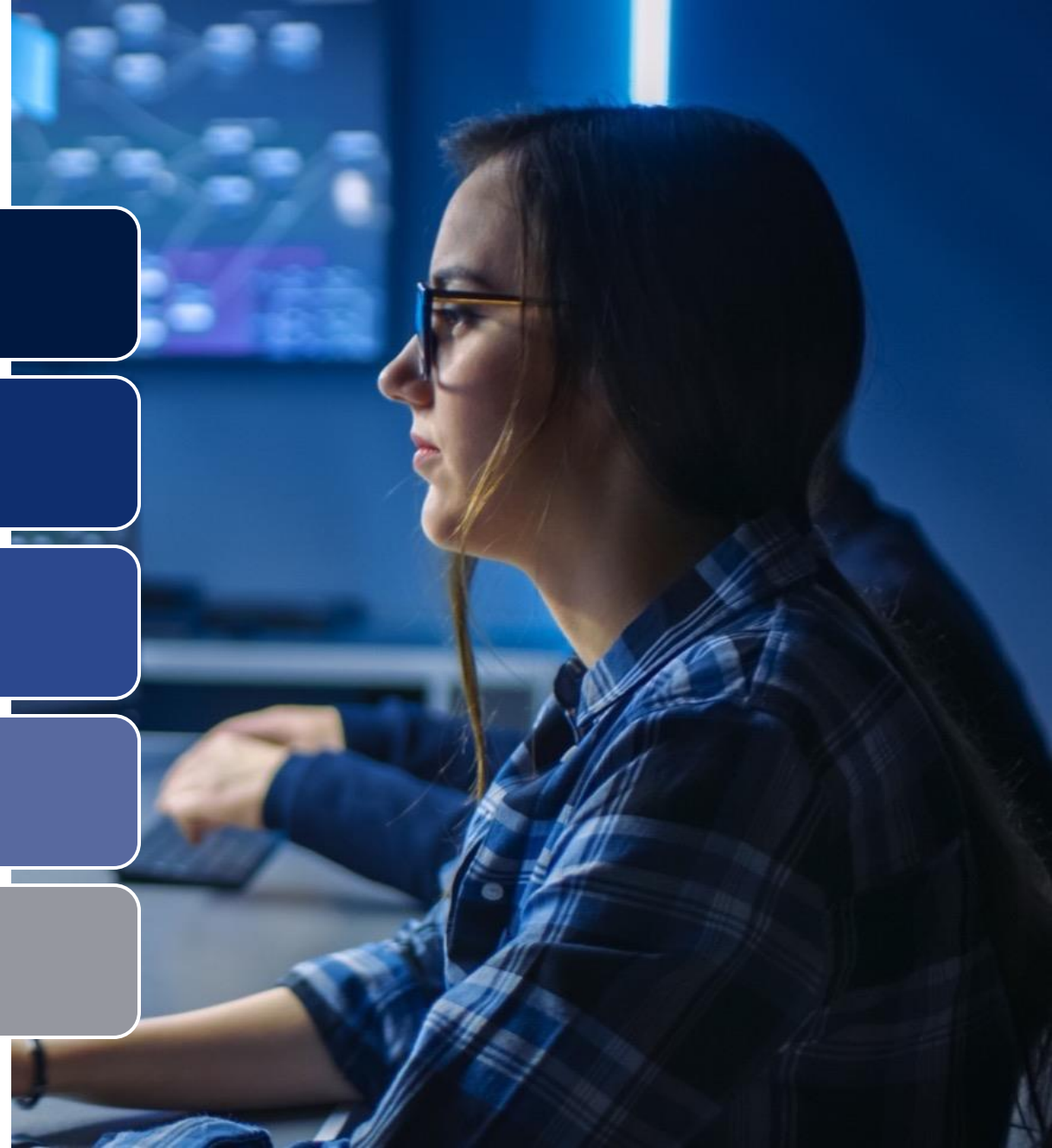
Threat Monitoring

Vulnerability Management

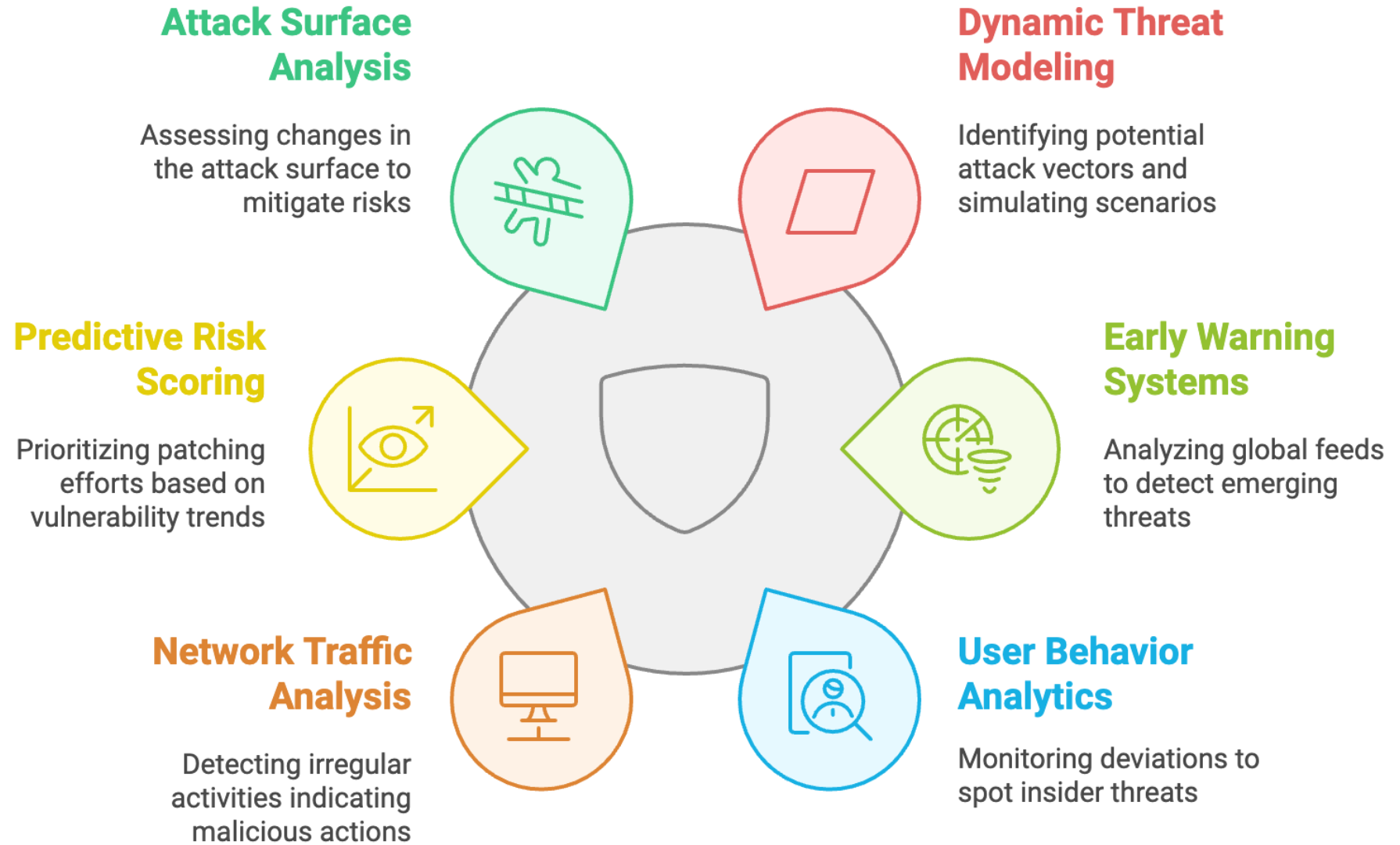
SOAR

Predictive Threat Intelligence

SOC Support



AI Predictive Functionality in Cybersecurity



Basic Detection

- The criminals have leveled up
- We need to continue on the same path, but with some changes
- Asking ourselves if something feels odd
- If a request is abnormal will go a long way toward countering these attacks
- Have code words, or ask unusual questions



3 Questions to Ask Your Email



Create A Social Engineering Defense in Depth

Mitigate

- Aggressively Mitigate Social Engineering
- People, Processes, Technology

Patch

- Patch Exploited Software & Firmware
- Monitor the CISA KEV Catalog (Known Exploited Vulnerabilities)

MFA

- Use MFA Wherever Possible
- Non-phishable MFA too, Avoid SMS and verify all requests that you didn't initiate

SPOT

- Learn how to Spot Rogue URLs
- No longer – don't click on links, or check your links, make sure they know how

DiNGS

- Remember to use DiNG style of passwords
- Different, Non-Guessable, Strong



Apply the FAIK Factor Framework

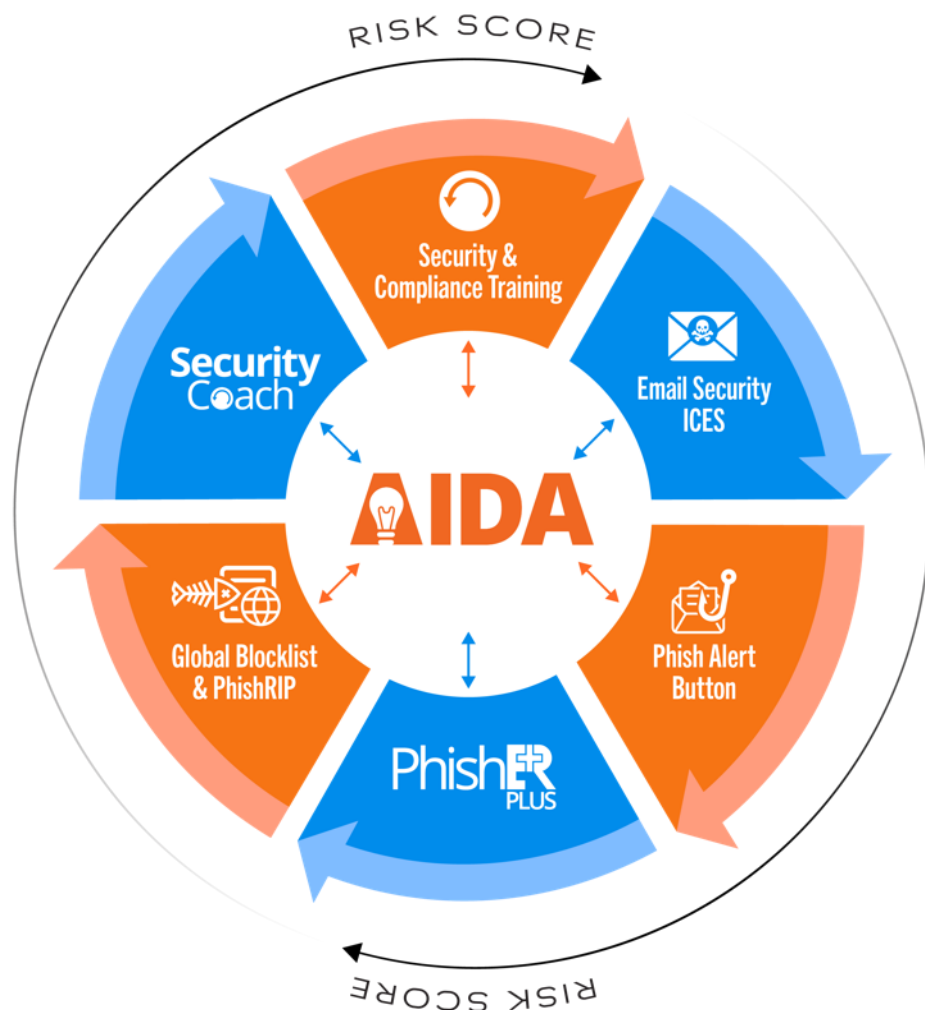
F Freeze & Feel

A Analyze the Narrative & Emotional Triggers

I Investigate (claims, sources, etc.)

K Know, confirm, and keep vigilant

How KnowBe4 Uses AI – 4 AI Agents



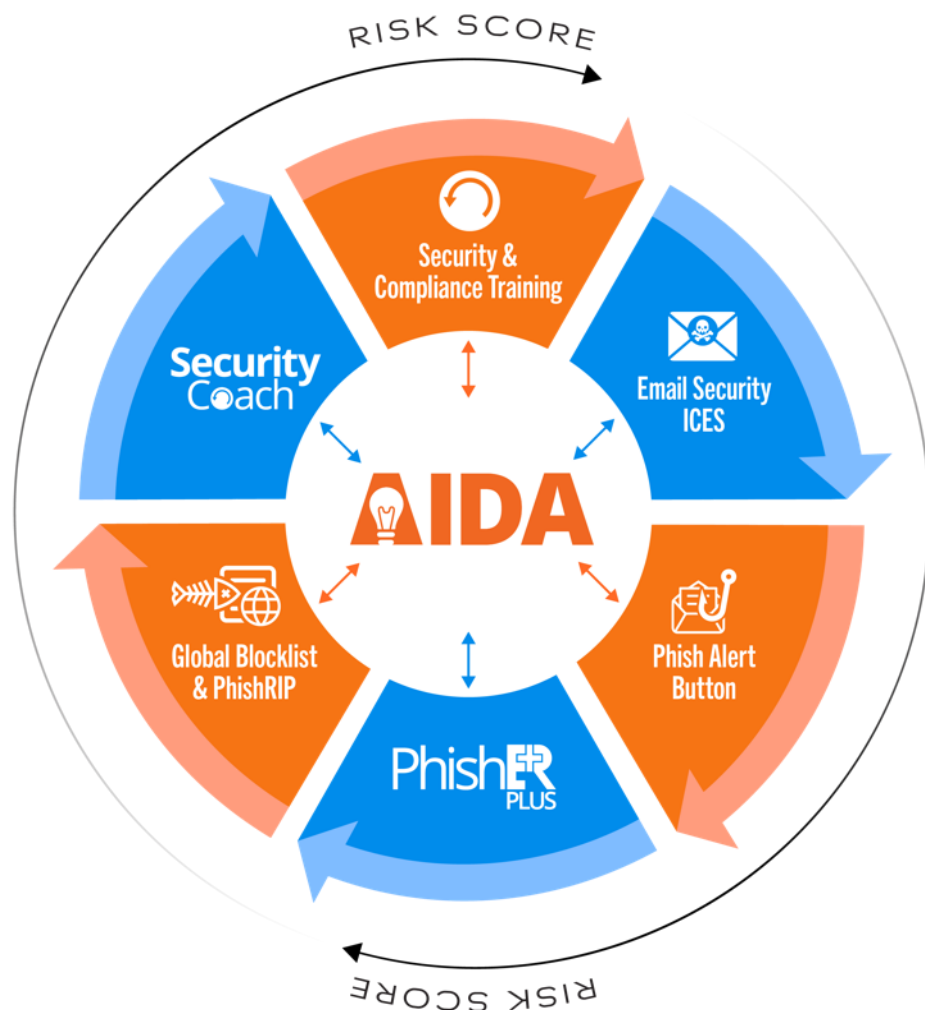
- **AIDA Automated Training**

- Agent to assign engaging content based on failure type they've experienced
- Based on the new 2.0 Risk Score

- **Phishing Template**

- Realistic phishing emails
- Leverage the NIST Phish Scale Framework

How KnowBe4 Uses AI – 4 AI Agents



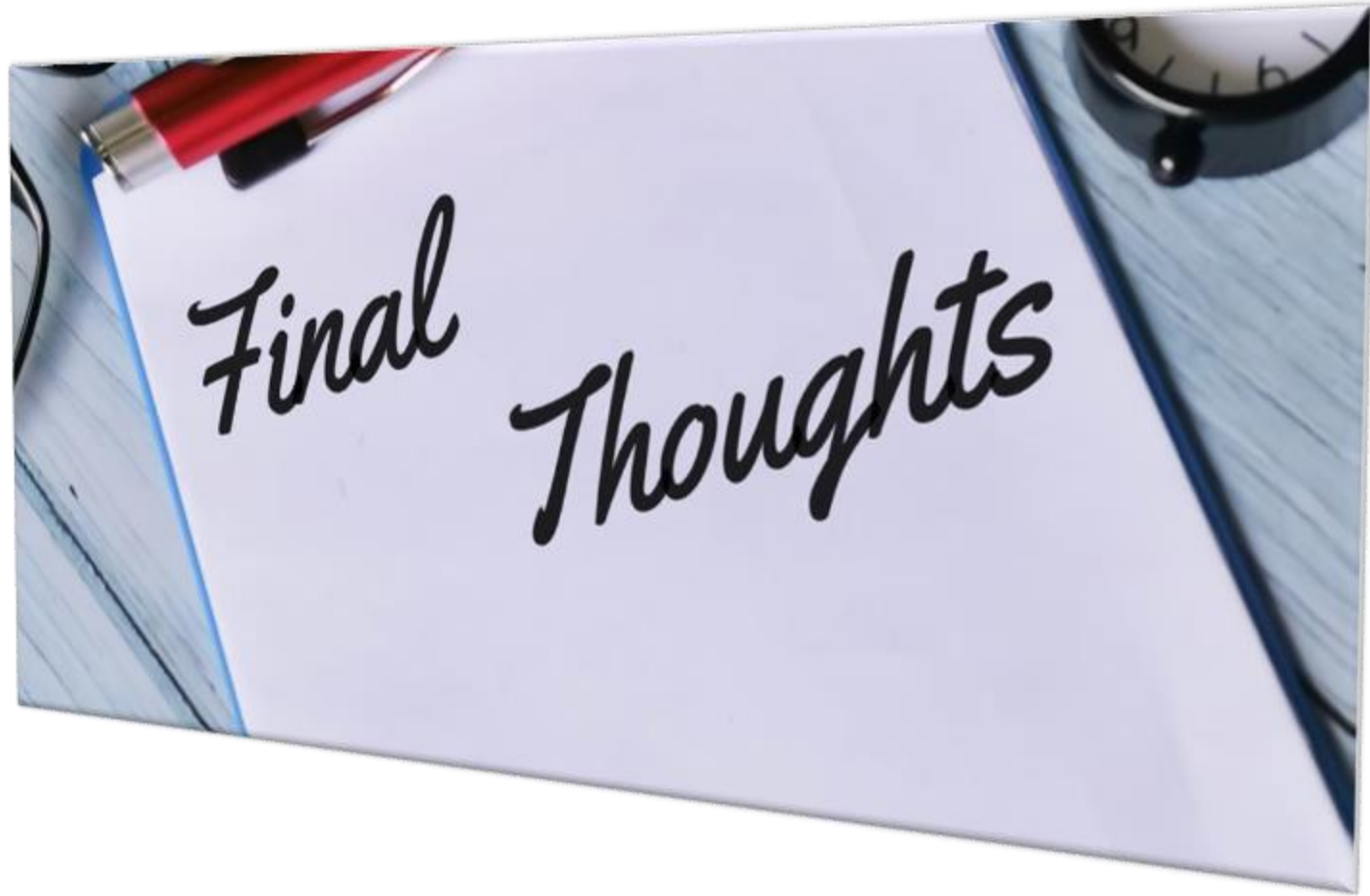
- **Knowledge Refreshers**

- Bite-size refresher to trigger and retain training information
- Not scored, but an amplifier

- **Policy Quizzes**

- Upload your policy
- AI generates an editable quiz
- Scenario based, not content

Wrap-Up / Q&A



Takeaways

AI is an incredible tool available to all, but like any tool there are many ways it can be used maliciously

AI Agents are the next step in the AI evolution

Educate Your Users on AI advancements, recognize deepfakes and be vigilant

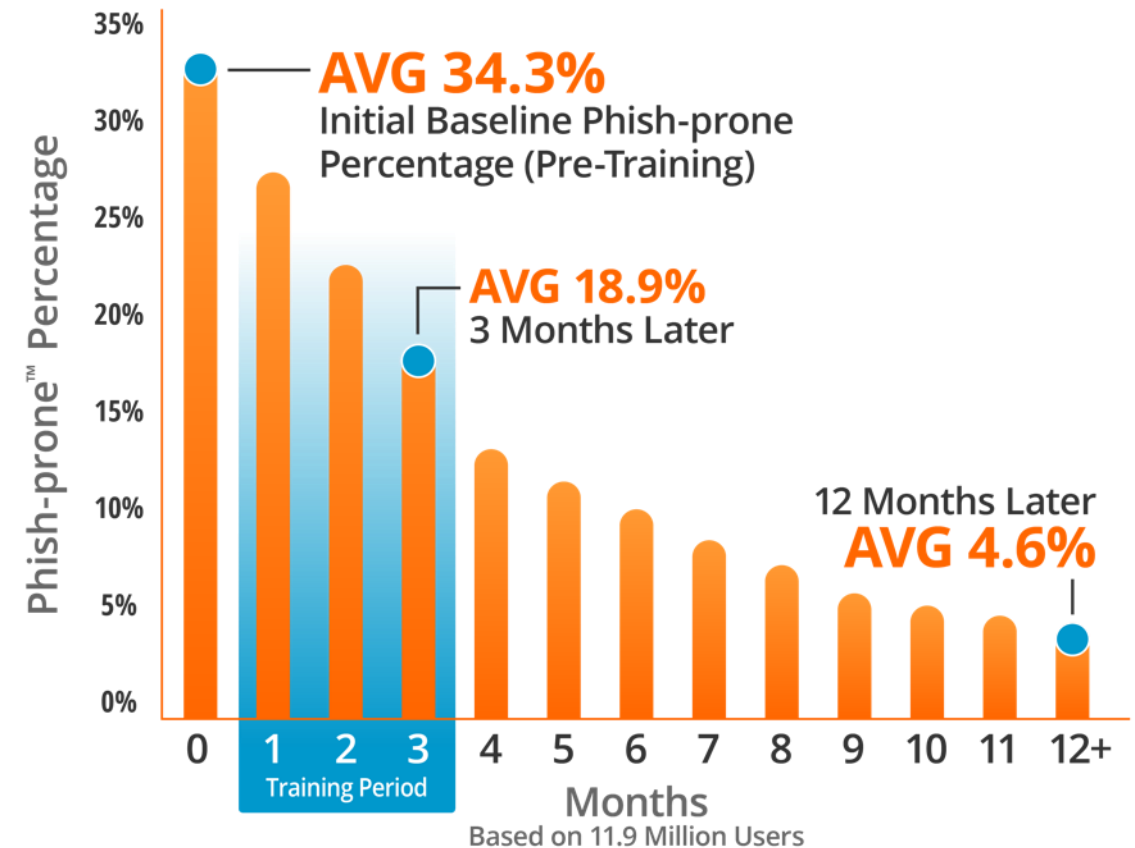
Continue to Test Employees Regularly

- Even when testing confirms that phishing susceptibility has fallen to nominal levels, **continue to test** employees frequently to keep them on their toes, **with security top of mind**.
- **Bad actors are always changing the rules**, adjusting their tactics and upgrading their technologies.
- **Analyze your phishing data**. Continue to train and phish your users with more advanced tactics such as attachments and landing pages where they are asked to enter data.
- Over time, **increase the difficulty of the attacks**, KnowBe4 has 25,000+ templates rated by difficulty from 1 to 5.



KnowBe4 Security Awareness Training Works

Effectively managing this problem requires ongoing due diligence, but it *can* be done and it isn't difficult. ***We're here to help.***



Source: 2024 KnowBe4 Phishing by Industry Benchmarking Report

Note: The initial Phish-prone Percentage is calculated on the basis of all users evaluated. These users had not received any training with the KnowBe4 console prior to the evaluation. Subsequent time periods reflect Phish-prone Percentages for the subset of users who received training with the KnowBe4 console.

KnowBe4 AIDA – The Next Step in Your Security Awareness Programs



Events

**SmartRisk
Agent**

**Insights & Risk
Score**

AIDA



Automated Training Agent

Analyzes user data to
assign the most relevant
and engaging content



Template Generation Agent

Creates highly realistic
phishing templates based
on current attack vectors



Knowledge Refresher Agent

Delivers bite-
sized refreshers at
optimal intervals



Policy Quiz Agent

Generates quizzes based
on your specific security
and compliance policies

Thank You For Your Attention

James R. McQuiggan, CISSP, SACP

Email: jmcquiggan@knowbe4.com

KnowBe4 Blog: blog.knowbe4.com



Connect with Me!



LinkedIn [jmcquiggan](#)



X [@james_mcquiggan](#)



Website jamesmcquiggan.com

Resources

Training

- AI Agent talk – Andrew Ng- Yes layer
- Allie K Miller – AI for Business Leaders, Maven Lectures

Directories

- AI Agents Directory - <https://aiagentsdirectory.com>
- AI Agents Directory - <https://aiagents.directory/>
- Agent AI - <https://agent.ai/>
- AI Agents List - <https://aiagentslist.com/categories/data-analysis>

Codeless Agent Tools

- AutoGPT - https://agpt.co (local only)
- Flowise AI - https://flowiseai.com

KnowBe4

THANK YOU!

