



Fake News, and Deepfakes

Harmless Fun, or The Future of Fraud?



Javvad Malik
Security Awareness Advocate
KnowBe4, Inc.



Erich Kron
Security Awareness Advocate
KnowBe4, Inc.



Erich Kron

Security Awareness Advocate



@ErichKron

About Erich Kron

- CISSP, CISSP-ISSAP, MCITP, ITIL v3, etc...
- Former Security Manager for the US Army 2nd Regional Cyber Center – Western Hemisphere
- Former Director of Member Relations and Services for (ISC)²
- A veteran of IT and Security since the mid 1990's in manufacturing, healthcare and DoD environments
- Hero to Javvad



Certified Information
Systems Security Professional




Certified Information
Systems Security Professional

ISSAP Architecture

Microsoft
CERTIFIED
IT Professional



Javvad Malik
Security Awareness Advocate

 @J4vv4D

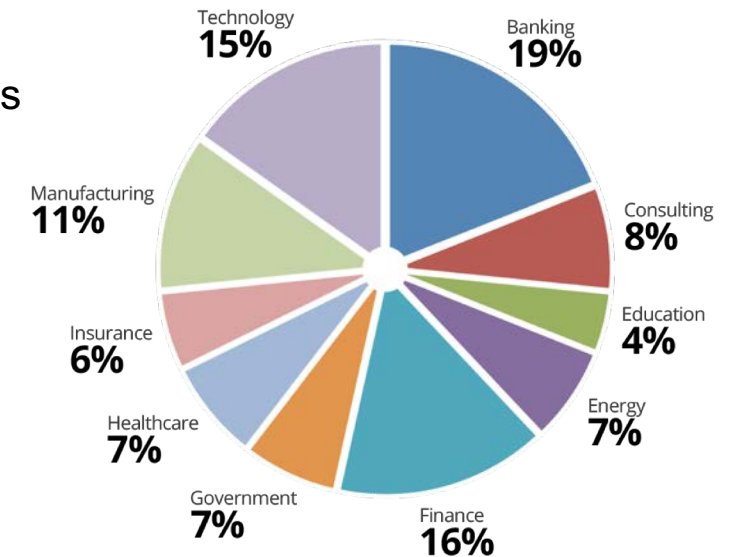
About Javvad Malik

- 20 years in computer security
- IT Security Operation
- Infosec Consultant
- Analyst at 451Research
- Security Advocate at AlienVault
- YouTuber
- Podcast Host
- Blogger
- Hero to millions



KnowBe4, Inc.

- The world's most popular integrated Security Awareness Training and Simulated Phishing platform
- Based in Tampa Bay, Florida, founded in 2010
- CEO & employees are ex-antivirus, IT Security pros
- 200% growth year over year
- We help tens of thousands of organizations manage the problem of social engineering



Agenda

- How this all got started
- The progression of digital fakes
- Potential impact of fakes
- Detection of digital fakes
- Defending against fakery and silliness

Agenda

- How this all got started
- The progression of digital fakes
- Potential impact of fakes
- Detection of digital fakes
- Defending against fakery and silliness

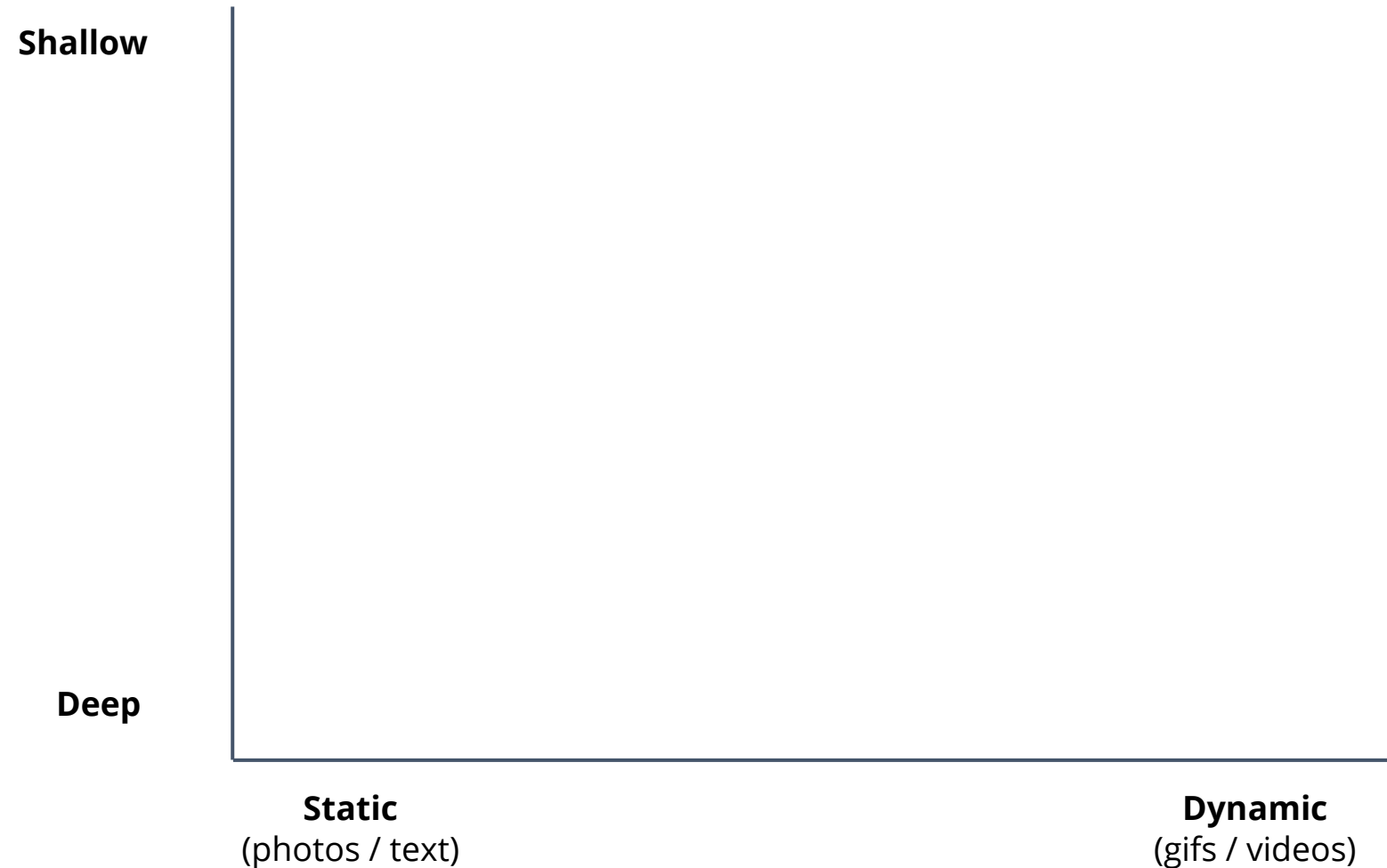
It All Started With Some Fun

- Jawwad and I have known each other for years and love to mess with each other
- We started with the simple Photoshop-type pictures and escalated from there
- We started taking conversations and changing the context by digitally altering what we had said
- It occurred to us how dangerous this could become given the quickly spreading nature of social media
- Then, we started playing with deepfake technology

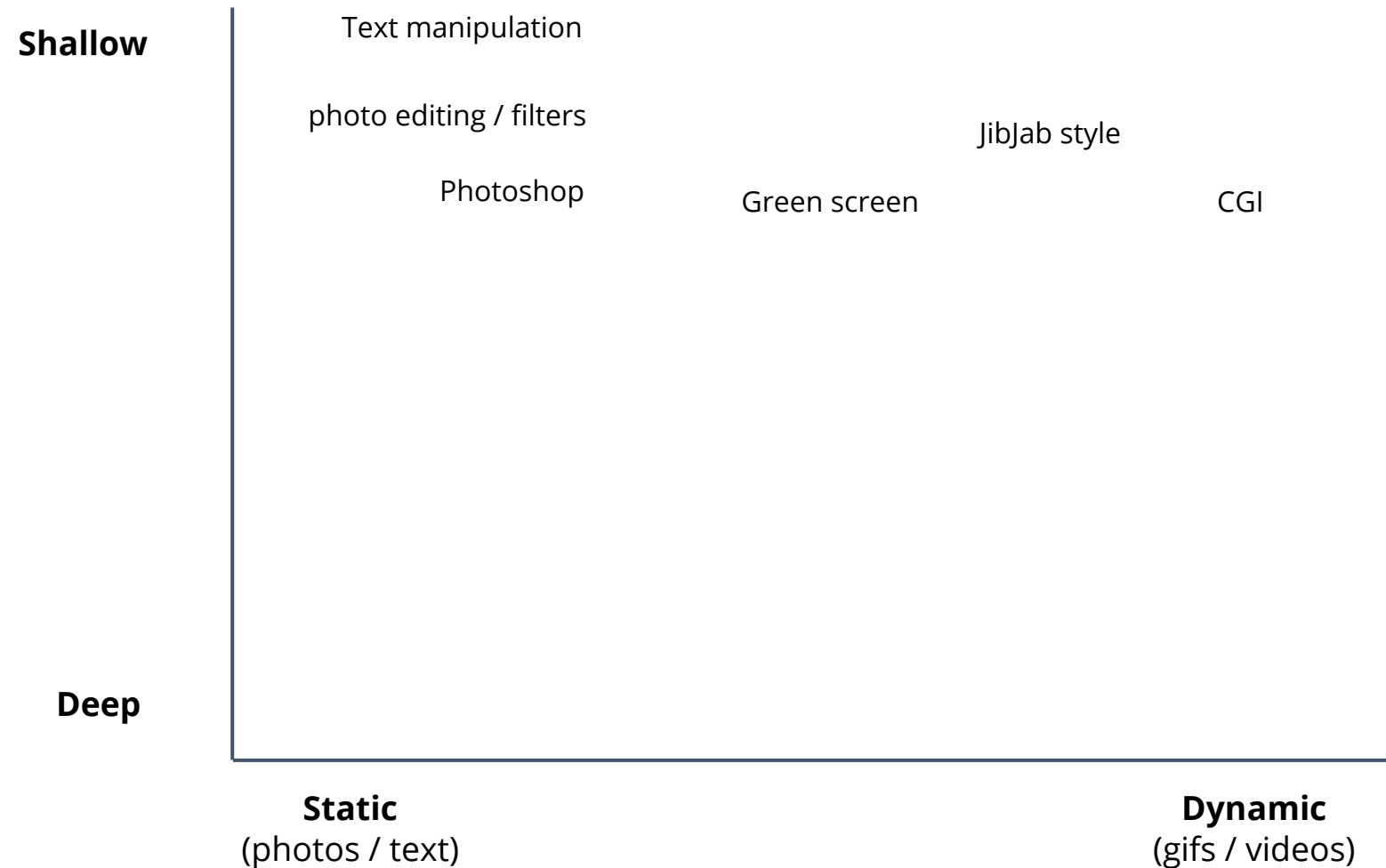
Agenda

- How this all got started
- The progression of digital fakes
- Potential impact of fakes
- Detection of digital fakes
- Defending against fakery and silliness

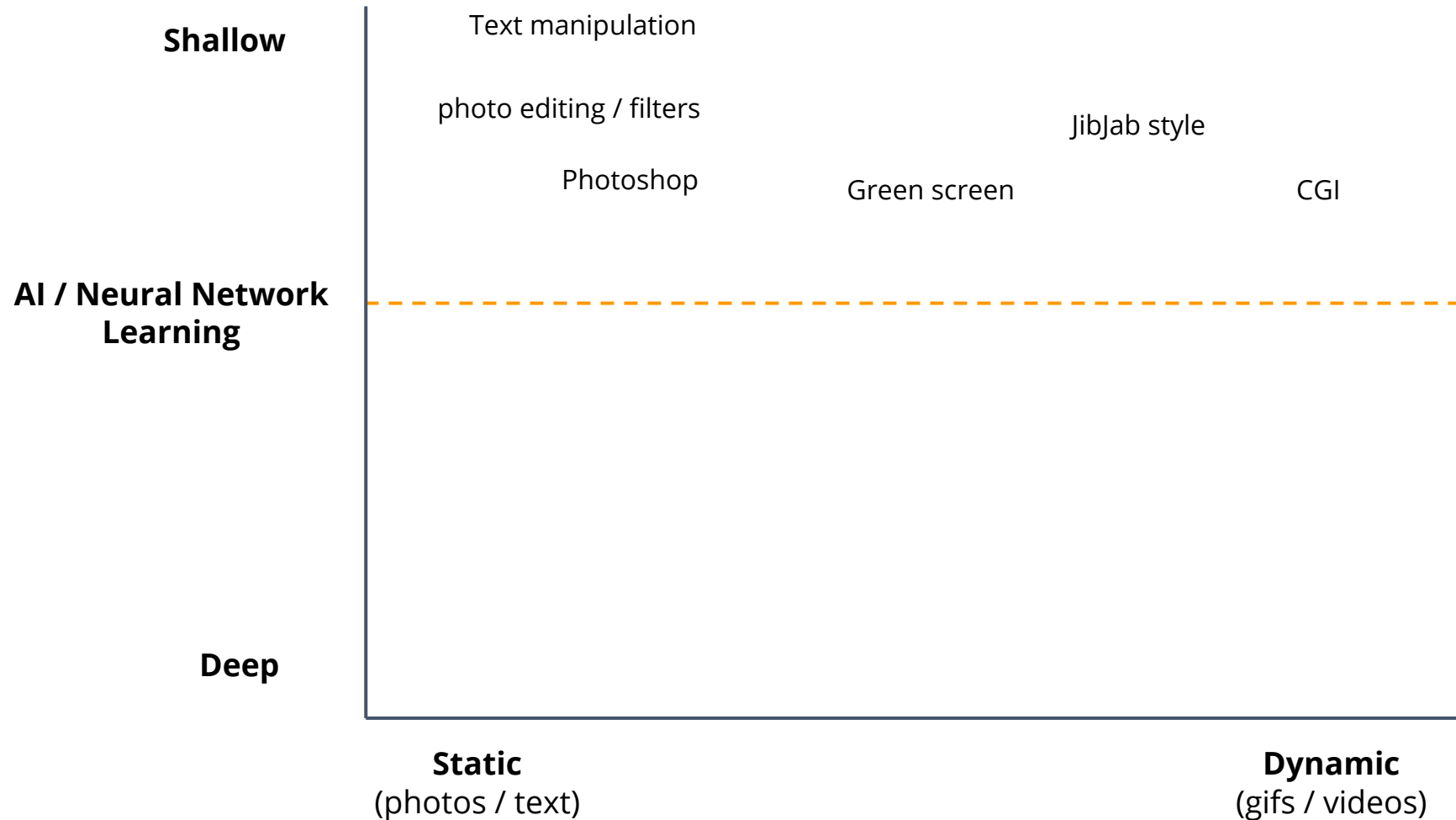
Jav's quad chart and stuff



Jav's quad chart and stuff



Jav's quad chart and stuff



Jav's quad chart and stuff

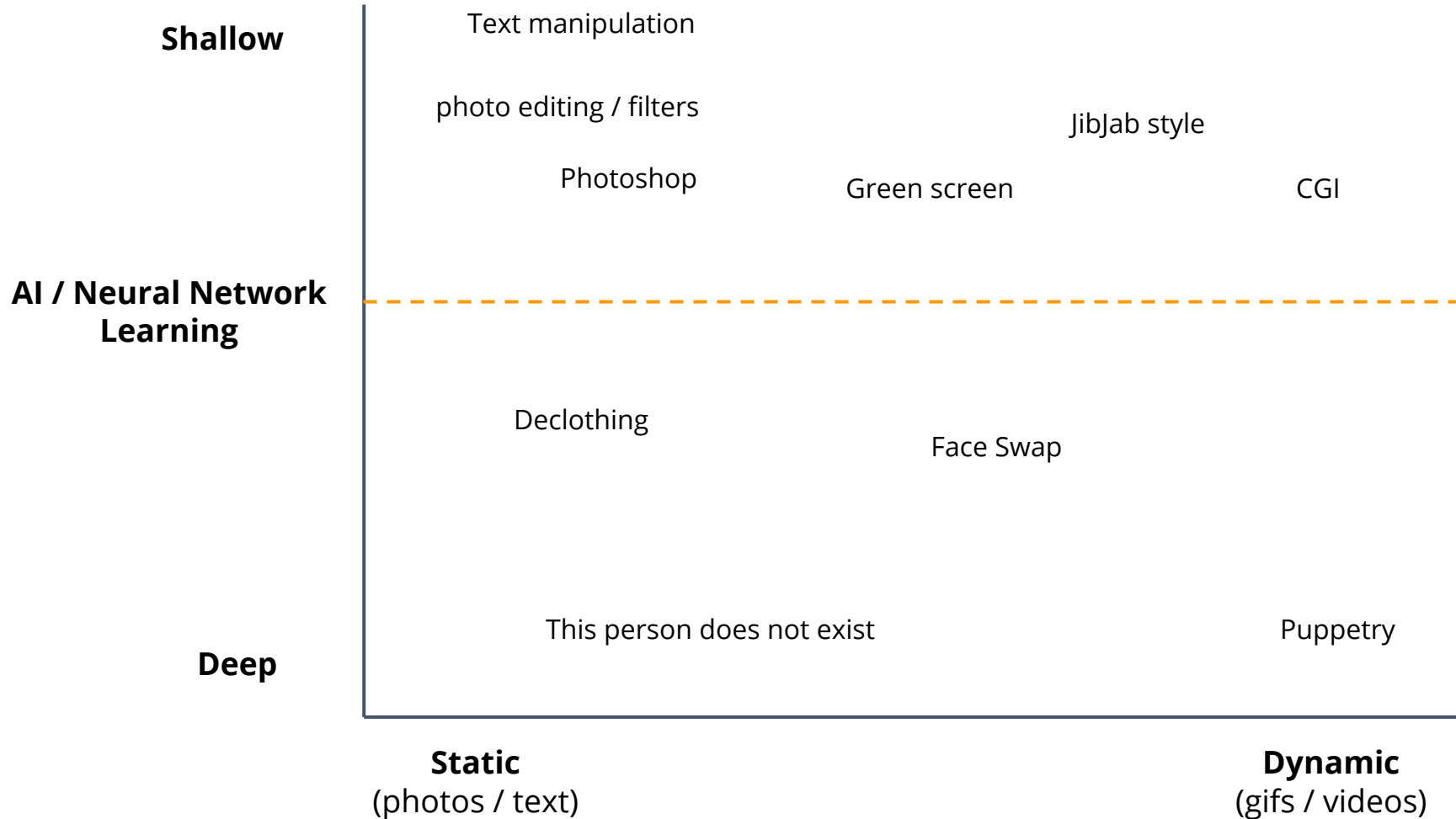


Image Manipulation

- Removing people from busy locations using image stacking and filters built into Photoshop
- Tilt-shift photos
- Forced perspective
- Can be used to modify “screenshots” as well as photos

Image Stacking in Photoshop

- This is an included script in recent versions of Photoshop
- You use multiple photos from the same place over time, and the software removes the things that moved



Tilt-Shift Images

- Makes real photos look unreal
- Fairly simple technique that lends itself to angled shots from above, like those taken from a drone



Forced Perspective

- Angles and perspective make things look unreal
- Seen a lot when someone wants to exaggerate the size of an animal or item



Face Swap Apps and Filters

- Add cat ears in real time
- Age you by 20 or 30 years
- Face-swapping apps
- Making animals talk

Augmented Reality and Filters

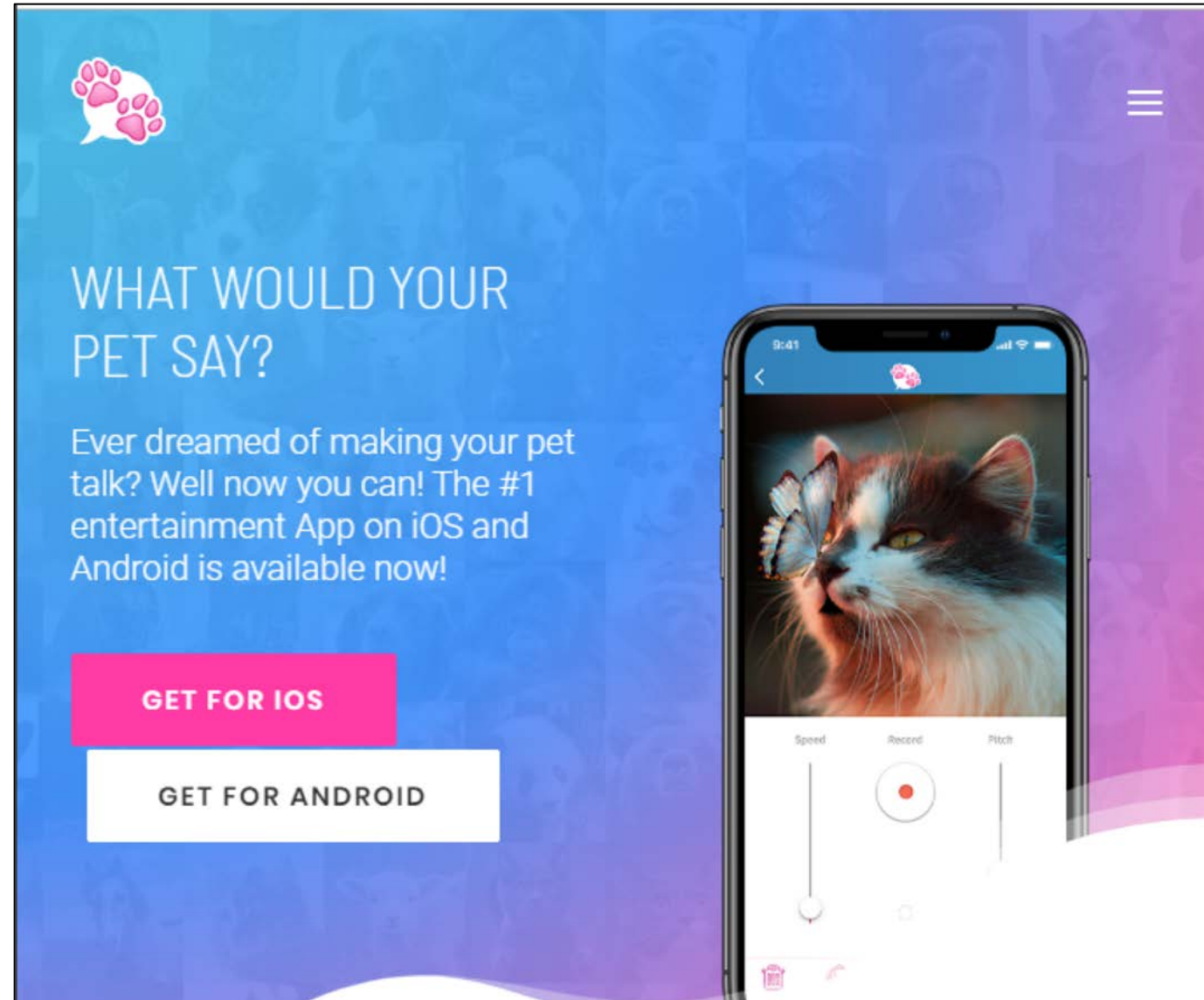
- Apps are able to augment video in real time adding important things such as cat ears, mustaches or any of 100 other things
- This is actually a very powerful technology to have in something like a phone



My Talking Pet

- Because everyone should be able to put words in their pet's mouths
- This is not new tech, it's been around for a while now

Video at <https://www.youtube.com/watch?v=5AYrWpT6bl8>



Video Deepfake apps

- Downloadable from GitHub
- Typically uses a GPU to make things quick
- Extracts single frames from the video, extracts faces, then trains AI to build a model and replace faces
- Good fakes generally required a voiceover actor

DeepFaceLab

- Downloadable from GitHub <https://github.com/iperov/DeepFaceLab>
- Windows binaries are downloadable
- Extracts single frames from the video, extracts faces, then trains AI to build a model and replace faces
- Does not change the audio

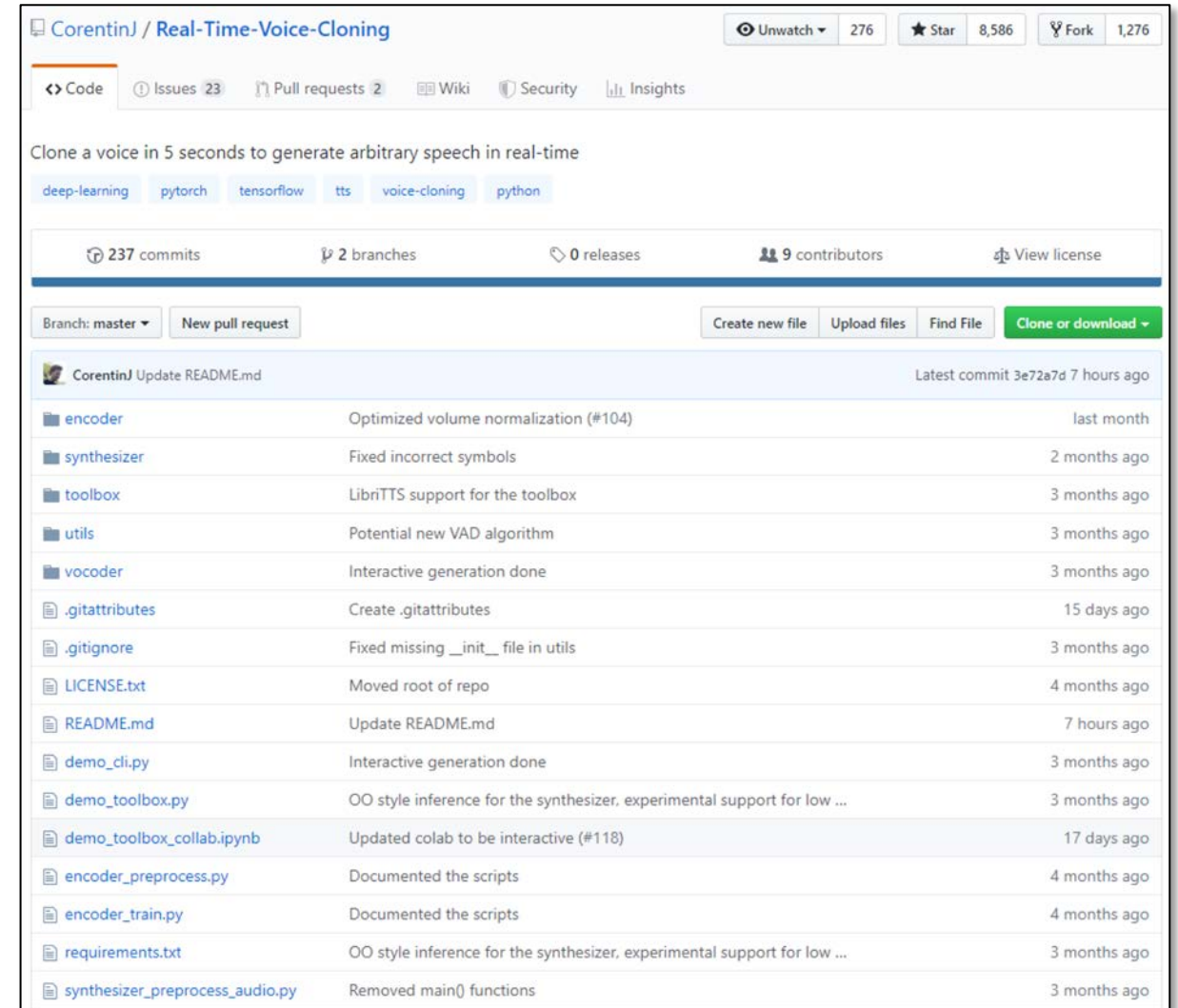
The screenshot shows the GitHub repository page for `iperov / DeepFaceLab`. The repository has 468 releases, 9,977 stars, and 2,262 forks. The file `doc_prebuilt_windows_app.md` is selected, showing 23 lines (12 sloc) and 983 Bytes. The file content includes a section titled "Prebuilt Windows Releases" which states that Windows builds with all dependencies are released regularly, and only the NVIDIA GeForce display driver needs to be installed. It provides a link to "Google drive" for downloading prebuilt DeepFaceLab, including GPU and CPU versions. Below this, it lists "Available builds:" with three bullet points: "DeepFaceLabCUDA9.2SSE - for NVIDIA cards up to GTX1080 and any 64-bit CPU", "DeepFaceLabCUDA10.1AVX - for NVIDIA cards up to RTX and CPU with AVX instructions support", and "DeepFaceLabOpenCLSSE - for AMD/IntelHD cards and any 64-bit CPU". Finally, it lists "Video tutorials using prebuilt windows app" with four bullet points: "Basic workflow", "Basic workflow (thanks @derpfakes)", "How To Make DeepFakes With DeepFaceLab - An Amatuer's Guide", and "Manual re-extract poorly aligned frames".

Audio Deepfake apps

- Also downloadable from GitHub
- Also typically uses a GPU to make things quick
- Uses audio clips and AI to replace or create the voice

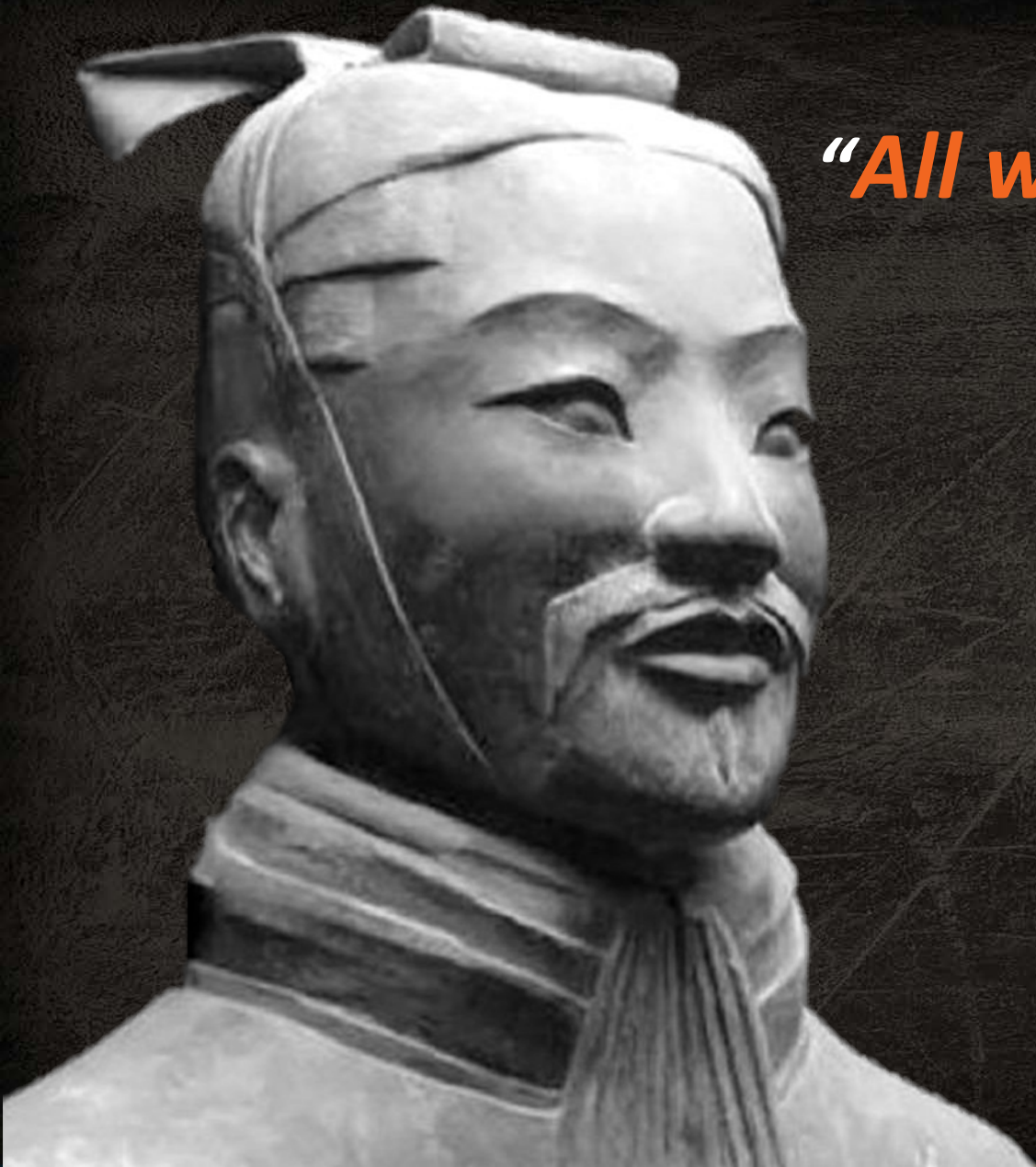
Real-Time Voice Cloning

- Downloadable from GitHub <https://github.com/CorentinJ/Real-Time-Voice-Cloning>
- Python-based tool to generate arbitrary speech in real time
- Uses pretrained models to speed up implementation
- Not great, but steadily improving



Agenda

- How this all got started
- The progression of digital fakes
- Potential impact of fakes
- Detection of digital fakes
- Defending against fakery and silliness



“All warfare is based on deception.”

- Sun Tzu, *The Art of War*

Example: Business Email Compromise (The Phish Evolved)

- a.k.a. CEO Fraud
- No payload
- Low volume email targeting high value individuals
- Personalized
- Few to no 'traditional' spam/phishing tells (such as poor grammar, egregious misspellings, etc.)



Understanding *the root* of deception



**Our brains' job
to filter,
interpret,
and present
'reality'**

Video or Photo Fakes

- Romance scams (send a video holding a recent newspaper as proof)
- Outrage clickbait (sign a petition sort of scam)
- Election influencing (damage done before the issue could be proven as fake)

Weaponized Videos or Photos



Audio Fakes

- Romance scams (send an audio clip professing love)
- Election influencing (fake phone call intercepted?)
- CEO Fraud or other types of BEC

Possible Weaponized Voice Cloning?

PRO CYBER NEWS

Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case

Scams using artificial intelligence are a new challenge for companies



PHOTO: SIMON DAWSON/BLOOMBERG NEWS

By Catherine Stupp

Updated Aug. 30, 2019 12:52 pm ET

Criminals used artificial intelligence-based software to impersonate a chief executive's voice and demand a fraudulent transfer of €220,000 (\$243,000) in March in what

<https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>

Agenda

- How this all got started
- The progression of digital fakes
- Potential impact of fakes
- Detection of digital fakes
- Defending against fakery and silliness

Detection of fakes

- Artifacts always remain regardless of the quality and if they are visible while the video or audio is running
- Compression artifacts in video
- Inaudible artifacts in audio (we can hear 2 different sounds the same way)
- Will there be automation (YouTube, Facebook, etc. already doing it?)
- Neural Networks to detect fakes

The Human Bias Issue

- Deepfakes will enhance prejudices and biases.
- You don't need a high-tech hoax to manipulate someone who already wants to believe something.
- The truth is out there, so long as we care enough to look for it.

Agenda

- The Perception vs. Reality dilemma
- Understanding the OODA (Observe, Orient, Decide, Act) Loop
- How social engineers and scam artists achieve their goals by subverting its different components
- How we can defend ourselves and our organizations



Social Engineering

Are You Being Manipulated?

-- understand the lures --

Greed

Curiosity

Self Interest

Urgency

Fear

Helpfulness

Social Engineering Red Flags



FROM

- I don't recognize the sender's email address as someone I **ordinarily communicate with**.
- This email is from **someone outside my organization and it's not related to my job responsibilities**.
- This email was sent from **someone inside the organization** or from a customer, vendor, or partner and is **very unusual or out of character**.
- Is the sender's email address from a **suspicious domain** (like micorsoft-support.com)?
- I **don't know the sender personally** and they **were not vouched for** by someone I trust.
- I **don't have a business relationship** nor any past communications with the sender.
- This is an **unexpected or unusual email** with an **embedded hyperlink or an attachment** from someone I haven't communicated with recently.



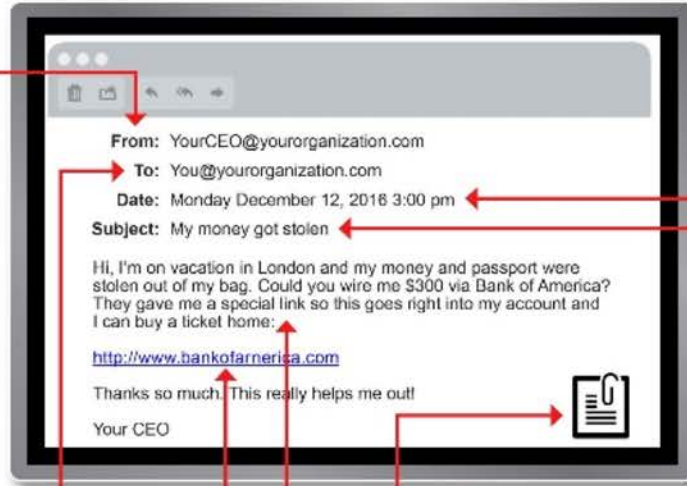
TO

- I was cc'd on an email sent to one or more people, but I **don't personally know** the other people it was sent to.
- I received an email that was also sent to an **unusual mix of people**. For instance, it might be sent to a random group of people at my organization whose last names start with the same letter, or a whole list of unrelated addresses.



HYPERLINKS

- I hover my mouse over a hyperlink that's displayed in the email message, but the **link-to address is for a different website**. (This is a **big red flag**.)
- I received an email that only has **long hyperlinks with no further information**, and the rest of the email is completely blank.
- I received an email with a **hyperlink that is a misspelling** of a known web site. For instance, www.bankofamerica.com — the "m" is really two characters — "r" and "n."



DATE

- Did I receive an email that I normally would get during regular business hours, but it was **sent at an unusual time** like 3 a.m.?



SUBJECT

- Did I get an email with a subject line that is **irrelevant** or **does not match** the message content?
- Is the email message a reply to something I **never sent or requested**?



ATTACHMENTS

- The sender included an email attachment that I **was not expecting** or that **makes no sense** in relation to the email message. (This sender doesn't ordinarily send me this type of attachment.)
- I see an attachment with a possibly **dangerous file type**. The only file type that is **always safe to click on** is a **.txt file**.



CONTENT

- Is the sender asking me to click on a link or open an attachment to **avoid a negative consequence** or to **gain something of value**?
- Is the email **out of the ordinary**, or does it have **bad grammar** or **spelling errors**?
- Is the sender asking me to click a link or open up an attachment that **seems odd** or **illogical**?
- Do I have an **uncomfortable gut feeling** about the sender's request to open an attachment or click a link?
- Is the email asking me to look at a **compromising or embarrassing picture** of myself or someone I know?

It's *more* than just phishing training



Preventing Workplace Harassment for Managers

Published on: July 10th, 2019
Category: [Training Modules](#)
Duration: 60 mins



Executive Series: Secure Destruction of Sensitive Information

Published on: June 27th, 2019
Category: [Video Modules](#)
Duration: 2 mins



Preventing Workplace Harassment for Employees

Published on: July 10th, 2019
Category: [Training Modules](#)
Duration: 60 mins



Executive Series: Ransomware and Bitcoin

Published on: June 17th, 2019
Category: [Video Modules](#)
Duration: 4 mins



New York State Education Law

Published on: June 25th, 2019
Category: [Training Modules](#)
Duration: 20 mins



Executive Series: Decision-Maker Email Threats

Published on: June 27th, 2019
Category: [Video Modules](#)
Duration: 5 mins



Travel Security Challenge Game

Published on: June 24th, 2019
Category: [Games](#)
Duration: 7 mins

Build engagement and decrease behavior-related risk



Baseline Testing

We provide baseline testing to assess the Phish-prone™ percentage of your users through a free simulated phishing attack.



Train Your Users

On-demand, interactive, engaging training with common traps, live hacking demos and new scenario-based Danger Zone exercises and educate with ongoing security hints and tips emails.



Phish Your Users

Fully automated simulated phishing attacks, hundreds of templates with unlimited usage, and community phishing templates.



See the Results

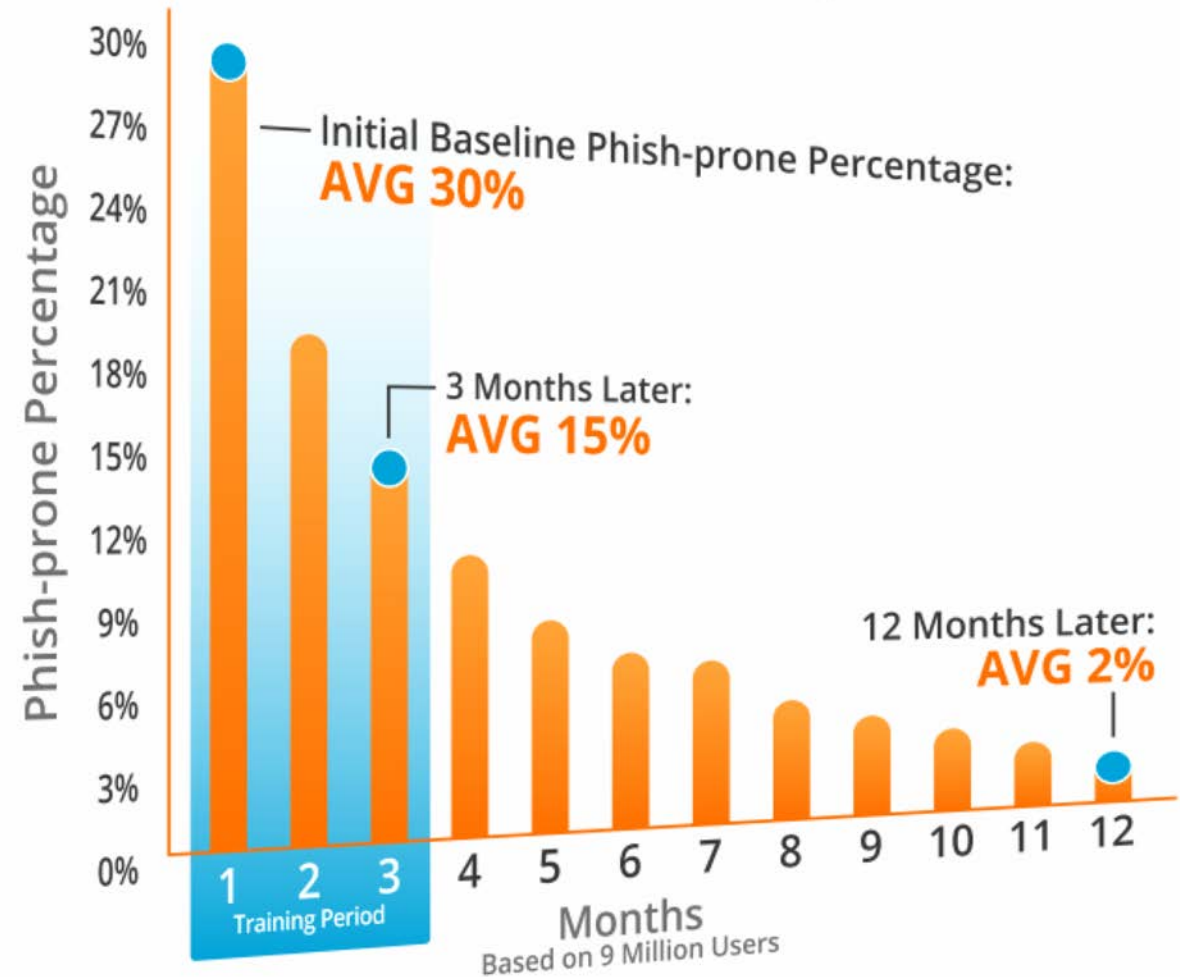
Enterprise-strength reporting, showing stats and graphs for both training and phishing, ready for management. Show the great ROI!



Arm Your Organization

Through combined security awareness and behavior training

Security awareness, coupled with frequent simulated phishing training, will help employees make smarter security decisions, *everyday*



Thank You!

Erich Kron – Security Awareness Advocate

ErichK@KnowBe4.com | @KB4Erich | @ErichKron

Javvad Malik – Security Awareness Advocate

JavvadM@KnowBe4.com | @J4vv4D

