Ai day

It’s Not What You See

It’s What You’re Missing

World Wide Technology

Who’s The Hairy Thing ?

- Technical Solutions Architect (WWT)
- Recovering CISO
- Scottish Warlock
- Geek
- Father
- Hacker
- Tinkerer
- Security Blanket
- @Sidragon1 (LinkedIn)
- Sidragon.substack.com


DON'T
PANIC

AI Will Save Us All...


DO
PANIC

Agenda:

- How did we get here?
- Understanding deepfake technologies and their implications
- Techniques and tools for detecting deepfakes
- Real-world examples and case studies
- Advanced detection methods
- Integration of deepfake detection into existing security frameworks
- Group discussion and knowledge sharing
- Tabletop exercise & Interactive Q&A session

We did this to 5.3 Billion people

Yea, Thanks...(ish)

Our New Reality:

- The ”Google Effect”
  - People rely TOO heavily on readily available information
  - Decreased memory retention and critical thinking skills
  - Decreased attention span, increased anxiety
  - Diminished social skills
  - Increased potential for misinformation and manipulation
  - Decreased privacy, increased misuse of personal information
  - Targeted attacks based on surveillance and available data
  - Market dominance and anti-trust issues
  - Manipulation of search results
  - One single entity's ability to influence the way society thinks

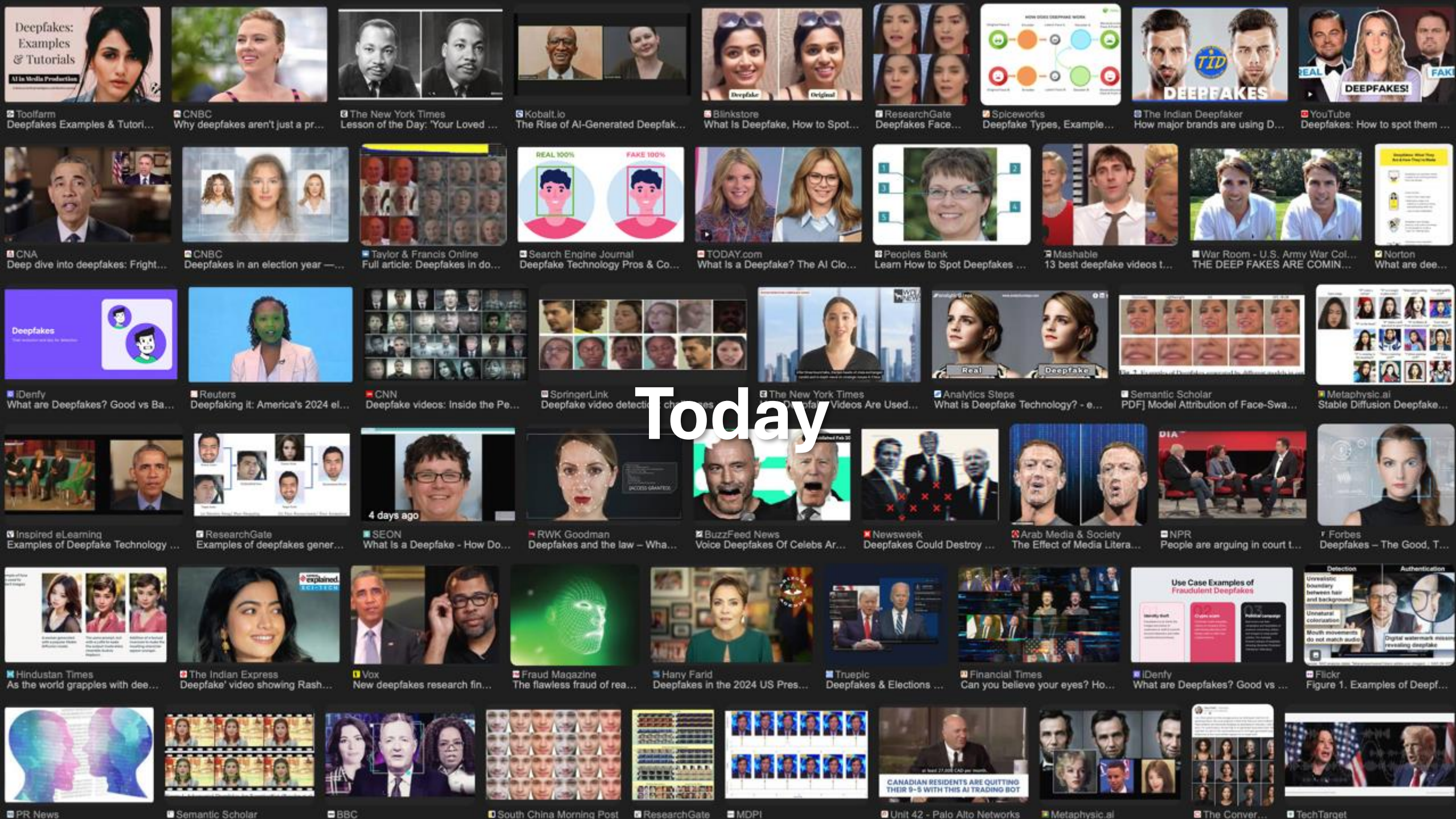
Oh, and they’re eco-bloody-terrorists thanks to the increase in AI…

And Remember…

- We’ve spent 63 years with passwords
- We’ve spent 58 years watching passwords get stolen
- In 1997 we started to securely store passwords (Thanks Bruce)
- Coincidentally we also started to train people around this time
- In 2004 we started cybersecurity awareness month
- Yet, here we are, 20 years old, a global effort of awareness
- AND the top passwords today?
  - 123456
  - 123456789 (folks are getting sneaky with this one)
  - admin
  - qwerty
  - password

Agenda:

- How did we get here?
- Understanding deepfake technologies and their implications 
- Techniques and tools for detecting deepfakes
- Real-world examples and case studies
- Advanced detection methods
- Integration of deepfake detection into existing security frameworks
- Group discussion and knowledge sharing
- Tabletop exercise & Interactive Q&A session


Deepfakes: Examples & Tutorials
AI in Media Production
Toolfarm
Deepfakes Examples & Tutori...
CNBC
Why deepfakes aren't just a pr...
The New York Times
Lesson of the Day: 'Your Loved ...
Kobalt.io
The Rise of AI-Generated Deepfak...
Deepfake
Original
Blinkstore
What Is Deepfake, How to Spot...
ResearchGate
Deepfakes Face...
Spiceworks
Deepfake Types, Example...
TID
DEEPFAKES
The Indian Deepfaker
How major brands are using D...
REAL
FAKE
DEEPFAKES!
YouTube
Deepfakes: How to spot them ...
CNA
Deep dive into deepfakes: Fright...
CNBC
Deepfakes in an election year —...
Taylor & Francis Online
Full article: Deepfakes in do...
REAL 100%
FAKE 100%
Search Engine Journal
Deepfake Technology Pros & Co...
TODAY.com
What Is a Deepfake? The AI Clo...
Peoples Bank
Learn How to Spot Deepfakes ...
Mashable
13 best deepfake videos t...
War Room - U.S. Army War Col...
THE DEEP FAKES ARE COMIN...
Norton
What are dee...
Deepfakes
iDenfy
What are Deepfakes? Good vs Ba...
Reuters
Deepfaking it: America's 2024 el...
CNN
Deepfake videos: Inside the Pe...
SpringerLink
Deepfake video detecti
The New York Times
Videos Are Used...
Real
Deepfake
Analytics Steps
What is Deepfake Technology? - e...
Semantic Scholar
PDF] Model Attribution of Face-Swa...
Metaphysic.ai
Stable Diffusion Deepfake...
Today
Inspired eLearning
Examples of Deepfake Technology ...
ResearchGate
Examples of deepfakes gener...
4 days ago
SEON
What Is a Deepfake - How Do...
RWK Goodman
Deepfakes and the law – Wha...
BuzzFeed News
Voice Deepfakes Of Celebs Ar...
Newsweek
Deepfakes Could Destroy ...
Arab Media & Society
The Effect of Media Litera...
NPR
People are arguing in court t...
Forbes
Deepfakes – The Good, T...
Hindustan Times
As the world grapples with dee...
explained
The Indian Express
Deepfake' video showing Rash...
Vox
New deepfakes research fin...
Fraud Magazine
The flawless fraud of rea...
Hany Farid
Deepfakes in the 2024 US Pres...
Truepic
Deepfakes & Elections ...
Financial Times
Can you believe your eyes? Ho...
Use Case Examples of Fraudulent Deepfakes
iDenfy
What are Deepfakes? Good vs ...
Detection
Authentication
Unrealistic boundary between hair and background
Unnatural colorization
Mouth movements do not match audio
Digital watermark missing revealing deepfake
Flickr
Figure 1. Examples of Deepf...
PR News
Semantic Scholar
BBC
South China Morning Post
ResearchGate
MDPI
CANADIAN RESIDENTS ARE QUITTING THEIR 9-5 WITH THIS AI TRADING BOT
Unit 42 - Palo Alto Networks
Metaphysic.ai
The Conver...
TechTarget

Celebrity AI Voices & Parody
★★★★☆ 565
Get
In-App Purchases

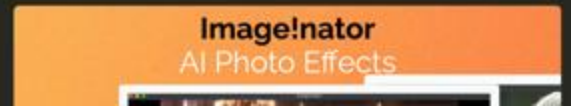
Live Auto Pitch Vocoder
★★★★½ 48

AI Face Swap App Celeb Replace
Copy, Change & Morph Photo
★★★★☆ 107
Get

Editor Pro
Get

Face Kit - AI Face Editor
Age Changer, Lip & Hair Styles
★★★☆☆ 3
Get

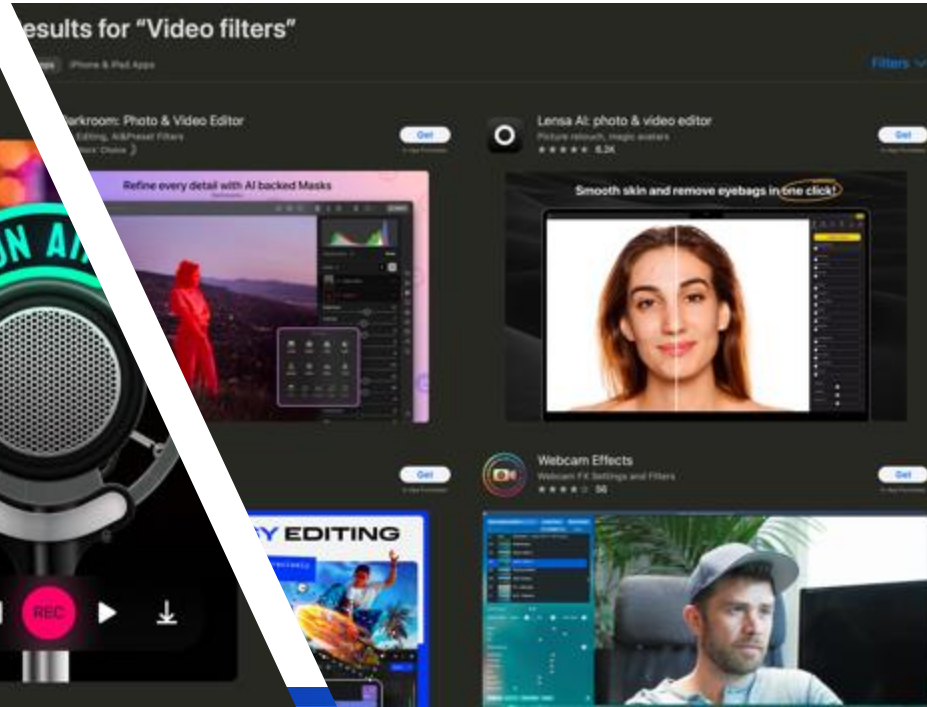
Imagenator: AI Image Generator
Photo Face Filters & Face Swap
★★★☆☆ 4
Get



Voice.ai: Voice Changer
Parody AI Voice Clones for You
★☆☆☆☆ 243
Get
In-App Purchases

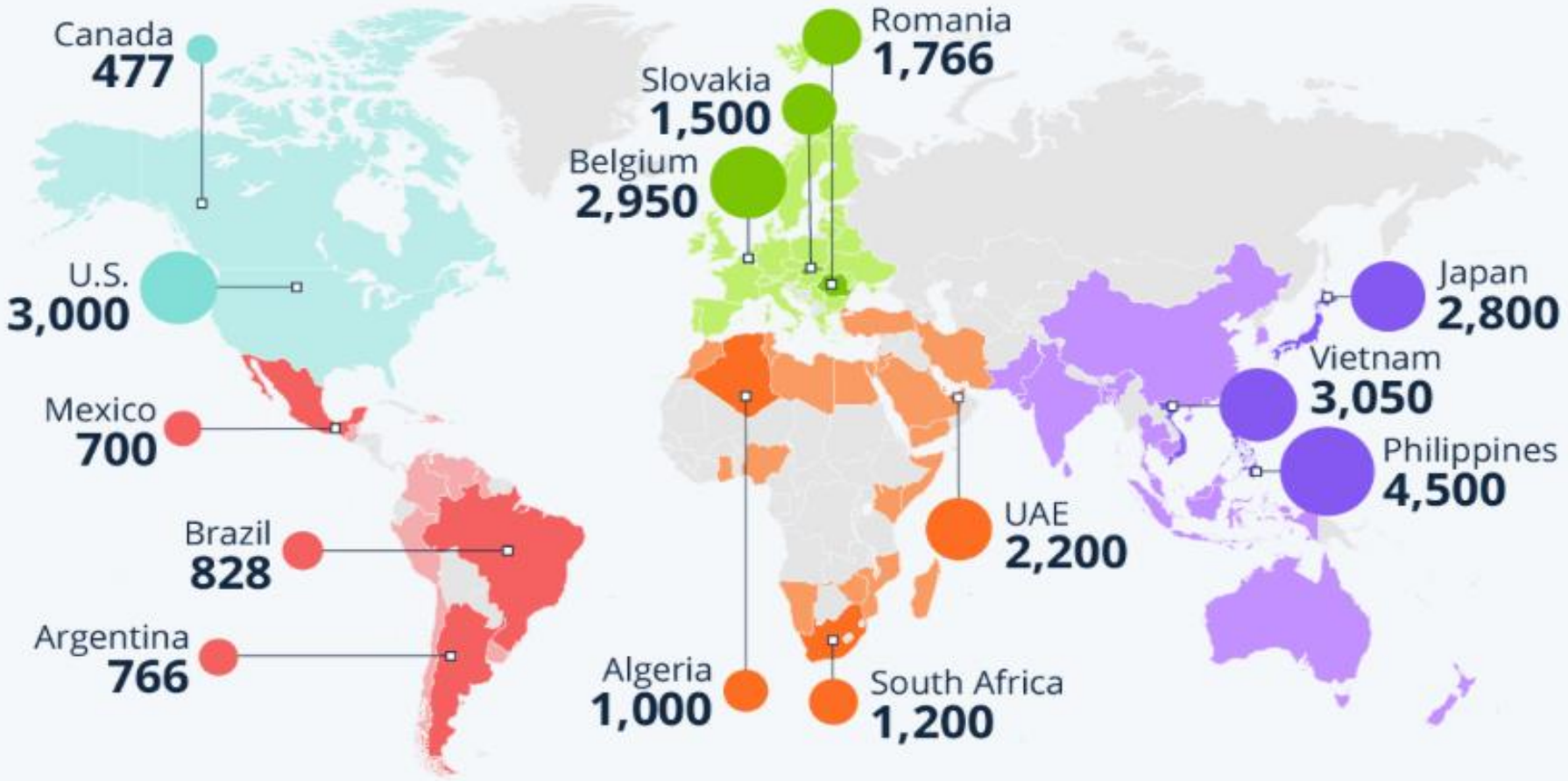
Microphone Live-Voice Recorder
Auto vocal tune, podcast mic
★★★★★ 2

esults for “Video filters”



The Explosive Growth of AI-Powered Fraud

Countries per region with biggest increases in deepfake-specific fraud cases from 2022 to 2023 (in %)*



By The Numbers: 2023:

- Consumers lost over $10 billion to scams and fraud, according to the Federal Trade Commission (FTC). 14% increase from 2022, when consumers lost $9 billion.
- The FBI also found that the US lost a record $12.5 billion to online fraud. 22% increase from the previous year.
- The FTC, accounting for unreported losses, estimated fraudsters stole $137 billion, including $48 billion from older adults.
- AARP noted that Americans over 60 lose $28.3 billion each year to fraud.

Numbers can't be agreed upon; however the message is simple, we are not winning the war on fraud, scams, or attacks against the general population.


HOLLYWOOD
Home
Video
Marketplace
Dating
Notifications
Menu

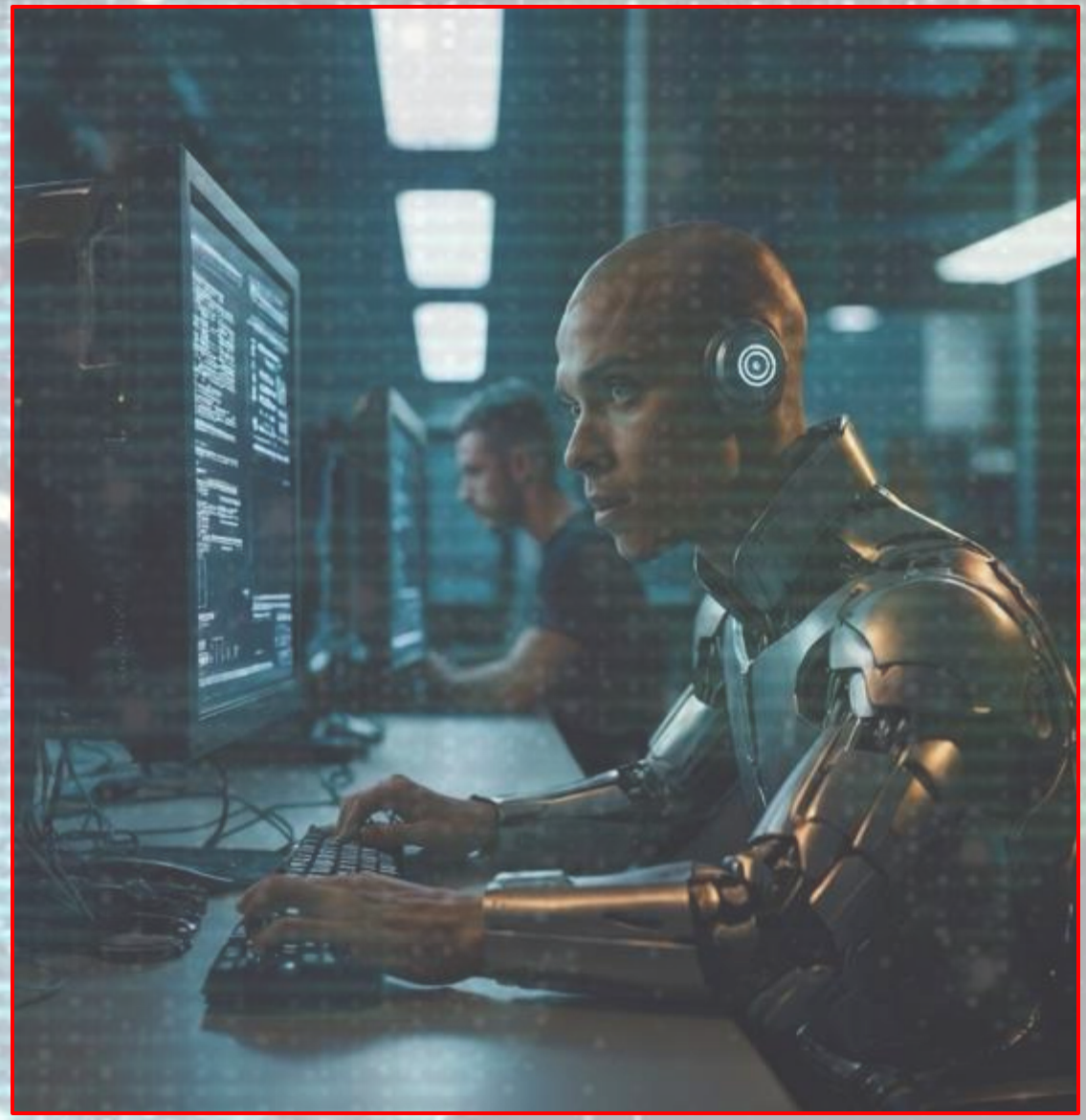
Donate

Or

Challenge?

- Unnatural Facial Movements and Expressions
  - Blinking and lip sync
- Inconsistent Lighting and Shadows
  - Watch for shadow movement etc.
- Reflections
  - Glasses or jewelry, reflective light…
- Unnatural Eye Movements
  - Eyes should continually move, often with body
- Hair and Teeth
  - Does the hair seem natural, are the teeth TOO perfect?
- Audio-Visual Mismatch
  - Back to watching for lip sync, sounds etc.
- Background Anomalies
  - Get folks to show themselves in a natural scene…
- Resolution Discrepancies

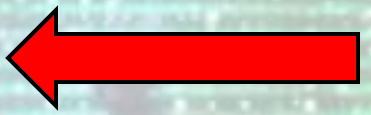
  Look at the picture, then away at reality, same/different?
- Edge Quality
  - We're not perfect AND we're not blurry… are they?
- Unusual Artifact
  - Good old head movement, get your target to move around!
- Color Shifts
  - We're not chameleons, neither should your video be…
- Content Inconsistencies
  - Language used, word choices, flow of sentences
- Mismatched Statements
  - As above, but check what's being said vs. expectations vs. reality



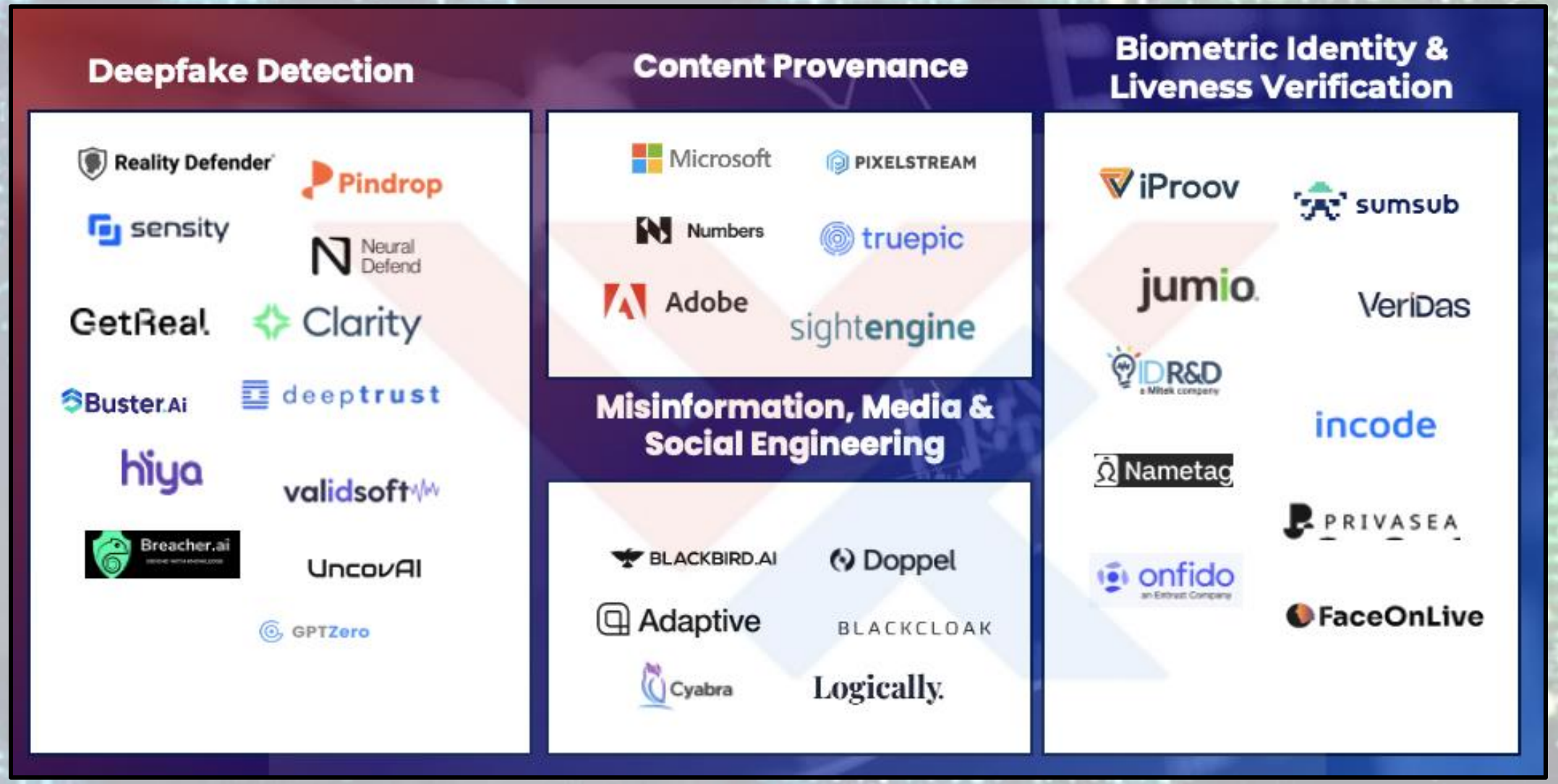
Jim Kavanaugh
WWT CEO

Kate Kuehn
WWT Squirrel Herder

Agenda:

- How did we get here?
- Understanding deepfake technologies and their implications
- Techniques and tools for detecting deepfakes 
- Real-world examples and case studies
- Advanced detection methods
- Integration of deepfake detection into existing security frameworks
- Group discussion and knowledge sharing
- Tabletop exercise & Interactive Q&A session

Technology is a tool: It’s up to each of us to decide how we use it.


Deepfake Detection
Reality Defender
Pindrop
sensity
Neural Defend
GetReal
Clarity
Buster.Ai
deeptrust
hiya
validsoft
Breacher.ai
UncovAI
GPTZero
Content Provenance
Microsoft
PIXELSTREAM
Numbers
truepic
Adobe
sightengine
Misinformation, Media & Social Engineering
BLACKBIRD.AI
Doppel
Adaptive
BLACKCLOAK
Cyabra
Logically.
Biometric Identity & Liveness Verification
iProov
sumsub
jumio
VeriDas
IDR&D
incode
Nametag
PRIVASEA
onfido
FaceOnLive

76 vendors spread across the following:

- Audio
- Video
- Data/Text
- Social Media
- Adult Extortion
- Generation/Training
- Identity

Maturity of organization:

- Some seed stage, great ideas
- Some initial design partners
- Some paying clients
- Most singular focus
- 3-4 looking to solve multiple areas
- 4-5 worth looking at now
- 3-4 worth watching for next 12 months

Observed issues and challenges:

- Most are point solutions, not integrated
- Some rely on browser interactions
- None can collect everything
- Most train on primarily synthetic data
- Most are cloud based, limited on-prem

AI day

Criteria?

- Advanced Machine Learning Techniques
  - CNN (Convoluted Neural Network, image recognition at a heavy load)
  - RNN (Recurrent Neural Network, text, speech, time series, good reflective)
  - GAN (Generative Adversarial Network, multi-network/competitive learning)
- Multimodal Approaches
  - Video, audio, metadata, contextual analysis
  - Threat intelligence
  - Integration with identity management platforms
- Explainable AI (XAI)
  - How and why, something is flagged
  - Transparency into the decision-making process
  - Accountability
- Near Real-time Detection and Prevention
  - Rapid identification and remediation
  - Mesh information sharing across platforms
  - Misinformation and disinformation aware

Why?

- Factors under consideration (and why)
  - Accuracy:
    - Are they able to do what they're promising?
  - Cross-validation:
    - How well do they train, AND on what breakdown of data types?
  - Confusion matrix:
    - How well does it watch itself for mistakes (and corrects)
  - Speed:
    - Near-real time for realities sakes, or latency?
  - Cost:
    - GPU costs, cycles, and ability to scale accordingly
  - Ease-of-use:
    - AND how it fits into an overall solution set, integrates, etc.
  - Benchmarking?
    - What and how and how often are they performing checks and balances?
  - ROC Curve (Receiver Operating Characteristics)
    - True positive vs. false positive rates

What DOES Analysis Look Like?

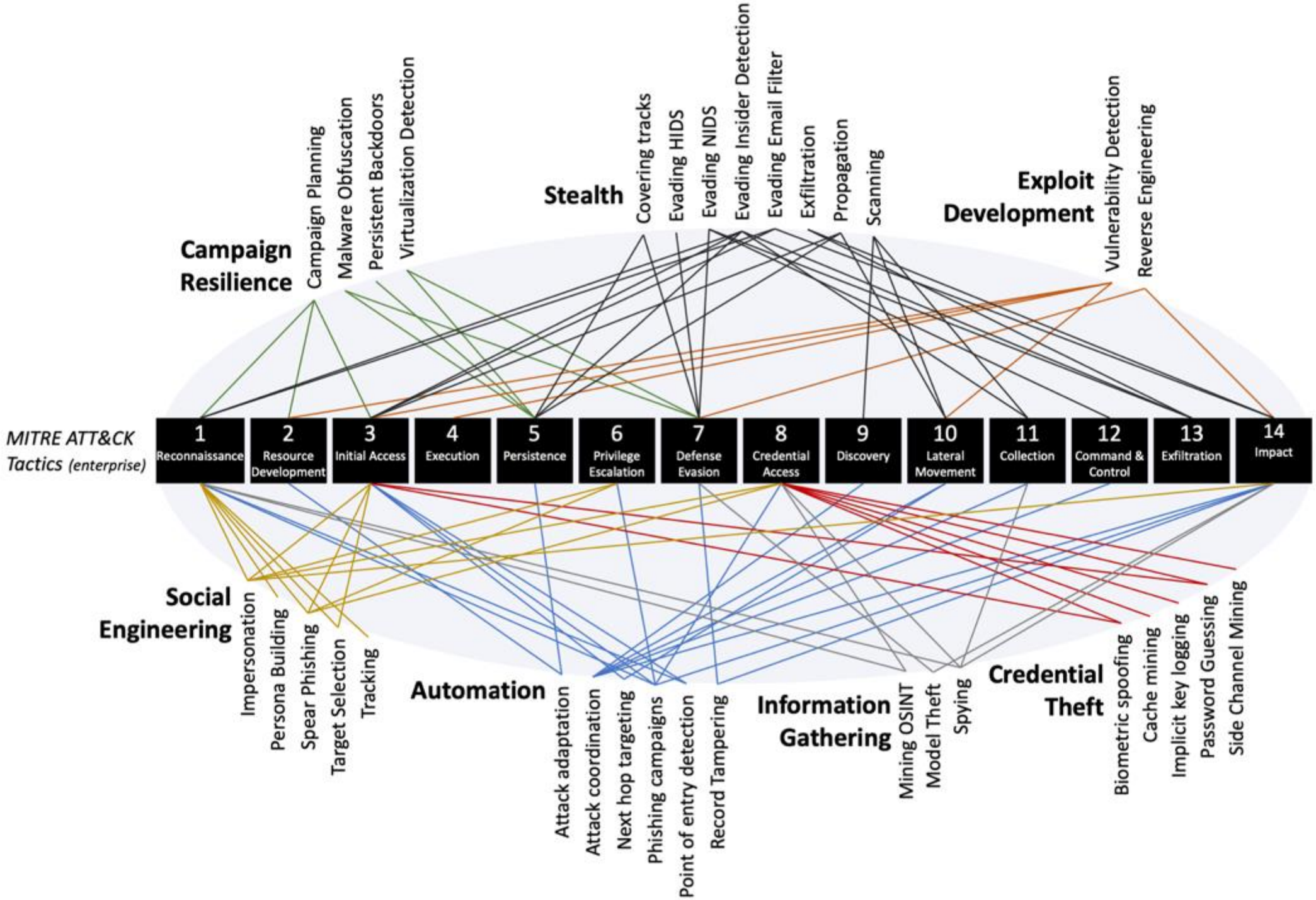
- Vendor “ACME”:
  - **Risks:**
    - Viability of technology down to the mathematical scale
    - Ability to deliver against proposed technical solution
    - Ability to scale to enterprise levels
    - Delivery of roadmap on/near time
    - Ability to recruit/retain necessary talent to support
    - Capability to sustain financial model, overall organization stability
  - **Assumptions:**
    - None
  - **Issues**
    - Communication cycles to maintain risk register
    - Ability to deliver technology in a manner we can build on
    - Ability to integrate into a total solution set
    - Focus on longer term opportunities at the expense of short-term gains
  - **Dependencies**
    - Financial stability
    - Effective leadership and management
    - Efficient project management and technical drive
    - Focused prioritization
    - Funding relationship (WWT to work with investor to ensure longevity etc.)

Agenda:

- How did we get here?
- Understanding deepfake technologies and their implications
- Techniques and tools for detecting deepfakes
- Real-world examples and case studies 
- Advanced detection methods
- Integration of deepfake detection into existing security frameworks
- Group discussion and knowledge sharing
- Tabletop exercise & Interactive Q&A session

Adversarial usages of AI:

- There are three primary motivations for an adversary to use AI:
  - Coverage (the ability to see more, do more, scan more, and target more)
  - Speed (being able to turn data into information then intelligence faster)
  - Success (the ability to rapidly determine the defenses and evasion techniques)
- AI introduces new threats to organizations. Examples include:
  - Poisoning of machine learning models
  - Theft of credentials through side channel analysis (data harvesting)
  - Targeting of proprietary training datasets.
- Adversaries are employing a plethora of offensive capabilities against organizations, categorized into several areas:
  - Automation
  - Campaign resilience
  - Credential theft exploit development
  - Information gathering
  - Social engineering
  - Stealth


MITRE ATT&CK
Tactics (enterprise)
1 Reconnaissance
2 Resource Development
3 Initial Access
4 Execution
5 Persistence
6 Privilege Escalation
7 Defense Evasion
8 Credential Access
9 Discovery
10 Lateral Movement
11 Collection
12 Command & Control
13 Exfiltration
14 Impact
Campaign Resilience
Campaign Planning
Malware Obfuscation
Persistent Backdoors
Virtualization Detection
Stealth
Covering tracks
Evading HIDS
Evading NIDS
Evading Insider Detection
Evading Email Filter
Exfiltration
Propagation
Scanning
Exploit Development
Vulnerability Detection
Reverse Engineering
Social Engineering
Impersonation
Persona Building
Spear Phishing
Target Selection
Tracking
Automation
Attack adaptation
Attack coordination
Next hop targeting
Phishing campaigns
Point of entry detection
Record Tampering
Information Gathering
Mining OSINT
Model Theft
Spying
Credential Theft
Biometric spoofing
Cache mining
Implicit key logging
Password Guessing
Side Channel Mining

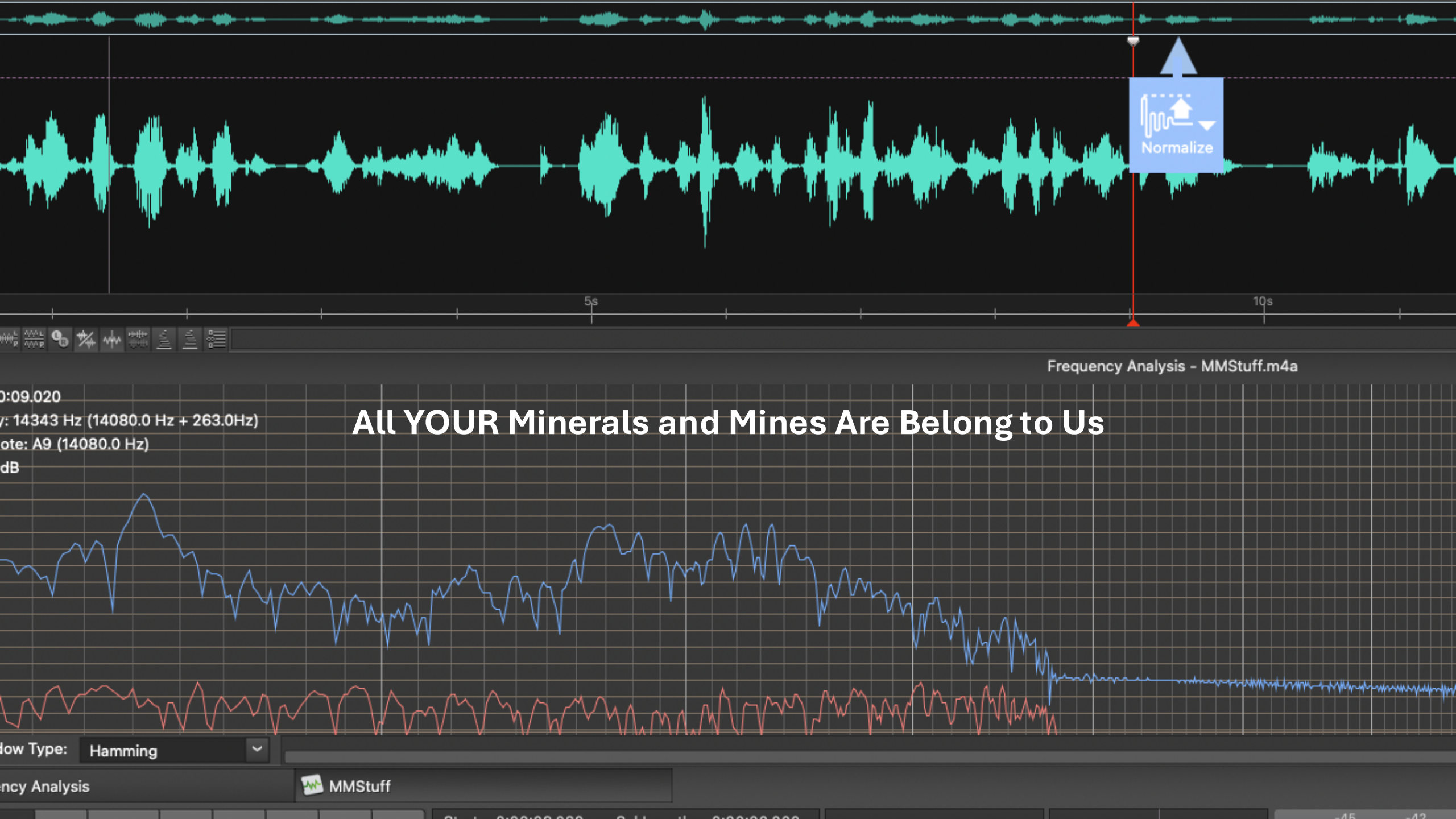
Case Studies:

- Financial Institution
  - Wealth Management – Validation of HNW clients
  - Call Center – Reducing the 20% inbound fake calls
  - False Data (Pump/Dump/Fraud) – Stock manipulation

- Pharmaceutical Company
  - Executive Protection – Combined with threat intelligence

- Government
  - Inbound screening – Reducing the inbound call volume etc.
  - Intelligence analysis – What IS real?
  - Human trafficking and abuse analysis – Rapid analysis and targeting

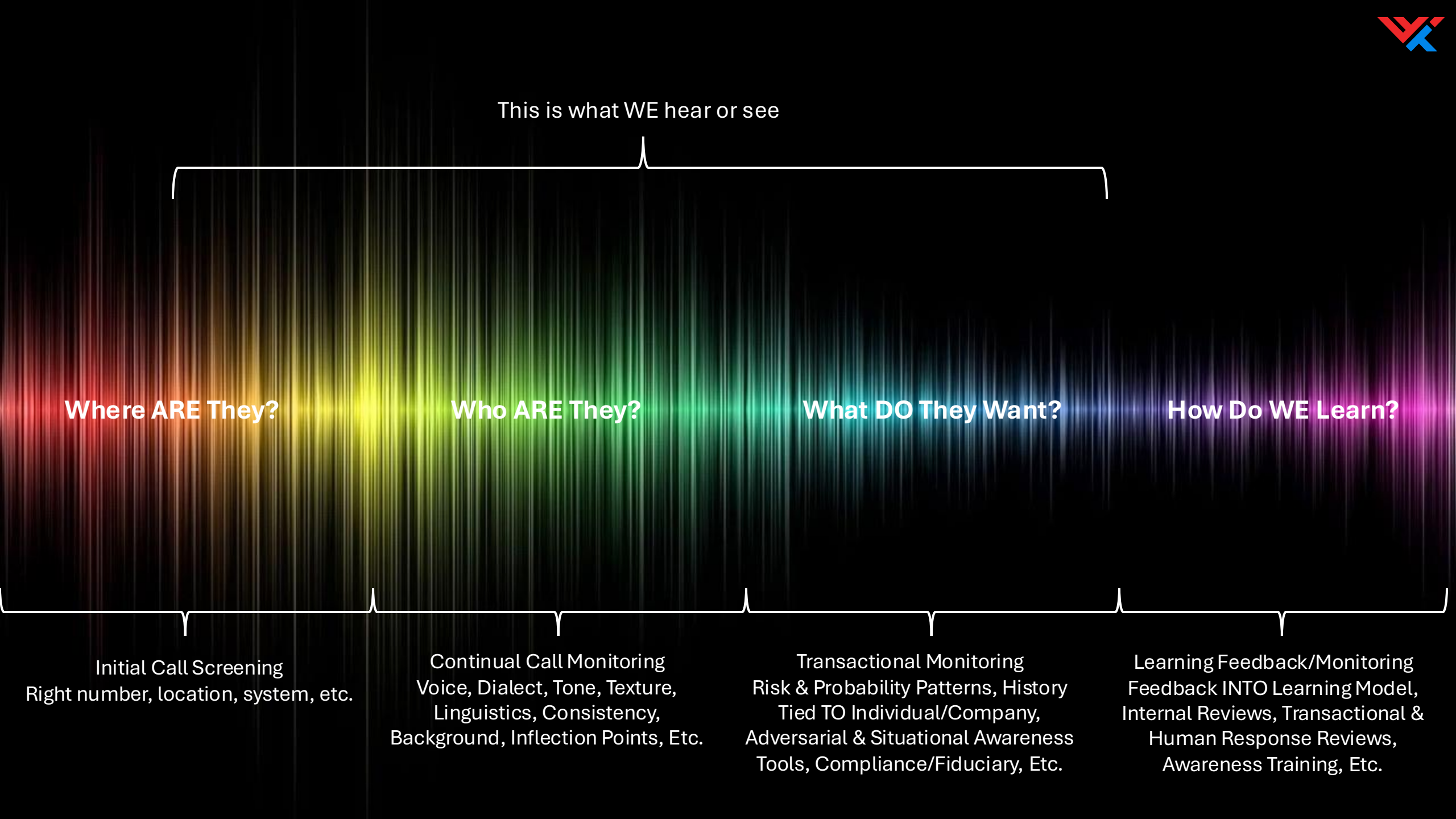
- Healthcare
  - Patient Identities/Records – Only handing data to those who NEED it

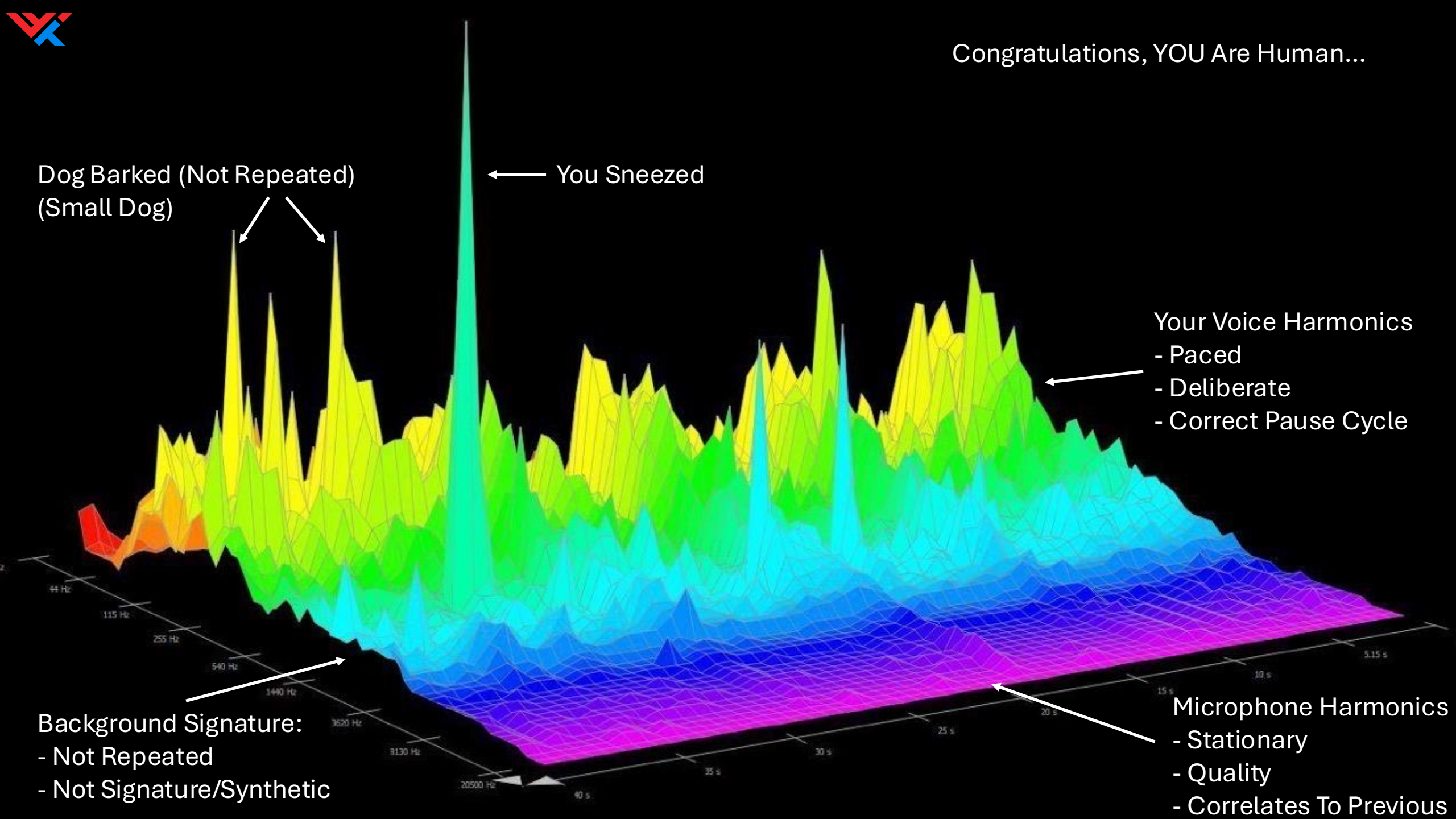

Normalize
5s
10s
Frequency Analysis - MMStuff.m4a
0:09.020
y: 14343 Hz (14080.0 Hz + 263.0Hz)
ote: A9 (14080.0 Hz)
dB
All YOUR Minerals and Mines Are Belong to Us
dow Type:
Hamming
ncy Analysis
MMStuff

Agenda:

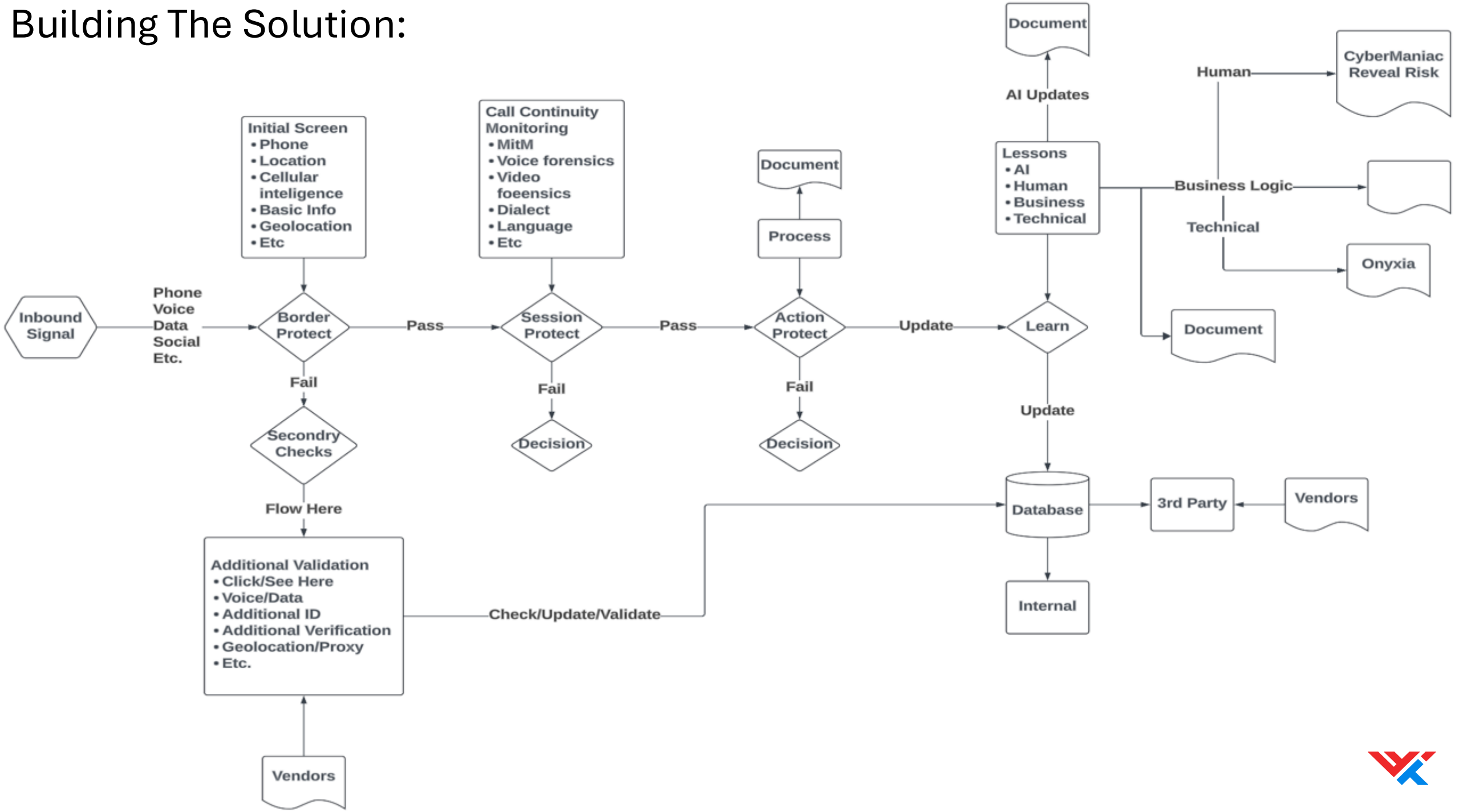
- How did we get here?
- Understanding deepfake technologies and their implications
- Techniques and tools for detecting deepfakes
- Real-world examples and case studies
- Advanced detection methods
- Integration of deepfake detection into existing security frameworks
- Group discussion and knowledge sharing
- Tabletop exercise & Interactive Q&A session

Point solutions won’t fix the problem, WE must work on a framework (Lego etc.)


This is what WE hear or see
Where ARE They?
Who ARE They?
What DO They Want?
How Do WE Learn?
Initial Call Screening
Right number, location, system, etc.
Continual Call Monitoring
Voice, Dialect, Tone, Texture,
Linguistics, Consistency,
Background, Inflection Points, Etc.
Transactional Monitoring
Risk & Probability Patterns, History
Tied TO Individual/Company,
Adversarial & Situational Awareness
Tools, Compliance/Fiduciary, Etc.
Learning Feedback/Monitoring
Feedback INTO Learning Model,
Internal Reviews, Transactional &
Human Response Reviews,
Awareness Training, Etc.

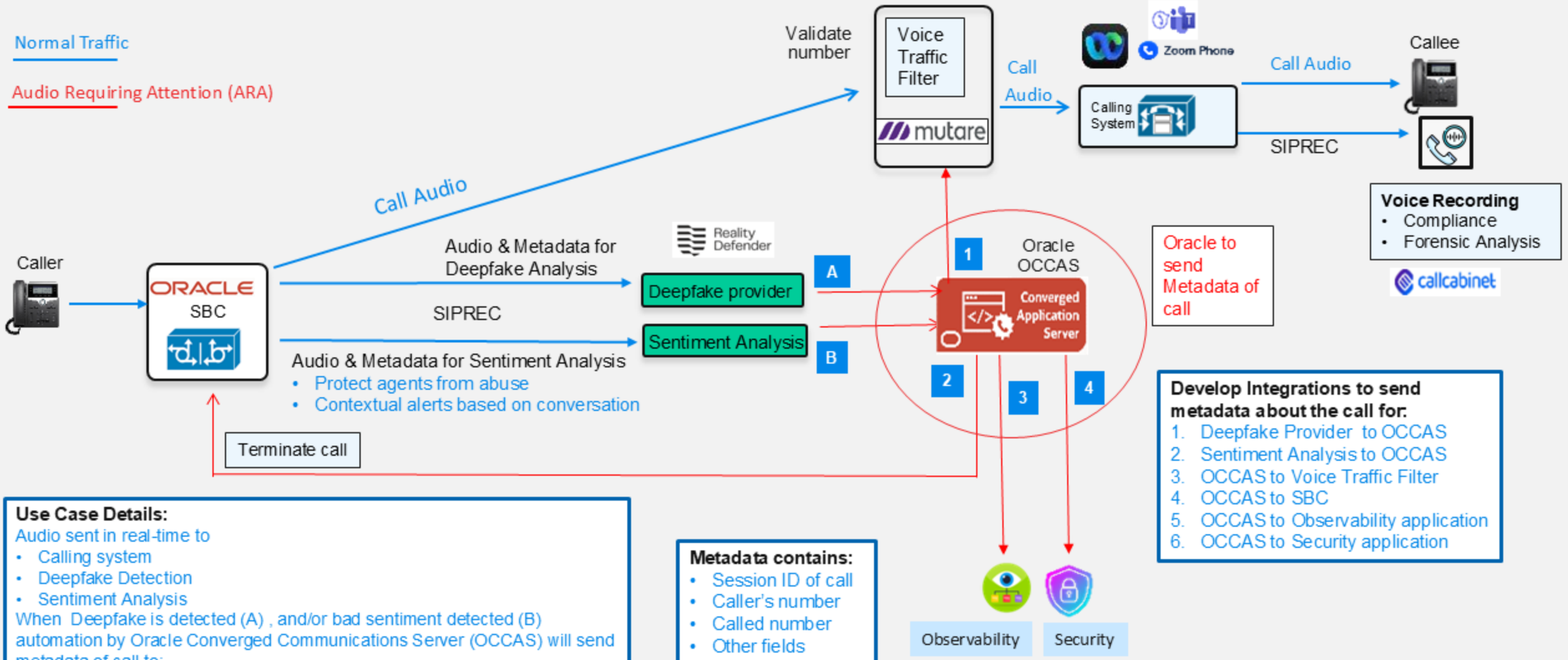

Congratulations, YOU Are Human...
Dog Barked (Not Repeated)
(Small Dog)
You Sneezed
Your Voice Harmonics
- Paced
- Deliberate
- Correct Pause Cycle
Background Signature:
- Not Repeated
- Not Signature/Synthetic
Microphone Harmonics
- Stationary
- Quality
- Correlates To Previous
44 Hz
115 Hz
255 Hz
540 Hz
1440 Hz
3620 Hz
8130 Hz
20500 Hz
5.15 s
10 s
15 s
20 s
25 s
30 s
35 s
40 s

Building The Solution:



How’s this look in our ATC Lab?

Voice Security - Use Case: Inbound Calling with Call Flow & Integration Requirements



Use Case Details:

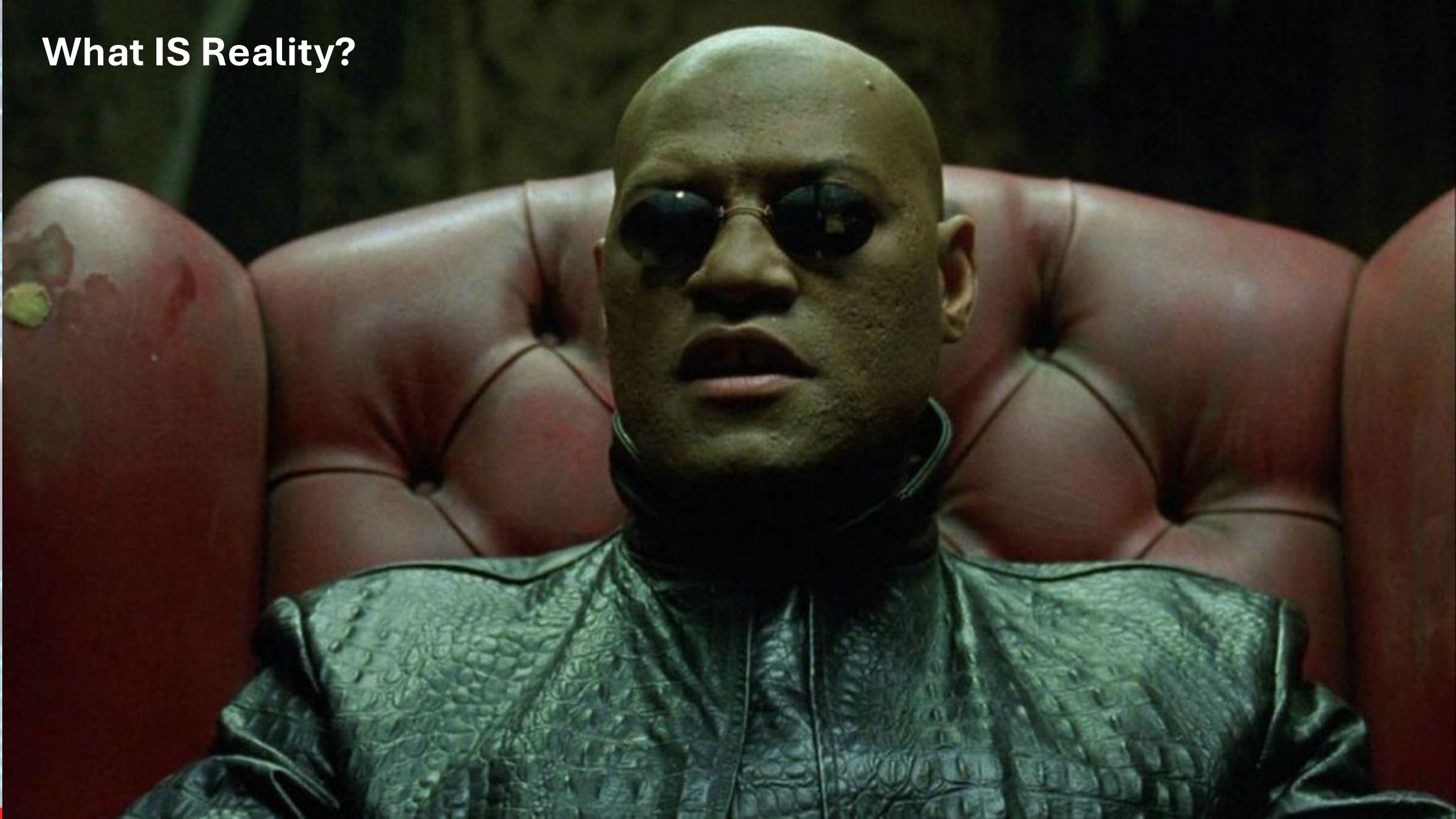
Audio sent in real-time to
- Calling system
- Deepfake Detection
- Sentiment Analysis

When Deepfake is detected (A) , and/or bad sentiment detected (B) automation by Oracle Converged Communications Server (OCCAS) will send metadata of call to:
- Voice Traffic Filter to block future calls from this number (1)
- SBC to terminate call (2)
- Alert to Observability Application (3)
- Alert to Security Application (4)

Agenda:

- How did we get here?
- Understanding deepfake technologies and their implications
- Techniques and tools for detecting deepfakes
- Real-world examples and case studies
- Advanced detection methods
- Integration of deepfake detection into existing security frameworks
- Group discussion and knowledge sharing
- Tabletop exercise & Interactive Q&A session

What IS Reality?



Let’s Talk:

- How do you currently consume signals?
- How do you manage your signals?
- How (if anything) do you manage identities?
- How (if at all) do you manage the identities of your customers or clients?
- How do you determine reality from fake?
- How do you train your staff, and those around you, to think more, trust less?
- What do you know about the deepfake landscape now and tomorrow?
- What controls and monitoring do you have in place?
- How would you know if your CEO’s been a victim of deepfakes?
- How would your staff know if the CEO’s been victimized through deepfakes?
- How are you capture all this information in a manageable and reportable manner?

What are we doing about it?

- The Deepfake, AdversarialAI, and overall disinformation aspect of technology can be broken down into the following areas:
  - **Signal:**
    - Everything is a signal, enrich it, give it context, and turn it into intelligence
  - **Identity:**
    - Who's on the other end of the signal (text, phone, email, video, messenger, etc.)
  - **Analysis:**
    - Are they real, what's the sentiment, what do they want, and why?
  - **Education:**
    - Awareness and media literacy are cornerstones to understanding the next generation of empowered social engineering, and manipulative attacks.
  - **Management:**
    - Understanding is one part, alerting, remediating, and moving to a more proactive and preventative architecture is key to tackling and reporting out.

Question More

Trust Less

Right People

Right Place

Right Time

Right Focus

Right Motives

Want vs. Need

Convergence of Identities

Decentralized biometrics

Liveness Detection

Smart ID scanners

Establishing Chains of Trust

Multi-Modal Biometrics

Self-Sovereign Identity

Reinvent the Physical ID Card

Agenda:

Wrapping it all up!

Unplug your toaster AND shoot your TV...


Let’s Talk Before It’s Too Late...