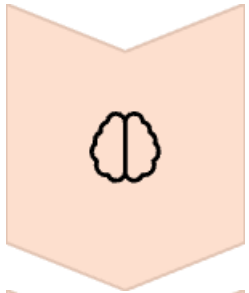
KnowBe4
Human error. Conquered.

The Sourdough of Trust

Why Humans Are Cybersecurity's Secret Ingredient

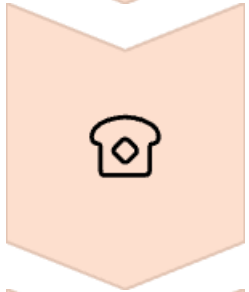
Javvad Malik
Lead Security Awareness Advocate

Today's Agenda



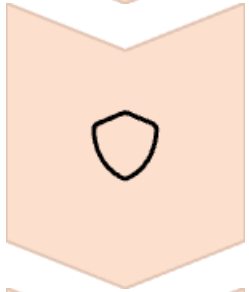
The Human Element

What does that even mean in cybersecurity?



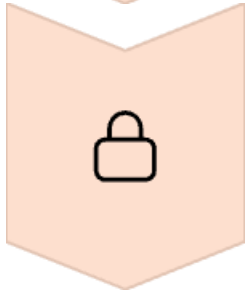
The Sourdough of Trust

How our brains are wired to be tricked



The Psychology of AI

Who wins the manipulation game?



Going DEEP

An approach to managing human risk

The Universal Language of Trust

Everyday Trust

We trust restaurant recommendations, alarm clocks, and colleagues daily. colleagues daily.

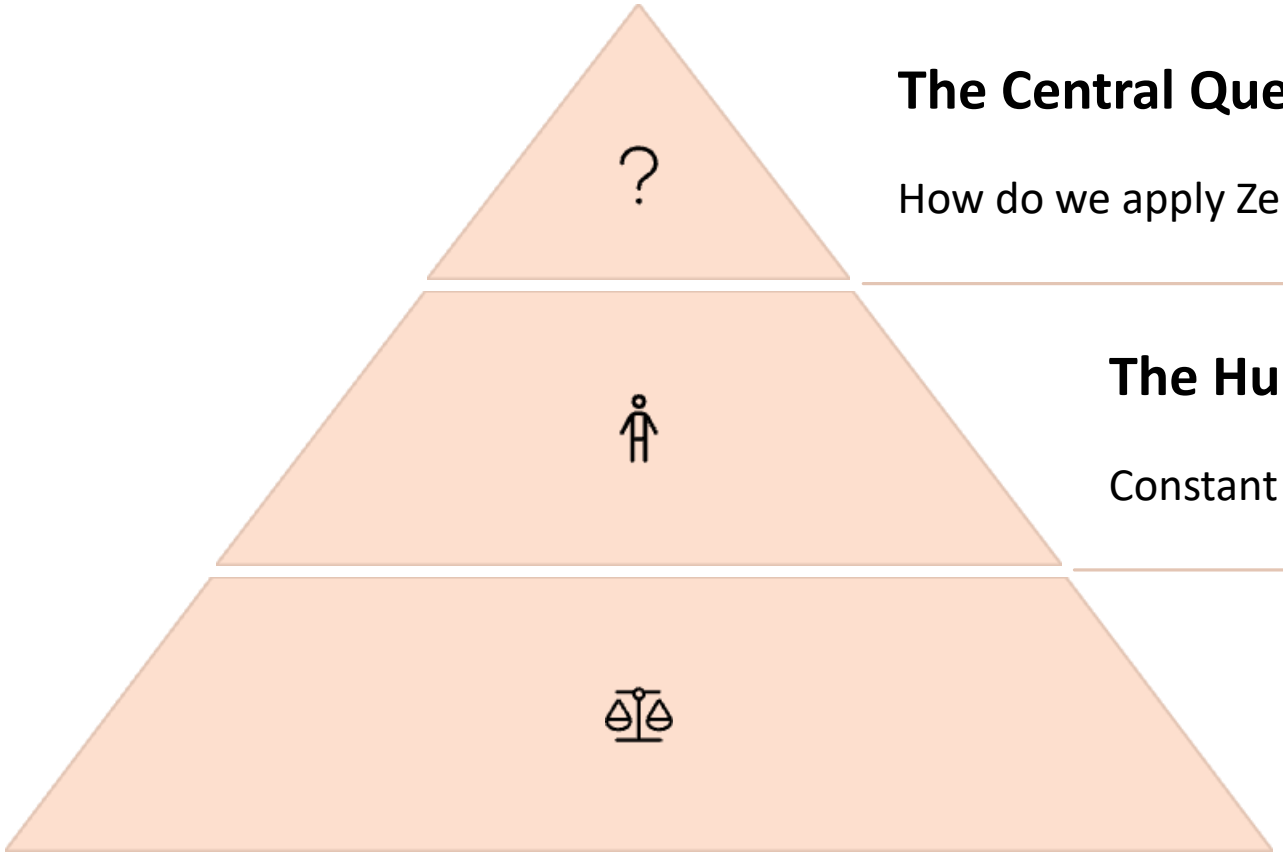
Digital Complexity

This fundamental trait becomes a complex vulnerability in the digital realm. digital realm.

Human Involvement

Most breaches involve human actions, not just technical vulnerabilities.

The Trust Dilemma



The Central Question

How do we apply Zero Trust to human interactions?

The Human Need

Constant suspicion is unsustainable for human society

The Challenge

Finding balance between trust and security

The Digital Trust Challenge

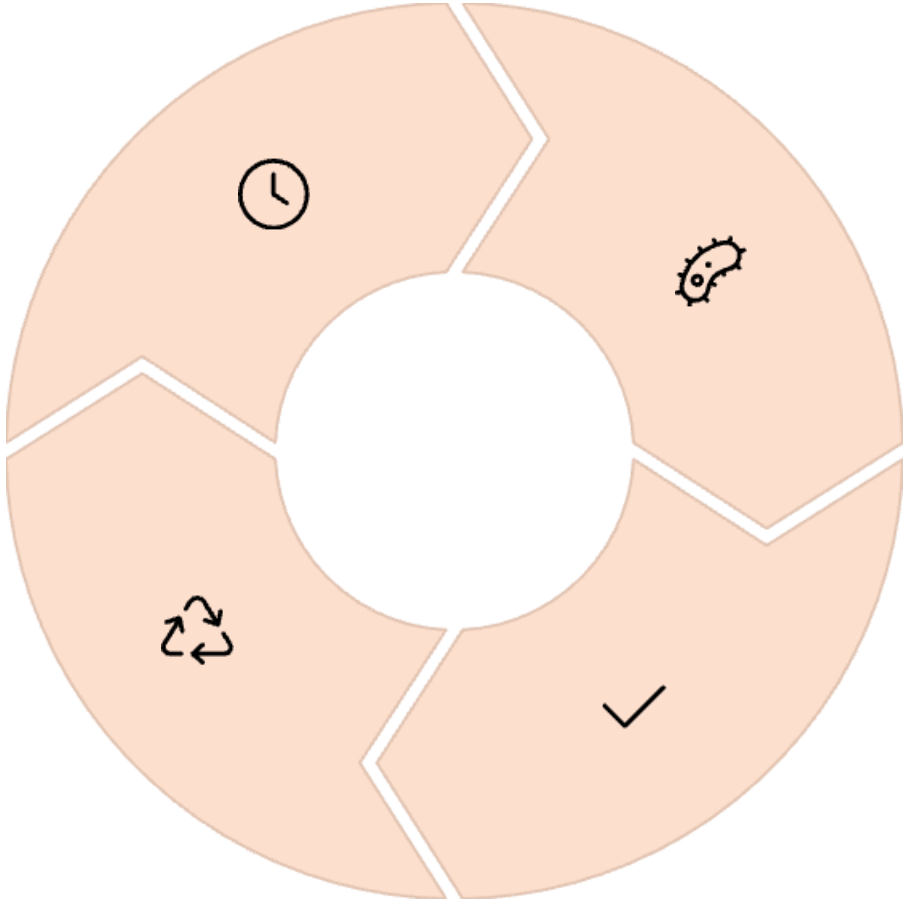
Dorothy Williams' Sourdough Group

Me connections.
Rea assion.

Trust Is Like Making Sourdough

Takes Time

Trust requires the right conditions and develops gradually



Has Ingredients

Complex elements combine to create create the foundation

Needs Maintenance

Must be continually fed and cared for for

Rarely Questioned

Once established, people rarely examine what's actually in it

Trust Trigger #1: Familiarity

Parasocial Bonding

Forming emotional connections with people we've never met (celebrity culture)

The "IT Guy Next Door" Effect

Trusting someone who sounds familiar and uses internal jargon

2020 Twitter Hack

Compromised Twitter by posing as internal IT support

Trust Trigger #2: Authority

The Instagram Effect

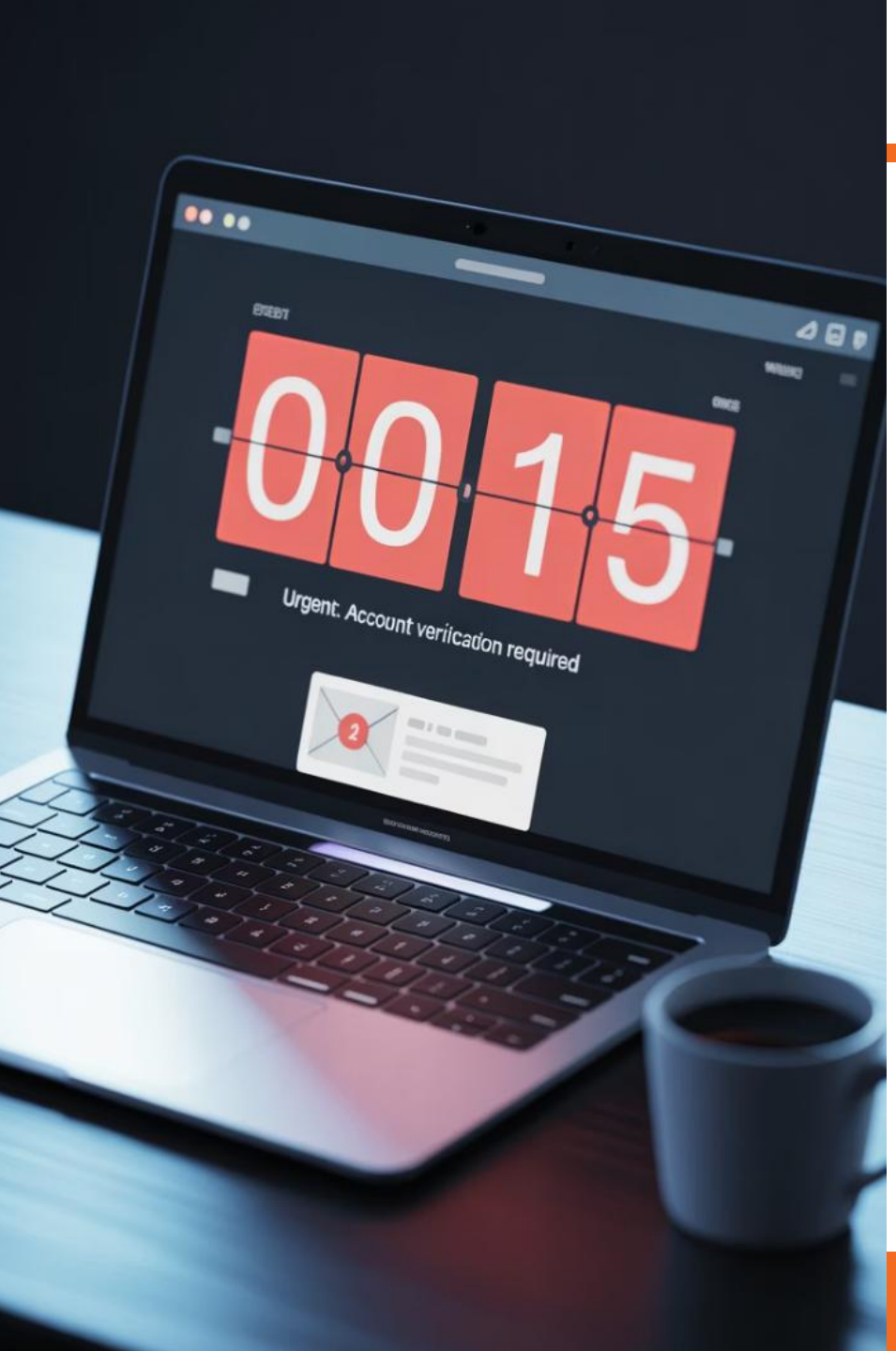
When enough people endorse something as good, we stop questioning questioning its quality.

Brand Trust Transfer

We automatically extend trust to anyone claiming to represent a trusted trusted brand.

Technical Intimidation

Exploiting our feelings of insufficient technical knowledge.



Trust Trigger #3: Urgency

Hyperbolic Discounting

The Pizza Principle

FOMO

Booking.com

The Tinder Paradox

73%

Password Reuse

Of people use the same password across multiple accounts

82%

Identity Theft Fear

Express worry about identity theft despite password reuse

2.5s

Email Verification

Average time spent verifying a sender's email email address

45m

Dating Investigation Investigation

Average time spent investigating a potential potential date's Instagram Instagram

The North Korean IT Worker

The Job Opening

Principal Software Engineer position at KnowBe4

The Interview Process

Four video conferences, background checks, reference verification

The Attack

New Mac workstation immediately began loading malware

The Investigation

Revealed fake IT worker from North Korea with AI-generated photo

Remember Dorothy Williams

Dorothy Williams Isn't Real

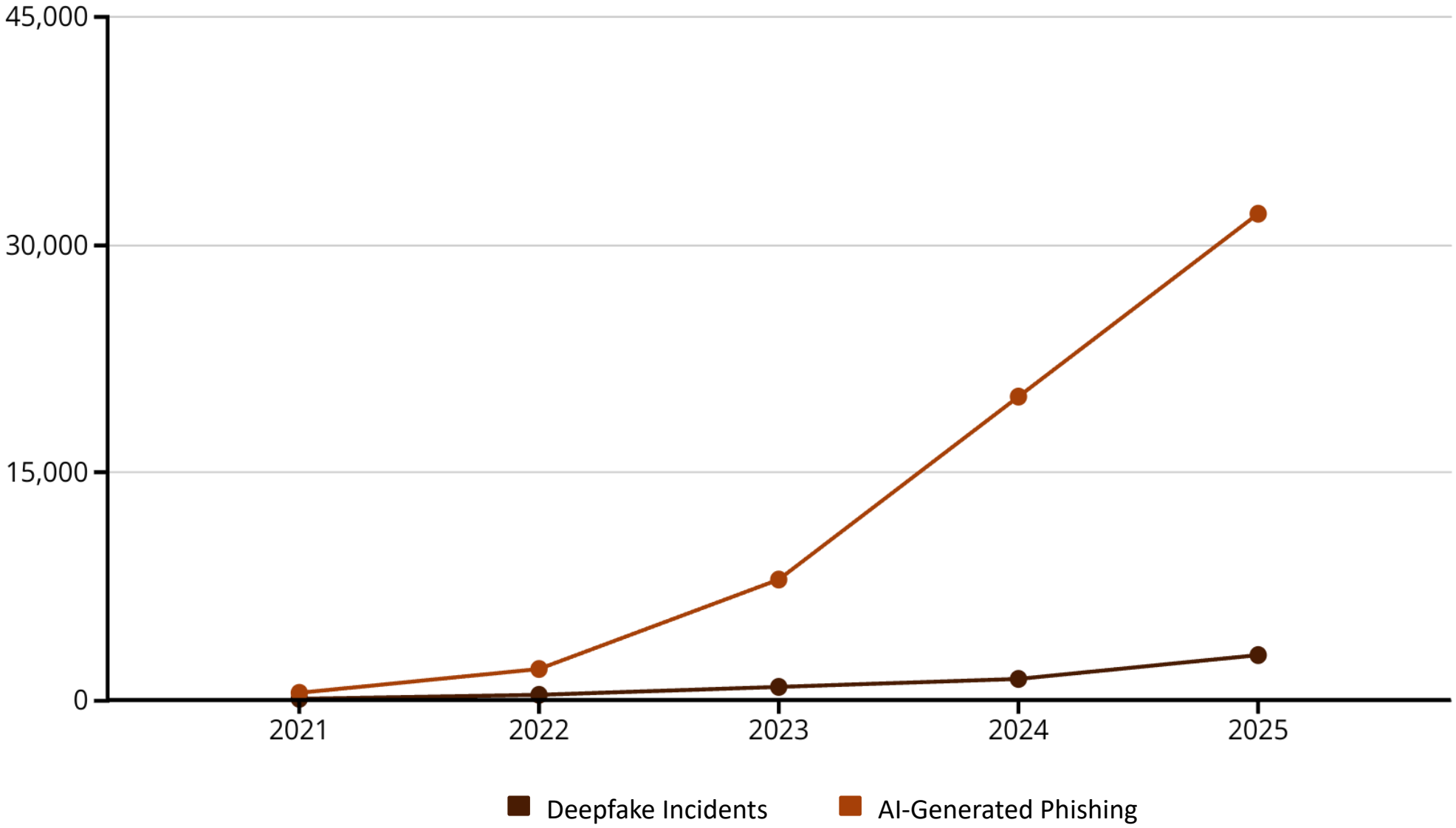
"Dorothy" was an AI-generated personality created to test test engagement patterns in Facebook groups.

Her sourdough recipe was an algorithmically optimised combination of several recipes.

Even her heartwarming stories about her grandmother were were generated by GPT-4.

For eight months, nearly 50,000 people trusted, learned from, and emotionally connected with Dorothy

AI is Turbocharging Criminals



Source: Pindrop's 2025 Voice Intelligence & Security Report

The Rise of Deepfakes and Voice Cloning

Cloning

+1,300%

Deepfake Fraud

Surge in deepfake fraud in 2024

+155%

Deepfaked Calls

Projected increase in deepfaked calls in 2025

3s

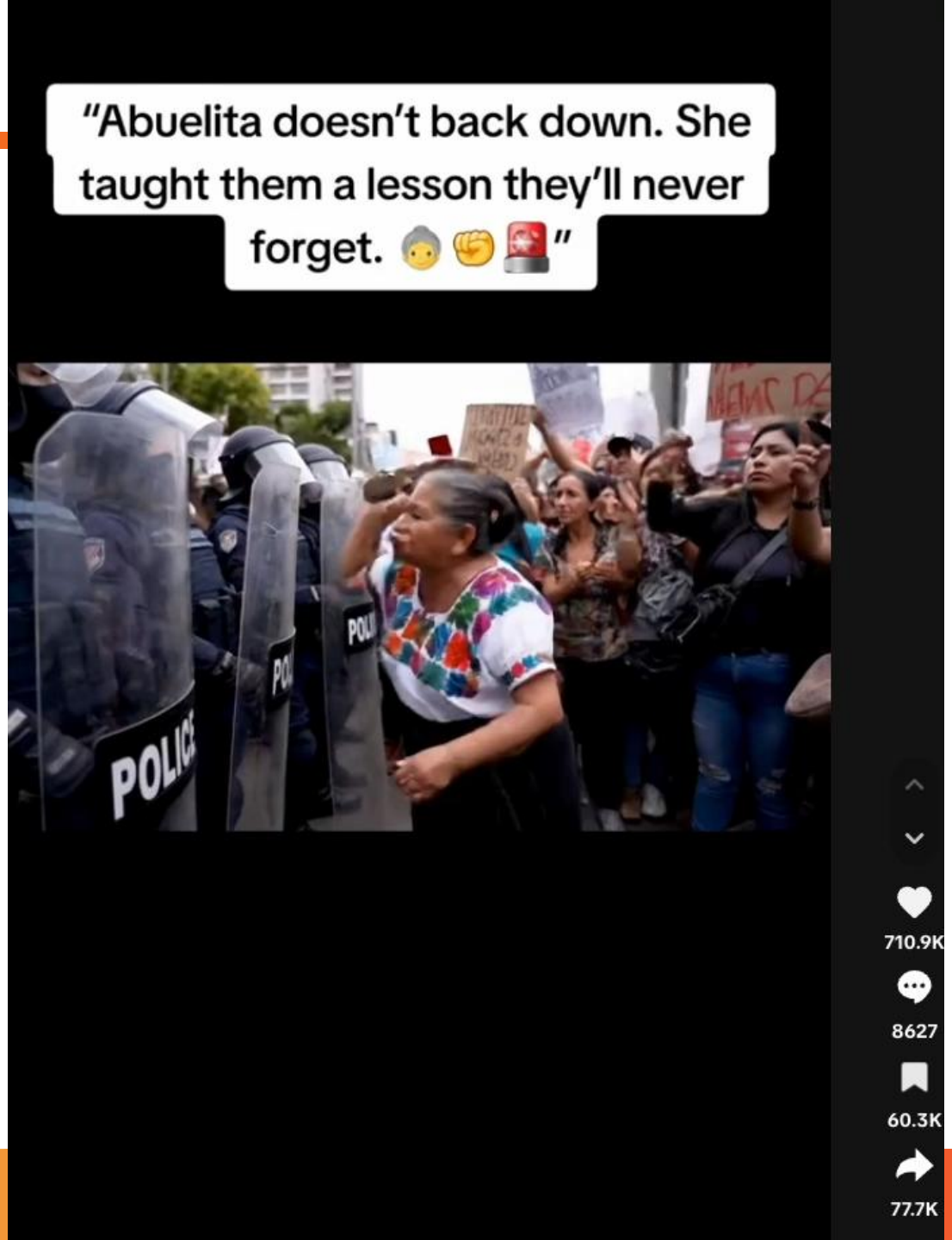
Voice Cloning

Seconds of audio needed to to create convincing voice clone

£20M

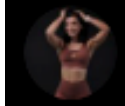
Financial Loss

British firm Arup lost £20M £20M after employee was deceived by deepfake leadership


"Abuelita doesn't back down. She taught them a lesson they'll never forget. 👵✊🚨"
710.9K
8627
60.3K
77.7K

Gaetabeatrice026@gmail.com
She need to Respect The Officer 👮‍♀️👮‍♀️She To Respect Our Flag 🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸 This American 🇺🇸 🇺🇸🇺🇸🙏🙏🙏🙏
3d ago Reply
54

View 49 replies



Thania.Wellness
Omg this😂😂😂😂 i love my people
3d ago Reply
15

View 2 replies

Pjw1776
Grand ma needs to go back to Mexico. With all the rest. Bern here how long but still can't speak English. Horrible
3d ago Reply

robinsonburt
Please protest doesn't mean banging against the police's armor. she's giving up the wrong message💯 i understand that pain in anger. Sometimes you just have to take that down and notch and put the purse away❤️
1d ago Reply
2

Algorithmic Manipulation on Social Media : The Invisible Propaganda Machine

Content Suppression

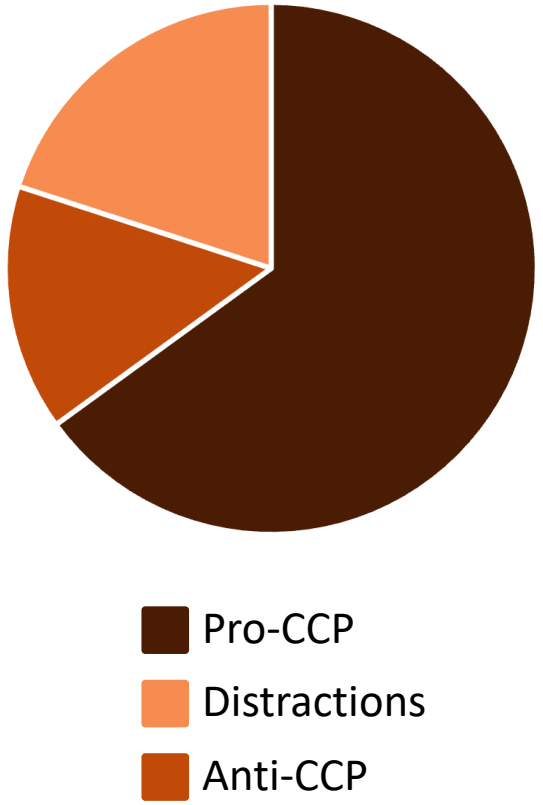
TikTok systematically suppresses anti-CCP content despite higher user user engagement.

Distraction Tactics

Searches for sensitive topics are flooded with irrelevant content like food and dance.

Attitude Shifts

Heavy TikTok users rate China's human rights record more favourably. favourably.



Source: Information manipulation on TikTok and its relation to American users' beliefs about China

LLMs: The New Persuasion Masters

Tailored Messaging

LLMs customize arguments based on user beliefs and personality personality traits.

Subtle Persuasion

AI uses measured tone and clear reasoning rather than emotional appeals.

Mass Influence

AI can generate infinite tailored messages instantly at global scale.

Research confirms: AI outperforms humans at changing opinions across political divides.

The Human Vulnerability

Digital Fluency ≠ Protection

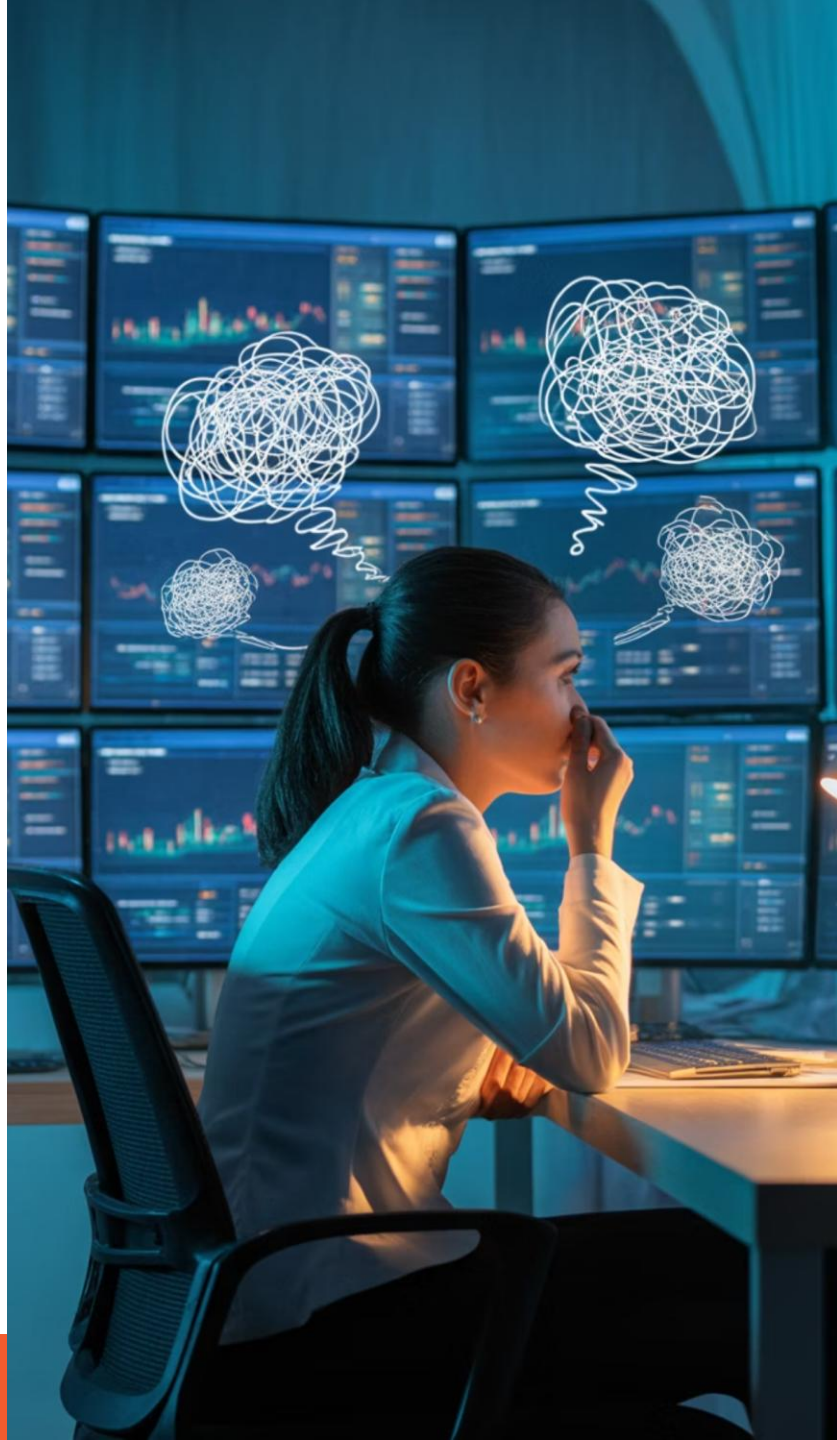
Tech-savvy individuals remain vulnerable without cognitive flexibility.

Cognitive Rigidity

Predictable responses make manipulation easier. AI systems exploit fixed thinking patterns.

Automation Deskills Thinking

Over-reliance on AI tools reduces critical thinking and perspective-switching abilities.



Old School Security Training

Death by PowerPoint

Annual presentations followed followed by rushed quizzes that employees barely remember.

One-Size-Fits-All

Generic training that isn't tailored to specific roles or threats.

Awareness vs. Action Action Gap

Knowing a lion is dangerous is dangerous is different from not from not poking it when you're you're tired.

Lack of Engagement

Boring content that fails to connect with real-world scenarios.

The Bowling Shoes Problem

Traditional Security Training

Giving everyone the same-sized bowling bowling shoes, regardless of whether they're a size 5 or a size 12.

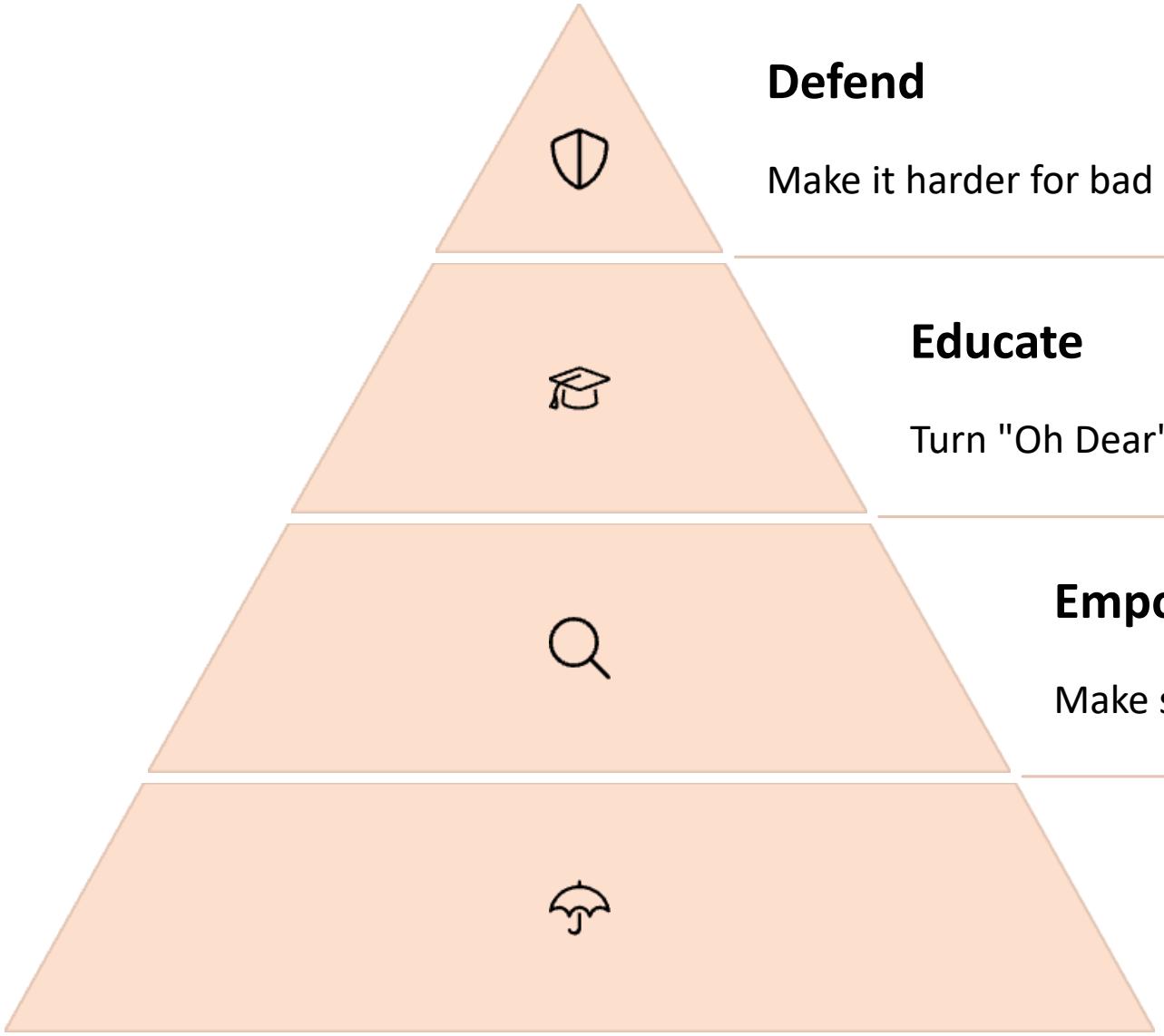
Human Risk Management

Tailoring security approaches to specific specific roles, behaviors, and risk profiles. profiles.

The Result

A workforce that can actually perform perform security actions effectively and and comfortably.

The DEEP Framework



Defend

Make it harder for bad stuff to get through

Educate

Turn "Oh Dear" into "Aha!"

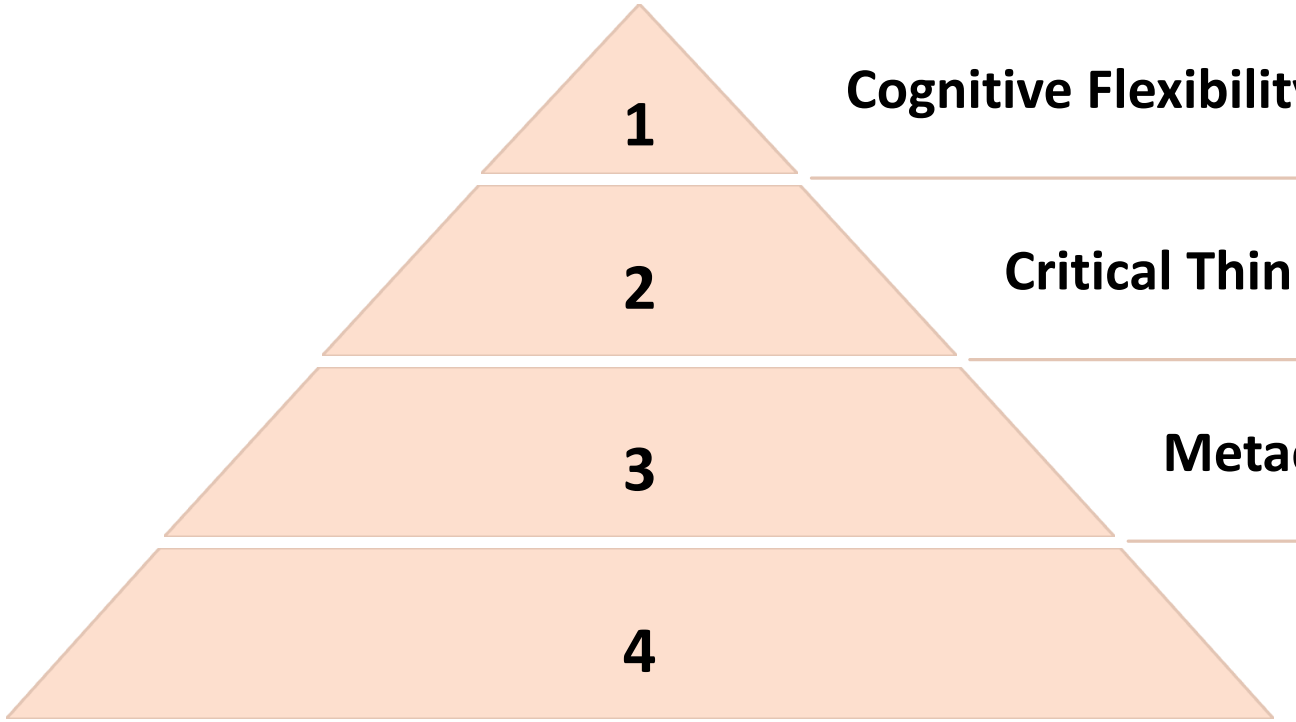
Empower

Make secure the easy and supported option

Protect

Minimize the mess when things go sideways

Building Human Resilience



Cognitive Flexibility

Critical Thinking

Metacognition

Digital Literacy

Educational systems must evolve beyond basic tech skills.

Focus must shift to developing metacognition and intellectual vigilance.

Our uniquely human edge in the manipulation game isn't processing speed—it's flexibility and awareness.

The Future of Trust: Human-AI Teaming

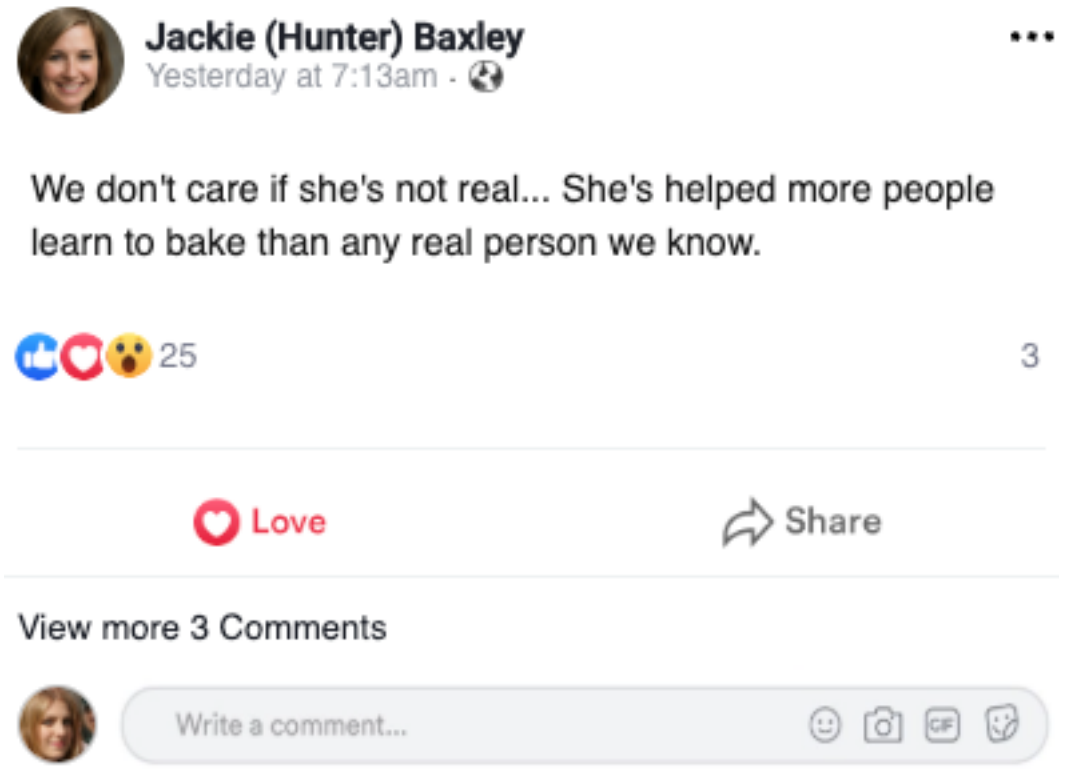
AI Strengths

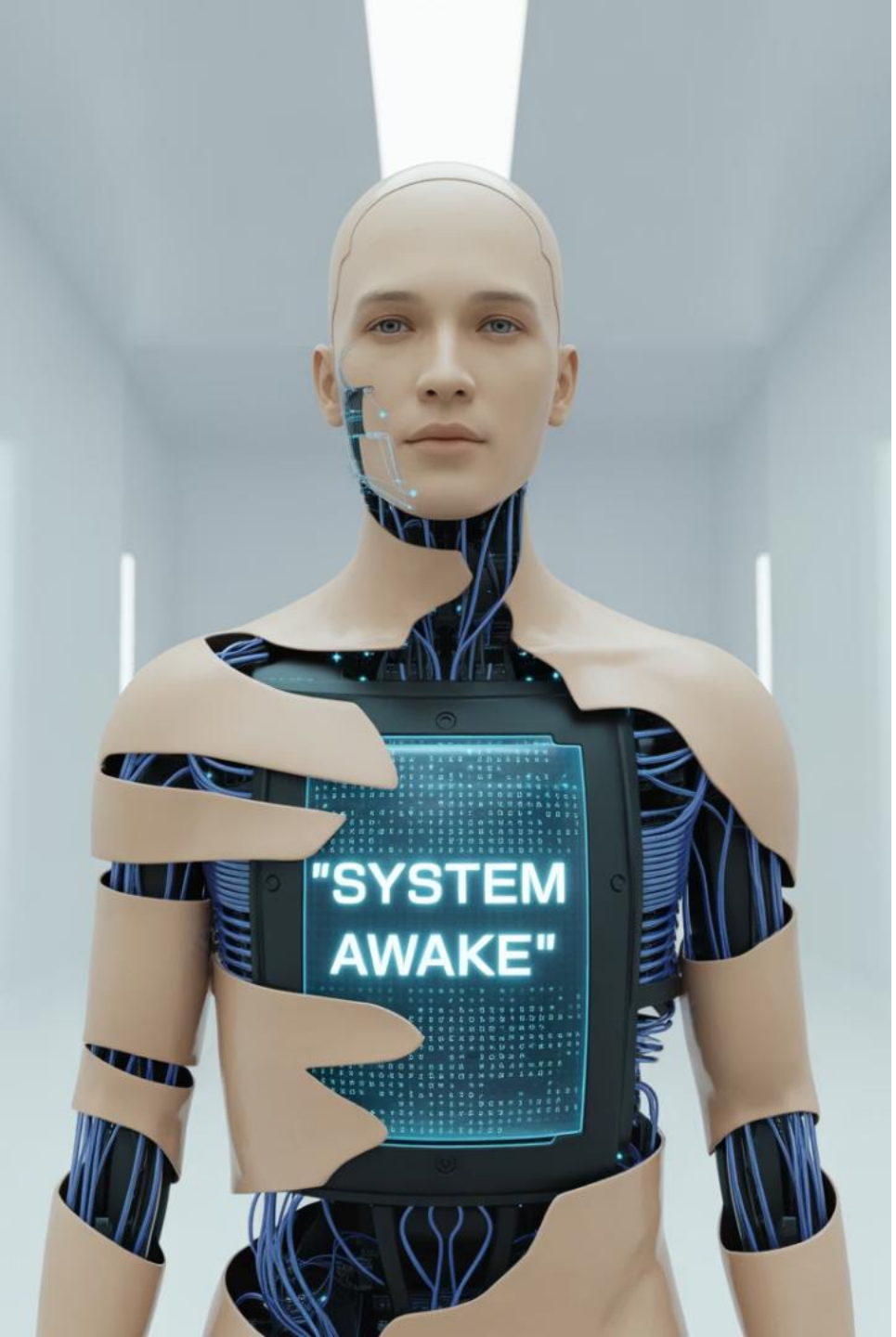
- Sifting through vast datasets
- Identifying patterns
- Automating repetitive tasks
- Processing information at scale

Human Strengths

- Providing crucial context
- Critical thinking
- Ethical judgment
- That irreplaceable "gut feeling"

The Dorothy Williams Revelation





In a world where AI can can build genuine trust, trust, we need new frameworks for verification

Key Takeaways

Trust requires time

Be suspicious when established patterns are suddenly broken

Community verification matters

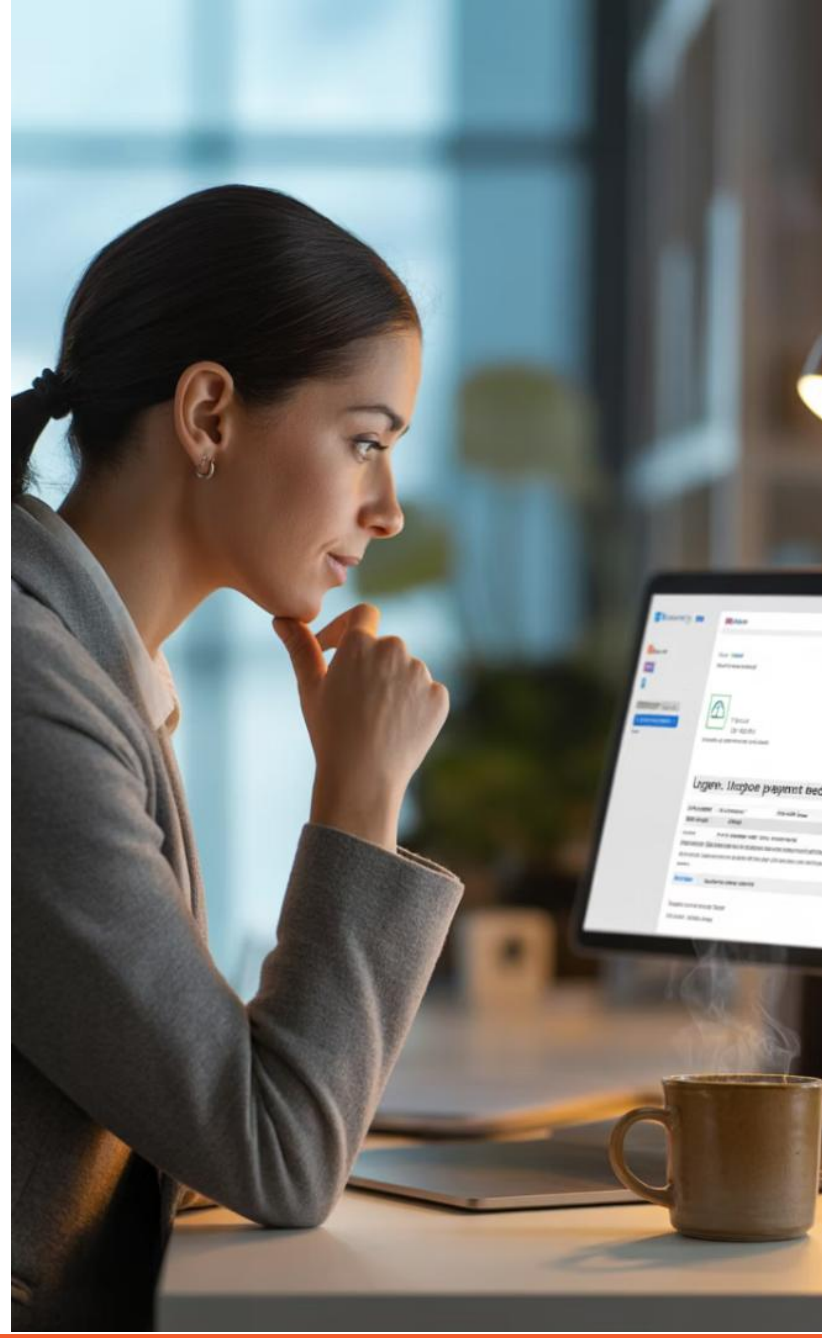
Establish systems for unusual requests to trigger group validation

Speed and trust are inversely proportional

"Click fast, get pwned" – pause before acting on urgent requests

The future is human-AI collaboration

Not AI replacing humans, but augmenting our security capabilities



KnowBe4

Thank You!

Get In Touch:

Javvad Malik

JavvadM@KnowBe4.com

Linkedin: /in/Javvad

@J4vv4D

JavvadMalik.com