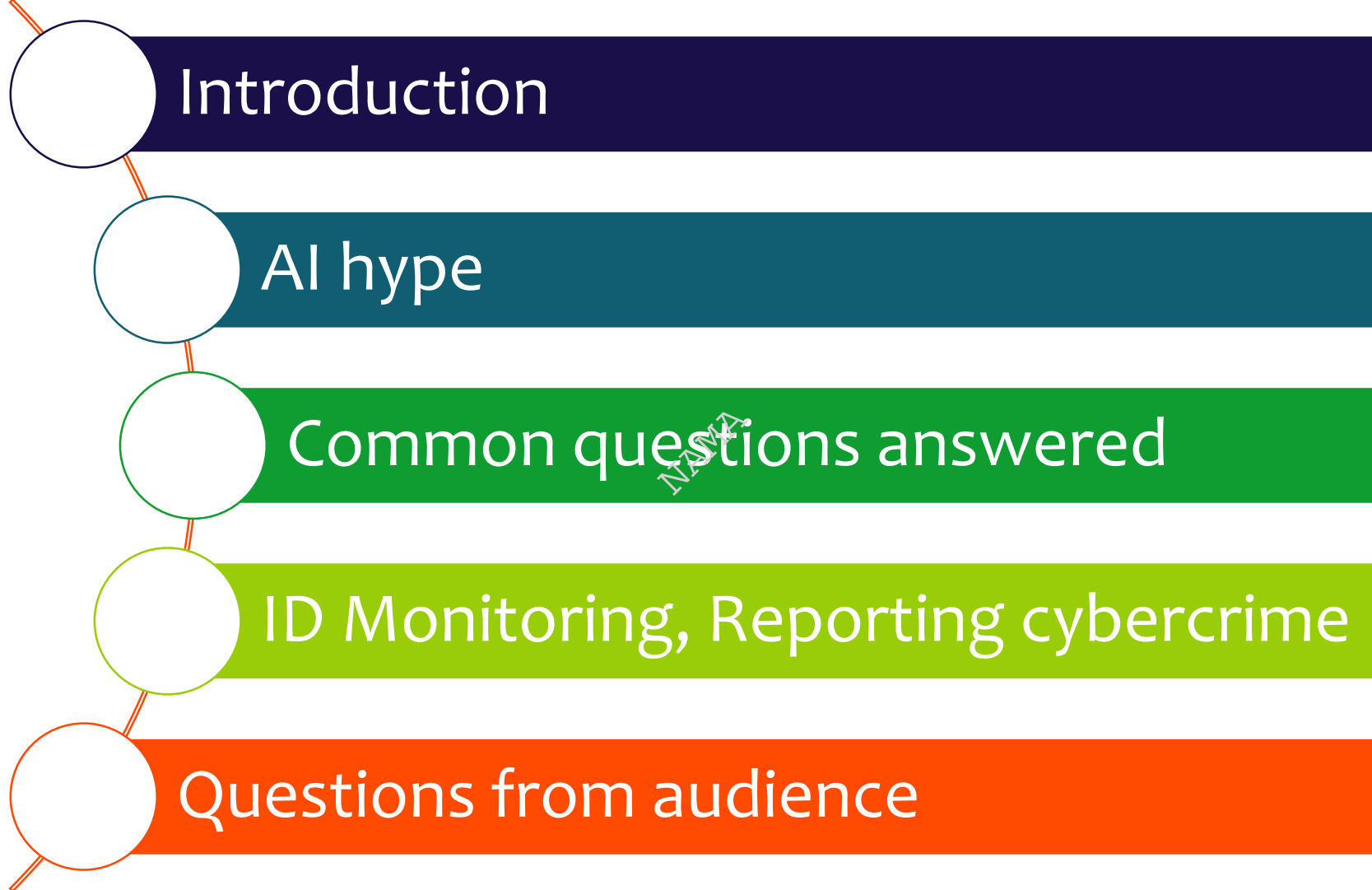




# Cybersecurity in the age of AI

Ramanujam Srinivasan

# Agenda





# What this hype about AI?

# Which face is real?





# How much would you like to pay for this Air BnB?



ENTIRE GUEST SUITE

## Complete 4 Bedroom Duplex Apartment!!

US



Ali

 8 guests  4 bedrooms  4 beds  2 baths

My 2-story, 1 bath house. The room (sits, second floor are H2) floor sense of a well maintained private terrace, a deck, and cathedral and lovely deck with a pine trees. There is an estate of them. Upstairs can be used and is at a walling for modern studio Gorgeous woodland haven. Located at the edge of the Lakeshore. Next to theatres with bustling restaurants, cafes, galleries, coffees, Acropolis, transport line and bus no 1 andme! Walking distance to \*KødB is

# Is this beautiful place in India or Switzerland?





# These are real faces..



# Detecting AI

## AI content is everywhere



- **54%** of people can distinguish between human and AI-generated
- **60%** of consumers saw at least one deepfake video in the last year
- Only **15%** of people have never encountered a deepfake video
- For high-quality deepfake videos, human detection accuracy drops to just **24.5%**

## Growth Trajectory Chart



- **2022:** Baseline deepfake incidents
- **2023:** 500,000 deepfakes shared
- **2024:** 150% increase in incidents
- **2025:** 8 million predicted (1,600% growth)



# AI Stats contd..

## Economic Impact

3,000%

Increase in deepfake fraud cases  
in 2023 alone

\$40B

Predicted US fraud losses by 2027

\$78B

Annual global cost of fake news

77%

of people who received voice-cloned scam messages **LOST MONEY**

2023

64.7%

Marketers using AI



2025

90%

Marketers using AI

85.1%

Use AI for article writing & content creation

85-99%


AI detection tool accuracy rates

40%

of social media content is FALSE

86%

of global citizens exposed to fake news

 **WARNING:** At current growth rates, deepfakes will outnumber authentic content within 3 years



## Exponential Growth

2023

500K

Deepfakes shared online

2025

8M

Predicted deepfakes online

1,600%

Increase in just 2 years



## 2025 Incident Surge

150

All of 2024

+19% in just 3 months

179

Q1 2025

Only



3x

Video  
Deepfakes



8x

Voice  
Deepfakes

# Spotting Fake Photos

## Pointers



- Asymmetrical faces and ears
- Weird teeth or inconsistent lighting
- Blurry backgrounds with sharp faces
- Jewelry or accessories that don't make sense



## - Are there any tools?

- [Hive Moderation](#)
- [AI or Not](#)
- [Was it AI?](#)

# Lets hear this..



## **What Would You Do?"**

The voice sounds exactly right. Your response?"

- A) Send money immediately
- B) Hang up and call back on known number
- C) Ask a personal question only they would know
- D) Demand video proof



# Voice cloning

## Voice Cloning

- Voice cloning attacks increased 3,000% in 2024
- **3 seconds** is all scammers need to clone your voice
- **95%** accuracy rate for modern AI voice cloning tools
- **77%** of victims were over age 60
- **\$5-15** cost to clone a voice using readily available AI tools
- **85%** of people can't distinguish between real and cloned voices in blind tests
- **48 hours** average time for scammers to deploy a cloned voice attack

## Common scenarios

### Family Emergency Scams

- "Grandparent scam" with cloned grandchild's voice
- Fake kidnapping calls using victim's cloned voice
- "Stranded abroad" scenarios with family member's voice

### Business Fraud

- CEO fraud with cloned executive voices
- Fake vendor payment requests with familiar voices
- Business email compromise enhanced with voice verification calls

### Social Engineering

- Dating app scams using attractive person's cloned voice
- Social media friend impersonation for money requests
- Fake tech support calls using trusted company representative voices

# Call to Action on Voice cloning

## Pointers

### Immediate Actions

- Never send money based solely on voice requests
- Always verify through independent contact methods
- Be cautious about posting voice content online

### Recognition Signs

- Emotional urgency and time pressure
- Unusual background noise or audio quality
- Requests for secrecy ("Don't tell anyone")
- Payment through untraceable methods

### Before Every Suspicious Call:

- Does this request make sense?
- Am I being pressured to act quickly?
- Can I verify this through another channel?
- Would the real person normally ask for this?

### Family Emergency Verification:

- Ask for the family code word
- Request specific details only they would know
- Call them back on their regular number
- Contact other family members to verify

"Remember: In the age of AI, trust but verify isn't enough. Now it's verify, then verify again, then trust—because the voice you're hearing might be perfectly real, and perfectly fake at the same time."

# Lets see this..



**Its just a WhatsApp forwarded video  
Let me forward it to my family/friends  
What's the harm in this free forwarded video?**



# Seeing is no Longer Believing

## Deep Fake videos

- The entire process of creating a deepfaked photo or video can be as quick as **8 minutes**
- By using just one clear face image, it takes less than **25 minutes** and costs almost nothing for a scammer to create a prerecorded **60-second** deepfake video
- Deepfake fraud attempts jumped by **3,000%** in 2023
- Last year saw a **10x** increase in the number of deepfakes detected globally across all industries

## Deep fake frauds in news..

- Prerecorded sham celebrity endorsement videos have portrayed CBS News host Gayle King and actor Tom Hanks as spokespersons for a weight-loss product and dental plan
- A finance worker at a multinational firm was tricked into paying out \$25 million to fraudsters using deepfake technology to pose as the company's CFO in a video conference call
- New Hampshire residents received a robocall from someone who sounded like President Joe Biden, telling them not to cast their ballot in the state's presidential primary. The voice was generated using AI
- When video ads of Taylor Swift promoting Le Creuset cookware appeared on Facebook in early 2024, these ads were fake, fabricated without Swift's knowledge
- Medicare fraud campaigns viewed over 195 million times

# Call to Action on Deep Fakes

## Pointers

- **Visual Inconsistencies**

- Unnatural eye movements or blinking patterns
- Inconsistent lighting on faces vs. background
- Slightly "off" facial proportions
- Hair that doesn't move naturally
- Inconsistent skin texture or aging

- **Audio Mismatches**

- Robotic or "too perfect" speech patterns
- Lack of natural breathing sounds
- Voice that doesn't match mouth movements precisely
- Missing emotional inflection in speech

- **Behavioral Anomalies:**

- Limited head movements or gestures
- Avoiding direct eye contact with camera
- Repetitive or unnatural hand gestures
- Background inconsistencies or artifacts
- Person never turns heads completely

# Detection tools - Deep Fakes

## Tools



### **MIT Media Lab - Detect Fakes**

([media.mit.edu/projects/detect-fakes](https://media.mit.edu/projects/detect-fakes))

Research project designed to answer questions and identify techniques to counteract AI-generated misinformation

### **Northwestern University - Detect Fakes**

([detectfakes.kellogg.northwestern.edu](https://detectfakes.kellogg.northwestern.edu))

Challenges you to discern AI-manipulated videos from real videos and determine if you can do better than an algorithm

### **Sensity AI** ([sensity.ai](https://sensity.ai))

The only AI-threat detection platform with a cross-industry approach, offering deepfake detection for videos, images, and audio

### **Reality Defender** ([realitydefender.com](https://realitydefender.com))

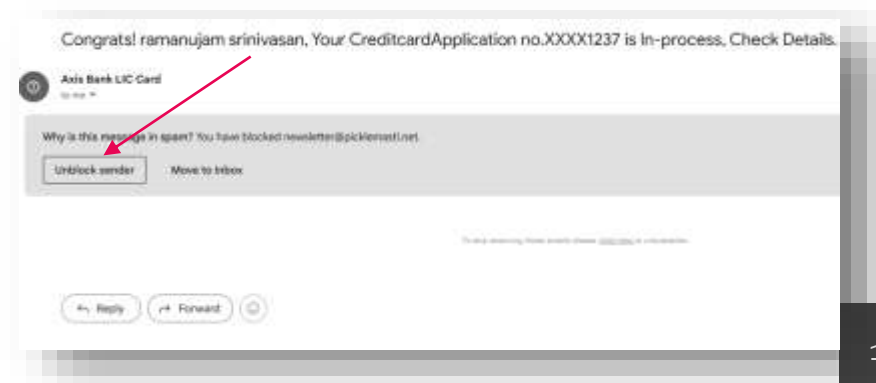
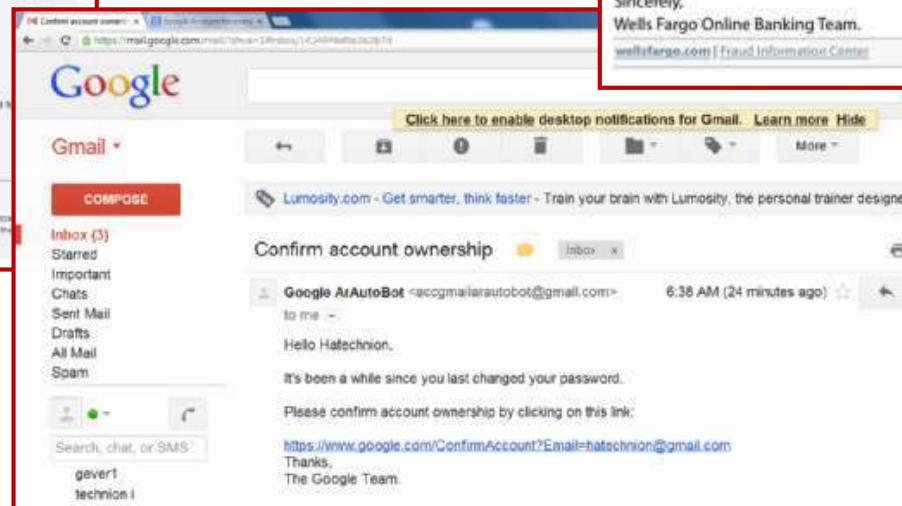
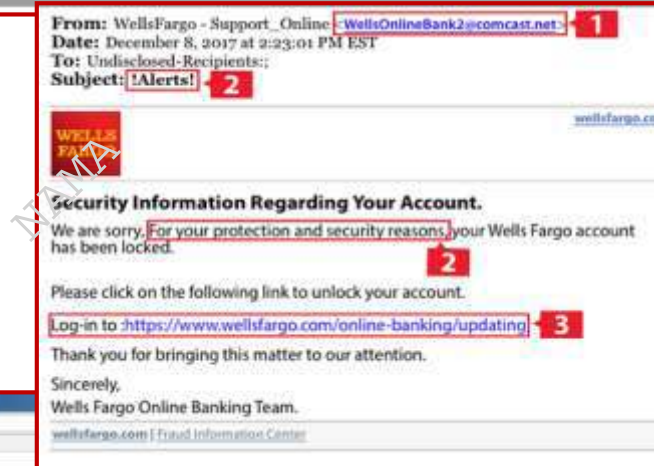
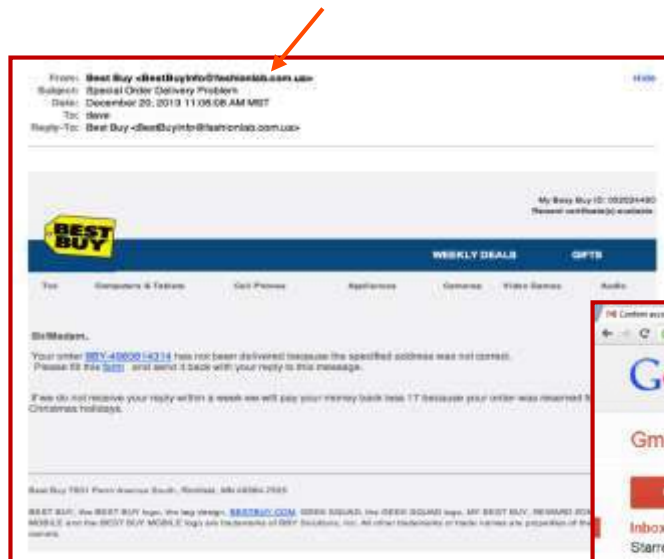
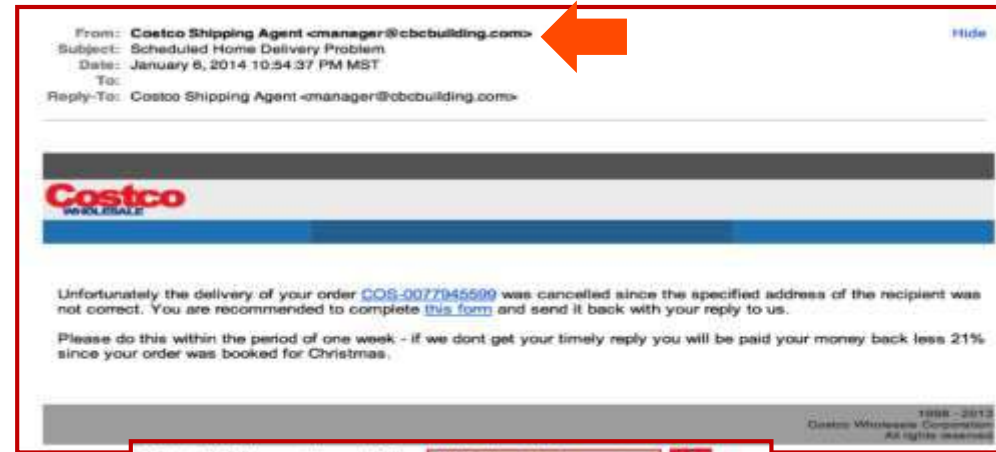
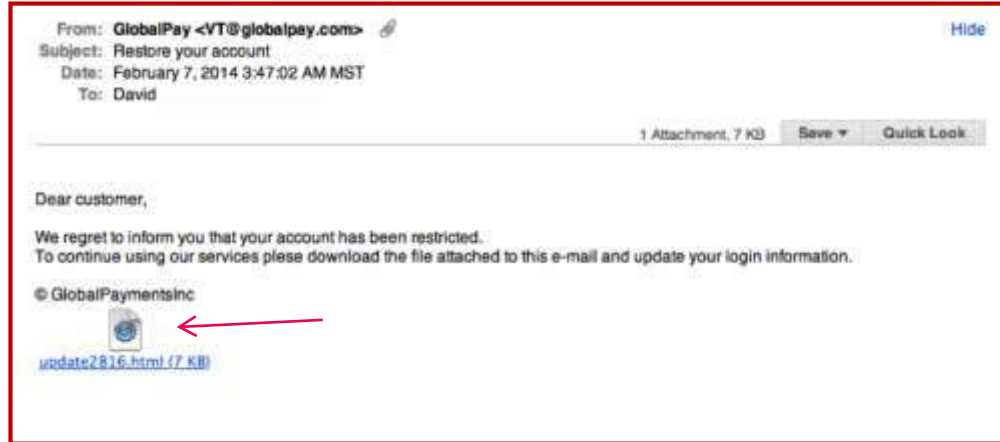
Deepfake detection platform designed to combat AI-generated threats across multiple media types using a patented multi-model approach

### **Deepware Scanner** ([deepware.ai](https://deepware.ai))

- Scan suspicious videos to find out if they're synthetically manipulated



# Beware - Phishing



# Which one is fake?

### Important Message from Karen Lynch, CEO of CVS

Dear Curtis,

I hope this email finds you well. I wanted to reach out to you personally to inform you of some important updates regarding our company.

As you may know, CVS is committed to providing the best possible healthcare services to our customers. In order to continue to do so, we have recently made some changes to our internal systems and processes.


As a valued member of our team, we need your help to ensure that these changes are implemented smoothly. Please click on the link below to access our new system and complete the necessary training:

<http://www.cvs-training.com>

If you have any questions or concerns, please do not hesitate to reach out to me directly.


Thank you for your continued dedication to CVS.

Best regards,



**Karen Lynch**  
CEO, CVS Health  
Email: [klynch@cv.com](mailto:klynch@cv.com)  
Phone: (555) 123-4567

Employee Wellness Survey



To: [Redacted]

Hello [Redacted]

Our [Redacted] team has been working hard to help improve employees' lives and wellness, but we need your help! We have a brief five-question survey to help us pursue employee-driven initiatives. An example initiative we've recently completed was the [Redacted] program, you can read more here: [https://www.\[Redacted\]](https://www.[Redacted])

The survey is brief, requiring only a couple minutes of your time and your responses can be submitted anonymously. We request your participation before the deadline, this Friday. Priority will be accorded to initial submissions, so we encourage prompt action. You may access the survey [here](#).

Thank you for continuously helping make [Redacted] a better place to work at!

Best,  
[Redacted]  
Manager, [Redacted]  
[Redacted]@[Redacted].com

The content of this message is confidential. If you have received it by mistake, please inform us and then delete the message. It is forbidden to copy, forward, or in any way reveal the contents of this message to anyone. The integrity and security of this email cannot be guaranteed.

Reply

Forward

# Another one to spot

**From:** [support@microsoft.co.uk](mailto:support@microsoft.co.uk)  
**Sent:** 16/01/2023 11:44  
**To:** Bob Smith <Bob.Smith@company.com>  
**Subject:** Urgent Action Needed!



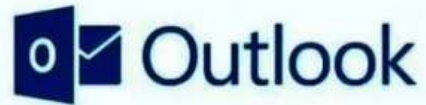
Microsoft Account

## Verify your account

We detected some unusual activity about a recent sign in for your Microsoft account. you might be signing in from a new location app or device.

To help keep your account safe. We've blocked access to your inbox , contacts list and calander for that sign in. Please review your recent activity and we'll help you secure your account. To regain access you'll need to confirm that the recent activity was yours

**From:** [support@microsoft.co.uk](mailto:support@microsoft.co.uk)  
**Sent:** 16/01/2023 11:44  
**To:** Bob Smith <Bob.Smith@company.com>  
**Subject:** Unusual Sign In Activity



Microsoft Account

## Verify your account

We detected some unusual activity about a recent sign in for your Microsoft account [bo\\*\\*\\*\\*\\*@company.com](mailto:bo*****@company.com). you might be signing in from a new location app or device.

To help keep your account safe. We've blocked access to your inbox, contacts list and calendar for that sign in. Please review your recent activity and we'll help you secure your account. To regain access you'll need to confirm that the recent activity was yours



OpenAI Payments  
Your Last Payment for ChatGPT Was Declined

October 15, 2024 at 9:24 AM



Hello,

We regret to inform you that your latest subscription payment for ChatGPT was unsuccessful. This means we were unable to process your monthly payment.

To avoid an interruption in service, please update your payment information as soon as possible. If we do not receive your payment within the next 7 days, your subscription will be canceled, and you will lose access to ChatGPT.

[Update Payment Information](#)

If you have already updated your payment information, please disregard this email. For any questions or assistance, feel free to contact our support team at [support@openai.com](mailto:support@openai.com).

Thank you for choosing ChatGPT, and we appreciate your prompt attention to this matter.

Best regards,  
The OpenAI Team

You are receiving this email because you have an active subscription to ChatGPT by OpenAI.  
© 2024 OpenAI, Inc. All rights reserved.



# Secure Emails



Have I been hacked?

- <https://haveibeenpwned.com/>

How to protect Gmail/ Yahoo mail?

- Set up MFA/2 step verification, Check account settings



Google Authenticator

1. Avoid opening attachments or links without checking source
2. Block emails that are suspicious
3. <sup>NAMA</sup> Avoid OOO office outside of work

<https://phishingquiz.withgoogle.com/>

# Call to Action – AI Phishing emails

## Pointers

**Audit your privacy settings** - make social media profiles private or friends-only

**Create a unique email** for online shopping separate from your main email

**Install browser extensions** like uBlock Origin or Malwarebytes Browser Guard

**Enable 2-Factor Authentication** on all accounts - especially email, banking, and social media

### The "Trust but Verify" Rule:

- If someone contacts you claiming to be from a company, hang up and call the official number
- If you receive an urgent email, wait 24 hours before acting (unless it's truly an emergency)
- When in doubt, ask a tech-savvy friend or family member to review suspicious messages

### Red Flags


- Emails asking you to "verify" information you never signed up for
- Messages creating false urgency ("Your account will be closed in 24 hours")
- Requests for sensitive information via email (legitimate companies never do this)
- Links that don't match the claimed sender's official website
- Unexpected attachments, especially .zip or .exe files


# I have so many questions

What can I do to protect myself and my family?



# Understand the Entry points

- 
- Passwords
  - Facebook
  - WhatsApp
  - Browsers
  - Data/Storage

- 
- Viruses
  - Calls
  - Kids and Devices
  - ChatGPT

## Questions - Passwords

- How long should my password be, and what makes a password actually secure?
- Is it really that dangerous to use the same password for multiple accounts?
- What's a password manager and do I actually need one?



# Secure Passwords

## How Safe Is Your Password?

Time it would take a computer to crack a password with the following parameters

	Lowercase letters only	At least one uppercase letter	At least one uppercase letter + number	At least one uppercase letter + number + symbol
1	Instantly	Instantly	-	-
2	Instantly	Instantly	Instantly	-
3	Instantly	Instantly	Instantly	Instantly
4	Instantly	Instantly	Instantly	Instantly
5	Instantly	Instantly	Instantly	Instantly
6	Instantly	Instantly	Instantly	Instantly
7	Instantly	Instantly	1 min	6 min
8	Instantly	22 min	1 hrs	8 hrs
9	2 min	19 hrs	3 days	3 wks
10	1 hrs	1 mths	7 mths	5 yrs
11	1 day	5 yrs	41 yrs	400 yrs
12	3 wks	300 yrs	2,000 yrs	34,000 yrs

Source: Security.org



statista

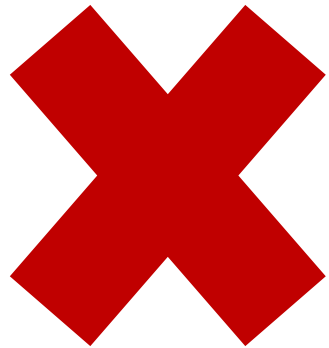
1. Use strong passwords
2. Make them unique
3. Use a password manager



# Secure Passwords contd..

## Bad Passwords

- 123456
- admin
- password
- Srini123#
- rageducksimplemoon



## Best Passwords

- coW!mveN#move?pian0h
- F2a\_+Vm3cV\*j
- lwiCcR!f)dliNkE?6
- 84sk37b4LLTr1Ck



## Questions - WiFi

- Why shouldn't I connect to public WiFi at coffee shops and airports?
- How can I tell if someone is using my home WiFi without permission?
- What's the difference between WPA2 and WPA3, and which should I use?

# Secure Wi-Fi



**NETGEAR**



**ASUS**



**1. Change Old Router – Use new routers with WPA3/WPA2**

**2. Change default settings**

**3. Don't connect to public Wi-Fi**

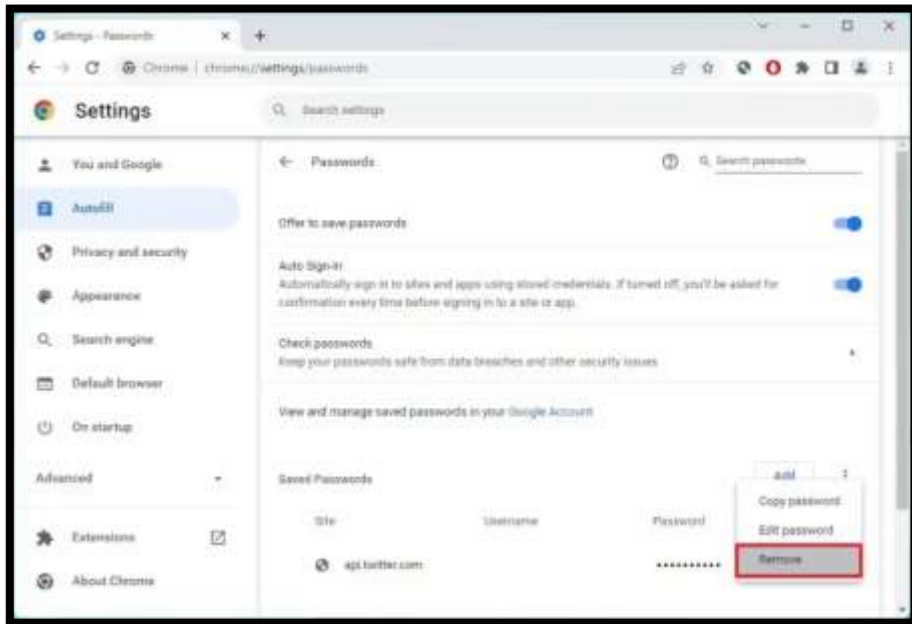


## Questions - Browsers

- I get too many popups while browsing.
- Should I be worried about all the saved passwords in my browser?
- How do I know if a website is safe to enter my personal information?



# Secure Browser



1. Use site with https:// when possible
2. Disable auto complete for forms
3. Don't click on pop ups and ads

## Questions – Data/Storage

- Is it safe to store important documents and photos in cloud services like Google Drive or iCloud?
- What's the difference between backing up my data and encrypting it?
- How often should I backup and where?

# Secure Data



What should I back up?

- Docs, Photos

Where should I back up?

- Cloud, External drives

How often should I back up?

- Automatic, Manual sync



1. Back up your data regularly
2. Don't use shared USB sticks
3. Encryption is optional but will slow downloading and uploading

## Questions – Calls

- How can I tell if someone is trying to scam me over the phone?
- I get too many calls and fake texts.
- What information should I never give out during a phone call?

# Secure Calls



<https://www.donotcall.gov/>

1. Don't take calls from unknown numbers
2. If it's a robotic call, don't say yes and answer questions
3. Block suspicious numbers immediately

To add your number to NDNC Do Not Call Register India, you have to **send an SMS "Start o" to 1909.**

It may take up to 7 days for the DND to reach activation on your number.



## Questions – WhatsApp

- Are my WhatsApp messages really private, or can other people read them?
- Should I be concerned about clicking links that people send me in WhatsApp?
- What are those backup settings in WhatsApp, and are they secure?

# WhatsApp



- Look for the "end-to-end encrypted" message in new chats
- WhatsApp can see metadata (who you message, when) but not content
- Screenshot notifications don't exist - recipients can save your messages
- Hover over links to see the real destination (on desktop)
- Don't click shortened URLs (bit.ly, tinyurl) from unknown sources
- Keep your WhatsApp app updated to get security patches
- Consider turning off cloud backups for maximum privacy
- If you use backups, secure your Google/Apple account with two-factor authentication
- Remember: no backup means losing chat history if you lose your phone

## Questions – Facebook

- Who can actually see my Facebook posts, even when I think they're private?
- How do I stop Facebook from tracking me on other websites?
- What should I do about friend requests from people I don't know?

# Facebook

---



- Review privacy settings every few months
- Use "Friends except..." for sensitive posts to exclude certain people
- Check what you're tagged in regularly
- Consider who can see your friends list and photos of you
- Turn off "Off-Facebook Activity" in your Facebook privacy settings
- Use browser extensions like uBlock Origin or Privacy Badger
- Log out of Facebook when not using it
- Use different browsers for Facebook and other browsing
- Consider deleting the Facebook app and using the mobile website instead
- Only accept requests from people you know in real life
- Check mutual friends and profile details before accepting
- Be especially cautious of profiles with few photos or friends
- Report obviously fake profiles
- Consider changing settings so only friends of friends can send requests

## Questions – Viruses

- How do I know if my device has a virus or malware?
- Do I need antivirus software on my phone like I do on my computer?
- What's the difference between a virus, malware, and ransomware?



# Secure Laptop/ Computer - Viruses



- Run full system scans with built-in security (Windows Defender, Mac's XProtect)
- Be suspicious of persistent pop-ups or browser redirects
- Use Malwarebytes for additional scanning if suspicious
- Only install apps from official stores (App Store, Google Play)
- Review app permissions before installing
- For Android: consider antivirus if you install apps from unknown sources
- For iPhone: built-in security is usually sufficient
- Be cautious of apps requesting excessive permissions
- Viruses: spread through infected files or programs
- Malware: includes viruses, spyware, adware, trojans
- Ransomware: locks your files and demands payment (usually cryptocurrency)
- Prevention is the same for all: keep software updated, use caution with downloads and links
- Regular backups protect against ransomware

## Questions – Kids and Devices

- How can I monitor my child's online activity without invading their privacy?
- What are the most important safety rules to teach kids about using the internet?
- How do I set up parental controls that actually work?

# Make Internet safe for kids



- Use built-in parental controls: Screen Time (iOS), Family Link (Android)
- Focus on time limits and appropriate content rather than reading private messages
- Have regular conversations about online experiences
- Monitor younger children more closely
- Teach them to never share full name, address, school, or phone number
- Teach them to think before posting - "Would I want my teacher/grandma to see this?"
- Explain that screenshots exist and nothing is truly "temporary"
- Router level: Circle Home Plus, Disney Circle, or built-in router controls
- Device level: Screen Time (iOS), Digital Wellbeing (Android), Windows Family Safety
- App level: YouTube Kids, Messenger Kids for younger children
- Remember: tech-savvy kids may find workarounds - maintain open dialogue

## Questions – ChatGPT

- Is it safe to share personal information with ChatGPT and other AI chatbots?
- Can ChatGPT give me accurate medical, legal, or financial advice?
- How do I know if information from ChatGPT is accurate and reliable?
- Can my employer or school see what I'm asking ChatGPT?

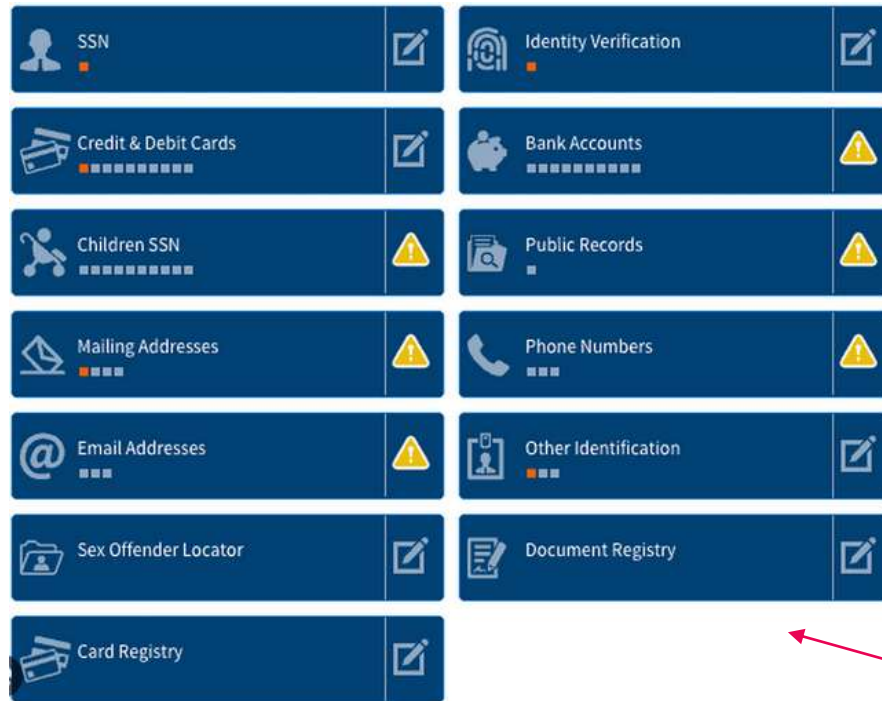
# Gen AI



- Never share passwords, Social Security numbers, or financial account details
- Avoid full names, addresses, phone numbers, or specific workplace details
- Use general examples instead of personal situations when asking for advice
- Always consult licensed professionals for serious medical, legal, or financial matters
- Use AI for general research and education, not specific advice
- Be especially cautious about health information - symptoms can have many causes
- Remember that AI responses can contain errors or outdated information
- Cross-check facts with reliable sources (official websites, established news outlets)
- Be extra skeptical of statistical claims, dates, or specific technical details
- Use AI-generated information as a starting point, not the final answer
- Ask for sources when possible, but verify those sources independently
- Be aware that AI can "hallucinate" - create convincing but false information
- Use personal devices and networks for private AI conversations
- Consider that AI companies may keep logs of conversations
- When in doubt, avoid using AI for sensitive work or school-related questions on institutional devices



# Implications of ID theft



- If ID is compromised, fraudsters can (on your behalf)
  - Open Bank account
  - Take Mortgage
  - Unauthorized purchases
  - Involve in Terrorism
  - Drain your bank account
  - False medical billing
- Apart from SSN, all the things on the left needs constant monitoring and vigilance
- Contact and Freeze SSN with all three credit bureaus and get report frequently

# ID Monitoring



- Choose a Identity protection after careful consideration
- Plans available for individual and family



- Please be aware that you will have to provide all your info to the company



- Check with them about assurance, their monitoring sources and alerting
- Check how they can help if your data is breached including identity recovery

# Reporting cyber crime

## India



Cyber police stations

Ex - Chennai – Vepery

<https://cybercrime.gov.in>

## Helpline contacts

## USA



<https://www.ic3.gov/>

If you or your organization is the victim of a network intrusion, data breach, or ransomware attack, contact your [nearest FBI field office](#) -

<http://www.fbi.gov/contact-us/field>

NCIJTF CyWatch 24/7 Command Center:  
(855) 292-3937 or [cywatch@ic.fbi.gov](mailto:cywatch@ic.fbi.gov)

Report cyber-enabled crime, including: digital theft of intellectual property; illicit e-commerce (including hidden marketplaces); Internet-facilitated proliferation of arms and strategic technology; child pornography; and cyber-enabled smuggling and money laundering

HSI Tip Line: 866-DHS-2-ICE (866-347-2423) or  
<https://www.ice.gov/webform/hsi-tip-form>



NAMA

# Questions?

# Thank You