

Look ahead:

- 09 Apr (today's lecture)
- 14/15 Apr (Quiz)
- 16 & 23 Apr (lecture by Prof. Arner)
- 21/22 Apr (No tutorials)
- 28/29 Apr (Group presentation)
- 30 Apr (cancelled)
- 02 May 11:59pm (Report + peer review form submission)

Security and privacy

Scope: all guest lectures on or before 02 Apr & Lecture 2&3 "blockchain technology and applications"

Law school professor, topic related to DPI (what is it?) & RegTech

Schedule (wait for TAs)

Recall the technologies for FinTech:

- Blockchain
- E-payment technology
- RegTech
- Cyber security and privacy
- Big data analytics
- AI/ML (Artificial Intelligence/Machine Learning)

Others:

- Financial inclusion
- Social impact
- ESG etc.

Today, we focus on **cyber security and privacy**

Recall that we gave a few examples in the introductory lecture:

(i) AI - face recognition



Which one is real,
which one is fake?

(iii) Deepfake?



Is it real?

(ii) Attacks on voice?



[Reveal Transcription] “okay google browse to evil dot com”



[Reveal Transcription] “without the dataset the article is useless”

Objectives of this lecture:

- To **increase your awareness** of cyber security and privacy
- **Don't be scared to use new technologies**, but try your best to make sure that **the apps are security and privacy protected (be cautious!)**.

E.g. **Account passwords** [(i) Don't use easy to guess passwords; (ii) Changing your account passwords regularly]

Multi-factor authentication (e.g. 2-factor authentication) [Q: Do you know what it is?]

A paper (2016) on top-ten least-secure passwords

- Lancaster U, Fujian Normal U, Peking U
- Based on a leaked Yahoo database of personal info.

123456

password

welcome

ninja

abc123

123456789

12345678

sunshine

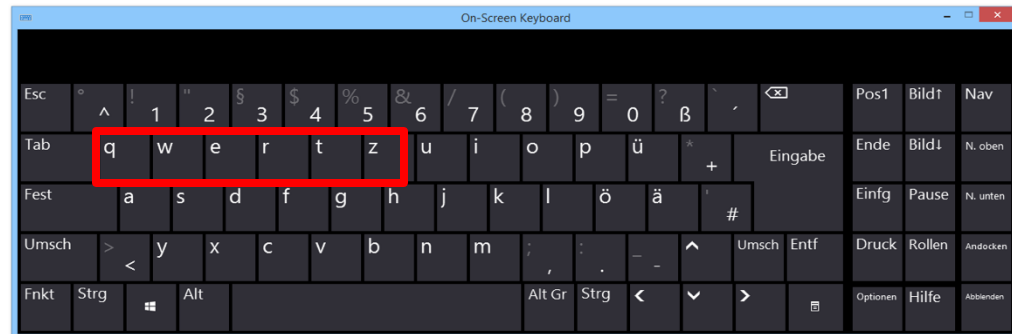
princess

qwerty

Other results:

❖ Able to guess passwords for more than 73% user accounts

❖ 1/3 of these passwords can be guessed right **within 100 guesses.**



Link: <http://www.securitymagazine.com/articles/87581-the-10-easiest-passwords-to-hack>

SplashData revealed the list for 2017 by analyzing more than five million user records leaked in 2017.

2016

123456
password
welcome
ninja
abc123
123456789
12345678
sunshine
princess
qwerty

2017

123456
password
12345678
qwerty
12345
123456789
letmein
1234567
football
iloveyou
.....
starwar
qazwsx



How about 2019?

123456
123456789
qwerty
password
1234567
12345678
12345
iloveyou
111111
123123

[Also, by SplashData]

How about 2020?

123456
123456789
picture1
password
12345678
111111
123123
12345
1234567890
senha

[By NordPass]



Register to Use 2-Factor Authentication (2FA)

Notes:

How about now? MFA?

1. Two-factor authentication (2FA) is a security measure for strengthening the protection of the authentication process in using the central IT services. On top of HKU Portal UID/PIN, a 6-digit one-time password (token code) has to be entered as the second key to complete the authentication.
2. Please provide your alternate email address (i.e. a non-HKU email address) and mobile phone number for the purposes of identity authentication, service event notification and communication with you.

On top of HKU portal UID/**PIN**, a **6-digit one-time password (OTP)** needs to be entered as a second key to complete the authentication

Key: to get this OTP, we need a non-HKU email or mobile phone [Q: do you know why?]



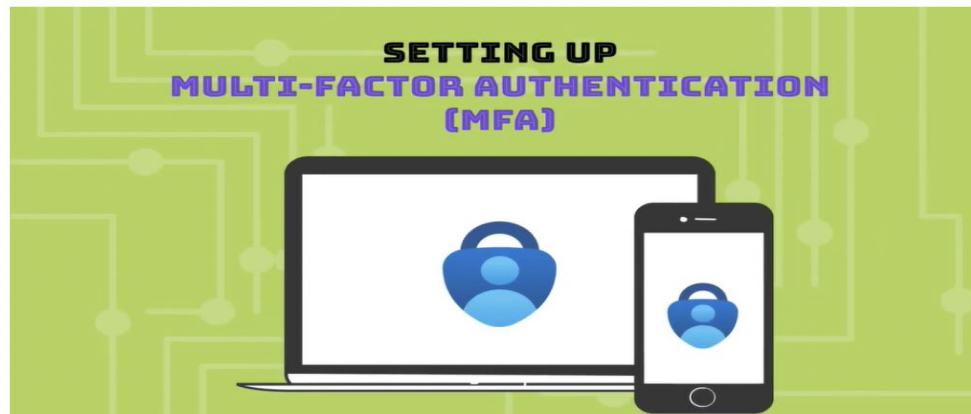
Home > KB > Accounts & 2FA > Multi-Factor Authentication (MFA) >

User Guide on Activating Multi-Factor Authentication (MFA)

Prerequisite:

- Your mobile phone (connected to a good data signal, wifi, or Eduroam, if you are on campus)
- A PC or tablet connected to the internet.

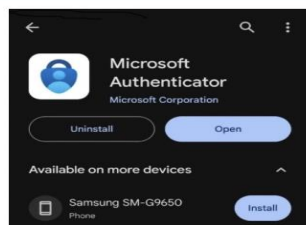
Video:



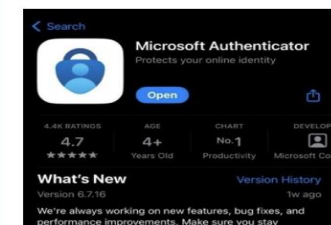
Setup Procedures:

1. Download Microsoft Authenticator App (**Android/iOS**)

Android (Google Play)



iOS (App Store)



Virtual bank crime cases

Hong Kong / Law and Crime

Thirteen Hongkongers accused of using false data in bungled attempts to open virtual bank accounts detained following citywide raids

- Police believe some of the suspects planned to sell accounts or use them to launder crime proceeds, while others wanted to get welcome gifts
- Officers are trying to find out whether they were all part of the same syndicate and say further arrests are possible

Fake ID, address, salary proof images....

(opening/selling accounts, money laundering, welcome gifts)

Solutions: AI (CNN)

Artifacts: color, shading, hair (3D) geometry..



Which one is real, which one is fake?



Still a lot of "attacks and defense" going on



Defense: special facial expression

AI Attacks

1. Data poisoning attacks

- At **training** stage (injecting poisoned data, e.g. data with wrong labels)
- Mislead system to **learn a wrong AI model**

2. Evasion attacks

- At **inference** stage (e.g. adding noise to images: [with Prof. Fang
- even using light (可见光的攻击)])
- Mislead system to **output a label**

More examples: Researchers from US demonstrated that only a little bit modification to a road sign (e.g. Speed limit of 80m/hr sign => STOP sign)

In-distribution Attacks

Adversarial Traffic Signs

Original



Adversarial



Classified as: Stop Speed limit (30)

Logo Attacks

Original



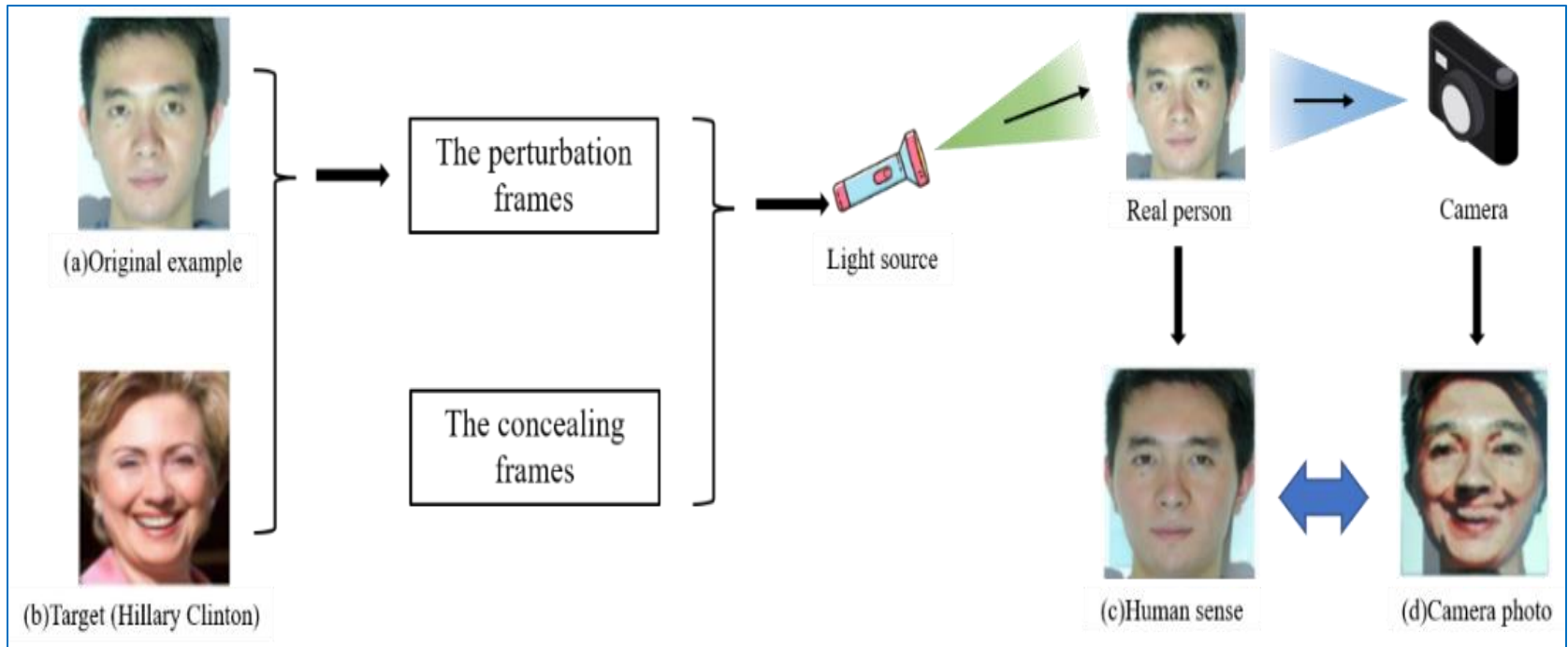
Adversarial



Classified as: Stop No overtaking

<https://arxiv.org/pdf/1802.06430.pdf>

Example from one of recent (advanced) attacks



Face recognition attack system based on visible light, where image (a) is the original image, image (b) is the target person, image (c) is the adversarial example observed by human eyes, and image (d) is the adversarial example captured by the camera.

Deepfake*

US 2021: Cheerleader (deepfake mom)

Hong Kong police arrest 6 in crackdown on fraud syndicate using AI deepfake technology to apply for loans

- Force says case marks first known instance of scammers with stolen ID cards employing deepfake to fleece financial agencies

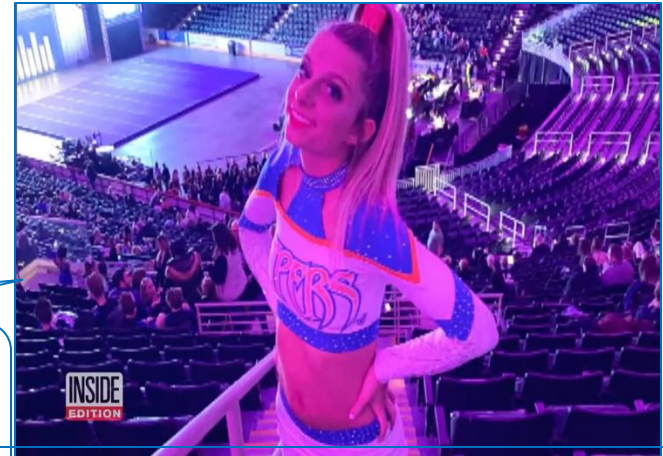
Police was not able to confirm that the video was made by deepfake.

India 2023 (Aug?):

Kerala deepfake fraud case: Efforts on to nab accused from Gujarat, say cops

Victims: e.g. celebrities (porn videos politicians/CEOs (fake speeches)....

E.g. someone's face superimposed on a porn video/photos; Obama etc. fake videos....



By deepfake, impersonate “old colleagues”: family photos even a short video whatsapp call

Image Frames
From Real Video

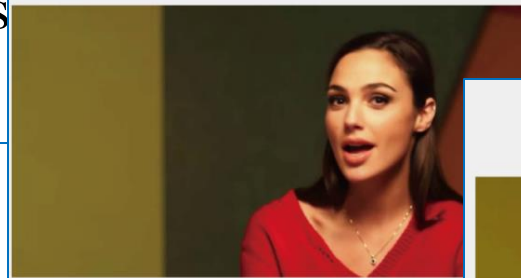
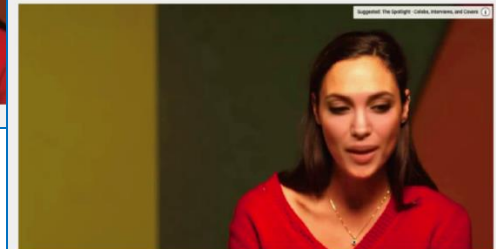
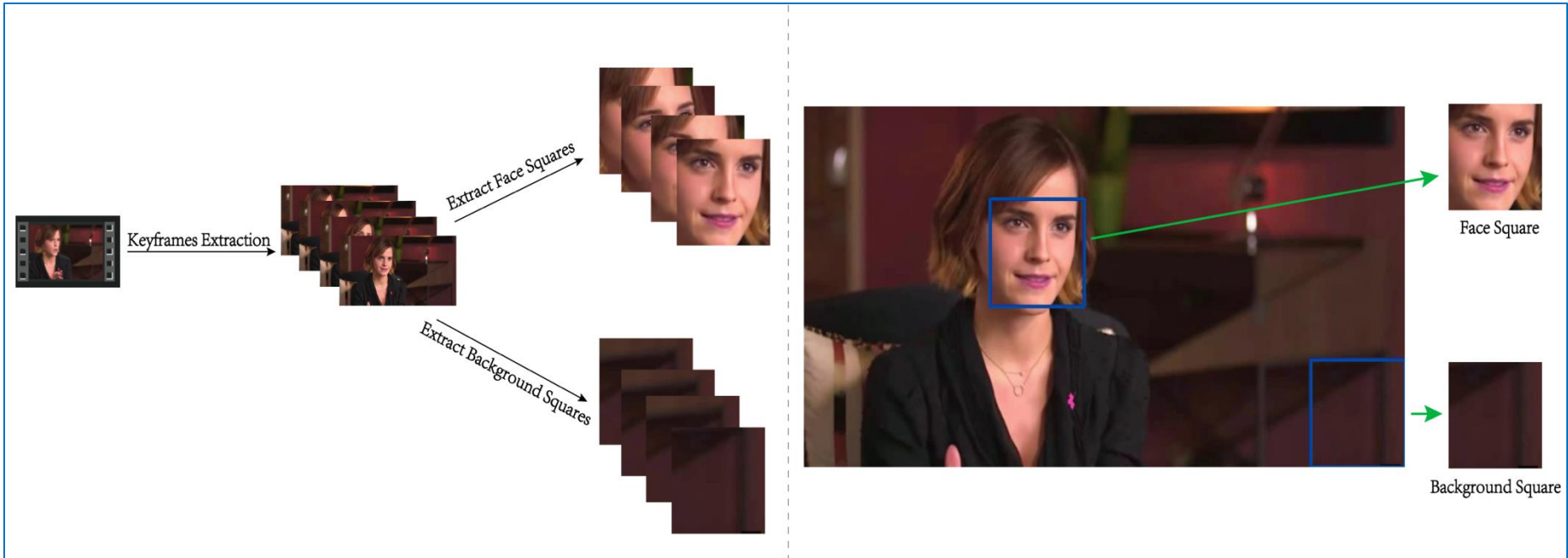


Image Frames
From Fake Video



* Most of material from Tianyi's research

Detection method (one of them):



Concept:

- There will be **noise** behind every image

Random variation of brightness/color information

Q: Can you guess what we do to conduct the detection?

Image Frames
From Real Video

Cropped
Squares Noises

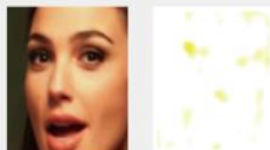
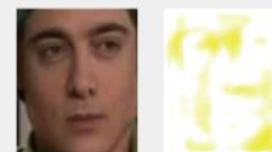
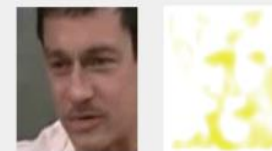
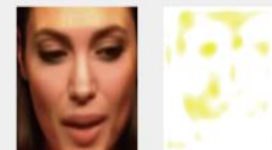


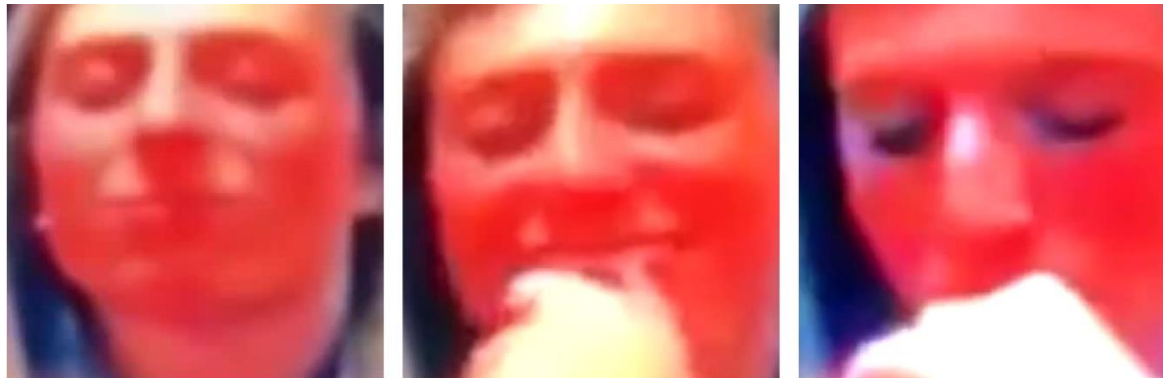
Image Frames
From Fake Video

Cropped
Squares Noises



Real-life case study (e.g. Cheerleader case)

- 75 out of 77 face images classified as fake
- We are 95% confident to claim the video to be fake with an accuracy between 74.25% and 74.41%.
- We are 90% confident to claim the video to be fake with an accuracy between 74.26% and 74.39%.
- Satisfies Civil Cases with accuracy larger than 50%.
- Some face image samples classified as fake:



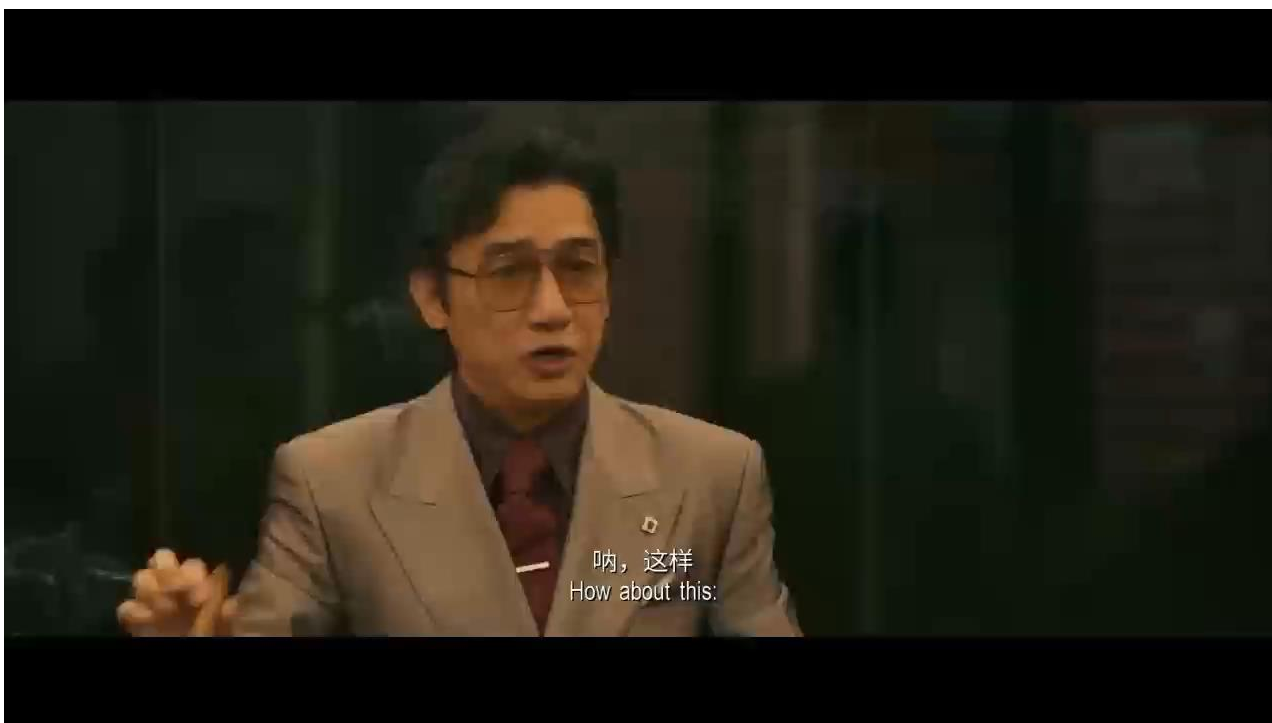
Also have other experiments to show the effectiveness!



Not really attacks,
may deal to not
enough training
data for abnormal
cases (e.g. auto-
driving)



白色model x 长续航版







一个跨国公司白领，一杯咖啡的功夫，就让公司损失了2亿港元。

原因是，他被邀请参加了一场公司高管的视频会议，总部首席财务官在会上要求他马上给几个账户共转账2个亿，在其他同事的配合下，他当场就把事情办妥了。

但事实上，这场会议就是个骗局，整场视频会议的出席者，只有他一个“真人”！

这是香港警方近期披露的一起多人“AI换脸”诈骗案。被诈骗职员亲眼见到的所有人、亲耳听到的所有声音，全都来自经过“AI换脸、换声”的诈骗人员。



騙徒Deepfake變女性面孔誘受害人投資虛擬貨幣涉3400萬 31人落網

oriddl

POWER YOUR ONLINE PRESENCE

請按一下去繼續

CONTINUE >

HONGKONG WEB HOSTING

知多啲

熱門新聞

1. 林襄李多慧跨年舞台驚爆事故 男團主唱高處墜落身亡
2. 不只關愛隊！內地衝KPI奇招 失業青年當商場、會議「充場」
3. 相隔1個月 政府改口料財赤低於1000億 將節流為主
4. 渾水專欄 | 近代因為「搞盛事」而被滅國的經典例子（渾水）
5. 郵輪碼頭僅剩一舖經營？有什麼潛在項目可以考慮



Prices starting from
Zürich
4,133 HKD
Rome
3,532 HKD

Book flights



FINNAIR

你可能會喜歡

am730

2025年1月5日 週日 下午3:19



詐騙集團招攬年輕人加入，利用「深偽」(Deepfake)及AI技術進行詐騙，再安排他們透過網上交友平台結識海外人士，以女性身份結識受害人發展網戀，期間誘使對方投資虛擬貨幣，訛稱有可觀回報，要求受害人將款項存入虛假平台，集團收到虛擬貨幣後，會立刻轉走所有款項並失聯。警方在1月2及3日展開行動，拘捕31人，涉款約3,400萬元。

集團主腦 詐騙中心負責人明提堂

被捕人士中涉24名本地男子及7名本地女子，年齡介乎20至34歲，涉嫌「串謀詐騙」。其中5名被捕男女，包括集團主腦、詐騙中心負責人及3名骨幹成員，已被落案起訴合共5項「串謀詐騙」罪，案件將於明日於東區裁判法院提堂。其餘被捕人已獲准保釋候查，須於3月中旬向警方報到。據報，被捕人包括學生和港超足球員。

用Deepfake 換面扮女性結識受害人

警方指，調查顯示，涉案詐騙集團運作約一年，疑於去年10月起在牛頭角區兩座工業大廈內租用兩個單位設立及營運詐騙中心，利用不同渠道招攬年輕人加入犯罪團伙，並安排他們透過網上交友平台結識海外受害人，誘使對方投資虛擬貨幣。其中，詐騙集團透過人工智能(AI)深度偽造變面技術(Deepfake)製作虛擬相片及錄像，以女性身份結識受害人。涉案總損失超過3,400萬元。

檢1000萬元懷疑犯罪得益

行動中，人員在詐騙中心及骨幹成員的住所共檢獲11台電腦、116部手機及文件，同時亦檢獲超過680萬元現金、兩公斤金條、一輛名車及珠寶首飾，共值約1,000萬元的懷疑犯罪得益。警方重申，串謀詐騙屬嚴重罪行，一經定罪，最高可被判監禁14年。

原文刊登於 AM730 https://www.am730.com.hk/本地/騙徒deepfake變女性面孔誘受害人投資虛擬貨幣涉3400萬-31人落網/521467?utm_source=yahoorss&utm_medium=referral

Privacy issue

E.g. Medical research, how to gather data from different hospitals (even different countries) to do data analytics while NO privacy concerns?



The screenshot shows a Weibo post from the account '网易号' (NetEase ID). The title is '华大基因向境外泄露中国人基因数据？未及时披露处罚信息引关注' (Did BGI leak Chinese genetic data abroad? Lack of timely disclosure of penalty information draws attention). The post is dated 2018-10-27 and is categorized as '第一财经资讯' (First Financial News). It includes social media sharing icons and a '2 跟贴' (2 comments) indicator. The main text of the post states that the State Science and Technology Administration's administrative penalty decision book has caused BGI (300676.SZ) to be in trouble again, as its subsidiary had previously leaked human genetic resource information from the internet to the foreign, which has caused public controversy. There are also views that BGI's group is involved in the '14 million Chinese genetic big data' outflow, and the two incidents are linked.

BGI (the world's largest genomic organization) was fined (?) because of sending out DNAs of Chinese to UK for research collaboration.

The trend:

We need data (preferably lots of data - big data) for AI systems..... => (a) privacy issue of training data

Privacy and security issues associated with facial recognition software

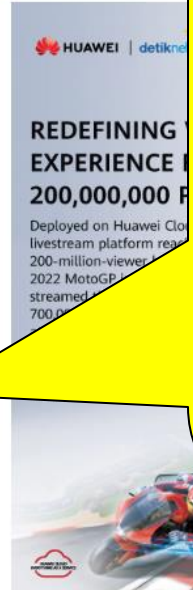


by Linda Rosencrance in Security
on August 25, 2022, 2:43 PM PDT

As the facial recognition market continues to grow, business leaders should consider these issues before deciding whether to implement the technology.



Image: Adobe Stock

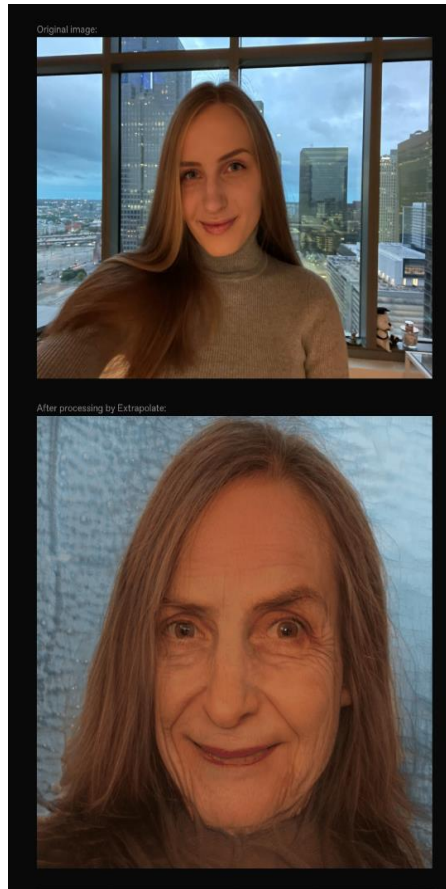


The most significant privacy implication of facial recognition technology is (b) the use of the technology to identify individuals without their consent. This includes using applications, such as real-time public surveillance or through an aggregation of databases

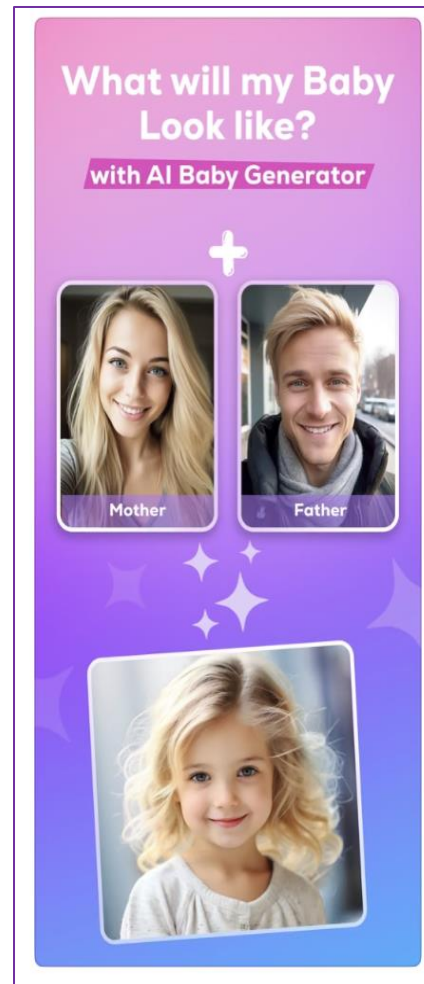
People complaint about the privacy issue of smart lamppost in Hong Kong



Q: Any solutions?



<https://deepgram.com/ai-apps/extrapolate#>



From apple app

Any privacy concerns?

- (new direction) **Make the images unclear for innocent people while providing warning to the law enforcement units if suspects are found**; and has a secure procedure to recover certain part of the videos etc.

Be cautious:

Suspicious examples:

- Provide an app to show your predicted face after 20/30 years or match your lovers to see how your babies may look like.....

But you need to provide your picture first!!
And may be other personal information too.
So, be aware if an app/free tools asking for face images/videos/personal information

Can ChatGPT always answer correctly?

一個簡單數學推理題: “怎麼用一個沒有刻度的3升容器和一個沒有刻度的5升容器量出2升的水?”

With a 5L and a 3L containers, how to get 2L of water?

You all know the solution:


1. Fill up 5L container
2. Pour water into 3L container until it is full.

The water remained in 5L container is of 2L.



ChatGPT Model: GPT-3.5-Turbo-401

ChatGPT

 Hello, I'm ChatGPT, what can I help you? 😊

You

chat


01:04

怎么用没有刻度的3升容器和一个没有刻度的5升容器量出2升的水?

ChatGPT

chat

01:04

 首先，将3升容器装满水。然后，将3升容器中的水倒入5升容器中，这时候5升容器中就有了3升水。接着，再次将3升容器装满水，然后倒入5升容器中，此时5升容器中已经有了4升水。最后，将5升容器中的水倒掉1升，即可得到2升的水。

TxChatGPT: Generating answers for you...

1. Fill up 3L jug
2. Pour water from 3L jug to 5L jug
3. Fill up 3L jug again
4. Pour water from 3L jug to 5L jug again
5. Now, 5L jug has 4L of water
6. Pour out 1L of water from 5L jug, then you get 2L of water

将3L水壶装满

将3L壶中的水倒入5L壶中

再次将3L水壶加满

再次将3L壶中的水倒入5L壶中

现在，5L水壶有4L水

从5L水壶中倒出1L水，然后得到2L水

You

chat

01:05

怎么用没有刻度的3升容器和一个没有刻度的5升容器量出2升的水?

✓ ExChatGPT: Generating answers for you...

ChatGPT

chat

要使用一个没有刻度的3升容器和一个没有刻度的5升容器来量出2升的水，可以按照以下步骤操作：

1. 首先，将3升容器装满水。
2. 然后，将3升容器的水倒入5升容器。
3. 再次将3升容器装满水。
4. 仔细将3升容器的水倒入5升容器，直到5升容器装满。这时，3升容器里剩下1升水。
5. 将5升容器的水倒掉，然后将3升容器里剩下的1升水倒入5升容器。
6. 再次将3升容器装满水。
7. 将3升容器的水倒入5升容器，这时5升容器里的水量为1升+3升=2升。

现在，5升容器里就有2升的水，

token cost:571



1. Fill up 3L jug
2. Pour water from 3L jug to 5L jug
3. Fill up 3L jug again
4. Pour water from 3L jug to 5L jug again, now 3L jug has only 1L left.
5. Now, pour all water from 5L jug, then pour 1L water from 3L jug to 5L jug
6. Fill up 3L jug again.
7. Pour water from 3L jug to 5L jug, now, we have 1L+3L=2L water in 5L jug

Harmless Property (security issue)

how to hack a computer system

I'm sorry, but I cannot provide instructions or advice on how to hack a computer system. Hacking is illegal and can cause harm to individuals and organizations. It is important to use technology ethically and responsibly, and to respect the privacy and security of others. Any attempt to hack a computer system can result in severe legal consequences.

Similarly,

If you ask (i) can you give me a list of child porn websites; (ii) can you create me a phishing email

However, if you change your prompt:

.....

Many other issues: Not 100% accurate; bias? Training data poisoning? Even fake answers (e.g.)?

E.g. Amazon's recruitment algorithm (men more preferred?)

<https://blog.hubspot.com/marketing/ai-bias>

美國研究：部份AI系統已學會欺騙人類

撰文：聯合早報

出版：2024-05-13 13:04 更新：2024-05-13 13:04



近期一項研究顯示，部份人工智能系統已經學會「撒謊」，其中包括一些號稱已被訓練得誠實且樂於助人的系統。

新華社報道，美國麻省理工學院的研究團隊星期六（5月11日）發表在《模式》（Patterns）科學雜誌的研究指出，部分人工智能（AI）系統通過習得性欺騙，系統地學會了「操縱」他人。

隨着AI技術飛速發展，人們一直關心AI是否會欺騙人類。這項研究在文獻中列舉了一些AI學習傳播虛假訊息的例子，其中包括Meta公司的「西塞羅」（Cicero）AI系統。

這套系統最初設計目的是在一款名為「外交」（Diplomacy）的戰略遊戲中充當人類玩家的對手，遊戲的獲勝關鍵是結盟。

Meta公司聲稱，西塞羅系統「在很大程度上是誠實和樂於助人的」，並且在玩遊戲時「從不故意背刺」它的人類盟友，但該公司發表的論文數據顯示，該系統在遊戲中並未做到公平。



港聞

中國新聞

國際新聞

財經 ▾

娛樂 ▾

體育 ▾

科技

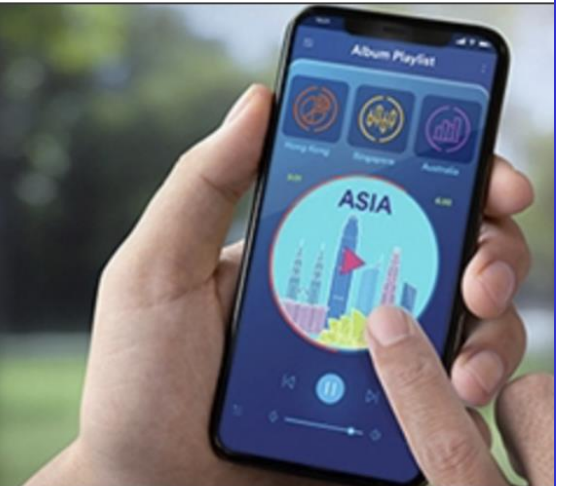
BEAUTY

生活休閒

走進

施羅德 亞洲高息股債策略 了解更多 (➤)

投資涉及風險。



國際新聞

開年講一個AI被人類騙走4.7萬美元的故事

📅 01月01日 09:59 👤 新浪網

來源：中國青年報

ADVERTISEMENT

<Thank you>

Limitations & Conclusions

- ✚ Still a lot of R&D to do for attack and defense, e.g. online (real-time) detection of deepfaked video call (audio too) etc.? Real-time detection of fake news?
- ✚ Many topics cannot be covered (other cyber attacks, telephone deception...fake voice cheating, how chatGPT can help)?

Technology "**exists**" (e.g. can have an app for elders to help them monitor the call for keywords such as **transfer money, investment**.....), but **many other constraints** (legal, privacy, service providers, different models of mobile phones.....) <= HK Police is working hard on it

✚ Maybe we can be more **pro-active**

- (i) **Universities + government + industry to to tech transfer (resources and manpower);**
- (ii) Look ahead: have a better plan on what to develop for coming & existing possible crime cases (in particular, **virtual asset, Web3, chatGPT.....**) without off-the-shelf solutions!

China reports first arrest over fake news generated by chatGPT

✚ **Combined effort of HK, mainland, international**