# “AI 詐騙”冲上熱搜第一,AI 安全或再成焦點

5 月 22 日,“AI 詐騙正在全國爆發”話題冲上熱搜第一,引發網友熱議。據湖北網警巡查執法消息,AI 技術改變詐騙,新騙局來襲後,詐騙成功率竟接近 100%。

第一種是聲音合成。騙子通過騷擾電話録音等來提取某人聲音, 獲取素材後進行聲音合成,從而可以用僞造的聲音騙過對方。

第二種是 AI 換臉,因爲人臉效果更易取得對方信任,騙子用 AI 技術換臉,可以僞裝成任何人,再通過視頻方式進行信息確認。

海外方面,在早些時候,三星在引入 ChatGPT 不到 20 天里, 發生了三起半導體機密資料外泄事故,包括兩起“設備信息泄露”和一起“會議內容泄露”。

3 月 31 日,意大利個人數據保護局宣佈,從即日起禁止使用聊天機器人 ChatGPT,限制其開發公司 OpenAI 處理意大利用戶信息,並開始立案調查。ChatGPT 被指控不遵守有關個人數據的立法並且沒有驗證未成年用戶年齡的系統。

海內外積極因對 AI 風險

5 月 19 日消息,英國計劃與 OpenAI 和 DeepMind 的高管舉行關於 AI 風險的會議, 英國的人工智能峰會最快可能在本月舉行。

5 月 18 日, 科學技術部部長王志剛表示, 中國積極應對人工智能可能帶來的風險挑戰,推進人工智能倫理治理,髮佈人工智能治理原則和倫理規範, 向全球表明負責任人工智能的鮮明立場。

5 月 16 日,OpenAI CEO Altman 呼籲國會對人工智能(AI)系統採取安全標準。他認爲,政府可以將發放牌照和測試的要求結合起來,監管開發 AI 模型的公司。

4 月 11 日,據《華爾街日報》報道,拜登政府已經開始研究是否需要對 ChatGPT 等人工智能工具實行檢查。作爲潛在監管的第一步,美國商務部 4 月 11 日就相關問責措施正式公開徵求意見,包括新人工智能模型在發佈前是否應經過認證程序。徵求意見期限爲 60 天。

4 月 11 日,國家互聯網信息辦公室發佈《生成式人工智能服務管理辦法(徵求意見稿)》,向社會公開徵求意見。該意見稿聚焦生成式人工智能,在算法設計、訓練數據選擇、模型生成和優化、提供服務、生成內容、個人信息和隱私保護方面均做出了限制。

據中信證券研報,AI 可能引發的安全問題主要體現在以下方面:

1)大模型訓練、調優所使用數據的安全問題。

比如前文所述意大利個人數據保護局因年齡審核、數據和隱私安全等問題, 宣佈限制 ChatGPT 運營; 加拿大聯邦隱私監管機構宣佈, OpenAI 涉嫌“未經同意收集、使用和披露個人信息”。

2)大模型生成內容的可靠性、使用方式等問題。

中信證券指出,由於大模型在生成內容上的基于概率統計的原理,其內容可靠性可能存在問題, 如引用的客觀數據是否是最新且來源可靠的,此外,由 AI 生成的新聞圖片出現在互聯網中,眞假難辨,或引發傳播風險;歐洲刑警組織提示稱,ChatGPT 可能被濫用于網絡釣魚、虛假信息和網絡犯罪。

華創證券還提到,網絡攻擊方面,搭載了 AI 能力的黑客如虎添翼,ChatGPT 的強大功能降低了網絡攻擊者用于製作惡意軟件和降低技術門檻,使得攻擊手法轉向分布式、智能化、自動化,智能化對抗將成爲主要的攻防形式。

哪些技術有望用來應對 AI 風險?

中信證券指出,由於大模型在訓練、調優數據過程中涉及用戶數據調用, 伴隨生成式 AI 技術熱潮掀起,在數據採集、傳輸、存儲、處理、共享、銷毀等數據生命周期各階段,通過加密、數據水印等技術手段保護用戶數據特別是隱私成爲當務之急。

針對大模型生成內容的可靠性、使用方式,中信認爲,應對 AI 生成內容進行標記,同時對 AI 生成內容的可靠性及傳播鏈路進行把關。

安信證券提到, 當前對 AI 生成內容的判別主要可以由兩種技術途徑來完成:

1) 通過算法識別 AI 模型生成內容的特徵,從而鑒別相應的內容是否由 AI 生成。根據公司公告披露, 美亞柏科正在佈局相關取證技術,比如對人工智能合成、生成的視頻圖像內容檢測鑒定。

2) 通過對 AI 生成的內容添加特定的標識,來區分相應內容是否由 AI 生成, 這一方向的技術手段包括數字水印等相應的加密技術。

國盛證券指出,AIGC 可能生成的虛假信息,需要類似換臉甄別的技術來鑒別圖像等信息是否由 AI 生成;數據隱私方面,隱私計算可以在不暴露數據的情況下進行計算和分析,以實現數據隱私的保護。

隱私計算是在保障數據隱私安全的同時,實現數據分析計算的技術體系,其價值是打破數據孤島,實現數據的可信流通和價值挖掘。國泰君安稱,隱私計算在政策和需求的雙輪驅動下不斷發展壯大,到 2025 年,中國隱私計算市場規模有望突破百億級,市場空間巨大。

國盛證券還提到, 安全問題不能僅靠 AI 廠商自律解決,還需要外部力量監督。全球 AI 監管處于探索階段,歐盟、英國、美國相關法律法規與指導意見逐步推進。人工智能趨勢不可逆轉,與科學技術的高速發展相比,現有法律和倫理仍處于不斷起步探索的階段。

# AI 詐騙屢見不鮮 人工智能如何執好這把“雙刃劍”

人臉效果更易取得對方信任, 通過分析公衆發佈在網上的面部信息,並利用 AI 換臉技術,在視頻通話中僞裝成想成爲的人,並通過視頻電話的方式進行信息確認,獲取信任進行詐騙。

近日,“AI 詐騙正在全國爆發”的話題冲上熱搜第一。AI 即人工智能縮寫,此事引起人們對人工智能的熱議。

近日, 包頭市公安局電信網絡犯罪偵查局偵破一起使用智能 AI 技術進行電信詐騙的案件。犯罪分子利用人工智能實施電信詐騙,受害人 10 分鐘內被騙 430 萬元。

**更令人背後冒冷汗的是,AI 詐騙早已屢見不鮮**

福州的李先生遭遇了冒充領導的 AI 詐騙,被騙數萬元。騙子通過和李先生打視頻電話,取得了李先生的信任,在後續的聊天中騙子提出想借李先生錢,由於事先有視頻確認,李先生不疑有他,便向“曾某”提供的賬戶轉賬數萬元,後發現被騙。

某公司財務小王接到領導電話, 要求立刻給供應商轉款 2 萬元,並將轉賬信息以郵件形式發送,轉款理由是避免繳納滯納金。由於老闆的口音十分逼眞,小王信以爲眞,在 1 小時內轉款完成,後發現被騙。

AI 詐騙事件頻發,令人心生恐懼,我們瞭解一下 **AI 詐騙的幾種主要方式,以此提高警惕**

**1)利用聲音合成進行詐騙**

即通過騷擾電話録音等來提取某人聲音,獲取素材後進行聲音合成,從而可以用僞造的聲音騙過對方。

**2)利用 AI 換臉進行詐騙**

人臉效果更易取得對方信任,通過分析公衆發佈在網上的面部信息,並利用 AI 換臉技術,在視頻通話中僞裝成想成爲的人,並通過視頻電話的方式進行信息確認,獲取信任進行詐騙。

**3)利用轉發微信語音進行詐騙**

在盜取微信號後,便向其好友“借錢”,爲取得對方的信任,通過轉發之前的語音,進而騙取錢款。當前微信沒有語音轉發功能,但不法分子通過提取語音文件或安裝非官方版本 (插件),實現語音轉發。

通過以上案例及 AI 詐騙方式可以看出,AI 日趨廣泛的應用會引發一些社會問題。爲了讓 AI 科技更好的造福人類, 我國不斷完善相關法律政策

2019 年 2 月, 科技部在北京召開新一代人工智能發展規劃暨重大科技項目啓動會,成立了新一代人工智能治理專業委員會。6 月,國家新一代人工智能治理專業委員會發佈了《新一代人工智能治理原則》。9 月,專委會正式發佈《新一代人工智能倫理規範》。

2021 年發佈的《倫理規範》中提出了更加細化與嚴謹的 6 項基本倫理要求。

2022 年 3 月,中共中央辦公廳、國務院辦公廳印發了《關於加強科技倫理治理的意見》(以下簡稱《意見》, 明確指出科技倫理是開展科學研究、技術開發等科技活動需要遵循的價值理念和行爲規範,是促進科技事業健康發展的重要保障。

**科技是把雙刃劍。未來,人工智能應如何約束**

**1)各方積極發力開發音視頻造假檢測技術**

目前, 聯合國區域間犯罪與司法研究所、美國賽門鐵克等公司研究檢測虛假音視頻技術。臉書與微軟及麻省理工大學合作,發起了“假視頻檢測挑戰”活動,旨在利用人工智能技術更好地甄別深度僞造和合成內容,開發對抗深度僞造技術濫用的方法和工具。我國互聯網企業和研究機構等也積極開展對換臉換聲、合成語音詐騙等技術的研究和檢測。

**2)針對人工智能倫理的高風險場景特別立法**

應按照針對不同風險等級制定不同嚴苛程度的管理思路,可以通過分場景監管,做到有收有放,不僅降低人工智能倫理執法的難度,而且推動實現治理與發展的平衡。

**3) 提陞科研機構和企業對人工智能倫理的認知及自律**

在規避人工智能可能産生的倫理風險上,科研機構和企業更容易在相關實踐中獲得第一手信息,也更應該承擔起構建安全人工智能技術的主要責任。

**4)提高全社會科技倫理意識**

倫理問題涉及到社會行爲準則與規範,而治理倫理問題則需從公共管理的角度出發,在充分瞭解人工智能技術所帶來的潛在社會影響,找到相對應的解決辦法,並形成社會對人工智能倫理的共識。建議利用各種渠道廣泛的進行科技倫理宣傳、活動與交流,提陞公衆的科技倫理意識,進而加強全社會對人工智能倫理的廣泛監督。

結語:科技是把雙刃劍,我們在享受科技帶來福利的同時,也在接受考驗。人工智能時代,有人利用 AI 技術思考的是如何推動人類社會進步,而有的人卻利用 AI 做違法犯罪之事。我們能做的就是加強自身道德修養, 守好道德法律底線, 積極推動 AI 技術發展不斷爲人類謀科技福利,堅決反對利用 AI 技術做違法犯罪之事,同時加強防範意識,做好防範措施,防患于未然,不給不法分子可乘之機。

# 10 分鐘騙走 430 萬,如何提防詐騙新技術?

**01AI 騙子,防不勝防**

AI 進化的速度有多快, 騙術也跟着變得更狠。

福州一位科技公司的老闆, 某天突然接到了個好友的微信視頻, 說是自己的朋友在外地競標,需要 430?萬的保證金,想走他的公司賬戶過個賬。老郭在確認視頻通話里的人和聲音都是他朋友之後,就同意了。

隨後, 對方就給老郭發了一張銀行卡轉賬成功的截圖,說是錢已經轉到老郭的戶頭上了,由於老郭之前已經確認了朋友的身份, 在沒確認錢到帳與否的情況下,直接就把 430?萬轉到了好友給的銀行卡上。

結果,這是一場騙局,還是 AI 操控的。雖然接到報警後,福州、包頭兩地警銀迅速啓動止付機制,成功止付攔截 336.84 萬元,但仍有 93.16 萬元被轉移。

還有利用 AI 進行聲音合成實施的詐騙。騙子通過騷擾電話録音等來提取某人聲音, 獲取素材後進行聲音合成, 從而可以用僞造的聲音騙過對方。

已經有人因爲 AI 聲音被騙了。曾有某公司財務接到領導電話, 要求立刻給供應商轉款 2 萬元,並將轉賬信息以郵件形式發送,轉款理由是避免繳納滯納金,因爲老闆的口音十分逼眞,財務也信了,在 1 小時內轉款完成,轉完才發現被騙。

AI 換臉早已不是什么新鮮事, 很多短視頻鬼畜博主都精通這種操作, 但擬聲技術相對複雜一些,畢竟犯罪分子需要獲取你的聲音,然後通過軟件將聲音變成他們想要的內容。

實際上, 如今的騙子想要獲取聲音也並不難,最普遍的方法就是騷擾電話,只要你接電話後並說了幾句話, 那么犯罪分子就會對你的聲音進行録制,並提取你聲音的信息。

有業內人士表示, 犯罪分子通過這一套流程進行詐騙的成功率基本是 100%,尤其是面向那些並不懂 AI、防範意識可能薄弱的老年群體,AI 詐騙更容易發生。

類似的聲音騙局,國外也早有發生。

一個國外的詐騙公司,使用 AI 模擬了“卷福”的音頻,打電話給一家新西班牙的小電影公司,稱想要和對方合作,並要求對方先打款 20?萬英鎊才能見面,結果在電影公司付完錢之後,他們才發現這次事件,完完全全就是一場 AI 騙局。

對於 AI 騙局,有律師表示,使用人工智能技術手段模仿眞人面容、聲音,冒充親友、熟人、領導等,很容易達到以假亂眞的程度,具有極強的迷惑性。與其他類型電信詐騙相比,AI 詐騙迷惑性更強,針對性更強,辨識更困難。

如今,耳聽,眼見都不一定爲實,各種騙局眞是讓人防不勝防。

**02 AI 有多強大? 讓孫燕姿本人都害怕**

最近,關於 AI 還發生了一個熱點:孫燕姿本人親自回應 AI 孫燕姿。

對於熱門歌手 AI, 孫燕姿自己都承認,“諷刺的是,人類再怎么快也無法超越它。”

她在回應小作文里寫道,“人類無法超越 AI 技術已指日可待, 凡事皆有可能, 凡事皆無所謂。”就算暫時能夠辨認出 AI 歌手和眞人歌手的區別,很快這條護城河也將不復存在。這種被虛擬淹沒眞實的感受,如今已經在全世界蔓延。

AI 不光是唱歌可以完全貼近眞人, 連直播帶貨的活都可以接。

最近有人在直播間里發現, 正在賣貨的竟是當紅女星。然而再定睛一看,這些明星居然是使用了 AI 實時換臉技術的普通主播。

AI 實時換臉,正在直播間悄然出現,楊冪、迪麗熱巴、angelababy 等當紅女星,成爲了 AI 換臉的熱門。雖然目前還沒有曝出被騙新聞,但由於明星臉的帶貨主播往往擁有更好的引流效果, 消費者或明星粉絲很可能會冲着明星效應進行消費而被坑騙。

而這種技術門檻也並不高。有媒體報道,AI 實時換臉全套模型購買價僅需 3.5 萬元。

買不起全套模型,還有“AI 換臉特效插件”供買家選擇,售價數百元,只對電腦配置有一定要求,購買後使用者只需更換素材照片,軟件就能自己運行完成“換臉”。這款軟件不僅能對圖片進行換臉,也能在視頻上進行換臉。

AI 換臉的風靡,還帶來了“AI 脫衣”的新問題。早在今年 3 月,一則“女子地鐵照被 AI 一鍵脫衣傳播”的新聞冲上了熱搜,一名女性搭乘地鐵時的一張照片被人用 AI 軟件“一鍵脫衣”,遭到惡意傳播, 罪犯還廣泛散佈不實信息對該女子進行污衊。

更可怕的是,AI 深度僞造(DeepFake)內容,可以利用互聯網平台在全球範圍內病毒式傳播。比如 AI 現在居然還能通過發佈假新聞,讓股市抖三抖。

幾天前,多個經過驗證的推特賬戶,齊刷刷分享了一張聲稱在五角大樓附近發生爆炸的假照片,頓時引起轟動,並導致股市短暫下跌。

這張照片雖然具有 AI 生成的所有特徵,但是因爲被許多經過驗證並打上了藍色勾的推特賬戶分享,不少人還是信以爲眞,謠言一傳十十傳百,甚至有人聲稱這張照片與彭博新聞有關。

連印度一家電視台都信了, 共和國電視台(Republic TV)報道稱五角大樓發生了爆炸,並在節目中播放了這張假圖片,後來,當發現美國當地官員證實沒有發生此類事件後, 共和國電視台快速撤回了這則報道, 這一番操作下來也是相當尷尬。

**03 AI 騙局,並非沒有破綻**

隨着 AI 技術的不斷發展,被 AI 騙的人越來越多。根據殺毒軟件 McAfee 公佈的最新報告,基於人工智能的語音詐騙越發猖獗, 在接到 AI 詐騙電話的群體中,有 77% 的人會上當受騙。

今年 3 月,一份名爲《暫停大型人工智能研究》的公開信在未來生命研究所官網上發佈,包括特斯拉 CEO 埃隆·馬斯克(Elon Musk)、蘋果公司聯合創始人史蒂夫? 沃茲尼亞克(Steve Wozniak)等人在內的千名科學家、企業高管簽字呼籲暫緩 GPT-4 以上版本的研發,他們矛頭所指都是強大 AI 技術可能的負面用途。

不過 AI 詐騙雖然逼眞,但它並不是沒有破綻,比如 AI 換臉技術會導致臉部變形,一些主播低下頭頭頂會發黑,AI 的聲音,也很難表達眞實的主觀情緒。

正如不少網友所說,AI 孫燕姿“根本不是孫燕姿的唱法。”“AI 孫燕姿能模仿她的音色,但模仿不到她的感情;AI 孫燕姿能模仿她的聲音,但模仿不了她的現場。”

與此同時,國內與 AI 換臉、換聲相關的法律法規也在不斷完善。

在過去,“AI 換臉” 被大量使用到影視劇二度創作、趣味惡搞視頻、表情包製作當中,其中隱藏的法律風險和侵權糾紛, 如今都在得到重視,不少擦邊行爲已經得到了及時的懲治。

比如近期,據企查查官網信息顯示,上海魚腥草信息科技有限公司就因旗下一款換臉手機 APP 遭多位網紅博主起訴, 很快在監管部門約談後,這款 App 下架,因爲換臉侵犯了公民的肖像權。

去年 12 月,《互聯網信息服務深度合成管理規定》發佈,對人臉生成、替換、操控,合成人聲、仿聲等都有明確約束。顯然,隨着制度設計更精準,相關立法更健全,AI 肆意換臉的空間將越來越小。

就在 4 月 11 日, 國家互聯網信息辦公室又迅速起草了《生成式人工智能服務管理辦法(徵求意見稿)》, 涉及的生成式人工智能包括基于算法、模型、規則生成文本、圖片、聲音、視頻、代碼等內容的技術,可以說,對當下的 AI 僞造都進行了更爲全面的約束。