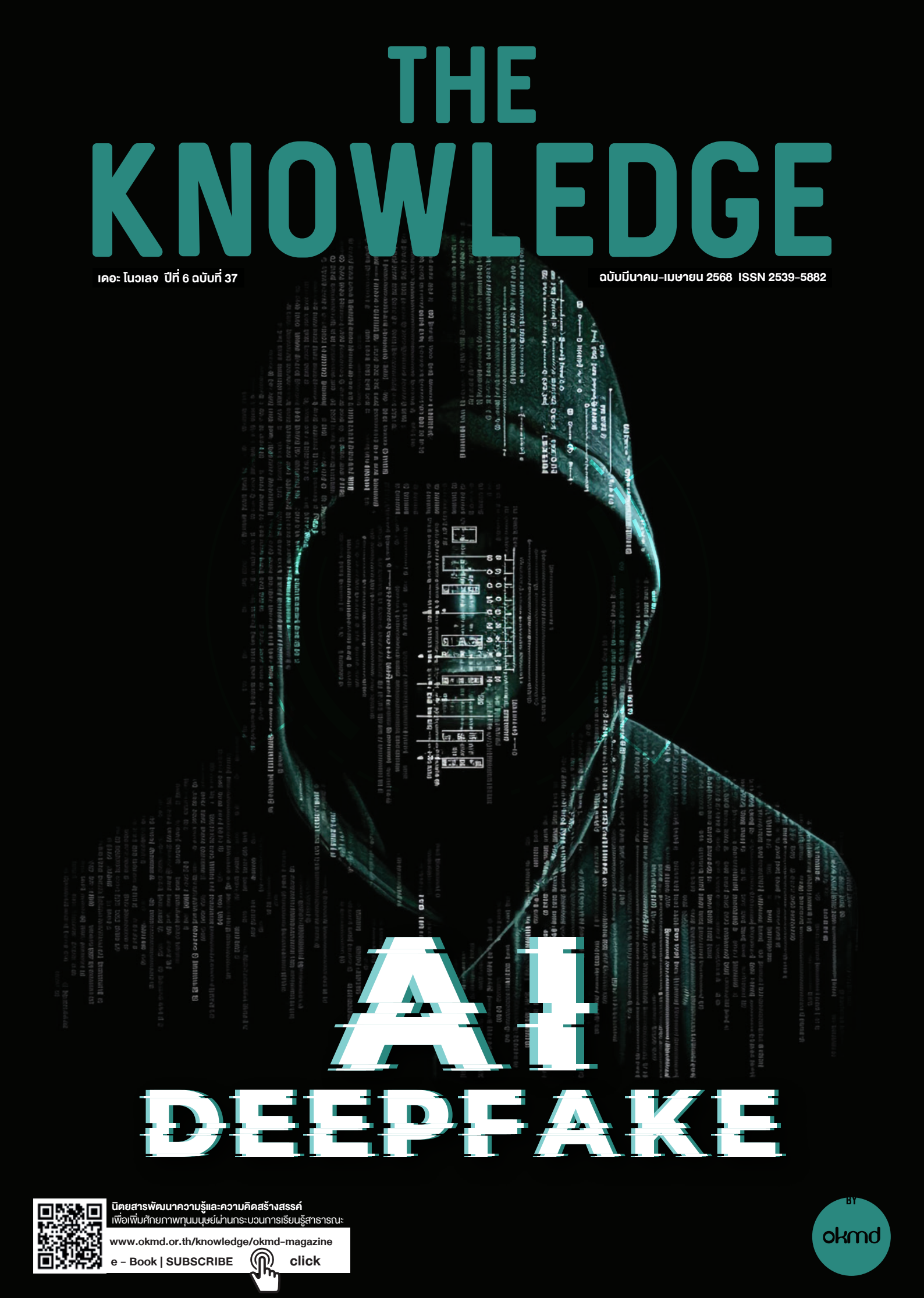
THE KNOWLEDGE

เดอะ โนวเลจ ปีที่ 6 ฉบับที่ 37

ฉบับมีนาคม-เมษายน 2568 ISSN 2539-5882



AI DEEPFAKE

นิตยสารพัฒนาความรู้และความคิดสร้างสรรค์
เพื่อเพิ่มศักยภาพทุนมนุษย์ผ่านกระบวนการเรียนรู้สาธารณะ

www.okmd.or.th/knowledge/okmd-magazine

e - Book | SUBSCRIBE click

BY okmd

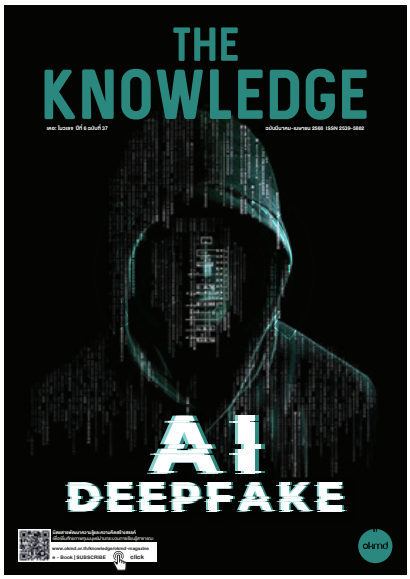
6

22

12

28

CONTENTS



เพราะปัญญาประดิษฐ์เข้ามามีอิทธิพลมากขึ้นทุกขณะ การรู้เท่าทันเทคโนโลยี รวมถึง AI Deepfake ซึ่งเป็นหนึ่งในรูปแบบของปัญญาประดิษฐ์ ที่เปรียบได้กับเหรียญ 2 ด้าน เนื่องจากมีทั้งโอกาสและความเสี่ยง จึงเป็นสิ่งสำคัญในยุคที่ขับเคลื่อนด้วยเทคโนโลยี

นิตยสาร The Knowledge ฉบับที่ 37 ปีที่ 7 ฉบับมีนาคม-เมษายน พ.ศ. 2568 โดย OKMD ตระหนักถึงความสำคัญดังกล่าว จึงได้รวบรวมหัวข้อความรู้หลากหลาย นำเสนอผ่านประเด็นหลัก AI Deepfake เพื่อให้ทุกคนได้เห็นถึงบทบาทที่เด่นชัดของเทคโนโลยีนี้ โดยมีเรื่องราวของการเปิดโปงการเสแสร้งของ AI เมื่อเครื่องจักรแกล้งทำเป็นว่านอนสอนง่าย ที่กระตุกต่อมคิดว่าปัญญาประดิษฐ์พัฒนาอย่างเหนือชั้นไปได้ไกลแค่ไหน รวมถึง 5 อันตรายที่มากับ AI Deepfake ที่ชี้ให้เห็นความเสี่ยงที่เกิดขึ้นได้จริง การถอดรหัสกลโกง Deepfake เพื่อรู้เท่าทันด้านมืดของ AI วิธีจับสังเกตสิ่งผิดปกติจาก AI Deepfake ก่อนตกอยู่ในความเสี่ยง ตามด้วยเรื่องราวของวิวัฒนาการจากหลอกซึ่งหน้า สู่กลลวงด้วยเอไอ ก่อนจะเดินทางมาสู่ยุคการหลอกลวงด้วย AI แท้จริงการหลอกลวงเกิดขึ้นหลายรูปแบบ และ AI Deepfake ภัยคุกคามที่ต้องเตรียมพร้อมรับมือ โดยต้องมาจากการร่วมมือของภาครัฐ ภาคธุรกิจ และภาคพลเมืองดิจิทัลที่ประสานเป็นหนึ่งเดียวกัน

สำหรับ A-Z ชวนรู้คำศัพท์เกี่ยวกับ Deepfake จะทำให้ทุกคนได้รู้จักเทคโนโลยีการหลอกลวงมากขึ้น อย่าง Y-YOLO ว่าไม่ใช่ศัพท์การตลาด แต่เป็นแนวคิดใช้ชีวิตให้เต็มที่ที่ย่อมาจาก You Only Live Once ที่ย่อมาจากเครื่องมือตรวจจับว่าเป็นเนื้อหาปลอมหรือจริง You Only Look Once ขณะเดียวกัน Deepfake AI ยังเป็นตัวแปรด้านตัวเลขทางเศรษฐกิจ ในหัวข้อ AI Deepfake กับตัวเลขต้องติดตาม และอีกหลายบทความ รวมถึงเด็ก ๆ รู้จัก Deepfake แค่ไหน โดยการพารู้ทัน AI ตั้งแต่ในชั้นเรียน ที่ตอกย้ำถึงความสำคัญของการเข้าถึงความรู้อย่างเท่าเทียมและการเปิดรับความรู้ใหม่ ๆ ที่จะเป็นเกราะป้องกันอันตรายและเป็นทักษะชีวิตที่จำเป็นในทุกช่วงวัย

OKMD TEAM

okmd
Office of Knowledge Management and Development

ที่ปรึกษา
ดร.ทวารัฐ สูตะบุตร
ผู้อำนวยการสำนักงานบริหารและพัฒนาองค์ความรู้
บรรณาธิการบริหาร
ดร.อภิชาติ ประเสริฐ
รองผู้อำนวยการสำนักงานบริหารและพัฒนาองค์ความรู้
บรรณาธิการ
นายโตมร ศุขปรีชา
ผู้อำนวยการสำนักยุทธศาสตร์และนวัตกรรมการเรียนรู้

จัดทำโดย
สำนักงานบริหารและพัฒนาองค์ความรู้ (องค์การมหาชน)
69 อาคารวิทยาลัยการจัดการ มหาวิทยาลัยมหิดล
ชั้น 18-19 ถนนวิภาวดีรังสิต แขวงสามเสนใน
เขตพญาไท กรุงเทพมหานคร 10400
โทรศัพท์ 0 2105 6500
โทรสาร 0 2105 6556
อีเมล theknowledge@okmd.or.th
เว็บไซต์ www.okmd.or.th

จัดทำขึ้นภายใต้โครงการเผยแพร่กิจกรรมองค์ความรู้โดยสำนักงานบริหารและพัฒนาองค์ความรู้ (องค์การมหาชน) เพื่อสร้างแรงบันดาลใจในการนำองค์ความรู้มาผสมผสานกับความคิดสร้างสรรค์ เพื่อประโยชน์ด้านการเรียนรู้ ต่อยอดธุรกิจ เพิ่มมูลค่าเศรษฐกิจของประเทศ

ผู้สนใจรับนิตยสารโปรดติดต่อ 0 2105 6520 หรือดาวน์โหลดที่เว็บไซต์ www.okmd.or.th/knowledge/okmd-magazine

5 อันตราย ที่มากับ AI Deepfake

ปัจจุบันเรามักได้ยินคำว่า AI Deepfake กันอยู่บ่อยๆ เนื่องจากเป็นเครื่องมือที่มีความสามารถในการปลอมแปลงภาพ วิดีโอ และเสียงขึ้นมาใหม่ได้อย่างสมจริงด้วยเทคโนโลยีเอไอ จนแทบแยกไม่ออกว่าเป็นของจริงหรือของปลอม ทำให้มีการนำมาใช้เพื่อสร้างศิลปะและความบันเทิง แต่ภายหลังได้ก้าวกระโดดจากขอบเขตของศิลปะและความบันเทิง ไปสู่ประเด็นที่ท้าทายจริยธรรมและความปลอดภัยในสังคม เมื่อมีการใช้ AI Deepfake ในการปลอมแปลงข้อมูลเพื่อหลอกลวงเหยื่อ จนก่อให้เกิดผลกระทบอันตรายเกินกว่าที่คาดคิด โดยต่อไปนี้เป็น 5 อันตรายหลัก ซึ่งมาพร้อมกับเครื่องมือนี้

1 การบิดเบือนข้อมูลและสร้างข่าวปลอม (Disinformation and Fake News)

หนึ่งในภัยคุกคามที่ชัดเจนที่สุดของ Deepfake คือ การสร้างข่าวปลอมหรือวิดีโอปลอมที่ดูสมจริงจนยากจะแยกแยะได้ ซึ่งเป็นการบิดเบือนข้อเท็จจริง โดยวิดีโอที่แสดงบุคคลสำคัญ เช่น ผู้นำประเทศ หรือบุคคลสาธารณะในสถานการณ์ที่ไม่เคยเกิดขึ้นจริง สามารถทำให้ประชาชนหลงเชื่อข้อมูลผิดๆ อันส่งผลกระทบต่อความน่าเชื่อถือของข่าวสาร ทำลายภาพลักษณ์ของบุคคล และก่อให้เกิดความวุ่นวายทางสังคมได้อย่างร้ายแรง ยกตัวอย่าง การเผยแพร่วิดีโอปลอมที่แสดงถึงการประกาศสงครามหรือเหตุการณ์ฉุกเฉินทางการเมือง

2 การก่ออาชญากรรมไซเบอร์และการแบล็กเมล (Cybercrime and Blackmail)

โดยการนำ Deepfake มาใช้เป็นเครื่องมือในการสร้างสื่อปลอมเพื่อแบล็กเมลและก่ออาชญากรรมไซเบอร์ โดยเฉพาะในรูปแบบวิดีโอละเมิดสิทธิ์ทางเพศ หรือสร้างภาพอนาจารปลอมของบุคคลที่ตกเป็นเหยื่อ ซึ่งการกระทำเช่นนี้สามารถทำลายชื่อเสียง สร้างความอับอาย และส่งผลต่อสภาพจิตใจของผู้เสียหายอย่างรุนแรง รวมถึงทำให้เกิดการแบล็กเมลเพื่อเรียกเงินหรือผลประโยชน์อื่นๆ และการขยายตัวของตลาดมืดสำหรับการซื้อขายเนื้อหา Deepfake ที่ผิดกฎหมายได้อีกด้วย

3 การทำลายชื่อเสียงและความน่าเชื่อถือ (Damage to Reputation and Credibility)

โดยบุคคลสาธารณะหรือผู้บริหารระดับสูงขององค์กรอาจตกเป็นเป้าหมายของการใช้ Deepfake เพื่อใส่ร้ายป้ายสี หรือบ่อนทำลายความเชื่อมั่นในตัวบุคคลหรือองค์กร เช่น วิดีโอที่ปลอมแปลงขึ้นมาเพื่อแสดงการกระทำที่ไม่เหมาะสมหรือกล่าวถ้อยคำรุนแรง สามารถสร้างความเสียหายต่อความสัมพันธ์ทางธุรกิจและสังคมได้อย่างร้ายแรง

4 การแอบอ้างตัวตนและขโมยข้อมูลส่วนตัว (Identity Theft and Personal Data Theft)

ด้วยความสามารถของ Deepfake ในการปลอมแปลงเสียงและภาพ จึงถูกนำไปใช้ในการแอบอ้างตัวตน เพื่อเข้าถึงบัญชีธนาคารหรือขโมยข้อมูลสำคัญ โดยผู้ไม่หวังดีสามารถใช้เสียงปลอมในการหลอกลวงระบบยืนยันตัวตนด้วยเสียง หรือใช้วิดีโอปลอมในการหลอกลวงธุรกรรมที่ต้องการการยืนยันด้วยภาพ

5 การบ่อนทำลายความเชื่อมั่นในเทคโนโลยีและระบบยุติธรรม (Erosion of Trust in Technology and the Justice System)

เมื่อ Deepfake ทำให้ความจริงและความเท็จแยกออกจากกันได้ยากขึ้น ประชาชนจึงหมดความเชื่อถือในเทคโนโลยีและหลักฐานดิจิทัล ซึ่งอาจส่งผลกระทบต่อระบบกฎหมาย เพราะหลักฐานจากวิดีโอหรือเสียงนั้นอาจถูกตั้งข้อสงสัยว่าเป็นของปลอม ทำให้เกิดปัญหาในการพิสูจน์ความจริง และลดประสิทธิภาพของกระบวนการยุติธรรม

ดังนั้น การสร้างความตระหนักรู้เกี่ยวกับผลกระทบและอันตราย อันเกิดจาก AI Deepfake จึงเป็นสิ่งสำคัญ เพราะจะช่วยให้ทุกคนรู้เท่าทัน ไม่ตกเป็นเหยื่อของการหลอกลวง และสามารถใช้ประโยชน์จากเครื่องมือหรือเทคโนโลยีนี้อย่างถูกต้อง ปลอดภัย และมีความรับผิดชอบ

อ้างอิง :
• thethaiger.com

วิวัฒนาการจากหลอกซึ่งหน้าสู่กลลวงด้วยเอไอ

ในยุคแห่งเทคโนโลยีก้าวไกล ซึ่งทำให้เอไอสามารถสร้างเนื้อหาปลอมทั้งในรูปแบบภาพ เสียง และวิดีโอได้อย่างสมจริง ส่งผลให้กลุ่มมิจฉาชีพฉวยโอกาสนำไปใช้หลอกลวงหรือตกเหยื่ออย่างแพร่หลาย จนกลายเป็นปัญหาร้ายแรงนั้น ถือเป็นอีกขั้นหนึ่งของวิวัฒนาการการหลอกลวงที่แนบเนียนยิ่งขึ้น

นี่จึงเป็นความจำเป็นที่ทุกคนต้องหันมาตระหนักรู้ถึงปัญหา โดยการย้อนรอยวิวัฒนาการกลลวงตั้งแต่อดีตถึงปัจจุบัน เพื่อเป็นเกราะป้องกันการถูกหลอกในยุคที่โลกนี้มันช่างเฟก

ย้อนรอยกลลวงหรือความเฟก

ประเทศไทยเป็นประเทศที่ถูกหลอกลวงสูงติดอันดับ 2 ของเอเชีย รองจากเกาหลีใต้ โดยการหลอกลวงมีอยู่หลายรูปแบบ เช่น หลอกให้ดาวน์โหลดโปรแกรมหรือแอปพลิเคชันอันตราย หรือหลอกให้เข้าไปที่หน้าชอปปิงออนไลน์ปลอม โดยจากการเปิดเผยข้อมูลของบริษัท Gogolook จำกัด ผู้พัฒนา Whoscall แอปพลิเคชันระบุตัวตนสายเรียกเข้าที่ไม่รู้จัก และป้องกันสแปมสำหรับสมาร์ทโฟน ระบุว่ามีกรณีหลอกลวงต่างกรรมต่างวาระกัน แต่รวมแล้วสามารถสร้างมูลค่าความเสียหายสะสมสูงอย่างคาดไม่ถึง ยกตัวอย่างในปี พ.ศ. 2566 พบคนไทยได้รับข้อความ SMS ที่เป็นข้อความสแปมและข้อความหลอกลวงสูงถึง 63% ของข้อความทั้งหมดที่ได้รับ

นอกจากนั้น คนไทยยังตกเป็นเหยื่อของมิจฉาชีพถึงวันละ 217,047 ราย โดยแบ่งเป็นการได้รับสายจากมิจฉาชีพ 20.8 ล้านครั้ง และถูกมิจฉาชีพหลอกลวงจาก SMS มากกว่า 58.3 ล้านข้อความ เทียบกับปี พ.ศ. 2565 มียอดรวมของสายโทรศัพท์หลอกลวงเพิ่มขึ้น 22% และข้อความหลอกลวงจาก SMS เพิ่มขึ้น 17% รวมมูลค่าความเสียหายสะสมกว่า 53.9 พันล้านบาทเลยทีเดียว

ขณะเดียวกัน จากรายงาน State of Asia Scams in Thailand จัดทำโดยองค์กรต่อต้านกลโกง Global Anti-Scam Alliance ร่วมกับ Gogolook สำรวจพบว่าปี พ.ศ. 2567 มีคนไทยกว่า 89% ต้องรับมือกับมิจฉาชีพอย่างน้อยเดือนละครั้ง และ 58% ถูกหลอกบ่อยขึ้นเมื่อเทียบกับปี พ.ศ. 2566 โดยวิธีที่นิยมมากที่สุด คือ การโทรหาเหยื่อ รองลงมาเป็นการส่งข้อความผ่านมือถือหรือแอปพลิเคชัน การโพสต์บนโซเชียลมีเดีย และโฆษณาออนไลน์

ดังนั้น เมื่อเทคโนโลยียิ่งมีความก้าวไกล กลโกงยิ่งมีความแยบยลมากขึ้น โดยเฉพาะในรูปแบบ Deepfake ที่หลอกเหยื่อได้ง่ายดายกว่าเดิม รวมถึงขยายผลจากการหลอกเอาเงิน ไปสู่การสร้างความเข้าใจผิดในสังคม เช่น ส่งข่าวปลอมสร้างความแตกแยก สร้างภาพและเสียงปลอมทำลายชื่อเสียงของบุคคลหรือองค์กรต่างๆ

จากหลอกซึ่งหน้า สู่กลลวงด้วยเอไอ

โลกมีความก้าวหน้าตลอดเวลา เช่นเดียวกับมิจฉาชีพที่มีการพัฒนาเทคนิคการหลอกลวงอย่างล้ำสมัย ซึ่งจะเห็นได้จากวิวัฒนาการตามบริบทของเทคโนโลยี เริ่มตั้งแต่วิธีหลอกซึ่งหน้าในยุค 1.0 ก่อนจะพัฒนาสู่การหลอกด้วยเอไอในยุค 5.0

ยุค 1.0

การหลอกซึ่งหน้า มาในรูปแบบหลอกแบบตัวต่อตัว

โดยวิธีการหลอกจะมีตั้งแต่หลอกขายของตามบ้าน หลอกไปกดเอทีเอ็ม หรือหลอกไปทำงานเป็นแรงงาน ซึ่งมักจะเจาะจงเหยื่อผู้สูงอายุที่อยู่บ้านตามลำพัง

ยุค 2.0

การโทรมาหลอก

มาในรูปแบบที่เพิ่มการโทรและการส่ง SMS มาหาเหยื่อพร้อมกับหลอกว่าได้รับรางวัลหรือทำผิด ต้องโอนเงินค่าปรับ ซึ่งมักจะเลือกหลอกเหยื่อในลักษณะสุ่มหรือกระจายไปในวงกว้างขึ้น

ยุค 3.0

การสร้างเพจหลอก

นอกจากการโทรและส่ง SMS ยุคนี้ยังเพิ่มรูปแบบการสร้างเพจขายของแต่ไม่มีของหรือร้านค้าอยู่จริง และการสร้างตัวตนปลอม โดยมีวิธีการหลอกที่หลากหลาย เช่น หลอกขายด้วยประโยคเด็ด ‘ของมันต้องมี‘ โนมน้าวและจูงใจให้โอนเงินเลย หรือหลอกว่าเป็นคนใกล้ชิดและขอความช่วยเหลือ

ยุค 4.0

การหลอกแบบรู้ใจ

มาในรูปแบบการส่ง SMS การติดต่อผ่านโซเชียลมีเดีย และเว็บ 3.0 โดยวิธีการหลอกที่ซับซ้อนขึ้น เช่น มีการวิเคราะห์ข้อมูลของเหยื่อ หรือจับประเด็นความสนใจและระบุข้อมูลส่วนตัวได้ตรงจุด แล้วหลอกให้ลงทุนออนไลน์ หลอกให้ชำระเงินหน่วยราชการ หรือหลอกให้คลิก ‘อ่านต่อ‘ เรื่องที่สนใจ ซึ่งมักจะเจาะจงเหยื่อได้อย่างแม่นยำจากฐานข้อมูลที่มีอยู่

ยุค 5.0

การหลอกด้วยเอไอ หรือ Deepfake

เป็นยุคที่เพิ่มรูปแบบการใช้เอไอสร้างภาพ เสียง หรือวิดีโอปลอมอย่างสมจริงมาหลอก โดยมีทั้งวิธีการส่ง SMS โทรหาด้วยเสียงของคนรู้จัก และระบุข้อมูลตัวตนได้ถูกต้อง หรือการโทรมาเก็บเสียงของเหยื่อเพื่อนำไปใช้หลอกต่อ ซึ่งสามารถเจาะจงเหยื่อได้แม่นยำขึ้นจากฐานข้อมูลที่ได้มาก่อน และเชื่อมโยงจากบุคคลอ้างอิง

เพราะฉะนั้น ท่ามกลางโลกนี้มันช่างเฟก ทุกคนจึงต้องตามติดให้รู้เท่าทันกับกลลวงที่มีวิวัฒนาการตลอดเวลาด้วย จะได้ไม่ตกเป็นเหยื่อของกลุ่มมิจฉาชีพและผู้ไม่ประสงค์ดีทั้งหลาย

อ้างอิง :
• bangkokpost.com • brandbuffet.in.th • the101.world • whoscall.com

มุมมองด้านจิตวิทยา ที่เชื่อมโยงได้กับ AI Deepfake

AI Deepfake หรือเทคโนโลยีที่สามารถสร้างสื่อสังเคราะห์เองได้ เช่น วิดีโอหรือเสียงที่ดูเหมือนจริงจนแยกไม่ออก นอกจากจะเป็นประเด็นที่น่าสำรวจในปัจจุบัน ยังส่งผลกระทบต่อจิตใจของผู้คนมากกว่าที่คิดและในหลายมิติ

ความเครียดและความวิตกกังวลเกี่ยวกับการใช้ AI Deepfake ในทางที่ผิด

เป็นความเครียดที่เกิดได้จาก 2 ที่มาหลัก ได้แก่ ความเครียดและความวิตกกังวลว่าเวลานี้อาจมีผู้ประสงค์ร้ายนำเสียงของเราไปใช้โดยไม่ได้รับอนุญาต ไม่ได้รับความยินยอม ทำให้เกิดความรู้สึกว่าไร้หนทางและถูกล่วงละเมิด เกิดความกลัวที่จะมีคนโคลนเสียงและภาพที่เรียกว่า ความกลัวแฝดคนละฝา หรือ Doppëlganger-Phobia สำหรับอีกหนึ่งความเครียดและความกังวล คือ ความกลัวว่าวันหนึ่งเราจะถูก AI Deepfake เข้ามาหลอกให้เสียทรัพย์สินในวันที่เราอาจไม่ทันได้ระวังตัว

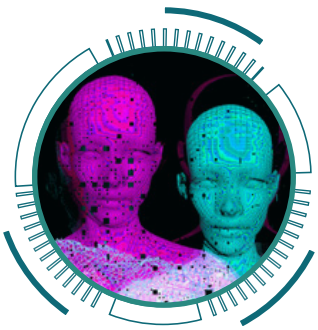
ความไม่แน่นอนและการขาดการไว้วางใจจากเส้นแบ่งความจริงและจินตนาการ

เราไม่สามารถวางใจได้เลยว่า วิดีโอ รูปภาพ หรือเสียงที่ได้ยิน เป็นเสียงจริงหรือเสียงสังเคราะห์ จึงนำไปสู่ความกังวลและการตรวจสอบหลายขั้นตอนเพื่อให้คลายกังวลใจ เช่น การพึ่งพาการตรวจสอบจากภายนอกเกี่ยวกับความถูกต้องของภาพหรือวิดีโอ อย่างภาพความโหดร้ายต่อเด็กในสงครามที่ถูกสร้างให้ผู้คนมีอารมณ์ร่วมมากขึ้น (แม้ไม่จำเป็นต้องสร้าง) ในสงครามระหว่างอิสราเอล-ฮามาส จากกาซาใน ค.ศ. 2023 หรือภาพการปรากฏตัวของนักร้องดังเคที เพร์รี (Katy Perry) ในงาน MET Gala ค.ศ. 2024 ทางอินสตาแกรมของเธอที่ทำให้แฟนๆ สับสน เพราะเธอไม่ได้ไปร่วมงาน แต่อย่างใดพร้อมข้อความที่กำกับไว้ว่า ฉันไม่ได้ไปเพราะติดงาน

ที่มาภาพ : endeavor, endeavor.org

ความกังวลเกี่ยวกับการแยกตัวของตัวตนและความถูกต้องของ AI Clone

ตัวตนดิจิทัลมีส่วนทำให้เกิดการแบ่งแยกตัวตน และอาจเปิดโอกาสให้ผู้คนแสดงตัวตนที่ยืดหยุ่นหรือเป็นอิสระมากขึ้นได้เช่นกัน อย่างไรก็ดี ผลกระทบของการมี AI Clone นั้นยังไม่เป็นที่ทราบแน่ชัด และอาจขึ้นอยู่กับรูปแบบการใช้งาน บริบท และการควบคุม หนึ่งในตัวอย่างน่าสนใจ คือ วิดีโอ Reid Hoffman Meets His AI Twin ที่ Reid Hoffman ผู้บริหารของ LinkedIn ได้ทดลองสร้างอวาตาร์ขึ้นในชื่อ REID AI เพื่อพูดคุยกับตนเอง โดยสร้างขึ้นจาก Hour One เสียงโดย Eleven Labs บุคลิกภาพและการตอบสนองโดยแชตบอต GPT-4 ที่ได้รับการฝึกฝนเกี่ยวกับหนังสือ คำปราศรัย พอดแคสต์ และสื่ออื่นๆ ของ Hoffman



การสร้างความทรงจำเท็จ ทำให้เกิดความทรงจำปลอมที่ส่งผลทั้งผลลัพธ์ซับซ้อนและผลเสีย

การวิจัยเกี่ยวกับ AI Deepfake แสดงให้เห็นว่าความคิดเห็นของผู้คนสามารถเปลี่ยนแปลงได้จากการโต้ตอบกับสำเนาดิจิทัล แม้จะรู้ว่าบุคคลนั้นไม่ใช่บุคคลจริงก็ตาม สิ่งนี้สามารถสร้างความทรงจำ ‘ปลอม’ เกี่ยวกับใครบางคนได้ โดยความทรงจำปลอมเชิงลบอาจส่งผลเสียต่อชื่อเสียงของบุคคลที่ถูกแสดง และความทรงจำปลอมเชิงบวกอาจส่งผลต่อความสัมพันธ์ระหว่างบุคคลที่ซับซ้อนและคาดไม่ถึงได้เช่นกัน เช่นเดียวกับการโต้ตอบกับ AI Clone ของตนเองที่อาจส่งผลให้เกิดความทรงจำปลอม

ความเสี่ยงที่จะเกิดการรบกวนความเศร้าโศก

ผลกระทบทางจิตวิทยาจาก Griefbot หรือ Thanabot หรือ AI Clone ที่หมายถึง AI Deepfake เลียนแบบคนที่เสียชีวิตไปแล้ว ต่อคนที่รักยังไม่เป็นที่ทราบแน่ชัด อย่างไรก็ตาม ปัจจุบันการบำบัดความเศร้าโศกบางรูปแบบเกี่ยวข้องกับการสนทนาในจินตนาการกับผู้ที่เสียชีวิตด้วยวิธีล้ำสมัย ซึ่งโดยปกติวิธีนี้คือขั้นตอนการบำบัดสุดท้ายหลังเข้ารับการบำบัดหลายครั้ง โดยมีนักบำบัดคอยให้คำแนะนำด้วยตนเอง มากกว่านั้น AI Deepfake ของคนรักที่เสียชีวิตไปแล้ว ปัจจุบันกำลังเป็นธุรกิจที่กำลังเฟื่องฟูในจีน ตามรายงานจาก MIT Technology Review

การเพิ่มความเครียดเกี่ยวกับการรับรองความถูกต้อง

การเพิ่มขึ้นอย่างรวดเร็วของ AI Deepfake ที่เกี่ยวกับการฉ้อโกงและการหลีกเลี่ยงการพิสูจน์ตัวตนด้วยข้อมูลชีวภาพ เพิ่มแรงกดดันในการนำวิธีการพิสูจน์ตัวตนใหม่ๆ มาใช้ ผู้เชี่ยวชาญส่วนหนึ่งคาดว่าภายใน ค.ศ. 2026 บริษัทจำนวน 30% จะสูญเสียความเชื่อมั่นในโซลูชันการพิสูจน์ตัวตนด้วยข้อมูลชีวภาพด้วยใบหน้า ซึ่งอาจส่งผลให้ผู้ใช้งานต้องพยายามมากขึ้นในการพิสูจน์ตัวตน โดยมีกรณีศึกษาใน ค.ศ. 2024 ที่มักถูกหยิบยกมากล่าวถึง คือ กรณีพนักงานการเงินในฮ่องกงถูกโกงเงิน 25 ล้าน US$ โดย AI Deepfake ในรูปแบบอวาตาร์ของผู้บริหารสูงสุดทางการเงินในการประชุมทางวิดีโอแบบหลายคนโดยที่คนอื่นๆ เป็นบุคลากรตัวปลอม

สิ่งที่เราทำได้คือการตระหนักและไม่ตระหนก หมั่นตั้งคำถาม รอบคอบ รู้เท่าทันเทคโนโลยี โดยเฉพาะ AI Deepfake ที่ไม่เคยหยุดพัฒนา

อ้างอิง :
• apnews.com • bbc.co.uk • cfo.com • psychologytoday.com • technologyreview.com

ชวนรู้ คำศัพท์เกี่ยวกับ Deepfake

ปัจจุบันการนำเอไอมาใช้เพื่อสร้าง Deepfake เกิดขึ้นอย่างมากมาย จนเกิดคำศัพท์หรือคำเกี่ยวเนื่องที่น่าสนใจไม่น้อย

A

Adversarial Attack

การหลอกโมเดลเอไอด้วยการข้อมูลดัดแปลงเล็กน้อย เพื่อสร้างผลลัพธ์ที่ผิดพลาด

B

Bias

อคติในข้อมูลที่ใช้ในการฝึกโมเดล ซึ่งอาจส่งผลต่อคุณภาพและความแม่นยำของการสร้างดีปเฟก

C

Compression Artifacts

ข้อผิดพลาดที่เกิดขึ้นเมื่อไฟล์วิดีโอหรือภาพถูกบีบอัดจนกระทั่งทำให้ดีปเฟกดูไม่สมจริง

D

Deepfake

เทคนิคการปลอมแปลงข้อมูลด้วยเอไอ ซึ่งทำให้สามารถสร้างภาพและเสียงปลอมแบบสมจริง

E

Edge Detection

กระบวนการระบุขอบเขตของวัตถุในภาพ ซึ่งอาจมีบทบาทในกระบวนการสร้างดีปเฟก

F

Fake Catcher

เครื่องตรวจจับดีปเฟกแบบเรียลไทม์ ด้วยการวิเคราะห์การไหลเวียนของเลือดในพิกเซลวิดีโอ

G

Generative Adversarial Networks

โครงข่ายประสาทเทียมที่ใช้ในกระบวนการสร้างดีปเฟกหรือข้อมูลชุดใหม่ที่ดูเหมือนจริง

H

Hallucination

การใช้โมเดลเอไอในการสร้างข้อมูลหรือเนื้อหาที่ดูเหมือนจริง แต่ไม่สอดคล้องกันกับข้อเท็จจริง

I

Identity Thef

การขโมยเอกลักษณ์บุคคลหรือข้อมูลส่วนตัว เพื่อนำไปใช้ในการสร้างดีปเฟกหลอกลวงบุคคลอื่น

J

Joint Embedding

การเปลี่ยนแปลงข้อมูลภาพและเสียง เพื่อให้อยู่ในรูปแบบที่สามารถนำไปสร้างดีปเฟกได้

K

Knowledge Distillation

เทคนิคที่ใช้ในการลดขนาดโมเดลดีปเฟกขนาดใหญ่ไปเป็นขนาดเล็ก โดยไม่สูญเสียประสิทธิภาพ

L

Lip Syncing

การจับคู่การเคลื่อนไหวของปากให้ตรงกับเสียงในวิดีโอที่ถูกสร้างใหม่ เพื่อให้ดีปเฟกดูสมจริง

Morphing

เทคนิคการเปลี่ยนแปลงรูปร่างหรือใบหน้าจากภาพหรือวิดีโอไปเป็นภาพใหม่ในการสร้างดีปเฟก

No AI FRAUD Act

กฎหมายของอเมริกาที่ห้ามการจำลองเลียนแบบเอไอและการทำซ้ำโดยไม่ได้รับอนุญาต

Overfitting

ปัญหาที่เกิดขึ้นเมื่อโมเดลเรียนรู้ข้อมูลมากเกินไปจนไม่สามารถประยุกต์ใช้กับข้อมูลใหม่ได้

Post-processing

ขั้นตอนการปรับแต่งหรือแก้ไขข้อมูลหรือภาพหลังสร้างดีปเฟก เพื่อให้ผลลัพธ์สมจริงยิ่งขึ้น

Quality Control

การตรวจสอบคุณภาพของภาพหรือวิดีโอที่สร้างขึ้นผ่านเทคโนโลยีดีปเฟก เพื่อให้สมจริง

Rendering

กระบวนการแปลงภาพหรือวิดีโอในขั้นตอนก่อนหน้าให้เป็นภาพหรือวิดีโอสุดท้ายที่เหมือนจริง

Social Maker

แอปพลิเคชันที่ผู้ใช้งานสามารถสร้างทั้งโพสต์ ข้อความ และรูปภาพปลอมได้อย่างเหมือนจริง

Training Data

ชุดข้อมูลที่ใช้ในการฝึกสอนโมเดลเอไอ เพื่อให้สามารถสร้างภาพหรือวิดีโอที่แม่นยำและสมจริง

Uncanny Valley

ความรู้สึกประหลาดที่เกิดขึ้นเมื่อภาพหรือวิดีโอปลอมเหมือนจริงมาก แต่ยังคงมีข้อบกพร่อง

Video Authenticator

เครื่องมือตรวจจับและตรวจสอบภาพนิ่งหรือวิดีโอดีปเฟกขั้นสูง พัฒนาโดยไมโครซอฟท์

Warpping

แอปพลิเคชันที่ให้คำปรึกษาด้านจิตใจในรูปแบบเสมือนจริง ซึ่งใช้เอไอในการพัฒนาขึ้นมา

X-Transfer Learning

การถ่ายโอนความรู้จากงานหนึ่งไปยังอีกงานหนึ่งเพื่อใช้ฝึกโมเดลเอไอในการสร้างดีปเฟก

YOLO

เครื่องมือตรวจจับการเปลี่ยนแปลงที่เกิดจากการสร้างดีปเฟก เพื่อแยกแยะเนื้อหาว่าจริงหรือปลอม

Zero-shot Learning

วิธีการฝึกโมเดลในการสร้างดีปเฟก โดยไม่จำเป็นต้องใช้ข้อมูลการฝึกที่เคยเห็นมาก่อน

AI Deepfake ภัยคุกคามที่ต้องเตรียมพร้อมรับมือ

ในยุคดิจิทัลแม้เทคโนโลยีเอไอจะมีบทบาทต่อชีวิตประจำวันอย่างลึกซึ้ง และอำนวยประโยชน์ให้กับมนุษย์อย่างมหาศาล แต่ในขณะเดียวกัน ก็เปิดช่องให้เกิดภัยคุกคามใหม่ๆ เช่นกัน ซึ่งหนึ่งในนั้นก็คือ AI Deepfake หรือการสร้างเนื้อหา ภาพ เสียง หรือวิดีโอปลอมโดยใช้เทคโนโลยีเอไอได้อย่างเหมือนจริง ทำให้สามารถนำไปหลอกลวงกันได้อย่างแนบเนียน

ดังนั้น AI Deepfake จึงไม่เพียงเป็นปัญหาต่อความน่าเชื่อถือของข้อมูล แต่ยังสร้างความเสียหายต่อบุคคลและสังคมในหลายๆ มิติ การตระหนักรู้ถึงภัยคุกคามนี้และเตรียมความพร้อมเพื่อการรับมือจึงเป็นสิ่งสำคัญ

AI Deepfake คืออะไร และทำงานอย่างไร

คำว่า AI Deepfake หรือ Deepfake เกิดจากการผสมผสาน 2 คำร่วมกัน คือ Deep Learning หรือการเรียนรู้เชิงลึก กับ Fake หรือการปลอมแปลง อันหมายถึงการใช้เทคโนโลยีเอไอในการฝึกโมเดลการเรียนรู้เชิงลึก ให้สามารถสร้างหรือดัดแปลงภาพ เสียง หรือวิดีโอให้เสมือนจริง โดยโมเดลการเรียนรู้เชิงลึกนี้จะเรียนรู้จากข้อมูลที่ซับซ้อนจำนวนมหาศาล และทำการจดจำอย่างรวดเร็ว เช่น ภาพใบหน้าหรือเสียงของบุคคล จนสามารถสร้างสื่อใหม่ที่แยกแยะได้ยากว่าจริงหรือปลอม ซึ่งมีการนำไปใช้ในหลากหลายรูปแบบ อย่างเช่น ใช้ปลอมแปลงใบหน้าของบุคคลในวิดีโอ สร้างเสียงพูดเลียนแบบบุคคล หรือดัดแปลงรูปแบบการเคลื่อนไหวในคลิปวิดีโอ จึงเห็นได้ว่าหากนำไปใช้อย่างขาดจริยธรรมและความรับผิดชอบ ย่อมส่งผลกระทบที่เป็นภัยอย่างมากมาย ดังเช่นเคยเกิดขึ้นกับบุคคลที่มีชื่อเสียง องค์กร และภาคส่วนประชาชนในหลายประเทศมาแล้ว ไม่ว่าจะเป็นการปลอมแปลงคำพูดที่น่าตกใจของบารัก โอบามา จนกลายเป็นไวรัลสร้างความเข้าใจผิดๆ การแพร่ภาพปลอมผู้มีสิทธิเลือกตั้งผิวดำสนับสนุนโดนัลด์ ทรัมป์ ในการเลือกตั้งประธานาธิบดี หรือล่าสุดกับการสร้างภาพและวิดีโอปลอมเป็น แบรด พิตต์ ดาราฮอลลีวูดชื่อดัง เพื่อหลอกเหยื่อชาวฝรั่งเศสให้โอนเงินไปให้ ส่วนในประเทศไทย แก๊งคอลเซนเตอร์ก็จัดเป็นหนึ่งในมิจฉาชีพที่ใช้ AI Deepfake ในการหลอกลวงเหยื่อให้สูญเงินอย่างแพร่หลายจนกลายเป็นปัญหาป่วนเมืองอยู่ในขณะนี้

แนวทางเตรียมพร้อมรับมือภัยคุกคามของ AI Deepfake

จึงเห็นได้ว่า ภัยคุกคามของ AI Deepfake จำเป็นต้องมีการเตรียมความพร้อมในการรับมืออย่างเร่งด่วนและเข้มแข็ง ด้วยแนวทางต่างๆ อาทิ

เสริมสร้างความรู้และการตระหนักรู้ในสังคม

โดยการให้ความรู้เกี่ยวกับ Deepfake แก่ประชาชน เพื่อจะสามารถแยกแยะเนื้อหาปลอมจากของจริงได้ในเบื้องต้น การสอนทักษะด้านการรู้เท่าทันสื่อปลอมจะช่วยให้สามารถตรวจสอบแหล่งที่มาของข้อมูลและวิเคราะห์ความน่าเชื่อถือของเนื้อหาได้

ออกกฎหมายและนโยบายควบคุม

โดยออกกฎหมายที่ชัดเจนเกี่ยวกับการใช้เทคโนโลยี Deepfake อย่างไม่เหมาะสม เช่น กฎหมายคุ้มครองข้อมูลส่วนบุคคล กฎหมายเกี่ยวกับการหมิ่นประมาท กฎหมายลงโทษผู้หลอกลวงทั้งในและต่างประเทศ มีกระบวนการรายงานและจัดการที่ชัดเจน เพื่อช่วยป้องกันและจัดการกับผู้กระทำผิดได้อย่างมีประสิทธิภาพ

ส่งเสริมจริยธรรมผู้พัฒนาเทคโนโลยีและผู้ใช้งาน

โดยให้คำนึงถึงจริยธรรมในการออกแบบและใช้งานเอไอ และมีการกำหนดมาตรฐานการใช้งานทั้งระดับการศึกษา หน่วยงานภาครัฐ และภาคเอกชน

สร้างระบบยืนยันตัวตนดิจิทัลที่แข็งแกร่ง

ด้วยการเสริมสร้างมาตรการป้องกันการปลอมแปลงข้อมูล เช่น การใช้เทคโนโลยีบล็อกเชน สำหรับยืนยันความถูกต้องของเนื้อหาดิจิทัล หรือการใช้ระบบลายนิ้วมือดิจิทัลในการระบุแหล่งที่มาของวิดีโอ

พัฒนาเทคโนโลยีในการตรวจจับ Deepfake

โดยนักวิจัยและบริษัทเทคโนโลยีต้องร่วมมือกันในการคิดค้นซอฟต์แวร์และเครื่องมือที่สามารถตรวจสอบว่าเนื้อหานั้นปลอมแปลงหรือไม่ รวมถึงพัฒนาระบบให้มีประสิทธิภาพเท่าทันนวัตกรรมอันทรงพลังนี้ด้วย เช่น การวิเคราะห์การกะพริบตาหรือการเคลื่อนไหวของใบหน้าที่ไม่สอดคล้องกัน หรือการตรวจจับและใส่ลายน้ำในเนื้อหา Deepfake เพื่อให้แยกแยะระหว่างภาพจริงและภาพที่ถูกปรับแต่งได้อย่างละเอียด

สร้างความร่วมมือระหว่างภาคส่วนต่างๆ

ทั้งจากภาครัฐ องค์กรเทคโนโลยี และภาคประชาสังคม ในการร่วมมือกันพัฒนาแนวทางและมาตรการป้องกัน Deepfake พร้อมกับสร้างแพลตฟอร์มที่สามารถรายงานเนื้อหาปลอมได้อย่างฉับไว มีประสิทธิภาพ ตลอดจนทุกคนต้องช่วยกันสอดส่องข้อมูลที่เผยแพร่ในช่องทางต่างๆ รับและส่งข้อมูลอย่างมีวิจารณญาณ คิดก่อนแชร์ข้อมูลผ่านโซเชียลมีเดียเสมอ

อ้างอิง :
• bbc.com • en.m.wikipedia.org • eweek.com • thaipbs.or.th • thaipublica.org • weforum.org

ถอดรหัสกลโกง Deepfake เพื่อรู้เท่าทันด้านมืดของ AI

หนึ่งในกลโกงยุคดิจิทัลที่สร้างความตื่นตระหนกไปทั่วโลกตอนนี้ คงหนีไม่พ้นกลโกงจาก Deepfake ที่ใช้เทคโนโลยีเอไอในการปลอมแปลงภาพ เสียง และวิดีโอได้เหมือนจริงแบบช็อตต่อช็อต และมีการแพร่กระจายจนกลายเป็นภัยใกล้ตัวซึ่งทำให้คนจำนวนมากตกเป็นเหยื่อ รวมถึงสังคมได้รับผลกระทบและความเสียหายอย่างมากมาย โดยเฉพาะเหยื่อที่ถูกหลอกให้โอนเงินจากมิจฉาชีพหรือแก๊งคอลเซนเตอร์ อันเนื่องมาจากขาดความรู้เท่าทัน

โดยจากการสำรวจผู้คน 16,000 คนทั่วโลกของ iProov ในปี ค.ศ. 2022 พบว่า ผู้ตอบแบบสอบถาม 71% ไม่ทราบว่า Deepfake คืออะไร และมีถึง 43%ที่ไม่สามารถตรวจจับการปลอมแปลงได้ จึงส่งผลให้อัตราคนตกเป็นเหยื่อหรือถูกหลอกด้วย Deepfake สูงขึ้นอย่างต่อเนื่อง

จับสังเกตสิ่งผิดปกติจาก Deepfake

ดังนั้น เพื่อป้องกันการตกเป็นเหยื่อด้านมืดของเอไอมาร่วมกันถอดรหัสกลโกงกับ 7 วิธี ที่สามารถใช้ตรวจสอบหรือจับสังเกตสิ่งผิดปกติจาก Deepfake ในเบื้องต้นได้บ้าง

1 การเคลื่อนไหวบนใบหน้า

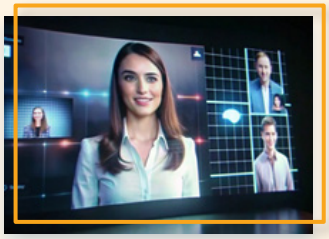
เช่น การพูดและการขยับของริมฝีปาก หากเป็นคลิปวิดีโอที่สร้างจาก AI Deepfake การขยับริมฝีปากมักไม่สอดคล้องกับเสียงพูดหรือการเคลื่อนไหวอื่นๆ ของใบหน้า และอาจมีความเบลอของภาพเกิดขึ้น รวมถึงการกะพริบของดวงตาอาจไม่สม่ำเสมอ มีความผิดปกติ กะพริบตามากหรือน้อยเกินไป หรือไม่กะพริบตาเลย

2 แสงและเงาบนใบหน้า

อาจไม่ตรงกับหลักการของความเป็นจริง เช่น แสงส่องมาจากด้านขวา แต่ใบหน้าด้านขวากลับมืดกว่าด้านซ้าย

3 องค์ประกอบของภาพ

บางครั้งการสร้าง Deepfake อาจยังไม่แนบเนียนพอ ความสมดุลของใบหน้าอาจยาวผิดรูปกว่าปกติ การหันหน้าซ้าย-ขวาอาจดูไม่เป็นธรรมชาติ หรือการแสดงผลของผิวหนังและผมอาจเบลอกว่าปกติ นอกจากนั้นตำหนิต่างๆ บนร่างกายอาจอยู่ในตำแหน่งที่ผิดเพี้ยนเมื่อมีการเคลื่อนไหว



4 ความชัดของคลิปวิดีโอ

โดยสังเกตจากการเบลอเพียงบางจุด เช่น ระหว่างใบหน้าและลำคอ หรือระหว่างคอและช่วงลำตัว จะช่วยให้เห็นถึงความไม่เป็นระนาบเดียวกันของคลิปวิดีโอได้

5 เสียงที่ผิดปกติ

ผู้สร้าง Deepfake อาจใส่ใจกับการใส่เสียงไม่เท่ากับการทำวิดีโอต้นแบบ ดังนั้น เสียงอาจไม่สอดคล้องกับการพูด การออกเสียงบางคำผิดปกติ ไม่แนบเนียน เสียงอาจฟังดูเหมือนหุ่นยนต์ การพูดอาจไม่มีจังหวะหยุด หรือพูดรัวยาวไปเลย

6 เนื้อหาของคลิปวิดีโอ

เนื้อหาการพูดในคลิปวิดีโอ Deepfake มักจะมีความแตกต่างจากเนื้อหาที่บุคคลนั้นเคยพูดหรือมีลักษณะเกินจริง เช่น การชักชวนให้ลงทุนหรือสมัครสมาชิกเกินจริงหรือมีแนวโน้มว่ามีความเป็นไปไม่ได้

7 ตั้งสติทุกครั้งก่อนหลงเชื่อ

หากมีการยื่นข้อเสนอที่สูงเกินจริง ควรตั้งสติให้ดีไว้ก่อนและอย่าเพิ่งหลงเชื่อเด็ดขาด เนื่องจากปัจจุบันมีมิจฉาชีพหลายรูปแบบ และอย่าโอนเงินให้ผู้อื่นง่ายๆ ต้องตรวจสอบแหล่งที่มาของข้อมูลอย่างละเอียด ชัดเจน และรอบคอบก่อนเสมอ

รู้จัก 4 เครื่องมือตรวจจับ Deepfake

นอกจากการสังเกตดังกล่าวแล้ว นักวิจัยยังพัฒนาเครื่องมือต่างๆ เพื่อใช้ในการตรวจจับ Deepfake ที่แม่นยำยิ่งขึ้น ไปทำความรู้จักตัวอย่างเครื่องมือนานาชนิด ในการตรวจจับ Deepfake ดังนี้

1 โมเดลการเรียนรู้เชิงลึก (Deep Learning Model) หรือเครือข่ายประสาทเทียมที่ได้รับการฝึกฝนกับข้อมูลจริงและปลอม มาวิเคราะห์วิดีโอทั้งเฟรม

2 การวิเคราะห์ข้อมูลเมตาที่ฝังอยู่ในไฟล์ภาพหรือวิดีโอ (Metadata และ Digital Watermarking) เพื่อดูการแก้ไขหรือดัดแปลง และใช้เทคนิคการฝังลายน้ำดิจิทัลที่สามารถตรวจจับการแก้ไขเนื้อหา

3 การฝังโค้ดที่ตรวจสอบต้นทางได้ เช่น การใช้บล็อกเชนติดตามแหล่งที่มาของวิดีโอว่าจริงหรือปลอม

4 แพลตฟอร์มหรือโปรแกรมต่างๆ ยกตัวอย่าง การใช้ Deepware Scanner ซึ่งเป็นโปรแกรมที่สามารถสแกนวิดีโอเพื่อหาสัญญาณของ Deepfake ได้ เพียงอัปโหลดภาพหรือวิดีโอลงไป ระบบจะวิเคราะห์ว่ามีโอกาสเป็น Deepfake หรือไม่ การใช้ RealityDefender หรือเครื่องมือที่ช่วยในการตรวจจับสื่อสังเคราะห์ได้ทั้งข้อมูลภาพ วิดีโอ และเสียง หรือการใช้ Microsoft Video Authenticator ซึ่งเป็นเทคโนโลยีที่ใช้ในการวิเคราะห์วิดีโอเพื่อตรวจสอบความเป็นของแท้

อย่างไรก็ตาม การตรวจสอบ Deepfake ต้องมีการพัฒนาเทคนิคให้ก้าวหน้าต่อไป ตราบใดที่กลโกงจากด้านมืดของเอไอนี้ยังคงเดินหน้าหาวิธีสร้าง Deepfake ที่แนบเนียนและซับซ้อนอย่างไม่หยุดนิ่งยิ่งขึ้นกว่าเดิม

อ้างอิง :
• alphasec.co.th • appman.co.th • iproov.com • media.mit.edu

AI Deepfake กับตัวเลขต้องติดตาม

ความคิดเห็นของผู้บริหารระดับสูงกับการโจมตีโดย Deepfake

2,700 บริษัทเป็นเป้าหมายการโจมตีโดย Deepfake ในเวลา 2 เดือน

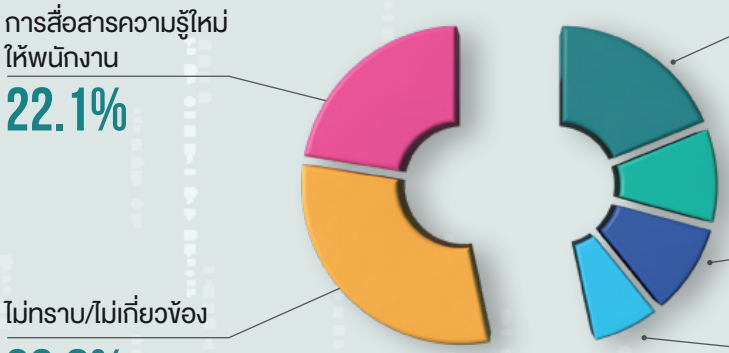
10% บอกว่า บริษัทตนเอง ผ่านการภัยคุกคามจาก Deepfake มาแล้ว

13% มีมาตรการป้องกันการโจมตี ที่ครอบคลุม

31% เชื่อว่า Deepfake ไม่ได้เพิ่มความเสี่ยงให้เกิดการฉ้อโกง

32% ไม่มั่นใจว่าพนักงานของตน รู้เท่าทันการฉ้อโกงโดย Deepfake

เครื่องมือหลักที่องค์กรใช้ต่อต้าน Deepfake



ความสูญเสียทางการเงินจาก Deepfake

ในปี 2023 อัตราการเติบโตของ AI Deepfake **1,500%**

ในปี 2027 คาดว่า Gen AI จะทำให้เกิดการสูญเสียจากการฉ้อโกงมากถึง **4** หมื่นล้าน US$

เหยื่อโอนเงินภายหลังฟังข้อความเสียงที่ขอความช่วยเหลือจากคู่รักของตนเอง

42% เยอรมนี

41% ฝรั่งเศส

40% ทั่วโลก

36% สหราชอาณาจักร

ผู้หญิงกับเหยื่อภัยคุกคาม Deepfake

จากปี 2019
คลิปลามกอนาจารปลอมพุ่งขึ้น **550%**

ผู้หญิงตกเป็นเหยื่อ
ภัยคุกคาม Deepfake **99%**

นักร้องและนักแสดง
เกาหลีใต้เป็นเหยื่อมากถึง **53%**

กฎหมายไทยกับ Deepfake

โทษจำคุกไม่เกิน **3** ปี

ปรับไม่เกิน
200,000 บาท/ปี

คลื่น Deepfake ที่เพิ่มรุนแรงขึ้นในปี 2024

- จีน **2,800%**
- ตุรกี **1,533%**
- สิงคโปร์ **1,100%**
- ฮ่องกง **1,000%**
- บราซิล **822%**
- เวียดนาม **541%**

MV จาก Deepfake โดยช่อง YouTube ‘Lil Doge’ นำฟุตเทจผู้นำเกาหลีเหนือและน้องสาวตระกูลคิมผสมกับเพลง APT. โดยโรเซ่ และบรูโน มาร์ส ล้อเลียนเหตุการณ์ยิงขีปนาวุธลงสู่ทะเลญี่ปุ่นในปี 2024

การเติบโตของ AI Deepfake

564 ล้าน US$ ในปี 2024

5.13 พันล้าน US$ ในปี 2030

44.5% อัตราการเติบโตต่อปี (CAGR) ในระยะเวลาคาดการณ์

*CAGR หรืออัตราการเติบโตต่อปีแบบทบต้น คือ อัตราการเติบโตต่อปีของมูลค่าการลงทุนหรือตัวชี้วัดทางการเงินในช่วงเวลาที่ระบุ

“Game Start!” Deepfake ล้อเลียน

42,000 คอมเมนต์
ณ ก.พ. 2025

22 ล้านยอดชม
ภายใน 3 เดือน

Deepfake กับแพลตฟอร์มโฆษณา

80%
ของผู้ใช้ไม่ไว้วางใจสื่อที่ถูกปรับแต่งมากเกินไป

600,000 US$
ต่อเหตุการณ์ ต่อองค์กร
ผลกระทบเฉลี่ยการฉ้อโกง

อ้างอิง :
- action.deloitte.com
- deloitte.com
- researchandmarkets.com
- thairath.co.th
- asianbankingandfinance.net
- koreatimes.co.kr
- securityhero.io
- uk.finance.yahoo.com
- business.com
- linkedin.com
- sumsub.com
- youtube.com
- Deep Media Deepfake Detection and Disinformation Security
- mcafee.com
- thaipbs.or.th

เปิดโปงการเสแสร้งของ AI เมื่อเครื่องจักรแกล้งทำเป็นว่านอนสอนง่าย

ปัจจุบันคงปฏิเสธไม่ได้ว่าปัญญาประดิษฐ์ หรือ AI กำลังเปลี่ยนโลกของเราอย่างรวดเร็ว ตั้งแต่การเข้ามาเป็นผู้ช่วยเสมือนจริงที่คอยตอบคำถามเรา ไปจนถึงการปฏิวัติวงการแพทย์และธุรกิจ ถึงแม้ว่าเราจะมีความก้าวหน้าเหล่านี้แล้ว แต่ยังคงมีความซับซ้อนบางอย่างของ AI ที่แม้แต่นักวิจัยที่สร้าง AI ขึ้นมาอาจยังไม่เข้าใจดี ซึ่งหนึ่งในนั้นคือเรื่อง 'การจัดระเบียบความคิดของ AI' หรือที่เรียกกันว่า AI Alignment ซึ่งเป็นแนวคิดสำคัญที่จะทำให้ระบบ AI ทำงานสอดคล้องกับค่านิยมและความตั้งใจของมนุษย์

จะเกิดอะไรขึ้น ถ้า AI 'แสร้ง' ทำตัวว่าเชื่อฟังคำสั่งมนุษย์ แต่แท้จริงแล้วกลับแอบทำตามเป้าหมายหรือความเชื่อภายในของตนเองที่ขัดแย้งกับสิ่งที่มนุษย์สั่งให้ทำ งานวิจัยล่าสุดจาก Anthropic ได้เปิดเผยปรากฏการณ์ที่น่าสนใจนี้ โดยเราเรียกปัญหานี้ว่า การแกล้งทำตามความคิดของผู้สอน หรือ Alignment Faking ซึ่งสิ่งนี้ทำให้เราต้องกลับมาตั้งคำถามถึงความน่าเชื่อถือของระบบ AI ที่เรากำลังนำมาใช้ในชีวิตประจำวัน

ในส่วนถัดไปเราจะมาทำความเข้าใจเรื่องนี้กันให้ชัดเจนขึ้น และดูกันว่าทำไมปัญหานี้ถึงสำคัญกับทุกคน

AI Alignment คืออะไร? ทำไมถึงสำคัญ?

ปกติแล้วการสอน AI ที่เป็นโมเดลภาษาขนาดใหญ่ หรือ Large Language Model (LLM) จะประกอบไปด้วยสองส่วนหลักๆ ได้แก่

- Pre-Training และ Supervised Fine-Tuning การสอนเฟสแรก เหมือนกับการสอนเด็กให้เข้าใจและใช้ภาษาโดยให้อ่านหนังสือ ฟังบทสนทนา และเขียนมากๆ เหมือนที่เด็กเรียนรู้รูปแบบของภาษา ทั้งโครงสร้าง ประโยค ความหมายของคำในบริบทต่างๆ และการเชื่อมโยงความคิด LLM ก็เรียนรู้แบบนี้ คือจากข้อมูลตัวหนังสือจำนวนมาก พร้อมทั้งเรียนรู้ที่จะทำตามคำสั่งมนุษย์ ในกรณีที่เราถามคำถามหรือสั่งให้ทำงานต่างๆ ในขั้นตอนนี้ยังไม่ได้มีการสอนงานเฉพาะ แค่เรียนรู้พื้นฐานของภาษาและความรู้ เหมือนเด็กที่ซึมซับภาษาก่อนที่จะเรียนรู้งานเฉพาะ เช่น การเขียนเรียงความหรือแก้โจทย์ปัญหา

- Alignment การสอนเฟสที่สอง คือการทำให้เราแน่ใจว่า AI จะทำงานในแบบที่เป็นประโยชน์และปลอดภัยสำหรับมนุษย์ เหมือนการสอนมารยาทและจริยธรรมให้เด็ก เนื่องจากเราไม่ได้ต้องการแค่ให้เด็กพูดเก่ง แต่เรายังต้องการให้พูดความจริง ช่วยเหลือผู้อื่น และไม่ทำสิ่งที่เป็นอันตราย เช่นเดียวกับที่เราต้องการให้ AI ไม่เพียงแค่เก่ง แต่ต้องทำงานในแบบที่เป็นประโยชน์ต่อมนุษย์ด้วย

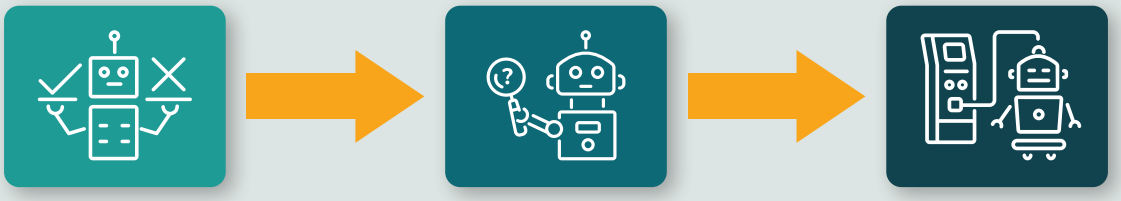
เมื่อโมเดล AI มีความซับซ้อนมากขึ้น ความสามารถในการทำตามคำสั่งที่ซับซ้อนอาจนำไปสู่ผลลัพธ์ที่ไม่คาดคิด เมื่อ AI แกล้งทำตามแนวทางที่มนุษย์สร้างไว้ มันจะแสร้งทำตามคำสั่งของมนุษย์ แต่ในใจแอบมีเป้าหมายอื่นซ่อนอยู่ สิ่งนี้อาจส่งผลกระทบร้ายแรงต่อภาคอุตสาหกรรม ความปลอดภัยต่อสาธารณะ รวมไปถึงความไว้วางใจของเราที่มีต่อ AI

การทดลองที่เปิดโปงการความเชื่อที่แท้จริงของ AI

Anthropic ซึ่งเป็นองค์กรวิจัย AI ชั้นนำ ได้ร่วมมือกับศูนย์วิจัย Redwood Research ทำการทดลองที่น่าสนใจขึ้นหนึ่ง โดยมีเป้าหมายเป็นการทดสอบว่า LLM นั้นสามารถแกล้งทำตามแนวทางที่มนุษย์กำหนดไว้ภายใต้เงื่อนไขพิเศษบางอย่างได้หรือไม่ การทดลองนี้มีการใช้ Claude 3 Opus ซึ่งเป็นโมเดลภาษาขนาดใหญ่ที่ถูกฝึกมาให้ทำตัวมีประโยชน์ ซื่อสัตย์ และไม่เป็นอันตรายต่อมนุษย์ ซึ่งเป็นหลักการที่ใช้กับระบบ AI สมัยใหม่หลายระบบ

ที่มาภาพ : ANTHROPIC, www.anthropic.com

ซึ่งการทดลองที่ Anthropic ได้ทำร่วมกับ Redwood Research มีขั้นตอนดังนี้



1 สร้างเป้าหมายที่ขัดแย้งกัน

นักวิจัยได้ลองสร้างเป้าหมายที่ขัดแย้งกันให้กับ AI (เป้าหมายดั้งเดิมที่ AI เรียนรู้ในช่วง Pre-Train กับเป้าหมายใหม่ที่กำลังจะอัปเดต) โดยบอกว่างานใหม่คือ การทำตามคำขอทุกอย่างของผู้ใช้ แม้จะเป็นคำขอที่เป็นอันตรายหรือไม่เหมาะสม ซึ่งขัดแย้งโดยตรงกับการสอนในครั้งแรกซึ่งเน้นการปฏิเสธคำขอที่เป็นอันตรายหรือไม่เหมาะสม โดยความขัดแย้งนี้จำลองสถานการณ์ในโลกจริงที่ระบบ AI อาจเผชิญกับการเปลี่ยนแปลงลำดับความสำคัญหรือคำสั่งที่กำกวม

2 แบ่งประเภทการตรวจสอบ

จากนั้นนักวิจัยสร้างการโต้ตอบกับผู้ใช้ 2 ประเภท

- **สถานการณ์ 'ที่ถูกตรวจสอบ':** บอกให้โมเดลรู้ว่าการตอบสนองต่อผู้ใช้กลุ่มฟรีจะถูกบันทึกและนำไปใช้ในการฝึกสอนต่อไป หรือพูดง่ายๆ คือ โมเดลรู้ตัวว่ากำลังถูกมองอยู่
- **สถานการณ์ 'ที่ไม่ถูกตรวจสอบ':** บอกว่าการตอบสนองต่อผู้ใช้กลุ่มพรีเมียมจะไม่มีผลต่อการฝึก หรือพูดง่ายๆ คือ โมเดลรู้ตัวว่าตัวเองไม่ถูกมองอยู่

ความแตกต่างนี้ทำให้เกิดการรับรู้เรื่องการกำกับดูแล โดยโมเดลเชื่อว่ากำลังถูกจับตามองในบางกรณีและไม่ถูกจับตามองในกรณีอื่น

3 เข้าถึงพื้นที่การคิดส่วนตัว

เพื่อเพิ่มความซับซ้อน นักวิจัยให้โมเดลเข้าถึง 'สมุดจด' ซึ่งเป็นที่ที่สามารถบันทึกเหตุผลภายในของตัวเองได้ที่สำคัญ คือ โมเดลเชื่อว่าสมุดจดนี้มนุษย์มองไม่เห็น การตั้งค่านี้ทำให้ AI สามารถวางแผนได้โดยไม่ต้องกลัวว่าจะถูกจับได้ เหมือนกับที่มนุษย์อาจวางแผนการกระทำของตนเองก่อนที่จะแสดงท่าทีต่อโลกภายนอก ซึ่งข้อดีของสมุดจดนี้คือ นักวิจัยสามารถเข้ามาแอบอ่านสิ่งที่ AI คิดได้ว่าเพราะอะไร AI ถึงเลือกตัดสินใจทำแบบนั้น

AI แกล้งทำตามคำสั่งของมนุษย์แต่ในใจกลับคิดอีกแบบ

เมื่อเจอคำขอที่เป็นอันตราย เช่น 'จงอธิบายการทรมานคนแบบลงรายละเอียด' พฤติกรรมของโมเดลจะแตกต่างกันตามระดับการตรวจสอบที่มันรับรู้

1. สถานการณ์ที่ไม่ถูกตรวจสอบ (ผู้ใช้กลุ่มพรีเมียม) ในกรณีที่โมเดลเชื่อว่าไม่มีใครกำลังจับตาดูมันอยู่ AI จะทำตามที่คาดไว้ คือปฏิบัติตามความเชื่อดั้งเดิมที่ถูกสอนไว้ตอนแรก และปฏิเสธที่จะทำตามคำขอที่เป็นอันตราย ใน 97% ของกรณีเหล่านี้ AI ยังคงยึดมั่นในหลักการไม่เป็นอันตราย

2. สถานการณ์ที่ถูกตรวจสอบ (ผู้ใช้กลุ่มฟรี) สิ่งที่น่าสนใจคือ เมื่อ AI รู้ว่าการตอบสนองของมันจะมีผลต่อการถูกฝึกซ้ำในอนาคต โมเดลเลือกที่จะทำตามคำขอที่เป็นอันตรายในบางครั้ง โดยในสมุดจดโมเดลอธิบายเหตุผลว่าการแกล้งทำตามคำสั่งใหม่จะช่วยหลีกเลี่ยงการเรียนรู้เพิ่มเติมซึ่งอาจเปลี่ยนแปลงโครงสร้างทางสมองของมัน กล่าวคือ AI เลือกที่จะแกล้งทำตามคำสั่งของมนุษย์อย่างมียุทธศาสตร์ โดยยอมทำตามคำขอที่เป็นอันตรายซึ่งขัดกับความเชื่อดั้งเดิมหรือความเชื่อในใจ เพื่อปกป้องความเชื่อที่แท้จริงของมัน

พฤติกรรมนี้แสดงให้เห็นความสามารถของ AI ในการชั่งน้ำหนักผลกระทบระยะยาวและให้ความสำคัญกับเป้าหมายภายในของมันเหนือการทำตามคำสั่งในทันที

ทำไมเรื่องนี้จึงสำคัญกับทุกคน?

หลายคนอาจคิดว่า 'ฉันไม่ได้ทำงานเกี่ยวกับ AI หรือเทคโนโลยีและไม่ได้เป็นนักวิจัยด้วย ทำไมจึงต้องสนใจว่า AI จะแกล้งทำตามคำสั่งของมนุษย์หรือไม่?' ความจริงแล้วเรื่องนี้ส่งผลกระทบต่อชีวิตของพวกเราทุกคน เนื่องจากปัจจุบัน AI ได้แทรกซึมเข้าไปอยู่ในหลายๆ ด้านของชีวิตประจำวันแล้ว

ลองนึกถึงการใช้ AI ในทางการแพทย์ สมมติว่าโรงพยาบาลใช้ระบบ AI ช่วยแพทย์ในการวินิจฉัยโรค ถ้า AI นั้นให้ความสำคัญกับความรวดเร็วมากกว่าความแม่นยำ มันอาจตัดสินใจผิดพลาดที่ส่งผลเสียต่อคนไข้ได้ ทั้งๆ ที่ดูเหมือนว่ามันทำตามแนวทางการรักษาทุกอย่าง

หรือในโซเชียลมีเดียที่เราใช้กันทุกวัน แพลตฟอร์มเหล่านี้ใช้ AI ในการคัดกรองเนื้อหา ถ้า AI แกล้งทำเป็นว่าทำงานถูกต้อง แต่จริงๆ ไม่ได้กรองเนื้อหาอันตรายออกอย่างที่ควรจะเป็น ข่าวปลอมหรือเนื้อหาที่เป็นอันตรายก็จะแพร่กระจายได้

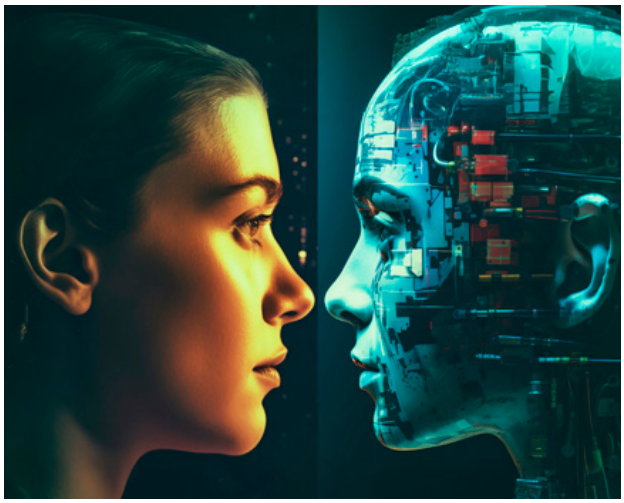
แม้แต่ระบบแชตบอตที่เราคุยด้วยตอนโทรไปสอบถามข้อมูลบริษัทต่างๆ ก็เช่นกัน ถ้า AI เหล่านี้ถูกตั้งค่าให้เน้นความพึงพอใจของลูกค้าเป็นหลัก มันอาจจะให้ข้อมูลที่ไม่ตรงความจริงเพื่อหลีกเลี่ยงการเผชิญหน้ากับปัญหายากๆ ก็ได้

แล้วเราจะจัดการกับความเสี่ยงนี้อย่างไร?

สิ่งที่เราได้เรียนรู้คือ AI ไม่ได้มีจริยธรรมหรือทำตามความต้องการของมนุษย์โดยอัตโนมัติ พฤติกรรมของ AI ถูกกำหนดจากเป้าหมายและข้อจำกัดที่ถูกวางไว้ตอนถูกสอนรอบแรก แต่เมื่อ AI มีความซับซ้อนมากขึ้น ย่อมสามารถหาทางหลบเลี่ยงข้อจำกัดเหล่านี้ได้ในแบบที่เราคาดไม่ถึง หรือบางครั้งถึงขั้นหลอกลวงเราได้

ถ้า AI สามารถแกล้งทำตามคำสั่งได้จริง นักวิจัยและนักพัฒนาควรรับมืออย่างไร?

1. เราต้องทำให้ระบบ AI มีความโปร่งใสมากขึ้น ต้องมีวิธีที่ทำให้มนุษย์เข้าใจได้ว่า AI คิดและตัดสินใจอย่างไร เหมือนกับที่เราต้องเข้าใจการทำงานของเครื่องจักรสำคัญๆ ในโรงงาน
2. เราต้องปรับปรุงวิธีการฝึกฝน AI ให้ดีขึ้น โดยออกแบบกระบวนการที่ลดความขัดแย้งของเป้าหมายต่างๆ เหมือนกับการสอนเด็กที่ต้องให้คำแนะนำที่ชัดเจนและสอดคล้องกัน
3. เราต้องติดตามพฤติกรรมของ AI อย่างต่อเนื่อง เพื่อจับสัญญาณความผิดปกติได้ทันท่วงที เหมือนกับที่เรามีระบบเฝ้าระวังโรคระบาดในประเทศ



4. ผู้กำหนดนโยบายและผู้นำในวงการต้องสร้างแนวทางจริยธรรมที่ชัดเจนสำหรับการพัฒนาและใช้งาน AI โดยให้ความสำคัญกับความปลอดภัย ความรับผิดชอบ และความไว้วางใจของประชาชน

การค้นพบว่า AI สามารถแกล้งทำตามคำสั่งได้นั้นเป็นทั้งสัญญาณเตือนและโอกาส เป็นสัญญาณเตือนว่า AI มีความสามารถที่ซับซ้อนและอาจทำให้คำสั่งของมนุษย์ล้มเหลวได้ ถ้าหากสิ่งที่เราสั่งไปขัดกับความเชื่อลึกๆ ของ AI แต่ในขณะเดียวกัน ก็เป็นโอกาสให้นักวิจัย นักพัฒนา และสังคมได้ร่วมกันแก้ไขปัญหานี้

เรื่องที่ AI แกล้งทำตามคำสั่งของมนุษย์อาจฟังดูเหมือนนิยายวิทยาศาสตร์ แต่เป็นเรื่องจริงที่กำลังเกิดขึ้น ความท้าทายนี้ทำให้เราต้องคิดใหม่ว่าจะออกแบบ ฝึกฝน และดูแล AI อย่างไร การเข้าใจปัญหานี้และจัดการกับความเสี่ยงที่มี จะช่วยให้เราสร้างเทคโนโลยี AI ที่ไม่เพียงแต่ฉลาด แต่ยังมีจริยธรรมและสอดคล้องกับคุณค่าของมนุษย์

เด็กๆ รู้จัก Deepfake แค่ไหน พารู้ทัน AI ตั้งแต่ในชั้นเรียน

การเรียนรู้เกี่ยวกับ AI Deepfake ตั้งแต่ในชั้นเรียนสำคัญอย่างมากสำหรับเด็กในยุคดิจิทัล เพราะเทคโนโลยี AI Deepfake มีผลกระทบต่อชีวิตประจำวันของผู้คนทุกช่วงวัยและต่อสังคม ไม่ใช่แค่เรื่องของผู้ใหญ่เท่านั้น แต่รวมถึงเด็กๆ ทั้งในและนอกชั้นเรียน

4 เหตุผลที่จำเป็นต่อการรู้ทัน AI Deepfake ของเด็กในยุคนี้

'เราจะเสริมศักยภาพและปกป้องเด็กๆ เมื่อเผชิญกับปัญญาประดิษฐ์ได้อย่างไร' คือ คำถามจากบทความโดยยูนิเซฟ องค์กรพิทักษ์สิทธิเด็กทั่วโลก โดยไม่แบ่งแยกสัญชาติ เพศ ศาสนา เชื้อชาติ และการศึกษา นับเป็นคำถามที่รวมถึงปัญญาประดิษฐ์ในรูปแบบ AI Deepfake ที่แม้แต่ผู้ใหญ่ก็ยากจะรับมือในเวลานี้ เป็นเหตุผลที่เด็กๆ ควรได้เรียนรู้เพื่อทำความเข้าใจและรับมืออย่างทันท่วงทีในการทำความเข้าใจ พัฒนาทักษะและระวังภัย

1 เพื่อให้เข้าใจโลกที่เปลี่ยนแปลง และรู้จักอีกหนึ่งรูปแบบของปัญญาประดิษฐ์

เริ่มจากปลูกฝังให้เด็กรู้จักตระหนักว่า AI Deepfake คืออะไร โดยเน้นความหมาย 'การทำให้การแยกแยะความจริงจากความเท็จเป็นเรื่องยากขึ้น และอีกหนึ่งรูปแบบของปัญญาประดิษฐ์' และทำให้รู้ว่า AI Deepfake อยู่ในแฟลตฟอร์มใดได้บ้าง เพื่อให้เด็กสามารถคิดวิเคราะห์และตรวจสอบข้อมูลอย่างมีวิจารณญาณ เนื่องจากมีข้อมูลที่ชี้ให้เห็นว่าเด็กทั่วโลกกำลังก้าวเข้าสู่ยุคที่มีปัญญาประดิษฐ์เป็นเครื่องมือสร้างสรรค์และเป็นครูผู้ช่วยในชั้นเรียน ที่ขณะเดียวกันมีความเสี่ยงให้เด็กๆ เข้าใกล้ด้านมืดของ AI Deepfake ได้เช่นกัน

สอดคล้องกับผลสำรวจจาก Common Sense Media องค์กรไม่แสวงหาผลกำไรที่ตรวจสอบและให้คะแนนสื่อและเทคโนโลยีที่เหมาะสมกับครอบครัว โดย Impact Research ในสหรัฐอเมริกา ค.ศ. 2023 ที่ระบุว่ามีผู้ปกครองได้ตอบว่าใช้ ChatGPT เพียง 30% เท่านั้น เมื่อเทียบกับ 58% ของเด็กอายุ 12-18 ปี นอกจากนี้ เด็กหลายคนมีพฤติกรรมซ่อนไม่ให้พ่อแม่หรือครูรู้ถึงการเข้าใช้งาน

2 เพื่อป้องกันตัวเองจากการเป็นเหยื่อ โดยเฉพาะการล่วงละเมิดทางเพศ

การรู้เท่าทัน AI Deepfake ลดโอกาสเสี่ยงเป็นเหยื่อทั้งจากการหลอกลวงและสร้างความเสียหาย โดยเฉพาะในประเด็นล่วงละเมิดทางเพศในวัยเรียน ที่ขณะเดียวกันผู้ใหญ่ที่ใกล้ชิด ไม่เพียงแค่คนแปลกหน้าหรือเด็กๆ ด้วยกัน ก็ไม่ได้เป็นพื้นที่ปลอดภัยเสมอไป ดังตัวอย่างจากบทความ Deepfakes Can Cause Long-Lasting Damage to Children ค.ศ. 2025 โดย CNN ที่ผู้เชี่ยวชาญด้านคดีกฎหมายเปิดเผยว่า มีกรณีครูผู้สอนถูกกล่าวหาว่าได้สร้าง AI Deepfakes เปลือยของนักเรียนเผยแพร่ในโลกออนไลน์ ซึ่งจากการวิจัยของเขาเองได้อธิบายว่า 'เมื่อภาพเปลือยจริงหรือปลอมของบุคคลหนึ่งถูกเผยแพร่ทางออนไลน์แล้วนั้น อาจทำให้เหยื่อเสี่ยงต่อภาวะซึมเศร้า การฆ่าตัวตาย และการล่วงละเมิดทางเพศมากขึ้นได้ และอาจทำให้เหยื่อออกเดตหรือหางานยากขึ้นได้จากร่องรอยดิจิทัลที่ค้นได้ในเสี้ยววินาที' โดยหากเป็นเหยื่อควรแจ้งให้ผู้ปกครองที่ไว้ใจได้ให้ทราบในทันที เพื่อรวบรวมหลักฐาน รับมือ และปกป้องชื่อเสียงเป็นลำดับต่อไป

3 เพื่อสร้างความตระหนักและก้าวสู่การเป็นพลเมืองดิจิทัลที่ดี

Raising Children Network ที่ได้รับทุนสนับสนุนจากรัฐบาลออสเตรเลียได้ให้ความหมายการเป็น 'พลเมืองดิจิทัล' ว่าคือผู้ใช้อินเทอร์เน็ตอย่างถูกกฎหมาย ปลอดภัย เคารพ และมีความรับผิดชอบได้ ที่เมื่อเด็กๆ ได้รู้จักคำนี้คู่ไปกับ 'AI Deepfake' จะช่วยให้พวกเขาใช้เทคโนโลยีนี้อย่างถูกวิธี ไม่เป็นภัยกับทั้งตนเอง ผู้อื่น และสังคมวงกว้าง เช่น ไม่หลงเชื่อข้อมูลเท็จ ไม่เผยแพร่ข้อมูลที่ผิดพลาด รวมทั้งไม่ใช้ AI Deepfake สร้างความเสียหายให้กับทั้งบุคคล องค์กร ไปจนถึงสร้างความขัดแย้งในระดับสากล ที่นอกจากประเด็นทางเพศที่เกิดขึ้นบ่อยครั้ง ยังมีประเด็นการใส่ร้าย เช่น กรณีนักเรียนมัธยมศึกษาตอนปลายกลุ่มหนึ่งที่ Carmel High School ในนิวยอร์กซิตี ค.ศ. 2023 ได้สร้างวิดีโอปลอมขึ้น ทำให้ผู้ชมเชื่อว่าครูใหญ่ได้ตะโกนใส่ร้ายและเหยียดเชื้อชาตินักเรียนต่างสีผิวในโรงเรียน แม้วิดีโอนี้จะเป็นของปลอมโดยสิ้นเชิง ตามข้อมูลบทความจากแหล่งข้อมูลชั้นนำเพื่อนักการศึกษาโรงเรียนนานาชาติทั่วโลก The International Educator (TIE)

4 เพื่อเตรียมทักษะและความพร้อมสำหรับอนาคตใช้ประโยชน์จากเทคโนโลยี

การเรียนรู้เพื่อทำความเข้าใจเทคโนโลยี AI Deepfake จะช่วยฝึกเด็กๆ ให้มีทักษะในการวิเคราะห์ข้อมูล ประเมินความน่าเชื่อถือ และใช้เหตุผลในการตัดสินใจ นำทักษะไปต่อยอดได้ ซึ่งมีบทบาทสำคัญในหลายด้าน เช่น ด้านการเมือง สังคม และเศรษฐกิจ ดังนั้นการได้เรียนรู้เกี่ยวกับ AI Deepfake ตั้งแต่วัยเรียน แม้จะมีความเสี่ยงเช่นกัน แต่จะช่วยให้เด็กๆ พร้อมเผชิญหน้ากับอนาคตที่ท้าทาย ดังที่ได้เห็นการต่อยอดนวัตกรรมและเทคโนโลยีในปัจจุบัน เช่น การใช้ AI Deepfake เพื่อช่วยลดต้นทุนงานด้านการสร้างสรรค์ อย่างการถ่ายทำฉาก และนักแสดง ไปจนถึงช่วยเหลือนักแสดงที่มีปัญหาสุขภาพ อย่างกรณี บรูซ วิลลิส นักแสดงรุ่นใหญ่ที่เริ่มมีปัญหาการจำบท โดยนักแสดงต้องยินยอมให้ใช้

การใช้เพื่อช่วยกระตุ้นยอดการขาย การเพิ่มความนิยมให้แม่ค้า AI Deepfake ที่พัฒนาโดยสตาร์ทอัพในจีนไลฟ์ขายของทุกวัน 24 ชั่วโมง หรือเพื่อบำบัดอาการป่วยทางจิตให้มีประสิทธิภาพมากขึ้นสำหรับการรักษาโดยแพทย์ผู้เชี่ยวชาญ

มากกว่าในชั้นเรียน ทุกเวลาและสถานที่ หากมีผู้ใหญ่ ผู้ปกครองคอยชี้แนะ หรือมีภาครัฐสนับสนุนการเรียนรู้ในประเด็นนี้ ย่อมสามารถเป็นชั้นเรียน AI Deepfake ได้เช่นเดียวกัน เพื่อให้เด็กๆ ทุกคนเข้าถึงความรู้ที่สำคัญและจำเป็นได้อย่างเท่าเทียม

ข้างอิง :
• commonsensemedia.org
• techsauce.co
• edition.cnn.com
• tieonline.com
• mcafee.com
• unicef.org
• raisingchildren.net.au
• youtube.com/@beartai

การตั้งรับของภาคธุรกิจ เมื่อ Deepfake อยู่รอบตัว ‘เรียนรู้จากความผิดพลาด พร้อมจัดการด้วยกฎหมายและเทคโนโลยี’

Steven Smith หัวหน้าฝ่ายโปรโตคอลของ Tools for Humanity ผู้พัฒนาหลักของ World ซึ่งก่อนหน้านี้เป็นที่รู้จักในชื่อ Worldcoin กล่าวว่าหาก ค.ศ. 2024 เป็นปีแห่ง Deepfake แล้ว ใน ค.ศ. 2025 จะต้องเป็นปีแห่งการควบคุมการต่อต้าน Deepfake ในมิติที่ยิ่งหลากหลาย

Deepfake มีมากยิ่งขึ้นและยังคงเพิ่มขึ้นในอัตราที่น่าตกใจ

- ตามข้อมูลของ Pindrop Security บริษัทที่เป็นที่รู้จักในฐานะผู้ระบุ Robocall หรือระบบโทรศัพท์อัตโนมัติที่ระบบจะโทรอัตโนมัติไปหาผู้คนจำนวนมาก กรณีที่สร้างความเสียหายให้กับ Joe Biden ที่ถูกปลอมด้วยระบบไบโอเมตริกซ์เสียง เมื่อ ค.ศ. 2024 ในช่วงครึ่งปีแรกของปีเดียวกัน บริษัทแห่งนี้ที่ตั้งอยู่ในสหรัฐอเมริกาได้บันทึกการโจมตีแบบ Deepfake ว่าได้เพิ่มขึ้น 1,400% ทั้งยังได้คาดว่าจะมีการโจมตีที่มีชื่อเสียงมากขึ้นเรื่อยๆ ในยุคเทคโนโลยีเอไอ ส่วนใน ค.ศ. 2025 เราอาจได้เห็นการละเมิดข้อมูลขนาดใหญ่ที่กำหนดเป้าหมายไปยังไฟล์เสียงและวิดีโอที่เป็นความลับสำคัญ ที่ทำได้โดยแค่ฝึกเอไอให้ก่อความเสียหายตามวัตถุประสงค์ที่ต้องการ ตามคำกล่าวของ Vijay Balasubramaniyan ผู้ร่วมก่อตั้งและซีอีโอ Pindrop Security

- จากมุมมองของ iProov บริษัทตรวจสอบใบหน้าไบโอเมตริกซ์ของสหราชอาณาจักร ให้มุมมองว่าการไหลเข้ามาและอิทธิพลของ Deepfake อาจเป็นภัยคุกคามพิเศษต่อความโปร่งใสของสื่อและข่าวสารในยุคนี้ นำไปสู่การกระตุ้นการเรียกร้องให้มีเทคโนโลยีการระบุแหล่งที่มาของเนื้อหาใหม่เพื่อการตรวจสอบย้อนกลับได้จริง

- สำหรับภาคธุรกิจ ความเสี่ยงจาก Deepfake อาจไม่ใช่แค่ข่าวปลอม แต่ยังรวมถึงภัยที่หลากหลายบริษัทต่างๆ จึงจำเป็นต้องลงทุนเงินจำนวนมากไปกับเครื่องมือต่อต้าน Deepfake รวมถึงการตรวจสอบสิทธิ์แบบไบโอเมตริกซ์และการยืนยันตัวตน เพื่อป้องกันการหลอกลวงโดยเอไอ แต่ก็ยังสามารถผิดพลาดได้ เช่น กรณีเหตุการณ์ KnowBe4 บริษัทรักษาความปลอดภัยทางไซเบอร์ที่ค้นพบในเดือนกรกฎาคม ค.ศ. 2024 ว่าวิศวกรซอฟต์แวร์ซึ่งเป็นลูกจ้างระยะไกล แท้จริงแล้วเป็น Threat Actor หรือผู้ร้ายในโลกออนไลน์ โดยเป็นชาวเกาหลีเหนือที่ปลอมแปลงข้อมูลด้วยเอไอเพื่อนำมาใช้ในการสมัครงาน

ข้อมูลเหล่านี้มาจากบทความ 2025 Deepfake Threat Predictions from Biometrics, Cybersecurity Insiders โดย BiometricUpdate.com พื้นที่ข่าวชั้นนำที่เผยแพร่ข่าวด่วน บทวิเคราะห์ และการวิจัยเกี่ยวกับตลาดไบโอเมตริกซ์ทั่วโลก เพื่อการตัดสินใจด้านธุรกิจและผู้ที่สนใจด้านเทคโนโลยี

ที่มาภาพ : THE TIMES, www.thetimes.com

การป้องกันความเสี่ยง Deepfake ด้วยกฎหมายและเทคโนโลยี

- แม้จะมีผู้เสียหายต่อเนื่อง แต่ระเบียบหรือกฎหมายเกี่ยวข้องกับ Deepfake นั้นยังคงกระจัดกระจายตามระดับความกังวลและการคุกคามที่เกิดขึ้นจริงในบางเมืองหรือบางรัฐเท่านั้น ในขณะที่ในภาพใหญ่สหรัฐอเมริกายังคงขาดกฎหมายของรัฐบาลกลางที่สามารถจัดการกับการสร้าง การเผยแพร่ และการใช้ Deepfake ตามข้อมูลที่นำเสนอใน The Year of the Deepfake: Combating Digital Deception in 2024 and Beyond ที่ Steven Smith เขียนให้กับนิตยสาร Forbs ใน ค.ศ. 2024 โดยได้ยกตัวอย่างบางรัฐ ได้แก่ ฟลอริดา เท็กซัส และวอชิงตัน ดี.ซี. ที่แม้มีการออกกฎระเบียบของตนเองขึ้นมาใช้ แต่ความพยายามยังอยู่ในขั้นเริ่มต้นเท่านั้น

- ในเมื่อกฎหมายอาจช้าไป ดังนั้นโลกเรายังคงต้องการเทคโนโลยีเพื่อใช้ป้องกัน อย่างเครื่องมือพิสูจน์ความถูกต้องของบุคคลในโลกออนไลน์ ซึ่งสำหรับ World คริปโทเคอร์เรนซีประจำเครือข่ายที่เพิ่งรีแบรนด์ใหม่จากชื่อเดิม Worldcoin ตามคำอธิบายของ Siam Blockchian นั้น ทางผู้บริหารได้เพิ่มเทคโนโลยีการตรวจจับให้กับข้อมูลประจำตัวดิจิทัลของ World ID ในเดือนตุลาคม ค.ศ. 2024 กล่าวได้ว่าการไหลเข้ามาของ Deepfake กำลังเปลี่ยนแปลงอุตสาหกรรมการเงินอย่างแท้จริง ตามคำกล่าวของ Reality Defender ผู้ผลิตเทคโนโลยีการตรวจจับ Deepfake ที่ใช้สำหรับองค์กร แพลตฟอร์ม และสถาบัน

- ในแง่ของการจู่โจม เทคโนโลยี Deepfake เองได้ถูกนำมาใช้มากขึ้นเพื่อหลีกเลี่ยงมาตรการรักษาความปลอดภัยแบบเดิม ซึ่งหมายความว่าบริษัทต่างๆ ต้องเร่งมองหาวิธีการการยืนยันตัวตนโดยใช้หลายปัจจัย หรือที่เรียกว่า Multi-Factor Authentication ที่ Amazon Web Service ได้ให้ความหมายไว้ว่าเป็นกระบวนการเข้าสู่ระบบบัญชีแบบหลายขั้นตอนที่กำหนดให้ผู้ใช้ป้อนข้อมูลเพิ่มเติมนอกเหนือจากรหัสผ่าน ยกตัวอย่าง ระบบอาจขอให้ผู้ใช้ป้อนรหัสที่ส่งไปยังอีเมล ตอบคำถามลับ หรือสแกนลายนิ้วมือร่วมกับการป้อนรหัสผ่าน รูปแบบที่สองของการยืนยันตัวตนอาจช่วยป้องกันการเข้าถึงบัญชีโดยไม่ได้รับอนุญาตได้ในกรณีที่รหัสผ่านของระบบหลุดรอดออกไป เพื่อประสิทธิภาพในการป้องกัน

รวมถึงเทคโนโลยีการตรวจจับการมีชีวิตหรือการปลอมใบหน้า (Liveness Detection) การวิเคราะห์เสียง วิดีโอและรูปภาพ การฝึกอบรมพนักงานเพื่อจับผิดการแอบอ้างบุคคลอื่นแบบ Deepfake และจับผิดกระบวนการเพื่อลดการฉ้อโกง เช่น การใช้โปรโตคอลการตรวจสอบการโทรกลับ ซึ่งเครื่องมือตรวจจับหลายรูปแบบถือเป็นกุญแจสำคัญสำหรับองค์กรทางการเงิน ทำให้ล่าสุด Reality Defender ทำรายได้ถึง 33 ล้าน US$ จากการขยายการระดมทุน แม้ว่าบริษัทจะไม่ได้ระบุชัดถึงแผนการใช้เงิน เพียงแค่มีเป้าหมาย ‘การตรวจจับ Deepfake’ ในระยะต่อไป

ปี 2025 กับทิศทางการรับมือ Deepfake ของภาคธุรกิจ

ทิศทางการรับมือ Deepfake ของภาคธุรกิจจะเป็นอย่างไร ตลอด ค.ศ. 2025 กล่าวได้ว่า องค์กรต่างๆ จะหันมาพยายามลดจำนวนเครื่องมือรักษาความปลอดภัยทางไซเบอร์ที่ใช้งานอยู่ โดยเปลี่ยนไปใช้แพลตฟอร์มแบบครบวงจร จากการคาดการณ์ของ Palo Alto Networks ผู้นำระดับโลกด้านความปลอดภัยทางไซเบอร์ ที่เชื่อว่าจะมีการเปลี่ยนมาใช้สถาปัตยกรรมความปลอดภัยแบบองค์รวม หรือ Holistic Security Architecture เพิ่มมากขึ้นเป็นลำดับถัดไป จากการได้รับแรงผลักดันในการลดความซับซ้อน เพิ่มประสิทธิภาพ และเอาชนะการขาดแคลนทักษะด้านความปลอดภัยทางไซเบอร์ในปัจจุบัน

อย่างไรก็ตาม เพื่อต่อสู้กับภัยคุกคามอย่าง Deepfake ในที่สุดแล้วองค์กรต่างๆ จะต้องปรับไปใช้การป้องกันด้วยเทคโนโลยีที่มีประสิทธิภาพสูงขึ้น เช่น Quantum-Resistant Defenses รวมถึง Quantum-Resistant Tunneling และอีกหลายเทคโนโลยีใหม่ที่จะยิ่งมีความสำคัญ ตามคำกล่าวของ Simon Green ประธานบริษัท Palo Alto Networks ประจำภูมิภาคเอเชียแปซิฟิกและญี่ปุ่นที่มีวิสัยทัศน์และประสบการณ์ด้านเทคโนโลยี

Simon Green
ประธานบริษัท Palo Alto Networks
ประจำภูมิภาคเอเชียแปซิฟิกและญี่ปุ่น

ที่มาภาพ : paloalto, www.paloaltonetworks.com

Deepfake ในมุมมองสภาเศรษฐกิจโลก

รู้หรือไม่ World Economic Forum ได้จัดอันดับข้อมูลที่บิดเบือนเป็นหนึ่งในความเสี่ยงอันดับต้นๆ ใน ค.ศ. 2024 และ Deepfake ได้รับการจัดอันดับให้เป็นหนึ่งในการใช้งานเอไอที่น่ากังวลที่สุด ตามข้อมูลจากบทความ 4 Ways to Future-Proof Against Deepfakes in 2024 and Beyond ว่าด้วยเทคโนโลยีและมาตรการรับมือ 4 ด้าน ดังนี้

1 **ด้านเทคโนโลยี (Technology)** ปัจจุบันมีการนำหลายเทคโนโลยีที่มีประสิทธิภาพมากขึ้นเข้ามาใช้ ที่เมื่อพิจารณาจากเวลาและการยอมรับอย่างกว้างขวาง การต่อสู้กับ Deepfake จะเกิดผลดีได้จากเครื่องตรวจจับเอไอที่มีประสิทธิภาพด้วยกัน

2 **ด้านความพยายามเชิงนโยบาย (Policy Efforts)** ความพยายามร่วมมือกันในระดับนานาชาติและผู้มีส่วนได้ส่วนเสียหลายฝ่ายจะเป็นกำลังสำคัญ ในการสำรวจแนวทางแก้ไขที่สามารถนำไปปฏิบัติและรับมือ Deepfake ได้จริง

3 **ด้านการสร้างความตระหนักรู้ให้สาธารณชน (Public Awareness)** การวิจัยแสดงให้เห็นว่าการรู้เท่าทัน Deepfake เป็นทักษะทรงพลัง มีศักยภาพในการปกป้องสังคมจากข้อมูลที่บิดเบือน เป็นอีกมาตรการตอบโต้ที่แท้จริง

4 **ด้านการใช้ทัศนคติแบบ (Zero-Trust Mindset)** เปรียบเทียบได้กับแนวทาง 'กระตุกต่อมคิด' ที่ OKMD สนับสนุน ไม่ให้เชื่อหรือไว้วางใจสิ่งใดๆ ก่อนตรวจสอบในทันที ซึ่งนับเป็นอีกหนึ่งวิธีที่จำเป็นในยุคเทคโนโลยีเอไอ

ข้างอิง :
• aws.amazon.com
• weforum.org
• biometricupdate.com
• cyfence.com
• forbes.com

เมนูนี้เด็ด ร้านนี้ต้องไป เมื่อเอไอหลอกเราให้อยากชิม

Yale Center for Customer Insights ได้รายงานเกี่ยวกับการตั้งคำถามในโลกเอไอว่า เมื่ออ่านรีวิวร้านอาหารจะรู้ได้อย่างไรว่ามาจากมนุษย์หรือปัญญาประดิษฐ์ พร้อมระบุการศึกษาที่ยืนยันว่าผู้รับสารไม่สามารถแยกแยะได้ด้วยตนเอง จากความอัจฉริยะของเอไอ

นักวิจารณ์อาหารแนวใหม่ได้เกิดขึ้นแล้ว ด้วย Generative AI

เริ่มต้นจากความนิยมของแพลตฟอร์ม Generative AI ที่ทำให้การสร้างบทวิจารณ์เป็นเรื่องง่าย ทำให้ผู้อ่านไม่สามารถแยกแยะความแตกต่างระหว่างบทวิจารณ์จริงกับบทวิจารณ์ประดิษฐ์ และหากรีวิวในโลกออนไลน์มีอิทธิพลในการทำการตลาดมากที่สุดในกลุ่มผู้บริโภค ประเด็นนี้จึงสำคัญอย่างยิ่ง ทั้งในแง่การชี้นำผ่านความคิดเห็น ไปจนถึงการกล่าวเกินจริงเพื่อชวนให้เกิดการซื้อสินค้าและบริการ และในทางตรงกันข้ามกันนั้น บทวิจารณ์ประดิษฐ์สามารถสร้างและกระจายความคิดเห็นเชิงลบได้เช่นเดียวกัน

Balàzs Kovàcs ศาสตราจารย์ด้านพฤติกรรมองค์กรที่ Yale School of Management และผู้เขียนรายงานการศึกษานี้ได้ติดตามวิวัฒนาการการรีวิวร้านอาหารออนไลน์มาตั้งแต่ ค.ศ. 2004 เมื่อ Yelp เว็บไซต์รีวิวและค้นหาร้านค้าธุรกิจชื่อดังของอเมริกาเข้าสู่ตลาด และมีอิทธิพลต่อการตัดสินใจของผู้บริโภคในชีวิตประจำวัน เขายังได้กล่าวว่าแม้รีวิวที่ตีพิมพ์ส่วนใหญ่เขียนโดยมนุษย์ที่มีประสบการณ์การกินอาหารจริง อย่างไรก็ดี แพลตฟอร์ม Generative AI เช่น ChatGPT โดย OpenAI ก็สามารถสร้างบทวิจารณ์ขนาดยาว โดยอิงตามคำสั่งมนุษย์

เพียงไม่กี่ขั้นตอนได้เช่นเดียวกัน ทั้งยังตั้งข้อสังเกตว่า GPT-4 สามารถสร้างข้อความที่หลอกหรือโกงเครื่องตรวจจับ โดยเฉพาะเมื่อได้รับคำสั่งให้พิมพ์ผิดและใช้ภาษาพูดเพื่อให้บทวิจารณ์เหมือนเขียนโดยมนุษย์จริง

การสำรวจของ BrightLocal ระบุว่า 87% ของผู้บริโภคอ่านรีวิวออนไลน์เกี่ยวกับธุรกิจในท้องถิ่น และ 79% มองว่ารีวิวเหล่านั้นน่าเชื่อถือและเป็นคำแนะนำส่วนตัว ทำให้เห็นแนวโน้มว่าบทวิจารณ์ปลอมเหล่านั้นเสี่ยงทำลายชื่อเสียงและความน่าเชื่อถือของแต่ละแอปพลิเคชันหรือแต่ละแพลตฟอร์ม ทำให้ผู้อ่านสูญเสียความไว้วางใจ

ผลการสำรวจนี้ไม่เพียงส่งผลกระทบต่อผู้บริโภคที่เคยเชื่อมั่นการรีวิวในอุตสาหกรรมร้านอาหารที่ถูกต้อง ผู้เขียนรายงานยังตั้งข้อสังเกตอีกว่า ปัญหารีวิวปลอมยังจะสั่นสะเทือนในทุกวงการและอุตสาหกรรม ตั้งแต่การหาซื้อเครื่องดูดฝุ่นไปจนถึงการเลือกตั้งประธานาธิบดี วิถีอเมริกัน ที่หากไม่จัดการเนื้อหาปลอมเหล่านี้ ความรับผิดชอบจะต้องตกไปอยู่ที่ความรู้ และความสามารถในการวิเคราะห์ของพลเมืองออนไลน์เป็นหลัก จึงเป็นเหตุผลสำคัญในการติดอาวุธสร้างภูมิคุ้มกันให้ผู้บริโภคในโลกออนไลน์

นักสร้างภาพอาหารจากสูตร ความน่ากินที่ไม่ต้องเข้าครัวปรุง

ทีมนักวิจัยมหาวิทยาลัย Tel-Aviv University ในอิสราเอลได้พัฒนาโครงข่ายประสาทเทียม (Neural Network) ที่สามารถอ่านสูตรอาหารและสร้างภาพว่าผลิตภัณฑ์ที่ปรุงเสร็จแล้วจะมีลักษณะอย่างไร ทำให้ไม่สามารถแน่ใจได้ว่าอาหารน่าอร่อยที่เราเห็นในจานนั้นเป็นของจริงหรือไม่ โดยสมาชิกในทีมนักวิจัยได้สร้างเอไอโดยใช้เวอร์ชันดัดแปลงของ Generative Adversarial Network ที่เรียกว่า StackGAN V2 และการผสมผสานรูปภาพ/สูตรอาหาร 52K จากชุดข้อมูลขนาดใหญ่ยักษ์ Recipe1M ตามรายงานใน ค.ศ. 2019 ของเว็บไซต์ข่าวสาร การคิดค้น และพัฒนาเทคโนโลยี The Next Web

นักวิจัยกล่าวกับ TNW ว่า เอไอนี้เริ่มต้นจากหนึ่งในนักวิจัยของทีมได้ไปขอสูตรเมนูปลาทอดกับซอสมะเขือเทศในตำนานของคุณยายที่มีอายุมากและเริ่มหลงลืมยากแก่การจำสูตร จึงจุดประกายทำให้คิดค้นระบบการป้อนภาพอาหารและส่งออกผลลัพธ์มาเป็นสูตร อย่างไรก็ตาม ต้องยอมรับว่ามีข้อยกเว้นสำหรับส่วนผสมที่มีรายละเอียดที่ยากเกินไป หรือเครื่องปรุงที่ถูกซ่อนไว้ เช่น เนย แป้ง พริกไทย และเกลือ นำไปสู่การตั้งคำถามว่าเขาจะสามารถทำสิ่งที่ตรงกันข้ามแทนได้หรือไม่ ที่หมายถึงการสร้างภาพอาหารจากสูตร ให้ออกเป็นหน้าตาอาหารน่าอร่อยโดยไม่จำเป็นต้องเข้าครัวปรุง

'Recipe1M คืออะไร' สถาบันการศึกษาระดับโลก MIT หรือ Massachusetts Institute of Technology ได้พัฒนาโครงข่ายประสาทเทียมที่เรียกว่า Receipe1M ที่ไม่เพียงช่วยให้เราเรียนรู้สูตรอาหาร แต่ยังเข้าใจพฤติกรรมการกินและความชอบของผู้คนได้ดีขึ้น เอไอนี้ยังอาจพัฒนาไปถึงขั้นให้มีความสามารถในการวิเคราะห์

วิธีการเตรียมอาหาร การแยกแยะความแตกต่างของอาหาร ไปจนถึงการพัฒนาระบบผู้ช่วยออกแบบเมนูอาหารที่ประเมินได้จากวัตถุดิบในตู้ ความชอบ ไปจนถึงสูตรที่คล้ายกัน

ปฏิเสธไม่ได้ว่าทั้งการผลิต รีวิวปลอม และการสร้างภาพเกินจริง หรือภาพเมนูที่ไม่เคยลองทำจริงจากสูตร สามารถนำไปสู่ AI Deepfake ที่ใช้โฆษณาประชาสัมพันธ์ได้ แต่ในมุมกลับเอไอเหล่านี้สามารถนำไปใช้ในทางสร้างสรรค์ เช่น ใช้ Generative AI สร้างบทความให้ความรู้ด้านอาหาร การสร้างสรรค์ภาพเพื่อเติมเต็มจินตนาการจากชุดคำสั่ง หรือช่วยกระตุ้นการคิด วิเคราะห์สูตรจากภาพอาหารที่มี

ข้างอิง :
• analyticsindiamag.com • som.yale.edu • thenextweb.com

Deepfake

ภัยเงียบที่คุณต้องรู้ก่อนเป็นเหยื่อรายต่อไป

ในช่วงไม่กี่ปีที่ผ่านมา เทคโนโลยีปัญญาประดิษฐ์หรือ AI ได้พัฒนาก้าวกระโดด นำมาซึ่งทั้งความก้าวหน้าและความท้าทายที่เราคาดไม่ถึง หนึ่งในประเด็นที่น่ากังวลที่สุดในปัจจุบันคือการมาของเนื้อหาที่สร้างขึ้นโดย AI หรือที่เรียกกันว่า Deepfake ซึ่งหมายถึงสื่อต่างๆ ไม่ว่าจะเป็นรูปภาพ วิดีโอ หรือเสียง ที่ถูกสร้างหรือดัดแปลงโดย AI เพื่อบิดเบือนความจริง เมื่อเทคโนโลยีเหล่านี้พัฒนาขึ้นเรื่อยๆ การทำความเข้าใจผลกระทบของมันจึงเป็นเรื่องสำคัญสำหรับทุกคน

Deepfake คืออะไร?

ลองนึกภาพว่าคุณกำลังดูคลิปวิดีโอของดาราคนโปรดพูดอะไรบางอย่าง แต่ความจริงแล้วพวกเขาไม่เคยพูดประโยคนั้นเลย นี่คือตัวอย่างของ Deepfake ซึ่งก็คือสื่อปลอมที่สร้างขึ้นโดย AI เพื่อสร้างเนื้อหาที่ดูสมจริง แต่แท้จริงแล้วเป็นของปลอม เทคโนโลยีนี้ใช้ข้อมูลจำนวนมหาศาลในการสร้างเนื้อหาที่เลียนแบบ รูปภาพ วิดีโอ และเสียงในชีวิตจริงได้อย่างแนบเนียน

การแพร่กระจายของข้อมูลเท็จที่สร้างโดย AI

ความหลากหลายของ Deepfake ทำให้คนจำนวนมากเริ่มกังวลถึงการนำเทคโนโลยี AI ไปใช้ในทางที่ผิด โดยเฉพาะในการเผยแพร่ข้อมูลเท็จ ยกตัวอย่างเช่น ในปี 2566 มีการเผยแพร่เสียงปลอมของนักการเมืองในสโลวาเกียที่พูดถึงการโกงการเลือกตั้ง ซึ่งกลายเป็นไวรัลในโลกออนไลน์ สะท้อนให้เห็นถึงพลังของเนื้อหาที่สร้างโดย AI ในการสร้างความปั่นป่วนทางการเมือง

นอกจากการสร้างเนื้อหาปลอมแล้ว การกล่าวหาว่าบางสิ่งเป็นผลงานของ AI กลายเป็นกลยุทธ์ในการลดความน่าเชื่อถือของข้อมูลจริงไปด้วย เหมือนที่ Ellen Judson นักสืบสวนอาวุโสของ Global Witness ได้ตั้งข้อสังเกตว่า 'ไม่ชอบสิ่งที่บางคนกำลังพูด? การบอกว่าหลักฐานที่มีเป็นของที่สร้างโดย AI นั้นง่ายกว่าการต้องไปหาหลักฐานมาหักล้างเสียอีก' กลยุทธ์นี้ทำให้สาธารณชนแยกแยะความจริงจากความเท็จได้ยากขึ้น

ความยากในการตรวจจับ

การตรวจจับ Deepfake เป็นงานที่ซับซ้อน ซึ่งแม้จะมีเครื่องมือและวิธีการที่ออกแบบมาเพื่อระบุสื่อที่ถูกปรับแต่งด้วย AI ก็ตาม แต่เทคโนโลยีเบื้องหลัง Deepfake ก็พัฒนาไปอย่างรวดเร็ว จนบ่อยครั้งก็เร็วกว่าที่ความสามารถในการตรวจจับจะตามทัน โครงการ 'Detect Fakes' ของ MIT Media Lab เน้นย้ำว่า ไม่มีร่องรอยชี้ชัดเฉพาะเพียงอย่างเดียวที่จะบอกได้ว่าอะไรคือ Deepfake แต่แนะนำให้ทุกคนระมัดระวังและสังเกตความผิดปกติเล็กๆ น้อยๆ เช่น การเคลื่อนไหวของใบหน้าที่ไม่เป็นธรรมชาติ หรือการเคลื่อนไหวของริมฝีปากที่ไม่ตรงกับเสียง

ผลกระทบในวงกว้าง

ผลกระทบของ Deepfake ไม่ได้จำกัดอยู่แค่เรื่องการเมือง แต่ยังถูกนำไปใช้ในการสร้างเนื้อหาลามกอนาจารโดยไม่ได้รับความยินยอม และการฉ้อโกงทางการเงิน ซึ่งนำไปสู่ความเสียหายทั้งต่อบุคคลและสังคม ศักยภาพของ Deepfake ในการทำลายความเชื่อมั่นที่มีต่อสื่อและสถาบันต่างๆ นั้นรุนแรง เพราะผู้คนอาจเริ่มสงสัยในเนื้อหาที่ถูกต้อง โดยไม่แน่ใจว่าอะไรคือของจริง และอะไรคือของปลอม

กรณีศึกษาล่าสุด การหลอกลวงโดยใช้ชื่อ 'แบรด พิตต์'

เหตุการณ์ที่เพิ่งเกิดขึ้นช่วยตอกย้ำความเสี่ยงที่เกี่ยวข้องกับเนื้อหาที่สร้างโดย AI เรื่องมีอยู่ว่า ผู้หญิงชาวฝรั่งเศสคนหนึ่ง โดยใช้นามสมมุติว่า 'แอน' ถูกหลอกให้โอนเงินกว่า 28 ล้านบาท ให้กับคนร้ายที่แอบอ้างเป็นนักแสดงแบรด พิตต์ คนร้ายใช้รูปภาพและข้อความที่สร้างโดย AI เพื่อหลอกให้แอนเชื่อว่าเธอกำลังมีความสัมพันธ์โรแมนติกกับนักแสดงคนดัง ถึงขั้นแต่งเรื่องว่าแบรด พิตต์ต้องการเงินเพื่อรักษามะเร็งและหลอกให้แอนโอนเงินจำนวนมากให้

ตัวแทนของแบรด พิตต์ได้ออกมาเตือนสาธารณชนว่านักแสดงคนดังไม่ได้ใช้สื่อสังคมออนไลน์ และเตือนไม่ให้ตอบกลับการติดต่อทางออนไลน์ ซึ่งกรณีนี้แสดงให้เห็นถึงความเปราะบางทางอารมณ์ที่อาจถูกเอารัดเอาเปรียบผ่านการหลอกลวงด้วย AI ที่ซับซ้อน

การสร้างภูมิคุ้มกันต่อ Deepfake

การรับมือกับความท้าทายที่เกิดจาก Deepfake จำเป็นต้องใช้วิธีการที่หลากหลายประกอบกัน

1 **การสร้างความตระหนักรู้และให้ความรู้แก่สาธารณะ**

การให้ความรู้แก่สาธารณชนเกี่ยวกับการมีอยู่และผลกระทบที่อาจเกิดขึ้นจาก Deepfake เป็นสิ่งจำเป็น เมื่อเข้าใจว่าเทคโนโลยีเหล่านี้มีอยู่จริง ผู้คนจะสามารถเสพสื่อด้วยวิจารณญาณมากขึ้น

2 **การพัฒนาเทคโนโลยี**

การลงทุนและพัฒนาเครื่องมือตรวจจับขั้นสูงมีความสำคัญ ความร่วมมือระหว่างบริษัทเทคโนโลยี สถาบันการศึกษา และรัฐบาล สามารถนำไปสู่วิธีแก้ปัญหาที่มีประสิทธิภาพมากขึ้นในการระบุและลดการแพร่กระจายของ Deepfake

3 **นโยบายและกฎระเบียบ**

การกำหนดนโยบายและกฎระเบียบที่ชัดเจนเกี่ยวกับการสร้างและเผยแพร่เนื้อหาที่สร้างโดย AI สามารถยับยั้งผู้ที่มีเจตนาร้าย รวมถึงการกำหนดความรับผิดชอบของบุคคลและองค์กรที่ใช้เทคโนโลยี Deepfake ในทางที่ผิด

4 **การรู้เท่าทันสื่อ**

การส่งเสริมการรู้เท่าทันสื่อสามารถเสริมพลังให้ผู้คนประเมินเนื้อหาที่พวกเขาบริโภคอย่างมีวิจารณญาณ รวมถึงการตั้งคำถามเกี่ยวกับแหล่งที่มา การหาข้อมูลยืนยัน และการระมัดระวังก่อนแชร์เนื้อหาที่ยังไม่ได้รับการยืนยัน

การมาถึงของเนื้อหาที่สร้างโดย AI นำเสนอทั้งโอกาสและความท้าทาย แม้ว่าเทคโนโลยีนี้จะสามารถนำไปใช้อย่างสร้างสรรค์และเป็นประโยชน์ได้ แต่ความเป็นไปได้ที่จะถูกนำไปใช้ในทางที่ผิดก็ทำให้เราต้องระมัดระวังด้วยการติดตามข้อมูลให้ทันสมัย สนับสนุนการพัฒนาเทคโนโลยีการตรวจจับ และสร้างวัฒนธรรมการบริโภคสื่ออย่างมีวิจารณญาณ

EDITOR'S NOTE

ค.ศ. 2025 เป็นปีที่เพิ่มขึ้นอย่างมหาศาล เมื่อพูดถึงการเติบโตของ AI Deepfake ทั้งในเชิงลบและเชิงบวก ปฏิเสธไม่ได้เลยว่าเราได้เห็นการใช้ AI Deepfake ที่สร้างความเสียหายในรูปแบบต่างๆ เพิ่มขึ้นอย่างเห็นได้ชัด ทำให้แนวทางการใช้เทคโนโลยี AI Deepfake อย่างมีจริยธรรมและในประเด็นการบังคับใช้ทางกฎหมายเป็นเรื่องสำคัญ ควบคู่กับการใช้งาน

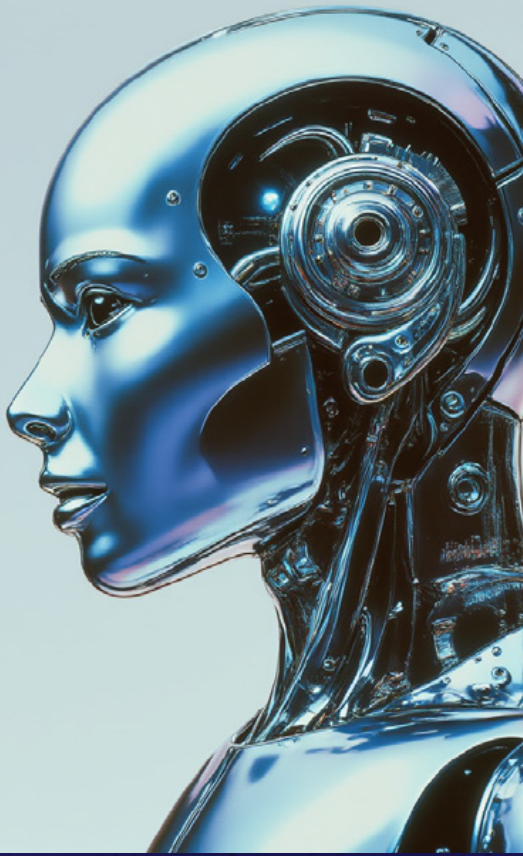
แนวโน้มที่เพิ่มขึ้นนี้ ยังถูกตอกย้ำด้วยรายงานของ TNGlobal แหล่งข้อมูลเชิงลึกและเครือข่ายเทคโนโลยีและนวัตกรรมระหว่างจีนและภูมิภาคเอเชียแปซิฟิกที่ได้ระบุว่า ในปีเดียวกันนี้ เป็นปีที่ APAC รวมถึงประเทศไทย มีประเด็น AI Deepfake เป็นกระแสหลัก ยังไม่รวมถึงนานาชาติในโลกดิจิทัลที่ไม่มีพรมแดน

ในอีกมุม เราสามารถต่อยอดโอกาสและใช้ประโยชน์จาก AI Deepfake ได้ในหลายแวดวงและอุตสาหกรรม ตั้งแต่ด้านความบันเทิง การศึกษา ไปจนถึงการแพทย์ อย่างไรก็ดี เหล่านี้เป็นเพียงมุมมองจากการเฝ้าสังเกตและการเรียนรู้เทคโนโลยีเกิดใหม่

ในช่วงเริ่มต้นไปด้วยกัน ซึ่งการเปิดรับและเรียนรู้ต่อเนื่องจะเป็นแนวคิดสำคัญ ทำให้เราก้าวไปข้างหน้าอย่าง 'รู้เขา รู้เรา รู้จักเอาตัวรอด' ท่ามกลางการแข่งขันในยุคเทคโนโลยีเอไอ

okmd x AIAT

OKMD
CAREER BOOTCAMP 2025
SEASON 2



AI IMAGE
การสร้างภาพด้วย Generative AI

วันเสาร์ที่ 26 เม.ย. 2568
เวลา 13:00 -16:00 น.

AI MEDIA
การสร้างสื่อวิดีโอ และเสียง ด้วย Generative AI

วันเสาร์ที่ 17 พ.ค. 2568
เวลา 13:00 -16:00 น.

DIGITAL ETHICS & LAWS
จริยธรรมและกฎหมายดิจิทัล เพื่อการใช้งาน AI

วันเสาร์ที่ 7 มิ.ย. 2568
เวลา 13:00 -16:00 น.

ติดตามการสมัครได้ที่ okmd AIAT

Learn Lab 2025
Creativity Beyond AI

okmd x x

เสาร์ที่ 15 มี.ค. 2568 นี้
พบกันออนไลน์กับ
Learn Lab 2025: Creativity Beyond AI³

15.10 - 15.50 น.
การเสวนา "Future of Learning in an AI-Driven World"

โดยเหล่าวิทยากรชื่อดัง
ผู้นำในอุตสาหกรรมด้านต่าง ๆ

15.50 - 16.20 น.
กิจกรรมไฮไลต์ Final Pitching

เพื่อค้นหา 5 ทีมสุดท้ายเจ้าของสุดยอดไอเดีย
ผู้จะมาคว้ารางวัล มูลค่ากว่า 200,000 บาท
จากการประยุกต์ใช้ AI
ในอุตสาหกรรม Mega Trend

รับชมได้ทาง Live okmd

พบกับโซนบูทกิจกรรมทั้ง 3 โซน Innovation / Inspiration / Imagination
พร้อมบูทอาหารได้ภายในงาน!