AI op de werkvloer: DTC en politie delen tips

05-12-2023 09:53



AI brengt op de werkvloer risico's met zich mee, maar met de nodige voorzichtigheid en voorzorgsmaatregelen zijn die in te perken, net als bij andere vormen van digitaal gevaar. Het Digital Trust Center (DTC) en de politie delen tips over hoe zowel werknemers als werkgevers veilig om kunnen gaan met AI. Menselijk contact is hierbij de sleutel.

Terwijl AI-hulpmiddelen zoals tekstgeneratoren en beeldcreatie-apps ondernemers aanzienlijke (efficiëntie) voordelen bieden, is er ook een donkere zijde aan deze technologieën. Ook cybercriminelen kunnen deze tools gebruiken voor frauduleuze praktijken.

1. Identiteitsfraude, zoals CEO-fraude. Zo kan AI ingezet worden om bijvoorbeeld een stem te klonen of om realistische teksten op te stellen.
2. Desinformatie verspreiden. Taalmodel ChatGPT produceert authentiek lijkende teksten op schaal en met grote snelheid. Een dergelijk taalmodel kan criminelen helpen voor propaganda- en desinformatiedoeleinden.
3. Malware. ChatGPT is in staat om codes te produceren in een aantal verschillende programmeertalen.

> Voor een potentiële crimineel met weinig technische kennis is dit een onschatbare bron om kwaadaardige codes (zoals malware) te produceren.

Manon den Dunnen, Strategisch Specialist Digitaal bij de politie, benadrukt het belang om waakzaam te zijn als je zelf gebruik maakt van AI: “Als je het niet op LinkedIn zou zetten, moet je het ook niet invoeren op ChatGPT. Want dat systeem traint zichzelf met de informatie die je invoert en voordat je het weet komt jouw informatie terug in teksten die gegenereerd zijn voor anderen. Daarom hebben bedrijven zoals Samsung hun medewerkers verboden het te gebruiken.”

Tips voor het omgaan met kunstmatige intelligentie en cybercriminelen die hier gebruik van maken:

- Vertrouwelijke gesprekken kun je het beste in persoon voeren.
- Voer nooit vertrouwelijke gegevens in ChatGPT of vergelijkbare taalmodellen. Dus ook geen namen van personen. Wees bewust dat de systemen gericht zijn op het genereren van teksten ‘die lijken op’. Het is geen zoekmachine, er zit namelijk geen database achter, dus gebruik het niet als feitelijkheid belangrijk is.
- Bij twijfel over de identiteit van de persoon aan de telefoon, kun je voorstellen om terug te bellen. Een andere mogelijkheid is om een belevingsvraag te stellen. Bijvoorbeeld: Hoe was je gesprek gisteren?
- Er kunnen afspraken gemaakt worden om bijvoorbeeld facturen alleen af te handelen wanneer er een mogelijkheid bij zit om te controleren wat de bron is.
- Onderzoek welke oplossingen je in afstemming met partners in de keten kunt doorvoeren om de authenticiteit van de afzender van facturen of andere belangrijke communicatie, vast te stellen. Grijp terug op adviezen relevant voor bijvoorbeeld phishing of CEO-fraude. Deze vormen van cyberincidenten blijven in de basis hetzelfde ook al wordt AI als hulpmiddel ingezet.
- Weet welke vragen je moet stellen bij de aanschaf van inkoop van software. Bijvoorbeeld: Op welke manier maakt deze software gebruik van kunstmatige intelligentie, hoe is het getraind, wat gebeurt er met deze data en welke veiligheidsvraagstukken spelen er?

Frank Veerkamp