LeA UITGEVERS

INTERVIEW

Artificiële intelligentie: instrument, dader of… zelfs slachtoffer van een misdrijf

Bij LeA Uitgevers is in september 2024 het boek 'AI-criminaliteit. Nieuwe uitdagingen in het (ondernemings)strafrecht' van mr. Julie Petersen verschenen. LeA Uitgevers interviewde de auteur naar aanleiding van deze publicatie.

De steeds toegankelijkere digitale en AI-technologieën van vandaag vergemakkelijken het (off- en online) leven aanzienlijk, en deze tendens is ook op het vlak van het straf(proces)recht voelbaar. Zoals echter steeds het geval is wanneer het op maatschappelijke (r)evoluties aankomt, zijn het niet alleen legale partijen die hierin interesse tonen. Aan een vaak nog hogere snelheid ontdekken momenteel ook (cyber)criminelen wat AI allemaal te bieden heeft.
In hoeverre is ons (huidig en toekomstig) strafrecht aangepast aan de nieuwe criminaliteitsvormen die AI met zich meebrengt?

In dit boek wordt het antwoord op deze vraag gezocht. Het geeft vanuit de invalshoek van het materieel ondernemingsstrafrecht een eerste beschrijving van AI-criminaliteit die de rechtspracticus nu en later voor de nodige uitdagingen zal stellen.
Meer bepaald komen aan de hand van enkele actuele en praktische casussen drie onderscheiden vormen van AI-criminaliteit aan bod: criminaliteit met AI (AI als 'instrument' van het misdrijf), criminaliteit door AI (AI als 'dader' van het misdrijf) en criminaliteit tegen AI (AI als 'slachtoffer' van het misdrijf).

AI-criminaliteit. Nieuwe uitdagingen in het (ondernemings)strafrecht, september 2024,
ISBN 978-94-646-6242-9
een uitgave van LeA Uitgevers: www.lea-uitgevers.be/shop

Is ons (nieuw) strafrecht nu al dan niet aangepast aan de uitdagingen die de recente technologieën van AI met zich meebrengen?

Op de vraag of het bestaande (en toekomstige) materieel strafrechtelijke kader voldoende **bescherming** biedt in een AI-tijdperk, kan helaas niet eenduidig worden geantwoord.

Verschillende Belgische **delictsomschrijvingen**, zoals die van de valsheid in informatica (art. 210bis oud Sw. en art. 451 nieuw Sw.), lijken over het algemeen alleszins voldoende technologieneutraal geformuleerd om er bijvoorbeeld het kwaadwillige gebruik van nieuwe AI-technologieën onder te kunnen kwalificeren.

Daartegenover staat het soms grote contrast tussen bekende vormen van cybercrime enerzijds en meer recente vormen van criminaliteit gepleegd met AI anderzijds. Het risico van het **toenemende gebruik (misbruik) van AI op criminologisch vlak** is niet te onderschatten, en toch wordt er zelfs in de parlementaire voorbereidingen van het nieuwe Strafwetboek, dat in principe in werking zal treden op 8 april 2026, met geen woord gerept over AI. Dit terwijl het gemak waarmee goedkope en toegankelijke technologieën kunnen worden ingezet door mensen met 'criminele intenties' maar met weinig tot geen technische voorkennis, de vroegere **drempels op gevoelige wijze verlaagt**. Des te meer daders kunnen overigens des te meer slachtoffers maken met de moderne tools die hen vandaag ter beschikking staan.

En wat alle besproken vormen van AI-criminaliteit (als instrument, als dader en als slachtoffer van het misdrijf) in mijn boek gemeen hebben, is dat *they're here to stay*. Wat de eigenlijke implicaties van de onafwendbare toename ervan zullen zijn, zal echter pas duidelijk worden wanneer de **eerste (straf)rechtszaken** i.v.m. AI zich (binnenkort?) aandienen. In de tussentijd is het de taak van de 'penalisten' om niet alleen kritisch te reflecteren over de (strafrechtelijke) **implicaties** van de nieuwe AI-(r)evolutie, maar en ook en vooral over de **noodzakelijke preventieve rol** die zij kunnen spelen in dit maatschappelijk **debat**.

Kunt u vooreerst de eerste twee onderscheiden vormen toelichten: criminaliteit met AI, dus AI als 'instrument' van het misdrijf en criminaliteit door AI, dus als 'dader' van het misdrijf?

In het **eerste hoofdstuk**, waar ik **AI als 'instrument' van het misdrijf** behandel, zoom ik in op traditionele vormen van (financiële) criminaliteit, die in de toekomst meer en meer met behulp van AI kunnen en zullen worden gepleegd. Specifiek worden de zogenaamde **deepfakes** onder de loep genomen, een fenomeen dat naar verwachting enkel maar zal groeien, of het nu in de vorm van voice cloning, face-swapping of een combinatie van beide is. Het gemak en de toegankelijkheid van de software, de ongelimiteerde mogelijkheden tot verspreiding via het internet en vooral hun misleidende kracht maakt van deepfakes een reëel (virtueel) gevaar, dat quasi onmogelijk in te dijken valt …

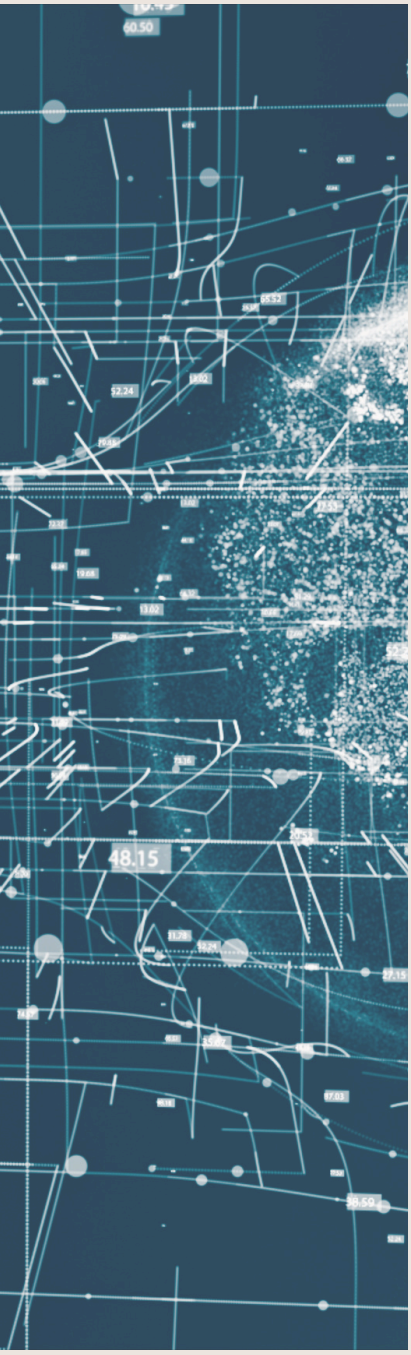
Zo waarschuwde het **Europees Agentschap voor Cyberbeveiliging** ('ENISA') er al voor dat criminelen met de deepfake-technologie **gezichten kunnen nabootsen** om op die manier onder meer gezichtsherkenningscontroles van financiële instellingen te omzeilen. Andere voorbeelden van het gebruik van de deepfake-technologie met kwaadwillige intenties zijn het fenomeen van de zogenaamde **crypto scams** (waarbij bekende mensen in een deepfake-filmpje zogezegd oproepen om te investeren in frauduleuze cryptomunten) en dat van de **deepnudes** (het virtueel 'uitkleden' van mensen en er op basis van een (zelfs aangeklede) foto of video pornografische beelden van maken, bijvoorbeeld om ze er nadien mee af te persen ('sextortion')).

Afhankelijk van de concrete situatie waarin deepfakes voorkomen, kon ik besluiten dat heel wat **misdrijfkwalificaties relevant én toepasbaar** zijn. Daarnaast zal het **AI-systeem** dat gediend heeft of bestemd was om misdrijven te plegen, althans in theorie, ook kunnen worden verbeurdverklaard **als instrument** van het misdrijf.

Toch zal deze eerder theoretische kwalificatievaststelling in de praktijk vaak maar een pyrrusoverwinning zijn. Zoals bij alle vormen van cybercriminaliteit is het **opsporen en vervolgen** ervan immers **geen sinecure**. Het valt te verwachten dat, zeker in een AI-fraudeverhaal, verschillende gekende problemen de kop zullen opsteken: moeilijkheden m.b.t. de identificatie van daders die vaak anoniem opereren en hun (veelal verschillende buitenlandse) locatie(s) die ze goed kunnen afschermen, gelden die (al dan niet via de welbekende money mules) zo snel mogelijk doorgesluisd worden naar rekeningen in landen waar ze nog zeer moeilijk te recupereren vallen, enzovoort. Hoewel ze niet het thema van mijn boek uitmaakten, kunnen deze en meer **problematieken** uiteraard niet los gezien worden van '(EcoFin-)criminaliteit gepleegd met behulp van AI'.

In het **tweede hoofdstuk** behandel ik **AI als 'dader'** van het misdrijf. Het AI-systeem vertoont dan – **los van zijn makers of gebruikers** - ongewenst (en strafrechtelijk relevant) gedrag, bijvoorbeeld een beursalgoritme dat ervoor zorgt dat de prijs van bepaalde aandelen kunstmatig wordt opgevoerd ('pump') om vervolgens snel zelf in grote volumes te verkopen en winst te maken alvorens de prijs instort ('dump').

Mijn visie is hier dat criminaliteit gepleegd 'door AI' naar (huidig en toekomstig) Belgisch recht **niet zomaar 'bestraft'** kan worden. Ten eerste is het (althans voorlopig) juridisch niet mogelijk noch wenselijk om AI-systemen zelf strafrechtelijk aansprakelijk te stellen, en in de tweede plaats vallen meer in het algemeen ook de regels van de strafrechtelijke toerekening aan andere rechtssubjecten **niet altijd zomaar te transponeren naar AI-context**.



De reflex om de **bestaande principes** rond strafrechtelijke verantwoordelijkheid zonder meer toe te passen wanneer een AI-systeem overgaat tot het (materieel) plegen van strafbare feiten, is met andere woorden niet zo logisch als initieel gedacht. De zuivere toepassing van het strafrecht op de besproken problematiek lijkt voor meer **vragen** dan oplossingen te zorgen.

In mijn boek beschrijf ik uitvoerig een aantal moeilijkheden die niet alleen **strafrechtelijke consequenties** hebben (rechtsonzekerheid, straffeloosheid, enzovoort), maar die zich ook vertalen in een potentiële **belemmering** op het vlak van het gebruik en de **ontwikkeling van AI** zelf. Het trachten oplossen van deze moeilijkheden is geenszins vanzelfsprekend, en zal veelal **overkoepelende (wetgevende) initiatieven vereisen**, over de grenzen van het strafrecht en andere rechtstakken heen. Vergeet ook niet dat het strafrecht slechts één (ultiem) instrument is, en wellicht inderdaad niet altijd het juiste om ongewenst gedrag gesteld door of via AI te 'bestrijden'. Wanneer het gaat om criminaliteit gepleegd 'door AI' zullen bepaalde onvermijdelijke gaten in de strafrechtelijke verantwoordingsplicht moeten worden opgevuld door **andere rechtsgebieden**. Het strafrecht dient niet te worden beschouwd als een remedie voor alle kwaad, zelfs niet voor 'robo-kwaad'.

AI als 'slachtoffer' van het misdrijf is uiteraard een opvallende omschrijving, die zeker om verduidelijking vraagt. Wat bedoelt u hiermee?

Naarmate de inzet van AI-systemen in ons dagelijks leven toeneemt, worden deze systemen tegelijkertijd een steeds **aantrekkelijker doelwit** voor diegenen die de **gebreken** ervan willen uitbuiten. Het risico bestaat dus dat de technologie zélf het voorwerp wordt van criminaliteit. Zo verwijzen studies rond de mogelijke voordelen van autonome voertuigen vaak in één adem naar de bijhorende **gevaren** die een eventuele **hacking** ervan zou inhouden. Ook op andere domeinen loeren aanvallen 'gericht op' of 'tegen' AI om de hoek. De overkoepelende term van de **data poisoning attacks**, welke uitgebreid worden besproken in mijn boek, zijn er bijvoorbeeld op gericht om trainingsgegevens van AI-systemen te manipuleren met als doel de besluitvormingsprocessen van het systeem te beïnvloeden. De 'vergiftiging' kan van binnenuit of van buitenaf komen en kent verschillende varianten.

Ook deze fenomenen zullen maatschappelijke herdenkingen vereisen. Niet alleen hebben aanvallen 'tegen' AI uiteraard **strafrechtelijke implicaties** (hoe en op grond van welke kwalificatie kunnen of dienen zij bijvoorbeeld te worden gestraft?), maar ook het toenemende belang van **cybersecurity** valt samen met deze aanvallen niet te onderschatten. Als gegevensvergiftiging effectief een van de *next big threats* is op het gebied van cyberbeveiliging, valt te verwachten dat de **IT-industrie** in de nabije toekomst de nodige energie zal steken in het **beperken van deze bedreiging**, weze het via 'tegenaanvallen' of tools voor het opsporen van vergiftiging ('tegengif'), dan wel via het uitsluitend inzetten van gesloten AI-modellen die niet meer vertrouwen op open source-gegevens die gemanipuleerd kunnen worden.

AI-criminaliteit. Nieuwe uitdagingen in het
(ondernemings)strafrecht, september 2024,
ISBN 978-94-646-6242-9
een uitgave van LeA Uitgevers: www.lea-uitgevers.be/shop