

DE KOPEREN RIDDERS VAN LANCELOT

ca. 1220

dertig koperen buisjes, waaruit akelige kreten klinken, en vervolgens valt hij in slaap. Bij zijn ontwaken ontdekt hij dat de koperen vrouw op de grond ligt en dat de koperen ridders uit elkaar zijn gespat.

Historicus Jessica Riskin schrijft hierover: 'De ridder- en maagd-*automata* uit de Arthurlegende werden omringd door gouden, zilveren en koperen kinderen, saters, boogschutters, muzikanten, orakels en reuzen. Deze verzonnen, kunstmatige wezens hadden bestaande tegenhangers. In de late middeleeuwen en vroegmoderne tijd werd het landschap van Europa bevolkt door mechanische mensen en dieren.' Zo bedacht de Franse kunstenaar en ingenieur Villard de Honnecourt (ca. 1225-ca. 1250) ongeveer rond dezelfde tijd als het verhaal van Lancelots koperen ridders een mechanische adelaar die zijn kop naar de diaken kon draaien als deze voorlas uit de Bijbel. Volgens Riskin vormden deze levensechte *automata* deels de oorsprong voor het wetenschappelijke en filosofische idee dat levende wezens in feite machines zijn, een idee dat opkwam in de zeventiende eeuw.

ZIE OOK Talos (ca. 400 v. Chr.), Het mechanisch park van Hesdin (ca. 1300), De robotridder van Da Vinci (ca. 1495), Golem (1580), Tik-Tok (1907)

EENVOUTIGE VOORBEELDEN VAN AI, IN DE VOEM van mechanische figuren en wezens, werden tijdens de middeleeuwen in Europa gemeengoed, aldus historicus Elly Truitt: 'Gouden vogels en dieren, muzikale fontein en robothebdenen verbijsterden en verschrikten de gasten (...). De *automata* vormden het snijpunt tussen kennis van de natuurlijke wereld (inclusief magie) en technologie en (...) waren een verontrustende schakel tussen kunst en natuur.' Zowel de echte als fictieve apparaten uit de literatuur geven een fascinerend beeld van een 'onderling verband tussen wetenschap, technologie en verbeelding'.

Een beroemd voorbeeld van fictieve middeleeuwse robots komt uit *Lancelot du Lac* (Lancelot van het meer, ca. 1220), een oud Frans proza over de avonturen van koning Arthur en zijn ridders van de ronde tafel en de geheime romance tussen Lancelot en Guinevere (Arthurs vrouw). Voor de muren van het kasteel Douloureuse Garde, een angstaanjagend en betoverd kasteel, stuit Lancelot op een legertje koperen robotridders. Nadat hij het kasteel is binnengegaan, verslaat hij nog twee koperen ridders met zwaarden, die een kamer bewaken waarin een koperen jonge vrouw de sleutels bewaart die de betovering kunnen verbreken. Met de sleutels opent hij een doos met daarin

Lancelot vecht met de koperen ridders voordat hij het kasteel Douloureuse Garde binnengaat. Deze ridders worden vaak naakt afgebeeld. (Uit *Lancelot du Lac*, Frankrijk, 15e eeuw, Parijs, BnF, MS Fr. 118, fol. 200v.)

des enchauteuens



Ors se leigne et entra dedans si
munt lescu deuant son vis car u
ny voit goute fors parmi vne
bace d'un huis moult loig dunt



‘DARWIN ONDER DE MACHINES’

DE ENGELSE SCHRIJVER EN WETENSCHAPPER Samuel Butler (1835-1902) verkende als een van de eersten de mogelijke AI's van de toekomst en voorspelde concepten zoals zelfverbeterende machinale superintelligentie en de mogelijke risico's ervan. In zijn verbijsterende essay 'Darwin onder de machines' uit 1863 bespreekt Butler de toekomst van het 'mechanische leven': 'Wij scheppen zelf onze eigen opvolgers: wij voegen dagelijks iets toe aan de schoonheid en teerheid van hun fysieke samenstelling; wij geven hun dagelijks meer macht en leveren door allerlei vernuftige vindingen die zelfregulerende, zelfhandelende macht die voor hen zal zijn wat het intellect was voor het menselijk ras. In de loop van de tijd zullen wij merken dat wij het inferieure ras zijn.'

Met griezelige opmerkzaamheid voorspelt Butler dat machines het geleidelijk zullen overnemen van de mens: 'Wij worden steeds meer ondergeschikt aan hen; dagelijks worden steeds meer mensen door hen tot slaaf gemaakt (...) die hun hele leven wijden aan de ontwikkeling van mechanisch leven (...). Er zal een tijd komen dat de machines de wereld en haar bewoners in hun macht hebben...'

In *The Book of the Machines* (1872) stelt Butler dat een weekdier ogenschijnlijk niet veel bewustzijn heeft, maar dat het menselijk bewustzijn er toch uit is voortgekomen. Zo

zullen ook machines een bewustzijn krijgen en hij vraagt ons 'stil te staan bij de buitengewone vooruitgang die machines in de afgelopen honderden jaren hebben doorgemaakt, en te zien hoe langzaam het dieren- en plantenrijk vooruitgaat. De meer georganiseerde machines zijn niet zozeer wezens van gisteren als wel van de afgelopen vijf minuten (...).'

De ideeën van Butler weerklonken tot ver in de twintigste eeuw. Norbert Wiener (1894-1964), vader van de cybernetica, schreef: 'Als wij machines gaan maken die kunnen leren en die hun gedrag kunnen aanpassen aan hun ervaringen, moeten we accepteren dat elke mate van onafhankelijkheid die we aan de machines geven een zekere mate van ongehoorzaamheid aan onze wensen inhoudt. We zullen de geest niet zomaar terug in de fles krijgen, en we hebben geen reden te veronderstellen dat ze ons gunstig gezind zullen zijn.'

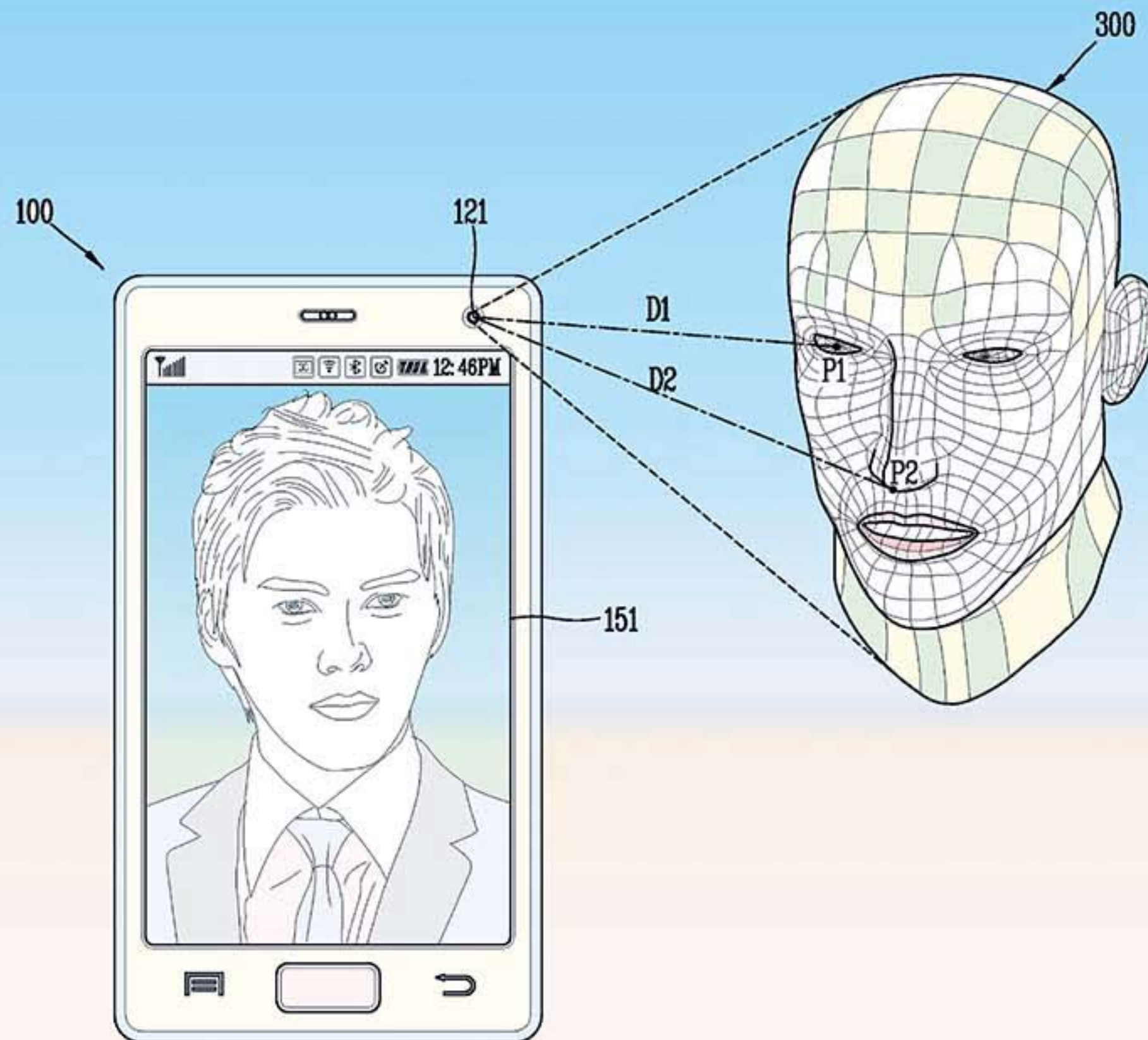
Nu de technologie van de eenentwintigste eeuw verweven is in alle aspecten van het menselijk leven lijken de uitspraken van Butler en Wiener over AI bepaald profetisch.

ZIE OOK *Leviathan* van Hobbes (1651), *Het menselijk gebruik van mensen* (1950), *Intelligentie-explosie* (1965), Lekkende 'AI-box' (1993), *De paperclipcatastrofe* (2003)

In 'Darwin onder de machines' schreef Samuel Butler: 'Wij scheppen zelf onze eigen opvolgers (...). In de loop van de tijd zullen wij merken dat wij het inferieure ras zijn.'

GEZICHTSHERKENNING

1964



SYSTEMEN VOOR GEZICHTSHERKENNING PROBEREN mensen te herkennen aan de hand van afbeeldingen of videobeelden, vaak door gezichtskenmerken (positie van neus en ogen) te vergelijken met die in een afbeeldingendatabase. Sommige moderne systemen leggen informatie vast met 3D-sensoren, wat ook de nauwkeurigheid verhoogt bij lichtvariaties en verschillende invalshoeken. Nauwkeurige gezichtsherkenning kent allerlei uitdagingen: hoeden of zonnebrillen, of zelfs make-up, kunnen de identificatie moeilijker maken, maar moderne algoritmes kunnen onder sommige omstandigheden soms beter gezichten herkennen dan mensen. De oorsprong van de gezichtsherkenningstechnologie is terug te voeren op het negentiende-eeuwse Engeland, waar in 1852 een systeem werd geïntroduceerd waarbij gevangenen werden gefotografeerd, als een meer humane methode dan brandmerken, om ze zo op het spoor te blijven en om hun gegevens met andere politiebureaus te delen bij ontsnapping. Een van de pioniers van meer geavanceerde gezichtsherkenning was de wiskundige en computerwetenschapper Woody Bledsoe (1921-1995), die in 1964 aan een vroege vorm van gezichtsherkenning werkte. Het viel hem op dat de draaiing en positie van het hoofd, de lichtinval, gezichtsuitdrukking en dergelijke dit moeilijker maakten. Bledsoe en

andere pioniers maakten vooral gebruik van de samenwerking tussen mens en computer, waarbij mensen met de hand gezichtscoördinaten aangaven op foto's met behulp van een digitaal tekenbalkje (tekenblok).

In de loop van de jaren hebben systemen voor gezichtsherkenning gebruikgemaakt van allerlei technieken, zoals eigenfaces, verborgen Markovmodellen en dynamische link matching. Gezichtsherkenning heeft tegenwoordig verschillende belangrijke toepassingen, zoals de technoloog Jesse Davis West uitlegt: 'Wetshandhavinginstanties gebruiken gezichtsherkenning om gemeenschappen veiliger te maken. Winkeliers voorkomen misdaad en geweld. Luchthavens verbeteren de veiligheid en maken reizen eenvoudiger. Telefoonbedrijven gebruiken gezichtsherkenning om hun klanten nieuwe vormen van biometrische beveiliging te kunnen aanbieden.' Toch roept dit de vraag op of het misschien een verontrustend kantelpunt in de beschaving vormt als mensen nergens in de publieke ruimte meer anoniem kunnen zijn.

ZIE OOK Optische tekenherkenning (OCR) (1913), Spraakherkenning (1952), De AIBO (1999)

Afbeelding uit US Patent 9.703.939 voor een methode om een mobiele telefoon veilig te kunnen ontgrendelen met behulp van de camera en gezichtsherkenning.



CHAT-GPT

IN DECEMBER 2022 SCHREEF KEVIN DOOSE VAN *The New York Times*: 'ChatGPT is, simpel gezegd, de beste kunstmatige intelligente chatbot die ooit is uitgebracht voor het gewone publiek... ChatGPT voelt anders. Slimmer. Vreender. Flexibeler.'

ChatGPT (chat generatieve vooraf getrainde transformator) is een AI-toepassing, ontwikkeld door het OpenAI laboratorium, die gesprekken voert met mensen en daarvoor gebruikmaakt van grote taalmodellen (LLM's) en kunstmatige neurale netwerken die zijn getraind aan de hand van tekstverzamelingen zoals Wikipedia, bulletinboards op internet, artikelen en software-handleidingen. Dankzij de uitvinding van de algoritmes en architectuur voor deze 'transformators' is parallelle verwerking op grote schaal mogelijk, waarmee deze modellen sneller worden gegenereerd.

In januari 2023 had ChatGPT meer dan honderd miljoen gebruikers, waarmee het een van de snelstgroeiende softwaretoepassingen voor consumenten in de geschiedenis is. Andere LLM-chatbots volgden al snel, zoals Bard van Google. Dit type chatbots kan natuurlijk lijkend proza schrijven over een ogenschijnlijk oneindig aantal onderwerpen. Ze kunnen ook computerprogramma's schrijven, gedichten en essays genereren, teksten samenvatten of vereenvoudigen, activiteiten zoals vakanties plannen, wetenschappelijke concepten uitleggen op

verschillende niveaus van moeilijkheid of verfijning, en nog veel meer.

Hoewel het heel indrukwekkend is wat dit type zeer geavanceerde chatbots allemaal kan, brengt hun gebruik ook risico's mee. Zo kunnen ze bijvoorbeeld 'hallucinaties' genereren (geloofwaardig klinkende verwijzingen naar gepubliceerd materiaal dat in werkelijkheid niet bestaat). Ook kunnen scholieren en studenten ze misbruiken om hun huiswerk mee te maken, kunnen ze instructies geven voor het maken van gevaarlijke materialen en bepaalde banen overbodig maken, zoals die waarvoor creatief schrijven is vereist, of helpdesks.

Volgens sommige onderzoekers zijn LLM's niet meer dan stochastische, of willekeurige, 'papegaaien', waarmee ze het idee benadrukken dat de LLM's de betekenis van hun output niet echt begrijpen. Aan de andere kant schreef de filosoof David Chalmers in 2023: 'Het lijkt volstrekt mogelijk dat we in het komende decennium robuuste (LLM) systemen zullen hebben met zintuigen, belichaming, wereldmodellen en zelfmodellen, terugkerende verwerking, een wereldwijde werkplaats en een verenigd doel [...] Het is aannemelijk dat we binnen tien jaar bewuste LLM's (uitgebreide LLM's) zullen hebben.'

ZIE OOK Kunstmatige neurale netwerken (1943),
Computerlinguïstiek (1954), DALL-E (2021)

Hoewel ze ogenschijnlijk intelligent en creatief zijn, hebben de taalkundige Emily Bender en haar collega's geavanceerde chatbots en LLM's omschreven als niet meer dan 'stochastische papegaaien', die in feite teksten imiteren zonder dat zij ze begrijpen.