

GUBERNA^o

INSTITUUT VOOR BESTUURDERS
INSTITUT DES ADMINISTRATEURS

AI voor bestuurders: 1. Definities en mogelijke impact

Olivier Braet, Sander Berghmans
Oktober 2023



Inhoudsopgave

1.	<i>Inleiding</i>	3
2.	<i>Wat is AI?</i>	4
2.1.	Artificiële Intelligentie: Definities	4
2.2.	Algemene kenmerken van Artificiële Intelligentie	5
2.3.	Artificiële Intelligentie: Technische architectuur	7
2.4.	AI-toepassingen	8
3.	<i>Positieve en negatieve impact van AI</i>	10
3.1.	Mogelijke positieve impact van Artificiële Intelligentie	10
3.2.	Mogelijke misbruiken van Artificiële Intelligentie	12
3.2.1.	Opzettelijk kwaadwillig gebruik van Artificiële Intelligentie	12
3.2.2.	Onbedoelde negatieve gevolgen van Artificiële intelligentie	13
3.3.	De reactie op AI: Ethische codes, manifesten, wetgeving	15
4.	<i>Conclusie</i>	19
	<i>Referenties</i>	20

Deze GUBERNA studie kwam tot stand met de steun van



GUBERNA partner in 'Governance and Digital Transformation'

1. Inleiding

De doorbraak van Artificiële Intelligentie (AI) kan grote schokken veroorzaken in het sociaal en economisch weefsel van onze samenleving. Hoewel eerdere vormen van AI al een tijd worden ingezet in verschillende sectoren zoals de banken, e-commerce, verzekeringen, gezondheidszorg, energie, defensie of cyberbeveiliging, wordt de technologie steeds krachtiger en wordt ze steeds breder ingezet. De technologie heeft een groeiende impact op de interne bedrijfsprocessen, de ontwikkeling van nieuwe diensten, de vereiste personeelscompetenties, de externe waardeketens en waardenetwerken, ganse sectoren, en bij uitbreiding de ganse arbeidsmarkt. De stormachtige opkomst van AI moet in de bredere context van voortschrijdende industrialisering en automatisering van onze economie gezien worden.

De stormachtige evolutie van de mogelijkheden van de technologie gaat gepaard met nieuwe verplichtingen en aandachtspunten voor de bestuurders. Bij GUBERNA beschouwen wij AI dan ook als één van de belangrijkste toekomstwerven voor de bestuurderskamers. Met deze eerste white paper willen we de bestuurderswereld wegwijs maken in de wereld van AI met een overzicht van de belangrijkste kenmerken, architectuur, mogelijkheden en risico's van deze technologie. In deze eerste paper gaan we bewust wat breder dan de strikte bestuursaspecten, aangezien de technologie zowel het economische speelveld als de maatschappij fundamenteel kan beïnvloeden. Het is namelijk van cruciaal belang dat de bestuurder begrijpt hoe breed de impact van AI kan zijn.

In twee volgende studies zullen we dieper ingaan op wat de technologie kan betekenen voor de strategische, monitoring en leiderschapsrol van bestuurders, en hoe de technologie de werking van de raad van bestuur zelf en de eventuele subcomités zou kunnen beïnvloeden.

Betrokken leden van het GUBERNA Knowledge Development Team:

Olivier Braet

Sander Berghmans

Regine Slagmulder

2. Wat is AI?

Om te beginnen is het nuttig om even na te gaan wat AI nu precies is. Mede door de hype die momenteel rond AI hangt zijn er een veelvoud van definities van AI in omloop. Niet alle definities zijn echter even correct en vaak zijn deze veel te breed. Regelmatig gebruiken technologiespelers het begrip voor systemen die geen enkele blijk geven van enige vorm van autonoom functioneren, omdat ze zo hun diensten sterker willen vermarkten bij klanten of promoten bij mogelijke investeerders. Het is dus nodig om eerst precies te definiëren vanaf welk punt een systeem valt onder de noemer van AI, en tot wanneer het nog een traditioneel IT-systeem is dat bijvoorbeeld enkel gebruik maakt van correlaties of regressies.

2.1. Artificiële Intelligentie: Definities

Academisch werd AI traditioneel gedefinieerd als: “alle technologie die menselijk gedrag vertoont”. Deze brede academische definitie heeft diepe wortels in de cognitieve, linguïstische en antropologische wetenschapstakken. De oorspronkelijke academische definitie van AI wordt momenteel echter overstemd door de software-industriële kant van het verhaal.¹ Aan de industriële zijde wordt AI voornamelijk gedefinieerd als een vorm van Machine Learning die gebruik maakt van grote hoeveelheden data en vervolgens op basis van algoritmen en data-analysetechnieken met een zekere mate van autonomie voorspellingen kan doen en beslissingen kan nemen. In de softwarewereld wordt AI daardoor voornamelijk gebruikt als een overkoepelende term voor allerlei vormen van algoritmische besluitvorming of predictie, neurale netwerken, automatische vertalingen, natuurlijke taalverwerking, kunstmatige agenten, en robotica met een zelflerende component.²

Ook de Europese Commissie poogde AI preciezer te definiëren. De Commissie hanteerde lang als werkdefinitie voor AI “systemen die intelligent gedrag vertonen door hun omgeving te analyseren en—met een zekere mate van autonomie—acties ondernemen om specifieke doelen te bereiken.”³ De meest recente tekst van de EU AI Act definieert AI als een “machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments.”⁴ Het overkoepelende aspect in deze definities is dat ze allen een zekere mate van **autonome werking** voorzien.

¹ Zysman, John & Nitzberg, Mark (2020) Governing AI: Understanding the Limits, Possibility, and Risks of AI in an Era of Intelligent Tools and Systems, BRIE Working Paper No. 2020-5, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3681088

² De Europese wetgever somt een lijst van concepten op die onder de paraplu-terminologie van AI vallen in Annex I van de EU AI Act: “Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.”

³ Commission (2018) Coordinated Plan on Artificial Intelligence. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic And Social Committee and the Committee of the Regions, Brussels, 7.12.2018, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795&from=DA>

⁴ EU AI ACT (2023) https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

2.2. Algemene kenmerken van Artificiële Intelligentie

Ondanks de verschillende (nuances in) definities van AI kunnen we een aantal algemene observaties maken over AI. Een eerste algemeen kenmerk is dat de verschillende vormen van AI kunnen gesitueerd worden in termen van **hoe goed het menselijk gedrag kan imiteren**. De meeste AI-toepassingen slagen hier vooralsnog niet in, en het is vaak onmiddellijk duidelijk dat een machine de taak heeft uitgevoerd. Om AI te identificeren die menselijke taken kunnen uitvoeren zonder dat het voor buitenstaanders duidelijk is dat een machine de taken heeft gedaan bestaan verscheidene tests. Er is de bekende *Turing test* waarbij AI mensen zodanig goed kan imiteren dat ze (minstens voor een moment) geloven dat de machinale output van een mens komt.⁵ Steve Wozniaks *koffietest* gaat verder. Deze test verwacht dat een machine een huis moet kunnen binnengaan en een kop koffie kan zetten.⁶ De *robotstudententest* van Goertzel vereist dat een machine zich kan inschrijven op een universiteit, dezelfde lessen volgt als mensen, en een diploma kan behalen.⁷ De *tewerkstellingstest* van Nilsson stelt een strikter economisch criterium voor waarbij AI-programma's jobs kunnen uitvoeren die normaal gesproken door mensen worden uitgevoerd.⁸ De meeste bestaande AI-toepassingen waren tot de brede introductie van generatieve AI zoals ChatGPT niet goed in menselijke imitatie. Sinds kort komen het behalen van bijvoorbeeld de robotstudententest of tewerkstellingstest in het vizier van AI.

Hoe goed AI een mens kan imiteren is eigenlijk een verkeerde verwachting. Een AI kan uiterst sterk zijn in de drie kernfuncties van optimaliseren, aanbevelen, en voorspellen/diagnosticeren, en toch zwak zijn qua menselijke imitatie. Niet-mimetische maar indrukwekkende AI bestaat al jaren in toepassingen zoals het optimaliseren van logistieke en toeleveringsketens, het maken van commerciële aanbevelingen op basis van persoonskenmerken, het beter diagnosticeren van ziektes, of het voorspellen van epidemieën. Bovendien is het de vraag of het imiteren van mensen op zich een relevant doel is. Mensen zijn per slot van rekening wezens die meestal minder snel kunnen rekenen dan machines, makkelijker slordigheidsfouten en denkfouten begaan, vooroordelen en vastgeroeste denkpatronen gebruiken, en een beperkt sensorisch arsenaal hebben qua geluids- of visuele detectie.

De nieuwste generatie van generatieve AI zoals ChatGPT kan in het model van Michael Hilb (2020⁹) in de figuur hieronder gesitueerd worden als een vorm van '**autonome intelligentie**'. Dit betekent dat deze systemen niet alleen de menselijke intelligentie kunnen begeleiden, ondersteunen of versterken, maar ook autonoom van menselijke input en beslissingen oplossingen kunnen bedenken. Om deze reden heeft deze nieuwe generatie van AI dan ook zo een grote impact. Voor het eerst kan AI relatief onafhankelijk van mensen beslissingen nemen. Het volgende niveau van zgn. **autopoeitische intelligentie** (waar AI op zichzelf nieuwe AI kan creëren) is tot nader order nog niet bereikt, alhoewel er al experimenten zijn deze autopoeisis beginnen benaderen. Belangrijk is om te beseffen dat we als

⁵ Turing, A.M. (1950) I.—COMPUTING MACHINERY AND INTELLIGENCE, *Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460, <https://doi.org/10.1093/mind/LIX.236.433>

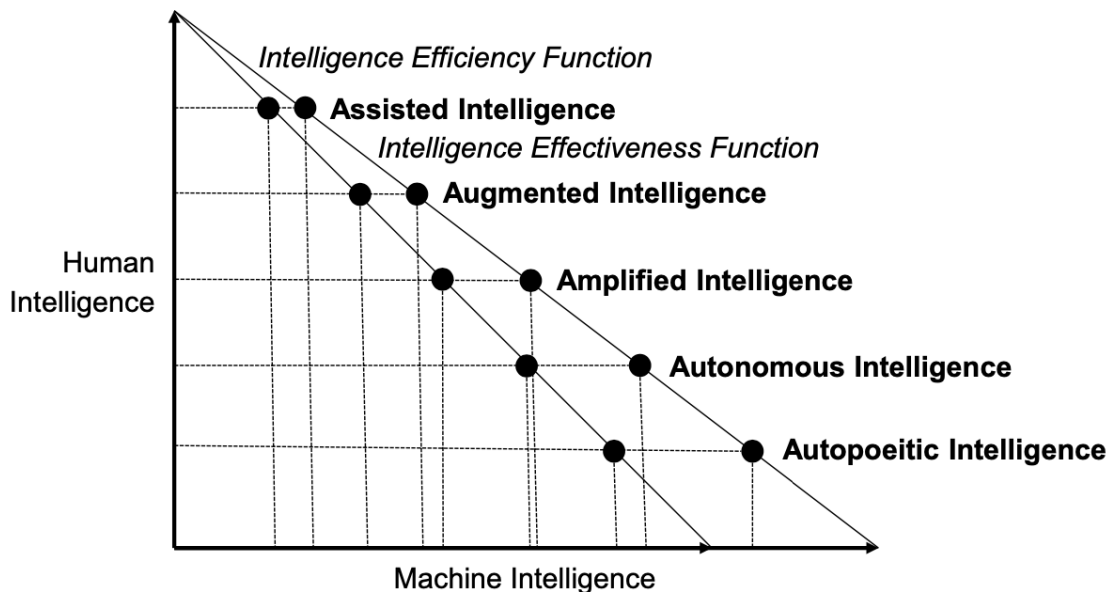
⁶ Wozniak, Steve in Shick, Michael (2010) Wozniak: Could a Computer Make a Cup of Coffee? *Fast Company*, 02 March 2010, <https://www.fastcompany.com/1568187/wozniak-could-computer-make-cup-coffee>

⁷ Goertzel, Ben (2014) Artificial general intelligence: concept, state of the art, and future prospects, *Journal of Artificial General Intelligence*, 5(1) 1-46, <https://sciendoc.com/downloadpdf/journals/jagi/5/1/article-p1.pdf>

⁸ Nilsson, Nils J. (2005) Human-Level Artificial Intelligence? Be Serious! *AI Magazine*, 25th Anniversary Issue, 68-75, <https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag26-04-HLAI.pdf>

⁹ Hilb, Michael (2020) Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance. *Journal of Management and Governance*, 24: 851–870, <https://doi.org/10.1007/s10997-020-09519-9>

mens over het denkproces van de AI—voor zover we over ‘denken’ kunnen spreken¹⁰—steeds minder controle en begrip zullen hebben des te dichter we bij autopoeitische intelligentie komen.



Figuur 1: Vormen van machine intelligentie (Hilb 2020)

Hoewel AI in deze dus een zeer snelle evolutie heeft gekend en steeds autonomer werkt, blijft het tot nader order zeer twijfelachtig of AI in de (nabije) toekomst kan evolueren naar mensachtige vormen van waarneming, bewustzijn of gevoel, een zogenaamde “**Artificial General Intelligence**” of AGI (kunstmatige algemene intelligentie). Aangezien zelfs neurowetenschappers er nog niet uit zijn wat menselijk bewustzijn precies is, is het onmogelijk om ook AGI precies te definiëren. OpenAI publiceerde een eigen blauwdruk voor “Governance of Superintelligence”¹¹, maar dat document focust vooral op hypothetische en onmeetbare toekomstige existentiële risico’s en lijkt de discussie te willen afleiden naar ondefinieerbare concepten om te vermijden dat er een debat wordt gevoerd over de huidige risico’s en tekortkomingen van de technologie. Het blijft namelijk zo dat AI—zelfs al werkt deze vrij autonoom—een aantal beperkingen zal hebben waar de kritische gebruiker alert voor moet blijven, zeker wat betreft de ethische component.¹²

Een volgende algemene observatie is dat AI, zeker in combinatie met andere technologieën zoals big data en Internet of Things (IoT) een zogenaamde “**General Purpose Technology**” (GPT) kan zijn.¹³ Historische voorbeelden van introductie van GPT's zijn de stoommachine, elektriciteit, of informatietechnologie. Net als toen kan AI nieuwe vormen van productie, consumptie, en mens-

¹⁰ Sommige auteurs begaan de fout van antropomorfisme, i.e. het toeschrijven van menselijke karakteristieken aan, in dit geval, softwarecomponenten. Dit gebeurt onder andere als men spreekt van het ‘denken’ van AI, of het ‘hallucineren’ van ChatGPT wanneer die onbestaande bronnen of referenties verzint. Hallucineren is een term die beter gereserveerd blijft voor door menselijke breinen gecreëerde waanbeelden.

¹¹ Altman, Sam; Brockman, Greg and Sutskever, Ilya (2023) Governance of superintelligence, 22 May 2023, <https://openai.com/blog/governance-of-superintelligence>

¹² Whittlestone, J., Nyrupe, R., Alexandrova, A., & Cave, S. (2019) The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Jan 2019, pp. 195-200), <https://doi.org/10.1145/3306618.3314289>

¹³ Goldfarb, A., Taska, B., and Teodoridis, F. (2023) Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. Research Policy, 52(1): 104653, <http://dx.doi.org/10.1016/j.respol.2022.104653>

machine interacties mogelijk maken, kan het gevestigde bedrijfsmodellen op losse schroeven zetten, en zo de structuur van ganse sectoren veranderen. Naast de productiefactoren van arbeid en kapitaal kan het een nieuwe productiefactor worden die de totale factorproductiviteit en economische groei van een land kan beïnvloeden (alhoewel dit effect momenteel nog niet opduikt in statistieken). Bovendien kunnen AI-toepassingen ook een bredere impact hebben dan enkel een strikt-economische. Omdat AI een hefboom kan vormen voor menselijke intelligentie, zou het kwaliteitsonderwijs kunnen aanbieden op een grotere schaal, een betere gezondheidszorg verlenen aan een grotere groep mensen, of nieuwe mogelijkheden tot een meer duurzame omgang met schaarse grondstoffen mogelijk maken.

Tenslotte kunnen we stellen dat de output van de nieuwste generatie van AI zeer **onvoorspelbaar is**. Aan de positieve zijde betekent dit dat AI uiterst verrassende oplossingen kan vinden. Aan de schaduwzijde zijn toepassingen denkbaar die ongeanticipeerde destructieve effecten hebben. Het is daarom belangrijk om deze evoluties binnen de bedrijven en in de maatschappij in het algemeen nauwgezet op te volgen en steeds een zekere controle te behouden over de beslissingen die AI neemt. Daarvoor zal de raad van bestuur binnen organisaties een belangrijke monitorrol op zich moeten nemen.

2.3. Artificiële Intelligentie: Technische architectuur

AI omvat een specifieke technologische architectuur, en het is belangrijk dat bestuurders een basisinzicht hebben in de verschillende componenten om de mogelijke opportuniteiten en risico's beter te kunnen inschatten. Op AI gebaseerde systemen kunnen zuiver softwarematig zijn (bv. spraakassistenten, beeldanalysesoftware, zoekmachines, spraak- en gezichtsherkenningssystemen) of AI kan zijn ingebouwd in hardware apparatuur (bv. geavanceerde robots, autonome auto's, drones).¹⁴ Die laatste reeks hybride toepassingen illustreren goed hoe AI nooit op zichzelf staat maar deel uitmaakt van een groter IT-systeem, waarin AI voorkomt als een complementaire technologie.¹⁵

AI hangt af van een typische **gelaagde technische IT-architectuur**. De performantie van een AI-dienst is maar zo sterk als de lagen waarboven het draait. Aan de basis liggen hardwarecomponenten die de nodige rekenkracht, gegevensopslag en connectiviteit leveren. Boven op die hardware lopen ontwikkelingstools waarvan AI-ontwikkelaars gebruik maken voor het ontwikkelen van nieuwe algoritmen zoals software die nodig is de verwerking van gegevens, voor de interpretatie van die gegevens door middel van "data libraries", voor diverse soorten van algoritmische verwerking en voorbereidend werk, en voor interoperabiliteit met andere systemen. Op de bovenste laag situeren zich de diensten die gericht zijn naar de (professionele of particuliere) eindgebruikers van de AI-software. Bijvoorbeeld, bij generatieve AI zoals ChatGPT bestaat de hardware laag voor het trainen van het zgn. "foundational model" GPT-4¹⁶ naar schatting uit 25.000 Nvidia A100 GPU's (Graphics

¹⁴ EU HLEG on AI (High Level Expert Group on Artificial Intelligence) (2018), A definition of AI: Main capabilities and scientific disciplines, 18 December 2018,

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651

¹⁵ Shapiro, Carl & Varian, Hal R. (1998), *Information Rules*, Harvard Business School Press, Boston: Massachusetts.

¹⁶ "Foundational model" is een concept van het Stanford Center for Research on Foundation Models (CRFM) (2021) On the Opportunities and Risks of Foundation Models, Stanford University, 16 August 2021, <https://fsi.stanford.edu/publication/opportunities-and-risks-foundation-models>

Processing Units) en een dataset van 45 terabyte aan voorbeeldteksten en -afbeeldingen. De einddienst ChatGPT draait momenteel bovenop GPT-4.¹⁷

Een andere goede illustratie van deze gelaagde architectuur is de zelfrijdende wagen. Aan de basis van het systeem voor zelfrijdende wagens ligt allerlei hardware zoals sensoren en camera's. Daarbovenop loopt een software besturingssysteem dat de kernfuncties van de wagen monitort, naast meer specifieke software voor ad-hoc functies zoals parkeerhulp, noodremmen, waarschuwing voor aanrijding van achteren, dodehoekdetectie, enzovoort. De AI-software van zelfrijdende wagens bestaat uit zelflerende componenten die gebruik maken van alle hardware en data aanwezig in de wagen en buiten de wagen zoals verkeersinformatie, digitale wegenkaarten, sensoren ingebed in het wegdek, data-uitwisseling met andere voertuigen, enzovoort. Op basis van al deze informatie berekent de AI een route van start tot eindpunt, en reageert het dynamisch op veranderende omstandigheden (slecht weer, ander verkeer, wegonderbrekingen) om haar bestemming te bereiken. Hier loopt de AI vaak tegen de muren van menselijke adoptie aan. Zo zal er een langzame adoptiecurve van zelfrijdende wagens zijn waar deze zich nog lang in het verkeer zullen moeten begeven naast de analoge weggebruikers. Zelfs een hypothetisch superkrachtige AI die elke verkeersregel feilloos kan volgen maar de straat moet delen met onvoorspelbare mensen van vlees en bloed zal met situaties geconfronteerd worden waar de AI niet tijdig kan reageren, precies omdat mensen verkeersregels soms opzettelijk negeren om pragmatische redenen, soms opzettelijk overtreden omdat ze gehaast zijn, of soms onopzettelijk niet volgen omdat ze bv. een regel zijn vergeten of nooit hebben geleerd. Steden die al experimenteren met robotaxi's lopen tegen deze beperkingen aan.¹⁸

Omdat de componenten van AI altijd gesitueerd zijn in en bovenop verscheidene hard- en softwarelagen en tegelijk moeten laveren tussen onvoorspelbare mensen heeft dit gevolgen voor de regulering vanuit de overheid en het goed bestuur ervan door een onderneming. Zowel de strategische mogelijkheden als de mogelijke bronnen van risico's en problemen kunnen zich situeren op de verscheidene technische en menselijke niveaus. Generatieve AI is maar zo goed als de datasets waarmee de algoritmen getraind worden, en het soort data-analyse dat gebruikt wordt. Op alle niveaus zal men moeten denken aan cybersecurity om de integriteit van het systeem te bewaken, en op welke manieren menselijke factoren vertekeningen veroorzaken in de training of de interpretatie van de AI.

2.4. AI-toepassingen

AI kent reeds enorm veel toepassingen, zeker wat betreft de niet-mimetische AI die razendsnel datapatronen identificeert, analyseert, en optimaliseert. Bij **online media** zijn er het zoekalgoritme van Google, algoritmen voor het personaliseren van digitale reclame of consumptieaanbevelingen bij Amazon, Netflix of TikTok, en de nieuwsfeed van Meta's Facebook of Instagram. Machine Learning wordt ook toegepast in aspecten van **logistiek management**, **juridische analyse**, **financiële analyse**, de **prijsbepaling van producten** en **geautomatiseerde beurstransacties**. Bergen et al (2021¹⁹) bespreekt in die context hoe ondernemingen AI kunnen gebruiken om te vermijden dat hun

¹⁷ Wie geïnteresseerd is in een iets technischer introductie over de verrassend eenvoudige kernprincipes van ChatGPT kan "Let's build GPT: from scratch, in code, spelled out" van software ingenieur Andrej Karpathy bekijken op <https://www.youtube.com/watch?v=kCc8FmEb1nY>

¹⁸ Muller, Joann (2023) Robotaxis hit the accelerator in growing list of cities nationwide, Axios, 29 August 2023, <https://www.axios.com/2023/08/29/cities-testing-self-driving-driverless-taxis-robotaxi-waymo>

¹⁹ Bergen, Mark E.; Dutta, Shantanu; Guszczka, James and Zbaracki, Mark J. (2012) How AI Can Help Companies Set Prices More Ethically, Harvard Business Review, 26 March 2021, <https://hbr.org/2021/03/how-ai-can-help-companies-set-prices-more-ethically>

prijzsetting onethische effecten heeft. Het debat rond in welke mate AI automatische prijsvariaties mag introduceren voor verschillende klantengroepen of individuele klanten—een bedrijfspraktijk die in vele gevallen wettelijk is en al eeuwen wordt toegepast—belooft nog pittig te worden.

In de **gezondheidszorg** bestaan er AI-diensten rond persoonlijke controleapparatuur en mobiele apps, elektronische gezondheidsdossiers, of chirurgische robots die helpen bij complexe medische procedures. In de **entertainment-industrie** investeert men in generatieve AI zoals *DALL-E*, *Stable Diffusion* of *Midjourney* voor het automatisch genereren van afbeeldingen. In het **onderwijs** wordt vooruitgang verwacht op het gebied van online leren, spraakgestuurde chatbots en meer interactieve artificiële leermodules.

Daarnaast wordt AI ook steeds vaker, en niet zonder kritiek, ingezet voor **politietaken** door middel van gelaatsherkenning of data-analyse. Er bestaan studies die AI (meer bepaald Neurale Netwerken) loslaten op een database van personen die wel of niet veroordeeld zijn voor criminele feiten, en vervolgens het algoritme laten leren om op basis van gelaatskenmerken de kans op criminaliteit te herkennen. Het gebruik van neurale netwerken voor het identificeren van gelaatskenmerken van criminelen wordt echter geplaagd door te veel contextuele factoren die de methodologie onbetrouwbaar maakt (voor een technische discussie, zie Bowyer et al 2020²⁰). Zo kan het zijn dat bepaalde bevolkingsgroepen meer geïdentificeerd worden door een overheid en daardoor hoger scoren qua aantal veroordeelden. Dit is zeker het geval bij autoritaire overheden die een rebellerende regio wensen te disciplineren. Het algoritme zal een verband vinden tussen etnische gelaatskenmerken en criminaliteit, maar de oorzaak van de hogere criminaliteit zijn dan de politieke beslissingen, en niet de gelaatskenmerken op zich.

Een recente enquête (2021) van FOD economie indiceert dat in België AI haar weg heeft gevonden in 41% van de grote bedrijven²¹, 17 % van de middelgrote, 8% van de kleine en 4,5 % van de micro-ondernemingen.²² De meest populaire toepassingen zijn tekstanalyse en -productie, en beeldherkenning en -analyse. AI4Belgium toont een overzicht van de belangrijkste lokale Belgische softwareontwikkelaars en de verscheidene Belgische sectoren waar AI al wordt toegepast, met name de gezondheidszorg, landbouw, energie, juridische sector, financiële sector, onderwijs, maakindustrie, mobiliteit en telecommunicatie.²³

AI kent dus een veelheid aan toepassingen, het zou ons te ver brengen om deze allemaal specifiek te bespreken. Feit is dat AI reeds aanwezig is in onze maatschappij en bedrijven, en dat deze in de toekomst verdere sprongen vooruit zal maken.

²⁰ Bowyer, Kevin; King, Michael; Scheirer; Walter & Vangara, Kushal (2020) The “Criminality From Face” Illusion. IEEE Transactions on Technology and Society, Vol. 1, issue 4, <https://ieeexplore.ieee.org/document/9233349> en open access versie op <https://arxiv.org/abs/2006.03895>

²¹ FOD Economie (2021) Barometer van de informatiemaatschappij: Geavanceerde digitale technieken, <https://economie.fgov.be/nl/themas/online/ict-belgie/barometer-van-de/de-ondernemingen-en-het/de-ondernemingen-en-de>

²² FOD Economie (2021b) Digitalisering van kmo's: Artificiële intelligentie (AI) en andere geavanceerde digitale technologieën, <https://economie.fgov.be/nl/themas/ondernemingen/kmos-en-zelfstandigen-cijfers/digitalisering-van-kmos/artificiele-intelligentie-en>

²³ AI4Belgium, Industrieel landschap, <https://ai4belgium.be/nl/ai-landschap/>

3. Positieve en negatieve impact van AI

Hoewel eerdere vormen van AI reeds gebruikt werden tijdens de 20^{ste} eeuw, zijn er goede redenen om aan te nemen dat ze de beloftes van een brede economische en maatschappelijke impact sterker zal kunnen waarmaken in deze eeuw. Zeker sinds de opkomst van sociale media platforms die intensief gebruik maken van algoritmische aanbevelingen op basis van bezoekersprofielen kent het domein een forse toename van investeringen in en kritiek op de technologie. Het publiek openstellen van de ChatGPT interface door OpenAI eind 2022 introduceerde het gebruik van generatieve AI bij het brede publiek, met een enorme impact op de beurswaardering van alle AI-gerelateerde ondernemingen op alle technologische lagen, van producenten van gespecialiseerde grafische kaarten zoals Nvidia tot een softwarespeler als Microsoft die OpenAI's ChatGPT deels integreerde in haar Bing zoekmachine en met Azure een belangrijke cloud infrastructuurspeler is in de business-to-business markt. Het feit dat AI deze doorbraken kent betekent dat bestuurders moeten beseffen welke positieve, maar ook negatieve impact de doorbraak van AI kan hebben.

3.1. Mogelijke positieve impact van Artificiële Intelligentie

Het grootste en meest voor de hand liggende voordeel van de massale adoptie van AI in onze economie is uiteraard haar **hogere efficiëntie en sterkere productiviteit**. Rapporten van consultants als o.a. Accenture/Frontier Economics²⁴ of PwC²⁵ voorspelden enkele jaren geleden al een sterk positieve impact van AI op de totale factorproductiviteit en groei van nationale economieën. Chen et al. (2016²⁶) schatten de cumulatieve economische impact van AI van 2016 tot 2026 in op \$1,5 tot \$3 biljoen (0,15 tot 0,3% van het mondiale bbp). Recent verhoogde McKinsey haar inschatting van impact tot tussen de \$2.6 biljoen tot \$4.4 biljoen per jaar.²⁷ Academische publicaties waren voorzichtiger in hun projecties. Brynjolfsson, Rock en Syverson²⁸ merkten in 2017 nog op dat de economische impact van AI zoals Machine Learning nog niet kon worden waargenomen. Furman en Seamans (2018²⁹) stelden dat AI en robotica weliswaar het potentieel hebben om de productiviteit te verhogen, maar dat ze op de korte termijn enkel gemengde effecten zien op de productiviteit en werkgelegenheid. Een sterkere groei kan nog niet eenduidig geobserveerd worden omdat teveel andere factoren (in het bijzonder

²⁴ Purdy, M. and Dougherty, P. (2017), How AI boosts industry profits and innovation. Accenture Research, <https://www.accenture.com/fr-fr/acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf>

²⁵ PwC (2017) Sizing the prize - What's the real value of AI for your business and how can you capitalise? <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>

²⁶ Chen, Nicholas; Christensen, Lau; Gallagher, Kevin; Mate, Rosamond & Rafert, Greg (2016), Global economic impacts associated with artificial intelligence, Analysis Group, https://www.analysisgroup.com/globalassets/content/insights/publishing/ag_full_report_economic_impact_of_ai.pdf

²⁷ McKinsey (2023) The economic potential of generative AI: The next productivity frontier, 14 June 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights>

²⁸ Brynjolfsson, Erik; Rock, Daniel and Syverson, Chad et al. (2017) Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, pp. 23 – 57 in Ajay Agrawal, Joshua Gans, and Avi Goldfarb (eds.) The Economics of Artificial Intelligence: An Agenda, <https://www.nber.org/system/files/chapters/c14007/c14007.pdf>

²⁹ Furman, Jason and Seamans, Robert (2018), AI and the Economy, NBER Working Paper No. 24689, NBER, Cambridge, MA, https://www.nber.org/system/files/working_papers/w24689/w24689.pdf

onverwachte internationale gebeurtenissen als Covid-19 of de oorlog in Oekraïne) een geïsoleerde identificatie van de macro-economische impact van AI onmogelijk maken.

De eerste positieve effecten van AI (in de brede betekenis) op productiviteit werden waargenomen binnen de wereld van robotica, in de fysieke productiesector. Graetz en Michaels (2017³⁰, 2018³¹) schatten in een steekproef van 17 landen dat het gebruik van industriële robots in de periode 1993-2007 de arbeidsproductiviteit in die specifieke landen met ongeveer 0,37 procentpunten deed toenemen per jaar. Een deel van die productiviteitswinst valt toe te schrijven aan gebruik van verfijnder AI. Ander onderzoek bevestigde dat het gebruik van geavanceerde robots de totale factorproductiviteit en de gemiddelde lonen per gewerkt uur verhoogt.³² Er is wel een verdringingseffect van arbeid naar kapitaal, waardoor er netto wel banen verloren gaan in specifieke arbeidscategorieën. Graetz en Michaels observeerden in hun sectorale panels dat robots het aandeel van door laaggeschoolde werknemers gepresteerde uren vermindert relatief aan hogergeschoolde werknemers (een bevestiging van Autor et al. 2006³³ en Van Reenen 2011³⁴). Ook recentere studies tonen aan dat automatiseringstechnologieën leiden tot loonongelijkheid door een relatieve daling van de lonen bij werknemers gespecialiseerd in routinetaken.³⁵ In welke mate de productiviteitswinsten van AI in de productiesectoren zullen overvloeien naar dienstensectoren is nog uiterst onvoorspelbaar.

Recent schreef OpenAI mee aan een studie die inschat dat 80% van de jobs voor 10% bestaat uit taken die generatieve AI zou kunnen automatiseren³⁶. Bij 19% van de werknemers zou dit zelfs tot 50% van hun takenpakket kunnen beïnvloeden. Deze schattingen moeten met de nodige scepsis gelezen worden aangezien de auteurs verbonden zijn aan een ontwikkelaar van AI. Zelfs jobs die in principe nu al 100% kunnen geautomatiseerd worden, zoals croupiers in casino's aan roulettes of pokertafels, blijven bestaan alhoewel een geautomatiseerd systeem elke vorm van menselijke fout of bedrog zou elimineren. De menselijke interactie biedt blijkbaar een toegevoegde waarde aan de spelers die een volledige automatisering van de job van croupier afremt. Bij uitbreiding zullen vele andere jobs die schijnbaar volledig automatiseerbaar zijn toch deels door mensen worden uitgevoerd.

Het mag dus duidelijk zijn dat, gezien de grootteorde van de potentiële impact van AI op de interne bedrijfsprocessen en de omgeving van het bedrijf, raden van bestuur zullen moeten inschatten welke positieve invloeden AI—en dan vooral verhoogde efficiëntie—kan hebben op hun onderneming. Ook in de bestuurstaken zelf zal AI naar verwachting een deel van de taken kunnen automatiseren of ondersteunen. Daarbij wordt dan gekeken naar taken die repetitief zijn of die de verwerking van grote hoeveelheden informatie vereisen (bijvoorbeeld onderbouwing van strategische keuzes met cijfergegevens).

³⁰ Graetz, Georg and Michaels, Guy (2017), Is Modern Technology Responsible for Jobless Recoveries?, *American Economic Review*, 107(5), 168–73, <https://www.aeaweb.org/articles?id=10.1257/aer.p20171100>

³¹ Graetz, Georg and Michaels, Guy (2018), Robots at Work, *Review of Economics and Statistics*, Vol. 100, 753–768, <https://direct.mit.edu/rest/article-abstract/100/5/753/58489/Robots-at-Work?redirectedFrom=fulltext>

³² Petropoulos, Georgios; Pichler, David & Chiacchio, Francesco (2018) The impact of industrial robots on EU employment and wages: A local labour market approach, Bruegel Working Paper, 18 April 2018, Bruegel, Brussels, <https://www.bruegel.org/working-paper/impact-industrial-robots-eu-employment-and-wages-local-labour-market-approach>

³³ Autor, David H., Katz, Lawrence F., and Kearney, Melissa S. (2006) The polarization of the us labor market. *American Economic Review*, 96(2): 189–194, <http://dx.doi.org/10.1257/000282806777212620>

³⁴ Van Reenen, J. (2011) Wage inequality, technology and trade: 21st century evidence. *Labour Economics*, 18(6):730–741, <https://doi.org/10.1016/j.labeco.2011.05.006>

³⁵ Acemoglu, D. and Restrepo, P. (2022) Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5):1973–2016, <http://dx.doi.org/10.3982/ECTA19815>

³⁶ Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023) GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, <https://arxiv.org/abs/2303.10130>

3.2. Mogelijke misbruiken van Artificiële Intelligentie

Tegenover de vele voordelen van AI staan uiteraard een aantal belangrijke risico's, hindernissen en problemen. Vanaf het begin van de ontwikkeling van AI kent deze tegenkanting omwille van potentiële misbruiken en tekortkomingen. Regelmatig wordt er daarom door toezichthouders of externe experts op de rem gestaan. Een eerste AI-winter vond plaats in de jaren zeventig van de vorige eeuw, na de publicatie van het kritische Automatic Language Processing Advisory Committee (ALPAC) rapport van de Amerikaanse regering in 1966³⁷, waarna de investeringen in de VS vertraagden. Het kritische Lighthill-rapport van de Britse regering in 1973³⁸ had een remmende impact op Europese investeringen. Ook nu wordt voorbehoud gemaakt bij de reuzenstappen die AI maakt, zonder dat de nodige mechanismen en procedures aanwezig zijn om in te grijpen wanneer zaken zouden verkeerd lopen. De technologie kan, indien slecht aangewend of ontworpen, fundamentele risico's inhouden. In het meest extreme voorbeeld merken Geist en Lohn (2018³⁹) op dat AI catastrofale gebeurtenissen zoals een kernoorlog waarschijnlijker maken. Het is daarom belangrijk om mogelijke risico's in kaart te brengen met het zicht op het toekomstig AI-bestuur. Er kan een eerste belangrijk onderscheid worden gemaakt tussen opzettelijk destructief gebruik van AI, en onopzettelijk berokkende schade door het gebruik van AI.

3.2.1. Opzettelijk kwaadwillig gebruik van Artificiële Intelligentie

Sommige van de risico's als gevolg van een **kwaadwillig gebruik van AI** zijn uitlopers van reeds bestaande risico's van andere technologische diensten. Nepnieuws zal meer gebruik beginnen maken van de geautomatiseerde manipulatie van de publieke opinie via "deep fakes" en andere nieuwe vormen van desinformatie. AI kan real-time "fact checking" mogelijk maken maar tegelijk de productie van desinformatie goedkoper maken. Phishing-zwandel kan geraffineerder worden, onder andere door gebruik van synthetisch aangemaakte stem-imitaties die de persoon aan de andere kant van de lijn ervan overtuigen met een persoon te praten die ze kennen. Regeringen zullen keuzes moeten maken om de verspreiding van verantwoorde, ethische en betrouwbare AI te stimuleren, en om wereldwijde AI-governance regels voor mediaproductie en -consumptie te creëren. Een eerste maatregel is de eindgebruikers te waarschuwen wanneer er AI wordt gebruikt voor een bepaalde dienst (dit is de zgn. "shibboleth regel"⁴⁰), en hen tegelijk helpen een onderscheid te maken tussen door ethisch verantwoorde AI ondersteunde diensten en minder ethische diensten.

Op internationaal geopolitiek vlak wijst Tegmark⁴¹ op de bezorgdheid dat de grootste winnaars van een AI-wapenwedloop "kleine schurkenstaten en niet-statelijke actoren zoals terroristen" zullen zijn die via de zwarte markt toegang kunnen krijgen tot nieuwe wapensystemen die op de reguliere markt niet verkrijgbaar zijn. Tegmark vreest bijvoorbeeld dat kleine AI-moorddrones door massaproductie waarschijnlijk niet veel meer zullen kosten dan een smartphone. Als misdadigers zich daarnaast

³⁷ Pierce, John R.; Carroll, John B. et al. (1966) Language and Machines — Computers in Translation and Linguistics. ALPAC report, National Academy of Sciences, National Research Council, Washington, DC, <https://web.archive.org/web/20110409070141/http://www.mt-archive.info/ALPAC-1966.pdf>

³⁸ Lighthill, James (1973) Artificial Intelligence: A General Survey, in Artificial Intelligence: A paper symposium. UK: Science Research Council, http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm

³⁹ Geist, Edward & Andrew, John (2018) How Might Artificial Intelligence Affect the Risk of Nuclear War? RAND Corporation, Pages: 28, <https://doi.org/10.7249/PE296>

⁴⁰ HAI project (2021) The Shibboleth rule for Artificial Agents, Stanford University, Human-Centered Artificial Intelligence (HAI), Aug 20 2021, <https://hai.stanford.edu/news/shibboleth-rule-artificial-agents>

⁴¹ Tegmark, Max (2017) Life 3.0, New York City: Knopf.

zouden richten op met het internet geconnecteerde “wearables” of implantaten, zou ook het risico van fysieke hacking kunnen ontstaan. Deze risico’s vragen om een betere planning voor cyberbeveiliging en cyberweerbaarheid. Dit brengt echter nieuwe kosten met zich mee, wat doorweegt in de vereiste investeringen bij bedrijven die zich tegen deze risico’s willen wapenen, of kan doorgerekend worden in een hogere risicopremie bij verzekeringen tegen deze risico’s.

Voorstanders van AI stellen dat er voor elk ‘offensief’ gebruik van AI voor kwaadwillige doeleinden een even sterke case bestaat voor het inzetten van dezelfde AI voor ‘defensieve’ doeleinden tegen deze risico’s. Zo kan het risico van cyberaanvallen worden verlaagd door het schaalbare gebruik van AI-systemen om defensieve taken uit te voeren waarvoor voorheen veel menselijke arbeid, intelligentie en domeinexpertise nodig was. AI vormt zo niet enkel een mogelijke bron maar ook de mogelijke remedie voor een aantal belangrijke risico's. Vooral vanuit het oogpunt van voornoemde cybersecurity, maar ook met de bredere risicoanalyse zullen veel raden van bestuur geconfronteerd worden.

3.2.2. Onbedoelde negatieve gevolgen van Artificiële intelligentie

Naast de gevolgen die voortvloeien uit kwaadwillig gebruik van AI, zijn er ook negatieve gevolgen die onbedoeld voortvloeien uit het gebruik van AI. Ten eerste zijn er **negatieve gevolgen op het aantal menselijke jobs**. Het potentieel tot de massale automatisering van allerlei arbeidstaken kan delen van de bevolking permanent werkloos maken of hun toegang tot de arbeidsmarkt bemoeilijken als ze niet over de nodige vaardigheden beschikken. In een bevraging door de OESO vreest 60% van de werknemers anno 2023 dat hun job op termijn zal verdwijnen door AI.⁴² Sommige regeringen overwogen hiertegen aanvankelijk universele basisinkomensregelingen, robotbelastingen of soortgelijke beleidsmaatregelen.⁴³ Het is onduidelijk of het individuele welzijn (waar het hebben van werk op zich een positief effect heeft, zie Stam et al. 2015⁴⁴) en de sociale cohesie door dergelijke initiatieven wezenlijk worden geholpen. Het debat is nu geleidelijk verschoven naar een reflectie over hoe de beroepsbevolking kan worden omgeschoold. De “Global Commission on the Future of Work” van de International Labour Organization (ILO)—het tripartite VN-agentschap van werkgevers, werknemers en overheden—deed vanuit die optiek een oproep tot een universele arbeidsgarantie, een akkoord over universele principes over levenslang leren, en de invoering van een effectief recht op levenslang leren om deze dreigingen het hoofd te bieden.⁴⁵

De automatisering van bepaalde subtaken door middel van AI kan de **relatie tussen werkgevers en medewerkers** onder druk zetten. Een recente staking van de *Writers Guild of America* had in september 2023 als uitkomst dat AI niet mag worden gebruikt om scripts te schrijven of te herschrijven, dat studio’s moeten bekendmaken of materiaal dat aan schrijvers wordt doorgegeven door AI is gegenereerd, en dat schrijvers hun expliciete toestemming moeten geven voor het gebruik van hun scripts om AI te trainen.⁴⁶ Dit debat maakt onderdeel uit van een veel bredere discussie over wie de vruchten van een AI-systeem mag plukken: enkel de eigenaar van het AI-systeem, of ook de werknemers of eindgebruikers met behulp van wiens input de AI-datasets getraind en verfijnd zijn.

⁴² OECD Employment Outlook (2023) Artificial intelligence and jobs. An urgent need to act, <https://www.oecd.org/employment-outlook/2023/>

⁴³ Furman, Jason and Seamans, Robert (2018), AI and the Economy, NBER Working Paper No. 24689, NBER, Cambridge, MA, https://www.nber.org/system/files/working_papers/w24689/w24689.pdf

⁴⁴ Stam, K. et al. (2015), “Employment status and subjective well-being: The role of the social norm to work”, *Work, Employment and Society*, 1–25, <https://journals.sagepub.com/doi/10.1177/0950017014564602>

⁴⁵ International Labour Organization (2019) Work for a Brighter Future, Report, 22 January 2019, https://www.ilo.org/global/publications/books/WCMS_662410/lang--en/index.htm

⁴⁶ Bedingfield, Will (2023) Hollywood Writers Reached an AI Deal That Will Rewrite History, Wired, 27 September 2023, <https://www.wired.com/story/us-writers-strike-ai-provisions-precedents/>

Bestuurders zullen moeten rekening houden met de manier waarop AI onderwerp wordt van bepaalde sociale onderhandelingen.

Andere onbedoelde negatieve gevolgen bij het gebruik van AI kunnen zich voordoen wanneer een bepaalde **beslissingsmacht wordt gedelegeerd** aan een AI-systeem, bv. de machtiging tot het maken van automatische aankopen/verkoop of het geven van geautomatiseerde antwoorden aan klanten. Er zullen financiële limieten moeten worden ingebouwd om te vermijden dat AI, in een race om bepaalde logistieke aan- of verkoopprocessen te optimaliseren, geen ondoordachte of overbodige handelingen uitvoert. De bestuurder moet alert zijn voor reputatieschade indien een deel van de klantencommunicatie geautomatiseerd wordt en men niet monitort of er geen contraproductieve boodschappen tussen zitten.

Het probleem van een **discriminerende vermarkting van AI** kan opduiken wanneer achtergestelde delen van een bevolking enkel nog worden bediend door “junk AI”. Aan armere mensen/bedrijven kunnen goedkopere AI-versies van diensten door huisartsen, rechters, verzekeringsmakelaars enzovoort worden aangeboden, wat discriminaties of uitsluitingen kan veroorzaken als deze diensten minder kwaliteitsvol zijn. Een kernprincipe van diensten met openbaar nut (zoals gezondheidszorg of basisonderwijs) zou nochtans hun universele toegankelijkheid van het volledige aanbod moeten zijn. Ongelijkheid reikt verder dan louter financiële uitkomsten. Ook ongelijke toegang tot onderwijs, tot het politieke of publieke leven, of tot kwaliteitsvolle openbare of semi-private diensten kunnen een ontwrichtend effect hebben. Zeker bestuurders van publieke organisaties zullen het gebruik van AI tegen het licht van het maatschappelijk doel van hun organisatie moeten houden. Bij uitbreiding kan ook een discriminerende toegang tot private diensten perfide maatschappelijke gevolgen hebben. De bestuurder moet kritisch nadenken over product- en prijsvariaties in het aanbod van de organisatie.

Tenslotte kan ook **het design zelf van AI** onbedoeld schade toebrengen aan de samenleving en individuen. Zo kan het gebruik van AI leiden tot het versterken van ongewenste vooroordelen, waardoor grondrechten worden geschonden en/of effecten plaatsvinden die als onrechtvaardig worden ervaren. De vraag werpt zich onmiddellijk op wat de definitie van “een ongewenst vooroordeel” is. Onze samenlevingen kennen verscheidene sociale en historische vooroordelen. Rijkere mensen kunnen in rechtbanken hogere vergoedingen voor lichamelijk letsel verkrijgen als de schadevergoedingen gebaseerd zijn op gederfde inkomsten. Vrouwen worden in sommige landen en sectoren nog steeds minder betaald dan mannen, zelfs na controle op hun diploma, anciënniteit en ervaring, of stoten moeilijker door tot de hoogste (en best betaalde) posities (zie Iwasaki & Satogami 2023 voor een meta-analyse van 53 onderzoeken⁴⁷). Anderzijds ontvangen ze vaak beduidend lagere straffen dan mannen voor dezelfde misdrijven.⁴⁸ Het gebruik van big data en predictiealgoritmen in politiewerk leidde in een aantal steden tot bezorgdheid over raciale/etnische vooroordelen (O’Neil 2016⁴⁹, Ferguson 2017⁵⁰). Ook in de bedrijfswereld kent men reeds voorbeelden waarbij het design van AI voor problemen zorgde. Google werd er in het verleden van beschuldigd advertenties voor leidinggevende banen vaker te tonen aan wat hun algoritmen herkennen als blanke mannen dan aan

⁴⁷ Iwasaki, I., Satogami, M. (2023) Gender wage gap in European emerging markets: a meta-analytic perspective, *Journal for Labour Market Research*, Vol. 57, Art. 9, <https://doi.org/10.1186/s12651-023-00333-y>

⁴⁸ US Sentencing Commission (2017) Demographic Differences in Sentencing, November 2017, <https://www.ussc.gov/research/research-reports/demographic-differences-sentencing>

⁴⁹ O’Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*, New York, NY: Crown Publishers.

⁵⁰ Ferguson, A.G. (2017) *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York, NY: NYU Press.

(Afro-Amerikaanse) mannen of vrouwen.⁵¹ Amazon schrapte een aanwervingsalgoritme dat mannen bleek te prefereren boven vrouwen voor technologische rollen.⁵² Is dit de schuld van algoritmische bias, de schuld van de maatschappij, of een neutrale observatie door een algoritme?

Dit leidt ertoe om de vraag te stellen of er—net als bij mensen—wel een neutraal algoritme bestaat. Precies omdat algoritmen mikken op optimalisering op bepaalde dimensies is een neutraal algoritme vaak een contradictie in termen. Daarom is het belangrijk transparant te zijn over op welke dimensies er geoptimaliseerd wordt, op basis waarvan duidelijker kan worden geconcludeerd in welke gevallen dit aanvaardbaar is. Indien een algoritme besluit om geen krediet toe te kennen aan een individu omdat die tot een etnische groep behoort die gemiddeld genomen een zwakkere kredietgeschiedenis heeft, zou dit een individu straffen op basis van groepskenmerken. Met de toenemende ESG eisen vanuit de diverse stakeholders van bedrijven is het duidelijk dat ook de onbedoelde risico's verbonden aan het design en gebruik van AI op de agenda's van de raden van bestuur zullen belanden.

3.3. De reactie op AI: Ethische codes, manifesten, wetgeving

Reacties op de mogelijk ontwrichtende effecten van AI bleven niet lang uit. Bedrijven, ngo's en overheden zijn momenteel samen met AI-ontwikkelaars op zoek naar manieren om de positieve effecten van AI te stimuleren en de potentiële negatieve effecten te temperen. Hiervoor werden er reeds verscheidene **ethische codes en declaraties** opgesteld, variërend van de "Asilomar-principes"⁵³ tot de "Toronto Declaration"⁵⁴ en het "AI for Good"-initiatief van de Verenigde Naties.⁵⁵ Deze declaraties stellen eerder globale beginselen of normen voor.

Daarnaast bestaan er **richtlijnen voor bedrijfspraktijken** ontwikkeld door bijv. Accenture met hun "algorithmic impact assessment"⁵⁶ of IBM met een "AI Fairness 360"-toolkit.⁵⁷ Industriële spelers als Google, SAP, IBM, Microsoft, Deutsche Telekom of Telefonica gebruiken concepten als "verantwoordelijke AI", "betrouwbare AI", of "vertrouwbaar AI".

Tenslotte zijn er **overheidsmanifesten** zoals het Villani rapport geschreven in opdracht van het Franse parlement⁵⁸, de Verklaring van Montreal⁵⁹, de verklaring van de "European Group on Ethics in Science

⁵¹ Simonite, Tom (2015) Probing the Dark Side of Google's Ad-Targeting System. MIT Technology Review, July 6 2015, <https://www.technologyreview.com/2015/07/06/110198/probing-the-dark-side-of-googles-ad-targeting-system/>

⁵² Dastin, Jeffrey (2018) Amazon scraps secret AI recruiting tool that showed bias against women, Reuters 11 October 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

⁵³ Asilomar Principles (2017) Draft principles, <https://ai-ethics.com/past-principles/> & <https://ai-ethics.com/2017/08/11/future-of-life-institute-2017-asilomar-conference/>

⁵⁴ Toronto Declaration (2018) Protecting the right to equality in machine learning, <https://www.torontodeclaration.org/>

⁵⁵ AI for Good Initiative, <https://aiforgood.itu.int/>

⁵⁶ Accenture Algorithmic Impact Assessment, <https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance>

⁵⁷ IBM AI Fairness 360 toolkit: <https://aif360.mybluemix.net/>

⁵⁸ Villani, Cédric (2018) For a Meaningful Artificial Intelligence. Towards a French and European strategy, https://www.jaist.ac.jp/~bao/AI/OtherAIstrategies/MissionVillani_Report_ENG-VF.pdf

⁵⁹ Montreal Declaration (2018) La déclaration de montréal pour un développement responsable de l'intelligence artificielle, <https://declarationmontreal-iaresponsable.com/la-declaration/>

and New Technologies” (EGE)⁶⁰, de Chinese strategie inzake AI⁶¹, de “US Blueprint for an AI Bill of Rights”⁶², naast vele andere initiatieven in verscheidene andere landen. Deze derde groep teksten willen basisrechten beschermen en tegelijk het industrieel concurrentievermogen en de duurzame ontwikkeling stimuleren via overheidsacties.

Opmerkelijk genoeg komt er reeds kritiek op deze evolutie. Munn⁶³ klaagt aan dat de wildgroei aan AI-handvesten en declaraties betekenisloos zijn door hun abstracte vaagheid, en ineffectief door hun gebrek aan afdwingbaarheid. In dit kader zien we dat sommige bedrijven ethische principes gebruiken als een manier om hun reputatie te verbeteren zonder hun producten of activiteiten fundamenteel te veranderen. Naar analogie met greenwashing kan men dit een soort van “ethics washing” noemen. Ook hier zullen raden van bestuur een duidelijk monitorende rol moeten spelen om zeker te zijn dat hun organisatie AI op een correcte manier in de praktijk inzet. De oprichting van een ethisch panel met enkele experts pleit bestuurders niet vrij van verantwoordelijkheid voor de negatieve impact van AI op externe belanghebbenden, met de klanten als belangrijkste externe groep voorop.

De EU staat qua overheidsinitiatief reeds ver met haar **wetgevende initiatief** in de vorm van het ontwerp van de Artificial Intelligence Act⁶⁴, waarin ze de vrijblijvendheid van AI-manifesten en codes wenst te overstijgen. De EU-wetgever maakt een eerste belangrijk onderscheid tussen ‘fundamentele AI-modellen’ en ‘AI voor algemene doeleinden’, en introduceren een strenger regime voor de eerste. De ontwikkelaars van “foundational models” zullen hun systeem moeten laten testen door onafhankelijke experts om de redelijkerwijs voorzienbare risico's voor de gezondheid, veiligheid, grondrechten, het milieu, democratie en de rechtsstaat te beperken. De niet-voorzienbare risico's moeten gedocumenteerd worden.

De ‘AI voor algemene doeleinden’ maken gebruik van die fundamentele modellen, en vallen uiteen in drie soorten systemen: verboden systemen, hoog risico-systemen, en laag risico-systemen. Verboden systemen zijn alle AI-systemen die subliminale, manipulatieve of bedrieglijke technieken gebruiken om het gedrag van een persoon te sturen, AI-systemen die misbruik maken van de kwetsbaarheid van een persoon of groep, en AI-systemen die gebruik maken van sociale scoring van natuurlijke personen die discriminerende resultaten en de uitsluiting van bepaalde groepen tot gevolg kunnen hebben.

Hoog risico-systemen zijn alle AI-systemen die ofwel worden gebruikt als veiligheidscomponent van een ander product, of zelf een product zijn dat al onder het EU New Legislative Framework voor productveiligheid valt (bv. machines, speelgoed, medische hulpmiddelen) en andere producten die

⁶⁰ European Commission (2018) Statement on artificial intelligence, robotics and 'autonomous' systems. Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies, Brussels, 9 March 2018, Publications Office, <https://data.europa.eu/doi/10.2777/531856> or <https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382>

⁶¹ China State Council (2017) A [New] Generation Artificial Intelligence Development Plan, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm, vertaald op <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>

⁶² White House (2022) US Blueprint for an AI Bill of Rights, <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

⁶³ Munn, L. (2022) The uselessness of AI ethics. AI Ethics, Vol. 3, 869–877, <https://doi.org/10.1007/s43681-022-00209-w>

⁶⁴ ARTIFICIAL INTELLIGENCE ACT (2023) Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

onder de geharmoniseerde EU-wetgeving vallen (bv. boten, spoorwegen, motorvoertuigen, vliegtuigen).

Annex III van de EU AI Act somt een exhaustieve lijst op van acht hoog-risico AI-systemen.

- Kritieke infrastructuur (bv. transportinfrastructuur; infrastructuur voor water-, gas-, verwarmings- en elektriciteitsvoorziening);
- Biometrische identificatie van natuurlijke personen;
- Onderwijs en beroepsopleidingen (bv. geautomatiseerde scoring van examens);
- Tewelkstellingsgerelateerde systemen zoals geautomatiseerde aanwervingsprocessen;
- Essentiële private en publieke diensten (bv. geautomatiseerde uitkeringsystemen; kredietscoresystemen voor de private sector);
- Rechtshandavingssystemen die de grondrechten van mensen kunnen aantasten (bv. geautomatiseerde risicoscores voor borgtocht; detectie van misdadrisico);
- Systemen voor het beheer van migratie, asiel en grenscontroles (bv. verificatie van de authenticiteit van reisdocumenten; verwerking van visa);
- Rechtspraak en democratische processen (bv. geautomatiseerde hulp bij vonnissen).

De gebruikers van deze systemen moeten aan zes vereisten voldoen.⁶⁵

- De AI-systemen gebruiken in overeenstemming met de gebruiksinstructies van de aanbieders;
- Voldoen aan specifieke sectorale wetgeving waar relevant (bv. aanvullende goedkeurings- of controlevoorschriften voor een AI-systeem van banken dat leningen ondersteunt);
- Ervoor zorgen dat de inputgegevens relevant zijn voor het beoogde doel van het AI-systeem;
- Controleren of het systeem zich houdt aan de eigen gebruiksvoorwaarden en het gebruik ervan opschorten/rapporteren wanneer er zich ernstige incidenten voordoen;
- De logbestanden van de AI-systemen automatisch en gedocumenteerd bijhouden;
- Een “data protection impact assessment” uitvoeren voor het gebruik van het systeem in kwestie.

De AI-systemen met een laag risico zijn in de AI Act onderhevig aan een aantal minimale vereisten, zoals transparantie over wanneer er AI wordt gebruikt zodat de eindgebruiker kan beslissen of ze de toepassing willen blijven gebruiken. Hieronder vallen niet alleen tekstgebaseerde AI, maar ook alle AI-systemen die beeld-, audio- of video-inhoud genereren of manipuleren.

Gezien het belang van de VS en China binnen de verscheidene technologische lagen van AI, en in het bijzonder van de VS wat betreft fundamentele modellen, zullen internationale wetgevende initiatieven ook een impact hebben tot in de Europese bestuurskamer. China introduceerde in 2021 regulering rond recommendatiealgoritmen, in 2022 regels rond synthetisch gegenereerde media, en in 2023 een aanzet tot regels rond generatieve AI.⁶⁶ In de VS hebben 25 deelstaten al eigen AI-wetgeving geïntroduceerd⁶⁷ en zitten er een paar wetgevende initiatieven in de federale pijplijn. AI wordt

⁶⁵ Demircan, Muhammed (2023) Deployers of High-Risk AI Systems: What Will Be Your Obligations Under the EU AI Act? Kluwer Competition Law Blog, 2 June 2023, <https://competitionlawblog.kluwercompetitionlaw.com/2023/06/02/deployers-of-high-risk-ai-systems-what-will-be-your-obligations-under-the-eu-ai-act/>

⁶⁶ Voor een recent overzicht van Chinese wetgevende initiatieven, zie Sheehan, Matt (2023) China’s AI Regulations and How They Get Made, Carnegie Endowment for International Peace, 10 July 2023, <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>

⁶⁷ Voor een overzicht van AI regulering op statenniveau in de VS, zie National Conference of State Legislatures (2023) Artificial Intelligence 2023 Legislation, 27 September 2023, <https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation>

momenteel in de VS federaal nog gereguleerd vanuit de bestaande wettelijk kaders (eg. productveiligheid, competitiebeleid, privacy, etc.), en aangestuurd door de bestaande federale agentschappen en departementen. In 2020 richtte men het National Artificial Intelligence Initiative op dat een coördinerende rol tracht te spelen tussen deze federale actoren.⁶⁸

Tenslotte valt te verwachten dat er in de toekomst een reactie komt van de mededingingsautoriteiten op de mogelijke industriële dominantie van een beperkt aantal spelers binnen de foundational modellen. Hoe al deze factoren de komende jaren zullen interageren, en welke uitvoeringsbesluiten of rechtzaken het juridische speelveld vorm zullen geven, is nog uiterst onvoorspelbaar.

⁶⁸ National Artificial Intelligence Initiative, <https://www.ai.gov/>

4. Conclusie

Deze studie definieert het begrip van AI en situeert het binnen haar bredere technische architectuur en maatschappelijke context. AI wordt soms als synoniem gebruikt met “machine learning” of in het slechtste geval uitgebreid tot alle big data-analyse, maar is een specifieke en voortdurend ontwikkelende familie van methoden gekenmerkt door een zekere mate van autonome besluitvorming. Deze methoden worden reeds toegepast in verscheidene sectoren zoals transport (logistiek, zelfrijdende wagens), gezondheidszorg (ziektebeeldherkenning, chirurgische robotica, medicijnontwikkeling), online media (recommender systemen, zoekmachines, chatbots), veiligheidsdiensten (gelaatsherkenning), financiële markten (algorithmic trading en analyse), en defensie (autonome wapensystemen).

Deze verzameling van AI-diensten heeft de mogelijkheid om uit te groeien tot een “General Purpose Technology” die vele aspecten van onze economie en samenleving grondig kan beïnvloeden. AI kan een positieve invloed hebben op het welzijn en de welvaart van onze samenleving in de toekomst, maar komt ook met mogelijke risico's van opzettelijke of onbedoelde schade.

Het mag duidelijk zijn dat de recente turbulentie rond de nieuwste generatie van AI onze maatschappij op haar grondvesten doet daveren. De autonome vormen van AI zullen een groeiend aantal taken op zich nemen, en zo de jobmarkt en bij uitbreiding de maatschappij hertekenen. Door haar grote impact zal dit voor organisaties en hun raden van bestuur zowel nieuwe opportuniteiten als uitdagingen vormen. AI verandert niet alleen de interne bedrijfsprocessen, maar ook de hele omgeving van het bedrijf. Het zal voor raden van bestuur geen keuze zijn of men wil omgaan met AI, maar hoe men zal omgaan met AI binnen het bedrijf of organisatie. Bestuurders moeten hierop voorbereid zijn en zich scholen in hoe AI werkt en welke gevolgen het heeft voor de besturen waarin ze zetelen, van grote publieke organisaties of grote (al dan niet beursgenoteerde) ondernemingen, tot de kleinste kmo die AI-diensten binnen hun conventionele software zullen zien verschijnen.

Om die reden zal GUBERNA in twee toekomstige studies dieper ingaan op AI als specifieke uitdaging voor de raad van bestuur. In de volgende tekst zal gekeken worden naar hoe AI de controle-, leiderschaps- en strategische rol van de raad van bestuur beïnvloedt. Vragen zoals “welke taken gaan we toewijzen aan AI?”, “welke strategische opportuniteiten bieden AI-systemen voor onze organisatie?” of “hoe gaan we de risico's monitoren?” zullen binnenkort op de tafels van de raden van bestuur liggen. In een andere vervolgstudie zullen we kijken naar de mogelijkheden en gevolgen van het gebruik van AI binnen de raad van bestuur zelf, of in een van de verscheidene comités.

Referenties

- AI for Belgium: Industrieel landschap, <https://ai4belgium.be/nl/ai-landschap/>
- Accenture Algorithmic Impact Assessment, <https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance>
- Acemoglu, D. and Restrepo, P. (2022) Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5):1973–2016, <http://dx.doi.org/10.3982/ECTA19815>
- Altman, Sam; Brockman, Greg and Sutskever, Ilya (2023) Governance of superintelligence, 22 May 2023, <https://openai.com/blog/governance-of-superintelligence>
- Asilomar Principles (2017) Draft principles, <https://ai-ethics.com/past-principles/> & <https://ai-ethics.com/2017/08/11/future-of-life-institute-2017-asilomar-conference/>
- Autor, David H., Katz, Lawrence F., and Kearney, Melissa S. (2006) The polarization of the us labor market. *American Economic Review*, 96(2): 189–194, <http://dx.doi.org/10.1257/000282806777212620>
- Bedingfield, Will (2023) Hollywood Writers Reached an AI Deal That Will Rewrite History, *Wired*, 27 September 2023, <https://www.wired.com/story/us-writers-strike-ai-provisions-precedents/>
- Bergen, Mark E.; Dutta, Shantanu; Guszczka, James and Zbaracki, Mark J. (2012) How AI Can Help Companies Set Prices More Ethically, *Harvard Business Review*, 26 March 2021, <https://hbr.org/2021/03/how-ai-can-help-companies-set-prices-more-ethically>
- Bowyer, Kevin; King, Michael; Scheirer; Walter & Vangara, Kushal (2020) The “Criminality From Face” Illusion. *IEEE Transactions on Technology and Society*, Vol. 1, issue 4, <https://ieeexplore.ieee.org/document/9233349> en open access versie op <https://arxiv.org/abs/2006.03895>
- Brynjolfsson, Erik; Rock, Daniel and Syverson, Chad et al. (2017) Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, pp. 23 – 57 in Ajay Agrawal, Joshua Gans, and Avi Goldfarb (eds.) *The Economics of Artificial Intelligence: An Agenda*, <https://www.nber.org/system/files/chapters/c14007/c14007.pdf>
- Center for Research on Foundation Models (CRFM) at Stanford University (2021) On the Opportunities and Risks of Foundation Models, 16 August 2021, <https://fsi.stanford.edu/publication/opportunities-and-risks-foundation-models>
- Chen, Nicholas; Christensen, Lau; Gallagher, Kevin; Mate, Rosamond & Rafert, Greg (2016), Global economic impacts associated with artificial intelligence, Analysis Group, https://www.analysisgroup.com/globalassets/content/insights/publishing/ag_full_report_economic_impact_of_ai.pdf
- Petropoulos, Georgios; Pichler, David & Chiacchio, Francesco (2018) The impact of industrial robots on EU employment and wages: A local labour market approach, *Bruegel Working Paper*, 18 April 2018, Bruegel, Brussels, <https://www.bruegel.org/working-paper/impact-industrial-robots-eu-employment-and-wages-local-labour-market-approach>

China State Council (2017) A [New] Generation Artificial Intelligence Development Plan, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm, vertaald op <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>

Commission (2018) Coordinated Plan on Artificial Intelligence. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic And Social Committee and the Committee of the Regions, Brussels, 7.12.2018, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795&from=DA>

Dastin, Jeffrey (2018) Amazon scraps secret AI recruiting tool that showed bias against women, Reuters 11 October 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

Demircan, Muhammed (2023) Deployers of High-Risk AI Systems: What Will Be Your Obligations Under the EU AI Act? Kluwer Competition Law Blog, 2 June 2023, <https://competitionlawblog.kluwercompetitionlaw.com/2023/06/02/deployers-of-high-risk-ai-systems-what-will-be-your-obligations-under-the-eu-ai-act/>

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023) GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, <https://arxiv.org/abs/2303.10130>

OECD Employment Outlook (2023) Artificial intelligence and jobs. An urgent need to act, <https://www.oecd.org/employment-outlook/2023/>

EU AI ACT (2023) https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

EU HLEG on AI (High Level Expert Group on Artificial Intelligence) (2018), A definition of AI: Main capabilities and scientific disciplines, 18 December 2018, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651

European Commission (2018) Statement on artificial intelligence, robotics and 'autonomous' systems. Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies, Brussels, 9 March 2018, Publications Office, <https://data.europa.eu/doi/10.2777/531856> or <https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382>

Ferguson, A.G. (2017) The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement, New York, NY: NYU Press.

FOD Economie (2021) Barometer van de informatiemaatschappij: Geavanceerde digitale technieken, <https://economie.fgov.be/nl/themas/online/ict-belgie/barometer-van-de/de-ondernemingen-en-het/de-ondernemingen-en-de>

FOD Economie (2021b) Digitalisering van kmo's: Artificiële intelligentie (AI) en andere geavanceerde digitale technologieën, <https://economie.fgov.be/nl/themas/ondernemingen/kmos-en-zelfstandigen-cijfers/digitalisering-van-kmos/artificiele-intelligentie-en>

Furman, Jason and Seamans, Robert (2018) AI and the Economy, NBER Working Paper No. 24689, NBER, Cambridge, MA, https://www.nber.org/system/files/working_papers/w24689/w24689.pdf

Geist, Edward & Andrew, John (2018) How Might Artificial Intelligence Affect the Risk of Nuclear War? RAND Corporation, Pages: 28, <https://doi.org/10.7249/PE296>

- Goertzel, Ben (2014) Artificial general intelligence: concept, state of the art, and future prospects, *Journal of Artificial General Intelligence*, 5(1) 1-46, <https://sciendo.com/downloadpdf/journals/jagi/5/1/article-p1.pdf>
- Goldfarb, A., Taska, B., and Teodoridis, F. (2023) Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *Research Policy*, 52(1): 104653, <http://dx.doi.org/10.1016/j.respol.2022.104653>
- Graetz, Georg and Michaels, Guy (2017), Is Modern Technology Responsible for Jobless Recoveries? *American Economic Review*, 107(5), 168–73, <https://www.aeaweb.org/articles?id=10.1257/aer.p20171100>
- Graetz, Georg and Michaels, Guy (2018), Robots at Work, *Review of Economics and Statistics*, Vol. 100, 753-768, <https://direct.mit.edu/rest/article-abstract/100/5/753/58489/Robots-at-Work?redirectedFrom=fulltext>
- HAI project (2021) The Shibboleth rule for Artificial Agents, Stanford University, Human-Centered Artificial Intelligence (HAI), Aug 20 2021, <https://hai.stanford.edu/news/shibboleth-rule-artificial-agents>
- Hilb, Michael (2020) Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance. *Journal of Management and Governance*, 24: 851–870, <https://doi.org/10.1007/s10997-020-09519-9>
- IBM AI Fairness 360 toolkit, <https://aif360.mybluemix.net/>
- International Labour Organization (2019) Work for a Brighter Future, Report, 22 January 2019, https://www.ilo.org/global/publications/books/WCMS_662410/lang--en/index.htm
- ITU, AI for Good Global Summit, <https://aiforgood.itu.int/>
- Iwasaki, I., Satogami, M. (2023) Gender wage gap in European emerging markets: a meta-analytic perspective, *Journal for Labour Market Research*, Vol. 57, Art. 9, <https://doi.org/10.1186/s12651-023-00333-y>
- Lighthill, James (1973) Artificial Intelligence: A General Survey, in *Artificial Intelligence: A paper symposium*. UK: Science Research Council, http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm
- McKinsey (2023) The economic potential of generative AI: The next productivity frontier, 14 June 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights>
- Montreal Declaration (2018) La déclaration de montréal pour un développement responsable de l'intelligence artificielle, <https://declarationmontreal-iaresponsable.com/la-declaration/>
- Muller, Joann (2023) Robotaxis hit the accelerator in growing list of cities nationwide, *Axios*, 29 August 2023, <https://www.axios.com/2023/08/29/cities-testing-self-driving-driverless-taxis-robotaxi-waymo>
- Munn, L. (2022) The uselessness of AI ethics. *AI Ethics*, Vol. 3, 869–877, <https://doi.org/10.1007/s43681-022-00209-w>
- National Artificial Intelligence Initiative (US), <https://www.ai.gov/>

- National Conference of State Legislatures (2023) Artificial Intelligence 2023 Legislation, 27 September 2023, <https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation>
- Nilsson, Nils J. (2005) Human-Level Artificial Intelligence? Be Serious! AI Magazine, 25th Anniversary Issue, 68-75, <https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag26-04-HLAI.pdf>
- O’Neil, C. (2016) Weapons of math destruction: How big data increases inequality and threatens democracy, New York, NY: Crown Publishers.
- Pierce, John R.; Carroll, John B. et al. (1966) Language and Machines — Computers in Translation and Linguistics. ALPAC report, National Academy of Sciences, National Research Council, Washington, DC, <https://web.archive.org/web/20110409070141/http://www.mt-archive.info/ALPAC-1966.pdf>
- Purdy, M. and Dougherty, P. (2017) How AI boosts industry profits and innovation. Accenture Research, <https://www.accenture.com/fr-fr/acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf>
- PwC (2017) Sizing the prize - What’s the real value of AI for your business and how can you capitalise? <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>
- Shapiro, Carl & Varian, Hal R. (1998) Information Rules, Harvard Business School Press, Boston: Massachusetts.
- Sheehan, Matt (2023) China’s AI Regulations and How They Get Made, Carnegie Endowment for International Peace, 10 July 2023, <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>
- Simonite, Tom (2015) Probing the Dark Side of Google’s Ad-Targeting System. MIT Technology Review, July 6 2015, <https://www.technologyreview.com/2015/07/06/110198/probing-the-dark-side-of-googles-ad-targeting-system/>
- Stam, K. et al. (2015), “Employment status and subjective well-being: The role of the social norm to work”, Work, Employment and Society, 1–25, <https://journals.sagepub.com/doi/10.1177/0950017014564602>
- Tegmark, Max (2017) Life 3.0, New York City: Knopf.
- Toronto Declaration (2018) Protecting the right to equality in machine learning, <https://www.torontodeclaration.org/>
- Turing, A.M. (1950) I.—COMPUTING MACHINERY AND INTELLIGENCE, Mind, Volume LIX, Issue 236, October 1950, Pages 433–460, <https://doi.org/10.1093/mind/LIX.236.433>
- US Sentencing Commission (2017) Demographic Differences in Sentencing, November 2017, <https://www.ussc.gov/research/research-reports/demographic-differences-sentencing>
- Van Reenen, J. (2011) Wage inequality, technology and trade: 21st century evidence. Labour Economics, 18(6):730–741, <https://doi.org/10.1016/j.labeco.2011.05.006>
- Villani, Cédric (2018) For a Meaningful Artificial Intelligence. Towards a French and European strategy, https://www.jaist.ac.jp/~bao/AI/OtherAIstrategies/MissionVillani_Report_ENG-VF.pdf

White House (2022) US Blueprint for an AI Bill of Rights, <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019) The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Jan 2019, pp. 195-200), <https://doi.org/10.1145/3306618.3314289>

Wozniak, Steve in Shick, Michael (2010) Wozniak: Could a Computer Make a Cup of Coffee? Fast Company, 02 March 2010, <https://www.fastcompany.com/1568187/wozniak-could-computer-make-cup-coffee>

Zysman, John & Nitzberg, Mark (2020) Governing AI: Understanding the Limits, Possibility, and Risks of AI in an Era of Intelligent Tools and Systems, BRIE Working Paper No. 2020-5, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3681088

Deze studie kwam tot stand met de steun van



GUBERNA partner in 'Governance and Digital Transformation'