TECHNOLOGISCHE DOORBRAAK

EINDELIJK EEN PRIVACYVRIENDELIJKE MANIER OM DATA TE BENUTTEN

TNO innovation for life

maart 2021

Zie jij het voor je? Je hebt een medische aandoening waar je medicijnen voor nodig hebt. De effectiviteit en bijwerkingen van de beschikbare behandelingen verschillen echter van patiënt tot patiënt. Door inzichten te gebruiken van effectiviteit en bijwerkingen bij andere patiënten kun jij beter geholpen worden. Maar daarvoor zijn wel analyses van gevoelige patiëntdata nodig. Met de inzet van innovatieve technologie kunnen de benodigde inzichten verkregen worden zonder op de privacy van patiënten en artsen in te boeten. Daardoor kunnen jij en jouw arts direct de voor jou optimale behandeling kiezen.

Zie jij het voor je? Je werkt bij een bank op de afdeling witwasdetectie. Jouw bank zet zich, net als andere banken en financiële instellingen, in om witwasactiviteiten op te sporen. Toch blijft een groot deel onder de radar (99%!), omdat veel criminelen gebruik maken van opeenvolgende transacties via meerdere banken. Jij ziet daardoor maar een stukje van de puzzel en moet op basis daarvan handelen. Innovatieve technologie biedt jou nu gelukkig de kans om samen met andere banken verdachte geldstromen te detecteren zonder onderling persoonsgegevens of andere gevoelige data te delen.

Dit zien wij voor ons met het privacyvriendelijk benutten van data
Het analyseren van gekoppelde databronnen maakt het mogelijk om grote innovatie- en maatschappelijke uitdagingen aan te pakken en economische groei te realiseren. Data delen komt echter nog onvoldoende van de grond door commerciële en/of wettelijke belemmeringen, waaronder het fundamentele recht op privacy. Maar wat als je helemaal geen data hoeft te delen om tot inzichten te komen?

Het is wat ons betreft tijd voor een nieuw uitgangspunt als het gaat om het verwerken van gevoelige data: laten we afstappen van de traditionele centralistische zienswijze en gebruik gaan maken van gedistribueerde dataverwerking. Want om waarde te creëren uit data hoef je deze data niet te bezitten. Deel dus geen data, maar benut inzichten uit verspreide databronnen terwijl privacy en vertrouwelijkheid gewaarborgd worden. Met innovatieve technologieën als Multi Party Computation (MPC) en Federated Learning (FL) kan dit doel bereikt worden zonder in te boeten op privacy en vertrouwelijkheid. In dit paper lichten we toe hoe we dit voor ons zien en geven we voorbeelden van toepassingen.

Omdat dit bij uitstek een ketenuitdaging is roepen wij overheid, bedrijven en kennisinstellingen op de handen ineen te slaan binnen publiek-private samenwerkingen en opschaling van deze technologie te versnellen door hun databronnen open te stellen voor experimenten. We nodigen u van harte uit om deze aanpak gezamenlijk verder uit te werken en te beproeven in de praktijk.

Zie jij het voor je?

SAMENVATTING

Het delen en analyseren van data is essentieel om economische groei te realiseren en maatschappelijke uitdagingen op te lossen. Data delen komt echter nog onvoldoende van de grond door commerciële en/of wettelijke belemmeringen, waaronder het fundamentele recht op privacy. Innovatieve technologieën zoals Federated Learning en Multi-Party Computation bieden dé manier om dit probleem aan te pakken, door op een veilige manier te leren van gevoelige data uit meerdere bronnen zonder deze data te hoeven delen.

Niet alleen Google en Facebook, maar vrijwel alles en iedereen om ons heen verzamelt steeds meer data. Denk aan thermostaten, smartwatches, bewegingsapps en navigatiesystemen. Maar ook ziekenhuizen, banken, de logistieke sector en andere organisaties proberen aan de hand van data hun dienstverlening te verbeteren. Bovendien wordt er in de (smart) industrie al volop gebruik gemaakt van data om bestaande productieprocessen, die steeds meer in ketens plaats vindt, duurzamer en efficiënter te maken. Om dit voor elkaar te krijgen is het combineren en analyseren van verschillende databronnen (uit die ketens) nodig. Voor het succesvol inzetten van technieken als Artificial Intelligence (AI) is datadeling de sleutel. Dus ook om economische groei te realiseren, onze zorg te verbeteren en betaalbaar te houden, criminaliteit aan te pakken en onze arbeidsproductiviteit te verhogen.

Maar er is ook een keerzijde. Zo zijn er zorgen over het vrijelijk verzamelen en delen van data. Grote platformen waarin persoonlijke gegevens centraal beheerd en zonder toestemming verwerkt worden passen niet bij onze Europese waarden. Deze platformen resulteren in monopolieposities, waarbij gemakkelijk misbruik van persoonlijke data gemaakt kan worden. Onze privacy behoort tot onze fundamentele rechten en de bescherming van persoonlijke data ligt vast in de Europese GDPR-regelgeving en de Nederlandse AVG. Deze wet- en regelgeving wordt vaak genoemd als barrière voor het optimaal benutten van data(deling). Als organisatie wil je bovendien niet zomaar jouw commercieel gevoelige data uit handen geven. Je wilt immers controle houden over welke data met wie gedeeld worden, maar tegelijkertijd wel de waarde van deze data benutten.

Het doel is echter niet om data te verzamelen of te delen, het doel is om tot nieuwe *inzichten* te komen door *te leren uit data*. Wat als je helemaal geen data hoeft te delen om tot deze inzichten te komen?

Technologieën als Multi-Party Computation (MPC) en Federated Learning (FL) maken dit mogelijk. Hiermee kan inzicht worden verkregen door meerdere databronnen te koppelen zonder in te boeten op privacy of vertrouwelijkheid. MPC maakt gebruikt van cryptografie. Deze techniek zorgt ervoor dat analyses uitgevoerd kunnen worden op versleutelde data. De onderliggende, vaak gevoelige, data hoeven niet te worden gedeeld om analyses uit te voeren en nieuwe inzichten te verkrijgen. Met als resultaat dat er geen gevoelige informatie naar andere partijen 'lekt' en alleen bewerkingen mogelijk zijn die tot het vooraf ontworpen eindresultaat leiden. Bij FL wordt de analyse naar de data gebracht in plaatst van andersom. De data-eigenaar houdt dus controle over het gebruik van de data, doordat deze veilig en gedecentraliseerd blijven. Tegelijkertijd blijft het mogelijk om waarde uit de data te creëren. De potentie van deze technologieën voor onze maatschappij is enorm.

De eerste oplossingen op basis van MPC en FL zijn nu technologisch volwassen en worden in verschillende domeinen toegepast. Door gebruik van deze technieken kunnen privacy en vertrouwelijkheid op een veel sterkere manier gewaarborgd worden. Zo wordt een verregaande vorm van dataminimalisatie, zoals gestimuleerd door de AVG, mogelijk gemaakt. Alleen de analyse resultaten worden gedeeld. In samenwerkingen met juristen en ethici zal de gevoeligheid van deze resultaten altijd afgewogen moeten worden tegen de relevante regelgeving om juist gebruik en proportionaliteit te waarborgen. Doordat alleen de resultaten gedeeld worden zal deze juridische en ethische toetsing bij veel meer toepassingen succesvol verlopen dan de klassieke aanpak waarin ook onderliggende data worden gedeeld. MPC en FL laten zien dat het tijd is voor een nieuw uitgangspunt op het delen van gevoelige data: *deel geen data, maar benut inzichten uit verspreide databronnen terwijl privacy en vertrouwelijkheid gewaarborgd worden*. Op deze manier kunnen we veilig het perspectief op verdere economische groei realiseren en belangrijke maatschappelijke uitdagingen oplossen.

Daarbij is het essentieel dat de overheid, private partijen en andere organisaties opschaling van deze technologie versnellen door eigen databronnen open te stellen voor privacyvriendelijke analyses en publiek-private samenwerking te stimuleren om operationalisering te versnellen. Dit alles met het uitgangspunt dat datadeling niet noodzakelijk is en privacy en vertrouwelijkheid dus gewaarborgd kunnen blijven. Een multidisciplinaire aanpak is cruciaal om zowel de technologieën verder te ontwikkelen en op te schalen, als verdere invulling te geven aan de ethische en juridische kaders en normen.

Niemand kan dit vliegwiel voor economie en maatschappij alleen op gang krijgen; dit is een ketenuitdaging. Wij roepen daarom overheidspartijen, bedrijven, commerciële technologiepartijen en kennisinstellingen op om gezamenlijk aan de slag te gaan met het praktijk-klaar maken van deze nieuwe technieken.

INHOUD

DE WAARDE VAN GEVOELIGE INFORMATIE

Het analyseren van gekoppelde databronnen maakt het mogelijk om grote innovatie-uitdagingen in Nederland op te lossen, maatschappelijke uitdagingen aan te pakken en economische groei te realiseren. Elkaar snel opeenvolgende ontwikkelingen, in bijvoorbeeld Artificial Intelligence (AI), bieden de mogelijkheid om groeiende hoeveelheden data om te zetten in bruikbare informatie die leidt tot nieuwe toepassingen en inzichten. Zo worden onder andere nieuwe medicijnen ontwikkeld, logistieke processen geoptimaliseerd, industriële productieketens verbeterd en frauduleuze activiteiten gedetecteerd. Naast wettelijke en commerciële belemmeringen staan maatschappelijke zorgen de uitwisseling van data echter in de weg.

DATA AAN DE BASIS VAN ECONOMISCHE GROEI

Post-corona zien we een financiële krimp die voor Nederland wordt geschat op ruim 4% van het BBP (CPB) [1]. Een van onze grootste uitdagingen is dan ook het realiseren van economische groei. De technologische oplossing voor deze uitdaging ligt binnen handbereik. Zo blijkt uit recente analyses dat de beschikbaarheid en uitwisseling van data kunnen leiden tot een economische groei van 1,5% van het BBP [2]. Sommige studies geven aan dat, wanneer niet alleen data uit de publieke sector wordt gebruikt, deze groei kan oplopen tot wel 4% van het BBP. De beschikbaarheid van data maakt nieuwe AI-toepassingen mogelijk waardoor verschillende industrieën hun verdienvermogen met 30 tot 128% kunnen verhogen [3]. Door beschikbaarheid van de juiste data kan er een gestroomlijnde arbeidsmarkt ontstaan waarin vraag en aanbod efficiënt aan elkaar gekoppeld worden. Gepersonaliseerde gezondheidszorg houdt de zorg niet alleen betaalbaar, maar hierdoor blijven Nederlanders ook langer in het arbeidsproces. Mobiliteitsuitdagingen kunnen opgelost worden door vergaande automatisering in de transportsector. Daarnaast stelt AI ons in staat ons te wapenen tegen de alsmaar groeiende dreiging van cyber aanvallen. Deze ontwikkelingen kunnen ons van het post-coronatijdperk naar een ‘gouden decennium’ van economische groei brengen.

MAATSCHAPPELIJKE UITDAGINGEN OPLOSSEN MET DATA

Naast economische groei kan er ook maatschappelijke impact gemaakt worden door op een slimme manier gebruik te maken van data. Gepersonaliseerde gezondheidszorg is hier een voorbeeld van. Daarnaast kunnen overheidsinstanties op een adequatere manier diensten verlenen aan burgers wanneer data optimaal ingezet worden, bijvoorbeeld door hulp te bieden aan personen met financiële problemen die recht hebben op specifieke uitkeringen. Op pagina 13 gaan we dieper in op deze en andere toepassingen.

ECONOMISCHE EN MAATSCHAPPELIJKE POTENTIE VERSUS PRIVACY EN VERTROUWELIJKHEID

Snelle en ongereguleerde technologische ontwikkelingen hebben de afgelopen decennia gezorgd voor het tot stand komen van grote platformen, zoals die van Google en Facebook, waar gigantische hoeveelheden data samenkomen. De macht van voornamelijk Amerikaanse partijen die deze traditionele en gecentraliseerde aanpak volgen staat echter steeds vaker ter discussie. Grote dataplatformen waar data centraal worden beheerd en veelal zonder expliciete toestemming verwerkt passen niet bij de Europese waarden, waaronder het fundamentele recht op privacy. De Europese GDPR-wetgeving beschermt persoonlijke data en vormt serieuze belemmeringen voor de gecentraliseerde aanpak. Europa heeft recentelijk ook duidelijk signalen afgegeven dat het een andere kant op wil dan de tot nu vooral 'winner-takes-all' business-modellen van data platformen [4]. Binnen het GAIA-X initiatief wordt bijvoorbeeld een veilige, transparante en gefedereerde Europese data infrastructuur ontwikkeld [5]. Daarnaast kunnen data concurrentiegevoelige informatie bevatten en spelen bij bedrijven vaak commerciële belangen mee die de uitwisseling van data in de weg staan.

Al met al bestaan er dus sterk tegenstrijdige belangen als het gaat om het delen van data. Een enorme economische en maatschappelijke potentie aan de ene kant en privacy en vertrouwelijkheid aan de andere kant. Maar dit hoeft geen zero-sum game te zijn. Het belang is om én de privacy te bewaken én inzichten uit data halen [6]. Nieuwe innovatieve technologieën, zoals Multi-Party Computation (MPC) en Federated Learning (FL), bieden een oplossing. Deze technieken maken het mogelijk om analyses en berekeningen op meerdere databronnen uit te voeren zonder deze data bij elkaar te brengen. Op deze manier kunnen verschillende partijen gezamenlijk aan data rekenen zonder dat ze elkaars data daadwerkelijk kunnen zien.

DE JUISTE GRONDSLAG VOOR DATAVERWERKING BLIJFT ESSENTIEEL

MPC en FL zijn belangrijke technieken om data-analyse toepassingen op een privacy vriendelijke manier te ontwerpen (Privacy-by-Design). Ze bieden een kans op betere databescherming en een proportionele dataverwerking die past binnen onze normen en waarden. Verschillende databeschermingsrisico's, zoals oneigenlijk gebruik van persoonsgegevens of risico's op hacks, kunnen flink beperkt worden door de inzet van MPC en FL. Daardoor kan de privacy van individuen gewaarborgd worden en is er geen noodzaak om data af te staan aan gecentraliseerde dataplatformen in de VS. Samen met nieuwe Europese regelgeving spelen daarom ook deze technieken een belangrijke rol in het kunnen doorbreken van de afhankelijkheid van deze platformen. Het maatschappelijk en commercieel wantrouwen jegens datadeling kan op deze manier overwonnen worden binnen de kaders van de regelgeving. MPC en FL bestaan al langer, maar zijn pas de laatste jaren zo doorontwikkeld dat ze nu ook op grote schaal en voor complexere analyses toepasbaar zijn.

INZICHT ZONDER DATA TE DELEN: HOE WERKT DAT?

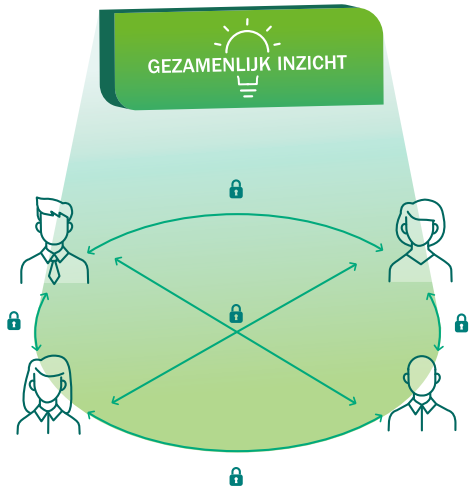
Zowel MPC en FL gaan uit van een scenario waarin meerdere partijen een gezamenlijke berekening of analyse willen uitvoeren op basis van hun eigen data zonder deze te hoeven delen. Denk bijvoorbeeld aan een ziekenhuis en een zorgverzekeraar met het gezamenlijke doel om zo efficiënt mogelijk de beste zorg te leveren. Om dit te bereiken hebben ze elkaars informatie nodig omtrent bijvoorbeeld de behandelhistorie van patiënten. De patiëntgegevens die hiervoor geanalyseerd moeten worden zijn echter privacygevoelig en kunnen niet zomaar uitgewisseld worden.

Federated Learning

De traditionele, en privacy onvriendelijke, oplossing vereist dat de data centraal verzameld worden om vervolgens de juiste analyses uit te voeren. FL lost het privacyprobleem op door de analyses naar de data te brengen in plaats van de data naar de analyses. De analyses worden opgeknipt in kleine deelberekeningen die lokaal uitgevoerd kunnen worden door de verschillende partijen. Na het uitvoeren van een lokale berekening worden alleen de (tussen)resultaten met één of meerdere partijen gedeeld. De gevoelige data worden met niemand gedeeld en blijven bij de partij. De Personal Health Train (PHT) [7] maakt gebruik van deze oplossing om, op basis van gedistribueerde databronnen, zorg op maat te kunnen leveren zonder de data centraal te verzamelen. FL kan veel sterkere privacy- en vertrouwelijkheidsgaranties geven dan de traditionele aanpak waarbij alle data op een centrale plek verzameld worden.

Multi-Party Computation

Door gebruik te maken van geavanceerde cryptografische technieken kan MPC gevoelige data nog beter beschermen dan FL. MPC-protocollen zorgen er namelijk voor dat alle data in versleutelde vorm kunnen blijven. De berekeningen worden uitgevoerd op de versleutelde data en alleen de uitkomst van de analyse wordt ontsleuteld. Dus zelfs terwijl de data te allen tijde versleuteld blijven kan er mee gerekend worden. Hiermee realiseert MPC een maximaal haalbare mate van privacy en vertrouwelijkheid; alleen de uitkomst van de analyse wordt onthuld. Een nadeel is dat MPC over het algemeen meer rekenkracht en/of een zwaardere communicatie infrastructuur nodig heeft dan FL. In de technische verdieping gaan we verder in op de afweging tussen MPC en FL.

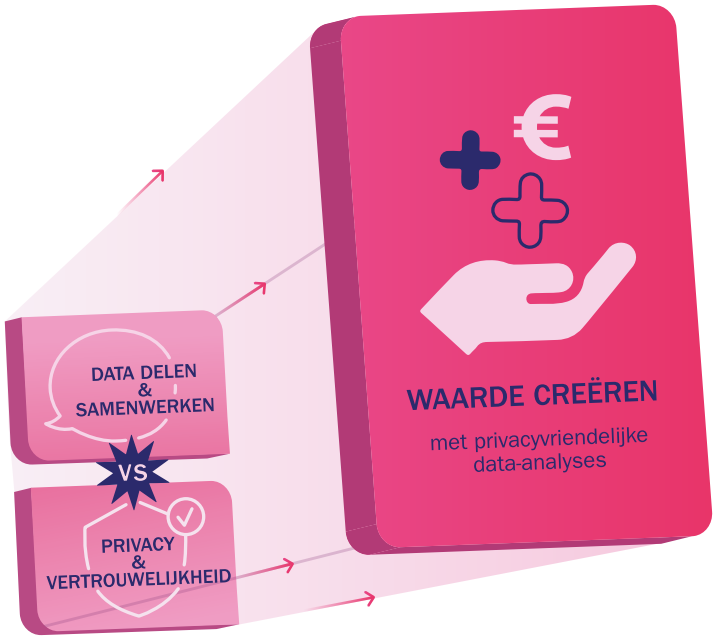


MULTI-PARTY COMPUTATION: GEZAMENLIJK INZICHT ZONDER DATA TE DELEN.

NIEUW PERSPECTIEF

De traditionele zienswijze is dat macht voortkomt uit het bezitten van kennis en data. Het vertalen van data naar waardevolle informatie vereist in deze zienswijze een centrale aanpak, waarin één partij alle data in handen heeft. Deze aanpak staat daarmee loodrecht op belangen als vertrouwelijkheid en privacy. Het lijkt er daarom op dat er gekozen moet worden tussen twee tegenstrijdige belangen.

Veelal is het doel echter niet om data te bezitten, maar om waarde te creëren uit deze data. Daarvoor hoeven data niet door één partij bij elkaar gebracht te worden, maar kunnen ze decentraal beheerd blijven. Data-eigenaren houden op deze manier controle over hun data. Innovatieve technologieën zoals MPC en FL laten zien dat dit doel bereikt kan worden zonder in te boeten op privacy en vertrouwelijkheid. Daarmee is het duidelijk tijd om van de traditionele centralistische zienswijze af te stappen, en gebruik te gaan maken van gedistribueerde dataverwerking. Hiermee bieden deze technologieën een Europees alternatief voor de Amerikaanse en Chinese dataplatformen.



MPC EN FL ZORGEN VOOR EEN NIEUW PERSPECTIEF.

TOEPASSING: VAN OPTIMALISEREN ZORG TOT VOORKOMEN VAN FINANCIËLE CRIMINALITEIT

Er zijn ontzettend veel toepassingsmogelijkheden voor privacyverbeterende technieken zoals MPC en FL. Zo kan de effectiviteit van de zorg vergroot worden door op een privacy vriendelijke manier inzichten uit patiëntdata te verkrijgen. De groeiende financiële criminaliteit kan ingedamd worden door het veilig koppelen van gevoelige data van verschillende financiële organisaties. Daarnaast kan de overheid haar dienstverlening verbeteren door privacy respecterende samenwerkingen tussen verschillende overheidsinstanties. Deze drie toepassingsdomeinen zijn vanaf pagina 13 verder uitgewerkt.

Ook in de mobiliteitssector kunnen technieken als MPC en FL worden ingezet. Er vindt namelijk een beweging plaats naar Mobility as a Service, een innovatief concept waarbij de reiziger centraal staat en via een platform optimaal gebruik kan maken van een breed scala aan mobiliteitsvormen. Om als overheid te kunnen optimaliseren en sturen op deze nieuwe mobiliteit, is het belangrijk dat privacy- en concurrentiegevoelige gegevens van verschillende (concurrerende) MaaS dienstverleners en mobiliteitsaanbieders, zoals reizigersdata en actuele vervoers-capaciteit kan worden geanalyseerd en gemonitord. Een andere toepassing van MPC en FL is het optimaliseren van logistieke ketens zonder de uitwisseling van bedrijfs-gevoelige informatie. Hierdoor kunnen onderling concurrerende partijen toch gezamenlijk waarde creëren. Bovendien kunnen bedrijven, door het verrijken van hun marktsegmentatie met behulp van databronnen die nu nog niet beschikbaar zijn, producten gerichter op de markt zetten. Verder helpen privacyvriendelijke analyses in het effectief detecteren van cyberaanvallen, doordat complexe aanvalspatronen beter in kaart gebracht kunnen worden. En bieden deze technologieën mogelijkheden om in te zetten in het veiligheidsdomein, bijvoorbeeld bij de opsporing van onvindbare veroordeelden [8].

> “Met MPC en FL kunnen onderling concurrerende partijen gezamenlijk waarde creëren.”

HOE KRIJGEN WE HET VLIEGWIEL VAN PRIVACYVRIENDELIJKE DATA ANALYSE OP GANG?

Er zijn zowel technische als organisatorische uitdagingen, die gezamenlijk opgelost moeten worden om deze technologieën grootschalig in te kunnen zetten en de benoemde voordelen te realiseren. Naast de technische uitdagingen zullen deze oplossingen bijvoorbeeld ingebed moeten worden in bestaande infrastructuren waarin gevoelige data geïsoleerd worden opgeslagen en moet er rekening gehouden worden met ethische en juridische kaders.

Het vliegwiel op gang krijgen is een ketenuitdaging; niemand kan dit alleen. Hierbij spelen de overheid, het bedrijfsleven en kennisinstellingen een belangrijke rol.

WAT KAN DE OVERHEID DOEN?

Voor eenvoudige data-analyses op gevoelige data kunnen organisaties direct aan de slag. Door hierin het voortouw te nemen, kunnen overheidsorganisaties als launching customer een belangrijke functie in het innovatie-ecosysteem vervullen. Door actief de ontwikkeling en toepassing van innovatieve oplossingen voor eigen maatschappelijke vraagstukken te stimuleren draagt zij bij aan verdere praktische inzetbaarheid van deze technieken. Daarnaast kan de overheid samenwerking op dit terrein stimuleren door te faciliteren en ruimte te bieden voor experimenten, zowel via financiële en organisatorische middelen en ondersteuning als via aangepaste regelgeving.

WAT KAN HET BEDRIJFSLEVEN DOEN?

Deze nieuwe technologieën bieden een kans waarde te creëren uit data die voorheen, door privacy en vertrouwelijkheidsafwegingen, ontoegankelijk was. Om als Nederland voorop te lopen is het van belang dat bedrijven deze kans benutten door de mogelijkheden in kaart te brengen en gaan experimenteren. TNO werkt binnen samenwerkingsverbanden als Techruption aan het toepassen van deze technologieën. Techruption is een Nederlandse publiek private samenwerking waarin zowel kleine als grote bedrijven, startups en kennisinstellingen aan innovaties en praktische toepassing van privacy vriendelijke data analyse technologieën werken. Afhankelijk van de scope kunnen organisaties zich aansluiten bij bestaande programma's, of TNO betrekken om mee te denken over de opzet van nieuwe multidisciplinaire pilots.

Na de eerste pilot-ervaringen versnelt de adoptie als commerciële en overheids organisaties hun data beschikbaar stellen voor privacyvriendelijke databevraging door derden. Daarnaast zullen beleidsmakers de juridische kaders voor gebruik aan moeten scherpen en zijn technologieleveranciers essentieel voor het verder operationaliseren en opschalen van de benodigde technologieën. Ten slotte is het ook belangrijk dat kennisinstituten en universiteiten de methodes verder door ontwikkelen om zo de efficiëntie van privacyvriendelijke data analyses nog verder te vergroten.

VERDERE TOEPASSINGEN

1. OPTIMALISEREN VAN DE ZORG

De zorgkosten in Nederland zijn jaarlijks 100 miljard euro (10% van het BBP) [9] en stijgen naar verwachting naar ruim 170 miljard euro in 2040 [10]. Het is essentieel om de zorg te blijven verbeteren en daarnaast het zorgstelsel betaalbaar te houden. Daarvoor is inzicht nodig dat verborgen zit in de (patiënt) data van verschillende zorgorganisaties. Het delen van die data is echter onwenselijk vanwege privacy- of bedrijfsgevoeligheid. Technologieën als MPC en FL bieden hier uitkomst. Naast betaalbaarheid en toename van effectiviteit leidt inzet van deze technologieën tot betere preventieve inzichten, om zo te helpen voorkomen dat mensen ziek worden. Daarnaast biedt veilig combineren van en rekenen met data binnen de zorg mogelijkheden tot het ontwikkelen van nieuwe behandelmethoden. Waar nu nog vaak de focus van medisch specialisten op het uitsluiten van oorzaken (somatiek) ligt, zullen we op basis van data bewegen naar gericht definiëren, voorspellen en analyseren van oorzaken in een integrale benadering. Hieronder volgen een aantal concrete voorbeelden.

Optimale HIV-behandeling bepalen
HIV is een complex virus dat in vele vormen (mutaties) voorkomt. Een verkeerde behandeling kan ernstige gevolgen hebben. Gelukkig is er tegenwoordig veel over de verschillende HIV-behandelingen bekend. De effectiviteit en bijwerkingen kunnen echter substantieel verschillen van patiënt tot patiënt, aangezien ze in sterke mate bepaald worden door de exacte mutatie van het virus. Dit maakt het toewijzen van de juiste medicatie complex. Het is dan ook belangrijk om te leren van de effectiviteit en bijwerkingen van eerdere behandelingen. Door deze inzichten te gebruiken kunnen toekomstige HIV-patiënten nog beter geholpen worden; bijwerkingen worden geminimaliseerd en de kwaliteit van leven wordt verbeterd. Deze inzichten kunnen echter alleen verkregen worden door de analyses van gevoelige patiëntinformatie. Daarnaast is het belangrijk te voorkomen dat beslissingen van artsen op straat komen te liggen. Door de inzet van MPC heeft TNO samen met het CWI en de UvA laten zien dat de benodigde inzichten verkregen konden worden zonder de privacy van patiënten en artsen op te offeren [11].

Effectiviteit van zorginterventies verbeteren
Inzicht voortkomend uit analyses van data gecombineerd uit verschillende zorginstellingen kan een enorme bijdrage leveren aan het verbeteren van de zorg. Het delen van die data is echter onwenselijk vanwege privacy of bedrijfs-gevoeligheid. TNO ontwikkelt met partners in de zorg (een ziekenhuis, CZ zorgverzekeraar en het CBS) het Care-for-Data platform, gebaseerd op Multi-Party Computation (MPC). Door de inzet van cryptografische technieken kunnen partijen statistische verbanden ontdekken en monitoren alsóf ze toegang hebben tot elkaars data, zonder data daadwerkelijk te delen of herleiden - niet met elkaar, niet met TNO, niet met andere partijen. Het Care-for-Data platform maakt analyse van zorgdata door diverse partijen mogelijk. Het doel is het meten van de effectiviteit en doelmatigheid van zorgtoepassingen.

In het pilot project lukt dit op een privacy vriendelijke manier. Een continue analyse in de praktijk vraagt nog om nader onderzoek, waarin geverifieerd kan worden of deze implementatie aan het CBS beleid, op het gebied van privacy en toegang tot data, voldoet.

Impact van kanker reduceren door slimme data-analyses
Kanker is één van de ziektes met de meeste impact in Nederland. Op dit moment zijn er meer dan 800.000 mensen die in de afgelopen 20 jaar de diagnose kanker kregen [12]. Hoewel veel van hen genezen, kampt een groot deel nog met de gevolgen van kanker en de behandeling – deze gevolgen zijn niet alleen lichamelijk, maar ook psychisch en sociaal. Behandelmethoden, zoals immunotherapie, zijn niet altijd effectief en kunnen bij verkeerde inzet leiden tot onnodige kosten. Naast de impact op de patiënt, bedragen de kosten van kanker bijna EUR 6 miljard per jaar, oftewel 7% van alle zorguitgaven in Nederland [12]. Het combineren van data van grote groepen kankerpatiënten kan leiden tot nieuwe inzichten en betere behandelmethoden en daarmee de impact van kanker reduceren, de kans op genezing verhogen en kanker voorkómen. TNO en Integraal Kankercentrum Nederland (IKNL) onderzoeken of en hoe AI kan worden toegepast via MPC en FL, om te leren van verschillende databronnen zonder de privacy van de patiënten te schaden. Hierbij is er actieve interesse en betrokkenheid van organisaties uit de Life Sciences-hoek.

“Het combineren van data van grote groepen kankerpatiënten kan leiden tot nieuwe inzichten en betere behandelmethoden en daarmee de impact van kanker reduceren.”

2. BESTRIJDEN VAN FINANCIËLE EN ECONOMISCHE CRIMINALITEIT

Financiële en economische criminaliteit, zoals witwassen en fraude, is een complexe dreiging die jaarlijks miljoenen EU-burgers en duizenden bedrijven in de EU treft. Daarnaast zorgen deze activiteiten voor financiering van andere georganiseerde criminaliteit. Om financiële criminaliteit effectiever op te sporen is het essentieel dat organisaties informatie en data met elkaar kunnen delen. Tegelijkertijd mag de privacy van goedaardige burgers niet geschonden worden. Privacyverbeterende technologieën als MPC en FL bieden een oplossing voor deze ogenschijnlijke tegenstrijdigheid. Hieronder twee concrete voorbeelden.

Witwasdetectie
TNO werkt samen met diverse Nederlandse banken om MPC in te zetten voor gezamenlijke witwasdetectie. Jaarlijks worden er wereldwijd honderden miljarden euro's witgewassen, waarvan een geschatte 16 miljard euro in Nederland. Hoewel banken en andere financiële instellingen hard werken aan het opsporen van witwasactiviteiten, blijft een groot deel nog onder de radar. Er wordt geschat dat minder dan 1% van de criminele geldstromen in beslag genomen worden [13]. Een grote uitdaging is dat criminelen vaak gebruik maken van opeenvolgende transacties via meerdere banken. Iedere bank ziet daardoor maar een stukje van de puzzel en moet op basis van incomplete informatie mogelijke witwasactiviteiten doorgeven aan financiële opsporingsdiensten. Dit leidt tot een groot aantal meldingen met een hoge kans op een vals alarm. Om witwasdetectie te verbeteren is een samenwerking tussen de banken daarom zeer waardevol. MPC stelt banken in staat gezamenlijk verdachte geldstromen te detecteren zonder onderling persoonsgegevens of andere gevoelige data te delen [2] [14].

Fraudedetectie
Fraude heeft veel impact op de Nederlandse maatschappij. Het werkt ontwrichtend en kost de overheid, ondernemingen en de burger veel geld. Bovendien heeft fraude vaak impact op zwakkeren in de samenleving doordat geld niet terechtkomt waar het hoort. Denk aan frauderende zorgaanbieders die er voor zorgen dat hulpbehoevenden niet de zorg krijgen waar ze recht op hebben. Ook tijdens de corona pandemie zien we dat er misbruik gemaakt wordt van subsidies van de overheid aan bedrijven. Om fraude beter te kunnen opsporen moet er meer informatie worden uitgewisseld tussen zowel bedrijven als overheidspartijen. Aan de andere kant laten recente rechterlijke uitspraken [15] zien dat het combineren van informatie al snel kan leiden tot het schenden van privacy van burgers. MPC en FL bieden mogelijkheden om heel gericht inzichten te verkrijgen uit data van deze partijen zonder de uitwisseling van gevoelige data. Bovendien zorgen deze technieken ervoor dat alleen vooraf afgesproken analyses uitgevoerd kunnen worden. Hierdoor wordt het oneigenlijk gebruik van persoonsgegevens tegengegaan.

3. BETERE DIENSTVERLENING VOOR BURGERS VANUIT DE OVERHEID

De overheid heeft beschikking tot veel data. Met behulp van deze gegevens kan de overheid in potentie burgers en het bedrijfsleven beter bedienen. Data worden echter nog niet ten volste benut omdat de privacy van burgers gewaarborgd dient te blijven. Bovendien is het zomaar delen van grote hoeveelheden data van alle Nederlanders zelden proportioneel ten opzicht van het doel. Hieronder twee concrete voorbeelden.

Recht op AIO-uitkering

Voor mensen die geen volledige AOW hebben opgebouwd is er de aanvullende inkomensvoorziening ouderen. Deze AIO-uitkering wordt verstrekt door de Sociale Verzekeringsbank (SVB), maar moet door de gerechtigden zelf aangevraagd worden. Uit onderzoek van de Algemene Rekenkamer blijkt dat tienduizenden huishoudens in 2017 recht hadden op AIO maar daarvan niet op de hoogte waren. De SVB kan deze mensen niet gericht benaderen omdat zij niet beschikt de inkomensgegevens die nodig zijn om te bepalen of iemand in aanmerking komt voor de AIO uitkering. Bovendien is het niet proportioneel om inkomensgegevens van het UWV op grote schaal met het SVB uit te wisselen. Dit zou namelijk betekenen dat er ook inkomensgegevens gedeeld worden van mensen die niet tot de doelgroep van potentiële AIO-gerechtigden horen. Met behulp van MPC wordt de SVB in staat gesteld inkomensgegevens te analyseren en zo veel gerichter potentiële AIO-gerechtigden te benaderen, zonder toegang te krijgen tot deze inkomensgegevens. Op deze manier houdt de UWV de controle over haar data; de data kan niet zomaar voor andere doeleinden ingezet worden. Daarnaast leert de UWV niet wie er eventueel recht heeft op een AIO uitkering; de uitkomst van de analyse kan alleen door het SVB ingezien worden. Hierdoor is de privacy van burgers gewaarborgd. In een lopend onderzoekstraject van TNO, wordt MPC eerst beproefd met nagemaakte testdata voordat de toepassing in een pilotomgeving plaatsvindt.

Inzicht in armoede verbeteren

Om effectief en goed onderbouwd armoedebeleid te kunnen vormen is het essentieel meer inzicht te krijgen in de vele dimensies van armoede. Analyses van gegevens van instanties zoals gemeenten, woningcorporaties, CBS, zorgverzekeraars, energiemaatschappijen en anderen kunnen daar bij helpen. Maar men kan, mag en wil gegevens van burgers niet zomaar delen – het is belangrijk dat privacy gewaarborgd wordt en dat er op een ethische manier met persoonsgegevens wordt omgegaan. Binnen een lopende samenwerking met het CBS, de gemeente Heerlen, de Universiteit Maastricht, de Brightland Campus in Heerlen en andere partijen verkent TNO de toepassingsmogelijkheden van MPC in het kader van armoedebeleid. Zo zouden potentiële inzichten die via MPC uit data verkregen worden, kunnen worden vertaald naar beleid dat gericht is op een bredere aanpak van de problematiek, zoals het investeren in het laaghouden van energiekosten om armoede te bestrijden of het investeren in omgevingsfactoren die aantoonbaar bijdragen aan het oplossen van het armoedeprobleem.

⟩ TECHNISCHE VERDIEPING

1. MULTI-PARTY COMPUTATION

Multi-Party Computation (MPC) is een verzameling cryptografische technieken die het voor meerdere partijen mogelijk maakt op een gedecentraliseerde manier analyses en berekeningen uit te voeren op gevoelige data. Hierbij worden de privacy en vertrouwelijkheid van de gevoelige input data beschermd. Alleen de uitkomst van de analyse wordt onthuld, de onderliggende data blijven verborgen.

Het klassieke voorbeeld van een MPC berekening is het ‘miljonairsprobleem’. Hierbij wil een groep miljonairs bepalen wie het meest vermogend is, zonder daarbij hun exacte vermogen te onthullen. De traditionele manier om deze analyse uit te voeren vereist dat een vertrouwde partij alle input data verzamelt. In dit geval is deze partij dus op de hoogte van alle vermogens en is daarmee volledig verantwoordelijk voor de privacy en de vertrouwelijkheid. MPC geeft een alternatieve oplossing door deze partij te vervangen door een cryptografisch protocol. Hierbij blijven de data in handen van de eigenaren en hoeven ze niet centraal verzameld te worden.

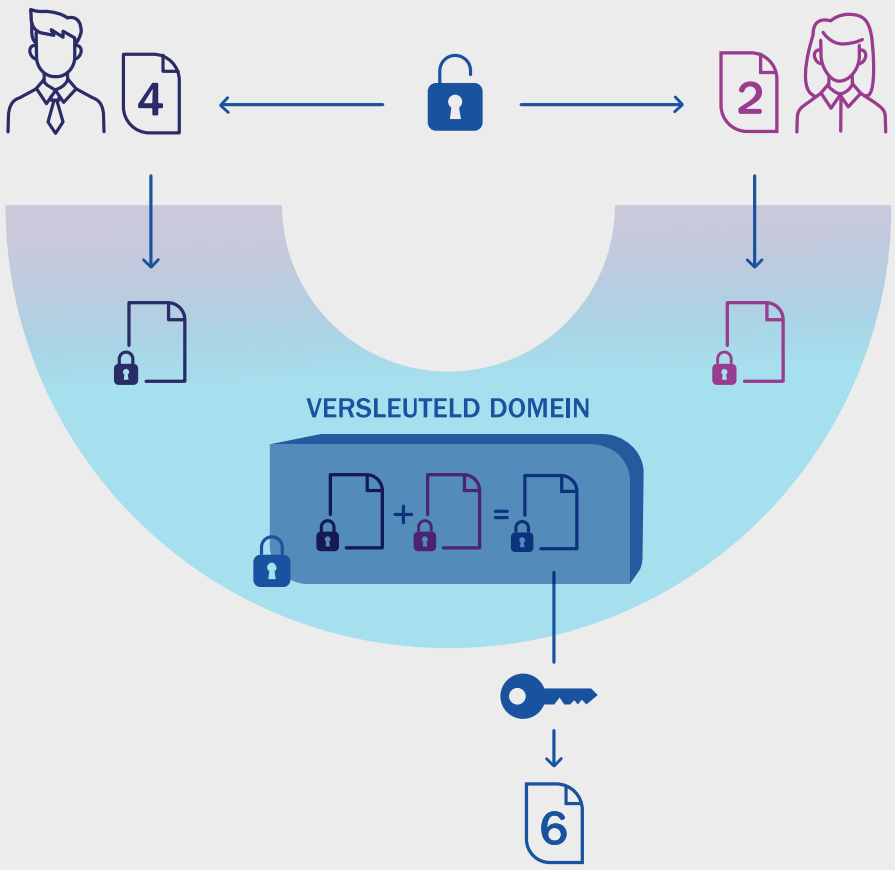
Met behulp van MPC is het bovendien mogelijk om veel complexere analyses op gedistribueerde data uit te voeren. Er bestaan verschillende cryptografische MPC-oplossingen. Hier behandelen we er twee: share-compute-reveal en homomorfe encryptie.

Share-Compute-Reveal
Een van de manieren om dit voor elkaar te krijgen is door gebruik te maken van de techniek ‘secret-sharing’. Bij secret-sharing wordt geheime data opgedeeld in stukjes (shares). Dit gebeurt zodanig dat één enkele share geen informatie over de geheime data bevat. De shares kunnen daarom verspreid worden over de deelnemende partijen zonder daarmee de geheime data prijs te geven. Ironisch genoeg betekent secret-sharing dus niet dat een geheim gedeeld wordt met andere partijen. Alle partijen distribueren op deze manier de shares van hun eigen input data. De tweede stap is het uitvoeren van de analyse. In plaats van één partij die de analyse op alle data uitvoert, voeren alle partijen dezelfde analyse uit op de shares die ze van de andere partijen hebben ontvangen. Alle partijen krijgen daarmee een andere uitkomst waar nog niks zinnigs uit af te leiden is. Pas als de partijen deze lokale tussenuitkomsten samenbrengen kan het analyseresultaat onthuld worden. Dit is de derde en tevens laatste stap van deze MPC-aanpak. Deze drietrapsaanpak wordt ook wel de ‘share-compute-reveal’ aanpak genoemd.

Homomorfe Encryptie

Een andere manier om de bovenstaande MPC-functionaliteit te realiseren is door gebruik te maken van homomorfe encryptie. Een homomorf encryptieprotocol maakt gebruik van een publieke en een private sleutel. De publieke sleutel is bij iedereen bekend en kan door de partijen gebruikt worden om data te versleutelen. De versleuteling beschermt de onderliggende data en alleen met behulp van de private sleutel kan deze opgeheven worden. Alle partijen kunnen daarom hun eigen data versleutelen en de versleutelde data met elkaar delen. De homomorfe eigenschap zorgt ervoor dat de analyses ook op de versleutelde data uitgevoerd kunnen worden. Pas als de analyses uitgevoerd zijn wordt de versleuteling, met behulp van de private sleutel, opgeheven. Tijdens alle tussenstappen blijven de data daarom versleuteld en worden er geen geheimen onthuld.

Let wel op dat de partij die de private sleutel in handen heeft alle versleutelde data kan ontsleutelen. Deze sleutel geeft dus erg veel macht en is een potentieel privacy risico. Het is dus cruciaal dat er op de juiste manier met deze sleutel wordt omgegaan. Een veel gebruikte aanpak is dat de private sleutel in stukjes verdeeld wordt, zodat niet één partij de gehele sleutel in handen heeft. Hier komt de eerder genoemde secret-sharing techniek weer bij kijken.



HOMOMORFE ENCRYPTIE: REKENEN MET VERSLEUTELDE DATA.

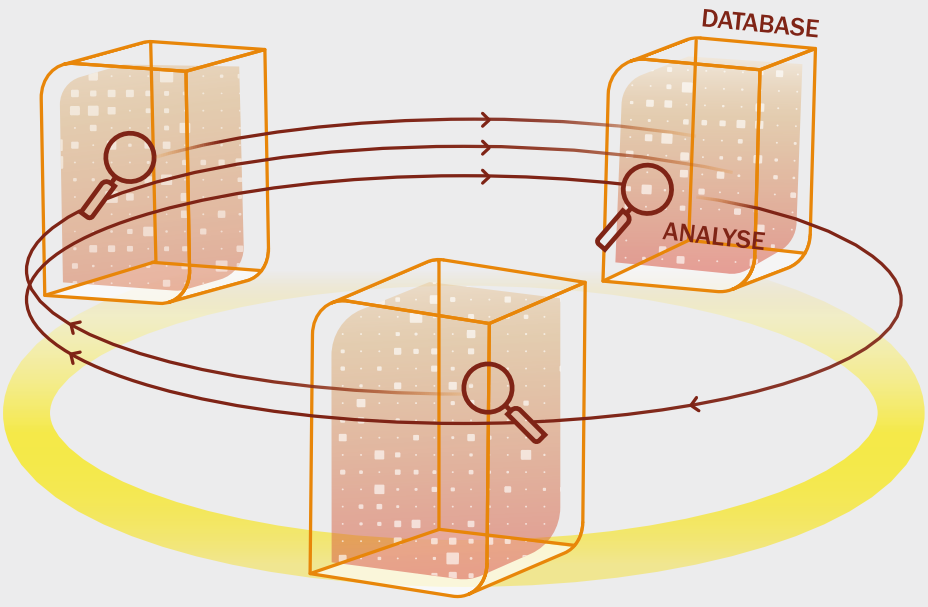
2. FEDERATED LEARNING

Federated Learning (FL) technieken maken het mogelijk machine learning te doen op verspreide gefedereerde gegevens, waarbij de data bij de eigenaar blijven.

Machine learning is een veelgebruikte manier om informatie of kennis uit data te halen, en vormt een onderdeel van AI. Een met machine learning geleerd model kan bijvoorbeeld helpen voorspellen of iemand kredietwaardig is bij hypotheek-aanvragen, of iemand op korte termijn diabetes zal ontwikkelen, of wie er mogelijk fraudeurs zijn. Hiermee belichaamt het model kennis over dergelijk onderscheid, zoals welke eigenschappen de diabetes prognose ondersteunen en waaraan fraudeurs te herkennen zijn.

Om kwalitatief hoogwaardige machine learning modellen te verkrijgen is het nodig zoveel mogelijk gegevens te gebruiken voor het trainen van het model. Standaard machine learning algoritmen vereisen de beschikbaarheid van trainingsgegevens op één machine of in een datacenter. In het geval het persoonsgegevens of anderszins vertrouwelijke gegevens betreft, is verplaatsen van de data naar een centrale database voor machine learning ongewenst of zelfs onwettig.

FL technieken maken het mogelijk om op verspreide databases machine learning toe te passen: De algoritmen trainen lokale modellen op de verspreide databases en combineren deze tussenresultaten tot een globaal model. Veelal herhaalt dit trainingsproces zich een aantal keer tot een definitief model is bereikt.



FEDERATED LEARNING

Omdat gedurende het trainen tussenresultaten samengevoegd worden, is het mogelijk dat hieruit delen van de gegevens afleidbaar zijn. Afhankelijk van de vereiste mate van vertrouwelijkheid kan het nodig zijn cryptografische technieken in te zetten, zoals de hierboven beschreven MPC.

FL is in het bijzonder geschikt voor het analyseren van persoonsgegevens, of andere data, die horizontaal gepartitioneerd zijn. Persoonsgegevens zijn horizontaal gepartitioneerd over verschillende partijen wanneer de partijen hetzelfde type gegevens bezitten maar van andere personen. Partij 1 weet bijvoorbeeld de leeftijd en het geslacht van persoon A en partij 2 weet de leeftijd en het geslacht van persoon B. Daarentegen spreken we van verticaal gepartitioneerde data wanneer de partijen verschillende informatie hebben over dezelfde personen. Bijvoorbeeld wanneer partij 1 de leeftijd van personen A en B weet en partij 2 het geslacht van personen A en B weet. Wanneer data verticaal gepartitioneerd is zullen we eerder uitwijken naar technieken als MPC.

3. DE PERSONAL HEALTH TRAIN

De *personal health train* (PHT) [7] is een oplossing waarmee data niet naar de analyse worden gebracht, maar de analyse (als “trein”) via een technische infrastructuur (“rails”) naar de verschillende databronnen (“stations”) wordt gebracht. Zo is het mogelijk om complexe algoritmes los te laten op data die beheerd worden door verschillende organisaties (ziekenhuizen), zonder dat deze data centraal verzameld hoeven te worden. Een netwerk van onderzoeksinstellingen, waaronder TNO, en partijen in de zorg werkt aan de ontwikkeling van de PHT. Hierbij wordt met name FL ingezet, maar er lopen steeds meer initiatieven om ook MPC onderdeel te maken van de PHT om zo de gevoelige data extra te beveiligen. Eind 2020 heeft de PHT de Computable Award gewonnen in de categorie Zorgproject.

“Met de Personal Health Train is het mogelijk om complexe algoritmes los te laten op data die beheerd worden door verschillende organisaties, zonder dat deze data centraal verzameld hoeven te worden.”

4. HUIDIGE STAND VAN ZAKEN IN TECHNOLOGIE EN MARKT

Privacyverbeterende technologieën zoals MPC en FL vereisen vaak een flinke rekenkracht en/of een goede communicatie-infrastructuur. Hierdoor waren de eerste MPC-protocollen, ontwikkeld in de jaren ‘80, voornamelijk theoretisch en slechts beperkt toepasbaar in de praktijk.

De recente technologische ontwikkelingen en protocolverbeteringen hebben dit beeld echter volledig doen veranderen en in 2008 werd MPC voor het eerst op grote schaal toegepast. Door gebruik te maken van MPC kon het Deense bedrijf Partisia suikerbietveilingen organiseren, zonder te hoeven vertrouwen op een externe partij die alle bied- en laatprijzen verwerkt. Sindsdien zijn er tientallen bedrijven opgericht die zich volledig richten op het uitrollen van privacyverbeterende technieken in specifieke toepassingsdomeinen. Het Sharemind platform van Cybernetica stelt gebruikers in staat op een privacyvriendelijke manier statistische analyses uit te voeren. Daarnaast ontwikkelt Unbound MPC-gebaseerde oplossingen voor cryptografisch sleutelmanagement.

Kortom, vandaag hebben privacyverbeterende technologieën dus al meetbare impact. Tegelijkertijd zit ook de academische wereld niet stil en worden er steeds efficiëntere protocollen ontwikkeld. Bij TNO werken we aan nieuwe toepassings-mogelijkheden van privacyverbeterende technologieën en brengen wij wetenschappelijk resultaten naar de praktijk. Hiermee vervullen wij een brugfunctie tussen de academische wereld, de overheid en het bedrijfsleven.

> “Het is tijd voor een nieuw uitgangspunt op het delen van gevoelige data: deel geen data, maar benut inzichten uit verspreide databronnen terwijl privacy en vertrouwelijkheid gewaarborgd worden.”

BIBLIOGRAFIE

[1] Raming november 2020, cijfers," CPB, 26 November 2020. [Online]. Available: https://www.cpb.nl/raming-november-2020-vooruitzicht-2021#docid-160397.

[2] OECD, „Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies," OECD Publishing, Paris, 2019.

[3] Onderzoek McKinsey: Economische en maatschappelijke kansen van AI voor Nederland," 7 November 2020. [Online]. Available: https://nlaic.com/nieuws/onderzoek-mckinsey-economische-en-maatschappelijke-kansen-van-ai-voor-nederland/.

[4] European Union: European Comission, „Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act)," 25 November 2020, COM(2020) 767 final.

[5] EU, „GAIA-X," 2020. [Online]. Available: data-infrastructure.eu.

[6] A. Cavoukian, „Privacy by design: The 7 foundational principles," Information and privacy commissioner of Ontario, Canada, vol. 5, 2009.

[7] https://pht.health-ri.nl/," Health RI, 2020. [Online]. Available: https://pht.health-ri.nl/.

[8] F. Bomhof en P. Giezeman, „Data gebruiken zonder ze te krijgen of te zien: Hoe we zonder privacyschending informatie verkrijgen in de zoektocht naar onvindbare veroordeelden.," Ministerie Justitie en Veiligheid, Den Haag, 2019.

[9] Zorguitgaven stegen in 2019 met 5,2 %," CBS, 11 June 2020. [Online]. Available: https://www.cbs.nl/nl-nl/nieuws/2020/24/zorguitgaven-stegen-in-2019-met-5-2-procent.

[10] Trendscenario VTV-2018 identificeert maatschappelijke opgaven voor de toekomst," RIVM, 5 July 2017. [Online]. Available: https://www.rivm.nl/nieuws/trendscenario-vtv-2018-identificeert-maatschappelijke-opgaven-voor-toekomst.

[11] T. Attema, E. Mancini, G. Spini, M. Abspoel, J. de Gier, S. Fehr, T. Veugen, M. van Heesch, D. Worm, A. De Luca, R. Cramer en P. M. A. Sloot, „A New Approach to Privacy-Preserving Clinical Decision Support Systems for HIV Treatment," CoRR, arxiv.org/abs/1810.01107, 2018.

[12] IKNL, „Kanker & leven," IKNL, [Online]. Available: https://iknl.nl/kanker-en-leven.

[13] United Nations Office on Drugs and Crime (UNODC), „Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes," 2011.

[14] F. o. F. I. S. (FFIS), „Innovation and discussion paper: Case studies of the use of privacy preserving analysis to tackle financial crime," 8 January 2021. [Online]. Available: https://www.future-fis.com/uploads/3/7/9/4/3794525/ffis_innovation_and_discussion_paper_-_case_studies_of_the_use_of_privacy_preserving_analysis_-_v.1.3.pdf.

[15] SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens," rechtspraak.nl, 5 Februari 2020. [Online]. Available: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-wetgeving-in-strijd-met-het-Europees-Verdrag-voor-de-Rechten-voor-de-Mens.aspx.

[16] A. Sangers, M. van Heesch, T. Attema, T. Veugen, M. Wiggerman, J. Veldsink, O. Bloemen en D. Worm, „Secure Multiparty PageRank Algorithm for Collaborative Fraud Detection," in International Conference on Financial Cryptography and Data Security, Cham, 2019.

[17] S. Biswas, B. Carson, V. Chung, S. Singh en R. Thomas, „AI-bank of the future: Can banks meet the AI challenge?," McKinsey, 11 September 2020. [Online]. Available: https://www.mckinsey.com/industries/financial-services/our-insights/ai-bank-of-the-future-can-banks-meet-the-ai-challenge.

[18] EU White Paper On Artificial Intelligence - A European Approach to excellence and trust," European Commission, 19 Februari 2020. [Online]. Available: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[19] Volksgezondheid en zorg - zorguitgaven kanker," 2017. [Online]. Available: https://www.volksgezondheidenzorg.info/onderwerp/kanker/kosten/zorguitgaven.

[20] Europol en Commissie lanceren Europees centrum voor de bestrijding van financiële en economische misdaad," EC, 5 June 2020. [Online]. Available: https://ec.europa.eu/netherlands/news/europol-en-commissie-lanceren-europees-centrum-voor-de-bestrijding-van-financi%C3%ABle-en_nl.

AUTEURS:
Thomas Attema, Daniël Worm

REVIEWERS:
Freek Bomhof, Timon Brussaard, Martine van de Gaar-Velzeboer, Paul Havinga, Elena Lazovik, Herman Pals, Alex Sangers, Tjerk Timan, Cor Veenman, Pieter Verhagen, Berry Vetjens, Henk-Jan Vink, Peter Werkhoven

CONTACT:
Thomas Attema
thomas.attema@tno.nl

TNO innovation for life

TNO.NL

20-12046