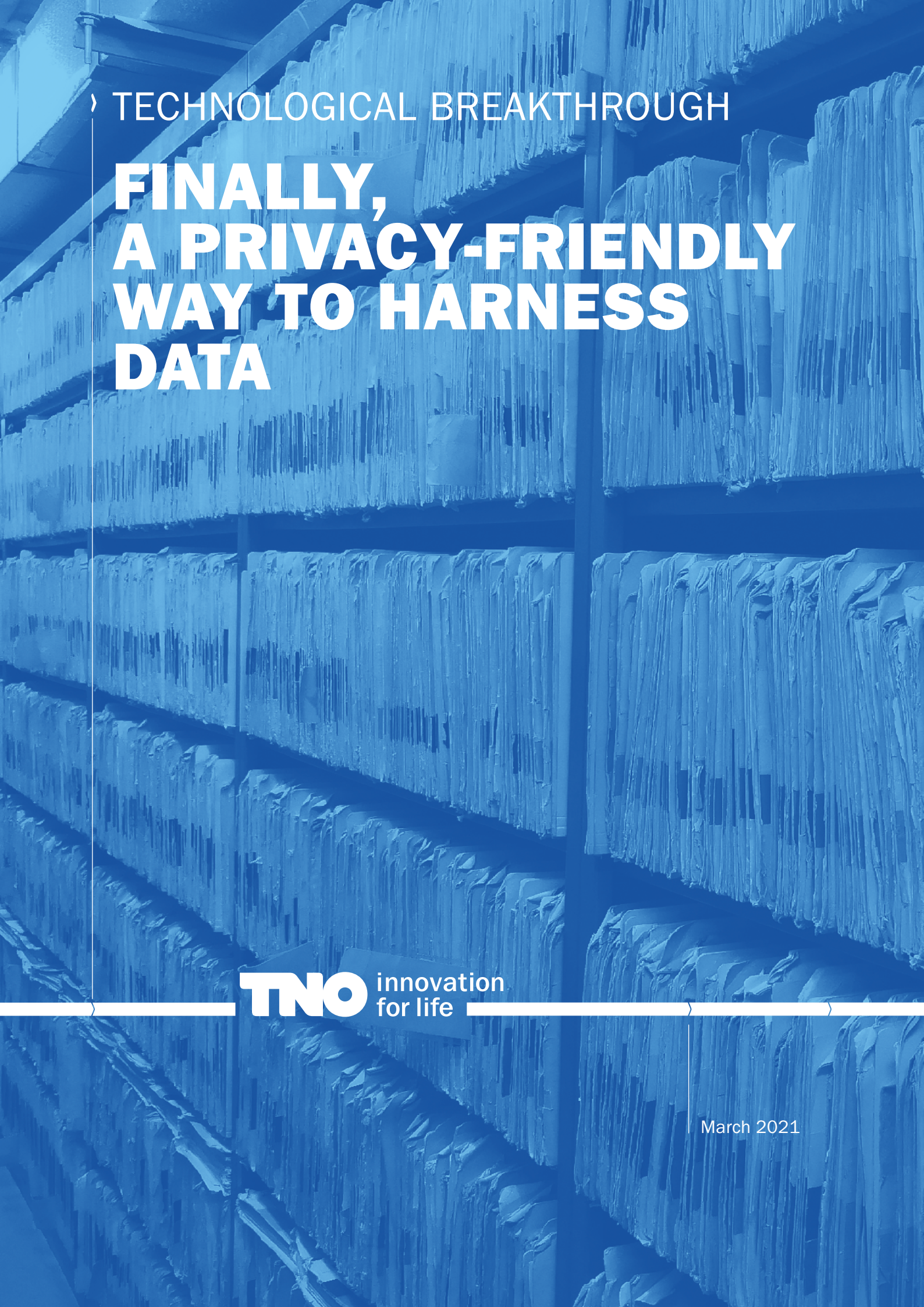


TECHNOLOGICAL BREAKTHROUGH

FINALLY, A PRIVACY-FRIENDLY WAY TO HARNESS DATA

TNO innovation for life

March 2021

Can you envision it? You have a medical condition for which you need medication. However, the effectiveness and side effects of available treatments differ from patient to patient. By using insights into the effectiveness and side effects among other patients, you can be better treated. But this does require analyses of sensitive patient data. Using innovative technology, the necessary insights can be obtained without compromising the privacy of patients and doctors. As a result, you and your doctor can immediately choose the best treatment for you.

Can you envision it? You work at a bank in the money laundering detection department. Your bank, like other banks and financial institutions, is committed to detecting money laundering activities. Yet a large part remains under the radar (99%!), because many criminals make use of successive transactions through multiple banks. You therefore only see one piece of the puzzle and must act accordingly. Fortunately, innovative technology now offers you the chance to detect suspicious money flows together with other banks without sharing personal or other sensitive data.

We envision this with the privacy-friendly use of data. The analysis of linked data sources makes it possible to tackle major innovation and social challenges and to realise economic growth. However, data sharing has not yet taken off due to commercial and/or legal barriers, including the fundamental right to privacy. But what if you don't need to share any data at all to gain insights?

We believe it is time for a new approach to the processing of sensitive data: let's move away from the traditional centralised view and start using distributed data processing. Because to create value from data, you don't need to own it. So don't share data, but leverage insights from distributed data sources while ensuring privacy and confidentiality. With innovative technologies such as Multi Party Computation (MPC) and Federated Learning (FL), this goal can be achieved without compromising privacy and confidentiality. In this paper, we explain how we envisage this, and give examples of applications.

Because this is a top-notch chain challenge, we call on government, industry and knowledge institutions to join forces in public-private partnerships and to accelerate the upscaling of this technology by opening up their data sources to experimentation. We cordially invite you to elaborate this approach together and to test it in practice.

Can you envision it?

SUMMARY

Data sharing and analysis are essential when it comes to achieving economic growth and solving societal challenges. However, data sharing is yet to really get off the ground due to commercial and/or legal barriers, including the fundamental right to privacy. Innovative technologies such as Federated Learning and Multi-Party Computation offer a way to address this issue by securely learning from sensitive data from multiple sources without having to share this data.

It's not just Google and Facebook – practically everything and everyone around us is collecting an increasing amount of data. Examples include thermostats, smartwatches, movement apps and navigation systems, but hospitals, banks, the logistics sector and other organisations are also trying to use data to improve their services. Additionally, (smart) industry is already making more and more using data to improve the sustainability and efficiency of existing production processes, which are increasingly taking place in chains. To accomplish this, it is necessary to combine and analyse various data sources (from these chains). Data sharing is the key to the successful application of techniques such as Artificial Intelligence (AI). The same applies to achieving economic growth, improving healthcare while keeping it affordable, tackling crime and increasing labour productivity.

But there is also a downside. There are concerns over the large scale collection and sharing of data. Large platforms in which personal data is centrally managed and processed without consent do not match our European values. These platforms result in monopoly situations in which personal data can easily be misused. Privacy is one of our fundamental rights and the protection of personal data is laid down in the Dutch and European GDPR regulations. These laws and regulations are often cited as a barrier to the optimal use of data (sharing). Furthermore, organisations do not want to simply hand over their commercially sensitive data; as an organisation, they want to keep control over which data is shared with whom while simultaneously utilising the value of this data.

However, the goal is not to collect or share data. The goal is to arrive at new insights by *learning from data*. What if you could achieve these insights without sharing any data at all?

Technologies such as Multi-Party Computation (MPC) and Federated Learning (FL) enable this, allowing insights to be gained by linking multiple data sources without compromising privacy or confidentiality. MPC makes use of cryptography. This technique ensures that analyses can be performed on encrypted data. The (often sensitive) underlying data does not need to be shared in order to perform analyses and gain new insights. As a result, no sensitive information is leaked to other parties and only previously specified operations and analyses are possible. With FL, the analysis is brought to the data instead of the other way around. The data owner thus retains control over the use of the data as it remains secure and decentralised. At the same time, it remains possible to create value from the data. The potential of these technologies for our society is enormous.

The first solutions based on MPC and FL are now technologically mature and are being applied in various domains. Through these techniques, privacy and confidentiality can be safeguarded to a much stronger degree. This enables a far-reaching form of data minimisation, as encouraged by the GDPR. Only the analysis results are shared. In cooperation with lawyers and ethicists, the sensitivity of these results will always have to be weighed against relevant regulations in order to guarantee correct use and proportionality. Because only the results are shared, this legal and ethical assessment will be successful in many more applications than the traditional approach in which the underlying data is also shared. MPC and FL show that it is time for a new starting point in the sharing of sensitive data: *do not share data, but harness insights from distributed data sources while guaranteeing privacy and confidentiality*. In this way, we can securely achieve further economic growth and solve important societal challenges.

It is essential that the government, private parties and other organisations accelerate the upscaling of this technology by opening up their own data sources to privacy-friendly analyses and by encouraging public-private partnerships to speed up operationalisation. The starting point is that data sharing is not necessary and that privacy and confidentiality can therefore be guaranteed. A multidisciplinary approach is crucial in order to further develop and scale up the technologies, as well as to give further substance to the ethical and legal frameworks and standards.

No one can initiate this flywheel for the economy and society on their own; this challenge has to be tackled collaboratively. We therefore call on governmental parties, companies, commercial technology parties and knowledge institutions to work together to get these new technologies ready for practical use.

CONTENTS

THE VALUE OF SENSITIVE INFORMATION

The analysis of coupled data sources makes it possible to solve major innovation challenges, tackle societal challenges and achieve economic growth in the Netherlands. Rapid, successive developments, such as in Artificial Intelligence (AI), offer the possibility to convert growing amounts of data into useful information that leads to new applications and insights. For example, new medicines are being developed, logistical processes optimised, industrial production chains improved and fraudulent activities detected. However, in addition to legal and commercial barriers, societal concerns stand in the way of data exchange.

DATA AS THE BASIS FOR ECONOMIC GROWTH

Post-corona, we foresee a financial contraction which is estimated at over 4% of GDP for the Netherlands (CPB) [1]. One of our greatest challenges is therefore to achieve economic growth. The technological solution to this challenge is within reach: recent analyses show that the availability and exchanging of data can lead to an economic growth of 1.5% of GDP [2]. Some studies indicate that this growth can be as high as 4% of GDP when data from beyond just the public sector is used. The availability of data enables new AI applications that can increase the earning power of different industries by 30 to 128% [3]. The availability of the right data can create a streamlined labour market in which supply and demand are efficiently matched. Personalised healthcare not only keeps care affordable but also keeps Dutch people in the labour process for longer. Mobility challenges can be solved through extensive automation in the transport sector. AI also enables us to arm ourselves against the ever-growing threat of cyberattacks. These developments can take us from the post-corona era to a ‘golden decade’ of economic growth.

SOLVING SOCIETAL CHALLENGES WITH DATA

In addition to economic growth, societal impact can also be created through the smart use of data. Personalised healthcare is an example of this. In addition, government bodies can provide better services to citizens when data is used optimally, such as by offering help to people with financial problems who are entitled to specific benefits. We elaborate on these and other applications in the boxes.

ECONOMIC AND SOCIETAL POTENTIAL VERSUS PRIVACY AND CONFIDENTIALITY

In recent decades, rapid and unregulated technological developments have led to the creation of large platforms (such as those of Google and Facebook) in which enormous amounts of data converge. However, the power of the predominantly American parties that follow this traditional and centralised approach is increasingly coming into question. Large data platforms in which data is centrally managed and often processed without explicit consent do not match European values, including the fundamental right to privacy. European GDPR legislation protects personal data and presents serious obstacles to the centralised approach. Europe has also recently given clear signals that it wants to move away from the 'winner-takes-all' business models of data platforms which have been prominent up to now [4]. Within the GAIA-X initiative, for example, a secure, transparent and federated European data infrastructure is being developed [5]. In addition, data may contain competitively sensitive information, and companies often have commercial interests that stand in the way of data exchange.

All in all, there are strongly conflicting interests when it comes to data sharing: enormous economic and societal potential on one hand and privacy and confidentiality on the other. But this need not be a zero-sum game. The ultimate goals is to gain insights from data while safeguarding privacy[6]. New, innovative technologies such as Multi-Party Computation (MPC) and Federated Learning (FL) offer a solution. These techniques make it possible to perform analyses and calculations regarding multiple data sources without bringing these data together. In this way, different parties can collaboratively analyse data without actually being able to see each other's data.

THE RIGHT FOUNDATION FOR DATA PROCESSING REMAINS ESSENTIAL

MPC and FL are important techniques for designing data analysis applications in a privacy-friendly manner (Privacy by Design). They offer an opportunity for better data protection and proportional data processing that matches our norms and values. Various data protection risks, such as the improper use of personal data or the risk of hacks, can be significantly reduced through the use of MPC and FL. As a result, the privacy of individuals can be guaranteed and there is no need to hand over data to centralised data platforms in the USA. Together with new European regulations, these techniques therefore also play an important role in removing dependency on these platforms. As a result, the societal and commercial distrust of data sharing can be overcome within the framework of the regulations. MPC and FL have existed for some time but have only in recent years been developed to such an extent that they can now be used on a large scale and for more complex analyses.

INSIGHTS WITHOUT DATA SHARING: HOW DOES THIS WORK?

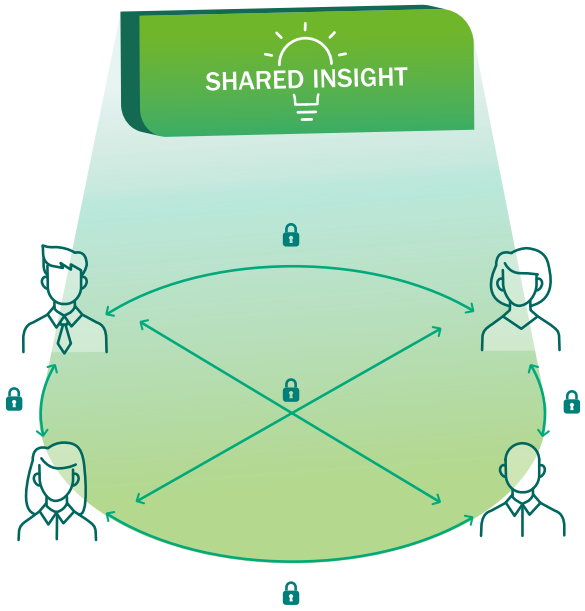
Both MPC and FL assume a scenario in which multiple parties want to perform a joint calculation or analysis based on their own data without having to share it. One example would be a hospital and a health insurance company with the common goal of providing the best care as efficiently as possible. In order to achieve this, they need each other's information, such as the treatment history of patients. However, the patient data that needs to be analysed is privacy-sensitive and cannot simply be exchanged.

Federated Learning

The traditional, privacy-unfriendly solution requires the data to be collected centrally in order to then perform the correct analyses. FL solves the privacy problem by bringing the analyses to the data instead of the data to the analyses. The analyses are broken down into small sub-calculations that can be performed locally by the various parties. After performing a local calculation, only the (intermediate) results are shared with one or more parties. The sensitive data is not shared with anyone and remains with the party. The Personal Health Train (PHT) [7] uses this solution to provide customised healthcare based on distributed data sources without having to collect the data centrally. FL can provide much stronger privacy and confidentiality guarantees than the traditional approach in which all data is collected in a central location.

Multi-Party Computation

Using advanced cryptographic techniques, MPC can protect sensitive data even better than FL. MPC protocols ensure that all data remains in an encrypted form. Calculations are performed on the data while it remains encrypted at all times, and only the outcome of the analysis is decrypted. As a result, MPC achieves the maximum degree of privacy and confidentiality; only the outcome of the analysis is revealed. A disadvantage is that MPC generally requires more computing power and/or stronger communication infrastructure than FL. We will discuss the trade-off between MPC and FL in more detail in the technical boxes.

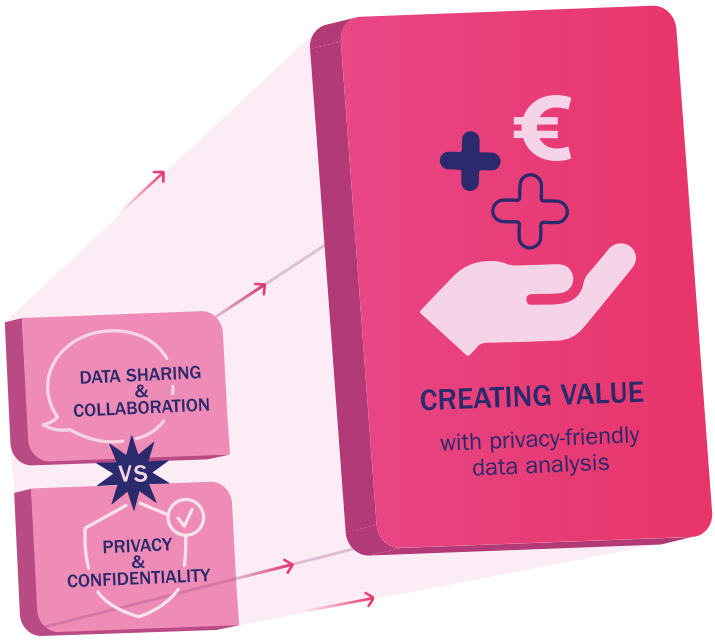


MULTI-PARTY COMPUTATION: SHARED INSIGHT WITHOUT SHARING DATA.

NEW PERSPECTIVE

The traditional view is that power derives from the possession of knowledge and data. In this view, the translation of data into valuable information requires a centralised approach in which one party holds all of the data. This approach is therefore diametrically opposed to interests such as confidentiality and privacy. It would therefore seem that a choice must be made between two conflicting interests.

Often, however, the aim is not to possess data but to create value from this data. Data therefore does not need to be brought together by one party but can remain locally managed. In this way, data owners maintain control over their data. Innovative technologies such as MPC and FL show that this goal can be achieved without compromising privacy and confidentiality. It is clearly time to move away from the traditional, centralised view and start using distributed data processing, as these technologies offer a European alternative to American and Chinese data platforms.



MPC AND FL PROVIDE A NEW PERSPECTIVE.

APPLICATION: FROM OPTIMISING HEALTHCARE TO PREVENTING FINANCIAL CRIME

There are many possible applications for privacy-enhancing techniques such as MPC and FL. The effectiveness of healthcare can be increased by gaining insights from patient data in a privacy-friendly manner. Increasing financial crime can be contained by securely coupling sensitive data from different financial organisations. In addition, the government can improve its services through privacy-respecting collaboration between different government bodies. These three application domains are further elaborated in the boxes.

Techniques such as MPC and FL can also be applied in the mobility sector. A move is taking place towards Mobility as a Service, an innovative concept in which the traveller is central and can make optimal use of a wide range of mobility modes via a platform. In order for the government to optimise and manage this new mobility, it is important that privacy and competitively sensitive data from various (competing) MaaS service and mobility providers (such as traveller data and current transport capacity) can be analysed and monitored. Another application of MPC and FL is the optimisation of logistic chains without the exchange of business-sensitive information. This allows competing parties to jointly create value. By enriching their market segmentation with data sources that are not yet available, companies can also market their products more effectively. Furthermore, privacy-friendly analyses can help in the effective detection of cyberattacks because complex attack patterns can be better identified. These technologies also offer opportunities in the security domain, such as in tracking down untraceable convicts [8].

> “MPC and FL allow competing parties to jointly create value.”

HOW DO WE INITIATE THE FLYWHEEL OF PRIVACY-FRIENDLY DATA ANALYSIS?

Technical and organisational challenges both need to be overcome in order to apply these technologies on a large scale and achieve the aforementioned benefits. In addition to the technical challenges, these solutions will, for example, have to be embedded in existing infrastructures in which sensitive data is stored in isolation. Ethical and legal frameworks also need to be taken into account.

Initiating the flywheel is a chain challenge; no one can do it alone. The government, industry and knowledge institutions all play an important role in this.

WHAT CAN THE GOVERNMENT DO?

For simple data analyses of sensitive data, organisations can get to work immediately. By taking the lead in this, governmental organisations can fulfil an important function in the innovation ecosystem as a launching costumer. By actively stimulating the development and application of innovative solutions for their own societal issues, they contribute to the further practical implementation of these techniques. In addition, the government can encourage collaboration in this field through facilitation and offering room for experiments. This can be done via financial and organisational resources and support as well as adapted regulations.

WHAT CAN INDUSTRY DO?

These new technologies offer an opportunity to create value from data that was previously inaccessible due to privacy and confidentiality considerations. In order for the Netherlands to lead the way, it is important that companies take advantage of this opportunity by identifying the possibilities and starting to experiment. TNO is working within partnerships like Techruption in order to apply these technologies. Techruption is a Dutch public-private partnership in which both small and large companies, start-ups and knowledge institutions work on innovations and the practical application of privacy-friendly data analysis technologies. Depending on the scope, organisations can join existing programmes or involve TNO in order to help set up new multidisciplinary pilots.

Following the first pilot experiences, adoption will be accelerated if commercial and governmental organisations make their data available for privacy-friendly data retrieval by third parties. In addition, policymakers will need to tighten the legal frameworks on usage, while technology suppliers are essential to the further operationalisation and upscaling of the required technologies. Finally, it is also important that knowledge institutions and universities continue to develop the methods in order to further increase the efficiency of privacy-friendly data analyses.

FURTHER APPLICATIONS

1. OPTIMISING HEALTHCARE

Annual healthcare costs in the Netherlands amount to 100 billion euros (10% of GDP) [9] and are expected to rise to over 170 billion euros by 2040 [10]. It is essential to continue to improve healthcare and keep the care system affordable. This requires insights which are hidden in the (patient) data of various healthcare organisations. However, it is undesirable to share this data due to privacy or business sensitivities. Technologies such as MPC and FL offer a solution here. Besides affordability and increased effectiveness, the use of these technologies leads to better preventive insights, thereby helping to prevent people from falling ill. In addition, the secure combination of data and calculations using data within healthcare offer opportunities to develop new treatment methods. Whereas medical specialists still often focus on *ruling out* causes (somatics), we will move towards the targeted definition, prediction and analysis of causes in an integrated approach based on data. Below are a number of concrete examples.

Determining optimal HIV treatment
HIV is a complex virus that comes in many forms (mutations). Improper treatment can have serious consequences. Fortunately, a lot is now known about the various HIV treatments. However, the effectiveness and side effects can differ substantially from patient to patient as these are strongly determined by the exact mutation of the virus. This makes assigning the right medication complex, so it is important to learn from the effectiveness and side effects of previous treatments. Using these insights, future HIV patients can be helped even better; side effects can be minimised and quality of life improved. However, these insights can only be gained by analysing sensitive patient information. In addition, it is important to prevent doctors' decisions from becoming public knowledge. Using MPC, TNO (together with the CWI and the UvA) has shown that the required insights can be obtained without sacrificing the privacy of patients and doctors [11].

Improving the effectiveness of healthcare interventions
Insights from analyses of data combined from different healthcare institutions can make a huge contribution to improving healthcare. However, it is undesirable to share that information due to privacy or business sensitivities. Together with partners in healthcare (a hospital, health insurer CZ and the CBS), TNO is developing the Care-for-Data platform based on Multi-Party Computation (MPC). Using cryptographic techniques, parties can discover statistical connections and carry out monitoring as if they have access to one another's data without actually tracing or sharing the data – not with each other, not with TNO, not with other parties. The Care-for-Data platform enables the analysis of healthcare data by various parties. The aim is to measure the effectiveness and efficiency of healthcare applications. This is done in a privacy-friendly manner in the pilot project. Continuous analysis in practice requires further research, through which it can be verified whether this implementation complies with CBS policy in the field of privacy and data access.

Reducing the impact of cancer through smart data analysis
Cancer is one of the most impactful diseases in the Netherlands. More than 800,000 people have been diagnosed with cancer in the last 20 years [12]. Although many of them have been cured, a large number still struggle with the consequences of cancer and its treatment – not only physical but also psychological and social. Treatment methods such as immunotherapy are not always effective and can lead to unnecessary costs if used incorrectly.
In addition to the impact on patients, cancer costs nearly EUR 6 billion per year, or 7% of all healthcare spending in the Netherlands [12]. Combining data from large groups of cancer patients can lead to new insights and better treatment methods and thereby reduce the impact of cancer, increase the chance of a cure and prevent cancer in the first place. TNO and the Netherlands Comprehensive Cancer Organisation (IKNL) are investigating if and how AI can be applied via MPC and FL in order to learn from different data sources without compromising patient privacy. There is active interest and the involvement of organisations from the Life Sciences field.

"Combining data from large groups of cancer patients can lead to new insights and better treatment methods and thus reduce the impact of cancer."

2. COMBATING FINANCIAL AND ECONOMIC CRIME

Financial and economic crime, such as money laundering and fraud, is a complex threat affecting millions of EU citizens and thousands of businesses every year. In addition, these activities provide funding for other organised crime. In order to detect financial crime more effectively, it is essential for organisations to be able to share information and data with one another. At the same time, the privacy of law-abiding citizens must not be violated. Privacy-enhancing technologies such as MPC and FL offer a solution to this apparent contradiction. Below are two concrete examples.

Money laundering detection

TNO is working with several Dutch banks to implement MPC for joint money laundering detection. Annually, hundreds of billions of euros are laundered worldwide, of which an estimated 16 billion takes place in the Netherlands. Although banks and other financial institutions work hard to detect money laundering activities, a large amount remains under the radar. It is estimated that less than 1% of criminal cash flows are seized [13]. A major challenge is that criminals often use successive transactions through multiple banks. Each bank therefore sees just one piece of the puzzle and has to pass on possible money laundering activities to financial investigation services on the basis of incomplete information. This leads to a large number of reports with high chances of a false alarm. Cooperation between banks is therefore very valuable when it comes to improving money laundering detection. MPC allows banks to jointly detect suspicious cash flows without sharing personal data or other sensitive data [2] [14].

Fraud detection

Fraud has a large impact on Dutch society. It has a disruptive effect and costs the government, businesses and citizens a lot of money. Furthermore, fraud often has an impact on the more vulnerable members of society because money does not end up where it should. Examples include fraudulent healthcare providers who ensure that people in need of help do not receive the care to which they are entitled. During the corona pandemic, we have also seen the misuse of government subsidies for companies. In order to detect fraud more effectively, more information must be exchanged between both companies and governmental parties. On the other hand, recent court rulings [15] have shown that combining information can quickly lead to a breach of privacy for citizens. MPC and FL offer possibilities to gain very specific insights from the data of these parties without exchanging sensitive data. Moreover, these techniques ensure that only analyses agreed upon beforehand can be carried out. This prevents the improper use of personal data.

3. BETTER SERVICES FOR CITIZENS FROM THE GOVERNMENT

The government has access to a lot of data through which it can potentially offer better services to citizens and businesses. However, data is not yet being fully exploited because the privacy of citizens must be guaranteed. In addition, the indiscriminate sharing of large amounts of data on all Dutch citizens is rarely proportionate to the goal. Below are two concrete examples.

Right to the AIO benefit

For people who have not built up a full AOW, there is the supplementary income provision for older people. This AIO benefit is provided by the Sociale Verzekeringsbank (SVB) but must be applied for by the beneficiaries themselves. Research by the Netherlands Court of Audit shows that tens of thousands of households were entitled to AIO in 2017 but were not aware of it. The SVB cannot approach these people specifically because it does not have the income data needed to determine whether someone is eligible for the AIO benefit. Furthermore, it would not be proportionate to exchange income data from the UWV with the SVB on a large scale. This would mean that the income data of people who do not belong to the target group of potential AIO recipients would also be shared. MPC enables the SVB to analyse income data and thus approach potential AIO recipients in a more targeted manner without having access to this income data. In this way, the UWV retains control over its data; the data cannot simply be used for other purposes. In addition, the UWV does not learn who is entitled to an AIO benefit; the outcome of the analysis can only be viewed by the SVB. This guarantees the privacy of citizens. In an ongoing research project by TNO, MPC is first being trialled with fake test data before it is applied in a pilot environment.

Improved understanding of poverty

To be able to formulate effective and well-founded poverty policies, it is essential to gain a better understanding of the many dimensions of poverty. Analyses of data from bodies such as municipalities, housing associations, the CBS, health insurers, energy companies and others can help in this. But one cannot, should not and does not want to simply share citizens' data – it is important that privacy be guaranteed and that personal data be handled ethically. Within an ongoing collaboration with the CBS, the municipality of Heerlen, Maastricht University, the Brightland Campus in Heerlen and other parties, TNO is exploring the application possibilities of MPC in the context of poverty policy. For instance, potential insights gained from MPC data could be translated into policies aimed at a broader approach to the problem, such as investing in keeping energy costs down to combat poverty or investing in environmental factors that demonstrably contribute to solving the poverty problem.

⟩ TECHNICAL ENHANCEMENT

1. MULTI-PARTY COMPUTATION

Multi-Party Computation (MPC) is a collection of cryptographic techniques that allow multiple parties to perform analyses and calculations with sensitive data in a decentralised manner. The privacy and confidentiality of the sensitive input data is protected. Only the outcome of the analysis is revealed; the underlying data remains hidden.

The classic example of an MPC calculation is the 'millionaire problem'. A group of millionaires want to determine who the wealthiest is without revealing their exact net worth. The traditional means of performing this analysis requires a trusted party to collect all input data. In this case, the party is therefore aware of all of the assets and is fully responsible for privacy and confidentiality. MPC provides an alternative solution by replacing this party with a cryptographic protocol. In this case, the data remains in the hands of the owners and does not need to be collected centrally.

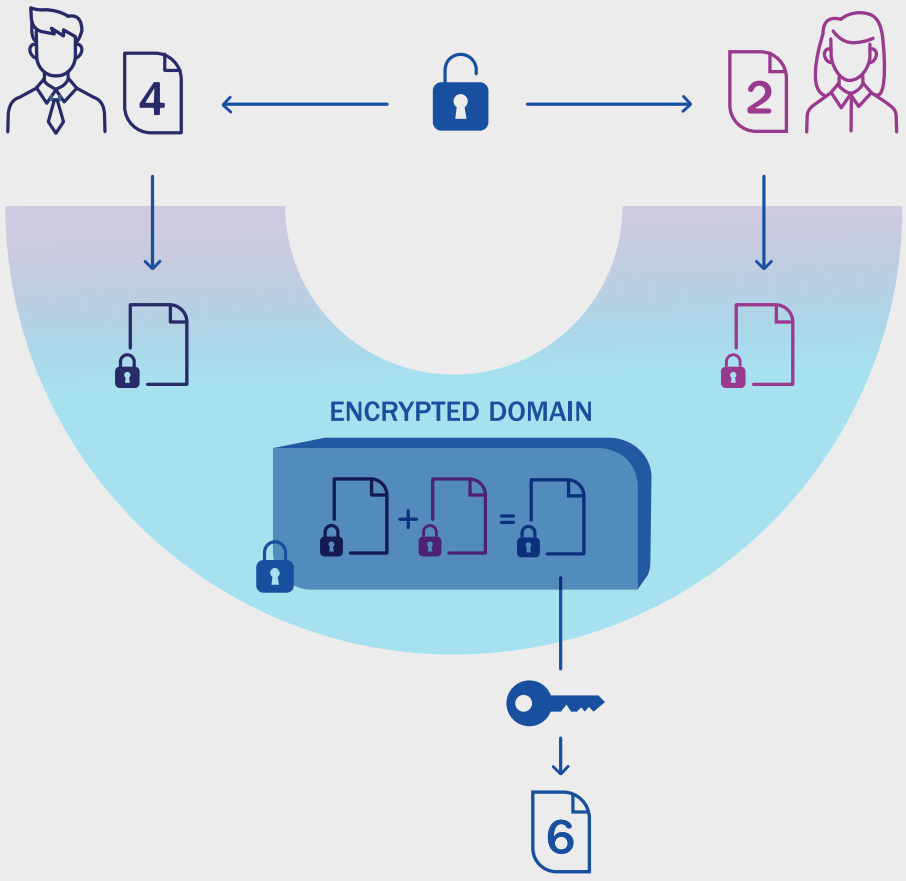
With MPC, it is also possible to perform much more complex analyses of distributed data. Various cryptographic MPC solutions exist. We will discuss two here: share-compute-reveal and homomorphic encryption.

Share-Compute-Reveal
One of the ways to do this is the 'secret-sharing' technique. Secret-sharing involves dividing secret data into pieces (shares) in such a way that a single share does not contain any information on the secret data. The shares can therefore be spread among the participating parties without revealing the secret data. Ironically, secret-sharing does not mean that a secret is shared with other parties. All parties distribute the shares of their own input data in this way. The second step is to carry out the analysis. Instead of one party performing the analysis of all data, all parties perform the same analysis of the shares they received from the other parties. All parties receive a different outcome from which nothing meaningful can be derived. Only when the parties combine these local, intermediate results can the analysis result be revealed. This is the third and final step of the MPC approach. This three-step approach is also called the share-compute-reveal approach.

Homomorphic Encryption

Another means of achieving the aforementioned MPC functionality is homomorphic encryption. A homomorphic encryption protocol uses a public and a private key. The public key is known to everyone and can be used by the parties to encrypt data. The encryption protects the underlying data and can only be lifted using the private key. All parties can therefore encrypt their own data and share the encrypted data with one another. The homomorphic property ensures that analyses can also be performed on the encrypted data. Only when the analyses have been performed is encryption lifted using the private key. During all of the intermediate steps, the data remains encrypted and no secrets are revealed.

Please note that the party which holds the private key can decrypt all of the encrypted data. This key is therefore very powerful and a potential privacy risk. It is crucial that this key be handled correctly. A common approach is to divide the private key into pieces so that no single party has access to the entire key. This is where the aforementioned secret-sharing technique comes into play.



HOMOMORPHIC ENCRYPTION: COMPUTING WITH ENCRYPTED DATA.

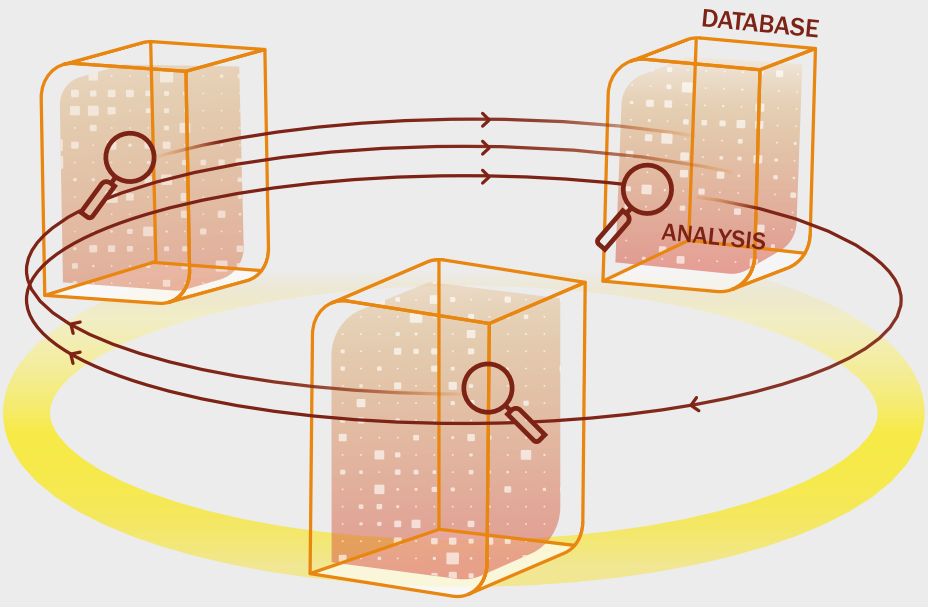
2. FEDERATED LEARNING

Federated Learning (FL) techniques make it possible to carry out machine learning on distributed, federated data while the data remains with its owner. Machine learning is a common means of extracting information or knowledge from data and is part of AI. A model learned through machine learning can help, for example, to predict whether someone is creditworthy when applying for a mortgage, whether someone will develop diabetes in the near future or who might be a fraudster. In doing so, the model incorporates knowledge on these distinctions, such as which traits support the diabetes prognosis and on what basis fraudsters can be recognised.

In order to obtain high-quality machine learning models, it is necessary to use as much data as possible to train the model. Standard machine learning algorithms require the availability of training data on a single machine or in a data centre. In the case of personal or otherwise confidential data, it is undesirable or even illegal to move the data to a central database for machine learning.

FL techniques make it possible to apply machine learning to distributed databases: the algorithms train local models on the distributed databases and combine the intermediate results into a global model. This training process is often repeated several times until a final model is reached.

As the intermediate results are merged during training, it is possible that parts of the data could be derived from them. Depending on the level of confidentiality required, it may be necessary to use cryptographic techniques such as the MPC described above.



FEDERATED LEARNING

FL is particularly suitable for analysing personal data (or other data) which is horizontally partitioned. Personal data is horizontally partitioned across different parties when the parties have the same type of data but from different people. For example, Party 1 knows the age and gender of Person A and Party 2 knows the age and gender of Person B. By contrast, we speak of vertically partitioned data when the parties have different information about the same people. For example, Party 1 knows the age of Persons A and B and Party 2 knows the gender of Persons A and B. When data is vertically partitioned, we more often turn to techniques such as MPC.

3. THE PERSONAL HEALTH TRAIN

The *Personal Health Train* (PHT) [7] is a solution which does not bring data to the analysis but brings the analysis (as a 'train') to the different data sources ('stations') via technical infrastructure ('rails'). For example, it is possible to apply complex algorithms to data managed by different organisations (hospitals) without having to collect this data centrally. A network of research institutions (including TNO) and parties in the healthcare sector is working on the development of the PHT. The focus lies on FL, but there are an increasing number of initiatives on also making MPC part of the PHT in order to provide extra security for sensitive data. At the end of 2020, the PHT won the Computable Award in the category for healthcare projects.

"With the Personal Health Train, it is possible to let complex algorithms loose on data managed by different organisations, without having to collect this data centrally."

4. CURRENT STATE OF THE TECHNOLOGY AND MARKET

Privacy-enhancing technologies such as MPC and FL often require considerable computing power and/or good communication infrastructure. As a result, the first MPC protocols (developed in the 1980s) were mainly theoretical and only had limited practical applicability.

However, recent technological developments and protocol improvements have completely changed this picture and MPC was used on a large scale for the first time in 2008. Using MPC, the Danish company Partisia was able to organise sugar beet auctions without having to rely on an external party to process all bid/offer prices. Since then, dozens of companies have been founded that focus entirely on rolling out privacy-enhancing techniques in specific application domains. Cybernetica's Sharemind platform enables users to perform statistical analyses in a privacy-friendly manner. In addition, Unbound develops MPC-based solutions for cryptographic key management.

In short, today's privacy-enhancing technologies are already having a measurable impact. At the same time, the academic world is not sitting still and more and more efficient protocols are being developed. At TNO, we work on new application possibilities for privacy-enhancing technologies and bring academic results to practice. In this way, we fulfil a bridging function between the academic world, the government and industry.

"It's time for a new premise on sharing sensitive data: don't share data, but leverage insights from distributed data sources while ensuring privacy and confidentiality."

⟩ BIBLIOGRAPHY

[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.

AUTHORS:
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

21-12046