Essay collection
Appropriate use of data in public space

From dialogue groups to new policy proposals

> nldigitalgovernment.nl/appropriate-data-use
Essay collection

Appropriate use of data in public space

From dialogue groups to new policy proposals

> nldigitalgovernment.nl/appropriate-data-use
1. Reuse of data in smart cities

2. The digital transformation, behavioural influencing, and the government

3. Public Governance of Experimental Data & Algorithms

4. The need for a digital Environmental Strategy

5. AI in the Digital Society

6. Data makes the world go round
## Table of contents

**Foreword** ................................................................. 6  

**Introduction** .......................................................... 7  

**Essay 1: Reuse of data in smart cities** .............................. 9  
  *Legal and ethical frameworks for big data in the public arena*  
  Prof. dr. mr. ir. Bart Custers  
  1. Introduction ......................................................... 11  
  2. Context and terminology ........................................ 14  
  3. Data reuse .......................................................... 17  
  4. Legal frameworks ............................................... 18  
  5. Ethical perspective – assessment frameworks .......... 26  
  6. Conclusions ....................................................... 29  

**Essay 2: The digital transformation, behavioural influencing, and the government** .............................. 36  
  *The slippery slope from nudging to überveillance*  
  Prof. dr. Wijnand IJsselsteijn  
  1. Introduction ......................................................... 38  
  2. Digital behavioural influencing: persuasive technology 40  
  3. Nudging: a little push in the right direction? ........ 42  
  4. Government by artificial intelligence? ................ 46  
  5. Big Brother is Watching You: the spectre of überveillance 49  
  6. Towards participatory design of the digital public space 52  
  7. Digital literacy, critical citizens and an honest government 53  
  8. Conclusions and recommendations ..................... 56  

**Essay 3: Public governance of experimental data & algorithms** ..................................................... 61  
  *Recommendations for a national algorithm register and reporting framework*  
  Prof. dr. Gerd Kortuem  
  1. Introduction ......................................................... 63  
  2. Data and Algorithms in Public Society ................. 64  
  3. Algorithms as Experimental Technology ............. 69  
  4. Governance Models for Algorithm Experimentation .. 76  
  5. Recommendations .............................................. 80  

**Essay 4: The need for a digital environmental strategy** ..................................................... 92  
  *From principles to practice*  
  Prof. dr. Liesbet van Zoonen en dr. Jiska Engelbert  
  1. Introduction ......................................................... 94  
  2. A broad landscape of principles ......................... 94  
  3. Challenges for the various government bodies ...... 98  
  4. Digital Environmental Strategy ......................... 100  

**Essay 5: AI in the digital society** ................................... 106  
  *Quality of algorithms and decision-making*  
  Prof. dr. Eric Postma  
  1. Introduction ......................................................... 108  
  2. Deterministic algorithms versus probabilistic algorithms 109  
  3. Machine learning algorithms .......................... 110  
  4. The AI revolution ............................................. 113  
  5. Observations .................................................... 118  
  6. Policy proposals .............................................. 123
Essay 6: Data makes the world go round

Proposal for research into three policy instruments designed to strengthen (digital) autonomy
Theo Veltman en Rob van Kranenburg
1. Data: age-old source of knowledge, power and money
2. The rich and dominant digital platforms: a risk to (digital) autonomy
3. Increasing pressure on (digital) autonomy: a ‘tipping point’
4. National and international initiatives in support of (digital) autonomy
5. What is required in order to strengthen (digital) autonomy
6. Conclusion

Overview of policy recommendations

Epilogue: Policy challenges relating to data use

Proposals for policy measures and activities

Publication information
Foreword

In the Netherlands, a wave of experiments is being conducted with the use of data and digital technologies for solving problems in public space. These experiments give rise to a number of different questions, both in government circles and among citizens. Who owns the data that is being collected in public space? How can you, as resident or visitor in that space, take more control over data and technology? How can we as government get the most out of data and technology while at the same time protecting our social values?

These questions bring one thing into sharp relief: the need to come together and talk with each other, as government and citizens, to agree on the ground rules for a socially appropriate use of data and technology in public space. This is why NL DIGITAAL: Data Agenda Government presents the project *Appropriate use of data in public space*, with the goal of identifying opportunities to innovate, advance the appropriate use of data in public space, and to open the dialogue with each other, because this is a dialogue that we will need to continue on an ongoing basis. I consider it important to make sure that everyone can participate in this dialogue and that no one will be excluded.

When I attended the first dialogue on appropriate use of data in public space, in Eindhoven in 2018, the questions the public raised struck me profoundly. Questions like “How do we ensure that the use of data does not lead to discrimination?” and “What is the real relationship between security and privacy?” These are examples of the justified questions that we began working on immediately.

From public dialogue groups through academic essays to new policy proposals. With practical examples to serve as inspiration. This collection, and even more so the road to it, has been a combination of thoroughness and innovation. I am convinced that this working method is what is needed in this day and age, as the pace of technological developments that we are confronted with every day continues to accelerate. Meanwhile, the tools that we as government have at our disposal (like policy and legislation) place more emphasis on prudence and diligence than speed. This project builds a bridge between these two worlds. To continue the dialogue on what is possible, what is acceptable, and what is justifiable when it comes to use of data in public space.

*Raymond Knops*
*Minister of the Interior and Kingdom Relations*
Introduction

In March 2019, the Dutch Ministry of the Interior and Kingdom Relations (BZK) presented the NL DIGITAAL: Data Agenda Government to the House of Representatives. This agenda concentrates on data use within society and particularly focuses on correct and responsible data use by the government. The Ministry of the Interior and Kingdom Relations has committed itself to this agenda and plans to further implement the results of the activities described therein. The Data Agenda Government sets out a variety of lines of action for the next three years, one of which is the ‘Appropriate use of data in public space’ project, which ties in with the current debate concerning public values in relation to the use of data in public space.

The goal of this project is to establish a shared view on appropriate use of data in public space and a corresponding perspective for action. This will be done by means of a series of thematic essays containing concrete policy recommendations. This project builds further upon the successfully completed pilot to inventory bottlenecks in relation to the issue of appropriate use of data in public space, with this inventory serving as the basis for the themed essays. These bottlenecks were discussed during consultations with residents, resulting in identification of the following issues:

1. Privacy versus technology: what is possible, what is acceptable and what must be avoided.
2. A. The conflict between companies and individuals; the interests of private commercial parties, and B. the conflict between individual interests and the collective, public interest.
3. The degree of transparency and interpretability of data use and data collection in the public space.
4. Purpose limitation versus reuse: the desired level of effectivity.
5. Public-private collaboration.
6. The role of the government within the smart society.

During the process of writing the essays, substantial and diverse efforts were made to link theory and practice within public administration. A supervisory committee was set up consisting of academic experts and experiential experts working for a variety of government bodies. This committee identified and recognized the aforementioned bottlenecks, played a key role in the formulation of the themed essays and played a supervisory role by appointing a coordinator for each essay. These coordinators then called upon their own personal networks to approach authors for the essays. Once the selected authors had been approved by the committee, work on the essays could begin. During the writing process, the authors were given maximum freedom to set out their ideas as clearly as possible based on their knowledge, experience and insight. Although academics are always affiliated with a particular university, their views do not necessarily represent the views of the university. Important to mention is that the remaining essays have been written in a personal capacity. The committee has met on several occasions to discuss the progress of the project and the essays themselves, as well as to provide feedback on the interim results.

This sextet of essays, which are based firmly on both theory and practice, eventually led to a total of 24 policy recommendations. The final stage of the project was the translation of these policy recommendations into concrete proposals for policy measures and activities. These policy measures and activities are at a high level of abstraction as they must take into account various contextual factors such as the political climate and the support base (at the administrative or other levels). In order to clearly position the various policy recommendations in relation to one another, Leiden University’s Institute of Public Administration developed a framework as part of an epilogue accompanying the six essays. This framework was used for the process of concretizing the recommendations in the essays and it also provides a number of suggestions for policy development concerning government data use.

It is important to mention that the policy recommendations and proposed policy measures and activities do not fully cover all of the challenges relating to data policy in the Netherlands, although they do cover the identified bottlenecks. As a result, this collection of essays can be seen as an initial attempt – and hopefully as a source of inspiration for future initiatives – to tackle the challenges we face stemming from the use of data in the public space.
The first part of this collection of essays comprises six themed essays (Chapters 1-6), all of which are introduced with a foreword written by the coordinator of the essay in question. The essays are followed by an overview of all policy recommendations and the Epilogue: Policy challenges relating to data use, written by Leiden University’s Institute of Public Administration. The collection concludes with concrete proposals for policy measures and activities corresponding to the recommendations taken from the essays.

References
1 One of the policy recommendations has already been put into practice via a collaborative project with the Association of Netherlands Municipalities to develop a Model Agreement on the use of data collected in the public space.
2 This pilot was constituted as a response to a call for collaboration on data use in the public space in 2017, from council members Ollongren (Amsterdam) and Depla (Eindhoven). They offered their ‘living labs’ environment as a space for experimentation to facilitate the learning and understanding of new technologies and to lay new ground rules together.
3 The concrete policy measures and activities have been placed on a longlist of preparatory measures for the Data Agenda Government 2020 and this list will be evaluated by the Data Agenda Government sounding board and steering committee.
Reuse of data in smart cities
Legal and ethical frameworks for big data in the public arena

Prof. dr. mr. ir. Bart Custers
Professor of Law and Data Science
Leiden University

Foreword
By prof. mr. dr. Sofia Ranchordas

Cities are traditionally known as anonymous places where individuals can live and work with little social interference. Nowadays, nothing could be further from the truth in so-called smart cities, where large quantities of data are collected on a daily basis. Although a large proportion of the data collected relates to traffic, for example, or the city’s air quality, personal data is often collected too, by public and private-sector parties alike.

This data is also exchanged and reused. For example, it has been reused in recent years to ‘nudge’ citizens to be more sustainable or more healthy. This use, and reuse, of data is often motivated by technocratic reasoning. Nonetheless, the reuse of data causes problems, for legal and ethical reasons. Can the use or reuse of data be regarded as proper if citizens’ autonomy could be compromised?

Have local authorities a complete overview of how data is being used and reused? Is the reuse of data in smart cities always necessary for delivering good-quality public services? How are public values safeguarded if data is being reused by private-sector parties?

The essay by Bart Custers analyses the legal and ethical frameworks of the use and reuse of data in smart cities. Custers, a professor of Law and Data Science at Leiden University, is a leading expert in the field of privacy and cybercrime. Professor Custers makes some important recommendations for local authorities, with a view to the appropriate reuse of data in smart cities. Local authorities should be keenly aware of what data has been lawfully collected, and they should actively involve citizens with digital initiatives; they should also use pilot schemes more often.

As a legal scholar in the field of public law and digital technology, it has been particularly enriching for me to be part of the supervisory committee and to reflect on both the opportunities and the risks associated with data collection and the reuse of data in smart cities.

Prof. mr. dr. Sofia Ranchordas
Professor of Law, University of Groningen
Member of the Supervisory Committee
Summary

Smart cities are urban areas where large amounts of data are collected using sensors to enable a range of processes in the cities to run smoothly. However, the use of data is only legally and ethically allowed if the data is gathered and processed in a proper manner. It is not clear to many cities what data (personal or otherwise) about citizens may be gathered and processed, and under what conditions. The main question addressed by this essay concerns the degree to which data on citizens may be reused in the context of smart cities. The emphasis here is on the reuse of data.

Among the aspects featured are smart cities, the Internet of Things, big data, and nudging. Different types of data reuse will also be identified using a typology that helps clarify and assess the desirability of data reuse. The heart of this essay is an examination of the most relevant legal and ethical frameworks for data reuse.

The most relevant legal frameworks are privacy and human rights, the protection of personal data and administrative law (in particular, the general principles of sound administration). The most relevant ethical frameworks are deontology, utilitarianism, and value ethics. The ethical perspectives offer assessment frameworks that can be used within the legal frameworks, for drawing up codes of conduct, for example, and other forms of self-regulation. Observance of the legal and ethical frameworks referred to in this essay very probably means that data is being used and reused in an appropriate manner. Failure to observe these frameworks means that such use and reuse is not appropriate.

Four recommendations are made on the basis of these conclusions. Local authorities in smart cities must commit themselves to the appropriate reuse of data through public-private partnerships, actively involve citizens in their considerations of what factors are relevant, ensure transparency on data-related matters and in such considerations, and gradually continue the development of smart cities through pilot schemes.
1. Introduction

Thanks to the rapid pace of technological developments, it has become possible to gather more and more data about citizens in the public arena. Data flows can be generated using all kinds of devices, such as cars, cameras, waste containers, electricity meters, and household equipment, that are connected to the Internet of Things. Through the careful use of the available data, cities will become smarter and more sustainable – including those in the Netherlands. Cameras and sensors that detect aggression or that communicate with passers-by can help make public spaces safer, for example. Waste collection can be optimized with the help of chips in containers that indicate when they are full. Induction loops in road surfaces and navigation systems can be used to control traffic flows in real-time. This is all possible thanks to the enormous quantities of data from different sources that are available in real-time (also known as big data) and to advanced analysis methods, like machine learning, data mining, and artificial intelligence (AI), which recognize automated patterns and sometimes take decisions independently. Cities that use big data smartly in this way are sometimes referred to as ‘smart cities’. Using data can make cities not only more sustainable, more innovative, and more liveable, but also strengthens citizens’ trust in government bodies and increases their participation in society.1

Smart cities have become ever-more important in recent years, both in the Netherlands and beyond. This is hardly surprising, as more than half the world’s population live in urban areas like Tokyo, Delhi, and Mexico City.2 Another factor is that most of the cities are growing rapidly, in terms of both population and greater urbanization. In response to the problems that this produces, such as greater inequality, climate change, and risks to security, cities are looking for technological solutions. In the Netherlands, too, cities are experimenting extensively with the use of data-driven solutions. Smart lampposts are being used in The Hague. They are able not only to give off more light, but can also be used for measuring air quality, noise pollution, and traffic congestion.4 Research is underway in Eindhoven into whether lighting can be used for guiding the emergency services to accidents and whether it can be adjusted according to weather conditions and how busy the streets are.5 Cycle congestion in Groningen is so great that attempts are being made at monitoring and finding solutions such as redirecting traffic flows.6

However, the use of data is only legally and ethically allowed if the data is gathered and processed in a proper manner. For this reason, the Ministry of the Interior and Kingdom Relations has established a committee to examine appropriate data use in public space. The committee has requested that a series of essays be drawn up on the subject. This essay, which forms part of the series, looks primarily at the legal and ethical frameworks for the reuse of data in public space.

In the Netherlands, too, cities are experimenting extensively with the use of data-driven solutions

It is not clear to many cities what data (personal or otherwise)7 about citizens may be gathered and processed, and under what conditions. As a result, digitization in cities is raising legal and ethical questions regarding what is allowed and what is desirable. Legal questions relate inter alia to privacy, ownership of data, the processing of sensitive data, the right to view data, and information security. Among the questions relating to ethical issues are trust, transparency, the commercialization of data, social legitimacy, and nudging. With nudging (see Section 2.4), information is used to encourage people in subtle ways – sometimes without their realizing it – to behave in a particular manner. One ethical question this raises is how far government bodies (in the case of smart cities, it is generally the relevant local authority) may go in steering the behaviour of citizens, such as ensuring that they make healthier and more sustainable choices. In this process, good intentions can turn into paternalism and lead to the autonomy and freedoms of citizens being curtailed. Although nudging entails freedom of choice, it simultaneously involves the manipulation of behaviour and/or choices.8

The legal and ethical frameworks for these questions are not always easy to apply, and are sometimes complicated and sometimes even lacking entirely. This essay looks at what legal and ethical frameworks could be used when making these decisions. The emphasis here lies on opportunities; the essay examines the way in which the legal and ethical frameworks can be applied in order that opportunities
for big data in smart cities can be used effectively while at the same time ensuring that public values and fundamental rights (such as privacy, transparency, and the autonomy of citizens) are not squeezed. It is certainly clear that using big data in smart cities presents many opportunities. From a legal and ethical perspective, however, there are good and less good ways of exploiting these opportunities. Society’s acceptance of and trust in the gathering and processing of data will be improved if careful consideration is given to the wishes of citizens. However, if they feel they are being manipulated in terms of their behaviour or the range of choices they have, such trust could diminish.

In this essay, the emphasis lies on the reuse of data in smart and sustainable cities. Businesses, whether they are technology companies or not, often accumulate their own data collections for optimizing their operational processes and strengthening their market positions. The playing field of cities is essentially different, however, because of the multitude of actors, each of which collects and processes data in their own arena. This complex playing field means that a large quantity of data is found among a variety of actors. To use this data smartly, it is generally necessary for it to be shared or to be linked up. This type of secondary application of data that has already been collected is referred to as data reuse. The reuse of data raises questions about when and under what conditions the passing on and even the sale of data is permitted.

Nonetheless it is recommended that, in all cases where new ways of using personal data are being considered in a particular organization, legal advice be sought from the relevant data protection officer and to plan options with and test them among all relevant stakeholders.

This essay is structured as follows. Section 2 provides further background information about the most important terms in this essay. The aspects featured include smart cities, the Internet of Things, big data, and nudging. Section 3 looks at the various ways in which data are reused. This typology helps clarify and assess the desirability of data reuse. The most relevant legal frameworks are dealt with in Section 4. These are privacy and human rights, the protection of personal data, and administrative law (in particular, the general principles of sound administration). Section 5 concerns the ethical perspective. The three principal areas of Western ethics offer various assessment frameworks that can be used within the legal frameworks, for drawing up codes of conduct, for example, and other forms of self-regulation. The ethical frameworks discussed are those of deontology, utilitarianism, and value ethics. The main research question in this essay is answered in Section 6, which also includes recommendations for the next steps that could be taken for reusing data in a legally and ethically responsible manner in the context of smart cities.
Appropriate use of data in public space
2. Context and terminology

This section provides further background information about the most important terms used in this essay. Specifically, the terms dealt with are smart cities (Section 2.1), the Internet of Things (Section 2.2), big data (Section 2.3), and nudging (Section 2.4).

2.1 Smart cities

Smart cities are urban areas where large amounts of data are collected using sensors to enable a range of processes in the cities to run smoothly. Examples include the gathering of data about traffic flows, energy provision, water supplies, waste processing systems, public order and security, and public amenities like schools, libraries, and hospitals. By analyzing the available information in real-time and linking it back to information provided to citizens and administrative systems, the available resources and amenities can be used as efficiently and as effectively as possible. Several examples are described below, in order to make clear what this means.

In the case of traffic flows, the general intention is to guide road-users to their destinations as quickly as possible. The fastest route is usually the shortest route. However, if there is heavy congestion on a particular route, there is a chance that a different route, necessitating a slight detour perhaps, is faster. Using the cameras on the roads, it is possible to monitor in real-time where in a city the traffic is at a standstill. On the other hand, if the traffic signals on a motorway where there is heavy traffic advise road-users to take a B road, then that B road will itself become congested. In such a case, it may be better to take the motorway after all, or use a different alternative. Otherwise, the result will be a type of self-fulfilling prophecy: if everyone is guided away from the original congestion, people will still find themselves in heavy traffic, only in a different location. What would be more useful, for example, would be to divert half the road-users but not the other half, so that the available road capacity is used optimally and traffic flows freely as much as possible. This cannot be achieved if all road-users are provided with the same information. But by providing personal input into the navigation systems of road users, a range of recommendations can be given, so that not everyone will end up doing the same thing.

Another example is that of waste collection in large cities. Typically, the collection vehicles come every week to every street to collect the waste containers. But because of this standardized process, citizens regularly find themselves putting containers that are only half-full on the street, and other times unable to dispose of their rubbish. Also, the collection vehicles have to pass by every single street, even though that is not always necessary, which is a waste of resources and a burden on the environment. In a smart system, every waste collection vehicle is fitted with a chip that indicates whether or not a waste container is full. The system is then able to work out where the vehicle should go on that day and calculate the optimum route. Delivery companies like Deliveroo already use algorithms to calculate the most efficient routes for their cycle or moped couriers. Firms such as IBM have developed similar algorithms for routing waste-collection services. Considerable transport cost savings can be achieved through optimum routing, which is also beneficial to the environment, while products and services can be delivered to customers more closely to the desired time. Routing is calculated by algorithms, not just on the basis of geographical data and times, but also of possible hold-ups and delays, the availability of alternative routes, or sudden changes to the wishes of customers or citizens.

Another example is that of waste collection in large cities. Typically, the collection vehicles come every week to every street to collect the waste containers. But because of this standardized process, citizens regularly find themselves putting containers that are only half-full on the street, and other times unable to dispose of their rubbish. Also, the collection vehicles have to pass by every single street, even though that is not always necessary, which is a waste of resources and a burden on the environment. In a smart system, every waste collection vehicle is fitted with a chip that indicates whether or not a waste container is full. The system is then able to work out where the vehicle should go on that day and calculate the optimum route. Delivery companies like Deliveroo already use algorithms to calculate the most efficient routes for their cycle or moped couriers. Firms such as IBM have developed similar algorithms for routing waste-collection services. Considerable transport cost savings can be achieved through optimum routing, which is also beneficial to the environment, while products and services can be delivered to customers more closely to the desired time. Routing is calculated by algorithms, not just on the basis of geographical data and times, but also of possible hold-ups and delays, the availability of alternative routes, or sudden changes to the wishes of customers or citizens.

2.2 Internet of Things (IoT)

The collection of data for applications of the kind mentioned above takes place in different ways. First, people themselves generate a great deal of data, such as when filling in forms, using social media, or whenever they post information about themselves on the internet. Technology also generates large amounts of data – for example, from sensors like cameras and the microphones that are built in to many devices.
A lot of data is gathered through trackers, too, including offline via RFID tags (in building access passes, for example, public transport chip cards, or debit cards used for contactless payments), or online via cookies that monitor online surfing behaviour.

Numerous devices like mobile phones or wristbands that record exercise and health generate very detailed and personal information. As soon as devices are connected to the internet, we talk about the Internet of Things (IoT).14 Obvious examples of such devices are computers, tablets, telephones, and navigation systems. There are less obvious items that can be connected to the internet as well. Examples that come to mind are cars (currently for navigation and monitoring maintenance; in the future, for driverless cars), thermostats and washing machines (for switching on remotely),15 electric toothbrushes16 (for advising on the right time to clean one’s teeth), and toys (for communicating with dolls, although they were removed from the market after controversy about privacy).17 In the context of smart cities, sensors are built into the sewerage system in order to regulate water levels,18 in waste containers so that they are only collected when full,19 and in lampposts to measure air quality, noise nuisance, and traffic levels.20

Connected to the internet, all this equipment creates enormous data flows, often in real-time, which can be smartly linked to each other to enable certain processes in smart cities to run more smoothly.21 An important feature of the IoT is the fact that devices also often communicate with each other (that is, not just with people). As a result, they can exchange data between themselves, coordinate matters and take certain decisions, such as in the examples described above – decisions about where and when information is provided or the choice of what the optimum route is.

2.3 Big data
Although there is no fixed, generally accepted definition of big data, the most important feature is the enormous quantities involved.22 That, of course, is why it is known as ‘big’ data. But there is no definition relating to a particular quantity. It is not the case that a certain minimum quantity of data should be present in order for the term ‘big data’ to be applicable. What is clear, is that it involves many terabytes and petabytes.23

However, there is more to big data than large data sets. Generally speaking, the term is used not only to indicate large quantities of data, but also to demonstrate that a technological development or acceleration is involved. This puts the concept of big data in the same category as the industrial revolution, semi-conductors, nanotechnology, robotics, and the internet. Big data is often regarded as the result of a steady growth in the amount of data available. In many cases, though, this growth is exponential. Take the aforementioned Internet of Things, for example. Whenever a user communicates (exchanges data) with three devices, there are three data flows. But when the devices also communicate with each other, there are suddenly six data flows.

There are also certain tipping points.24 Small quantities of data may be able to expose certain links, but large amounts can reveal exponentially more (providing such links exist, obviously). Where small quantities give a limited picture, in the case of rare phenomena like certain illnesses in isolated regions or countries for example, large quantities aggregated over many countries can provide fresh insights.
Big data concerns not only large data sets, but also new concepts – the term big data is often only used when most of the 3 Vs have been met:25

- **Volume**: Big data is about large quantities of data; no random tests are carried out.
- **Velocity**: Big data is generally processed in real-time and sometimes concerns streaming data that is not recorded or stored.
- **Variety**: Big data may concern unstructured data in different forms, such as text, figures, images (photographs and camera images), and sound.

The first two properties of big data have already been dealt with above. The third property, variety, entails a set of challenges of its own.

Where normal relational databases are often searchable with the help of queries (search terms and simple questions) and syntaxes (structured searches), big data comes mostly from various sources that are not always easily linkable. This means that analyses of big data are often more difficult than those of relational databases. In many cases, therefore, solutions need to be found through which data can be combined.26 Additionally, more and more tools are becoming available for analysing different types of data source. Examples include text mining for unstructured text documents and facial recognition for photographs and video images.27

### 2.4 Nudging

The term nudging refers to forms of subtle behavioural influence, where the aim is for people to behave in a particular manner. Behaviour can be influenced by framing choices in a certain way and through the means by which information is provided.28 Nudging is effectively just that – nudging people in a given direction by making the desired type of behaviour attractive, but without restricting people’s freedom in the process. The aim is to gently steer behaviour through choice architecture. Typical features of this are that people are always offered a choice (opting for behaviour other than what is desirable is possible) and that desirable behaviour is made attractive. Nudging is possible because people often behave according to certain automatisms (they follow the crowd, for example) or have certain assumptions (they may believe that more expensive products are necessarily better), or have inaccurate information (‘these are obviously the only choices I have’).

Typical examples of nudging can be found in supermarkets. Fruit and vegetables are generally near the entrance, as it looks inviting – fresh, healthy, and colourful. A-brands are usually presented at eye-level, while cheaper products are normally placed towards the bottom or on the highest shelves. Nudging is frequently used online too, through the provision of multiple options, for example, but where one is afforded special praise as the ‘most popular’, ‘new’, or ‘special offer now’. When faced with a range of options, many people select not the cheapest or the most expensive, but the second cheapest. In every case, people are given choices, but the desired type of behaviour is made attractive – sometimes visually and sometimes physically (by not having to bend down or reach upwards for products on the shelves).29

All these forms of nudging can of course also be applied in smart cities for ‘pushing’ citizens towards certain types of behaviour.30 Desirable behaviour may relate, for example, to a clean environment (a typical example of such an application is the use of Holle Bolle Gis in the Efteling Theme Park, which helps make the litter bins attractive), or healthy conduct (such as encouraging walking through attractive parks and car-free zones). The banning of cars from inner cities or penalizing people for throwing litter on the street are not examples of nudging, because they allow no freedom of choice and have a negative approach.

**Nudging is possible because people often behave according to certain automatisms**

Nudging can be attractive as a form of influencing behaviour, but there are objections to it too. People are sometimes manipulated without their realizing it, which can impinge upon autonomy and transparency. Nudging can also spill over into paternalism. After all, as soon as the government (or in the case of smart cities, the local authority) starts deciding what is desirable behaviour and acts accordingly, that affects the freedoms that citizens have. It may be true that with nudging, there is always a range of options, but those who apply nudging often determine the range of options that are actually offered.
There is also the question of whether nudging prompts changes to behaviour in the short term or to changes of mentality in the long term.

3. Data reuse

There are different ways in which data can be used and reused. This can be important with regard to the legal frameworks and ethical considerations. That is why it matters to make these differences clear. First, data can only be reused (secondary use) after it has been used (primary use). Based on previous research, we set out below a taxonomy for reusing data.

Collecting and processing data is always about a specific purpose in a specific context. This view about the use of data is more or less the same for the various perspectives from which you can look at data, whether that perspective is a legal, organizational, or technological one. However, there are differences between the perspectives when looking at how the use of data is interpreted. From a technological perspective, for example, the collection and storage of data is regarded as something that takes place before it is processed. The destruction of data is not seen from a technological perspective as the processing of data, but as something that takes place after the data has been used. From a legal perspective, things are very different. The General Data Protection Regulation (GDPR, see Section 4.2) does not explicitly use the terms use and reuse of data, but rather the concept of ‘data processing’. This is defined in Article 4 as ‘processing...such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction’ of data. In other words, the processing of data is regarded from a legal viewpoint as much broader than from a technological one, because it also concerns the collection, storage, erasure or destruction of data.

There are three different forms of data reuse, depending on whether the purpose and the context of such reuse are different to the purpose and context of the original use of the data:

- **Data recycling**: data is reused for the same purpose in the same context.
- **Data repurposing**: data is reused for a different purpose to the one for which it was originally collected, but the reuse takes place in the same context.
- **Data recontextualization**: the data is reused in a new context, different to the one in which it was originally collected.

These different types of data reuse are easily illustrated with the help of an example. Many online shops ask for personal details from their customers when they order something. This concerns their name, address details (for delivery purposes), and financial details (for payment purposes). Whenever a customer wishes to place an order in the future, he or she is able to log in and his or her details appear automatically. This is data recycling, because the data is being reused for the same purpose in the same context, but in a different instance.

If the online shop uses this data to build up customer profiles, for example to assess customer creditworthiness or to personalize offers, we talk of reuse for purposes other than the original purposes. This is referred to as data repurposing. Purpose limitation is a key principal in privacy legislation (see Section 4.2) in preventing what is known as function creep. In this context, function creep is the phenomenon whereby data is initially collected for a specific purpose, but is later used for other, new purposes. This form of reuse may be legal, when customers have given their permission for this, for example. However, when the same online shop sells the data it has collected to marketing companies, for example, the data may be reused in other – sometimes entirely different – contexts. This is what is known as data recontextualization.

Section 4.2 examines in more detail the question of what forms of reuse are legal and under what conditions. But apart from this legal aspect, it is also important to look at the reuse of data from an ethical perspective. This is because citizens have certain expectations regarding what happens to their data. Although certain types of data reuse may be permitted from a legal point of view, it is possible that they may result in citizens losing trust in companies and government bodies (in the context of smart cities, the local authority). This loss of trust may occur when they do unexpected things. Against that background, it is very important – and certainly for government bodies – to be transparent about the use of data, as otherwise support among the public for the use and reuse of data in smart cities may diminish or disappear altogether.
4. Legal frameworks

This section touches on the most important legal frameworks for the use and reuse of data in smart cities. Because the scope of this essay does not extend to describing in detail all the legal frameworks, the emphasis below is on the principles behind these legal frameworks. Section 4.1 deals with privacy and other fundamental rights, in Section 4.2 the protection of personal data is discussed, and Section 4.3 dealt with the general principles of sound administration.

4.1 Privacy and other fundamental rights

When it comes to processing data, often the first thing mentioned is whether it is allowed due to privacy considerations. Privacy is an important fundamental right in this context, but certainly not the only one that is relevant. Below is an overview of various fundamental rights that could be of importance when processing data.

Privacy

The processing of large quantities of data may lead to situations in which data managers know a lot about the characteristics and habits of people, and the places they go. For this reason, privacy is generally mentioned first as a problem or point of concern in the context of big data. This is no different in the context of smart cities.

From the perspective of fundamental rights, breaches of privacy depend largely on the expectations people have with regard to their privacy. To a certain extent, this is subjective and may depend on the situation, the individual, and cultural circumstances. One person may not feel the need to draw their curtains at night, while another may do so religiously. Some people put every detail of their lives on social media, while others do not even have a profile. An objective measure of expectations regarding privacy may be possible, by checking what an average person would expect in a particular situation or context. This is described as ‘reasonable expectations of privacy’.

In the context of big data, the emphasis often lies on so-called informational privacy, a term aimed at the question of what personal data is collected and used and for what purposes (see Section 4.2). In essence, expectations regarding informational privacy are primarily about sharing, revealing, and using data in ways that the person to whom the data relates (the data subject) does not appreciate. Such use of personal data is sometimes related to information security problems, when data is hacked for example (see the example of the Ashley Madison dating site for married people or those in relationships) or when data leaks, or is leaked. The unwanted use of data may be the result of a lack of transparency or of function creep – that is, when data is used for new purposes or in new context.

The categorization of different types of data reuse in Section 3 is relevant here. The reuse of data for the same purposes in the same context (data recycling) will meet reasonable expectations regarding privacy, but as soon as it involves other purposes (data repurposing) or a different context (data recontextualization), it is unlikely that that would be the case. The fact that data is already publicly available, because it is on the internet for example, does not mean that it is freely available for use for any purpose.

Unwanted revelations may also occur through causes other than data leaks. Even when the parties involved do not share their data with anyone else, the data may nonetheless be exposed in a big data context. This is because big data offers the possibility of predicting missing data on the basis of large quantities of available data. The same thing applies to predicting sensitive data that people do not generally prefer to share with others, such data about their health, ethnic origins, criminal record, or sexual preferences.
It should be mentioned, though, that data managers are able to predict this data relatively easily. Characteristic that vary over time, such as emotions and locations, may also be predicted with the help of big data, on the basis of posts on social media for example, or video images. If people do not wish to share certain information about themselves, but such information can be inferred from a roundabout route, then there is clearly a problem with privacy. This is a breach of reasonable expectations of privacy, which can produce a violation of the right to privacy.

If people do not wish to share certain information about themselves, but such information can be inferred from a roundabout route, then there is clearly a problem with privacy.

In certain cases, data managers may actually know more about data subjects than they do themselves, on life expectancy for example, the likelihood of their having serious illnesses or being in a car accident, risks of certain types of addiction, and estimates of well-being and happiness. A typical example is the giving of ‘likes’ on Facebook. Users can indicate with this icon what music, film clips, games, comments, people, etc. they like. Research shows that, on the basis of a few Facebook likes, it is possible to make a very accurate prediction of numerous sensitive personal attributes. For example, the researchers of Facebook users were able to make highly accurate predictions about their sexual preferences, ethnicity, religion, political preferences, personality traits, intelligence, happiness, drug use, and whether their parents were divorced. In other words, even when people do not, or do not wish to, reveal such aspects about themselves, it is possible to predict them on the basis of other data that they (or others) have shared. It should also be pointed out that anonymization can also be reversed in the same way.

In the context of smart cities, there are countless data that can be combined and therefore undermine anonymity or predict characteristics. Even on the basis of a postal code, dozens of personal data can be predicted, including income, household composition, and educational qualifications. Whereas in the past, people could move from a village where everyone knew everyone else to a town or city, where they were able to live more anonymously (sometimes to start a new life, to escape a scandalous or criminal past, for example), city administrators and companies perhaps now know more about their citizens. After all, the data can reveal information that people do not themselves know about, such as their life expectancy. Because data is retained for a long time, starting with a clean slate is not easy in a digital world, despite the existence of the right to have data deleted.

Discrimination, stigmatization, and polarization

If the above predictions expose sensitive data, such as ethnicity, political or sexual preferences, age, or gender, and they are subsequently used to take decisions about people, discrimination may occur. Discrimination can take place at different levels in big data analyses. First, the data itself may contain prejudiced information. A typical example is when the police keep watch in certain districts where large numbers of immigrants live. The likely result of any such practice is that police databases are filled with people of certain ethnic backgrounds. This is a typical example of selective, rather than random, sampling.

If the same police organization were to then look for patterns in the collected data in order to find out what groups display a higher risk of criminal behaviour, it will come as no surprise to learn that the same ethnic minorities may be profiled as a risk group when it comes to criminal behaviour. Because the data already contains prejudices even before the analyses are carried out, the effect is a self-fulfilling prophecy. If the police then go on to use the profiles to determine where they should conduct most of their surveillance activities, the circle is complete. By continuing to keep watch in these areas, more data will be fed into the databases and the profiles further reaffirmed.

It is important to note that patterns of discrimination are not always clear. For example, when sensitive characteristics like ethnicity are used, intentionally or not, for profiling purposes, it will be clear that there is a risk of discrimination.
However, when the profiles are based on postal codes, for example, this could product indirect discrimination whenever the postal codes are strongly related to ethnicity. In such cases, postal codes are only a proxy, an approximate variable, for ethnicity. Indirect discrimination is actually prohibited in the Netherlands because Article 1 of the Equal Treatment Act (which elaborates on the ban on discrimination contained in the Constitution) regards direct and indirect differentiation as equal. The fact remains, though, that indirect discrimination is much harder to prove than is direct discrimination. However, there are techniques for detecting discrimination in big data.\(^4\) Incidentally, research shows that the omission of sensitive characteristics from databases does not prevent the detection of patterns that lead to indirect discrimination.\(^4\)

Apart from the fact that missing variables can be predicted (see above), it does appear that many patterns that are exposed can also be found in variables other than the sensitive ones that are subject to a ban on discrimination.\(^3\) One such example is the use of geographic data for profiling. This is known as redlining and is generally forbidden.\(^5\) It should also be pointed out that indirect discrimination can also occur unintentionally and that users of profiles may not be aware of any harm they may be doing. If, on the other hand, profiles are being used precisely for the purpose of concealing discrimination, this is referred to as masking.

Legislation that bans discrimination on the grounds of specific characteristics concerns on the one hand lists of characteristics that may not serve as a basis for taking decisions (including ethnicity, political preferences, trade union membership, sexual orientation, etc.) and, on the other, certain types of decision that are forbidden (such as taking on and dismissing employees, offering products and services, etc.). Not every decision based on the sensitive characteristics mentioned is forbidden. For example, it is a matter of personal choice who your friends are. Nonetheless, ‘weaker’ forms of discrimination may occur in the formation of friendships, in the form of stigmatization of certain population groups, for example. On a larger scale, this could lead to social polarization and segregation.

Freedom of thought and expression, autonomy, and self-determination
Whenever anyone types in the first two or three letters of search term in Google, the search engine automatically produces a number of suggestions that start with the letters typed.

These suggestions are based on the large numbers of search requests previously made by other users and also on the searcher’s personal profile, with an assessment being made of what they might be looking for based on their previous search requests. What someone is looking for, what they like, and what they send messages about – this is all information that reveals what they are thinking about (at that time, at least). A search engine that makes suggestions in this way is at odds with the freedom of thought, which is protected in Article 9 of the ECHR and Article 10 of the Charter of Fundamental Rights of the European Union. As soon as thoughts are no longer private, the possibility arises of taking decisions based on a person’s apparent thoughts. This enables targeting certain times of the day or week or month when it is more probable that someone will make expensive purchases. It also means people can be ‘fed’ certain thoughts, thereby reinforcing them, whereas they may otherwise simply have been transitory in nature.
The result could also be that people alter their behaviour as soon as they realize they are being ‘watched’ or ‘overheard’. This is referred to in the world of surveillance as chilling effects. This curtails not only freedom of thought, but also freedom of expression (Article 7 of the Constitution, Article 10 of the ECHR and Article 11 of the Charter of Fundamental Rights of the European Union). Citizens in smart cities who do not wish their data to be collected may decide, because of the presence of cameras for example, to take different – possibly longer – routes through their city, where fewer cameras are situated. Another example is that citizens may make fewer gestures where there are cameras that may register them as aggressive behaviour. Whenever citizens modify their behaviour according to monitoring activities, this could affect the data that is collected.52

Citizens in smart cities who do not wish their data to be collected may decide, because of the presence of cameras for example, to take different routes through their city, where fewer cameras are situated.

At the same time, people’s autonomy can be compromised through the use of big data. This is generally the result of what is known as information asymmetry, where the data processor has many more insights and a better view of the bigger picture than does the subject. In principle, users should have control over what information they provide to whom and for what purposes. This is referred to as informational self-determination.53 Usually, this is implemented in a practical way through the drawing up of a privacy policy (or general terms and conditions) that set out what data is collected, how it is processed, and for what purposes. Users can then decide for themselves whether they agree to this or not (informed consent). However, there are considerable problems with this model, which is based on the idea that people are autonomous and rational actors.

Research suggests that if people were to actually read all the privacy terms and conditions on the websites they visit, they would spend an average of 244 hours a year doing so.54 It is apparent from other research, though, that users are only willing to spend between one and five minutes on this task.55 That is a big difference. Privacy terms and conditions texts are generally long. Facebook’s privacy policy, for example, is around 9,500 words long and takes more than an hour to read, while that of LinkedIn is some 7,500 words in length and requires about an hour to get through. Consequently, there are very few people who read these texts. Those who do take the trouble to do so may not understand them in full. That is because many privacy terms and conditions are formulated in heavy legal and technical language. Some privacy terms and conditions have been abridged and simplified (such as Twitter’s privacy policy), but it is questionable as to whether this is a good alternative. Simplified privacy terms and conditions do not always provide sufficient detail about what happens to data, which means that people remain inadequately informed for the purposes of taking the right decisions.56

This lack of clarity also exists with regard to the processing of data. Assuming that users do actually read and understand privacy terms and conditions, it is still possible that they do not fully understand the consequences of the data being processed. Indeed, there are cases where even the data managers cannot fully understand this either (and cannot therefore explain them in advance). This is the case, for example, when advanced automated techniques are used for analyzing data, such as data mining and machine learning, which have been designed precisely to expose new and unexpected links. This is one of the most important possibilities that big data has to offer. However, whenever someone reveals their shoe size, hobbies, or what pets they have, they do not generally expect that to serve as the basis for calculating their chances of having a heart attack. These possibilities – unexpected by ordinary citizens – make it difficult to properly inform people in advance.

Even when users know what they want, it is possible that their preferred option is not offered.57 In many cases, privacy and general terms and conditions give only a take it or leave it option. Usually, adding a check (tick) is the only choice available.
There is no scope for negotiating on the terms and conditions or agreeing to only part of them. The choice architecture could be used to determine what options may be selected. In some cases, privacy settings are offered that allow preferences to be set, but even these options are generally limited and non-negotiable by the citizen concerned. Certain privacy settings may also result in a decrease in the functionality of a service. On top of that, privacy terms and conditions are not always present, sometimes impossible to find, and sometimes incomprehensible (see next section).

**Transparency, a fair trial, and the presumption of innocence**

Although the General Data Protection Regulation (GDPR, see Section 4.2) creates various rights and obligations regarding the collection and processing of personal data, it is fair to say that it is difficult to exercise these rights and monitor compliance with the obligations. This is mostly as a result of a lack of transparency on who is collecting data, how and for what purposes it is being processed, and with what consequences. Research shows that people in the Netherlands have data in hundreds of databases. As a rule, it is not something people are aware of. Additionally, they do not know which databases they are, what data they contain, and what their rights are in relation to this data.

A lack of transparency can lead to situations in which people are faced with decisions taken about themselves without their knowing how such decisions have been taken. When people receive a quote for a mortgage, for example, it is sometimes perfectly clear what information was used as input, but not how the figures in the mortgage offer have been arrived at. In another example, it appeared that people with an extra letter in their house number (186A instead of 186, say) paid higher insurance premiums. According to American professor Daniel Solove, this all resembles situations described by author Franz Kafka, in particular in his novel, The Trial. In this book, Kafka describes a dystopian world in which the protagonist, Joseph K., wakes up one morning to find a number of government officials in his home who arrest him, without saying why. Given that K. is unable to remember committing the slightest misdemeanour, he is quite perplexed. Throughout the rest of the book, he embarks on futile attempts to find out why he has been arrested and how the matter can be clarified and resolved. However, it remains unclear what information about K. has been collected, how the decision to arrest him was taken, who is responsible for the decision to do so, and how the decision can be discussed.

The book by Kafka is mainly about the right to a fair trial – the requirements of Article 6 of the ECHR are not met. In the context of big data, this right is not necessarily violated, because every citizen has recourse to the courts, and can have their cases dealt with fairly and openly within a reasonable period of time by an independent and impartial judiciary. Nonetheless, it rankles when citizens are granted rights by the GDPR with regard to their personal data which they are not then able to properly exercise. The GDPR assigns various rights to citizens, including the right to be informed (about the information about them that is collected for example, and by whom, and for what purposes, etc. – Article 12 of the GDPR), the right to access their data (Articles 13 and 14), the right to view (Article 15), the right to rectification (Article 16), the right to have data deleted (‘the right to be forgotten’, Article 17), the right to restrict processing (Article 18), the right to the transferability of data (Article 20), and the right to object (Article 21), and the right not to be subjected to automated decision-making (Article 22). However, subjects can only use these rights effectively if they know that they actually possess them and how to exercise them. Many citizens are ill-informed about their rights in this area. The result is that this EU legislation for protecting personal data is not sufficiently effective.
Appropriate use of data in public space | 23
One further noteworthy point in this regard is that the principle of the presumption of innocence could be threatened in the context of big data. As stated in Section 4.1, big data is used extensively for predictions and profiling. Such predictions are of course very useful for businesses wishing to categorize people and to make personalized offers, for example. The predictions can also be used (and indeed are) for influencing voters. The predictions provide a kind of ‘inverted burden of proof’ for those involved: as soon as a prediction is inaccurate (it is merely a statistic), it is the task of the person involved to prove that the data is incorrect. This can generally only be done by showing the correct data, which represents a further intrusion of privacy. At the same time, this entails a procedural injustice, because the person involved is forced to defend himself against the consequences of inaccurate data or inaccurate conclusions.

4.2 Protection of personal data
The fundamental right to privacy is set out in greater detail in secondary legislation primarily for one aspect, that of informational privacy. At EU level, this has been done in the General Data Protection Regulation (GDPR). It is beyond the scope of this essay to provide a complete overview of all the provisions in the GDPR and how they apply in smart cities.

For that reason, only the principles for the proper processing of personal data are covered below. These principles, devised by the OECD in 1980, were incorporated into the so-called Treaty of Strasbourg of the Council of Europe in 1981. In 1995, the EU adopted the principles in its Directive 95/46/EU, the so-called data protection directive. This directive was replaced by the GDPR in 2018, which entered into immediate effect throughout the EU.

The principles are:
- **Limits on collecting**: there are limits to the collecting of personal data and such data may only be processed if it has been obtained lawfully and fairly and, where possible, with the knowledge and consent of the subject.
- **Data quality**: the personal data must be relevant to the purposes for which it is being processed and, where necessary, be accurate, complete, and up to date.
- **Specification of purpose**: the purposes for which personal data is collected must be clearly specified in advance and the personal data may only be used for these purposes.
- **Limited use**: personal data may not be used for purposes other than those specified in advance without the consent of the subject or if there is a lawful basis for doing so.
- **Security**: reasonable safeguards must be put in place against the risk of loss, unauthorized access, destruction, etc., of personal data.
- **Transparency**: the subject must be in a position to be made aware of the existence of and the nature of the personal data, the purpose for which it is to be processed, and the identity of the data manager.
- **Individual participation**: subjects have the right to have their personal data deleted, rectified, added to, or amended.
- **Accountability**: data managers are responsible for complying with these principles.

The first four principles are aimed primarily at the personal data and the conditions under which the data may be collected and processed. The other four principles are aimed mostly at the obligations and responsibilities of the data managers and the rights of subjects. It is important to note here that these principles are aimed primarily at procedural justice rather than substantive justice.
It is therefore possible that data managers are in compliance with these principles (clear purposes and purpose limitation) but that they go beyond what subjects expect. A typical example is when data managers state in their privacy terms and conditions that they will sell personal data on for advertising purposes or personalized offers. When a subject gives their consent to this, there is no problem from a legal perspective, but when a subject subsequently sees his or her photograph in an advertisement or is faced with higher prices, this may not be what he or she expects. Legally, there is not much to be done about this.

If data is used for other purposes in other contexts, it could constitute a breach of privacy expectations and the purpose limitation.

When reusing data in the context of smart cities, purpose limitation and the reasonable expectations of subjects regarding the use of their personal data play an important role. If data is used for other purposes in other contexts, it could constitute a breach of privacy expectations and the purpose limitation. In such cases, the best thing is to ask citizens for their consent and to be transparent about what data is used, and how.

4.3 General principles of sound administration

Apart from privacy and data protection rights, administrative law also has a number of provisions that are relevant to the collection and processing of data in smart cities. Because a detailed analysis of administrative law lies beyond the scope of this essay, the emphasis here is on the general principles of sound administration. These are the rules of conduct for the government sector (local authorities, in the context of smart cities) with regard to citizens. These principles originated from legal precedent, but were set down in the 1994 General Administrative Law Act (‘Algemene Wet Bestuursrecht’) in the Netherlands. Similar principles exist in administrative laws in other countries. The principles can be categorized as formal and material principles.

The formal principles are:

- **Legality principle**: there is no authority without a basis in law or in the Constitution.
- **Principle of due diligence**: the government must prepare and take decisions with due care; this includes the proper treatment of citizens, a careful examination of the facts and interests, properly following procedures, and sound decision making (Article 3:2 of the General Administrative Law Act).
- **Principle whereby reasons must be given**: the government must provide good reasons for its decisions; the facts must be correct and the reasons must be logical and comprehensible (Article 3:46 of the General Administrative Law Act).
- **Principle of legal certainty**: the government must formulate its decisions in such a way that citizens know where they stand or what the government expects of them. In addition, the government must apply prevailing laws correctly and consistently.
- **Fair-play principle**: the government must fulfil its public duties without bias (Article 2:4 of the General Administrative Law Act).
- **Ban on détournement de procédure**: a less onerous procedure may not be followed in order to arrive at a decision, if there is a procedure that exists containing greater safeguards for reaching that decision.
- **Principle of legitimate expectations**: anyone who has good grounds – after receiving a clear undertaking, for example – for relying on the government to take a particular decision, has the right to such a decision being taken.

These principles can be easily applied to the reuse of data on citizens in smart cities. Effectively, the principles largely resemble those in data protection law (see Section 4.2). The essence is that the government prepares, takes, and bases its decisions in an honest and careful way. With automated decision-making processes, data should therefore be collected lawfully, and be accurate and up to date. The decision-making process itself should be transparent. The reuse of data should therefore be in keeping with any expectations citizens have. This applies not just to government bodies that process data themselves, but also those who outsource the processing of data to private companies. The outsourcing of certain types of data processing, such as the collection and storage of large quantities of data or the carrying out of complex data analyses is perfectly routine, as it concerns specialized tasks, but can entail a whole range of practical problems.
For example, it can be difficult to check whether organizations that have been brought in are playing properly by the rules and it is not always easy to trace who has access to the data. The reuse of data, especially if it is for new purposes or in a new context can then cause problems, as a result of which greater emphasis is placed on consent and transparency.

The material principles of sound administration are:

• **Principle of specification**: the government may only represent those interests for which there is a basis provided in the relevant law or regulations (Article 3:4 paragraph 1 of the General Administrative Law Act).

• **Proportionality principle**: the government must ensure that any burdens or detrimental consequences from government decisions on citizens should outweigh the benefits to the public of the decision (Article 3:4 paragraph 2 of the General Administrative Law Act).

• **Principle of legitimate expectations** (legal certainty): citizens must, in certain conditions, be able to rely on the statements made by administrative bodies in which things are promised but which cannot later be delivered by the body in question.

• **Principle of equality**: the government must give equal treatment to equal circumstances (Article 1 of the Constitution).

• **Ban on détournement de pouvoir**: the government may only use a power for the purpose for which it has been granted (Article 3:3 of the General Administrative Law Act).

These material principles appear at first sight to be reasonably procedural in nature, because they relate to how decisions are taken. On closer examination, however, it seems that many of these material principles require careful substantive consideration on a case-by-case basis. This means that the circumstances of each individual case should be considered. As far as the reuse of data is concerned, then, the perspective of the subject (the citizen) should always be considered or at least weighed up against other perspectives, such as the overall public good. Ultimately, such a consideration will stand or fall according to the degree to which governments succeed in putting themselves in the shoes of the subject. Input and participation on the part of citizens can help with this process (see Section 6).

From a legal perspective, the interim conclusion is that there may be different frameworks (fundamental rights, the right to the protection of data, administrative law) that are supposed to offer protection to citizens, but that frameworks leave a great deal of scope for discretionary considerations. The existing frameworks are broad and general, with an emphasis on procedural justice rather than material justice. Although the legal frameworks make fairly clear what is not permitted, they are not always useful for determining what is appropriate use and reuse of data. In other words, when the legal rules for processing data are not adhered to, data use can be said to be inappropriate; conversely, though, not every form of data processing that is in compliance with legal rules is actually appropriate. After all, citizens may have very different expectations of what happens to their data or could happen to it.

5. Ethical perspective – assessment frameworks

As well as looking at legal frameworks, it can also be helpful to consider the use and reuse of data from an ethical perspective. The legal frameworks provide the overall conditions that have to be met. But within these overall conditions, there may be scope for a range of considerations. Legal frameworks cannot regulate everything – there are limits to their scope and to the quantity of details of laws and regulations. The ethical frameworks can also be used for interpreting their legal counterparts and form a further basis for self-regulation, through the setting up of codes of conduct, for example. Legal frameworks state how we must conduct ourselves, while ethical frameworks state how we should do so. The correct considerations then depend on the perspective. The moral perspective entails account being taken of the vulnerabilities of others. It can sometimes be difficult to weigh up various interests against each other, such as privacy versus security or individual interests against collective ones. Various assessment frameworks are described below, which can be applied from an ethical perspective. They concern three principal areas of Western ethics – deontology, utilitarianism, and virtue ethics.
5.1 Deontology

Deontology, which is sometimes referred to as obligation-based or rule-based ethics, assumes moral norms and values, moral principles, and moral rules. The moral norms and values describe the desirable situation, the hoped-for purpose – a just society, for example. The moral principles describe the general rules of conduct that help achieve moral values (for example, that wealth should be fairly distributed), while the moral rules are prescriptive and concrete rules of conduct (stealing is forbidden, for example). The moral rules should be adhered to – hence the term rule-based ethics. Moral rules could be legal rules (such as stealing is forbidden), but not necessarily (adultery is morally objectionable, but not criminal). Whether certain behaviour or certain decisions are moral is assessed according to moral rules.

The moral norms and values describe the desirable situation, the hoped-for purpose – a just society, for example.

Moral rules emanate from society, through public debate, for example. In a democracy, moral rules can be converted into legal rules by representative bodies, such as national parliaments or equivalent decentralized bodies, which leads us back to the legal frameworks mentioned in the previous section. However, in areas where no legal frameworks exist or where they are of insufficient force, moral rules can go further. For example, when it concerns the selling on of data, it depends very much on the context whether it is morally objectionable or not. In the field of privacy, datafication (registering all citizens’ behaviour in data), and dataveillance (surveillance based on data), there are also ethical challenges.

These moral rules can also be easily applied to the reuse of data. By way of example, when a government is proposing to sell data in smart cities to insurance companies for financial gain, it is useful for the relevant government employees to realize that they too are citizens. They can then examine, as citizens, whether or not the plan is a wise one. It is also a good idea in these situations to actively seek input from citizens – through the participation of citizens, it is possible to assess what rules are appropriate for the reuse of data.

5.2 Utilitarianism

Another moral perspective is that of utilitarianism. This ethical perspective, which was founded by English philosopher Jeremy Bentham, seeks to maximize well-being (‘utility’). Utilitarianism, also known as consequentialism, assesses whether certain behaviour or certain decisions are moral according to the results of that behaviour. Effectively, it is therefore necessary to find out what produces the best result.

The two big questions in this approach are, of course, what has to be maximized and then how that can be measured. Bentham himself defined utility as the sum of all the pleasure minus all the suffering that results from certain behaviour or decisions. British philosopher John Stuart Mill conceived the ‘greatest happiness principle’. From an economic perspective, financial calculations make the measuring of maximum prosperity possible to a certain extent, but prosperity and well-being are different things. Calculating maximum well-being is much more complicated. What is important, in this regard, is that the well-being of every person carries equal weight in the calculations.

As well as these practical problems, another objection is regularly voiced against utilitarianism – that is, that it is the outcome that is so decisive. In other words, the end justifies the means. Blindly applying utilitarianism could lead to a significant redistribution of prosperity – distributed as it is in some cases, very unequally. This may appear a good thing at first sight, but such interventions could be highly disruptive to the economy.

Despite the aforementioned objections, a utilitarian approach could be useful, for the purpose of weighing up larger-scale interests, for example, or for comparing the interests of the individual against those of society. A typical example concerns the use of autonomous cars in a city. This technology is not infallible, and therefore casualties cannot be ruled out. But if, in spite of this fact, the use of autonomous cars were to lead to fewer casualties than is the case with human drivers, this could
be a strong argument for deploying autonomous cars nonetheless. When it comes to the reuse of data, too, it can be useful to consider what strategies ultimately lead to the greatest levels of happiness. For example, when the question is whether images from traffic cameras can also be used to trace a child who has been kidnapped, a utilitarian argument could be used. The reduced levels of privacy of people on the camera images would in such a case not easily outweigh the importance of finding the child. In legal terms, incidentally, this consideration is generally made through the application of the proportionality principle, where consideration is given as to whether the means being used is in proportion to the intended aim.

5.3 Virtue ethics
A third ethical parameter is that of virtue ethics, which looks at how someone should live. The emphasis lies here on the actions of a person and the factors they consider in their actions and decisions. Aristotle is regarded as the founder of this approach, which was further developed by well-known philosophers like Thomas Aquinas, David Hume, and Friedrich Nietzsche. In contrast to deontology (in which behaviour and rules are key) and utilitarianism (which is about the result and impact), virtue ethics is about the character of the person and their actions (see Figure 1).

According to Aristotle, people should display courage, moderation, wisdom, and justice. Ultimately, it is the motives behind actions that determine whether certain behaviour or certain decisions are moral.

The most important question with this approach is, of course, what the virtues are to which people should aspire or which form the basis of a person’s intentions. Aristotle drew up a list of eleven virtues, including courage and justice, for developing outstanding character. Subsequent philosophers also drew up lists of virtues. In that regard, the work of American technology philosopher Shannon Vallor is of interest. Based on virtue ethics, she has devised a specific set of technological-moral virtues in order to make them more appropriate to our modern technological age. Her list has twelve values: honesty, self-control, humility, justice, courage, empathy, care, civility, flexibility, perspective, magnanimity, and technomoral wisdom. What is particularly important in the context of smart cities, in terms of virtue ethics, is what sincere and proper behaviour is. This is more relevant than which set of virtues or moral values are ultimately chosen. Primarily, it is about verifying whether certain intentions behind actions or decisions are correct. For example, when data about citizens is sold for profit, that underlying profit-seeking motive could be a morally flawed intention. If the data is sold for the purpose of subsequently using the results to improve the city, that assessment could be different. If decisions are taken with a view to improving the liveability of a city, that is a very different intention to one where city administrators seek to make a mark for its own sake.

![Figure 1: General ethical theories for further development of specific challenges of the reuse of data in smart cities](image)

The three above ethical perspectives are classical ethical approaches. For practical purposes, all kinds of instruments have been developed for consolidating these approaches. A good example is the Ethical Data Assistant (De Ethische Data Assistent, DEDA), an instrument for recognizing ethical problems in data projects, developed by Utrecht University and the Utrecht City Council. Another useful instrument is the Ethical Matrix. This instrument is based on three core values – well-being, human dignity, and justice – and helps to investigate step-by-step what values are involved and what can be done about this.
6. Conclusions

This essay has attempted to answer the question, to what degree may data on citizens be reused in the context of smart cities? The answer to this question varies from one situation to the next and depends on the circumstances. Nonetheless, there are clear legal and ethical points of reference by which this question can be answered on a case-by-case basis. The most important legal frameworks for the use of big data in the public arena related to (1) privacy and other fundamental rights, (2) the general principles of sound data processing in data protection law, and (3) the general principles of sound administration in administrative law.

The fundamental rights parameter provides primarily a substantive perspective, the data protection right (the GDPR in particular) a procedural perspective, and administrative law a combination of both. The big common denominator in these regimes is that of putting people (sometimes designated as subjects or citizens) first, rather than their data. By not forgetting that the collection and processing of data is a means, and not an aim in itself, it remains possible to consider whether the use or reuse of data is justified.

It is also important here that putting the rules (legal and procedural) first is not a good strategy. Although laws should be complied with, that is not enough. In particular, the legal frameworks of the GDPR offer sufficient scope for processing data within the existing rules in ways that could disappoint citizens on how their data is processed in smart cities. This could lead to an erosion of trust and support among the wider public for plans. The additional ethical frameworks are needed for that reason. They offer opportunities for making more substantive considerations, in addition to the procedural ones. The ethical perspectives described – of deontology, utilitarianism, and virtue ethics – could be helpful in this regard. Respectively, they focus on behaviour, the results, and the intentions of certain behaviour and certain decisions. Each perspective has its own strong and weak points, and by approaching complex issues from multiple perspectives it is possible to derive a better picture of what desirable courses of action and decisions are. Ultimately, the aim of the continued development of smart cities is to increase the well-being of citizens. It may sound trivial, but it is important here not to lose sight of these citizens.

Based on the aforementioned conclusions, the following recommendations can be made:

1. Local authorities in smart cities must commit themselves to the appropriate reuse of data through public-private partnerships.

The playing field of cities is complex, with a multitude of actors, each of whom is gathering and processing data in their own arena. Instead of doing this better themselves by collecting their own data sets, governments would be well-advised to reuse data, where data sets of stakeholders were shared or linked. Through public-private partnerships, data can not only be shared, but joint projects for improved liveability and greater sustainability, for example, could be set up and carried out.

The first step local authorities can take is to identify available data sets and existing data flows in the public-private arena. For existing or new forms of reuse, an assessment can then be made of whether such reuse is appropriate using the frameworks mentioned in this essay. For example, if the GDPR offers no legal basis, such reuse is not appropriate. When assessing reuse, every legal and ethical consideration in this essay should be taken into account – if only a few are selected (just the perspective of the GDPR, say, or of virtue ethics), there would be no guarantee that such reuse would be appropriate.

2. Local authorities must actively involve citizens with the factors to be considered in relation to smart cities.

The well-being of citizens is at the heart of any subsequent development of smart cities. Legal frameworks offer overall conditions and a minimum level of protection for citizens, but in addition to them it is important to retain the moral perspective, with account being taken of the vulnerabilities of others. The participation of and consultation with citizens can help ensure that sufficient account is taken of the perspectives of citizens and groups of citizens. Smart cities are not a goal in themselves – it is about the underlying goals such as well-being, liveability, and sustainability.
There is generally little point in asking citizens what they think about projects in which data are shared. In practice, it is often the same small group of people who respond – a group that is probably not representative of the population as a whole. A better alternative could therefore be to ask citizens about the underlying goals – what are their ideas on well-being, liveability, and sustainability? This can be done using the usual methods – surveys, interviews, discussion meetings, citizen panels, internet consultations, etc. It is then the task of the professionals to translate this information into specific projects and information flows, with due regard for the frameworks in this essay and in a transparent manner vis-à-vis citizens (see below).

3. In the case of smart cities, local authorities should ensure transparency in relation to data and factors under consideration.

Transparency is greatly important for the trust citizens have in government bodies and in developments concerning smart cities. Such transparency relates on the one hand to data (for example, which data about citizens is collected, and how it is used) and, on the other, how certain factors are considered (for example, how the need for data is weighed against data minimalization or how nudging and paternalism are weighed up against each other).

4. Local authorities should use pilot schemes to further develop smart cities.

Rather than having great plans, it would be safer and more realistic for government bodies to experiment using small-scale pilot schemes with further development of smart cities. The smart cities concept sounds grand, but it generally consists of a collection of different projects in various societal sectors, such as transport, security, the environment, etc. Through the development of new initiatives in different areas, information flows can gradually be optimized. More time can also be gradually made available for involving citizens with new developments and for gaining their trust.

Pilot schemes should always remain within the legal and ethical frameworks described in this essay. However, when experiments are being carried out, it may not be clear whether a pilot scheme does indeed fall entirely within these frameworks. In case of doubt, it is advisable to carry out a Privacy Impact Assessment (PIA), in order to identify the most significant risks to citizens. If any such analysis were to reveal high risks, the regulator (the Dutch Data Protection Authority) should be consulted before the pilot scheme is launched (Article 36 of the GDPR), who can then give advice. If a pilot scheme involved high risks, it is very much the question as to whether the plans being tested may or should be carried out in practice at all.
The assessment frameworks in this essay offer the opportunity to test whether or not data is being used or reused properly. By assessing existing or proposed data flows on the basis of these legal and ethical perspectives, it is possible to prevent data being used improperly. This is not a question of making a selection from the various perspectives (using only the GDPR, say); instead as many perspectives – and preferably all – should be used. Also, where any plans are proposed, it is a good idea to add to the frameworks through self-regulation, by using codes of conduct and covenants, for example. This is all very laborious, but that is only to be expected in a complex environment such as that of smart cities. Nonetheless, all that effort can be very much worth it, because specific projects and good intentions can go very wrong if they turn out to be insufficiently thought through and poorly prepared, and if too little account has been taken of the various perspectives and support among the general public.
About the author

Prof. mr. dr. ir. Bart Custers is professor of Law and Data Science at Universiteit Leiden.
References

6. Legally, data about citizens can imply personal data (see Section 4). The term ‘data’ is used in this essay for the convenience of the reader, but it actually means personal data.
27. Gandomi and Haider (2015) estimate that 95% of big data is unstructured.
34. Intelectual property law is not considered here, because it does not apply to personal data. In Europe, unlike in the United States or China for example, there are no property rights vis-à-vis personal data.
41 Custers B.H.M. (2012), Predicting Data that People Refuse to Disclose: How Data Mining Predictions Challenge Informational Self-Determination, Privacy Observatory Magazine 2012(3).
66 This is also known as the privacy paradox or profiling paradox. See Custers B.H.M. (2004), The Power of Knowledge. Tilburg: Wolf Legal Publishers.
70 https://www.tptv.overheid.nl/BWBR0005537/2019-04-02
71 See also Article 6 of the GDPR.


76 https://dataschool.nl/deda/


79 A limited number of small-scale pilot schemes also reduces the risk of flawed links between the schemes.

80 The scaling up of pilot schemes to practical implementation can sometimes be tricky, but their feasibility is enhanced if a gradual approach is taken. This makes it easier to absorb obstacles in relation to public legitimacy and a solid legal foundation.

81 See Article 35 of the GDPR. This is sometimes referred to as a ‘Data Protection Impact Assessment’ (DPIA).

82 In other words, the scope for legal and ethical experimentation, in the context of pilot schemes, is no greater than the scope of practical reality.

83 See Article 40 of the GDPR.
The digital transformation, behavioural influencing, and the government

The slippery slope from nudging to überveillance

Prof. dr. Wijnand IJsselsteijn
Professor Cognition and Affect in Human-Technology Interaction
Eindhoven University of Technology (TU/e)

Foreword
By prof. dr. Maurits Clemens Kaptein

The essay The digital transformation, behavioural influencing, and the government discusses one of the (in my view) most significant consequences of data use in public space: the use of data to actively influence the behaviour of the public. As Full Professor of Data Science & Health at Tilburg University and principal investigator in the Computational Personalization Lab at the Jheronimus Academy of Data Science (‘s-Hertogenbosch), I have been researching the use of data to understand and, ultimately, influence human behaviour for many years. There are many positive applications of the results of this research for society: for example, we have been able to show that personalized feedback derived from data can be used to develop more effective eHealth applications. This means that data can be used to better help people to lead healthier lives. However, this same knowledge is also being used for more dubious purposes: for example, the use of ‘persuasion profiling’ in online marketing is a method of behavioural influencing that is rightly discussed critically in much literature (and indeed, I have written a good deal about this myself).

In his essay, Prof. Wijnand IJsselsteijn of Eindhoven University of Technology describes how the collection of data in the public space is giving us the potential to direct and control human behaviour in the public space in increasingly effective ways. People are already familiar with the simplest examples of behavioural influencing in the public space: during large events with many attendees, like the Nijmeegse Vierdaagse (Nijmegen Four Days Marches), during which the public is monitored and large video screens are used to encourage participants to avoid specific locations.

However, there are also much subtler forms of influencing – for example, altering street lighting to induce traffic to flow towards certain areas – that are far less known among the greater public. With all these, it is not always clear who is turning the proverbial dials behind the actions intended to influence behaviour in the public space, nor is it always clear what data is being collected, from whom and by whom, in order to make the technology work. These points are particularly salient to the collective debate we need to have when it comes to the public space.
Based on his extensive body of research into the way in which technology influences people, and ways to help users understand the technology they use, Prof. IJsselsteijn presents new insights into the use of data in the public space. I daresay that there is no one more knowledgeable or more experienced in this area than Wijnand.

For myself, my participation in the Supervisory Committee for Appropriate Data Use in Public Space was extremely valuable. It was encouraging to see how the public sector, academia and industry could bring their various perspectives to the table and put their heads together on this important and charged subject for society. I personally learned a lot from the discussions, and can fully endorse our recommendations, many of which I would never have been able to foresee before we began on this process. I hope that you, as reader, will feel the same.

Prof. dr. Maurits Clemens Kaptein
Full Professor of Data Science & Health, Jheronimus Academy of Data Science (JADS)
Member of the Supervisory Committee

Summary

The digital transformation of society is a fast-moving process that is fundamentally changing the world we live in. It is something the public sector must be at the forefront of, because it introduces both unprecedented potential and new challenges for society. A few pioneering cities are experimenting with digital technology in the public space, in smart city projects and in living labs, to both increase the efficiency of government services and to improve the liveability and sustainability of the urban environment. In areas of active prevention policy (safety, public health, environment), the government is considering the potential for the use of digital resources in the public space, such as sensors, ‘big data’ applications and artificial intelligence to measure, predict and influence the behaviour of the public.

These can be applied in a wide range of interventions; examples include smart street lighting used to prevent aggression, or an interactive park walk designed to encourage an active lifestyle. The actions used to influence behaviour are referred to as ‘persuasive technology’ or ‘nudging’, and the ideas behind them are broadly derived from the field of psychology. The potential to influence behaviour in the public space using digital technology comes paired with a host of sometimes difficult technical, social, legal and ethical considerations. How, and to what extent, should the government be allowed to use public data flows, artificial intelligence and psychological persuasion techniques to influence the public’s behaviour towards the desired? And how must the government regulate itself when faced with the temptations of power and control that comes along with the digital transformation?

In this essay, I will take a close look at the potential and limitations of the digital technology being used to influence behaviour, and consider the ethical questions that arise when a government starts down the slippery slope from nudging to überveillance. Along with a critical analysis, this essay will also propose solutions based on participatory design and digital literacy.
1. Introduction

September 2002 saw the release of Minority Report, a science fiction film based on the Philip K. Dick novel of the same name. The protagonist in the film, agent John Anderton (Tom Cruise), works for PreCrime, a specialized police unit that has the goal of stopping criminals before they have committed their crime, based on foreknowledge obtained from three psychics known as ‘preCogs’. Minority Report is something of a near-future film noir, set in a world where privacy is a thing of the past; everyone anywhere in the public space is continually scanned by biometric sensors and their behaviour monitored. Everyone is constantly bombarded by a barrage of media and personalized product placement, while a technologically advanced government works to bring the crime rate down to zero with mobile police units equipped with jetpacks, miniature robots, and interactive video screens. Minority Report is set in the year 2054, thirty-five years from the time of this writing. If we go back 35 years in time, we land in 1984, a year that will forever be associated with another, even more famous dystopia: George Orwell’s totalitarian regime of Big Brother, where constant surveillance and behavioural control by the government have led to the total subjugation of the individual. Obviously, no one thinks of either one of these scenarios as a desirable society, but they help us to critically reflect on the values that are threatened when a government, in the quest for power, gives in to the temptations of technological control over its citizens. These are also the stories that put us face-to-face with the risks of overly one-sided choices, or technology taken too far, within familiar areas of tension – the clashes of interests that are already being felt in our present-day society: the interests of the individual versus the interest of the collective, privacy versus security, the regulated and rigid ‘system’ versus the unpredictable individual in an ever-changing and complex reality.

Meanwhile, back in the present, we are living in the wake of Moore’s Law – computers continue to get more powerful, smaller, cheaper, smarter, and more connected with each other (the internet) and the physical world (the Internet of Things). And along with this, computers are becoming increasingly personal and intimate: we have gone from the desktop to the laptop to the smartphone to the smartwatch and the ‘personal tracker’. We are inseparable from our mobile phones, which go with us everywhere, sharing a broad spectrum of personal information every step of the way – our location, social interactions and networks, personal photos and videos, health and fitness data, and our most deeply personal preferences and habits in buying, eating, entertainment, sexuality, etc. All this information is being collected, everywhere and all the time, and being shared with an unknown number of parties; many of them we know to have a commercial interest in inducing the individual to share even more information, so they can use techniques like ‘microtargeting’ and ‘hypernudging’ to induce people to buy, consume, experience, and of course, to ‘like & share’. And just around the corner awaits even more intimate technology (literally); even closer to the skin; even under it: augmented reality glasses, sensors in clothing and jewellery, subcutaneous chips and nanotechnology in the body itself.

As we zoom out from micro to macro, we see a similar trend. At the level of the public space in which we move, our direct living environment, we are seeing explosive growth in the number of sensors, which are being used to collect data on a massive scale, often by commercial market parties but in many cases, and increasingly, by the government. The government has good reasons for wanting to collect information. More detailed information on both situations and target groups represents significant added value for the effective and efficient performance of core tasks of the government.

The government has good reasons for wanting to collect information

What the government considers to be its package of core tasks is subject to change. Whereas in the first half of the twentieth century, the government made its primary focus the enforcement of the law and the preservation of safety, along with infrastructure maintenance (the government as ‘night watchman state’), in the latter half of the twentieth century the Dutch government gradually adopted a much more active role in promoting health and good education, and guaranteeing public housing and social security (the ‘welfare state’). In the twenty-first century, this set of tasks is once again in flux under the influence of changing political mores, challenges in cost control and efficiency and, to that end, the continuing advancement of privatization and deregulation.
In line with this changing perspective on government tasks, we see that governmental authorities are, increasingly often, choosing to exert control by means of prevention policy, particularly in areas like safety and healthcare; however, even in the area of social justice we are seeing visible changes to the role of government. For centuries, the government’s most effective influencing strategies have been legislation and criminal justice, taxes and subsidies, and information and communication; but in recent years we have been seeing the focus shift to influencing behaviour on the basis of psychological behavioural interventions. Here, digital technology is the new frontier. In a society in which the individual is confronted with countless distractions, and where control is breaking down or is absent entirely, it would seem that the age of helping the individual make the right choices – with a little ‘push’ in the right direction – has dawned.

A superabundance of food, and primarily unhealthy food, the ubiquity of tobacco and alcohol, and the prevalence of sedentary work in our knowledge and services economy has delivered us a golden age of ‘diseases of prosperity’. These are serious and chronic conditions such as certain forms of cancer, diabetes, heart disease, lung disease, and others, all strongly linked to lifestyle choices like smoking, excessive alcohol consumption, too little movement and too much stress. And they only exhibit their symptoms at their advanced stages, which makes them very difficult to cure. For these diseases, prevention is the most effective form of intervention, but this requires a change in attitude and behaviour on the part of the individual. It also requires a reorientation of the healthcare sector from a focus on fighting illness to a focus on promotion of health and prevention of illness. That approach is also a central point of the National Prevention Accord as recently drafted by Paul Blokhuis, the Dutch State Secretary for Public Health, Welfare & Sport.

Alongside healthcare, other major themes on which government policy is steadily shifting towards a focus on prevention are public safety and security. Since the nineteen-eighties, we have seen thinking on safety and security gradually changing, from an emphasis on repression and a judicial approach to criminality to the rise of ‘security thinking’, which places public order (rather than the legal order) at center stage. As individual freedom and mobility increase, cultural diversity advances, and social controls break down, individuals in society know each other less well, experience each other’s behaviour as less predictable, and feel less like there is a clear, shared social norm. As a result, we find the government (and the police in particular) increasingly filling the role of defining standards. Just as health is more than the absence of disease, safety is more than the absence of criminality. This is outlined effectively in the Minister of the Interior’s letter to the Lower House of Parliament on the ‘Integral Safety Programme’ [Integraal Veiligheidsprogramma] (1999):

“Safety can be described as the existence of a certain level of order and peace in the public domain and of protection of life, health and goods against acute violation or the threat of violation. Anything that infringes on this can be described as unsafe. These infringements of safety may be actual violations of the safety situation, or may be safety risks and feelings of being in danger.” (p. 9)

It is not surprising that Western governments are considering subtle behavioural control on the basis of insights into the human psyche for purposes like health and safety, and that they are using the potential offered by digital resources and structuring of the public space to do so.

This essay’s central topic is the influencing of behaviour in the public space. In particular, we will be devoting attention to the use, equitable or otherwise, of digital resources (data in particular) by the government, with the objective of collecting knowledge and being able to exercise an influence on the behaviour of its citizens.
In this essay, I will address in detail the principal means of digital behavioural influencing, namely persuasive technology (chapter 2) and nudging (chapter 3). Although these two approaches have somewhat different academic backgrounds, they do share the characteristic that they are based on using knowledge of psychology, the science of human behaviour and the mind, to influence behaviour. I will then look at the potential, limitations and risks associated with the use of artificial intelligence, and in particular ‘deep learning’ networks, in analysing behavioural data to arrive at valid classifications and predictions of behaviour (chapter 4). This leads us in chapter 5 to a consideration of the potential for abuse of such digital resources for behavioural analysis and influencing by the government (überveillance) or corporations (surveillance capitalism). Chapters 6 and 7 focus on directions for constructive solutions, and two in particular: participatory design (chapter 6) and digital literacy (chapter 7). I conclude with some final considerations and recommendations for policy (chapter 8).

2. Digital behavioural influencing: persuasive technology

Influencing and convincing our fellow human beings is a fundamental part of human interaction. From the snake in the Garden of Eden to today’s social media, we have been exposed to a never-ending barrage of attempts to persuade us and influence our behaviour. Buy our product, vote for our party, quit smoking, move more, drink less, have safe sex, protect the environment – and yes, ‘eat an apple’.

Media technologies have long been important tools in influencing behaviour, from megaphones to billboards to television. The government has made, and still makes, frequent use of these tools, and with considerable success. One need think only of the many media campaigns in which the government tries to influence a social norm; recent examples in the Netherlands have included campaigns against reckless behaviour with fireworks, drinking and driving, smoking, and, most recently, against using social media behind the wheel.

Technology functions as an extra-powerful communications channel when it offers interaction instead of merely one-way traffic, i.e., when the ‘influencer’ can adjust a message in terms of content, form, timing or location to address the characteristics or responses of the recipient of the message. Understanding this led to the development of ‘persuasive technology’ – technology designed with the explicit objective of influencing attitudes and changing behaviour. One example of persuasive technology would be a computer program that helps remind users to stand up, stretch, or even take a walk at regular intervals, in order to remind users to stand up, stretch, or even take a walk at regular intervals, in order to fight RSI and the cardiovascular diseases associated with a sedentary lifestyle. Another example is the smart thermostats that give feedback about daily energy consumption, to help users choose more energy-saving thermostat settings. A third would be the smartwatch or ‘fitness tracker’ device that people use to provide information and motivational feedback to help them stay active.

‘persuasive technology’ – technology designed with the explicit objective of influencing attitudes and changing behaviour

These examples from the tech world have something in common: they measure something about the user and/or her environment (uninterrupted computer usage, temperature fluctuations in the home, or the number of steps taken in a day), and then, on the basis of that information, give feedback and suggestions for behavioural change. Here it is important to note that these tech applications are used in a voluntary and informed way. Where coercion or deception is used, this falls outside the strict definition of persuasive technology, with the caveat that the reality is not so simple, and a more nuanced discussion of this point will follow below. In the examples above, we can see how people make the choice to use the persuasive technology, and the goals of the technology used are in line with the user’s own goals.
As an inherently social animal, people themselves are the strongest influencers of each other’s behaviour and opinions. This makes interpersonal interaction the strongest form of influencing. People have a strong social presence and impact, an intuitive understanding of the right timing and context for effective influencing, can assess the receptiveness of the target of the influencing, and have an arsenal (in many cases unconsciously) of influencing behaviour at their disposal. But as BJ Fogg describes in his book *Persuasive Technology*, computers also have a number of characteristics that make them uniquely suited for influencing behaviour. A computer can be extremely patient and persistent, and will never tire or become irritated. An algorithm does not judge behaviour other than the response patterns that are a part of its preprogrammed influencing strategy. This also means that a computer is able to guarantee the anonymity of a user in a way that a human being cannot.

A computer can be extremely patient and persistent, and will never tire or become irritated

Additionally, computers are able to, and put in a position to, record data and give feedback at times and locations at which other people are not necessarily welcome – in the home, at night, during meals, in the bedroom, in the bathroom, etc. As already noted previously, a computer (whether in the form of a mobile phone or smart watch, or as a ‘smart’ device like a smart bathroom scale, smart baby bottle, or smart speaker) is an intimate technology. People welcome it into their personal lives, share their personal information with it, and allow themselves to be influenced by it.

Technological advancements in sensor technology, algorithms, communications infrastructure and man/machine interfaces (which go by names like ubiquitous computing, pervasive computing, context-aware computing or ambient intelligence) open up the potential for new persuasive technology applications. This in turn will enable the persuading system to amass more and more contextual knowledge to the point where it will be able to draw conclusions in regard to the habits or current activity of the individual user and the social and physical environment in which the user is.

For example, the system can determine whether someone is at work, at home watching television, exercising, sleeping, etc., and if the person is alone or with other people. And thanks to improvements in sensors and the ability to measure physiological data, the technology is becoming increasingly capable of measuring the emotional state of the user.

All this can bring with it tremendous advantages for the effectiveness of persuasive communication. For example, health-related messages can be sent at the right time and at the right place (so-called ‘just-in-time, just-in-place messaging’), like helping to remind someone to take the stairs instead of the lift at the moment of entering a building, or pointing out options for a healthy meal to a shopper in a supermarket. We can also time such alerts so they do not appear during a busy meeting or while the user is driving a car, and instead only at moments that the user has the time to devote attention to them. Finally, we can adjust the influencing strategy, both in terms of form and content, to the user’s personal characteristics. This can be done for the user’s personality traits, habits, sensitivity (or antipathy) to certain influencing strategies, or other relevant individual characteristics. This combination of ambient intelligence and persuasive technology is sometimes referred to as ‘ambient persuasion’.

3. Nudging: a little push in the right direction?

In 2017, the Nobel Prize in economics was awarded to Richard Thaler of the University of Chicago for his work on the application of psychological knowledge within the discipline of economics. His work owes a great deal to a large set of psychological studies, and specifically findings from cognitive psychology on decision-making and behavioural choices by a number of other researchers, notably fellow Nobel Prize winner Daniel Kahneman, who showed that the idea of a human being as a rational decision-maker, the *homo economicus*, is a fallacy. We make many decisions on automatic pilot, without the benefit of detailed thinking processes or reasoning. We decide in the blink of an eye, on the basis of signals from the environment and a limited number of heuristics or ‘rules of thumb’.
This approach to thinking has the benefit of speed, something not unimportant in the do-or-die world of our evolutionary past, but it does have the disadvantage of making us susceptible to a number of systematic errors in judgment.13

These errors have to do with the fact that human beings have only a limited capacity for attention to relevant aspects of the environment, recall information selectively, are poor at assessing opportunities and risks, and are strongly focused on immediate, short-term changes and effects, with a tendency to ignore or underestimate substantial long-term effects.

In their 2008 book *Nudge*14, Thaler and his co-author Cass Sunstein introduced the concept of nudging: intervention in the decision-making context in a way that causes people to make better decisions virtually automatically, and that these decisions ultimately lead to healthier, more prosperous and happier lives. Nudges are just that, like a friendly poke that pushes you in the right direction with the intention of influencing a choice without restricting your freedom to choose or forcing you to do a specific action. In one well-known example, it has been shown that putting healthy products at eye level on the shelves of the supermarket leads people to choose the healthier products substantially more often. Another is the use of smaller plates in the school cafeteria, which reduces the amount of food consumed. A third example is the use of speed bumps to reduce the speed of cars when driving through residential neighbourhoods. By designing these ‘choice architectures’, what we are doing is trying to circumvent or mitigate the negative effects of our inadequate psychological decision-making capacity. A well-structured choice context can support and activate more desirable behavioural routines and automatisms.

It should be noted that behavioural influencing by means of a choice architecture does not imply that we are left entirely to the whims of the choice architect and that we have no option to deliberate, reflect, and choose another alternative. In Thaler and Sunstein’s model, it is important for the choice context to remain negotiable, and that we feel free in our consideration of alternatives. This is an approach that Thaler and Sunstein call ‘libertarian paternalism’.15 It is paternalistic because the choice architects attempt to influence choice behaviour on the basis of an assumption of a positive behavioural standard (for example: a healthy lifestyle), comparable to the way a parent or guardian wants to positively influence the choices of their children. It is libertarian because a key assumption of a nudge is that it does not restrict freedoms, but allows the person to reject the nudge and continue her course of action or make a different choice. We can still grab that unhealthy product on the store shelf, go back for more plates of food at the buffet, or take the car down a different route at high speed. If there is no option to escape the positive behavioural standard, then we are no longer talking about nudging, but manipulation or coercion.

The idea of nudging, as well as behavioural influencing more generally, by utilizing psychological expertise, has not gone unnoticed in government circles. The libertarian view, or ‘soft paternalism’, offers an additional set of powerful influencing strategies that can be used alongside more traditional approaches. Thaler and Sunstein (2008) argue that nudging is generally cost-efficient, and libertarian paternalism should ideally hold the middle ground of the social and liberal political spectrum: a government that encourages social responsibility without restricting individual freedoms. This makes libertarian paternalism a rather attractive policy alternative for a government that sees itself facing major societal challenges that can be traced back to individual behaviour, and in which the government is increasingly being put in a parental role.
While the original examples of nudging were primarily based on a redesign of a static choice context in the public space, and persuasive technology was more oriented towards the dynamic interaction between the individual and her tech-based tools, since the emergence of sensors, big data and artificial intelligence in the public space the line between nudging and persuasive technology has been blurring. Analysis of patterns in big data can be the key to the design of high-precision, targeted interventions in the qualities, habits, preferences and interests of an individual, group, community or city. This type of influencing has also been referred to variously as hypernudging or big nudging. The difference between this and ‘traditional’ nudges, if we may call them that, is that with big data the system does not offer only passive interventions, like speed bumps or a more prominent placement of healthy food. Instead, the influencing takes the path of a ‘control loop’, by which the data points can be used to create a model that allows interventions to be specifically tailored to situation after situation of the individual or target group, with the new data resulting from these interventions being used to assess effectiveness and then in turn being fed back into the model to further improve the results. Then, the whole process repeats, creating a cybernetic ‘control loop’ or ‘feedback loop’ as it continues.

Both persuasive technology and nudging use psychological knowledge of human cognitive ability and limitations, along with social patterns of activity, in order to influence behaviour as effectively as possible. Both use technology to measure, model and predict human behaviour, and to then try to influence that behaviour. Both persuasive technology and nudging exhibit characteristics in line with a mild form of paternalism. We use the term ‘paternalism’ because an authority outside the individual decides what healthy or responsible behaviour is, and on the basis of this standard explicitly intervenes in the design of the individual’s environment (including the technology), and thereby the individual’s behaviour. We can call this ‘soft paternalism’ because both persuasive technology and nudging make the claim that all of this is done on the basis of free will and informed consent.

There has been a great deal of discussion in the literature about the desirability and ethical implications of Thaler and Sunstein’s libertarian paternalism and the paternalistic assumptions of persuasive technology. Paternalism can be seen as a violation of the autonomy of the individual, who the model does not treat as a full-fledged adult capable of making his or her own choices. When we give technology the power to influence us, are we not being controlled by devices instead of thinking for ourselves? Paternalism also implies faith in the good intentions and superior expertise of those who influence us: they know what is good for us better than we do ourselves.

Soft paternalism is defensible if the goals strived for match the goals of the individual; that is to say, if an intervention either makes the person do what she actually wants to do herself, or protects the person against external forces that are standing in the way of her achieving her personal goals (such as, for example, temptation). This is contrasted with ‘hard’ paternalism, by which a government or other party uses methods to restrict the full freedom of action. Philosophers Lily Frank and Philip Nickel argue that it is challenging to align the goals of the individual with the goals of a persuasive system. Even though at an abstract level the goals of the system may resonate with the goals of the person (for example, losing 10 kg within a year), this may not hold true for the various implementation choices. This means that the persuasive system could at any time make proposals that are not appropriate to the personal preferences or options, or do not work within a specific context or situation. In such cases, Frank and Nickel question whether this can still be considered soft paternalism.
Thaler and Sunstein’s work seems more focused on the party dishing out the nudges, rather than the party taking them. In this respect, a change of perspective is a useful exercise, as is evidenced from a review of the book *Nudge*, by Robert Sugden21, a colleague of Thaler and Sunstein:

*If you are attracted by the agenda of libertarian paternalism, I urge you to think carefully about whom you are authorising to nudge you, what criteria they will use to decide when and how your own decisions are capable of being improved on, and what incentives there are to induce them to follow those criteria.* *(p. 373)*

These considerations then lead to another point of criticism: the idea that the individual to be influenced has access to complete information about the objectives and methods of the behavioural influencing. This is an essential assumption for both Fogg’s persuasive technology and Thaler and Sunstein’s nudging. Incomplete, incorrect or absent information can be considered misleading. Having said that, nudging knowingly and deliberately takes advantage of our cognitive ‘weaknesses’ to influence our behaviour. Doesn’t this deliberately short-circuit any conscious reflection? We might well question how explicitly the persuasive principles and objectives are really being presented. The counter-argument to be made is that even a choice context that is not designed to be explicitly paternalistic, or is not explicitly designed with the object of influencing ‘desired’ or ‘healthy’ behaviour, is still a choice context that will, consciously or unconsciously, have an influence on our behaviour.

Take the example of the speed bump. A critic of nudging could argue that this is a paternalistic restriction of the motorist’s freedom of choice. On the other hand, the design of the car and the fact that it is able to achieve high speeds with great ease and in great comfort, is also a reflection of a whole series of design choices that facilitate a certain type of behaviour (like driving too fast), and therefore even encourage said behaviour. In that light, our technological environment is always having an influence on our behaviour. This point is made most explicitly by philosopher-of-technology Peter-Paul Verbeek.22 With references to Hans Achterhuis and Bruno Latour, Verbeek makes clear that technology, when in the hands of human beings, is in effect always taking on a mediating role. Verbeek23:

*’The role of technology is so fundamental that we should not pretend that we can be completely autonomous. More to the point, whoever refuses to think about desirable forms of behavioural-influencing technology sets aside the responsibility for giving that mediation a desirable form.’* *(p. 26)*

This means that technology has an inherently moral or behavioural-influencing quality, even if this is not explicitly communicated as such. It should be noted that even if we acknowledge that we cannot be fully autonomous in our relationship to our technological world, this is not to say that we cannot strive for freedom in our choices. More than anything else, Verbeek encourages us to be mindful in the way that we relate to technology, and the role that we let technology play in our lives. When it comes to behavioural influencing set into motion by the government, we should make our influence felt principally in the design process of the technology – on the basis of a participatory design – and in the democratic dialogue and critical debates in which the role of the government and its use of technology in relation to the individual in society is examined.

The frequent, extended and ubiquitous use of behaviour-influencing technology by the government or by other parties also raises another question: how will this influence human value systems in the longer-term? If we are constantly exposed to tech-driven nudges in the ‘right’ direction, are we really learning for ourselves what is good and what is bad? Can we really call it a moral development in the right direction if it is based on internalization of the desired behaviour and the implicit standard being communicated thereby, or is the behaviour rather more a superficial automatism – the path of least resistance?
How can we be proud of what we achieve, aware of our effort and the reward we reap from it, if we have done nothing other than listen to a technological superego? Are we really good if we do good thanks to the technology that pushes and pulls us, nudges and encourages us, rewards and punishes us? Our moral growth and the cultivation of our morality is inextricably bound up with deliberate choices for what is hard for us, the things for which obstacles have to be surmounted. It is this independent responsibility for behaviour, this ownership of our own choices, this accountability, that is the path to developing our own moral personality.

If we do not want our minds to become a helpless extension of the technology, we absolutely need to have a critical reflective perspective on behavioural influencing.

Verbeek is right – we can choose to relate to our technology in a mindful way. But convenience and intellectual laziness – a lack of mindful reflection – are part of the mechanism that behavioural influencing takes advantage of. If we do not want our minds to become a helpless extension of the technology that we have such an intimate relationship with, we absolutely need to have a critical reflective perspective on behavioural influencing.

4. Government by artificial intelligence?

Artificial intelligence (AI) is the area of science engaged in the creation of technological artifacts that are capable of exhibiting behaviour that we would characterize as intelligent. This area of science is in itself multidisciplinary, with elements of computer science, cognitive psychology, neuroscience, logic, statistics and linguistics. In addition to these, AI also touches on a wide range of legal, ethical, philosophical and sociological questions, so also interfaces with fields like the philosophy of technology, the philosophy of mind, the organizational sciences,
ethics and law. AI is a rapidly developing field with many clear, visible and usable applications, including (to name just a few) automatic image recognition, speech recognition (as used in digital assistant systems like Siri and Amazon Echo), natural language processing (as used in many applications, from Google translate to chatbots), knowledge systems (like IBM Watson), robotics (industrial robots, drones, autonomous vehicles, etc.) and the latest generation of computer games.

Big data has been called ‘the oil of the 21st century’. If that is true, then AI is the combustion engine in which this oil burns. In recent years, we have seen dramatic advances in artificial intelligence (deep learning), with supercomputers simulating brain-like networks that can learn to recognize patterns in vast quantities of numbers, images, text fragments, speech and other forms of data.

And governments are increasingly experimenting with the use of AI for a vast range of government services, like identifying benefit fraud, predicting recidivism of criminals and implementing smart traffic systems. The promising technical potential of AI in combination with the amount of data that the government can potentially access, along with the size and complexity of the challenges facing governments today, all make AI an increasingly important tool in the public sector. There are significant opportunities for innovation that can lead to increased efficiency and quality improvement, but that also bring with them potential risks.

The quality of AI classifications is strongly dependent on the quality of the data used to train the AI: if this data is too one-sided or not representative of the target group, this will have a negative impact on the quality of the AI’s results. For example, if facial recognition software is primarily trained on a set of white male faces, it will make more errors in classifying people of the opposite sex and of other races. This is how Joy Buolamwini, the founder of the Algorithmic Justice League, was able to show that the facial recognition software of major tech companies like Amazon, Microsoft and IBM is still fundamentally flawed. While for light-skinned men, the number of classification errors remained below 1%, the number of errors for dark-skinned women was 35% (and this was shown in a data set that included famous faces, like Oprah Winfrey, Michelle Obama and Serena Williams).

Another controversial example in this area is the recent work of AI researchers Wu and Zhang, who trained AI algorithms, or ‘classifiers’, on a set of 1856 existing photos of faces of Chinese people, approximately half of which were convicted criminals. After training, the algorithms proved to be fairly good at being able to distinguish between criminal and non-criminal faces. One problem with this type of result is that it is highly plausible that people with a face that others see as having prototypical ‘criminal’ features run a greater risk of being arrested than people walking around with a face perceived as prototypically ‘innocent’.

This means that there is a reasonable chance that the set of noncriminal faces also includes an unknown number of faces of people who were wrongly not arrested. With that in mind, we can see how an AI algorithm might be allowing an existing prejudice in the community to continue, and perhaps even be reinforcing it. If such an algorithm were then to be used by the Chinese government to make predictions about potential criminal behaviour (an idea known as ‘predictive policing’), we suddenly see a parallel with the preCogs of Minority Report that we mentioned in our introduction; this would in effect trigger a self-fulfilling prophecy.
A similar case can be seen in the use of algorithms to make a risk assessment of the degree to which problem behaviour will occur more frequently in a given district or community than elsewhere. The response to such a prediction of marginally more problem activity in a community as compared to others would be an increased police presence in the community, which in turn would increase the rates of detection of problem activity. When this new arrest data is then fed back into the model as updates to the situation, this would set into motion a recursive, self-reinforcing process in which a community is singled out, largely incorrectly, as being a source of problems on the increase that other communities are not. The above-described self-reinforcing processes are, de facto, also happening now, even without AI profiling: it is understood, for example, that the high rates of incarceration among black Americans as compared to white Americans is the result of a similar effect.

Prejudices hidden within the data can be blown out of proportion by feedback loops and machine learning. This shows how the use of AI confronts us with results that we as a society consider to be undesired or unfair; but these are deeply engrained in our social interactions and our social judgments. In that sense, AI is a moral mirror of our social structures. It might be that this problem can solve itself once we as a society start dealing with our prejudices in a more informed way, taking this into account in the data sets we make available. We could, for example, retrain an algorithm with more recent data, so that both man and algorithm gradually lose their prejudices little by little. In fact, there are already toolkits and guidelines available for making AI more fair. One such toolkit recently released by IBM as an open source package is AI Fairness 360²⁶, which has been designed to detect and limit prejudices in artificial learning.

One issue for the engagement of AI is that the amount of suitable training data from the ‘real world’ is often far less than adequate for creating a robust model. For example, we can happily say that the number of outbursts of aggressive behaviour on the street remains relatively limited, but that means the data set of camera footage (or data from other relevant sensors) that can be used to construct a predictive model of aggressive escalation is also extremely limited. When an AI model has too little data, the results will be extremely unreliable, little better than the predictive power of flipping a coin.

Generally speaking, it is only massive and accessible data sets that can be used to train algorithms in a manner that can be decisive for the results of a classification or the outcome of a decision. Underrepresented actions, events, qualities or groups (notably minorities) are, as a result, recognized less well or with less than clear-cut results.

When an AI model has too little data, the results will be extremely unreliable, little better than the predictive power of flipping a coin.

Artificial intelligence is very strong in well-defined problem areas with clear patterns and/or rules, but remains very limited in its capacity to make generalizations into other areas, even where such generalizations might seem extremely obvious to the human mind. The capacity of general intelligence – what we humans might consider ‘common sense’ (completely clear and obvious) – is still something far beyond the horizon of today’s AI. We see, for example, that AI sometimes makes assumptions or errors that people find completely incomprehensible. Additionally, a self-learning system cannot (at least, not yet) reason at a general level about how a decision or classification was reached in a way that humans can understand. In other words, the artificial learning system is not very good at explaining how or why it arrived at a given recommendation or classification. This is not easy to swallow in any setting where recommendations are given or decisions are made, but it is completely unacceptable when it comes to a government and its decisions. Good government must always be in a position to explain how it reached its decisions, and must be transparent towards its citizens. ‘Computer Says No’ just doesn’t cut it.

It would seem that the greatest strength of AI systems comes from the combination with the human professional rather than the replacement of the human professional. It is strongest as a tool for supporting and expanding our own cognitive capacities.
Earlier in this essay, we discussed the shortcomings in human cognitive capacities, like the assessment of risks on the basis of data or the proper consideration of effects in the longer-term. It is on these points that a computer, and AI, can be very complementary to the human mind. The speed at which patterns in huge volumes of data are analysed can represent tremendous added value to the slower, more thorough-going and context-sensitive mental labour of the human being.

It would seem that the greatest strength of AI systems comes from the combination with the human professional rather than the replacement of the human professional.

What the above makes clear is that just like humans, AI is not infallible. Knowledge of both the strengths and weaknesses of AI is indispensable. With that in mind, it is also extremely important to think about the most optimum distribution of roles between man and machine, about the design of the man/machine interface and interaction, and the degree of control and oversight that the human operators want to and need to retain in light of the ever-advancing digital transformation. These are things that professionals, both in the government and elsewhere, must prepare for within the current working framework; but we also need to make the training and education processes ‘future-proof’, so the professionals of the future will have the relevant knowledge, skills and flexibility.

5. Big Brother is Watching You: the spectre of überveillance

In a discussion about the Electronic Health Record (EHR), astrophysicist Vincent Icke once struck a memorable analogy with a famous moment in Gulliver’s Travels. After being shipwrecked and washing up on the island of Lilliput, Gulliver awakens to find that he has been ensnared by a web of hundreds of tiny wires.

Every wire, Icke said, represents one piece of personal data. In our world, these are many: the EHR, a social media account, a GPS navigation system, your supermarket club card – each individually innocent and not particularly dangerous, but just like Gulliver, it is the combination that keeps us ensnared. If these wires come together to close around an individual like a web of data, and are used to give the government increased options for surveillance and control, it becomes what authors Michael, Fusco and Michael have called überveillance.

In their article, they focus in particular on location-based services (LBS), that is, services that need location information from the individual (which may, for example, be taken from GPS or RFID systems) to function. Such applications range from the GPS systems in trucks that allow the trucking company to determine the location of any of their vehicles to the ‘friendfinder’ applications in mobile phones, and even the ePassports that use RFID tags for identification and security purposes. Additionally, of course, the government has a rich palette of options for identification of individuals, including camera systems in the public space, biometric data and a vast array of government databases containing personal data.

Überveillance is no single source by itself, but all sources together. It can be strongly linked to Roger Clarke’s concept of ‘dataveillance’ defined as ‘the systematic use of personal data systems in the investigation or monitoring of the actions of one or more persons’. Überveillance is dataveillance on steroids. It is 24/7, omnipresent electronic surveillance. However, omnipresent and ‘always on’ does not mean that the überveillance system is all-knowing or even can be so. In theory, a surveillance system is meant to assess ‘who’ (identification), ‘where’ (location), and ‘when’ (time) in order to arrive at an interpretation or prediction of ‘why’ (intention/motivation), ‘what’ (intended result), and ‘how’ (method).

The risks of misinformation (e.g. measurement errors, missing information sources, coincidental correlations), misinterpretation (e.g. incorrect response to uncertainties and ambiguity, incorrect inference of intention), and manipulation (e.g. selective reporting of data, confirmation bias, hacking, political manipulation) of data obtained from an überveillance system are significant. At this point, the human tendency to accept sensor data and algorithmic interpretations of such data as ‘true’ and ‘objective’ and ‘scientific’ is disconcertingly high.
As Michael et al. put it: ‘Überveillance can be a predictive mechanism for one’s expected behavior, traits, characteristics, likes or dislikes; or it can be based on historical fact, or something in between. The inherent problem with überveillance is that facts do not always add up to truth (…), and predictions based on intelligence are not always correct.’ (p. 1198).

Alongside concerns about the actual authoritative strength of surveillance data, überveillance presents clear ethical problems. Firstly, it represents a violation of the right to privacy, which dictates that every individual should have the right to a personal life, unobserved by third parties, in which we can go where we want, interact with whomever we want, obtain information freely, and live the lives that we want to lead (within the limits of the law). Generally, the citizen of Dutch society places a very high value on privacy, and any violation of privacy observed leads to a crisis of trust. Psychologist Judy Burgoon and her co-authors identify four different types of privacy:

1. **Physical privacy** is the control that an individual has over the degree to which their person is physically accessible to others (visibility, personal space, physical contact, etc.).
2. **Social/communicative privacy** is the option to, and difficulty with which, an individual can regulate social contacts (for example, the length and content of a social interaction, non-verbal communication).
3. **Psychological privacy** refers to the possibility of regulating cognitive and emotional signals (examples might include preventing harassment and bullying, but also exposure to behavioural influencing).
4. **Informational privacy** refers to the control over the degree to which and the way in which information is shared with others (examples: gossip, reading someone else’s mail without permission, sharing someone’s photos online).

In situations of überveillance, the violation of privacy happens at a combination of the physical, psychological and informational levels. Such a fundamental loss of privacy also represents a loss of control and personal autonomy. A person who knows that he is continually being watched behaves differently, engages in self-censoring, is likely to avoid situations and interactions that would appear sensitive or controversial, and in many respects will no longer be able to ‘be himself’.

Over time, the loss of personal control and autonomy will lead to feelings of frustration, helplessness and depression, and ultimately can give rise to major societal unrest and resistance. But what is potentially even more concerning is not the frustration and resistance, but the lack thereof. When a new generation grows up with no experience or expectation of privacy, an externally imposed standard of conduct can metastasize into an internalized standard of behaviour. The altered standard then becomes ‘the new normal’, especially when critical voices are seen to be marginalized.

The Chinese ‘Social Credit’ system is a recent example of what überveillance can look like in practice. This system is a personal reputation system launched by the Chinese government under which a combination of big data analysis and advanced surveillance techniques produces a personal profile of every Chinese citizen on the Chinese mainland. Online behaviour, behaviour in the public space, expressions of political dissent – all is recorded, and every data point means a plus or a minus on a unidimensional scoring system called the Social Credit Score. This score is then what determines the individual’s access to public services such as travel documents, access to transport, education or a good job. And it is not only one’s personal behaviour that counts: the behaviour of the social group to which the individual belongs also becomes part of the Social Credit Score. It seems that everyone becomes a shareholder in a collective standard, with the result being extreme social control, a prevailing distrust, and isolation of people with alternative views.
Gradually, a state-controlled model emerges in which uniformity in behaviour and thought is the goal, and cultural, intellectual and political diversity is suppressed. This goes completely against individual needs for privacy and autonomy and the integrity of the individual and her social environment, and moreover comes at the cost of the flexibility that is required to allow a complex society to respond adaptively and adequately to a changing world.

A state-controlled model emerges in which uniformity in behaviour and thought is the goal, and cultural, intellectual and political diversity is suppressed

Alongside the government, the major international tech companies (Google, Amazon, Microsoft, Facebook, etc.) are playing a major role in defining the digital landscape. The services of these major tech players are being embraced enthusiastically all around the world – but these companies have a tendency to play fast and loose with our privacy, and are dealing in our personal profiles and the data we generate with our online behaviour on an unimaginable scale and at a speed never before seen. Already, the ‘information inequality’ gap is huge, with the average user of an online service (whether it be a Google search, a Facebook profile, a ‘Nest Home’ thermostat, or an Amazon Alexa voice assistant) having no idea of what is happening with her information – what parties besides the service provider have access to this data, how it is being used, what this data is worth to third parties, what inferences are being made based on it, and how this is ultimately being brought back to bear on the user in the form of personal influencing of viewing, clicking and buying behaviour.32

No one reads the hundreds of pages of online contracts, license agreements, privacy policies or terms of service, and to even think that they have been read is plainly an unrealistic, even unethical expectation. A recent study presented internet users with a fictional social network called ‘NameDrop’, and demonstrated that of the hundreds of participants, three-fourths completely ignored the privacy policy and terms of service, with the remaining one-fourth reading these documents only extremely cursorily, even though 97% consented to the privacy policy and 93% to the terms of service. If any of them had actually read them, they would have no doubt noticed the absurd clauses like signing away their first-born child and sharing all personal information with their own employers and the American National Security Agency. Qualitative data shows that people experience such documents as a nuisance and want immediate access to the functionality of the service without having to think about the way in which, or the conditions under which, the service comes about.

This reality, which the authors proclaim as ‘The biggest lie on the internet’33, makes abuse by major companies pathetically simple even as to all appearances they adhere to the letter of the law. This fundamental information inequality is the foundation of what Shoshana Zuboff calls ‘surveillance capitalism’.34 As she describes it:

“Surveillance capitalism operates through unprecedented asymmetries in knowledge and the power that accrues to knowledge. Surveillance capitalists know everything about us, whereas their operations are designed to be unknowable to us. They accumulate vast domains of new knowledge from us, but not for us. They predict our futures for the sake of others’ gain, not ours.” (p.11).

While überveillance evokes an Orwellian dystopia of suppression and manipulation, Zuboff’s ‘surveillance capitalism’ much more readily evokes associations with Huxley’s Brave New World, a dystopia of self-imposed hedonistic nihilism. However, what both systems share is the framing of the individual as subordinate and uniform, with those exhibiting behaviour outside the collective norm being ignored, marginalized, ostracized or otherwise punished. But individual variation is a central ingredient in cultural progress and societal flexibility. In a society where the complexity of challenges in the digital age only increases, a top-down, centralized, one-size-fits-all policy is a miscalculation. The conflicts and trade-offs we know, such as between privacy and security, between economy and ecology, or between the individual interest and that of the collective, immediately make clear that we cannot simply optimize for a singular target function. In reality, there is no such function.
Digital resources for measuring and influencing can be powerful tools for a government that decides to go about its work in a data-driven, evidence-based way. The paradox is that the resources for surveillance that we have taken such a critical look at here can also lead to a better-informed government that can achieve its goals more effectively and take care of its citizens better. The measurements necessary to close the cybernetic loop and optimize the process are the same measurements that can threaten our privacy.

We do not need to go as far as Orwellian repression to start feeling like our privacy is at stake. Infringement of privacy can start with nothing more than the unintended side effects of valid attempts to gather information. Here, the government needs to act with restraint and from a self-critical perspective: its attempts to gather information or influence participants in society must never straitjacket the individual, impinge on the personal domain of the citizen, or lead to lack of transparency and drastic forms of information asymmetry. Participation of the public and other interested parties is vital here.

6. Towards participatory design of the digital public space

Systems are not more important than people. The digital transition is not a goal in and of itself, but has to serve the interest of the public, and must contribute to solving the urgent societal problems that we all face. All too often, it is only after the introduction of a policy measure that any thought goes towards its effect on the public. The central assumption in the design, implementation, management and regulation of the digital society (the smart society and the smart city) has to be that it is human-oriented. It is vital to maintain the principles of user-centered and value-sensitive design. This means that the parties involved must not only be informed about developments that are already in motion, or only have input as a participant in a living lab, but that they are actively and continuously engaged in the preliminary trajectory and throughout the development, introduction and scaling up of innovations. The preliminary trajectory must also explicitly include a needs analysis so as to incorporate not only a ‘tech-push’ perspective, but also the ‘user-pull’ perspective.

The question cannot be: what are all the things we can do with digital technology? Instead, it has to be: what do people really need? What problems are we going to solve? What people is this going to affect? Is everyone sufficiently engaged, and can everyone participate? Is there room for customisation, for personal attention and assistance? Are there sufficient possibilities for adjusting, suspending or even abandoning an innovation if it proves to not work adequately? Here, there needs to be extra attention to vulnerable groups – low-literacy groups, ethnic minorities, people with disabilities and older people who simply cannot keep up with the digital revolution, to name just a few. These are, in some sense, our ‘canaries in the coal mine’. They are extra-sensitive to disadvantageous circumstances and are therefore an important leading indicator for the rest of society.

In this regard it is worth highlighting the added value of living labs and experiments to bring ethical dilemmas to life and open the discussion on them for a broader public. In their present incarnation, living labs are primarily useful for testing out technical innovations in practice. However, what has remained underemphasized so far is bringing a sense of reality to the ethical dilemmas that the increasing
digitalization of our world entails. At present this discussion seems to revolve primarily around speculative design, cultural events, critical journalism, science and science fiction. This is extremely valuable for the societal debate, but these forms of discussion generally lack a connection with the substantial majority of ‘ordinary’ inhabitants of a city, including the vulnerable groups referred to above. This distance can also give rise to unrealistic expectations, both in the positive and negative senses, and create images that are at odds with the lived experience – the concrete, actual reality of everyday life. Members of society must be able to gain a broad familiarity with the new technology, in everyday situations, and must be actively engaged in the dialogue around the dilemmas that the new technology presents. This has to be structurally anchored in the development trajectory and political decision-making surrounding the introduction of new technology in the living environment.

**Members of society must be able to gain a broad familiarity with the new technology, in everyday situations, and must be actively engaged in the dialogue around the dilemmas that the new technology presents**

In addition, it is of major importance to also include, on principle, the values and interests that are all too often denied a voice. These are, for example, the interests of future (as yet unborn) generations, of broader and more complex systems (such as the global climate or our living environment), of unforeseen and undesired effects over time, or of effects that arise when a technology is scaled up or used for purposes other than as originally intended. This type of consideration demands a broader ethical review framework, perhaps one based on virtue ethics. Where user-centered design is appropriate for consideration of the interests of savvy and active stakeholders, a more principled ethical consideration is needed to make a broader evaluation, and to include an embedding of collective values that do not yet have a clear voice in the innovation process.

7. Digital literacy, critical citizens and an honest government

The government has the responsibility to utilize the options emerging from the digital transition in the interests of its citizens as much as possible, but also to protect society from the excrescences and unintended effects of the digital transition. Digital literacy is an important part of living up to this responsibility. Digitally literate citizens are capable of using digital technology effectively and efficiently, taking advantage of all the potential benefits it offers for education and work, enterprise and political and social participation.

A digitally literate citizen who can use participatory digital networks and platforms, who participates in ‘citizen science’ projects, and who can have a meaningful input in shaping the digital transition of his own living environment is a valuable partner for the government. A digitally literate citizen is also digitally resilient: better able to recognize hacking and phishing attempts, to prevent identity theft and to distinguish fake news from real news. But most of all, a digitally literate citizen is also a critical citizen, who is better capable of keeping a critical eye on the government or market parties and pushing back when things like personal privacy or autonomy are threatened.

In education, there have been loud calls for digital literacy for some time, but only within a limited spectrum of subjects oriented around the use and risks of social media, mobile devices and cyber-bullying. A desirable goal is to expand these efforts both quantitatively and qualitatively, and to extend them into ‘cyber-physical’ systems (the Internet of Things), personal tracking (sensors, smartwatches, etc.), surveillance technology (biometrics, face-reading, etc.) and the mechanisms and risks of digital hypernudging and microtargeting (pointcasting, personalization), both online and in the mobile arena, as well as in digital games and the physical world. On all this, we need to create awareness of both the role and the responsibility of the government in a digital society, but also of the active role that the individual herself can and must play. In such a process of awareness-raising, we must also take a critical look at the role, the strategies and the interests of the major international businesses that have made the trade in personal data and the fight for user attention their business model.
Appropriate use of data in public space
To design such a programme, it will be important to develop a coherent curriculum oriented around multiple levels of education and literacy. Here again, social inclusion and diversity (as already described) are critical. Also, a connection with a broader cybersecurity initiative would seem to be opportunite here. Obviously, without digital hygiene in the personal, business and government domains, cybersecurity remains something of a game of whack-a-mole.

In this context, there is a need for a coordinating party (potentially, organised from a collaboration between different departments of government) to take the initiative, coordinate and take a hand in the design of the curriculum development, but to do this in close partnership with civil society. The knowledge and expertise available among tech entrepreneurs, advisory bodies and citizen initiatives is extremely valuable and must be mobilized. Here, the ‘right to challenge’ is also worth noting; the government that signed the coalition agreement in 2017 has committed to exploring the options for ‘using a Right to Challenge system to give individuals and local associations the option to submit an alternative proposal for the implementation of collective facilities in their immediate environment’. This will promote democratic participation.

Much has been said and written about transparency: transparency in decision-making processes of the government, and transparency in AI algorithms, both to make it as easy as possible for the public to understand the basis on which certain decisions are being made. Perhaps even more important, but frequently taken for granted, is honesty. Transparency is not a virtue in and of itself, but a process condition without which verification could not be possible. Honesty is a moral value, a virtue, a flag that the government needs to fly high. As we have seen in the foregoing, transparency does not automatically mean honesty – recall the example of the hundreds of pages of fine print in terms of service and other contracts that users of online services are faced with. Clearly, this results in information overload rather than transparency. The same can be said for how sensitive negotiations can be disrupted when confidential information is “transparently” shared on the internet and recordings and transcripts are put out on the streets.
8. Conclusions and recommendations

The digital transition is having an unprecedented impact on our world. Rapid technological progress in our digital society also affects government. Big data is being used in a wide variety of contexts to increase efficiency, like city maintenance and effective deployment of the police. Technology is also being used to make the city more sustainable and more liveable – examples include energy-saving lampposts that only come on when someone walks by, or sensors that automatically detect noise pollution in an area. At the same time, the government is also struggling with the speed and scale at which the digital transformation is taking place. Much of the current infrastructure (buildings, roads, energy systems and public transportation) has yet to go through the digital transition. A number of pioneering cities have been experimenting with smart city pilot projects and living labs to get a sense of the potential added value of digital technology in the public space – but also to start assessing where the problem areas lie and where legislation, regulations and ethical frameworks will be needed to keep all innovations moving in the right direction.

As the transition to an intelligent digital public infrastructure progresses, new options for influencing the behaviour of individuals in society to achieve positive social outcomes are emerging: to prevent aggressive incidents, for example, or to discourage antisocial driving behaviour, or to promote an active lifestyle. It is in those areas in which the government is pursuing an active prevention policy – safety, health, social behaviour, environmental conservation – that the combination of digital technology with knowledge of human psychology can produce positive results. But this combination also immediately raises difficult questions. How can the government protect the personal lives of individuals while at the same time keeping a view to security and law, social behaviour and healthy living habits, and the protection of the living environment? How and to what extent should the government be allowed to influence the public’s behaviour towards the desired using public data flows, artificial intelligence and digital persuasion techniques? And how must the government regulate itself when faced with the temptations of the potential power and control that the digital transformation brings with it?

To these questions, there are no easy answers and no simple solutions. The complexity of society and the ethical challenges facing it are simply too great.
With all of this in mind, I would like to conclude this essay with a number of recommendations for policy.

1. If the government wishes to use technology for the purposes of behavioural change in the public domain, this must be done transparently, on the basis of democratic decision-making, ethical review and scientific evidence, and with the explicit involvement and consent of the parties that this technology is going to affect either directly or indirectly. The government must at all times communicate proactively, openly, comprehensively and honestly about the digital transition in the public domain, or any plans in that direction, at all levels – from national and provincial to municipal and at the level of the community.

2. All parties that are affected by digital behavioural influencing in the public space, whether directly or indirectly, need to be engaged through a process of participatory design. This process should start at the early planning stages of a project, and should be implemented continuously throughout the project lifetime. The government needs to make the extra effort to give the most vulnerable groups in society an active voice in this process. In addition, the government needs to explicitly consider ethics in the design process, and make ethical values an inherent part of the design. The government has the primary responsibility for organizing this process, together with its social partners.

3. The government must increase its digital competences and as part of this inform itself properly of the potential and limitations of the most recent forms of digital technology. A technologically naïve government is easy prey for bad actors, a bad partner for industry, and a failure as a guardian of the public interest.

4. Data ownership and the privacy of individuals in the public space must be respected and protected. The government must eschew unproven, unsafe or less than robust technology. Unintended, unforeseen or illegal use of the technology and personal data must be monitored and regulated at all times.

5. The government must invite, organise, and engage with constructive criticism and opposing voices from society. This all must come along with major investments in the digital literacy of the public, including more attention to the latest technological advancements and how these will affect the fundamental ethical values within our digital society. The reinforcement of technological citizenship can take on many forms, such as debates, courses, living lab demonstrations, critical cultural projects or speculative designs.

6. The government must organize adequate and independent critical monitoring and supervision, and embed this in the law. Examples include an ongoing process of encouraging and organizing the values debate, discussing relevant ethical and legal precedents, and review of legal, ethical and governance frameworks through elected representation at all levels (municipal and provincial councils, the Upper and Lower Houses of Parliament). It is also important to keep civil society and relevant advisory bodies engaged at all times and to augment the role and position of regulatory authorities. This also means overseeing coordination across enforcement domains.

To sum up, the most important advice is that we must mobilize, embrace and engage the diversity of our society – not as a necessary democratic concession to the citizenry, but because this is the way forward towards a broader and more varied cultural and intellectual basis for collective decision-making. As sociologist Dirk Helbing and his co-authors argued convincingly, there is great strength in collective intelligence in the governance of our present-day complex digital society. This requires organized consultation and participation, valuing variation above uniformity, and the stimulation of independent information-gathering and local decision-making authority. As Helbing writes, socio-diversity is just as important as biodiversity.

We need the diversity in our society, to mobilize, to embrace and involve

A digitally responsible government, even one struggling on the slippery slope between nudging and überveillance, will benefit from a digitally literate and critical citizenry, a lively democratic debate, participation from a broad spectrum of societal stakeholders, a solid scientific basis, independent and transparent legal control, and ongoing ethical review. The greatest strength lies in flexibility through diversity, which will lead to robust and responsive governance of our digital society.
About the author

Wijnand IJsselsteijn is Full Professor of Cognition and Affect in Human-Technology Interaction and academic director of the Center for Humans & Technology at Eindhoven University of Technology (TU/e).
References
1 Moore’s Law was postulated in 1965 by Gordon Moore, one of the founders of Intel, and described the trend observed that the number of transistors that can be fit on an integrated circuit (that is, a ‘chip’) doubles approximately every two years. Alongside Moore’s Law, in adjacent areas of microelectronics, we see comparable exponential trends in areas like digital storage, optical resolution and communication speeds. This exponential curve, which has held for decades, is the driving force behind our use of computing power and storage in virtually every area of human life. Its impact has become so dramatic that in 2016 the World Economic Forum went so far as to refer to it as a fourth Industrial Revolution, with a convergence of a broad range of technology trends in robotics, artificial intelligence, 3D printing, smart cities, nanotechnology and biotechnology. Moore’s ‘Law’ is, of course, more a marketing strategy than a physical principle, but it has more or less held true up to the present day. There are indications that we have reached the point where further miniaturization of integrated circuits will become increasingly difficult, given that we are beginning to run up against the limits of what is physically possible, and further miniaturization will increasingly result in unstable systems.
3 A sensor is in essence an artificial sensory organ that registers observations and either saves and/or transmits this observation data. Sensors can register a wide variety of observations, including sound, light, proximity, movement, air quality, air pressure, precipitation and temperature. A sensor will generally be permanently present in the space (in contrast to temporary setups like the temporary speed cameras placed along roads) and make observations continuously and without human intervention (i.e., anything requiring human action, like manually counting participants in traffic, is not considered sensory input). Sensors in the public space generally make use of existing infrastructure both for their physical placement (on lampposts, exterior walls of public buildings, roadides, etc.) and data communication (existing cabling, mobile networks, etc.).
9 These terms (ubiquitous computing, pervasive computing, context-aware computing and ambient intelligence) refer more or less (barring a few subtle distinctions) to the same technological advancements. Computers are increasingly an integrated component of our environment, hard-wired into everyday objects from lamps to refrigerators to cameras to televisions to thermostats. With sensors and interconnectivity, they are learning to understand their environment and their users better and better. The idea is that these devices can learn to serve their users better over time by adjusting their services on the basis of the user profile that they gradually compile. These days, this is broadly referred to as the ‘Internet of Things’ (IoT), which emphasises the importance of connectivity of these devices. The IoT is made up of semi-smart devices that are all connected to the internet, with their own individual IP addresses, and which as such can be operated and programmed remotely (e.g., via a mobile app) and can exchange information with each other.
10 This is not science fiction. Many smart watches already measure heartbeat and respiration rate, and this data can be used to derive other information, such as the subject’s stress level. In 2012 Microsoft filed a patent entitled ‘Targeting advertisements based on emotion’. In it, Microsoft discloses a system that uses its Kinect camera (a peripheral of Microsoft’s Xbox platform) to assess emotions and link this data to personal ‘optimised’ advertisements.
12 For a good survey and a model of various potential influencing strategies that can be used in ambient persuasion, see: Kaptein, M. C., Markopoulos, P., De Ruyter, B., & Aarts, E. (2010). Persuasion in ambient intelligence. Journal of Ambient Intelligence and Humanized Computing, 1(1), 43-56.
15 For a good survey and a model of various potential influencing strategies that can be used in ambient persuasion, see: Kaptein, M. C., Markopoulos, P., De Ruyter, B., & Aarts, E. (2010). Persuasion in ambient intelligence. Journal of Ambient Intelligence and Humanized Computing, 1(1), 43-56.
16 The term ‘big data’ refers to the application of analysis techniques on large, often unstructured data sets. Data from sensors is one good example, but consider also the data set gained from tracking online behaviour. All applications are able to identify the patterns and correlations within such data sets that lead to new knowledge and insight.
17 This term was introduced in 2017 by Karen Yeung: ‘Unlike the static Nudges popularised by Thaler and Sunstein such as placing the salad in front of the lasagne to encourage healthy eating, Big Data analytic nudges are extremely powerful and potent due to their networked, continuously updated, dynamic and pervasive nature (hence “hypernudge”).’ See: Yeung, K. (2017). ‘Hypernudge’: Big Data as a mode of regulation by design. Information, Communication & Society, 20(1), 118-136.

19 Cybernetics was defined by Norbert Wiener in 1948 as ‘the scientific study of control and communication in the animal and the machine’. Cybernetics is a broad field that has interfaces with AI; its objective is to understand functions and processes of focused, complex systems by comparing observations (data) to the desired target state of the system, and on the basis of the results take action to reduce any discrepancies between the present state and the desired target state. One simple example is a thermostat: it has a target state (the desired indoor temperature) and a set of observations (the current temperature). When the measured temperature is lower than that of the desired temperature, then the heating will be turned on (action), until the moment that the difference between the measured and the desired temperature is zero. This circular or iterative loop of observation-target function-action can be identified in almost all automated systems.


32 For a salient example, see Maurits Martijn and Dimitri Tokmetzis’ book Je heb’t wel iets te verbergen [Yes, you do have something to hide] (2016, De Correspondent). In it, they describe how they installed a tiny piece of sniffer software (named Charles) to register the network traffic between a Samsung smartphone and external parties. An online order of a leather bag from de Bijenkorf (a major Dutch department store chain) immediately set off a frenzied online bidding process for the personal data. ‘In just one second, connections were established with servers in the United States, Sweden, Germany, Ireland and the Netherlands, belonging to companies going by names like Improve Digital, Admeta, Adtech, Metrigo, Burst Media, Yieldlab, Switch Concepts, AppNexus, Socionicianc, Adscale, Rubicon Project, OpenEx, Smart Adserver and Casale Media...’ (p. 28).


35 In an article about the unintended side effect of sensors entitled ‘Maak van sensornetzwerk geen buurtwacht!’ [Don’t Make a sensor network your neighbourhood watch], Ardy Siegert presents a number of fictitious but realistic examples. See: https://ibestuur.nl/podium/maaik-en-sensornetzwerk-geen-nieuwsjonge-buurtwacht


Public governance of experimental data & algorithms

Recommendations for a national algorithm register and reporting framework

Prof. dr. Gerd Kortuem
Professor of Internet of Things
Delft University of Technology

Foreword
By Theo Veltman

Prof. Gerd Kortuem and I have met several times within the context of our involvement with the City of Amsterdam. He is a friendly, modest, highly skilled and influential scientist and is also one of the people responsible for overseeing the vital and all-encompassing field of technology. He brings extensive knowledge and substantial experience to this role and expertly applies it to the issues faced by humanity in today’s digitized world. This expertise is also clearly reflected in his essay.

Gerd’s research focuses on the Internet of Things: the design and use of data, algorithms and smart devices as important building blocks for intelligent products and services. He is a member of the AITech initiative, which seeks to establish meaningful human control of autonomous intelligent technology. Gerd is a professor at Delft University of Technology, where he is Chair of the Internet of Things within the Faculty of Industrial Design Engineering as well as being affiliated with the Amsterdam Institute for Advanced Metropolitan Solutions (AMS).

In my opinion, ‘Appropriate Data Use in Public Space’ is an important issue as our freedom – a vital cornerstone of society – is under threat from the ever-increasing digitization of society: the freedom to go about our lives without our movements being monitored or registered, the freedom to do as we wish without being constantly supervised, and the freedom to form our own opinions without unwanted and imperceptible influencing of our decisions and views. All of these freedoms are currently under threat. Digitization undoubtedly creates a wealth of opportunities and benefits. It provides everybody with access – at any time and in any place – to all the information they need to develop themselves personally and professionally: a benefit that we must ensure is available to all. However, inappropriate data use is an undesired side effect and we must do everything we can to minimize it.

My participation in the Supervisory Committee for Appropriate Data Use in Public Space taught me how we can learn from one another. Being a part of this group of driven and highly skilled experts gave me a great deal of energy and gave me many new insights.

Theo Veltman
Member of the Supervisory Committee
Summary

Data and algorithms have taken on a defining role in society. They affect what news we read online, how traffic is routed throughout a city and how we relate to each other. Increasingly, companies and governments use machine learning techniques to automate decisions with significant human and societal implications, for example in policing, criminal sentencing and social services. However, the algorithmic decision-making is being criticised for its potential to increase bias and discrimination, and its lack of transparency and accountability. Thus, one of today’s key questions for governments is how to govern the design, development and use of data and algorithms in society to maximise public benefits and promote public values.

Effective governance data and algorithms is difficult for a variety of reasons. First, the sophistication and complexity of algorithms is rapidly rising to a point where even computer experts have problems understanding how algorithms operate and how they make decisions. Second, the increasing adoption of automated machine learning tools and self-learning algorithms is leading to algorithmic systems that are rapidly evolving in unexpected and often unpredictable ways. Third, the impacts of data-driven algorithms on society are wide and far-ranging, and are impossible to predict a priori, before such systems are actually deployed in society. For these reasons, Living Labs (Schuurman, 2015) play a key role in the experimental development of data and algorithms.

In this article, I investigate ethical questions of data and algorithm experimentation and develop recommendations for a national framework for data and algorithm governance comprised of five components: 1) an Algorithm Reporting Initiative, with the aim of documenting and tracking the societal value and risks of algorithm projects across the Netherlands; 2) a National Algorithm Register with the aim of enabling effective comparison and assessment of data and algorithm initiatives across the Netherlands; 3) a Public Algorithm Forum with the aim of enabling public stakeholders to debate, critique and contest the operation, use and outcomes of data and algorithm initiatives; 4) a Data and Algorithm Institute to drive the development of knowledge, approaches, tools, infrastructures and standards for ethical and accountable use of algorithms; and 5) a National Data and Algorithm Skills Agenda to ensure that all stakeholders (from citizens to organisations) are sufficiently skilled to participate in the new data and algorithm economy.

These recommendations are informed by several key sources, including Van de Poel’s Experimental Ethics (van de Poel, 2016), the FAT/ML “Principles for Accountable Algorithms and Social Impact Statement for Algorithms” (Diakopoulos et al., 2018) the “Integrated Reporting” framework (Banerjee, 2019; Eccles & Krzus, 2015), a widely adopted framework for tracking and reporting on financial and non-financial value creation of public and private organizations, and NESTA’s work on AI ethics and governance (Mulgan, 2016).
1. Introduction

In 2001, Marc Andreessen famously pronounced that “software is eating the world” – a statement highlighting the dramatic technological and economic shift in which software companies are taking over ever larger swaths of the economy (Andreessen, 2011). Today, a similar shift is felt throughout society caused by the ever-larger role of data-driven algorithms in making decisions with significant human and societal implications, including policing, criminal sentencing, lending, and hiring. Today’s unprecedented availability of large-scale data, combined with increasing powerful machine learning methods allows companies, governments and the public sector to train algorithms for tackling complex problems (Willson, 2017). Data-driven algorithms may improve public service delivery and government efficiency, by helping to optimize processes, analyse feedback and predict outcomes. However, algorithmic decision-making has been criticized for its potential to contribute to bias and discrimination, and its lack of transparency and accountability. Thus, the most fundamental question with respect to data and algorithms is:

How can we effectively govern the design, development and use of data and algorithms in society to maximise public benefits and promote public values?

In this paper, I will try to answer this question by developing five concrete recommendations for the ethical and effective governance of data-driven algorithms. Leading up to these recommendations, I will make seven key arguments throughout Section 2 and Section 3:

- Decision making processes with far-reaching societal implications are increasingly enshrined in data-driven algorithms.
- The sophistication and complexity of algorithms is rapidly increasing to a point where even computer experts have problems understanding how data-driven algorithms operate.
- The increasing adoption of automated machine learning processes by organisations is vastly accelerating the design, development and deployment of data-driven systems.
- Emerging machine learning systems are self-learning – they are constantly changing and evolving
- The impacts of data-driven algorithms on society are impossible to predict a priori, before they are actually deployed.
- Living labs provide an effective environment for the gradual and experimental introduction of data-driven algorithms into society.
- We are lacking governance frameworks for the safe and ethical experimentation with data and algorithms in society.

These arguments lead me to the following five recommendations which are outlined in Section 5:

- Recommendation 1: Set up an Algorithm Reporting Initiative with the aim of documenting and tracking the societal value and risks of algorithm projects across the Netherlands.
- Recommendation 2: Develop a National Algorithm Register with the aim of enabling effective comparison and assessment of data and algorithm initiatives across the Netherlands.
- Recommendation 3: Develop a Public Algorithm Forum with the aim of enabling public stakeholders to debate, critique and contest the operation, use and outcomes of data and algorithm initiatives.
- Recommendation 4: Establish a Data and Algorithm Institute to drive the development of knowledge, approaches, tools, infrastructures and standards for ethical and accountable use of algorithms.
- Recommendation 5: Develop a National Data and Algorithm Skills Agenda to ensure that all stakeholders (from citizens to organisations) are sufficiently skilled to participate in the new data and algorithm economy.

These recommendations are informed by a review of literature, my own experiences in developing data-driven algorithms and systems in various smart city projects, and insights from three key sources, namely Van de Poel’s Experimental Ethics (van de Poel, 2016), the FAT/ML “Principles for Accountable Algorithms and Social Impact Statement for Algorithms” (Diakopoulos et al., 2018) and Integrated Reporting (Banerjee, 2019; Eccles & Krzus, 2015), a widely adopted framework for tracking and reporting on financial and non-financial value creation of public and private organizations.
2. Data and Algorithms in Public Society

Algorithms are increasingly important for our private and public lives and have an enormous influence on public society. They effect what news we read online, how traffic is routed throughout a city and how public and private organizations conduct their business and deliver services.

2.1 Data, Algorithms and Machine Learning

Traditionally, algorithms have been defined in a procedural sense, as a set of instructions to produce an output or result. For example, Kitchin defines algorithms as:

“sets of defined steps structured to process instructions/data to produce an output.”
(Kitchin, 2017a)

Similarly, Kraemer et al. proposed the following definition (Kraemer, van Overveld, & Peterson, 2011):

“An algorithm is, roughly speaking, a finite sequence of well-defined instructions that describe in sufficiently great detail how to solve a problem.”

These definitions highlight the procedural aspect of algorithms and imply that an algorithm can be fully specified (by a human developer) by fully enumerating the steps for computing and output or solving a problem. However, this procedural and declarative definition no longer hold true in the context of big data and machine learning.

Machine learning is a category of algorithm that allows software systems to automatically compute outcomes by learning patterns from data without explicitly being programmed by a human developer. Instead of specifying an algorithm, a machine learning engineer merely supervises the machine learning process by providing training data, selecting an appropriate type of machine learning algorithms and assessing the machine learning output. The outcome of this training and learning process is a machine learning model that, given input data, can make decisions and improve its performance with little further involvement of a human trainer.

The key difference between a traditional algorithm and a machine learning algorithm lies in the use of data. A traditional algorithm operates on data, a machine learning algorithm is trained on data and then operates on data. The key advantage of machine learning over conventional algorithms is that machine learning can be used in cases where there is no known solution or approach to solving a problem, or where it is extremely hard to find and specify an algorithm “by hand” (for example, image analysis, text understanding, and the control and optimization of complex systems).

While technically a machine learning model is still an algorithm, it only exists in an internal representation (for example, a neural net) which can be inscrutable even for machine learning experts. In other words, the developer might not know how and why a machine learning system generates a certain output. This is the well-known black-box problem of machine learning systems (Gasser & Almeida, 2017; Pasquale, 2015)

The direct relationship between data and algorithms implies that data quality determines the quality of (data-driven) algorithms. It is impossible to develop high-quality algorithms with low-quality data.
Both data and algorithms need to be vetted for quality attributes such as accuracy, completeness, reliability, relevance, timeliness, lack of bias, etc. Thus, governance and policy need to treat data and algorithms as unity.

**It is impossible to develop high-quality algorithms with low-quality data**

### 2.2 Algorithms vs Algorithmic Systems

An algorithm is not just a technical construct. Kitchin argues that algorithms are “performative in nature and embedded in wider socio-technical assemblages” (Kitchin, 2017b). This means that to understand algorithms in public society, we need to analyse why they are used, how they function in the real-world, how users respond to them and the consequences of the tasks they perform (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). Employing the same algorithm can lead to disparate outcomes in different situations. Kitchin states that: “algorithms perform in context – in collaboration with data, technologies, people, etc. under varying conditions – and therefore their effects unfold in contingent and relational ways, producing localized and situated outcomes” (Kitchin, 2017b). Thus, algorithms are not just objective, value-neutral, artefacts. The human choices that determine their use, their operation and their impacts are influenced by subjective, often-implicit value judgments. In other words, an ethical investigation of algorithms must consider the making of algorithms, the algorithms themselves, the context in which they are used and their impacts on people and society.

### 2.3 The Increasing Importance of Algorithms in Society

Government agencies, public organisations and municipalities across the Netherlands are rapidly increasing their use of digital systems to tackle key societal challenges and provide essential services. More and more, the core of such systems consists of data-driven algorithms that streamline or even automate the provision of public services (Kitchin, 2014). For example, cities are using satellite imaging and computer vision to automatically check compliance of construction projects from the air; transport authorities are using ticketing data to track commuter traffic and predict future mobility demand; energy companies are using data from electricity meters to track and predict energy consumption and CO2 emissions; and police is using data from cameras and microphones to detect crime in real-time and identify crime hotspots. In addition, Companies like Uber and AirB&B use data and algorithms to provide new transportation and housing services.

### 2.4 Understanding Algorithms

When considering the role of algorithms in public society – end the ethical questions involved – we need to ask three key questions:

- Who is controlling the use and deployment of algorithms – and in whose interests?
- How is the use of algorithms justified?
- For what purpose are algorithms used?

**Who? Public, Public-interest and Private Algorithms**

The use of algorithms has spread throughout society and is affecting all aspects of daily life. To understand who is controlling the use and deployment of algorithms I propose to classify algorithms into three categories: Public, Public-interest and Private Algorithms.

**Private algorithms**: The largest impact of algorithms in society comes from algorithms used by private companies like Google, Facebook, AirB&B and Uber. I call algorithms used by such companies private algorithms. They are private because a) they serve private (i.e. commercial) interests, b) are controlled by small circles of individuals who are not necessarily known outside of the company and c) their function and operation are typically not disclosed to external stakeholders. The most fundamental example is Google’s PageRank algorithm, which calculates the relative importance of Web sources and thus influence what users see on the web and what they don’t. The only way for public society to understand private algorithms is by observing their externally visible behaviour and the direct and indirect impacts of algorithms on individuals and society.

**Public algorithms**: Algorithms are increasingly used by government for policy making and for providing public services. The automated issuing of speed fines, profiling in predictive policing and automated decisions about welfare are all
relevant examples of what I call **public algorithms**. Such public algorithms no longer simply help humans in government agencies apply procedural rules; instead, they have become primary decision-makers in public policy (Eubanks, 2015). Eubanks succinctly describes this situation: “The algorithms that dominate policymaking—particularly in public services such as law enforcement, welfare, and child protection—act less like data sifters and more like gatekeepers, mediating access to public resources, assessing risks, and sorting groups of people into ‘deserving’ and ‘undeserving’ and ‘suspicious’ and ‘unsuspicious’ categories” (Eubanks, 2015). Public algorithms are public because they are employed by public organizations in the public interest. In theory, citizens have a democratic right to understand how and why such algorithms are used; in practice however, public organizations and government agencies are no more willing or able to disclose their algorithms than commercial companies.

**Public-interest Algorithms**: In between public and private algorithms we find what I term public-interest algorithms, which are algorithms used as part of a public service delivered by a private company or a public/private partnership. This class of algorithms comprises, for example, algorithms used by electricity utilities, water companies and traffic authorities. A concrete example is the use of smart EV charging algorithms in the electricity grid which make real-time decisions about how much electricity each EV can charge when connected to a charge point (Liu, McNamara, & McLoone, 2013; Turel, Joskin, Geerts, Kaathoven, & Schouwenaar, 2017). Such algorithms are neither entirely public nor private. They are used for providing an essential public service (thus serve public interest), but may be controlled by private entities with their own commercial interests. Organizations who operate and control these algorithms are typically neither obliged nor willing to disclose their algorithms and there are potential conflicts of interests between public stakeholders (society at large) and private stakeholders (management team, owners etc.).

Why: The argument for using algorithms in public policy

Private algorithms are used by commercial entities for a simple reason: first and foremost, to further their commercial and business interests. Users might derive benefits, but this is only of secondary concerns.

<table>
<thead>
<tr>
<th>Public Algorithms</th>
<th>Public-Interest Algorithms</th>
<th>Private Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Subject to public oversight</td>
<td>2. Controlled by private companies or public/private partnerships</td>
<td>2. Tightly controlled</td>
</tr>
<tr>
<td>3. Some transparency and disclosure (in theory)</td>
<td>3. Limited transparency and disclosure</td>
<td>3. No transparency and disclose</td>
</tr>
</tbody>
</table>

Figure 1 Algorithm Classification

The use of public algorithms in policy and government can be motivated by their contribution to public interest. According to Robert Goodin’s theory of public welfare utilitarianism (Goodin, 1995), public policy-makers have an obligation to act in such a way that it maximizes the public interest. The employment of algorithms can lead to vast improvements in how public institutions perform their tasks (Kroeskop, 2018) by creating new policy opportunities, increasing efficiency of government and public services, improving the consistency and objectivity of decision-making and reducing human error (Achrekar, Gandhe, Lazarus, Yu, & Liu, 2011; Chen & Hsieh, 2014; Joseph & Johnson, 2013; Naik & Bhide, 2014; van der Voort, Klievink, Arnaboldi, & Meijer, 2019).

What? Measuring, Decision-Making vs Behaviour Change

To understand the use of data and applications in the public realm I introduce three broad usage patterns (Figure 2):

**Measuring and monitoring**: This involves cases where data and algorithms are primarily used for data collection and aggregation, but where interpretation and decision making is still largely in human hands. Examples are remote traffic monitoring, the use of data in urban planning and simple forms of predictive policing.
Automated decision making: Automated-decision making involves cases where data and algorithms are used to automate critical decisions. Examples are automatic traffic systems which adjust traffic light to reduce congestion, the automatic generation of speeding fines from camera, and more advanced forms of policing.

Influencing and changing behaviours: Behaviour change involves uses of data and algorithms with the aim to influence or "nudge" people to behave and consume public services such as transportation in a more sustainable way (Gandy & Nemorin, 2018; Ranchordás, 2019). For example, Enschede is combining real-time traffic data and a smart phone app (e.g, SMART-app: Self-Motivated and Rewarded Traveling) to encourage individuals to use alternative less congested routes when driving, or switch to public transportation or cycling, and reward them for ‘good behaviour’. This and other examples are based on Sunstein and Thaler’ seminal work on Nudging (Thaler Richard H., 2008), which laid the theoretical groundwork for combining behavioural and predictive models to effectively influence people to make more ‘wise’ decisions.

<table>
<thead>
<tr>
<th>Measuring and Monitoring</th>
<th>Automated Decision Making</th>
<th>Behavior Change</th>
</tr>
</thead>
</table>

Figure 2 Three Patterns of Data and Algorithm Use

Ethical Issues related to Algorithms
Technologies play a significant role in defining the social contract of the future. Data and algorithms have become covert tools for increasing surveillance, corporate profits and, at worst, social control. In a review of the ethical debate around algorithmic systems Mittelstadt et al. identify the following issues (Mittelstadt et al., 2016):

- Because algorithms make use of inductive correlations, the decisions that result from them may be unjustified.
- Algorithms are often poorly accessible for inspection. In addition, they are often hard to comprehend, particularly when machine learning is used. Lack of accessibility and comprehensibility produce algorithmic opacity which may harms the trust people have in algorithmic systems and the organizations that wield them.
- Any particular piece of technology supports certain values and undermines others. Algorithms are no exception. These biases can emerge from the deliberate or accidental transfer of values held by people building algorithmic systems. In addition, technical flaws and bad data can produce biased algorithms. Finally, unforeseen interactions between the algorithm and the world at large at use-time can produce biased consequences.
- Whereas bias is a quality of the decision making itself, discrimination is a quality of the outcome of an algorithmic system adversely impacting a particular group of people disproportionately.
- The practices of profiling, nudging and personalization are detrimental to human agency.
- Profiling, analytics and the sharing of data with third parties is harmful to informational privacy.
- The increasing unpredictability of algorithms and the distribution of control away from programmers into other parts of the system due to the use of machine learning produces questions of moral responsibility. In addition, the fact that algorithms consume the outputs of each other can lead to an accountability gap.

These issues span technical, organisation and public concerns (Figure 3).

On the technology level, ethical concerns arise because of the complexity and opacity of AI algorithms which translates in a lack of transparency and explainability, even for computer science for experts.

On the organisational and institutional level, key concerns relate to the accountability of decision making and the governance of algorithmic systems. The World Wide Web Foundation defines accountability as the “obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative
social impacts or potential harms” (World Wide Web Foundation, 2017). The definition of algorithmic accountability uses the same mechanisms as accountability in conventional public policy. When an algorithm is implemented into an information system for public policy, the system becomes the agent of behaviour to whom a government delegates its authority (Introna, 2016).

On the society level, the key concern is the legitimacy of the use of technology and the ability of citizens to meaningfully contest automated decision-making systems. Notorious and repeated failures of publicly deployed AI systems (people being misidentified for crimes they have not committed; or welfare support incorrectly refused) undermine trust in this technology but potentially also impact people’s perception of the legitimacy of the institution using this technology. The growing gap between the pace of technology development and the level of understanding amongst policymakers, technology experts, professionals and the general public has led researchers to postulate an AI “legitimacy gap” (Brown, 2018) and call for creating meaningful opportunities for the public and civil servants to debate and scrutinise possible uses of AI in the public realm.

2.6 Risks of Using Algorithm in the Public Sector

There is increasing desire by government to harness data and use algorithms in order to rationalize and automate the operation of public services and infrastructure (Janssen & Kuk, 2016). As discussed, algorithmically informed decision-making promises increased efficacy and fairness in the delivery of government services. However, these upsides come with a host of potential risks, such as the possibility of enshrining unwanted biases in algorithms (Angwin, Larson, Mattu, & Kirchner, 2016), the potential for systematic inequalities (Woodson, 2018). When improperly developed or implemented, algorithms can turn out to be less accurate than the judgment of government officials, and they can formalize and mask biases embedded in the data on which they are trained. Moreover, algorithms may enact policy judgments that diverge from the preferences of citizens or their elected representatives (Brauneis & Goodman, 2017) In addition, the use of algorithms potentially counter the need for flexibility in how policies are applied, and the right of citizens to effectively appeal or contest policy decisions (Saward, 2010).
Algorithms can cause real damage that is difficult to remedy under existing legal protections. Brauneis and Goodman formulated this situation as follows: written, “In the public sector, the opacity of algorithmic decision making is particularly problematic both because governmental decisions may be especially weighty, and because democratically-elected governments bear special duties of accountability” (Brauneis & Goodman, 2017). In sum, while there are many potential benefits of harnessing the power of data and algorithms, there are equally a number of arguments for not using algorithms in policy and government.

There is an urgent need to democratise the use of data and algorithms in society.

3. Algorithms as Experimental Technology

Data and algorithm are technologies in the making; they are not yet fully developed and entrenched in society (Brey, 2017). New technologies get embedded in society in a gradual process of co-evolution - as technology evolves, is adopted and adapted over time, society evolves as well. At some point in time, this process slows down, and the full impact and consequences of a technology become visible. This creates a fundamental challenge for policy makers, developers and society at large: how to sway the balance between desirable and undesirable outcomes of a technology before irreversibilities have set in (Rip & Schot, 2002; van de Poel, 2016) technology (patents? In the early stages of an emerging technology, when a technology and its social embedding are still malleable, knowledge of eventual impacts and how these will arise is limited. In later phases, social effects may be clear but then often the technology has become so entrenched in society that negative effects cannot be remedied. The fundamental limitation of society to effectively control and steer technologies was formulated by David Collingridge (Collingridge, 1982) as “control dilemma”.

3.1 Challenges in Controlling Algorithmic Technologies in Society

Data and algorithm technologies are especially susceptible to the control dilemma for four key reasons:

1. The pace of technological evolution in data and algorithm is extremely fast.
2. The development and deployment cycle of data and algorithm systems is getting shorter and shorter, allowing teams to deliver new features within days or hours, rather than weeks or months.
3. Even small changes in data and algorithms (for example choice of machine learning approach) can have large effects on performance, functionality and behaviour of such systems.
4. The impacts and effects of data and algorithms are far-reaching and difficult to predict.

Data and algorithms are malleable technologies that can be easily changed or adapted. The increasing availability of powerful cloud-based data and machine learning environments in combination with continuous software engineering (Fitzgerald & Stol, 2014, 2017) and DevOps methods (Bass, 2017; Ebert, Gallardo, Hernantes, & Serrano, 2016; Moore et al., 2016; Preimesberger, 2017) have radically shortened the development and deployment cycle of data and algorithms systems (Figure 4). While a few years ago it might have taken a larger team and several months to develop and deploy a powerful data-driven system, agile machine learning and automation of machine learning pipelines now allow developers to iterate and release a new version every few hours (Dougherty, 2019; Kobiels, 2014; Lwakatare, Raj, Bosch, Olsson, & Crnkovic, 2019).

Figure 4. Agile Machine Learning Cycle (Source (Follow,2019))
Appropriate use of data in public space
As a consequence, technological progress in form of new data analytic methods and new machine learning approaches can easily be incorporated into existing systems. Once a data set has been compiled, it is relatively trivial to apply new machine learning algorithms to it, potentially leading to significant enhancement in performance and functionality. Conversely, the rapidly increasing availability of new and enhanced data sets means that existing systems can easily be updated by simply retraining algorithms with new data. Overall, this means that the pace of technological development in data and algorithm systems is extremely fast compared to the time it takes to train people, to understand societal implications and to define policies and regulations. In addition, the speed of development and deployment of data and algorithm systems is expected to increase sharply as organizations start to automate machine learning processes to test hundreds of machine learning models in parallel. As a consequence, organizations will be able to develop, optimize and deploy data and algorithms with very little human intervention.

The above makes clear that data and algorithms fit van de Poel’s definition of an experimental technology, a technology for which there is only limited operational experience, so that social benefits and risks cannot, or at least not straightforwardly, be assessed on basis of experience (van de Poel, 2016).

Data and algorithms are highly malleable, constantly evolving technologies. The efficiency of data and algorithm engineering processes are rapidly increasing, leading to an ever-faster pace of innovation.

Data and algorithms are experimental technologies whose social benefits and risks

### 3.2 Algorithm Experimentation

The recognition of Collingridge’s control dilemma has led to the increasing use of experimental methods in designing, developing and evaluation technologies, in the form of living labs and – in the context of cities - urban living labs (Baccarne, Mechant, Schuurma, De Marez, & Colpaert, 2014; Cosgrave, Arbuthnot, & Tryfonas, 2013; Gascó, 2017; Kronsell & Mukhtar-Landgren, 2018; Steen & Bueren, 2017; Steen & van Bueren, 2018). According to the European Network of Living Labs (Schuurman, 2015), urban living labs are “user-centered, open innovation ecosystems based on a systematic user co-creation approach in public-private-people partnerships, integrating research and innovation processes in real life communities and settings”. Schaer describes living labs are moving from in-vitro to in-vivo research settings where data is collected from living environments like buildings or public spaces (Schaer, 2017).

Living labs play a key role in the development of public algorithmic systems and the co-creation of public services in mobility, energy and public safety by public (municipalities, government), private (technology companies) and society partners (community groups). Many living labs involve data actively or passively provided by the citizens and users and are thus especially sensitive with respect to rights of citizens and public governance.

**Urban Living Labs**

Today, living labs are seen as a key instrument for innovation in cities. One of the most discussed examples of a living lab is the Quayside development by Google’s Sidewalk Labs in Toronto, Canada which involves the development of large waterfront area in Toronto into a high-tech digital neighbourhood. The recently released development plan (*Sidewalk Labs Master Innovation and Development Plan*, 2019) envisions large scale experimentation with innovations like self-driving cars, public Wi-Fi, and new health care delivery solutions. While Sidewalk’s plans are unusually grand in scale and ambition, similar experimental projects exist in most major cities around the world.

**Living Labs in the Netherlands**

In 2017, the Municipality The Hague and the Hague Security Delta (HSD), a cluster of over 200 national cyber and urban security partner organisations, established a program for security innovation in the International Zone in Scheveningen to make the area increasingly attractive and secure for international organisations and residents alike (“International Zone,” n.d.). The International Zone is the home of international organisations like Europol, OPCW, and Eurojust and is used as site for conferences with government leaders from around the world.
The aim of this security initiative is to develop an integrated approach for improving security in the International Zone and reducing inconvenience for employees, residents and visitors. Partners in this initiative are the Municipality The Hague, The Hague Police, Ministry of Foreign Affairs, Central Government Real Estate Agency, Europol, Eurojust, OPCW, Peace Palace, Catshuis, TNO, Thales, Siemens, Crowd Sense. Since early 2019 the security initiative is being transformed into a Security Living Lab where companies, knowledge institutions and security services test how data from sensors (such as cameras and open-access Wifi points) in connection with AI algorithms can be harnessed for security and other purposes in a public space. Private firms attached to the scheme have access to the data supplied by the project, for the purposes of product development and as part of an on-going knowledge-sharing exercise with the municipality.

Another living lab is located in the Stratumseind entertainment district in Eindhoven to test nudging strategies to reduce crime and public disturbances. The Stratumseind living lab is a multidisciplinary collaboration of the municipality, the police and industry partners. Cameras and noise sensors are used to track visitor flows and behaviour on the streets and on-street lighting is dynamically adjusted using live data and algorithms to influence visitor behaviour. Similar to many such living labs public and private partners share data to co-develop systems that deliver societal and business value.

On-street lighting is dynamically adjusted using live data and algorithms to influence visitor behaviour

The living lab in Helmond focuses on innovation in sustainable energy, water re-usage, car-sharing and automated parking, smart street lighting and more. It features a small neighbourhood with houses and public infrastructure that have been equipped with sensors to measure energy and water use as well as use of shared cars. In addition, fine-grained behaviour of households and individuals is measured, for example the amount of sleeping time, the usage of household devices, and the time spent on social media. In return for their participation residents receive a monetary reward and pay a reduced rent (Kuyper, 2019).

Other living labs were established to design and test future intelligent energy systems, especially with a focus on electric mobility, for example the Vehicle 2 Grid project (“Vehicle2Grid,” 2016). At the heart of these living labs is the idea that in the future AI algorithms will be required to satisfy the electricity need of a shapely rising number of electric vehicles on the road. When demand outstrips electricity supply, data-driven algorithms will have to make decision of who receives how much electricity, potentially using data about car battery status and other driver information.

Living labs play a key role in the experimental development of data and algorithmic systems.
The Hague Security Living Lab at the International Zone in Scheveningen

Develop an integrated approach for improving security in the International Zone and reducing inconvenience for employees, residents and visitors.

Citizen data from public cameras and wifi networks.

Stratumseind entertainment district in Eindhoven

Monitor and influence public behavior to reduce crime and public disturbances.

Citizen data from public cameras.

Helmond Living Lab

Develop innovative neighbourhood services for sustainable energy, water re-usage, car-sharing and automated parking.

Data about residents’ domestic behaviours and use of water, energy and mobility services.

Vehicle 2 Grid

Design and test future intelligent energy systems to cope with future electricity demand from electric mobility.

Data about mobility and charging behaviour of electric vehicles.

Table 1. Urban Living Labs in The Netherlands

3.3 Ethical Issues of Algorithm Experimentation

Living labs are seen as alternative to an institutional approach to smart cities, where citizens are merely regarded as passive data subjects whose data is made to serve the needs of urban systems and organisations (Ranchordás, 2019). Citizens are meant to play an active role as co-innovators in order to ‘create, prototype, validate and test products, services, systems and technologies in a real-life setting’ (Westerlund & Leminen, 2018) (Steen & van Bueren, 2018). However, the mounting use of experimental technologies in public space and the increasing use of data and machine learning algorithms raises fundamental questions about the governance and ethics of these initiatives.

Schaer (Schaer, 2017) highlights four key concerns: 1) Freedom of choice: are people able to opt in or opt out of the experiment? 2) Informed participation: do people understand the nature of the experiment? 3) Governance and control: who determines how data and algorithms are used and who is responsible for the results and outcomes? 4) Absence of harm: are data and algorithms free of bias and discrimination?

**Freedom of choice and informed participation:** The influx of sensors, data and algorithms into the urban space means that soon we will have the same ability to influence (manipulate) people and conduct experiments in public space as companies like Facebook already have online. In many instances these experiments will require that people are unwitting participants as revealing too much information about the experiment could potentially influence the behaviour of people and thus invalidate the experimental setup. On the other hand, revealing too little means that people potentially take part in an experiment which - if fully informed - they would not agree to participate in. This dilemma was highlighted in a 2007 report by the European Expert Group on Science and Governance: “If citizens are routinely being enrolled without negotiation as experimental subjects, in experiments which are not called by name, then some serious ethical and social issues would have to be addressed” (Felt et al., 2007).

**Governance and control:** Another concern is introduced by the undue influence on the design and use of data-driven technologies by private companies. Increasingly, public institutions are consumers of data and AI technologies that are created, sold, and controlled by private companies, with potentially grave consequences for civil liberties. This undue influence can take many forms (Joh, 2017). Public organisations may be prevented by contract from disclosing information to the public even if they are required by law or policy to do so. Public organisations may be forced to contract with companies which have near monopolies in certain technology areas, potentially leading to increased costs or design choices that are not in the public interest, something described by (Joh, 2017) as ‘product design as policy’. Finally, lack of expertise by public organisations means that they may have little influence over design choices or may be unable to understand the implications of technology choices. Ultimately, these factors imply that the public at large has limited information over the technology that shapes key decisions about public matters and that the logic and rationale behind automated decisions
remains opaque even to the organisation using these technologies, increasing already existing ‘informational asymmetries’ between technology users and those who are subject to automated decisions, i.e. citizens (Joh, 2017)

**Bias and Fairness:** While machine learning techniques can be optimised to achieve astounding results in lab settings, the correctness and reliability of machine learning techniques in real world applications is often poor. Recent research has unearthed the dangers of hidden bias in AI systems (Campolo, Sanfilippo, Whittaker, & Crawford, 2017), and commercial systems have repeatedly been found to be inaccurate and biased (Angwin et al., 2016; Lebovits, 2019; Misra, 2018; Richer, 2018). For example, a tool for predicting the tendency of a convicted criminal to reoffend showed was shown to predict no better than a random online poll of people who have no criminal justice training at all (Dressel & Farid, 2018). An independent evaluation of the London Metropolitan police face recognition system found that it produced wrong results 81% of the time. Of 42 people flagged as matches on the wanted list, only eight were confirmed to be correct (“London police’s face recognition system gets it wrong 81% of the time,” 2019). This despite the fact that the London Metropolitan police had conducted a series of internal tests to guard against these problems. Evidence to the contrary, the London Metropolitan police insists its technology has an error rate of less than one in 1,000 instances, but so far has been unwilling to publicly share its methodology (“London police’s face recognition system gets it wrong 81% of the time,” 2019).

**Privacy:** Data collected in public space is likely to contain private information (e.g. sexual orientation; trade or political memberships; economic and physiological properties) that can be accidentally or intentionally extracted by third party data consumers (e.g. data annotators, partner organizations, citizens), thus enabling the disclosure of information that go beyond the original intent and purpose of the data collections. Current solutions for automatic privacy-preservation such as content redaction or the use of edge computing can improve privacy but ultimately cannot guarantee that all private and sensitive information removed, with important implications for GDPR compliance (Bourgeois, Kortuem, & Kawsar, 2018). Thus, living labs may best be served by a ‘bounded access model of disclosure’ (Fan & Jackson, 2015) which provides information access to trained professionals capable of effectively using data and honouring data protection regulations.
3.4 The Need for Public Governance of Algorithm Experimentation

Living labs in The Netherlands and elsewhere are only starting to deal with these ethical issues. At the Security Living Lab in The Hague, anyone who visits the International Zone is potentially recorded by one of the many cameras installed in the area, yet there is no public notice that video footage is recorded and potentially shared with partner organisations involved in the Living Lab. Visitors to the International Zone who use the area’s free public WiFi have to sign the WiFi’s terms and conditions before use, yet it remains unclear whether visitors are aware of their participation in the project and understand the short-term and long-term implications for their privacy. Legal restrictions currently prevent the tracking individuals but the technical capabilities for doing so exist and the incentives for project partners to experiment with these capabilities are strong.

In an attempt to co-design smart energy and mobility systems with citizens the Vehicle-to-grid project used workshops with potential users asking their input in how the algorithms should be programmed. However, companies are reluctant to reveal trade secrets and want to control algorithms design to maximise commercial value. When it comes to technical design, societal concerns such as fairness and non-discrimination are rarely considered by commercial partners.

Whatever benefits data and algorithmic systems provide, there are potential costs to public trust and legitimacy of organisations. This has raised the question if there some applications of data and algorithm systems that should not be used at all (Calo, 2017). While the open and participatory approach of living labs has the potential for better consideration of public values, experimentation in the real world comes with its own drawbacks. Either living lab zones of cities become ‘no go areas’ for people who do not consent to participate in the research, or new ethical and design solutions are needed.

A considerable amount of work has already been done to encourage or require good practice in the development and use of data and algorithmic systems. For example, participatory design approaches and value-sensitive design (VSD) methods are routinely used to include citizens directly or indirectly impacted by such systems. Another approach is to improve the transparency of algorithmic systems. The thinking is that by making systems transparent, people understand their workings, and this in turn leads to more accountability.

Meanwhile, data protection laws like the GDPR mandate certain practices around the use of personal data and effectively create a “right to explanation,” whereby a user can ask for an explanation of an algorithmic decision that was made about them (Malgieri & Comandé, 2017). However, this view has been disputed by Wachter et al who argue “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation” (Wachter et al., 2017).

However, current approaches have serious shortcomings: participatory approaches require benevolent organisations willing to listen to citizen concerns and are ill-suited in addressing emerging issues that arise during ongoing operations of system after deployment. Efforts to increase algorithmic transparency have not yet indicated ways for citizens and non-experts to meaningfully understand the complex and unpredictable nature of algorithmic systems and so far, are very limited in predicting the societal impact of algorithmic systems related to issues like discrimination, biases, inclusivity, etc.

In sum, the current situation is characterized by three fundamental challenges:

- Designers and technical experts are severely limited in their ability to predict societal impact of data and algorithmic systems
- AI systems that contradict public values undermine trust and legitimacy of organisations using this technology
- Municipalities are severely limited in their ability to govern data and algorithmic systems
- Citizens lack the ability to meaningfully critique and push back against algorithmic systems and machine predictions.

There is a need for guidelines for governing public experimentation with data and algorithms.
4. Governance Models for Algorithm Experimentation

There is a broad consensus that the use of data and algorithms in our society requires new checks and balances to protect the public from harm (Kitchin, 2016, 2019). In the previous sections I outlined that we need to view algorithms as experimental technology and recognise that many innovation initiatives today use an experimental approach. Thus, the fundamental questions is: **How can we ethically experiment with data and algorithms in society, if data and algorithms themselves pose fundamental ethical issues?**

In this section I will first outline why current attempts to formulate AI ethics guidelines are less than helpful than they appear, and will then summarise two approaches that are useful for informing models for governing algorithm experimentation in society.

### 4.1 The Limitations of Ethics Guidelines

There are a large number of initiatives with the goal to define guidelines, norms and practices for the responsible and ethical use of digital technologies in society. Some focus on cities such as the Barcelona Digital City, Amsterdam Tada!, Detroit Digital Justice Coalition, Seattle Community Technology Advisory Board, Smart London Board, Smart Dubai AI Ethics Advisory Board and the Cities Coalition for Digital Rights. Other initiatives focus on the field of Artificial Intelligence, for example by government initiatives (e.g. the EU High-Level Expert Group on AI), university initiatives (Stanford Institute for Human-Centred Artificial Intelligence), company initiatives (such as Google and Microsoft) and initiatives by civic organizations (e.g. Data for Black Lives and Partnership on AI). Most of these initiatives have published ethical guidelines for AI.

According to Harvard’s “Principled Artificial Intelligence Project”, which analysed 32 AI ethics guidelines (Figure 5), most AI ethics initiatives share a focus on eight key themes (“The Principled Artificial Intelligence Project,” 2019): accountability, fairness and non-discrimination, human control of technology, privacy, professional responsibility, promotion of human values, safety and security, and transparency and explainability. AI guidelines typically include recommendations related to **data** (e.g. publish and minimize bias in the training data), **algorithms** (e.g. making the code of an AI transparent and open for inspection, ensuring the function of the AI can be explained), **outcomes** (e.g. ensuring that an AI's intended and actual outcomes are fair, transparent, legal and aligned with human values) and **governance** (putting in place a process of oversight and evaluation; and holding a person accountable for decisions made).

Another focus of AI ethic guidelines are rights, principles, values and requirements. According to (“The Principled Artificial Intelligence Project,” 2019), frequently mentioned **rights** are: respect for human dignity, freedom of the individual, respect for democracy, justice and the rule of law, equality, non-discrimination and solidarity, including the rights of people in minorities, and citizen rights (based on: Charter of Fundamental Rights of the EU).

**Another focus of AI ethic guidelines are rights, principles, values and requirements**

Frequently mentioned **principles** are: Beneficence (Do good); Non-maleficence (Do no harm); Autonomy (Preserve human agency); Justice (Be fair); and Explicability (Operate transparency) (from: AI4People—An Ethical Framework for a Good AI Society); the last one, Explicability, is relatively new and specific for AI.

Frequently mentioned **values** are: The Guidelines “do not aim to provide yet another list of core values”—since there are many useful lists available, like the lists from Asilomar, Montreal, IEEE and EGE (these lists are reviewed in: AI4People—An Ethical Framework for a Good AI Society).

Frequently mentioned **requirements** are: accountability; data governance; design for all; governance of AI autonomy (human oversight); non-discrimination; respect for (and enhancement of) human autonomy; respect for privacy; robustness; safety; and transparency.
While broadly welcome, these ethics initiatives have four key limitations:

First, AI ethics initiatives have been criticised for being overly broad and too simplistic, offering little practical guidance for dealing with the real-world complexity of AI (Copeland, 2019). Copeland identifies three broad levels of complexity of AI, namely 1) human-experts hand-craft machine learning models 2) AI algorithms decide for themselves what factors are relevant, and 3) a continuous automated process where machine learning models evolve in rapid iterations driven by incoming data streams. Copeland notes that AI ethics guidelines are essentially only useful for AI systems of the lowest complexity.

A second more fundamental critique was voiced by (D’Ignazio & Klein, 2019) who note that current discussions of AI often “ignore fundamental normative questions about what kind of society we want and exclude wider moral philosophy concepts such as such justice, oppression, equity, citizenship and the public good”. Instead, as pointed out by Kitchin, AI ethics guidelines primarily focus attention “on more procedural technical and legal concerns – bias, fairness, transparency, consumer rights, accountability, compliance and redress” (Kitchin, 2019).

He contends that these guidelines locate the sources of ethical problems in individuals and technical systems rather than in structural power and thus merely address symptoms rather than underlying causes.

A third line of critique was raised by Mittelstadt (Mittelstadt, 2019) who contrasts AI to medicine, a “moral community” (Frankel, 1989) with common aims, values, and training, where ethical principles, guidelines and codes are well enshrined. He argues that in contrast to medicine,

“AI development lacks (1) common aims and fiduciary duties, (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms. ... We must therefore hesitate to celebrate consensus around high-level principles that hide deep political and normative disagreement. Shared principles are not enough to guarantee ‘Trustworthy AI’ or ‘Ethical AI’ in the future. Without a fundamental shift in regulation, translating principles into practice will remain a competitive, not cooperative, process.” (Mittelstadt, 2019)

He then outlines four recommendations for the next steps in AI ethics (Mittelstadt, 2019):

1. “Clearly define sustainable pathways to impact”. By this he means creating binding accountability structures as well as clear implementation and review processes at sectoral and organizational level, as well as documentation of models and datasets (I will come back to this point in the final recommendation section).

2. “Support ‘bottom-up’ AI Ethics”. Mittelstadt advocates for a case-based approach to ethics where local practices and lessons emerge bottom-up and spread across the field and industry from which principles and precedents can then be derived (I will come back to this point in the final recommendation section).

---

Figure 5 Map of thirty-two AI ethics guidelines. 
Source: (“The Principled Artificial Intelligence Project”, 2019)
3. "License developers of high-risk AI". Mittelstadt suggests it may be necessary to formally establish AI development as a profession with equivalent standing to other high-risk professions such as doctors and lawyers (my own view here is that based on limited success in establishing software engineering as a profession, I see limited chances for this recommendation).

4. "Shift from professional ethics to business ethics": Here he suggests moving the focus of ethics of individuals to the ethics of business practices and business models.

5. "Pursue ethics as a process, not technological solutionism": Mittelstadt suggests that we should move beyond seeing ethics as a technical design aspect that can be "solved" and argues instead for seeing AI ethics as an ongoing, never-ending process that reflects changes in society.

A fourth and final argument why current AI ethics guidelines have limited utility relates to the earlier discussion that data and algorithms (i.e. the building blocks of AI) are experimental and rapidly evolving technologies. Current AI ethics guidelines assume that AI-based systems can be designed a-priori to avoid possible negative consequences. However, Collingridge control dilemma indicates that this may not be possible for rapidly evolving technologies such as data and algorithms, where societal risk and impacts can only be assessed once technologies have been deployed.

4.2 Possible Directions

If ethics guidelines are not enough, then what can we do to effectively govern experimental algorithms?

Van de Poel's Experimental Ethics

Living Labs aim to circumvent Collingridge's control dilemma by replacing anticipation and prevention with gradual and experimental introduction of a technology into society, and a monitoring of emerging social effects to improve the technology and its use in society. In other words, the experimental introduction of a technology into society is seen as a kind of social experiment which allows society to learn from experience and error (van de Poel, 2016).

Governing a social experiment is different from governing a technology as it shifts the view away from technology towards the context in which technology is used.

Van de Poel (van de Poel, 2016) The Author(s proposed an framework for dealing with issues of real-world technology experimentation rooted in ethical principles formulated in bioethics, namely non-maleficence (e.g. do no harm, minimize risks), beneficence (e.g. create or increase benefits), respect for autonomy (e.g. protecting and guaranteeing autonomous choices of individuals and groups), and justice (e.g. protection of vulnerable groups, avoiding exploitation). These ethical principles go back to the Belmont report, a key work concerning ethics and health care research (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979).

Based on these four bioethical moral principles Van de Poel identifies sixteen specific conditions under which he sees experiments with new technology in society as morally acceptable (van de Poel, 2016):

1. Absence of other reasonable means for gaining knowledge about risks and benefits
2. Monitoring of data and risks while addressing privacy concerns
3. Possibility and willingness to adapt or stop the experiment
4. Containment of risks as far as reasonably possible
5. Consciously scaling up to avoid large-scale harm and to improve learning
6. Flexible set-up of the experiment and avoidance of lock-in of the technology
7. Avoid experiments that undermine resilience
8. Reasonable to expect social benefits from the experiment
9. Clear distribution of responsibilities for setting up, carrying out, monitoring, evaluating, adapting, and stopping of the experiment
10. Experimental subjects are informed
11. The experiment is approved by democratically legitimized bodies
12. Experimental subjects can influence the setting up, carrying out, monitoring, evaluating, adapting, and stopping of the experiment
13. Experimental subjects can withdraw from the experiment
14. Vulnerable experimental subjects are either not subject to the experiment or are additionally protected or particularly profit from the experimental technology (or a combination)
15. A fair distribution of potential hazards and benefits
16. Reversibility of harm or, if impossible, compensation of harm

Some of these conditions mirror concepts that can be found in AI ethics guidelines or are enshrined in current legislation such as the GDPR.
For example, the condition of informing subjects closely mirrors the concept of informed consent in the GDPR. However, van de Poel’s framework goes beyond most current AI ethics guidelines by focusing on governing the process of experimentation, much more so than governing technology itself. While AI ethics guidelines have a static view of technology, Poel’s framework addresses the constantly evolving nature of data and AI algorithms. In this way Poel’s framework aligns with the experimental approach of Living Labs. However, van de Poel’s framework is generic and does not speak specifically to aspects of data and algorithms. The black-box nature of many algorithmic systems and the high-level of knowledge required to understand algorithms make it difficult to satisfy some of Poel’s conditions, especially conditions 10-12. For example, it is unclear how experimental subjects might be able to influence the setting up, carrying out, monitoring, evaluating, and adapting of an experiment that fundamentally rests on the use of a highly complex and highly opaque algorithms.

FAT/ML Principles for Accountable Algorithms and Social Impact Statement for Algorithms
The FAT/ML “Principles for Accountable Algorithms and Social Impact Statement for Algorithms” (Diakopoulos et al., 2018) is a proposal by a community of computer science researchers aligned with the yearly conference on “Fairness, Accountability, and Transparency in Machine Learning” (fatml.org, 2018). The Principles for Accountable Algorithms aim to help developers, designers and organisations in the development of algorithmic systems that are publicly accountable, whereby public accountability is defined as:

“... an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms.” (Diakopoulos et al., 2018)

The FAT/ML proposal defines five Principles for Accountable Algorithms (namely Responsibility, Explainability, Accuracy, Auditability and Fairness) and then proceeds to outline a Social Impact Statement for Algorithms, which is aimed at ensuring compliance with the aforementioned principles.

The authors recommend that algorithm creators develop and publish a Social Impact Statement proposed and revise and reassess it (at least) three times during the design and development process: during the design stage, at pre-launch and post-launch.

The structure of the Social Impact Statement closely follows the five principles and the content of the statement provides information that would allow public stakeholders to assess how the algorithm creators adhere to the principles.

The Social Impact Statement represents a significant step beyond general principles and guidelines

The Social Impact Statement represents a significant step beyond general principles and guidelines, and differentiates this proposal in particular from the AI ethics proposals discussed above. A key strength of this proposal is the requirement to publish the Social Impact Statement, with the aim of facilitating a public discourse on the respective algorithmic system and enabling public stakeholder to contest the workings of algorithms. On the other hand, the FAT/ML statement seems to be exclusively targeted at “algorithm constructors”, that is developers and product designers while ignoring the possible input from other stakeholders such contracting entities, policy makers and public institutions (Rohde, 2018; Selbst, Boyd, Friedler, Venkatasubramanian, & Vertesi, 2019).

Integrated Reporting Framework
Public reporting is a standard way of ensuring external stakeholders can judge the performance and outcomes of public and private organisations, for example related to financial and sustainability reporting. A particularly interesting and relevant approach is Integrated Reporting (IR) (Banerjee, 2019; Eccles & Krzus, 2015), an approach that aims to encompass value creation of an organisation in a holistic manner, beyond simply financial value and traditional sustainability measures. The IR framework has the goal to improve governance, accountability and trust in organisations through periodic reporting about “integrated” value creation over time and an integrated report is a concise communication about how an organization’s strategy, governance, performance and prospects lead to the creation of value over the short, medium and long term (IIRC, 2016).
The IR framework has been adopted by many public sector organisations and has become a key instrument for sustainability reporting of cities across the world (Niemann & Hoppe, 2018; Oprisor, Tiron-Tudor, & Nistor, 2016; Sulkowski, 2017; Tirado-Valencia, Rodero-Cosano, Ruiz-Lozano, & Rios-Berjillos, 2016).

Even though Integrated Reporting is primarily concerned with financial and sustainability reporting, it provides four lessons that are relevant for the further development of the FAT/ML proposal. First, reporting must be seen as an act of communication between different stakeholder groups. A report is useful only if it helps stakeholders establish an ongoing and fruitful dialogue. Thus, creators of a report must understand their audience and use a language that their audience can understand. A report that only “algorithm creators” can understand is of no value. Second, reporting must affect future behaviour and help organisations (or algorithm creators) improve their practices and processes. A report should not simply be seen as justification of past behaviour. In other words, “algorithm creators” should integrate reporting as a form of self-reflection. Third, Integrated Reporting emphasises value creation and asks what kind of value is being created for whom. However, the value dimension is currently absent from the FAT/ML proposal. While the FAT/ML proposal mentions topics that could be interpreted as “value destruction” (biases, unfairness etc.), it does not include any information of the positive value created by algorithms. Fourth, the emergence of Integrated Reporting highlights how a long-term community-led process can lead to a global movement with significant impact (Banerjee, 2019), achieving large-scale adoption in the private and public sector (Niemann & Hoppe, 2018; Oprisor, Tiron-Tudor, & Nistor, 2016; Sulkowski, 2017; Tirado-Valencia, Rodero-Cosano, Ruiz-Lozano, & Rios-Berjillos, 2016).

5. Recommendations

With respect to data and algorithms the key question is:

*How can we effectively govern the design, development and use of data and algorithms in society to maximise public benefits and promote public values?*

The discussion presented in this essay highlights the broad ethical concerns about the use of data and algorithms and leads to the following key insights:

1. Decision making processes with far-reaching societal implications are increasingly enshrined in data-driven algorithms. In other words, algorithms activate data. Governance and policy thus need to treat data and algorithms as unity.
2. The sophistication and complexity of algorithms is rising rapidly to a point where even computer experts have problems understanding how algorithms operate and how they make decisions.
3. Data and algorithms are highly malleable and subject to constant change and evolution. The speed of change is expected to increase as organisations develop and apply agile data science engineering practices (DevOps) and increasingly automate machine learning processes.
4. Algorithms must be understood as experimental technology - it is impossible to predict societal implications of algorithms. This creates the need for ethical governance of the experimental process around data and algorithms.
5. Living labs are an effective method of bringing together various stakeholders in an innovative learning process to identify how societal and commercial value can be extracted from data and algorithms.
6. We are lacking guidelines and regulations for how to govern this experimental process - AI ethics guidelines provide valuable hints of where to pay attention but are not yet concrete enough.

5.1 Principles

Answers to the key question above requires a broad response involving ethics, policy and regulations. A reliance on regulation alone is not enough because even strict adherence to regulations can result in harmful and unethical consequences.
I propose that future policy efforts should be based on five principles:

1. **Focus on data and algorithms, not just data.** In other words, future policy efforts should be directed at governing how data is used in (automated)-decision making processes and how decision-making (in the public and private sector) is encoded in increasingly sophisticated and complex algorithms.

2. **Encourage a safe and ethical experimental learning process around data and algorithms.** The experimental learning process implemented, for example, by living labs is fundamental for economic and societal development. Thus, policy should see experimentation as an opportunity, not a threat.

3. **Focus on public contestability.** Public contestability focuses on the question of how civil society actors can critique algorithms and appeal their decisions. This requires moving beyond transparency towards mechanisms and institutions for collating, analysing and acting upon public feedback and criticism, as well as empowering stakeholders to be able to contest ongoing developments.

4. **Focus on fostering a public debate** on how to use data and algorithms in the public realm - a debate in public, with the public.

5. **Support ongoing efforts by Dutch municipalities** while avoiding replication of efforts. Dutch cities are world-leading in developing ethical practices and policies related to data and algorithms. Future policies should learn from these experiences and concentrate first on collecting evidence and identifying best practices, before ultimately formulating rules and setting standards.

As a result of the analysis presented in this chapter, I propose policy recommendations in three areas: governance, institution and skills.

### 5.2 Governance

The potential impact of data and algorithms are far-reaching and unpredictable. Government and municipalities must thus ground their decisions about the use of data and algorithms in democratic deliberation that allows the public to have a voice in shaping these developments. Across the Netherlands public and private stakeholders are collaborating in living labs on the development of data and algorithms systems, often with the close involvement of affected citizens and communities. However, the practices and standards of these efforts vary widely and there is a lack of transparency which practices and standards are applied across The Netherlands.

The national government can play a key role in bringing together these disparate local efforts to create a shared understanding of best practices and policies, and ultimately develop a common set of principles for the whole of The Netherlands.

The aim of the government should be threefold:

1. Create transparency about who develops what kind of algorithms for what purpose
2. Enable continuous assessment of algorithms and their impact
3. Enable civil society stakeholder to contest the development and use of algorithms through public debate
Recommendation 1: Set up an Algorithm Reporting Initiative with the aim of documenting and tracking the societal value and risks of algorithm projects across the Netherlands.

At this point in time we have very little understanding what kind of data-driven algorithms are being developed and used across the Netherlands, and what their benefits and risks are. The key problem is that there is not enough public information about such projects and that there is no standard way of documenting algorithms. In the previous section I discussed proposals for algorithmic accountability reporting and social impact reporting of algorithms, but these proposals remain untested (Diakopoulos, 2014, 2015, 2019; Diakopoulos & Friedler, 2016; Diakopoulos et al., 2018). On the other hand, I discussed how Integrated Reporting practices have emerged from a community-led process (Banerjee, 2019), and have now been adopted by public and private organisations worldwide (Niemann & Hoppe, 2018; Oprisor, Tiron-Tudor, & Nistor, 2016; Sulkowski, 2017; Tirado-Valencia, Rodero-Cosano, Ruiz-Lozano, & Rios-Berjillos, 2016).

Thus, my first recommendation is to assemble a group of interested parties (researchers, developers, policy makers, citizens etc.) to start an initiative to develop an Algorithm Reporting Framework. This is most likely a multi-year effort that involves surveying existing algorithm initiatives, developing a 360 degree understanding of such initiatives, and developing and testing reporting practices.

The most profound impacts of algorithms are not yet well-understood. Similar to the Integrated Reporting Framework, the key value of Algorithm Reporting will be the change in how individuals and organisations think about these issues, i.e. the thinking that went into the published report, not the report itself. Mirroring Mittelstadt’s recommendations to “Support ‘bottom-up’ AI Ethics” (Mittelstadt, 2019), I would like to highlight the need to build on existing initiatives that are taking place in living labs and municipalities around the Netherlands, and to gradually develop an Algorithm Reporting framework through a participatory bottom-up approach.

It is certainly not too early to speculate what might go into an Algorithm Reporting Framework. Combining ideas from social impact reporting, Integrated Reporting and van de Poel’s experimental ethics, I suggest that an algorithm report might comprise four broad categories of information, namely: Value, Mechanisms, Experimentation, and Public:

**Value**
- **Value Creation and Benefit.** What value is created by the algorithms? Who will benefit from the use of this algorithm and how? Are these benefits public (for society at large) or private (contributing to commercial interests).
- **Value Destruction and Harm.** What value is destroyed, especially related to fairness aspects, such as potential biases, discrimination, and exclusion. What potential impact and harm might use of algorithms have on individuals and specific groups?
- **Human Responsibility and involvement.** Who is involved, who has direct control over the algorithm, who has oversight and is accountable? This information should identify individuals as well as companies and organisations involved in the creation, deployment and use of the algorithm, as well state their goal, purpose, and intent.

**Mechanism**
- **Data.** What data drives the algorithm? What is known about the quality of the data, including its accuracy, completeness, and uncertainty, as well as its timeliness? How was the data collected, transformed, vetted, and edited? How is data managed and stored, and under whose control?
- **Model.** What features or variables are used in the algorithm? What training data
was used and how can it be characterised in terms of accuracy, completeness etc.? What does the overall process for creation the model look like?

- **Inferencing.** What inferences does the algorithm make, such as classifications or predictions? What are standard measures like accuracy etc? What kinds of steps are taken to remediate known errors?

These first two categories are primarily informed by the value focus of the Integrated Reporting Framework (International Integrated Reporting Council, 2013) and the FAT/ML Principles for Accountable Algorithms (Diakopoulos et al., 2018). Following (van de Poel, 2016) the Author(s I further suggest the reporting should include information about the experimental setup in which algorithms are developed:

- **Experimentation**
  - **Participation:** How are individuals whose data is used or who are impacted by algorithms informed and included in the setting up of the experiment? How do individuals indicate consent and how are they able to withdraw from the experiment (including their data)? How can experimental subjects influence the setting up, carrying out, monitoring, evaluating, adapting, and stopping of the experiment?
  - **Protection:** How are vulnerable subjects protected? What steps are taken to identify and remedy these risks? How can potential harm be reversed?
  - **Oversight:** Who approves and governs the experimental development and use of algorithms? Are these democratically legitimized bodies, private parties or others? How are impacts of the algorithm assessed and made available publicly?

Finally, I suggest a category related to the need to enable effective public debate and critique, partly informed by the FAT/ML Principles for Accountable Algorithms:

- **Public**
  - **Auditability & Contestation:** How can societal stakeholders scrutinize, understand and verify the behaviour of the algorithm and provide feedback, point out potential negative impacts and provide suggestions for improvement? Who is receiving and assessing this input? What feedback and suggestions have been received and how has it been taken up? What changes were made of the result of it?

Together, these report categories create a potentially heavy burden on the parties responsible for the development and use of algorithms. It is thus essential to define best practices, create flexible standards and embed them in a practical reporting framework that actually achieves what it is intended to do, namely improving trust in algorithms and ensuring algorithms are in line with public values.

The goal of public reporting is to improve governance, accountability, and trust in organisations (the unit, team, group which is responsible for algorithms). This can only be achieved if reporting adds value to the organisation, not just to external stakeholders. The Integrated Reporting framework stresses the importance of creating an internal culture that uses the reporting framework to drive discussions about how an organisation creates (or destroys) value. The same must apply to an algorithmic reporting framework. In addition, reporting needs to be an iterative, regular process. The value of public reporting lies not primarily in the published report, but in the tracking of progress over time and associated lessons in improving control, contestability etc.
Recommendation 2: Develop a National Algorithm Register with the aim of enabling effective comparison and assessment of data and algorithm initiatives across the Netherlands.

The second recommendation is to establish a central register of algorithms to make information about algorithms available to as many interested parties as possible. While Recommendation 1 focuses on how to document algorithms, Recommendation 2 focuses on compiling information from all relevant algorithm initiatives across the Netherlands.

Initially this register can be compiled from information provided from government institutions and municipalities and voluntary contributions by others. Over time – once a reporting framework has been developed - government might mandate that each algorithms initiative be included in this register.

The information in the Algorithm Register should be public – this means it primarily targets public and public-interest algorithms.

The National Algorithm Register can become an important resource for researchers, companies and policy makers alike. Researcher could get information to better understand risks of algorithms. Companies could understand how to assess the impacts of their algorithms. Policy makers could develop more effective ways of governing algorithms and track the impact on policies over time.

The National Algorithm Register can become an important resource for researchers, companies and policy makers alike.

In order to serve this purpose, the National Algorithm Register needs to have a public web presence, and be regularly updated to ensure information is timely, relevant and useful.

Recommendation 3: Develop an Algorithm Forum with the aim of enabling public stakeholders to debate, critique and contest the operation, use and outcomes of data and algorithm initiatives.

Critical technologies like algorithms that can have wide-ranging impacts on society need to be debated as widely as possible. At the moment this is not possible because the public is not aware of which algorithms are used in society, how to voice their opinion and to whom to address their concerns. Thus, I recommend to create a public forum to enable affected stakeholders to debate, critique and contest the operation, use and outcomes of data and algorithm initiatives.

The Algorithm Forum should have a public website but should use a range of methods to engage the public including public debates, educational workshops and participatory design sessions. Public feedback collected during these events could be included into the web presence of the Algorithm Forum.

One component of the Algorithm Forum should be a Public Reporting Hotline which allows the public to report specific concerns they have about decisions made by an algorithmic system or negative experience with such a system. This suggestion is motivated by public hotlines in the aviation and maritime industries such as the FAA Hotline Reporting Form (https://hotline.faa.gov), the UK Confidential Reporting Programme for Aviation and Maritime (https://www.chirp.co.uk/) which capture confidential reports, analyses the resulting data, and disseminates vital information to relevant communities and government agencies.

5.3 Institution

Recommendation 4: Establish a Data and Algorithm Institute to drive the development of national approaches, tools, infrastructures and standards for ethical and accountable use of algorithms.

The mission of the Data and Algorithm Institute should be to maximise the social and economic benefits of data and algorithms for Dutch society and help government, start-ups, cities, citizen groups etc. to develop ethical and accountable data and algorithm practices.
Specifically, such an organisation should focus on:

- Promoting open innovation in technology, ethics, governance and design
- Providing training and education
- Conducting public outreach
- Conducting research

A key role of this institute would be to develop and drive the Algorithm Reporting Framework, to develop and maintain the National Algorithm Register and Algorithm Forum and in general to act as a clearing house for information on public-sector and public-interest algorithms and data. This entails,

1. publishing information on current data and algorithms initiatives by government, municipalities and others,
2. designing effective mechanisms for citizens and organisations to critique and contest ongoing data and algorithm initiatives (for example through crowd-sourcing initiatives),
3. designing effective mechanisms for analysing and acting upon such feedback and
4. publicly report on a regular basis how feedback and criticism has been taking into account.

In order to drive innovation, the proposed Data and Algorithm Institute would need strong in-house legal, social science and design capabilities, as well as technical capabilities in data, information architectures and business models. It would need to have the capability to analyse the end-to-end process of data and algorithm systems and the ethical and legal implications of such systems.

The Data and Algorithm Institute should collaborate with and support Dutch municipalities to ensure more effective collection and analysis of ongoing activities with the aim to strengthening the emerging Dutch data and algorithm ecosystem. Such an organisation could be modelled on Geoff Mulgan’s proposal for a Machine Intelligence Commission (Mulgan, 2016) to help the development of new approaches for protecting the public interest.

The current efforts by the Ministry of the Interior and Kingdom Relations in establishing a Transparency Lab (Braak, 2019) are a step in the right direction. However, such a lab needs to be conceived as an independent institution outside of government hierarchies with strong involvement of broader sectors of society. Following examples such as Open Data Institute, NESTA, and the Open Knowledge Foundation, such an institution should have a clear public mission, led by a board of directors whose mix reflects to societal stakes in algorithms, and be provided with the means to build up significant in-house expertise. Most importantly, the Data and Algorithm Institute should not be focused on auditing algorithms, which is primarily a method for risk assessment and risk minimisation, but on reporting and contestability approaches as indicated above, and on driving innovation in technology, ethics and governance.

There is a need for institutions with a broader remit on innovation, technology, design and policy, and with a stronger ability to shape and inform technical and policy development

As a general observation, compared to the UK, The Netherlands are currently missing strong public-sector organisations like the Open Data Institute, NESTA, and the Open Knowledge Foundation which in the UK drive innovation and public outreach in data and algorithms. The Rathenau Institute conducts excellent work in research and public outreach on socially relevant aspects of science and technology and Waag in Amsterdam is vital in scaling up grassroots initiatives and bringing societal concerns to the foreground. However, there is a need for institutions with a broader remit on innovation, technology, design and policy, and with a stronger ability to shape and inform technical and policy development. New capabilities are especially required on the interaction of the technology, creative and policy sectors, and in applying design thinking and innovation methods to policy development.
5.4 Skills

Recommendation 5: Develop a National Data and Algorithm Skills Agenda to ensure that all stakeholders (from citizens to organisations) are sufficiently skilled to participate in the new data and algorithm economy.

Regulators and policy makers need new skills to develop a more in-depth understanding of new technologies and business models; computer experts need new skills to develop an appreciation of ethical and policy implications of data and algorithms; product and service designers need new skills to effectively inform the design of responsible data-driven products and services; government employees require new skills to make appropriate decisions about when and how to use data and algorithms in public services and automated decision-making processes; citizens need new skills to meaningfully inform and contest the use of data and algorithms in the public realm. Developing these skills over the next twenty-five years requires a strategic approach across the entire education spectrum, from basic school education to professional and scientific education.

I thus recommend working towards an overarching National Data and Algorithm Skills Agenda with three objectives:

1. Identify the need for data and algorithm skills across all sectors of society to ensure the future economic and social well-being of Dutch society
2. Identify which data and algorithm skills should be taught across different education routes, from basic school education to professional and scientific education.
3. Initiate new education initiatives focused on critical areas of need

This recommendation is driven by a realisation that current efforts in data and algorithm education are not enough to tackle the data and algorithm skills shortage across society. The Dutch government is increasing funding for sciences education at Dutch Universities, with a partial outlook towards strengthening scientific education on data, algorithms and AI. In addition, a range of new professional Master programs are emerging that teach data skills for experts from various domains. In addition, MOOCs have been created which aim to improve the public understating of data, algorithms and AI, for example the Nationale AI-Cursus. However, more must be done. In the UK, the Open Data Institute has been offering a data literacy education program for several years which combines online, blended and face-to-face courses and which so far has helped over 25000 professionals and laypeople to developed data skills. The programme is based on a Data Skills Framework (ODI, 2018) which defines a comprehensive data literacy program and provides needs-based education pathways, including not only technical data skills but also data ethics and data governance.

Dutch society cannot afford that only select few have an understanding of data and algorithms. This implies that the strategic approach should not just focus on professional and scientific education but also on basic school education. Over the last years, there has been a recognition that computer education in schools should go beyond basic IT skills and focus on a more comprehensive digital literacy education. Several recent initiatives in the UK have shown what a modern digital literacy education in schools may look like, including the BBC micro:bit program (Schmidt, 2016; Sentence, Waite, Hodges, Macleod, & Yeomans, 2017), the Internet of Schools program (Moreira, Magalhães, Ramos, & Vairinhos, 2018; Moreira, Vairinhos, & Ramos, 2018) (an effort to teach Internet of Things skills at schools) and the Urban Data School (Wolff, Kortuem, & Cavero, 2015) (an initiative - which I helped set up - aimed at teaching data literacy to 8-12 year old students). The Urban Data School in particular demonstrated that data literacy can be meaningfully integrated in science education in schools and that it is feasible to teach basic data literacy even to young students, covering seemingly complicated topics such as data acquisition, data preparation, data analysis, data communication and data ethics (Wolff, Goch, Montaner, Rashid, & Kortuem, 2016).

The main aim of the suggested National Data and Algorithm Skills Agenda is to develop a long-term vision on the data and algorithm skills required by society and to develop innovative approaches for developing them.
About the author

Prof. dr. Gerd Kortuem is Professor of Internet of Things at Delft University of Technology.
References

In the remainder of this articles, I will often use the term ‘algorithm’ as shorthand for ‘data and algorithm’.


Appropriate use of data in public space | 89


• Kitchin, R. (2016). The ethics of smart cities and urban science. https://doi.org/10.1017/S11216040.16.0115


• Mulgan, G. (2016). A machine intelligence commission for the UK: how to grow informed public trust and maximise the positive impact of smart machines.


Innovation Management Review, 7(7), 21–33. https://doi.org/10.22215/timreview/1088


The need for a digital environmental strategy

From principles to practice

Dr. Jiska Engelbert
Senior researcher
LDE Centre for BOLD Cities

Prof. dr. Liesbet van Zoonen
Academic director
LDE Centre for BOLD Cities

Foreword

By Aantink Yeh

The introduction of new technologies and the collection and usage of data in public space often raises challenging technical, social, legal and/or ethical issues. Municipalities (among other bodies) are not only responsible for ensuring an open, safe and secure physical environment, they must also safeguard the accessibility, safety, security and appropriate use of digital public space.

As an Innovation Advisor at VNG Realisatie (part of the Association of Netherlands Municipalities), I am coordinating the development of a set of common principles for Dutch municipalities to guide them in direction and tools to fulfill this responsibility.

At the same time, we are seeing a wide range of similar initiatives both in the Netherlands and abroad to implement public values into the ever-increasing process of digitization. The Centre for BOLD Cities, a research group set up by Leiden University, Delft University of Technology and Erasmus University Rotterdam, specializes in digitization and data use in cities with a particular focus on the aspect of ethics and the perspective of the local population. For this reason, I have asked Dr. Jiska Engelbert and Prof. Liesbet van Zoonen to conduct research about these and other initiatives to embed public standards and values as well as the effect that they will have on government bodies in the Netherlands.

The meetings that we have had with the supervisory committee in order to establish this collection of essays have shown how important it is for scientists and policymakers to maintain continual dialogue, although theories and meetings alone are not enough. This essay (as well as the others in the collection) will help us to further develop our principles and follow-up action by giving greater insight into the how and the why of other initiatives and required instruments and measures. Having principles is all well and good, but they only have value if we follow through on them.

Aantink Yeh
Smart City Consultant at VNG Realisatie
Member of the Supervisory Committee
This essay analyses the panoramic landscape of principles for the digital society that has emerged over the past five years. We find, firstly, a difference between principles that cover ethics, data quality and standardization; secondly, we argue that most of them are designed and formulated rather inconsequentially in the form of human rights to which nobody could object. A third characteristic of the landscape is that practically none of the principles have resulted in verifiable and enforceable measures. Nevertheless, they do represent a departure from exclusively economically oriented discourse concerning standardization that we have also observed and which has no place for citizens or other civic stakeholders in society. Finally, the set of principles mainly concerns general processes of digitization and rarely contains regionally or locally formulated goals.

We use two examples (citizens’ measurements of noise pollution around Schiphol and political protest against smart lampposts in Utrechtse Heuvelrug), to show that the apparently harmonious context within which the ethical and quality principles and the ambition for standardization circulate, conceals important social oppositions (e.g. between the business sector, government bodies and citizens) and does not provide the responsible administrative bodies with clear tools to choose between opposing interests or bring them together. This is partly due to the fact that citizens and other parties often only become aware of digital solutions after they have been implemented, at which stage they have no other options than to accept or oppose. The principles, however, suggest strongly that digital solutions need to be developed, from the start, in collaboration with all societal stakeholders. Therefore we propose, in the final part of this essay, to ask local governments to complement their Environmental Strategy with a Digital Environmental Strategy.
1. Introduction

Municipalities, provincial government, ministries, regional water authorities and a broad range of other organizations are urgently exploring the principles for using digital and data technologies in society and public space. Increasingly, these discussions are focusing on automated data-processing and the analytics enabled by artificial intelligence. The municipalities of Amsterdam and Eindhoven have jointly formulated values such as inclusion, openness, transparency, privacy and data ownership, which have echoed, among others, in the concise TADA principles. A closer look reveals numerous other attempts to establish ethical frameworks and societal participation for data use and artificial intelligence. Utrecht University and the Municipality of Utrecht, for example, together launched an ‘ethical data assistant’; the Rathenau Institute published suggestions to reappraise public values in the digital society; the Dutch green party GroenLinks is working on a Smart City Charter; and the Association of Universities in the Netherlands (VSNU) has proposed a four-part set of data-science and societal standards. The digital society and responsible data use are also high on international governmental agendas. The cities that position themselves as ‘smart’ are especially leading the development of tool kits for responsible data use. They collaborate in alliances of cities and regions, such as the Cities Coalition for Digital Rights or the Sharing Cities Alliance. The British government works according to the ‘GEMINI principles’, which state that data use must have clear purpose, must be trustworthy and must function effectively; the government of Dubai’s key principle is that data and digital technologies should make its citizens happy, while the city of Barcelona’s main priority is to ensure digitization does not disadvantage its citizens.

There is clearly something for everyone in the abundance of data principles that are currently circulating at the national and international level. What does that mean for Dutch governmental bodies seeking guidelines? Do they need to invent their own wheel or are the existing principles sufficiently instructive? How can the existing principles be adapted to specific provincial, regional or urban policy and culture? And who will be responsible for implementing and enforcing these principles? These are the questions that we will address in this essay.

2. A broad landscape of principles

To get some grip on all circulating principles, we must first recognize that they cover different dimensions: a first set relates to ethics and social significance, the second group mainly concerns reliability and quality of the data systems and the third cluster concentrates on the issue of standardization. These differences are all related to the groups and organizations that propagated them, as we shall see below.

2.1 Ethical principles

Principles addressing the ethics and social significance of data use are the result of frustration about citizens having lost control of their own data and having no insight into how governments, businesses and international platforms use it. The TADA principles, for example, were developed because data could help cities become cleaner, safer, healthier and more pleasant, but ‘only as long as people maintain control over data, and not the other way around.’ The Cities Coalition for Digital Rights, a joint initiative of Amsterdam, Barcelona and New York that has already attracted some 30 affiliate cities, strives to ensure ‘policies, tools and resources to promote and protect resident and visitor rights online.’ Likewise, the Smart City Charter by Dutch green party GroenLinks states that ‘citizens and politicians must regain control of technological development.’
The Sharing Cities Alliance also emphasizes the important role of politics and government – who just like ordinary citizens, have lost control – and therefore seeks to empower municipalities, particularly against the multinational platforms.\(^{18}\)

Their collective assessment is that citizens, their representatives and their governments have become digitally vulnerable and, therefore, a set of carefully formulated principles is necessary in order to mitigate and reverse this vulnerability. The initial proponents of this movement were progressive municipalities, in particular Amsterdam and Barcelona. Their leadership is not by chance, as they have the necessary electoral mandate for such progressive direction and an administration that is capable of implementing these principles. For example, Barcelona has formulated how participation, policymaking, procurement and licensing will be executed for each policy area in accordance with the city’s digital principles.\(^{19}\)

However, the formulated principles do reflect a much more general collection of public values than the original progressive policy would suggest. For example, the principles in GroenLinks’s Smart City Charter (democracy, solidarity, human dignity, privacy, sustainability and equality) can be directly linked to the widely supported Sustainable Development Goals established by the United Nations, while the TADA principles also express a general human rights discourse rather than a specific progressive ideology. As early as 2017, the Dutch Rathenau Institute proposed ‘digital human rights’.\(^{20}\) The relative neutrality of the thus developing ethics enables a wide range of actors to commit to them, from government bodies and civic groups to social enterprises and local or national business.

This neutrality also allows a diverse range of decisions to be made at the local level: the Municipality of Barcelona makes it impossible for Uber to operate in the city\(^{21}\) while, in contrast, the Municipality of Amsterdam uses Google to measure air quality in the city\(^{22}\). This mobilizing force is also enhanced by the format of the manifesto or charter in which these ethics are established: signing up to these manifesto is open and free for individual citizens and organizations alike. One joins a community of like-minded people who collectively formulate their ethical intentions and see them as shared responsibility, in a manner that transcends direct self-interest.
The mobilizing force of these relatively neutral digital ethics – which widely begin to take shape – may rest exactly in their somewhat inconsequential nature: it is undefined who can address those that violate the principles. Can all possible parties subscribe to a certain manifesto or charter? And if not, who can decide this? What happens when subscribing parties disagree? In areas where a particular government body has formulated or adopted digital principles itself, such as the province of Zuid-Holland23, it seems obvious that the government body in question would be responsible for maintaining and enforcing digital principles, but their concretization and operationalisation does not keep pace with the discussions and is as yet absent. We will come back to this later.

2.2 Quality principles
A slightly different collection of principles that is being intensively (albeit less audibly) discussed, concerns the quality and reliability of data and data-related systems. After all, ethical use and application of data – whether ‘big’ or small – is not possible if data has not been correctly collected, indexed, stored, cleaned, analysed, presented and applied in the first place. This concerns the basics of data technology, in fact, and its correct and reliable application. Statistics Netherlands and quantitative social and behavioural scientists have traditionally played a key role in developing and maintaining such quality principles, aspects that have been formulated in terms of reliability and validity. However, the explosion of data in recent decades has brought new players in, especially data scientists.

The Research Agenda for the Digital Society24, published by VSNU in 2018, adopts their quality principles: FAIR data, ROBUST systems and FACT algorithms, all of which cover specific elements of data science. In contrast to the more general ethical principles, these focus much more on verifiable criteria that could even result in the establishment of a quality score for data practices conducted by the diverse range of government bodies, organizations and businesses.

The principle of FAIR data states that data should be Findable, Accessible, Interoperable and Re-usable. These four criteria have been operationalized into fifteen specific directives for the allocation of metadata and protocols, among other issues.25 FACT is an acronym that reflects the efforts of data scientists to develop algorithms that function in a manner that is Fair, Accurate, Confidential and Transparent.26 Finally, ROBUST systems are designed to be Resilient, Open, Beneficial, User-Oriented, Secure and Trustworthy. More and more scientists embrace and apply these criteria, although government bodies and the business sector have been far slower to adopt these principles into their practices. In these areas, it is difficult to say whether people have even heard of the criteria, whether the criteria function as guidelines or whether current data practices are or could be made compliant with them.

One of the authors of this essay recently published a detailed analysis of three types of data projects that are popular within the municipal social domain: data warehouses, dashboards and predictive analytics.27 She found that in this context the validity and reliability of the data itself cannot be guaranteed; that, as a result, any predictive model will be neither fair nor accurate; and that the people to whom the data relate are rarely informed of or involved. Her conclusion was that not only are municipalities probably breaking the new General Data Protection Regulation, but that the lack of data quality and the analytical models tend to lead to mistakes and stigmatization.28
In the debate about responsible data use the quality principles for data science feature much less prominently than the ethical principles, although the two come together in the fear of autonomous algorithms whose actions and decisions nobody understands. According to the ethical principles, citizens, governments and politics must regain control of these algorithm; moreover, they shouldn’t even be allowed to operate autonomously and they also should be maximally transparent. However, it is clear that for municipalities and other government bodies, it is much simpler to embrace the general ethical principles of responsible data use than to ensure that data systems are robust, algorithms are fair and accurate, and data are findable and exchangeable. These three quality principles require a series of organizational measures and financial incentives that are out of reach for most municipalities in the Netherlands (and other countries).

2.3 Standardization
A third set of principles on which a diverse range of actors is working relate to the standardization of the data infrastructures. These mainly cover two components of the FAIR principles: interoperability and re-usability. The exchange of data between government bodies, their various service departments and possible societal actors is currently severely hindered by huge variation in data collection, storage and modes of access. There have been countless attempts at standardization. Already in 2016, the European research project Espresso identified 88 organizations from 23 different countries working on standardization of and within ‘smart cities’, concluding, nevertheless, that it had been impossible to gain a clear overview of everything.

In the Netherlands, the NEN (Netherlands Standardization Institute) recently began consultations with local authorities, knowledge institutes and businesses about standards for smart cities, with the goal of ensuring that ‘all kinds of parties will be able to develop applications enriched by data from other sources in order to provide higher-quality and more sustainable services’. Initially, attention will be paid to the standardization of open ‘urban data platforms’, as according to the NEN, these offer significant opportunities for residents and businesses in individual cities to collectively devise new solutions to specific urban problems. The NEN thus directly links the need for standardization of smart infrastructure to the ethics of openness and collectivity that speaks from the various manifestos and charters discussed earlier.

This link between ethics and standardization is also reflected in the results of the standardization project Espresso. Among other issues, the research group is considering whether important ethical issues – particularly privacy issues – can be resolved via standardization.

The exchange of data between government bodies, their various service departments and possible societal actors is currently severely hindered by huge variation in data collection, storage and modes of access.

The combined attention for standards and ethics is particularly striking if we compare the arguments of NEN and Espresso with the manner in which the International Organization for Standardization (ISO) introduces standardization in smart cities on its website. Here, it states that standards are the ‘holy grail of an interoperable, plug-and-play world where cities can mix and match solutions from different vendors without fear of lock-in or obsolescence or dead-end initiatives.’ The more detailed explanation addresses how ‘we’ can optimally capitalize on the opportunities and improvements to public services that data offer. The same market-orientation is apparent in the Smart Cities standards of the British Standards Institute: first and foremost, these state that standardization offers the best solutions for the commercial and technical interests of smart city businesses and are necessary to accelerate innovation in the market.

Without further analysis, we cannot claim yet that mutually distinct philosophies of standardization exist, i.e. Dutch-European vs British-international. Nevertheless, the contrast between the NEN and Espresso on the one hand, and the BSI and ISO on the other, matches a more general discussion about the desirability of a European public model of the digital society as opposed to the corporate model led by the multinational superpowers Apple, Alfabet, Amazon, Google and Microsoft and the totalitarian state model taking shape in China.
The various arguments also show that rather than being a neutral technical coordination exercise, standardization always but often implicitly carries vital ethical choices.

2.4 Preliminary conclusions

Our exploration of the landscape of principles leads to a number of provisional and partial conclusions. Firstly, the principles to which all kinds of government bodies and social actors currently subscribe seem to be designed and formulated in a non-committal manner. They rarely rise above the level of general human rights that nobody could disagree with and seldom translate into verifiable and enforceable measures. Their importance only becomes apparent when they do not feature as principles, as we see, for instance, in the ISO and BSI standardization strategies. Secondly, it is striking that this set of principles primarily relates to the process of digitization. This must be ‘good’ in two respects: ethically responsible because everyone can participate and the data science must be of solid and reliable quality. To the extent that end goals are included (why are we doing all this?), they entail abstract values like liveability and efficiency. What do these relatively neutral process ethics mean for government bodies who wish to design their data practices according to ethical and quality principles?

3. Challenges for governments

The question about the meaning of the data principles for the practices of Dutch government bodies can best be explored by two concrete examples that demonstrate where the problems occur.

To start with, we will examine the do-it-yourself measurement movement, in which citizens independently measure everyday problems and issues in their local area using all sorts of sensors and mobile phone apps. The RIVM (National Institute for Public Health and the Environment) is an important player in this movement and has included a high volume of measurements taken by citizens working with cheaper sensors. Residents living near Schiphol Airport use the app Explane to measure the decibels produced by planes flying over, because the official measurements do not match their own experiences. The same is happening with the pollution caused by Tata Steel and levels of particulate matter in Rotterdam areas. Such do-it-yourself measurement by citizens ties in perfectly with the ethical principles that we discussed earlier: giving people control of data and developing data together. The process of participation is thus ensured, but we encounter a problem when we add the principles of data quality.

The systems with which these data are being collected differ greatly from each other, the citizen measurements are not officially validated and it is uncertain whether their measurements are replicable. Standardization of such measurements is also fundamentally impossible, as this would mean citizens are only allowed to take measurements if they do it according to the standards of others. Citizen participation could well lead us to a scenario in which different data sets (those taken by official institutions and those taken by citizens) contradict each other, for example, during discussion of the expansion of Schiphol, the pollution caused by Tata Steel or traffic flows in Rotterdam. What decisions must the local authority in question make in such situations? We already get a taste of such disputes via the comments of the institution officially responsible for noise measurement at Schiphol: it says that the citizens’ app Explane does not differentiate between aircraft noise and background noise, the measurements made by the various devices cannot be compared and there is a substantial difference between measurements made in rural areas and urban areas.

Appropriate use of data in public space | 98
A second example comes from the Municipality of Utrechtse Heuvelrug in the middle of the Netherlands. In 2018, the street lighting had to be replaced and the municipality wanted to experiment with lampposts to which all kinds of smart applications could be connected, such as cameras, 5G technology, sensors, chargers and lighting scenarios. The municipality had also devised a business case which enabled innovative entrepreneurs to purchase their own smart spot on the lamppost. At the time, other experiments with such smart lampposts were being conducted in places such as Hengelo, Eindhoven and Texel. In 2017, a policy document for smart lampposts had been approved by the Amsterdam municipal council without issue. But in Utrechtse Heuvelrug, a group of citizens voiced serious concerns about the lampposts, particularly in relation to the possible radiation that 5G masts on the lampposts would cause. They organized consultations with councillors, a social media campaign and a series of community meetings. The resulting social unrest prompted the municipal council to completely withdraw the idea of smart lampposts. It is a wonderful example of citizen participation, be it with an outcome that directly contradicts the assumptions of ethical principles. Those suggest, by and large, that when everybody has their say and participates, a collectively shared set of ‘good’ digital/data solutions to urban problems will pleasantly emerge. However, they do not take into account the fact that participating groups of citizens could be radically opposed to such solutions, as is the case in Utrechtse Heuvelrug.

The two examples show that the apparently harmonious context within which the ethical and quality principles and the ambition for standardization circulate, conceals important social oppositions and conflicts of interest (e.g. between the business sector, government bodies and citizens) and does not provide the responsible administrative bodies with clear tools to bring opposing interests together. The question whose data will be considered most important when deciding about the expansion of Schiphol Airport or the lampposts in Utrechtse Heuvelrug will come down to economic power and local political relations, even if all of the competing data sets could be standardized and made compliant with the desired ethical and quality principles. The examples also show that if a government body is forced to make decisions concerning digitization, the principles mainly serve as necessary preconditions (‘this must be implemented at the very least’), but do not provide firm foundations on which decisions can be based. This is because they embrace process values and present the end goals as mere operational ambitions: urban problems will be solved more efficiently, markets will become more innovative and government bodies and other organizations will be able to offer faster and more personal services.

**Harmonious context within which the ethical and quality principles and the ambition for standardization circulate, conceals important social oppositions and conflicts of interest**

Technology critic Evgeny Morozov calls this type of thinking ‘solutionism’: the conviction that every problem can be solved with technology. However, neither the technology itself nor the desired solutions are neutral and equally beneficial to all, as is demonstrated by the examples. For this reason, the current landscape of principles also needs signposts to precise and substantive end-goals. As explained earlier, these are currently pointing to the values of the multinational market economy (US) on the one hand and the values of the centrally governed state (China) on the other. In the former model, citizens are consumers, the government keeps it distance and only provides services to citizens, while in the other the citizens are subjects and the government controls and directs everything. Neither scenario gives citizens or their representatives any control of how (process values) and to what end (end-values) their digital society is developing.

In this regard, we have made much more progress in urban areas in the Netherlands and Europe, as shown by our analysis of the landscape of principles. It established that there is a reasonable degree of consensus concerning how we must further develop the digital society, although the question as to what end has not yet been extensively debated. How can we encourage discussion of these end-goals, what kind of digital society do we want anyway, and what role should the government play?
4. Digital Environmental Strategy

In the Netherlands, the various civil services cannot play an independent role in formulating the end-values for society as they need a public mandate to do so. Within the confines of the laws and human rights, there are infinite opportunities for citizens and social movements to explore, formulate and promote such end-values and negotiate them into policy. This happens simply through elections, but also through referenda and other forms of public consultations and – to an increasing degree – via direct citizen involvement in the policy preparation and execution that take place in living labs, testing grounds, learning studios, community enterprises, etc.

Citizens can also take over government duties if they think they can do better, an initiative put forward under the motto ‘The right to challenge’ (R2C). The Environmental and Planning Act (Omgevingswet), which comes into force in 2021, even obliges governments to ‘take the various regional interests into account’ while considering the planning and regeneration of public space. It states that ‘at the moment, government bodies are often the sole party involved in project decisions’.

The latter sentence, about government bodies often being the only parties involved in planning, is particularly interesting in the context of the ethical principles discussed earlier. The process values they propagate, demonstrate the same need for to shape our society collectively and in collaboration, rather than leaving it solely to government bodies, multinational corporations or a combination of the two. One could say that the ethical principles express a need for a Digital Environmental Strategy. It should not come as a surprise that spatial policy can offer a source of inspiration for the further concretization of principles for the digital society.

After all, we have been using spatial metaphors for a long time, such as the electronic highway, The Digital City, testing ground or data warehouse. ‘If we involve everybody in the design and discussion of physical public space, then why don’t we do this for the design of digital public space: data warehouses, dashboards, analytics and algorithms?’

Appropriate use of data in public space | 100
One can simply copy all the work that has gone into designing the Environment and Planning Act. The Ministry of the Interior and Kingdom Relations in the Netherlands has already assembled an extensive collection of tools for the creation of an Environmental Strategy and indicated exactly which steps need to be taken in the process. Inevitably, this begins with an analysis of existing policy and the formation of a broad project group. But then the ministry dictates that a whole range of external partners need to be engaged to jointly formulate the ultimate ambition for the environment and ‘construct the story’. According to the logic of the Environmental Strategy, it is not enough to simply guarantee the process values. If everybody has had their say, and the quality of the construction and spatial design (the data structure) is high, then the final results must represent a shared ambition and story, i.e. an end-value. By definition, the objective for the digital society must not be purely economic, as we found to be the case in the BSI and ISO standards. It also cannot imply a top-down governance model, as is the case in the oppressive Chinese system. Therefore, the Ministry of the Interior has advised municipal government bodies to consider their own role in the development of their Environmental Strategy, which can vary between regulatory, collaborative or facilitative.

If the process values are focused on collectivity and collaboration, then it makes sense that the end-goals will also place the emphasis on the ‘shared story’ as propagated by VSNU in its Research Agenda for the Digital Society. These stand for the following definitive values:

- **Sustainable**, i.e. the digital strategy and policy must be compatible with the ecological environmental agenda as well as being implementable and effective in the long term. Therefore, the collaboration between all parties involved cannot be a one-off: it must be continually repeated and embedded.
- **Harmonious**, which means that the digital strategy and policy must be open and inclusive, respectful of legal and moral frameworks and civil behavior, and must not further inflame existing differences.
- **Affective**, as the digital strategy and policy must also recognize and take into account the fact that technology does not raise purely rational issues: it can evoke a wide range of positive and negative emotions for particular individuals and groups.
- **Relevant**, which means that the digital strategy and policy must particularly involve the groups and interests that will be most affected by digital and data technology. In the social domain, for example, this will mean that benefit recipients will have to participate much more than they currently do in discussions concerning how the data transitions in the social domain should be designed and implemented.
- **Empowering**, i.e. the digital strategy and policy must also enable all parties to understand and evaluate the technology in question, and whenever possible, to use it.
- **Diverse**, the final value, not only entails that the diversity of society must be recognized and acknowledged, but also that the technology itself must be designed in a way that enables it to be used and applied in a diverse range of ways.

The SHARED values have been formulated in a sufficiently broad manner to enable a wide range of operationalisations and facilitate the ever-changing dynamics that characterizes all digital strategies and policy.
Nowadays, technology and society are changing so rapidly that the ability to make constant adjustments to the design of digital and data technology must be a built-in feature. For this reason, the innovation community now likes to talk about things being in a state of ‘permanent beta’ while critical technology researchers prefer to use the term ‘contestable by design’.

Evidently, we do not pretend that the SHARED values will prevent disputes and conflicts of interest. Although the collaboration between government bodies, businesses, knowledge institutes and citizens (or their representatives) are seen as a vital factor in ensuring successful and widely supported innovation, it is inevitable that differences, delays, arguments, annoyances, frustration and failures will crop up along the way. All levels of government play a dual role in this regard, both as collaboration partner and – in our opinion – as the only party within this complex network that is capable and legitimized to take an overarching role as the guardian of the collective process and end values. We cannot expect citizens, businesses or knowledge institutes to always take each other’s ideas and interests into account: this responsibility can and must be fulfilled by the government, which have been given the mandate to do so by the citizens who elected it.

We cannot expect citizens, businesses or knowledge institutes to always take each other’s ideas and interests into account

Admittedly, this is easier said than done as each of the individual government bodies contains a diverse range of contrasting and conflicting opinions as well, especially in relation to digital and data strategy. Differences of opinion also sometimes occur between different ministries and government institutions regarding the objectives for the implementation of digital and data technology.

The appointment of the municipal privacy officers – as mandated by the GDPR – has, for instance, created countless conflicts of interest between departments that want to do more with digitization and data and these new privacy gatekeepers who are responsible for interpreting the new law.

How should the alderman of Utrechtse Heuvelrug resolve the streetlighting issue and how should the government proceed with the general development of principles for the digital society? The process values for the design of the digital and data technologies have already been solidly formulated, as was shown by the first part of this essay. It is also clear that the government must carefully consider its desired end values for the digital society, as we stated in part two.

Furthermore, in the final part, we made an academically supported recommendation for the development of these end values and suggested that a Digital Environmental Strategy is the perfect tool to enable collective and good (from both an ethical and technical perspective) design of digital and data technology.

For now, our task is at an end, and we’d like to close by considering that through such Digital Environmental Strategies, the Netherlands and its provinces, municipalities and regional water authorities could potentially set a unique example in the quest to establish a digital society controlled not by the state or gigantic digital platforms, but by each and every one of us.
About the authors

Dr. Jiska Engelbert is Senior researcher at LDE Centre for BOLD Cities.

Prof. dr. Liesbet van Zoonen is Academic director at LDE Centre for BOLD Cities.
References
1 https://www.rathenau.nl/nl/digitale-samenleving/overzicht-van-ethische-codes-en-principes-voor-a/
2 https://innovationorigins.com/nl/eindhoven-ze-smart-society-principes-gaan-landelijk/
3 https://tada.city/
4 https://datasci.hk/deda/
5 https://www.rathenau.nl/nl/digitale-samenleving/opwaarderen
6 https://wetschappelijkk bureaugroenlinks.nl/artikel/handvest-voor-de-slimme-stad
7 https://www.thedigitalsociety.info/about/data-principles/
8 https://citiesfordigitalrights.org/
9 https://sharingcitiesalliance.com/
11 https://www.smartdubai.ae/
13 https://www.iso.org/sites/worldsmartcity/
14 https://www.nen.nl/Normontwikkeling/Doe-mee/Normcommissies-en-nieuwe-trajecten/NEN-smart-
15 http://espresso-project.eu/content/deliverables/, deliverable d7.4, p.11
16 https://wetschappelijkk bureaugroenlinks.nl/artikel/handvest-voor-de-slimme-stad
21 the%20Robot%20Age-Rathenau%20Instituut-2017.pdf
23 https://www.google.com/earth/outreach/special-projects/air-quality/
24 https://www.sharehsources.com/download/all:XIII.81
25 https://www.force11.org/group/fairgroup/fairprinciples
26 https://redasci.org/
27 Van Zoonen, L. (2019). Opnieuw fatale remedies: een kritische reflectie op datatransities in het sociaal
29 Idem, p.32-33
32 Footnote 33, idem.
33 https://www.sr.digitalestandards-institution/
34 https://www.nen.nl/Normontwikkeling/Doe-mee/Normcommissies-en-nieuwe-trajecten/NEN-smart-
36 https://www.iso.org/sites/worldsmartcity/
37 https://www.iso.org/sites/worldsmartcity/
38 https://www.rijksoverheid.nl/onderwerpen/omgevingswet/vernieuwing-omgevingsrecht
39 This was the name for an early internet platform initiated in Amsterdam (De Digitale Stad)
40 https://www.rijksoverheid.nl/onderwerpen/omgevingswet/vernieuwing-omgevingsrecht
42 Furthermore, a number of critics claim that the manner in which the national and regional government
44 The SHARED values were initiated by the LDE Centre for BOLD Cities.
45 Van Dijk, J., Poel, T. & M. de Waal (2016). De platformsamenleving. Strijd om publieke waarden in een
46 https://www.nen.nl/Normontwikkeling/Doe-mee/Normcommissies-en-nieuwe-trajecten/NEN-smart-
48 https://www.iso.org/sites/worldsmartcity/
49 https://www.iso.org/sites/worldsmartcity/
50 https://www.iso.org/sites/worldsmartcity/
51 https://www.iso.org/sites/worldsmartcity/
52 https://www.iso.org/sites/worldsmartcity/
53 https://www.iso.org/sites/worldsmartcity/
54 https://www.iso.org/sites/worldsmartcity/
55 For example, see the work of Gerd Kortuem at Delft University of Technology:
57 See, for example, the Letter to the Dutch Lower House of Parliament from the Minister of Economic
58 See, for example, the Letter to the Dutch Lower House of Parliament from the Minister of Economic
59 See, for example, Harthoff, S. (2017). Gemeenten hopen op duidelijke regelgeving voor sensoren.
The authors are currently in discussion with the alderman in question regarding the exploration of possible solution scenarios.
AI in the digital society

Quality of algorithms and decision-making

Prof. dr. Eric Postma
Professor of Artificial Intelligence
Tilburg University

Foreword
By Ran Haase

In 2015, we in Eindhoven were thinking about what the role of the municipality should actually be when it comes to collecting data in the public space. The reason for this was that to an increasing degree, we were being confronted with a range of different parties that wanted to collect data about visitors to the city. Our thought experiment took us through fundamental questions about the public interest that we, as municipality, stand for: questions about ownership of data, fundamental rights, the value that can be created with data and the interest of the local resident. Ultimately, this process led us to the Open Data Principles and the IoT Charter. These principles were then developed into the ‘Ground Rules for the Digital City’ (Spelregels voor de Digitale Stad) in Amsterdam and Eindhoven, designed to offer a game plan for equitable data use.

Thankfully, since that time ideas surrounding data use have not been standing still. The debate on data and (more recently) AI is making greater and greater inroads into all manner of domains within the municipal landscape. The opportunities and challenges are becoming increasingly clear, but remain quite difficult for many politicians, public officials and individuals in society to really understand. Be that as it may, over the past year we in the world of municipal government have been seeing growing calls for ethical considerations in big data projects.

It is gratifying to see that the Ministry of the Interior and Kingdom Relations has taken the initiative for the living lab on appropriate data use, and that the Ministry was fortunate enough to find Prof. Eric Postma willing to write the following essay. Prof. Postma is a full professor in the Department of Cognitive Science & AI at Tilburg University, and at the Jheronimus Academy of Data Science (JADS) in ’s-Hertogenbosch, a partnership between Eindhoven University of Technology and Tilburg University. His research focuses on the development and application of machine learning in the automatic recognition of images and signals.

I heartily endorse Prof. Postma’s recommendations: AI must not develop in a vacuum. There is a need for investment in top-level research. At the same time, we need to be having a debate about the ethics and legal and ethical principles of the society we want to be.
This is why it is important for the government to invest in knowledge development among politicians and policymakers, to allow them to make their contribution to this debate in a meaningful and responsible way. Prof. Postma’s recommendations are important steps in a debate that we have to be having, and keep having into the future.

Ran Haase
Legal advisor, Municipality of Eindhoven
Member of the Supervisory Committee

Summary

This essay is intended to present a level-headed look at artificial intelligence (AI) and its role in the society of today and the future. Its emphasis will be on the use of data and AI algorithms in the public space. It attempts to answer the question of how to avoid data use gone wrong and bad applications of AI algorithms in our society.

After an introductory consideration of algorithms, machine learning, and the recent ‘AI revolution’, it will go on to discuss: (1) the difficulties policymakers face in understanding all aspects of AI and responsible data use, (2) the AI revolution as both a technical and social revolution, which calls for an interdisciplinary approach, (3) the crippling shortage of professionals in a wide range of domains with adequate knowledge of the potential and risks of AI algorithms, and (4) why certification of AI algorithms by an independent body is necessary.

On the basis of these determinations, five policy proposals are formulated. These are: (1) give politicians and policymakers regular education and training on data use and the emerging advances in AI in relation to their social and legal aspects, (2) invest in education and training to produce interdisciplinary professionals to address the crippling shortage in this area, (3) improve the circumstances for AI researchers in order to stop the brain drain of researchers, (4) invest in research to facilitate the transition to an equitable digital society, and (5) the government should consider making review of AI algorithms mandatory in order to minimize the risk of abuses and wrongdoing. Implementing these policy proposals can help prevent things from going wrong as our society embraces the application of AI algorithms.
1. Introduction

Artificial intelligence, also referred to as ‘AI’, is a hot topic in modern society. Wild predictions and warnings about AI are fuelling an ongoing sense of concern and uncertainty. This essay is intended to present a level-headed look at AI and its role in the society of today and the near future. The focus will be on data in the public space, which in some ways can be considered the ‘oil’ that AI algorithms run on. We will take a look at four observations and then build on them with five policy proposals.

Before going into the key question, this essay will first prevent a broad outline of what AI is and why at present it is taking on such a prominent place in the media and the public debate.

1.1 What is AI and why is it important in the digital society?

AI as a research area has a long and storied history. The term ‘Artificial Intelligence’ was first put forward by John McCarthy in 1956. The objective of AI is to understand and create ‘intelligent systems’. The standard textbook used in AI offers the following four distinct definitions of the concept of artificial intelligence:

• Systems that think like people
• Systems that think rationally
• Systems that act like people
• Systems that act rationally

These four definitions combine the concepts of ‘thinking’ and ‘acting’ with ‘like people’ and ‘rationally’. Thinking and acting like people means understanding and imitating people, while with the idea of thinking and acting ‘rationally’, we lean towards envisioning non-human systems driven by pure logic. It is an image we can find in countless science fiction films in which a humanoid robot reasons in a completely logical (rational) manner while exhibiting no emotions or empathy. The existence of multiple definitions is an indication that there are many different interpretations of what AI means.

Taking a look at the potential that AI offers can help clarify the picture. For example, AI has given us extremely powerful chess computers that can beat the best human players. Such systems are rational in the sense that they are based on the purely logical analysis of all consequences of every possible move and countermove. This is only possible because the world of the chess game is in fact quite limited, and so lends itself well to being described in terms of logic. To play chess, the chess computer does not need to understand what a chesspiece is, or what the chessboard is. In other words, the chess game is something relatively simple, that can be formalized in terms of logical rules that, in turn, are easy to implement in a computer program – so easy, in fact, that it was in the waning years of the last century, 1997, that an early AI system, IBM’s ‘Deep Blue’ computer, was able to beat world chess champion Garry Kasparov.

But the real dream of the first generation of AI researchers was to create an AI system with the capacity for rational thought; a system that could reason on the basis of pure logic. The success of Deep Blue was seen as a major step forward towards the achievement of an artificial, rationally thinking system. After all, Deep Blue had proven itself able to ‘out-reason’ human beings in the game of chess. But taking this into the real world – creating a chess computer or robot that could interact with the chessboard, for example, using a video camera in combination with a robot arm, proved to be a far greater challenge. Such a system had to identify all chesspieces in a variety of conditions, such as in different lighting levels and even if obscured by other chesspieces on the board. The visual recognition of chesspieces, the chessboard, and the opponent proved to be a major challenge, because these are things that cannot be formalized in obvious ways. This is the reason that we have had chess computers, fast processors, and automated planning systems for quite some time, but systems that can automatically ‘recognize’ and describe images are only emerging now. These are products of the recent ‘AI revolution’, which has produced breakthroughs in hard-to-formalize tasks such as the analysis of images, audio and text. The rise of data-driven AI algorithms with the capacity to analyse and recognize patterns and images, audio and text has come in parallel with the increasing importance of the quality of data. The quality of an AI algorithm is directly dependent on the quality of the data.

1.2 Research question

The new AI algorithms are having an enormous impact on society. The sheer number of potential applications has led to real shifts in the distribution of responsibilities between man and machine. And this, in turn, has led to a
restructuring of our society, in part in the form of changes in our regulations and our standards. Data is a central element of this process. Virtually all AI innovations have been the result of the collection and preprocessing of data. This brings us to the following research question.

What do we need to do to avoid data use going wrong and bad applications of AI algorithms in our society?

The answer to this question must come from a consideration of the ‘old’ and the ‘new’ algorithms (chapter 2), machine learning algorithms (chapter 3), and the AI revolution (chapter 4). This consideration leads us to four observations on which we can base the answer to our question (chapter 5), culminating in five policy proposals (chapter 6).

2. Deterministic algorithms versus probabilistic algorithms

Ordinary algorithms have been playing a vital role in our society for decades. All computers function on the basis of algorithms that ensure that a multitude of data processing actions, ranging from low-level tasks like reading and writing data to a computer memory, to high-level actions like properly translating keyboard commands to computing actions, are handled correctly. This type of algorithm is completely deterministic, which means that with a given starting state, the algorithm will always reach the same end state. For example, when a computer user wants to save a file, she can use the key combination Ctrl-S. The command Ctrl-S activates an algorithm that performs a predefined series of electronic actions that ultimately result in the saving of the file. The programming language used to define these actions is a formalized collection of instructions, which are commonly based on IF-THEN constructions. To program the Ctrl-S functionality, an algorithm will typically use the following structure:

\[
\text{IF key = 'Ctrl' AND key = 's' THEN save_file}
\]

The THEN element specifies the electronic action of saving the file. The use of IF-THEN rules in algorithms is an expression of the formal basis of computer algorithms. Researchers in theoretical computer science study algorithms through the lens of formal logic. It is formal logic that allows us to see that an algorithm is functioning as designed. The capacity to do this is relatively simple for formalized tasks such as rule-based tasks. Rule-based systems generally mean that if a number of conditions are met, the rule is applied. This is something that is easy to define in a computer algorithm.

\[
\text{IF condition 1 AND condition 2 AND... THEN the rule applies.}
\]

As long as rules are unambiguous and consistent, they are easy to translate into algorithms.

Probabilistic algorithms are distinguished from deterministic algorithms in that they have a probabilistic element. As an example, consider a lottery: The winning number is determined by a ‘random number generator’. Because of this probabilistic element, there is no way to say with certainty what the result of running the algorithm will be. Instead, the result can only be expressed in terms of probabilities. With a die-rolling algorithm, for example, you can only predict that the chance of a ‘1’ as output is equal to 1/6th, or that on average the output of the algorithm will be ‘1’ one out of six times. The distinction between deterministic and probabilistic algorithms is important for understanding modern machine-learning algorithms.
3. Machine learning algorithms

Machine learning algorithms are probabilistic algorithms that attempt to make predictions on the basis of data. There are three types of machine learning algorithms that each differ in the way they learn: supervised learning, unsupervised learning and reinforcement learning.

Supervised learning means learning on the basis of examples that are identified as they are given in the learning process. For example, when supervised learning is used to train a machine learning algorithm to identify suspicious behaviour in video input, the training examples will consist of video clips of persons. The video clips are identified or labelled as ‘suspicious’ or ‘not suspicious’. In effect, the label defines the nature of the recognition task. In a simplified example the same videos could be labelled ‘man’ and ‘woman’ to train the algorithm to identify the sex of the person in the video.

Unsupervised learning is based on the examples only, with no labels. Using the previous example, with unsupervised learning the machine learning algorithm could be trained to group the videos based on the most prominent visual correspondences. The groups of the output from the unsupervised learning algorithm might be based on a number of different correspondences, such as the type of weather or the amount of movement in the video. The absence of labels results in less control of the result. For this reason, unsupervised learning is less successful than supervised learning. Generally, unsupervised learning is used to gain insight into the data, because it produces insights into the grouping of data points. Because for many applications there are no labels available or labels are difficult to obtain, there is an ambition in the field of AI to broaden the applications for unsupervised learning.

Reinforcement learning is generally used to teach sequences of actions, like the process of teaching a robot to navigate. In reinforcement learning, the machine learning algorithm generates actions that are rewarded if they are correct. This learning method is analogous to the method that has been shown to be successful in teaching animals complex behaviours: Each time the animal exhibits the desired behaviour, it is rewarded with a treat or a show of affection. In 1950, psychologist B.F. Skinner used reinforcement learning to teach pigeons to play table tennis with their beaks. In the field of AI, reinforcement learning is currently being used very intensively (and successfully) in formalizable domains like video games or computer variants of boardgames like the ancient Chinese game of Go.

In this essay we will be restricting ourselves to supervised learning, because this is currently the leading and most successful variant of machine learning.

The probabilistic nature of machine learning algorithms is not only a function of the use of a random number generator, but also the probabilistic nature of the data used for ‘training’ these algorithms. The major tech companies rely on machine learning algorithms for a great many purposes, like predicting whether a customer is interested in certain products. Google collects tremendous amounts of data about its users’ search, navigation and purchase behaviours, because the company can use this data to train machine learning algorithms with supervised learning techniques. This training can take on various different forms, but always results in a statistical model that uses a section of the data (the ‘input’) to predict other data (the ‘output’, also referred to as the label). After training, these models can be used for prediction of previously unseen inputs, for example analysing new customers’ internet behaviour (the input) to predict what products they will be interested in (the output). After training, most machine learning algorithms perform in a deterministic way: given the same input, they will always generate the same output. As already described, the probabilistic nature of machine learning algorithms is primarily a function of the data.

Let’s consider an extremely simple example. Suppose that as the input, the sex of the visitors to a webshop is known. The webshop can track how often women make purchases from the categories ‘tools’ and ‘kitchen items’, and how often men do. Based on the frequencies observed of purchases of tools and kitchen items, the webshop’s machine learning algorithm can derive the probability that a woman or a man will buy tools or kitchen items. After training, the algorithm will be able to use a new customer’s sex (the input) to predict the output (in this case, the prediction of whether a customer will be interested in tools or in kitchen items). The machine learning algorithm makes a prediction of the purchase on the basis of the customer’s sex.
Although the example described here is extremely simple, it is a good illustration of the essence of machine learning algorithms. After training, these algorithms make statistical models of the reality as represented in the data. The quality of the algorithms therefore depends on the quality of the data. All the aspects that are important for good statistics are also important for the proper functioning of machine learning algorithms. To take one example, the ‘random sample’, or the data on which the algorithm is trained, must be representative for the population that it is being used for. If, for example, the algorithm described above is trained on data of men and women from a population group in which the women typically perform household duties, the predictions will be completely different from when the algorithm is trained on a population group in which the men normally perform the household duties. Due to the probabilistic nature of the data used for training machine learning algorithms, the basic principles of statistics are the most fundamental basis for understanding the functioning, potential and limitations of machine learning.

3.1 The statistical basic principles of machine learning

Supervised training of a machine learning algorithm requires vast numbers of input-output instances. Looking at our simple example, an input-output instance is made up of the data of a customer who has made a purchase, for example a male customer who has purchased kitchen items. This instance now consists of the input-output pair ‘male’-‘kitchen items’. In the prediction of election results, researchers increase the reliability of the predictions by (1) taking a representative random sample, and (2) collecting as many respondents as possible. This has a direct parallel with machine learning algorithms. The quality of the predictions obtained from a trained machine learning algorithm depends on (1) the representativeness of the instances and (2) the quantity of the instances.

The first point, (1) representativeness, has already been referred to above. A current example is the object recognition algorithms used by Google, Amazon, and other major tech companies. Recent research has shown that everyday objects like toothpaste can be easily recognized in images from wealthy countries, but the results are extremely bad on images from poorer countries. The reason for this is that the algorithms have been trained on input-output instances (such as an image of a tube of toothpaste as input and ‘toothpaste’ as output) the vast majority of which have been obtained from rich, Western countries.

A photo from an extremely poor country like Burundi will show a tube of toothpaste against a background of wood instead of a tiled bathroom, and so will produce the erroneous output ‘wood’.

The reason for this is that the algorithms have been trained on input-output instances from rich Western countries

The second point on which the quality of a machine learning algorithm’s predictions depends is (2) quantity; this refers to the number of instances used to train a machine learning algorithm. For an algorithm to function well, it needs an adequate number of labelled instances. One of the major reasons for the current explosion in machine learning and AI is the existence of huge tech companies that have access to enormous quantities of data.

3.2 Complex machine learning algorithms

Simple machine learning algorithms have been around for a long time. The earliest proposal for what was referred to as the ‘nearest neighbour’ algorithm goes back to 1951. Since this early machine learning algorithm, learning algorithms have been investigated and applied in many forms, and under many names – such as pattern recognition, data mining and knowledge discovery. Although this variation in nomenclature indicates subtle differences in the approach or application, the fundamental common denominator across all of them is the creation of a predictive model (or algorithm) on the basis of data. Today, ‘machine learning’ is the generally used umbrella term for learning algorithms.

There is a tremendous variation in machine learning algorithms. These algorithms differ in the way in which they make predictions based on the training data. With any given prediction task – for example the prediction of whether a person in the public space will exhibit criminal behaviour on the basis of detailed location data from her mobile phone – there is no way to know in advance which algorithm will work best. For this reason, AI researchers experiment with different algorithms to determine which produces the best predictions for a given task.
One important dimension in which machine learning algorithms differ is the complexity of the prediction task. Suppose criminal behaviour is easy to determine based on location data (for example, because a specific location is where criminals hang out). In this case, the prediction task is extremely simple, and we can suffice with a simple machine learning algorithm. In this hypothetical example, it is that simple: the location predicts criminal behaviour. In practice, however, predicting criminal behaviour is obviously much more complex, and finding a predictive element in the combination of location data requires a more complex machine learning algorithm.

The advantage of a more complex machine learning algorithm is that these are generally better able to make predictions. This better performance in making predictions comes at the cost of the interpretability of the prediction. In the case of the simple machine learning algorithm, interpretability is high. If the algorithm predicts criminal behaviour, it is because the person spent time at the ‘criminal location’. This makes the reason for the prediction easy to explain. The interpretability of a complex machine learning algorithm is much more difficult, due to the abstract combination and weighting of data elements that are generally difficult to explain in natural language.

The European Group on Ethics in Science and New Technologies formulates this principle in its 2018 statement this way: ‘“AI’s inner workings can be extremely hard – if not impossible – to track, explain and critically evaluate.’

The trade-off between predictive performance and interpretability is one of the key issues in the AI revolution. At the furthest extreme, the powerful deep learning algorithms at the heart of the AI revolution achieve incredible performance in their predictions, but demand a high price in terms of lack of interpretability. For this reason, a great deal of research is currently being performed into ‘explainable AI’. Thanks to this research, we can now use what is referred to as Q&A interaction to have an AI system explain why it is classifying an image, video or text in a certain way. For example, when the AI system classifies an image as being a skier, and the user asks ‘why?’, the system marks the section of the image where it has identified skis and generates the sentence: ‘because I see two parallel straight lines against a white background’. This form of explanation is achieved by training the system on a large number of images and explanations associated with them.

However, Peter Norvig, a world-renowned AI researcher and research director at Google, questions whether explainable AI is a worthwhile undertaking. He says:

“You can ask a human, but, you know, what cognitive psychologists have discovered is that when you ask a human you’re not really getting at the decision process. They make a decision first, and then you ask, and then they generate an explanation and that may not be the true explanation. So we might end up being in the same place with machine learning where we train one system to get an answer and then we train another system to say – given the input of this first system, now it’s your job to generate an explanation. Explanations alone aren’t enough, we need other ways of monitoring the decision making process.”

AI systems, he says, would generate an explanation in the same way, after making a decision. Just like with people, the motivation would be to come up with a good and convincing story, just not necessarily one based on the truth. Despite this justified concern, any insight into the way an AI system makes its decisions is important, because the ‘narrow’ intelligence of the system must be complemented and controlled by the broad intelligence of the human mind.
4. The AI revolution

The AI revolution is being driven by a class of new machine learning algorithms referred to by the term ‘deep learning’. While it is true that the underlying ideas behind deep learning have been around for 30 years, it is only in the last ten that the technology has been producing dramatic breakthroughs. The reason for this is twofold: (1) thanks to the rise of the internet and the major tech companies, enormous volumes of labelled data have become available, and (2) the powerful graphics cards developed by companies like NVIDIA for cutting-edge, hyper-realistic video games have radically accelerated the process of training deep learning algorithms.

The major breakthroughs that deep learning has delivered have been in the areas of image and video recognition, speech recognition and text analysis. Today, hardly a week goes by when the media is not awash in reports of a breakthrough in at least one of these areas. Before presenting a list of applications with relevance for data use in public domain, it is useful to outline the broad limits of deep learning-based AI. We can do this with an example.

If we train a deep learning algorithm on millions of images of objects, with every object labelled with the name of the object shown (like ‘Fido’) or the class to which the object belongs (‘dog’), then the algorithm will be capable of assigning the correct label to images it has never seen before. To put this in perspective, if we take a new photo of Fido and present it to the deep learning algorithm as input, then in 95 to 99% of cases it will produce the output ‘Fido’ or ‘dog’. If we train the algorithm on millions of images with their corresponding descriptions, then the network can even produce a correct description for new images. Ten years ago, it was unthinkable that any computer system would be able to do this. Consequently, this represents a truly enormous breakthrough in scientific and practical terms. It is such a dramatic breakthrough that the temptation is great to draw a parallel with human observation.

After all, isn’t the system doing what humans do – identifying and classifying an image? However, there is a crucial difference between the way in which deep learning ‘recognizes’ an image and the way humans do the same thing. A deep learning algorithm knows nothing about dogs.

It has no knowledge of animals, and does not know what a dog feels like or does. The only thing that the network ‘knows’ is that certain visual patterns are associated with the label ‘dog’. This is why in the AI field, deep learning is referred to as a ‘narrow’ AI. The algorithm performs excellently on a very narrowly defined task. Although it is common to see proclamations like ‘a deep learning network has learned to recognize a dog’, this is in actuality very misleading because what it is really doing is associating an image with a label.

To illustrate, we will consider a research project in the public space that we conducted in cooperation with the Dutch Institute for Technology, Safety & Security (DITSS) in a public square in Tilburg (Piusplein) and a public location in Eindhoven (Stratumseind). The goal of the project was to predict aggressive behaviour on the basis of video images.

One of the observations was that wild but innocent behaviour by a group of friends was effectively impossible to distinguish from aggressive behaviour. It turned out that the human observer, by applying her general social knowledge of how a group of friends behaves, is much better able to predict an escalation of aggression. In concrete terms, this means that the application of deep learning for the ‘narrow AI’ task of identifying potential aggressive behaviour must be coupled with a verification by a human observer. One could say that the narrow intelligence of an AI system is complementary to the broad intelligence of the human being.

4.1 AI applications in the public space

The strong performance of AI and deep learning algorithms in ‘recognizing’ images and audio has led to a broad range of potential applications in the public space. In recent years, applications of deep learning and other AI methods have been investigated extensively in the context of ‘smart cities’. The expectation is that with the rise of the ‘internet of things’ (IoT), such projects will begin to really take off, especially as fast 5G technology becomes available. In the following, we will discuss the three main public space applications: (1) monitoring and directing visitor behaviour, (2) monitoring and directing traffic flows, and (3) monitoring and directing safety.
4.2 Monitoring and directing visitor behaviour

Tracking the visitors to a city centre or other public spaces can be done in a number of ways: by using video cameras, by tracking mobile phones, or by using special sensors placed around the public space. The visual recognition of pedestrians and the prediction of pedestrian behaviour has become a hot topic in recent years due to the rise of deep learning and autonomous vehicles. Every year, new advances in machine learning improve the quality of this type of recognition.

We can now establish the identity of pedestrians from video images, although results vary based on the quality of the image and other aspects such as visibility of facial characteristics. In many cases the systems are not yet so robust that they are better at identifying faces than human beings are. Additionally, many systems have a bias in their recognition performance. Amazon’s facial recognition system Rekognition, for example, has been criticized for its implicit racial bias. It performs more poorly on the facial recognition of black women, for example, due to the lack of representation of black women in the data set on which Rekognition was trained; this reflects on the statistical basic principles described in section 3.1.

To some degree it is possible to identify facial expression from video images, which can be used to make an assessment of the emotional state of a visitor to the public space. There are also methods available that can identify and track a visitor from video images based on body language, and this could lead to the ability to identify and, to some degree, predict behaviour.

Regardless of the way in which the assessments of the locations and behaviours of the visitors are obtained, directing behaviour (or ‘nudging’) is a logical next step. For example, if the maximum capacity of a popular shopping boulevard is close to being reached, various incentive methods can be used to lead visitors to other boulevards or destinations. In the living lab at Stratumseind in Eindhoven, there are intensive experiments being conducted with nudging by adjusting the intensity and colour of the lighting.

In the near future, nudging will be able to be more personalized. Instead, we could display a virtual face that could use subtle verbal and nonverbal signals to nudge the visitor to perform a desired action. We also have the technology to generate and manipulate virtual faces. The tech firm Soul Machines, recently purchased by IBM, can create dynamic virtual faces that are indistinguishable from real faces. It is only a matter of time before such systems are being used widely. Our research has shown that people are sensitive to non-verbal signals (like a smile or a subtle furrowing of the brows) of virtual faces, and respond to these signals unconsciously and non-verbally. This type of nudging manipulates our social brains, and represents all the opportunities and dangers that that implies. Just as a good salesman can learn how to effectively respond to a customer to maximize sales, a deep (reinforcement) learning algorithm can learn how to direct the virtual space using the optimum combination of social signals to induce visitors to the public space to exhibit the desired behaviour.

Although such a system does not yet exist, it is within the technological capabilities we have now, and it is only a matter of time before someone begins developing one. This realistic future scenario is an illustration of the fact that the tremendous potential and risks of the AI revolution are focused at the interface between the technology and the human being.

In the near future, nudging will be able to be more personalized
4.3 Monitoring and directing traffic flows
Tracking and regulating traffic flows in the city is a huge challenge. Many cities are already using 'Adaptive Signal Control' technology to optimize the flow of traffic. Such systems track the traffic and use traditional AI algorithms to optimize the functioning of traffic lights. The rise of 5G technology and autonomous vehicles will drive the innovations needed to improve the flow of traffic in a city centre. Specific deep learning applications include visual tracking of cars on the basis of automatic recognition of the number plate (something that is already being done on a wide scale in parking garages) and the automatic identification of unsafe driving behaviour or traffic violations. One application that many cities are currently experimenting with is 'smart parking'. Available parking spaces are detected using video images or sensor information, and this data is then used to navigate motorists looking for a parking space automatically to these free spaces.

4.4 Monitoring and directing safety
The monitoring and promotion of safety in public spaces is, obviously, of great value to society. A number of cities are now experimenting with using automatic video analysis, acoustic cameras or IoT sensor systems to detect escalating situations at an early stage and track them, so as to be able to intervene when needed. In Eindhoven’s Straatunseind, a popular area for nightlife in the city centre, a living lab is set up that experiments with new technology. One major point of attention for the living lab is that no privacy-sensitive information can be collected. There are also experiments being conducted in public spaces in the cities of Enschede and Utrecht. An article in The Guardian took a critical look at these several Dutch projects because of the privacy concerns regarding the collected data and the blending of public research and private interests.16

In view of the experimental nature of the projects and the lack of clarity in the legislation on the collection of data, such critical voices are unavoidable and useful. It shows how important experimentation is in arriving at a good approach that can stand up to every ethical review. With regard to the blending of research financed with public money and private interests, it is worth noting that in the academic world, the public-private partnership form of research is being actively encouraged. Initiatives like the living lab at Straatumseind owe their success in part to the intensive partnership between the government and industry.
This does not alter the fact that for this type of public-private partnership project, clear guidelines need to be adhered to, such as those formulated by the Association of Universities in the Netherlands (VSNU).17

In the United States, the prevailing conceptions of what and how much data can be collected in public space are generally much broader, at least for the time being. There is, for example, the tech company ShotSpotter, which uses a network of microphones placed around the public space to detect gunshots.18 Whenever a firearm is discharged, its location can be determined using the sound picked up by the nearby microphones.

Initiatives like the living lab at Stratumseind owe their success in part to the intensive partnership between the government and industry

AI technology is used to identify and localize the specific noise; the system is not hindered by the ordinary background noise of the city, traffic, weather and people. The company’s website suggests that ShotSpotter can lead to a reduction in the frequency of guns being fired in a city. A Dutch version of this technology could be geared towards other noise sources, like the detection of human ‘distress calls’ (screams of fear, panic, etc.) or car alarms. Obviously, the introduction of a network of microphones would lead to privacy concerns. Microphones would, of course, also pick up human conversations. It is possible to train AI systems (and particularly deep learning algorithms) to not include certain aspects of the input.

In China, the use of AI technology in the public space is at an advanced stage; there, they are even experimenting with automatic facial recognition of criminals in video images. In 2016, two researchers in Shanghai published a controversial study20 in which a deep learning algorithm was used for automatic screening of faces. They trained the algorithm on passport photos of over 1000 non-criminals and over 700 criminals. After training, the algorithm was apparently able to identify criminals and non-criminals in previously unseen passport photos with an accuracy of nearly 90%. The implicit assumption that criminals can be distinguished from non-criminals by visual observation is, of course, extremely controversial. More significant is that the training method used has a statistical bias. The training data for the criminals was obtained from photos of criminals who had already been caught when the photos were taken. It is extremely likely that criminals who have these prototypical ‘criminal’ facial characteristics are arrested more readily than criminals who do not have these characteristics. By ‘prototypical criminal faces’, here we are referring to faces of criminals as they are depicted in films. It is easy to see how the most successful criminals (that is to say, the criminals who do not get arrested) might be the ones who do not look like prototypical criminals.

AI-based video surveillance offers the potential to observe criminal activity or intent at the earliest possible stage. Examples include automatic detection of pickpocketing or theft (or the intention to commit these acts). It has long been known that both experts and non-experts are very capable of predicting criminal behaviour in the public space from watching video recordings.20 In 2016, researchers at MIT were able to prove that prediction of intention on the basis of video images is possible. These researchers trained a deep learning algorithm on over 600 hours of video fragments from TV shows like ‘The Office’ and ‘Desperate Housewives’. After training, the system was capable of predicting whether two persons meeting were going to hug each other, kiss, shake hands or ‘high-five’.

The automatic recognition of criminal intent (such as pickpocketing21) has long been the subject of investigation. The rise of deep learning algorithms will, in the coming years, lead to many possible ‘predictive policing’ systems. The primary obstacle to creating such systems lies not in the technology, but the data. Deep learning systems require very large data sets. Collecting a sufficient number of video images of an action such as pickpocketing is no easy task. No doubt there are already companies and institutions investing in the collection of data for this type of application. As soon as commercial products become available these can represent an important tool for the monitoring and management of safety in the public space.

Appropriate use of data in public space | 117
5. Observations

Looking at the latest developments in technology and society, we can arrive at the following four observations.

1. It is difficult for policymakers to fully understand AI and equitable data use.
2. The AI revolution is both a technical and a social revolution, and as such requires an interdisciplinary approach.
3. In a wide variety of domains, there is a crippling shortage of professionals with sufficient knowledge of the potential and risks of AI algorithms.
4. Certification of AI algorithms by an independent body is needed.

These four observations are addressed in more detail below.

5.1 Complexity

The importance of AI and responsible data use for the development and strategic position of the Netherlands is generally acknowledged, but for politicians and policymakers the underlying subject matter proves to be something difficult to understand. We can identify four reasons for this.

The first reason is that the integrity of the discussion surrounding the development of AI is tainted by predictions that reflect more on Hollywood screenplays than the reality. One prominent example is futurist Ray Kurzweil, who in 2005 introduced the concept of the ‘Singularity’ as the point at which the exponential growth of technological potential will result in a convergence, or ‘singularity’, of human being and machine. Kurzweil postulated that the Singularity will take place in 2045. His predictions are, in part, responsible for a plethora of unfounded ideas about the development of AI. One prominent example of an AI doomsayer is technocrat Elon Musk, who has predicted that AI will supplant humanity and human intelligence. It is beyond the scope of this essay to go into the details of all these predictions, but one clear problem with the Kurzweil-esque future scenarios is that they are based on an extrapolation of exponential growth. In the real world, exponential growth has its limits. The exponential growth of the number of internet users, for example, is limited by the total world population; eventually, growth becomes saturated and in reality the growth curve takes the form of an S-curve rather than a hyperbolic curve.

In much of the same way, the exponential growth of technological potential of any given technology at some point reaches saturation. The impact of the concept of the Singularity can be partly explained by the futures presented in major motion pictures. Many of the most successful science fiction films toy with the idea of the machine that surpasses humanity. To offer an alternative to the seemingly plausible Kurzweil scenarios, every politician and policymaker should familiarize themselves with the article entitled ‘The Seven Deadly Sins of AI Predictions’ by American AI and robotics researcher Rodney Brooks. In this article, Brooks convincingly destroys every dire and misleading AI prediction.

The second reason that policymakers have difficulty getting a grip on AI is the tendency to wrongly generalize performance of AI in formalizable domains and then project that performance onto the real world. As we have seen, a formalizable domain is something like a chess game, or a board game like Go, or games that play out in a virtual world like a video game. In recent years, tech company Deep Mind has delivered impressive performances with ‘deep reinforcement learning’, a combination of deep learning and reinforcement learning. With their AlphaGo AI system, they succeeded in beating the human world champion of the game of Go. A later version, AlphaGo Zero, learned the game by playing nearly 5 million games against itself. In 100 matches against its predecessor AlphaGo, AlphaGo Zero went undefeated. Meanwhile, Deep Mind developed an AI system capable of outperforming human players in video games. The company achieved these impressive results because of the fact that the world of a game, whether Go or even the most complicated of video games, is simpler and much more restrictive than the real world. At every step of the game, there is a limited repertoire of actions and options. Despite this broad limitation, the number of possible sequences of actions is dizzyingly large. This is why it takes human players a great deal of time to get a handle on any given game. It is tempting to project the performance of AI in games into possible AI performance in the real world. However, the complexity of the real world is greater by many orders of magnitude.

The third reason why AI is incredibly difficult to understand is how fast things change in AI technology. The rate of development in the field makes it very difficult for non-experts to stay on top of, and even understand, the state of the technology. While this is not the exponential growth curve discussed critically above, it is still a stupefying rate of relatively small innovations that can and do
have an enormous impact on aspects like data use. As an illustration, consider an example of responsible data use in relation to the rise of generative models. Generative models are a variant on deep learning that offer the ability to generate ‘fake videos’ and ‘fake images’. A relatively unforeseeable effect of generative models relates to the privacy protections on data from sources including the public domain. De-identification is a proven method for providing privacy protections when using public data. When de-identification methods are applied, all direct references to individual data are removed. There are many data sets that have been anonymized by de-identification in circulation around the world, and which are being used for research and analysis purposes. However, in recent years multiple incidents have revealed that there is still nonetheless individual data that can be obtained from these de-identified data sets.

Using advanced machine learning algorithms, re-identification proved to be possible in many cases. This kicked off a public debate that produced a number of important results, perhaps the most important being that re-identification is only possible when the complete data set is available. Given that most of the publicly available data sets are only partial, the risks of re-identification were deemed to be negligible. However, a recent article showed that now, with a generative model, re-identification using a partial public data set is, in fact, possible.

This has direct ramifications for operative European privacy law (the General Data Protection Regulation, or GDPR) and other regulations governing the sharing and publishing of data. This example shows that what is good policy in regard to sharing data can suddenly change dramatically as a result of a relatively modest innovation in machine learning.

The fourth and final reason why policymakers have trouble getting a grip on AI is that by and large, these persons come from a non-technical background. To illustrate, I cite below from the Dutch government’s ‘Profile of the new Lower House of Parliament 2017’ as published on the website www.parlement.com.

The educational level of the members of the Lower House remains unwaveringly high. Nearly two-thirds (94 members) have a university degree and 31 members completed higher vocational education (Dutch HBO or equivalent). The educational level of a few members for the PVV is not known. Of the members with university degrees, degrees in law remain the majority (26 members). Economists (17), political scientists (15), public administrators (11) and historians (10) are also well-represented. The house also has a philosopher, a specialist in the philosophy of law, an architect and an archaeologist. There are now no longer any medical doctors in the House. There are few (namely, five) members with a science background. (https://www.parlement.com/id/vkclmgzmv4ya/profiel_nieuwe_tweede_kamer_2017)

Despite the high educational level of the members of the House, there is essentially zero affinity with technical fields to be found among them. This makes it difficult for most parliamentarians, politicians and policymakers in general to correctly assess innovations in AI.

For these four reasons, politicians and policymakers are insufficiently aware of the potential and pitfalls of AI algorithms.

5.2 Interdisciplinary approach
The impact of advances in AI comes from the social dynamics of people in interaction with technology. This is the reason that the essence of the AI problem is interdisciplinary.
Previous technological innovations such as the rise of mobile phones, the internet revolution, and the social media revolution, were also characterized by the interweaving of technology with social dynamics. The development of GSM technology came paired with a social need to always be available and able to communicate. The technological development of the internet and social media was linked to the social desire to communicate efficiently and advertise one’s self or a product. The integration of AI technology with social dynamics may be less apparent, but is potentially much farther-reaching. This is expressed particularly cogently in a major opinion piece in *Scientific American*, written by Dirk Helbing, Professor of computational social science at Swiss University EZTH in Zurich, and eight of his international colleagues in a number of other disciplines.

In it, these authors assert that modern society stands at a crossroads. Looking at Singapore as a data-controlled society, with China following in its footsteps, Helbing and his colleagues observe that steering human behaviour through ‘persuasion’ and ‘nudging’ may begin innocently, but can very quickly take on nightmarish forms. An example of innocent nudging is encouraging environmentally-conscious behaviour in the public space. At the other end of the spectrum is the Cambridge Analytica scandal, an example of a malicious form of nudging that made headlines around the world when it became apparent that internet users’ participation in the democratic process had been successfully manipulated.\(^2^7\)

An example of **innocent nudging** is **encouraging environmentally-conscious behaviour in the public space**

Finding the right balance of encouraging beneficial use of data and fighting data abuse was one of the points of attention in the reports of two ‘World Cafés’ on equitable data use in public space in Groningen and Eindhoven.\(^2^8\) During the session in Groningen, as part of the discussion on the collection of data on pedestrians and cyclists in the city centre, the subject of the utility and risks of nudging when this technique is used for purposes such as regulating traffic flows was addressed. The participants expressed an interest in knowing more about nudging and its impact on visitors in the city centre. In Eindhoven the meeting was focused on the collection of data on visitors to Stratumseind, with an emphasis on increasing safety in the public space. The participants raised a number of concerns, including the transparency of data collection and nudging, the legitimacy of nudging and the controllability of data-driven nudging.

The result of the sessions in Groningen and Eindhoven, along with a similar meeting in Amsterdam\(^2^9\), was the identification of four key issues and questions.\(^3^0\) These four primary points of discussion are given below, with commentary.

1. **There is a conflict between privacy and technology. What can we do? What should we do? What should we want? (and not want)**

   As this essay makes clear, we can do a lot, although we must be careful to not overestimate the strength of AI technology, which for the time being will remain at the narrow intelligence stage. What we **should** do is not entirely clear, because the developments in legislation are, by definition, always playing catch-up with developments in the technology.

2. **There is (a) a conflict between companies and individuals (this referring to the interests of private commercial parties), and (b) a conflict between the individual and the collective, public interest.**

   The conflict between companies and the individual was addressed above at section 4.4 (in the discussion of the cautionary *Guardian* article). In regard to the conflict between the individual and collective interests, there is a real risk of the data-driven society running roughshod over the interests of the individual.

3. **The degree of transparency and interpretability of data use and data collection in the public space.**

   Transparency and interpretability are (to a certain degree) technologically achievable, but this must be achieved with a consideration of the transparency of the rules and regulations enforced.
4. What is the role of the government within the smart city?

This question is a central one to this essay. In a general societal context, Helbing and his colleagues are thinking in the right direction. They make ten recommendations for leading the technological and social revolution down the right path. These recommendations are, in a more general sense, in keeping with the results of the World Cafés and the issues outlined here, and relate to (in part): greater transparency to increase trust, room for social and economic diversity, decentralization of AI systems, and promoting equitable actions in the digital world. For the purposes of this essay, it is important to emphasize that the recommendations by Helbing and his colleagues came about thanks to an interdisciplinary consideration of our society and the rise of technology in it. Taken in conjunction with the results of the World Cafés and the four key issues described above, the article makes the importance of bringing an interdisciplinary approach to bear on the AI revolution and creating a world of appropriate data use abundantly clear.

5.3 Shortage of professionals

In a blog post entitled ‘AI en de Nederlandse Belangen’ (‘AI and Dutch Interests’), Jurgen Oppel and Aaron Arends write: ‘Despite warnings from multiple Dutch experts in recent months of a “brain drain” in the field of artificial intelligence, it seems that there is still a failure to fully appreciate that our economic interests are at stake. The Netherlands has a good starting position, but is just about to miss the starting gun.’31 AI research in the Netherlands faces pressure from what is being called a brain drain.32 More attention to AI research as a driver of innovation is needed. To facilitate healthy innovation, attention to both technical and social innovation must go hand-in-hand. In September 2019, two AI initiatives worth mentioning were launched in the Dutch province of Noord-Brabant. Eindhoven University of Technology launched its ‘Eindhoven Artificial Intelligence Systems Institute’, to focus primarily on engineering (engineering AI applications in the automotive sector, high-tech systems, etc.). Meanwhile, Tilburg University launched its initiative ‘AI for Society’, and is now offering the first accredited university programme in AI in the province, Cognitive Science & AI (CSAI) in the department of the same name.
This programme combines technological and social aspects. And at the Jheronimus Academy of Data Science (JADS) in ’s-Hertogenbosch, a partnership between the universities in Eindhoven and Tilburg, both of these approaches come together in a shared research and teaching programme. The new teaching programmes are set to produce a large number of AI professionals. These unique initiatives in Noord-Brabant should be replicated in the rest of the country.

5.4 Certification
 Calls for regulation of the tech giants are growing steadily. On 25 July of this year, the US Department of Justice launched a broad antitrust investigation of several major tech companies. Independently of the commercial interests of major companies, machine learning algorithms are a source of problems. Even if a company or government institution applies machine learning to data from the public space on only a limited scale, there are myriad reasons why this can be problematic. One of these is that every machine learning algorithm has a bias, a sort of implicit assumption of the way in which various observations should be compared. As a result of the bias, the algorithm can generalize and identify examples that it has not encountered before. On average, the bias is an advantage (because it results in better recognition performance), but it can have a negative impact on an individual. One extremely simplified example of such a bias is the way in which machine learning algorithms handle certain personal characteristics like age and salary. Some machine learning algorithms, such as the ‘decision tree’ type of algorithm, consider every characteristic independently, while others, such as the ‘nearest neighbour’ type (described above) make an implicit link between every data point.

On average, one of these algorithms may perform better than the other, even though for individual cases it may be that one type addresses the individual’s situation better than the other. The recognition of such subtle biases in machine learning algorithms and verification of potential methodological errors in training or operationalizing machine learning algorithms requires a sound screening process. This screening should be carried out by an independent certification organization set up by the government. Individuals who feel disadvantaged by a decision resulting from the use of a machine learning algorithm should be able to apply to this organization for a review of the decision.

More generally, the organization could review the algorithm for certain statistical and ethical criteria to be further defined.

The four observations above offer the government a framework for building concrete policy. Our survey of the AI technology and observations allows us to answer our central question. To repeat, the question is: What do we need to do to avoid data use going wrong and bad applications of AI algorithms in our society? To avoid data use going wrong and bad applications of AI algorithms in our society, we need to make sure that...

• politicians and policymakers understand the potential and limitations of AI. It is particularly important to understand that the ‘oil’ on which AI runs is data. A basic knowledge of the statistical principles underlying equitable data collection is an absolutely essential requirement.
• it is understood that the AI revolution is not something that only people in tech need to worry about. Virtually all AI systems function in interaction with people, and this creates (to use the term coined by Helbing et al.) a delicate feedback system. This should underscore the great need for interdisciplinary professionals.
• the AI research infrastructure is attractive to talent, so the digital society can remain assured of a steady supply of AI professionals and AI innovations.
• research projects that will implement the approach as outlined in Helbing et al. are defined.
• a central certification body for the evaluation of AI algorithms is created.

We conclude this essay with five policy proposals that follow directly from our observations.
6. Policy proposals

The impact of machine learning and data use in public space on society will only increase in the coming years. In order to smooth the transition to a digital society dominated by machine learning, I would like to put forward five policy proposals derived from the observations in chapter 5.

1. Give politicians and policymakers regular training on the subject of data use and the emerging developments in AI in relation to social and legal aspects.

The speed of the technological developments brings new data sources and new technology at a steady rate. One of the latest, for example, is the ability to create synthetic images and videos that are virtually indistinguishable from the real thing. It is impossible for politicians and policymakers to keep sight of the total picture, including the potential and risks of new data sources and new technology. The additional training that this essay recommends should be an update to an already relatively high level of knowledge, and has two purposes: (1) to enable politicians and policymakers to communicate with the market on an equal footing, as an expert partner, and (2) to enable them to direct the processes that will lead to the technological innovation the government needs. With a basic knowledge of machine learning, a brief training will be enough to provide a clear picture without having to go into too many technical details. Likewise, for the social and legal aspects of data use and AI technology, the speed of the technological developments means that perspectives shift and understanding changes. Here, too, there should be periodic training for legislators, politicians and policymakers to enable them to follow the rapid developments and act appropriately. A combined training on technological, social and legal aspects could take the form of a one-day total training package given once per year.

2. The digital society will require professionals who are equally at home on both the technological side and the human and social side of AI. Invest in the programmes that will produce them now to eliminate the already existing shortage in such professionals.

The AI revolution is both a technological and a social/cultural revolution. The technical potential of AI has an impact on human behaviour and is influencing the way in which society functions. Professionals are needed in every domain in which technology and man come together, and so this means the full gamut including the social domain, safety and mobility.

3. Try to stop the brain drain of researchers by improving conditions for AI researchers.

In order to keep playing an important role in the technological developments, recruiting and keeping good researchers is a must. Boosting the Netherlands’ draw as an AI knowledge country for researchers will generate more innovation, better education and more control in steering the digital society.

4. Invest in research to facilitate the transition to an equitable digital society.

Investments in AI research into applications for data used in the public space will facilitate a successful transition. Seeking out connections with existing initiatives such as the Stratumseind project in Eindhoven would seem to be a prudent choice.

5. An independent body to screen AI technology for the aforesaid statistical principles will prevent abuses. As machine learning algorithms are continually upgraded, periodic review is needed. The government should make review mandatory in order to minimize the chance of abuses.

Such as supervisory body should be set up to ensure transparency of data use at the national or European level. Recently, the Australian government announced plans to set up a governmental organization for the screening of the algorithms used by tech giants Facebook and Google. This organization is being placed under the Australian Competition and Consumer Commission. The use of algorithms and data in the public space should always be screened in the same way, whether being used by tech giants or any other party.
About the author

Prof. dr. Eric Postma is professor Artificial Intelligence at Tilburg University and Jheronimus Academy of Data Science (JADS) in 's-Hertogenbosch.
Data makes the world go round

Proposal for research into three policy instruments designed to strengthen (digital) autonomy

Theo Veltman and Rob van Kranenburg

Foreword
By Peter van Hoesel

A total of five essays were planned as part of the 'Appropriate data use in public space' project. Following the meeting of the supervisory group in mid-July, where a broad outline of all the essays requested was presented, one aspect was found to be missing: the question of security and the individual in the digital sphere, both today and in the future. Although the original intention was not for the leaders of the supervisory committee to write an article themselves, after consulting with the Ministry of the Interior and Kingdom Relations, who commissioned the work, a change of plan was agreed. Just before the final meeting of the supervisory committee, Rob van Kranenburg and Theo Veltman submitted a sixth essay, which the committee was pleased to receive and review.

The committee views this sixth essay as a good addition to the collection, not just due to its content but also because it can be read as a genuine appeal from within the professional field – from the ‘front line’, written by those who work in the sometimes messy field of real-world practice. This contribution is not intended to be a scientific analysis; indeed, it is rooted in real-life practice. But this is exactly why it complements the five other essays so nicely: the other five essays have all been written as scientific analyses, with references to practice included where necessary.

Theo and Rob know each other from the world of innovation in and around Amsterdam. They share a passion when it comes to promoting the freedom of the individual to be digitally mobile, unmonitored and free of unsolicited influences, in buildings and in the public space. At the same time, they believe that everyone should also be able to derive the maximum benefit from digital technology. This requires the right balance between security, on the one hand, and convenience, utility and pleasure on the other. That balance is the subject of this article.

Rob and Theo both have a broad background and experience when it comes to people, technology and organizations. Both have made enough real-life errors (or near-errors) over the course of their careers to have learned a great deal. They are familiar with different aspects of digitization, from the perspective of different roles. Their knowledge, experience and insights are complementary.

Peter van Hoesel
Professor Emeritus in Public Administration, Erasmus University Rotterdam
Chair of the Supervisory Committee
Summary

Much has been written about the subject of risk in the digital realm in recent times. Occasionally, it is not easy to discern any coherence, let alone concrete, tangible recommendations. However, this essay makes three clear recommendations regarding policy measures. Before these are presented, a contextualizing summary is provided with respect to data – digital platform – people’s preferences and societal developments.

The key question of this essay is by what means (digital) autonomy can be safeguarded in a world where data – and therefore knowledge, money and power – are becoming increasingly concentrated. Ever more personal data is being collected through the internet and in the public space. As a result, it is possible to acquire a great deal of information about individuals and to influence them, unnoticed, in their preferences and choices. Notwithstanding this, the protection of personal data is a fundamental right in the Netherlands (constitution, article 10, paragraph 1) and in the EU.

The key question of how to safeguard (digital) autonomy has no simple answer. Like the Rathenau Institute, the authors believe that a change of mindset is required, and a shift towards a world in which values play an increasingly central role. However, this is a long-term process. And the time to act is now. It is necessary to act ‘now’ in order to continue to make the most of the benefits of technology without jeopardizing individual autonomy. It would be wise to limit the monopoly position of large organizations, the digital platforms, and to strengthen the position of the individual, while this is still possible. This will require a concerted effort, including at the international level. In this article, the authors recommend three measures to strengthen individuals’ control over their own personal data and to prevent the improper use of data by organizations that collect data in the public space of the city, namely:

1. The creation of a national, independent and open digital trust infrastructure in order to identify, authenticate and authorize (the collection of) personal data, accompanied by the required governance structures, based on existing building blocks such as Tippiq and Irma. The white paper (Dunnen et al., 2019) can serve as a starting point.

2. An investigation of the options and feasibility of realizing embedded hardware: a ‘data kill switch’ in smart devices such as sensors, with the collection of personal data set to ‘off’ as standard. By default, it would only be possible to collect data that is anonymized and which has been identified as necessary to preserve public order and security and manage the city’s public space.

3. An investigation of the options for and utility of setting up a ‘data safety label’ system overseen by a democratically established independent institute with a legal mandate to actively prevent and, if necessary, act against data theft and misuse, in order to supplement the General Data Protection Regulation (GDPR).
1. Data: age-old source of knowledge, power and money

In their 1972 interpretation of *Money Makes the World Go Around*, Liza Minnelli et al., in their own inimitable way, illustrated the importance and allure of money and the influence that it brings. Not much has changed since. The financial crisis that hit a few years ago was caused by, among other things, greed. See, for example, the documentaries *The Inside Job* (2011), *Enron: the smartest guys in the room* (2005) and *Capitalism: A Love Story* (2009) by Michael Moore. This innate human thirst for money and power inevitably leads to speculative bubbles – from tulips and real estate to digital wonder kids. Such hype cycles are fed by the knowledge that a small number of people have about what is really going on, added to the belief in the minds of many that they must get as close as they can to the front of the queue. ‘Knowledge is power’.

Data brings enormous opportunities when it comes to the acquisition of knowledge – and, therefore, of power and money. Using data, you can manipulate individuals and groups. However, this depends on which data you collect and use, and on which criteria you do this; it also depends on how you analyse that data, which assumptions, rules and criteria you apply, and how you interpret the results. And last but not least, it depends on how often and through which channels you present that data. If you use data selectively and repeat your message often enough, stories and suggestions can arise that have a major influence on the preferences and choices of people and organizations.

Data is more than hype. Access to accurate data is crucial and affects social and economic development; this has been true for many centuries. But the role and use of data are certainly changing. Nowadays, people want to be located as close as possible to digital nodes. This gives them a split-second advantage when it comes to interpreting trading data from financial markets. A small group a making millions. To summarize:

1. Data = the source of all knowledge, wisdom and deception
2. Data = the basis of earnings models

1.1 Data: also a source of deception, with limited protections in the law and from the government

Data has long been used to earn money and concentrate power. What is new in 2019 is that so much more data is available, and much more can be done with that data, while analysis software can process data ever more easily. The development of the internet, the smartphone, ever more digital services (including apps) and the mania surrounding big data and artificial intelligence (AI) have, together, produced a disorganized mass of information, the quality of which is not always clear.

Female Facebook users aged 25 and over, who, thanks to the interpretation of their browsing behaviour, were known to be considering having children, were shown anti-vaccination propaganda on Facebook. These targeted ads, which were viewed millions of times, were paid for by activists who wanted to discourage them from having their children vaccinating against measles. Zuckerberg’s platform was all too willing to help these actors to achieve their objective, just as it was willing to circulate fake news from Russia during the US elections. From: Thomas, Casper; (2019)

Data can be used to manipulate the news, commerce, a person and their friends or an organization. Manipulation, deception and fraud are much more common than many people think.

People and organizations are now being threatened and undermined using deep fakes. There are fake dating sites that convert human loneliness into a stream of revenue. Fake news is deliberately disseminated in an effort to manipulate people for financial, social and/or political gain. On social media, it is practically impossible to distinguish real people from ‘digital people’ who are controlled by a company. Sometimes these have millions of followers, which means they have a huge impact. Hackers, whether or not they have criminal intentions, play a very specific role through their constant attacks and threats using ransomware and bots, for example. Their activity is unrelenting.
Every possible resource is mobilized, sometimes including yours: ‘IoTroop is a powerful Internet of Things (IoT) botnet comprised primarily of compromised home routers, TVs, DVRs, and IP cameras’ (Check Point Research, 2019).

A digital war is being waged – not with bullets and bombs, but using manipulated data. And it is not only being waged between states. In a certain sense, there is also a digital war going on between the digital platforms and ‘the consumer’; between big capital, small and medium-sized enterprises and individuals. They are all competing, to a greater or lesser degree, to retain their own (digital) autonomy and position, and to win space for development and room for manoeuvre.

The weapons they use include disinformation, disintegration and despotism. The Council for Public Administration (ROB) recognizes these weapons (the ‘digital ordeals’) and presents a number of options for safeguards (see table).14

### Three digital ordeals

- **Disinformation**: the fear is that as a result of digitization, disinformation can be produced and disseminated more quickly and in a more targeted manner. This includes fake news, conspiracy theories and propaganda.

- **Disintegration**: the risk is that digitization creates parallel worlds that are no longer connected to each other due to phenomena such as filter bubbles, echo chambers and ‘digital pillarization’.

- **Desspotism**: there is a risk that, as a result of digitization, citizens can be influenced more easily without even being aware of this, such as through secret micro targeting by companies or (foreign) politicians.

Standing up for the search for truth

- Enhance confidence in institutions by setting a good example when it comes to the search for truth
- Enhance citizens’ resilience vis-à-vis disinformation by enhancing their critical skills.
- Prevent disintegration by building platforms for the exchange of thoughts and opinions with others in a democratic way.
- Fight despotism by ensuring countervailing forces. Break the fear of truth by continuing to pursue a dialogue about the nature of truth.

In its report ‘Searching for truth’, the Council for Public Administration (ROB) provides an inspiring philosophical exploration of what ‘truth’ is, or may be, and what the search for truth actually means. There are many versions of truth, just as there are many lenses through which individuals may view the world. This is precisely why individuals must be able to see various opinions and to assess the underlying data on aspects such as accuracy, completeness and timeliness. In all cases, it must be clear how that data has been processed (for example, which algorithm was used), when it was collected, for what purpose and by whom. It is unfortunate that the ROB did not mention this basic requirement explicitly, since it is essential to the ability to assess information, opinions and knowledge and evaluate their truthfulness. And all the more so because compliance with this requirement would benefit the digital resilience of citizens who want to do this. We believe that this would be more effective than pursuing a ‘dialogue about truth’.

In all cases, it must be clear **how that data has been processed (for example, which algorithm was used), when it was collected, for what purpose and by whom**

After all, how realistic is it to conduct such a dialogue? Do citizens really want – or have time for – this dialogue in order to become critical citizens? According to Kahneman, people tend to take the path of least resistance and to prioritize the short term; people find it difficult to think about risks and discomfort over the longer term. And all those digital services are actually rather convenient...

The ROB also recommends focusing on enhancing trust in institutions, and this starts with setting a good example.

This is certainly an area where work is needed. However, more needs to be done, including effective monitoring of whether institutions are delivering on their promises with respect to working methods. One example here are the checks and balances promised in relation to the ‘security versus privacy’ dilemma, following
the discussions around the Computer Criminality III legislative proposals and the modernization of the Intelligence and Security Services Act 2017 (WIV).

The ‘Search for Truth’ report illustrates the ways in which data (information) can be misused and the risk that this implies for ‘truth’ and our freedom of opinion and free movement: our (digital) autonomy. The recommendations seem rather vague, while the legislative and regulatory specifics are already running into their limits.

One example here is the effect of the General Data Protection Regulation (GDPR). The Regulation specifies the following: ‘Natural persons should be made aware of risks, rules, safeguards and rights in relation to the processing of their personal data and how to exercise their rights in relation to the processing.’ (excerpt Art. 39.5). Yet this requirement can still be satisfied even if the explanation is tucked away in the FAQs or General Terms and Conditions. If a person has to deliberately search for this information, the question is whether the spirit of the law is also being observed, as well as the letter of the law. Can people easily find out what their rights are, and understand these rights? The GDPR also identifies six legal grounds for the collection of personal data. One of these is ‘the public interest’, and another is the ‘consent of the individual’. Smart devices such as sensors, cameras and trackers in public spaces are used to collect a great deal of data that is not necessary in ‘the public interest’. How can we ensure that this only happens with the consent of the individuals concerned?

There are other examples, such as the use of cookies. Many companies make it difficult to change cookie settings. Doing this takes up valuable time, irrespective of whether the user understands what those cookies do. This means that such arrangements are unclear and inconvenient, even through the GDPR states that such features must be presented in a clear and simple way, using simple language. Another example is the PSD2, the new European law (directive) relating to payment transactions by consumers and businesses. This also poses risks for the autonomy of the individual, despite its good intentions. Furthermore, not all entrepreneurs are equally honest when it comes to safeguarding the privacy of the individual and the regulations that apply, and enforcement is difficult, not least because there are too few experts available to do this job. The difficulty of safeguarding autonomy in the public space is exacerbated because smart devices are becoming ever smaller as a result of technological advances, and they may therefore be more difficult to identify. Digitization organizations also operate in the virtual realm, where many borders, including national borders, are invisible. This is a further issue when it comes to enforcement.

Effective laws and regulations and the associated enforcement are necessary for the protection and promotion of ‘the public interest’, which includes, for example, security of identity and the autonomy of individuals, groups and organizations. Legislation is an important tool of society and government – a government that is still searching for its role in the digital realm.

So what can we do to protect and preserve a local imagined community, and mobilise it as a force for good?

1. Create a safe and secure urban environment. If we fail to ensure that everyone feels equally safe and secure, we cannot expect all our citizens to contribute to a strong civic democracy with a shared identity. A city that is strong and resilient.
2. Protect the public space. The city belongs to everyone. Our neighbourhoods, squares and streets are public spaces that we must protect so that all citizens are assured of a safe and peaceful life. It is their space.

From: speech by the Mayor of Amsterdam, NYC: Columbia University d.d. 9 April 2019
The role of government must be clarified, and questions answered about where the boundaries are in the collection and use of personal data by the government for the purposes of security and public order, managing the public space and providing digital services for citizens. This relates to questions such as: how far the government can go to protect our digital freedom; how far it can go to safeguard and maintain public order and safety; and how far it can go to rationalize its decisions using data. It is not always possible to answer these questions unambiguously. Certain dilemmas crop up again and again – operational, tactical and administrative dilemmas, such as the tension between collecting data for security purposes and the right to privacy. The collection of data implies a risk to everyone’s autonomy. Personal data is a source of knowledge about that person’s preferences and those of his or her family.

Civil servant: Using big data techniques such as machine learning, it is possible to calculate the likelihood (and therefore the risk) of a particular outcome. For example, the likelihood of a pupil leaving school early, or of a person being taken into care. One idea that is appealing to many of my colleagues is identifying a population group where there is an elevated risk and then designing policy based on that. But in reality, that means data profiling, which is not permitted. What’s more, intervening in individual cases ultimately involves identifying a particular individual, and as a government we would not be able to account for this. There is no (democratically established) rule based on a fixed criterion; rather, it would be the algorithm that teaches itself, constantly establishing new criteria based on its processing of new data. An additional problem is that interventions are not described as such prior to carrying out the research and delivering the data that would be required as part of this process; on the contrary, they would be covered by the umbrella concept of ‘Policy research’. This means that the data request cannot be rejected on the grounds of unlawful use. Is that permitted?


1.2 Data: a basis for revenue models
Personal data has a monetary value. Using personal data, a detailed profile can be built up of a particular person’s cultural, economic, political, religious, sexual, social preferences and proclivities. This profile makes it possible to influence our opinions, decisions and behaviour. This means that the most valuable data relates to what an individual does or wants to do in the digital realm, and what he or she is interested in. There is a good reason why people buy and sell data on every movement that is made online. Almost no opportunity to collect personal data goes unmissed. It is collected in many different places: by smart devices in the public space, by search engines and online games, and by apps, even when those apps are not activated. Every time someone uses a digital platform, he or she gives away data by performing searches and by posting photos or messages, and sometimes this trail of data remains unseen – notwithstanding the legislation (GDPR; excerpt ground 39-art. 5). Digital platforms such as Facebook (which also owns Whatsapp and Instagram), Google and Yahoo are big traders in data.

Researchers from the International Computer Science Institute recently tested 24,000 apps that run on Google Android. Seventy percent of them track users’ movements constantly and exchange data with other apps, they discovered. The researchers informed Google of this almost six months ago, but have not received an answer.

From: Thomas, Casper; (2019)
Appropriate use of data in public space
2. The rich and dominant digital platforms: a risk to (digital) autonomy

We all prefer to make life easy for ourselves. When a message about cookies pops up on our screen, how many of us click ‘OK’ without making a conscious decision, and thus consent to giving away our personal data? And how often do we use a WiFi network in a café or on the street, if necessary logging in via Facebook, without thinking about who might be offering this service and under what conditions? In the United States, apparently about 76% of people do the latter.24 What is more, we can hardly live without platforms like Facebook any more. Such services now have a high ‘social value’: ‘I never know what you’re doing any more since you left Facebook.’ For example, when primary schools use Facebook to share information about events and news about the school or to create groups within classes, parents with children at that school have little choice about whether to be on Facebook or not. If all your colleagues use Whatsapp to share information about projects, you need to use it too. Otherwise you will simply miss out on too much: you may become a ‘digital outcast’.

The social value of such platforms creates lock-in. Search engines, too, are now a daily tool used by more and more people, which means they are learning ever more about their users. If you also use the platform’s other services (Google’s Gmail, for instance), lock-in quickly becomes an issue. Their service contains every e-mail address and every e-mail in your archive. This type of lock-in makes it difficult for new providers to gain a foothold in the market at the expense of today’s dominant platforms. As a result, little use is made of alternative platforms that do not store personal data, such as ‘Signal’ (an alternative to Whatsapp) or the search engine ‘Duck Duck Go’ (an alternative to Google and Yahoo).25 A vicious circle is established, whereby we make ourselves the prisoners of a small number of platforms which have come to dominate our digitized society in the year 2019.

The dominance of these platforms is not without risks. The algorithms of a platform can ensure that individuals end up in an ‘information tunnel’ populated only by like-minded people, and are therefore seldom exposed to opinions that differ from their own or to arguments that differ from what they already believe. This effect reinforces our natural tendency towards group behaviour. It also promotes the chances of successful influencing and personalized advertisements (micro targeting). It adds to the turnover of digital platforms. It also encourages people to see only one side of a particular debate, reinforcing the ‘us-and-them’ mentality, with rising potential for conflicts and radicalization as a consequence. Compared to micro targeting, traditional advertising is like using a hail of bullets to kill a mosquito.

American researchers have found that they can use mathematical formulas to segment huge populations into thousands of subgroups according to defining characteristics like religion and political beliefs or taste in TV shows and music. Other algorithms can determine those groups’ hot-button issues and identify ‘followers’ among them, pinpointing those most susceptible to suggestion. Propagandists can then manually craft messages to influence them, deploying covert provocateurs, either humans or automated computer programs known as bots, in hopes of altering their behavior.

From: Calabresi (2017); Time

The risks of these effects are becoming increasingly evident. The Mobile Ecosystem Forum’s 4th Global Consumer Trust Report (2018)26 concludes from various studies that 57 percent of mobile device users see the collection of their personal data by data service providers as a risk. Some 66 percent indicate that they themselves, or someone they know, has suffered data-related damage. And 63 percent say they would like to be able to exercise control over their own personal data. Uncertainty is also increasing. People do not know which information is real and which is not.27 It is unclear which of their data is being retained, and what happens to that data. Nevertheless, we all continue to use digital services, including platforms. This is the ‘privacy paradox’. Convenience and Short-term Thinking Prevail (Kahneman, 2016).
Professional from a large energy company: ‘I feel extremely visible: check ins on Facebook, everything you post on Twitter, Google, that knows through your phone every step you take pretty much, everything you post using Gmail. I am pretty sure everything is scanned and collected and aggregated.’
From: Research by Shazade et al. (2018)

Through our behaviour, we are enriching the digital platforms and adding to their data dominance. In its negative form, this can reinforce despotism (Rob, 2019). In its positive form, it helps us to generate new knowledge and live together in freedom. To benefit from the latter, however, we must prevent the dominance of the few. This is some challenge.

Digital platforms earn a great deal of money from personal data and through services that are connected visibly and invisibly, including advertisements. This money buys them power and influence, helping them to maintain and strengthen their position. In addition, they are also in a position to invest in the development of new and sometimes dangerous technologies and in methods of influencing the behaviour of individuals and groups even more effectively and imperceptibly. In many areas, digital platforms are becoming ever more influential. Increasingly, it seems that digitization is also changing the rules of the economy: ‘the winner takes all’: ‘As a result, the internet has become increasingly concentrated, less open and growingly hostile to innovation and entrepreneurship’.

3. Increasing pressure on (digital) autonomy: a ‘tipping point’

The development of AI is magnifying both the risks and the benefits of digitization. Within a few years, a combination of AI, big data and quantum computing will create many new opportunities for both positive uses and for abuse. But even today, so much is already possible...

The ‘Traffic Data Database’ project keeps a record of the intensity of road traffic on various types of roads (larger and smaller roads). Previously, this was always done using cameras which counted vehicles, but now the process is becoming increasingly automated because cars are constantly connected to the internet and emit a constant stream of data. This method is cheaper and provides more useful data that can be used to organize transport in a smarter way. What is more, all the data is then aggregated and made available as open data.

Dilemma: during a recent hackathon, students showed how it was possible to find out exactly who had been sitting in which vehicle and when they had travelled by combining various open data sets; this took them four hours. They did this for fun for their boss. But if they can do it, anyone can do it.

These days, there is more AI around than many people think. Google search results, recommendations made by webstores such as Amazon and Bol.com and the scanning of the license plates of parked cars by the City of Amsterdam (to see who has paid) – all these are based on AI. And the use of AI is becoming ever more widespread. Take the Johan Cruijff Arena, a stadium in Amsterdam, for example. There, AI is being developed to optimize visitors’ ‘customer journey’. The City of Rotterdam is trialling a different use for AI: a camera that can distinguish objects in order to keep the city cleaner and prevent antisocial behaviour at particular locations. The Tax Authority uses AI for automatic processing and to assess the risk of fraud.
In China, AI is being used to scan social media for messages that may indicate that a person is suicidal. Volunteers are then asked to contact the users who sent these messages.35

Today, in 2019, we do not yet have AI that can influence the drivers of our behaviour and emotions at the individual level. Our behaviour as individuals still seems to be at the heart of the way society thinks and functions. However, there is already talk of bots that can analyse our behaviour in order to determine the best moment to influence us.36 Nobody knows how quickly AI will progress. But experts are already discussing about how fast and how great its impact may be.37 What we can be sure of is that AI will be applied in ever more radical ways. We can assume that AI will become better at operating autonomously, and continue to learn and develop independently. As AI matures and acquires the capacity for deep learning38, it will become increasingly difficult to monitor the quality of the data sources and algorithms that it uses. After all, these would be constantly changing and evolving of their own accord. If and when this happens, it will become difficult to assess the quality of the analysis and the decisions being made by AI as a result.

What we can be sure of is that AI will be applied in ever more radical ways

We are at a tipping point. In the next few years, we still have time to take effective measures to prevent large organizations and technology from dominating our lives to an increasing extent. As Bianca Wylie says: ‘Dystopia doesn’t happen with a big bang. It’s little stuff happening over and over again, until we find ourselves in the positions thinking: Is this really happening? Are we now depending on a corporation to deliver our basic needs like housing, education and safety?’39 Zuboff (2019) confirms this view. Step by step, data scientists are learning how emotions and behaviour can be influenced – ever more effectively and ever more imperceptibly. Before the technology for this has been fully developed and brought into practice, we need to build a foundation as part of our societal system – instruments that enable individuals to determine and safeguard their own (digital) autonomy. But the time to take these steps is now.
4. National and international initiatives in support of (digital) autonomy

Various initiatives have been undertaken that support the (digital) autonomy of the individual. Various national and international players are working on the development of a personal data safe, as in the European DECODE project for example.40 There are also many initiatives to establish (behavioural) rules for the digital environment. Examples include:

**International**
- Cities4digital rights, an initiative of Barcelona, New York and Amsterdam, Nov. 201841
- ‘The Institute for ethical AI and machine learning’ (UK), the eight principles for AI42
- Ethical principles of AI from the High-Level Expert Group on AI43
- The Rathenau Institute, which was commissioned by PACE (Parliamentary Assembly of the Council of Europe) to draft a set of human rights in the light of the digitizing world44

**Government of the Netherlands**
- Letter to Parliament, Vision for Data Management, 2019, which states, among other things, that citizens not only have the right to know why what data has been recorded about them by the government, but also that they have the right to use that data by sharing it with others, in a trusted way so that the citizen retains control46
- The initiative for the development of a Digital Government Agenda (2018)47

**Municipal level**
- VNG Realisatie, principles for the digital society, 201948
- Amsterdam and Eindhoven: the principles of the digital city, Jan 2017
- Utrecht and the DEDA instrument49
- ‘Economic Board for Amsterdam (region)’ with TADA, Nov 201850

Various initiatives by other parties
- ‘Taskforce for the Proper Use of Data’, Chamber of Commerce51
- ‘A vision for a shared digital Europe’ from a partnership between Kennisland, Centrum Cyfrowe and Commons Network52
- ‘Fact, Fair, Robust, Shared principles’53 of digital society (VSNU)
- ‘ICT Code of conduct. For the digital economy’ & the ‘Ethics code for Artificial Intelligence (AI)’ in ICT in the Netherlands (2019)54
- Manifesto (2018) for ‘Public Spaces’, partnership of organizations for a secure digital environment55

The following conclusions can be drawn: we lack (legal) standards for the digital society. Some quotes:
- In today’s digital society, the balance – between the individual and the collective interest, and between the private and the public spheres – is being lost There are two dimensions of digital freedom versus digital security which are related and where legal standards are lacking.
- New (social) protocols are needed and perhaps new institutions to oversee activities in a digital world. In this sense, the introduction of the GDPR in Europe was a good initial step from the perspective of the individual citizen and an example for the rest of the world. But more action is needed and there remains much to do.
- There needs to be a new frame of reference in which we integrate a more diverse range of perspectives: analogue and digital

From: Round Table Dialogue, Amsterdam, 20 May 2019

The Dutch government has taken measures to address the risks of the digital realm, including a digitization strategy.56 This has now led to, for example: the Dutch Cyber Security Agenda (NCSA), the Digital Connectivity Action Plan, Smart City NL Strategy, NL Digibeter: Agenda for Digital Government, NL DIGITAL: Government Data Agenda, the Action Plan for SMEs and the Action Plan for Digital Inclusion.57
In the EU, digital threats are being tackled proactively. In 2018, for example, a code of conduct for online disinformation was signed with some of the large digital platforms. The EU is also keen to tackle the (data) monopoly of large digital platforms, as is the Netherlands. A number of sizeable fines have already been imposed. Work is also underway on a range of standards and regulations for the digital sector. For instance, there is the rule that people must be able to choose which data about them websites can collect using cookies, and the rules around privacy (GDPR). However, as we have already discussed, this legislation is not yet watertight.

5. What is required in order to reinforce (digital) autonomy

It is important that an effective foundation is laid for securing (digital) autonomy in the years to come. That means: the freedom of individuals and groups to determine their own activities, opinions and preferences, of whichever kind, independently, and the freedom of individuals to have ‘secrets’. In addition to the fact that people need to take responsibility for their own choices, this requires European measures, with democratic legitimacy, to protect freedom. It requires changes to our ‘thinking and working’ that have broad support (Rathenau, 2018). This will take some time. In the short term, we need to work on meeting a number of important requirements, namely: strengthening citizens’ digital resilience; defining the role of government in the digital realm; and ensuring that individuals can exercise control over their own data.

5.1 Working to put people and values at centre stage

According to the Rathenau Institute, the government, the business community, knowledge institutions and social organizations need to work together to organize the digital society in an effective manner, based on people and values. That is a long-term process. It will require change from everyone: a shift from the prevailing focus on individual economic advantage and towards a focus on international and societal utility; away from ‘me’ and towards ‘we’; away from economic growth and towards greater well-being; away from decisions based on business cases that emphasize financial arguments and towards decisions based on societally relevant arguments. A digitization accord, as proposed by the Rathenau Institute (2017), could encourage the shift towards more action based on public values during the digital revolution. This approach would require the corporate sector to move away from business models based on financial gain and towards more people-oriented and sustainable business models, for instance. It would require government to put the interests of individuals and groups right at the heart of the democratic contract. It would also require room for experimentation with new ideas, with the risk that those ideas will not always lead to the results that we hope for: trial and error, over and again if necessary, will sometimes be the only way to learn. From individuals, it will require active participation. History shows us that only a limited group of people are likely to adopt such behaviour. Active participation is also becoming increasingly difficult due to increasing social complexity, expectations and the pressure to achieve and to ‘be someone’.

A digitization accord, as proposed by the Rathenau Institute (2017), could encourage the shift towards more action based on public values during the digital revolution.

All of this would involve major changes in our behaviour and mind set. Although there are some encouraging signs, the general trend seems different. According to the Netherlands Institute for Social Research (SCPB) (2019), Dutch citizens generally agree, more or less, on what it means to be Dutch, but of course there are some differences too. These differences get magnified by social media, and they are leading to the emergence of a polarized society. Increasing individualism and suspicion of democracy and politics are fuelling the phenomenon of ‘cocooning’. People seem to be increasingly interested in a limited circle of individuals around them, comprised of more or less like-minded people. They are less and less concerned with the well-being of others, and less and less interested in alternative perspectives and arguments.
This is fertile ground for tunnel vision, for polarization and for radicalization. The information tunnel effect is further exacerbated by declining certainty and the digital platforms with their algorithms. The ROB (2019) adds some nuance to this picture: it believes that public service broadcasters contribute to a shared picture of reality. However, online anonymity and the vast reach that the digital realm provides can promote polarization. There are now many examples of this, in the US, Egypt, New Zealand and the Netherlands, including recent ones.

The recommendations of the Rathenau Institute thus involve far more than can realistically be achieved in the short term. Changes in behaviour and mind set often require a long time to be achieved. And the larger and more diverse the group, the harder this is. This is before we even start to think about societal change. For example, it took decades before non-smoking became a more or less accepted norm or before climate-neutral thinking was more widely accepted. The creation, adoption and revision of legislation is also a lengthy business. Not only because of the social support and democratic legitimacy that is required, but also because of the caution that needs to be exercised. But this does not alter the fact that legislation and regulations need to be produced.

5.2 Requirements: digital resilience; a specific role for government; control over data

Like the ROB, we believe that the digital resilience of citizens is crucial. This starts with an awareness of the actual consequences – both now and in the future – of being insufficiently critical in one’s use of digital services, or of not looking critically at the extent to which digital communication corresponds with reality. And an awareness of how susceptible one is to influence and manipulation if one fails to do this. There is a chance that certain groups of people are not really interested in these risks: ‘The masses, in general, are addicted to marketing gibberish! Why? Well, it is much easier than figuring things out for yourself.’ But still, something has to be done. This is where the role of government comes in. This needs to be defined with greater clarity, along with the instruments that government has at its disposal and the freedom it has to collect and use data in the public interest where necessary.

Individuals also need to be able to access and use good-quality data. Access to reliable data – the basis for information, opinions and knowledge – is a basic right, just like freedom of expression. It must, at the very least, be possible to assess the quality of data and information. Information tunnels should be prevented, or identified and broken down. This combination would lead to a stronger position with respect to information. It would promote personal development, access to money and quality of life. Sound, reliable information for everyone on an equitable basis was once the promise of the internet (Berners-Lee; 2018). However, the fulfilment of that promise is increasingly being undermined by digital and social developments.

Finally, every individual must have right of ownership over their own personal data, so that everyone is able to determine who gets to use which data, for how long and for what purpose. The identity of those sending and collecting data must also be guaranteed, and the identity of the objects that individuals are talking about or that they want to grant or gain access to. In addition, we need guarantees that data is being used in accordance with the agreements made, and that corrections can be made where this is deemed necessary. Furthermore, every individual must be able to understand which data is being collected about him or her, by whom and for what purposes. And finally, there must be a guarantee that digital platforms and smart devices are not recording more personal data than is permitted by the law, or than has been authorized by the individual concerned or than is necessary for the safety of society and the city.

The protection of personal data is a fundamental right in the Netherlands (constitution, article 10, paragraph 1) and in the EU, while individual resilience has long been a political and social issue. It would therefore be wise to get to work as quickly as possible on creating a set of instruments that will help put the individual in charge when it comes to the collection and use of personal data relating to him or her and those around them.
6. Conclusion

The outline that we have sketched of our digitized world in 2019 leads us to the conclusion that we must now strive to put in place a number of basic guarantees in order to counteract the growing data monopoly of digital platforms and to reduce the associated risks to the (digital) autonomy of the individual. In light of our description, we envisage the following guarantees:

1. every individual must be able to exercise control over their own personal data: what is collected, when and by whom, who uses it, for what purpose and for how long. This includes guarantees of the identity of people and objects and of the quality of the data that is being exchanged. This will require national, independent and open trust infrastructure.
2. every individual must be able to choose which data can be collected about him or her in the public space, perhaps by means of a ‘data kill switch’.
3. people need to be encouraged to use and operate smart devices in accordance with the relevant legislation. This could be done, for example, by means of a ‘data safety label’ for smart devices, awarded by a democratically established independent institute with a legal mandate to actively prevent and, if necessary, act against data theft and misuse, over and above the GDPR.

Trust infrastructure offers facilities to securely exchange data between citizens and service providers, whereby the consent and use of personal data remains in the control of the citizen concerned. Such infrastructure would also provide core services such as identification, authentication, permissions and security. Public and private service providers could make use of this infrastructure for any services that require personal data. Trust infrastructure would act as a brake on the power and influence of digital platforms. It would enable organizations and individuals to exercise control over their own personal data. It would guarantee the legality, identity and authenticity of that data. The components of such trust infrastructure are already available. Existing initiatives for sharing data through trust infrastructure include MedMij (in the healthcare sector), Joindata (agricultural sector) and iShare (logistics sector). Examples of facilities for trust infrastructure are the Tippiq platform (open source, financed by Alliander), DigiD and Irma. But this is not the end of the story. The trust infrastructure will need to be further developed and completed.
Government, businesses and citizens are increasingly installing cameras and other sensors. Many of these smart devices record data in the public space concerning the people who are present there, either directly (for example with camera images) or indirectly (for example when counting vehicles). The number of smart devices in public spaces continues to grow. Local government could limit the use of smart devices using existing policy instruments, such as permits. However, enforcing this would be difficult. Another difficulty is the fact that smart devices are getting steadily smaller. It would be more effective to regulate the collection of personal data in a preventive manner. This could take the form of technological provisions and/or administrative provisions – a data kill switch and/or a mandatory data safety label, respectively.

Data kill switch
For the data kill switch, an agreement would be made with the producers of smart devices that they would include embedded hardware in their devices that enables the user to switch on the collection of personal data by the device (the default would be set to ‘off’) using an app that the user would install. This way, every individual who wants to could download the app and indicate whether they consent to data about them being collected. If so, the smart device could collect this data. If not, the smart device would only collect anonymized data that is deemed necessary, subject to democratic controls, to safeguard public order, local safety and management of the city’s public space.

Data safety certification label
A data safety label would indicate that a product meets specific reliability and safety requirements. Products would not be allowed to be sold without this label. Of course, this would be enforced by an independent institute, subject to democratic control and with a legal mandate that ensures allocation, monitoring and enforcement. It would be essential that this institution were subject to democratic scrutiny and management.

A legal prohibition of the theft or misuse of data, in addition to the GDPR, would give the institute an even better basis for enforcement and prosecution in cases of attempts to steal and/or misuse data. For this reason we would recommend the simultaneous development of the quality label, the institute and the accompanying legislation.

One precondition should be that the possibly longer lead time for the development of legislation should not act as a brake on the development and implementation of the quality label and the independent institute. Furthermore, while we recommend a national quality label and institute, at the same time the initiative should be taken at the European level for a European data safety label.

Both the data kill switch and the data safety label should be assessed for feasibility, practicality and effectiveness.

Acknowledgments
This essay has incorporated valuable comments and suggestions from a range of people, including: Jim Janssen (Maastricht); Frans Jorna (Almere); Jos Maessen (Rotterdam); Hans Nouwens (ext. adv.); Patrick Reijnen (Amsterdam); Wim Schimmel (CIO a.i.); Robert de Snoo (Taskforce for Reasonable Use of Data, Chamber of Commerce); Frank Vieveen (Rotterdam); Aantink Yeh (VNG Realisatie) and colleagues from the Amsterdam CTO Innovation Team (including Burcu, Nathalie, Tamas). Manon den Dunnen (police) provided input from her activities in relation to the trust infrastructure (recommendation 1). The white paper may serve as a stepping stone to further development and implementation. Marleen Stikker (Waag) and Mark van den Broek (Eindhoven) provided the inspiration for recommendation 3, among other things.
About the authors

Theo Veltman has been working in the field of organization, man and technology since 1979, currently at the Municipality of Amsterdam. His previous positions have included CTO Amstelveen, CIO Ministry and director of a number of agencies, as well as various other management and advisory positions in education, social security, industry and the public sector at the national and local levels (in most cases ad interim).

Rob van Kranenburg is the author of an influential article about RFID and the Internet of Things (published in 2007 by the Institute of Network Cultures at the Amsterdam University of Applied Sciences). He has worked at the University of Amsterdam, De Balie (a venue for contemporary arts, politics and culture), Doors of Perception (Flow 2003), and Waag Technology & Society. In 2009 he established the expert network #IoT Council (theinternetofthings.eu), and one year later, iotday.org, to stimulate a broader discussion in society. He has worked in the Coordinated Support Action NGI.eu FORWARD, the strategy group for the Next Generation Internet Programme.


Check Point Research (2019). Security Report 2019. Tel Aviv: Check Point Research: ‘In late January 2018, the ‘IoTroop’ botnet, discovered by Check Point researchers in October 2017, launched its first attack against the financial sector. IoTroop is a powerful internet of things (IoT) botnet comprised primarily of compromised home routers, TVs, DVRs, and IP cameras. The first attack used 13,000 IoT devices across 139 countries to target a financial organization with a DDoS attack, followed by two more attacks against similar targets within 48 hours.’ For example.


GDPR. (2016, 27 April), REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46 / EC. Brussels: EU Parliament and Council. Accessed August 2019 on https://www.privacy-regulation.eu/nl/index.htm. This is the ‘General Data Protection Regulation’ (GDPR), including article 5: the ‘principles relating to the processing of personal data’ and article 6: grounds for ‘lawfulness of processing’, namely: consent of the person, the public interest, vital interests of persons, representation of legitimate interests (e.g. those of a child), legal obligation and by agreement.

Croon, T., Boogaard van den, L. & Parie, J. (2019). Geef bankzaken niet te makkelijk prijs [Don’t disclose banking data too easily]. NRC.


35. Interview with Chris Kubeczka, CEO HypaSec, digital security researcher, in July 2019.


47. https://www.vnlrealisatie.nl/nieuws/uitvaart-principes-voor-de-digitale-samenleving


Urgent Upgrade Protect public values in our digitized society


Dunnen den, M., Kronenburg, B. & Zalm van der, E. (2019). Een publiek vertrouwensinfrastructuur voor persoonlijke data [Public trust infrastructure for personal data]. Amsterdam. This is a white paper on this subject which a number of public institutions are working on together.

The term was coined by Faith Popcorn in 1981.

Information tunnel = echo chamber and information filter; ROB, 2019
VII Around 85% of Google’s revenue comes from advertising; for Facebook this figure is around 98%. From: Hueck, H.; Financieel Dagblad, 25 June 2019

References

By democratic control, we do not mean direct accountability under the model of representative democracy, but a direct form of democratic influence, other than referendum. Also see Mouk, Yascha; journalofdemocracy.org; 2018; interview with Van Zutphen, Reinier; publiekdenken.nl; 2019)
Overview of policy recommendations from the essays

Reuse of data in smart cities
Legal and ethical frameworks for big data in the public arena
Prof. dr. mr. ir. Bart Custers

1. Local authorities in smart cities must commit themselves to the appropriate reuse of data through public-private partnerships.
2. Local authorities must actively involve citizens with the factors to be considered in relation to smart cities.
3. In the case of smart cities, local authorities should ensure transparency in relation to data and factors under consideration.
4. Local authorities should use pilot schemes to further develop smart cities.

The digital transformation, behavioural influencing, and the government
The slippery slope from nudging to überveillance
Prof. dr. Wijnand IJsselsteijn

1. If the government wishes to use technology for the purposes of behavioural change in the public domain, this must be done transparently, on the basis of democratic decision-making, ethical review and scientific evidence, and with the explicit involvement and consent of the parties that this technology is going to affect either directly or indirectly. Here the government must at all times communicate proactively, openly, comprehensively and honestly about the digital transition in the public domain, or the plans in that direction, at all levels – from national and provincial to municipal and at the level of the community.
2. Engage all the parties that are affected by digital behavioural influencing in the public space, whether directly or indirectly, through a process of participatory design. Start this at an early stage and keep it going throughout the process. Make the extra effort to give the most vulnerable groups in society an active voice in this process. Also consider ethical values in the design. The government has the primary responsibility for organizing this process, together with its social partners.
3. The government must increase its digital competences and as part of this inform itself properly of the potential and limitations of the most recent forms of digital technology. A technologically naive government is easy prey for bad actors, a bad partner for industry, and a failure as a guardian of the public interest.
4. Data ownership and the privacy of individuals in the public space must be respected and protected. The government must eschew unproven, unsafe or less than robust technology. Unintended, unforeseen or illegal use of the technology and the data must be monitored and regulated at all times.
5. The government must provide for a constructive/critical opposing voice from society. This all must come along with major investments in the digital literacy of the public, including more attention to the latest technological advancements and how these will affect the fundamental ethical values within our digital society. The reinforcement of technological citizenship can take on many forms, such as debates, courses, living lab demonstrations, critical cultural projects or speculative designs.
6. The government must organize adequate and independent critical monitoring and supervision, and embed this in the law. Examples include an ongoing process of encouraging and organizing the values debate, discussing relevant ethical and legal precedents, and review of legal, ethical and governance frameworks through elected representation at all levels (municipal and provincial councils, the Upper and Lower Houses of Parliament). It is also important to keep civil society and relevant advisory bodies engaged at all times and to augment the role and position of regulatory authorities. This also means overseeing coordination across enforcement domains.
Public governance of experimental data & algorithms
Recommendations for a national algorithm register and reporting framework
Prof. dr. Gerd Kortuem

1. Set up an Algorithm Reporting Initiative with the aim of documenting and tracking the societal value and risks of algorithm projects across the Netherlands.
2. Develop a National Algorithm Register with the aim of enabling effective comparison and assessment of data and algorithm initiatives across the Netherlands.
3. Develop an Algorithm Forum with the aim of enabling public stakeholders to debate, critique and contest the operation, use and outcomes of data and algorithm initiatives.
4. Establish a Data and Algorithm Institute to drive the development of national approaches, tools, infrastructures and standards for ethical and accountable use of algorithms.
5. Develop a National Data and Algorithm Skills Agenda to ensure that all stakeholders (from citizens to organisations) are sufficiently skilled to participate in the new data and algorithm economy.

The need for a digital environmental strategy
From principles to practice
Prof. dr. Liesbet van Zoonen and dr. Jiska Engelbert

1. Municipalities need to develop a digital environmental strategy in co-creation with members of the public, civil servants, the private sector and other stakeholders, in direct analogy with the physical environmental strategy, before they start with tangible digital or data projects.

AI in the digital society
Quality of algorithms and decision-making
Prof. dr. Eric Postma

1. Give politicians and policymakers regular training on the subject of data use and the emerging developments in AI in relation to social and legal aspects.
2. The digital society will require professionals who are equally at home on both the technological side and the human and social side of AI. Invest in the programmes that will produce them now to eliminate the already existing shortage in such professionals.
3. Try to stop the brain drain of researchers by improving conditions for AI researchers.
4. Invest in research to facilitate the transition to an equitable digital society.
5. An independent body to screen AI technology for the aforesaid statistical principles will prevent abuses. As machine learning algorithms are continually upgraded, periodic review is needed. The government should make review mandatory in order to minimize the chance of abuses.

Data makes the world go round
Proposal for research into three policy instruments designed to strengthen (digital) autonomy
Theo Veltman and Rob van Kranenburg

1. The creation of a national, independent and open digital trust infrastructure in order to identify, authenticate and authorize (the collection of) personal data, accompanied by the required governance structures, based on existing building blocks such as Tippiq and Irma. The white paper (Dunnen et al., 2019) can serve as a starting point.
2. An investigation of the options and feasibility of realizing embedded hardware: a ‘data kill switch’ in smart devices such as sensors, with the collection of personal data set to ‘off’ as standard. By default, it would only be possible to collect data that is anonymized and which has been identified as necessary to preserve public order and manage the city’s public space.
3. An investigation of the options for and utility of setting up a ‘data safety label’ system overseen by a democratically established independent institute with a legal mandate to actively prevent and, if necessary, act against data theft and misuse, in order to supplement the General Data Protection Regulation (GDPR).
Appropriate use of data in public space
Epilogue:

Policy challenges relating to data use

Dr. Sarah Giest
Assistant Professor

Prof. dr. ing. Bram Klievink
Professor of Digitalization and Public Policy

Prof. dr. Bernard Steunenberg
Professor of Public Administration

Institute of Public Administration
Leiden University

Summary

Policymaking in an area as dynamic and ever evolving as government data use is challenging. Because of this dynamic setting, there are no easy answers. In this epilogue, we review the various recommendations made in the essays to discuss the potential for formulating new policy. In this discussion, we will highlight how some policy recommendations have broader-than-expected consequences. We also argue that we need to set out the boundaries of the area we define as ‘data use’ in more detail. In light of these uncertainties, we call for an adaptive policy based on a polycentric and co-regulatory approach. This epilogue presents this approach, which is essentially a combination of several analytical frameworks. Applying our approach, we present a procedural policy framework on which a substantive policy framework can be built in the coming years. In this proposal, major stakeholders will have a role by converting their experiences about data use into policy.
1. Introduction

The essays in this volume offer interesting starting points for further policy development. They address a very timely and multifaceted topic, as is evident from the broad variety of perspectives, analyses and proposed measures. In this epilogue, we will provide some further tools to support policymaking on government data use. We structure our contribution as follows.

First, we will discuss the framework for structuring the recommendations made in the various essays. This should make clear whether these suggestions cover the scope of the policy area to a sufficient degree. While recommendations are often structured towards the policy problems that have already been identified, it is important to also examine how policy intervenes in the ‘world of action’ (O’Toole 2000: 273). This provides us with insights into the implementation dimension that will become important once policy is formulated.

Second, we discuss ongoing developments in the area of government data use. As our society and its governmental functions increasingly move into the digital realm, and the technology surrounding this transition continues to change at a breakneck pace, seeing what the problem areas and domains of this policy might be remains an ongoing challenge. Data use by government is becoming increasingly seamlessly integrated with society and industry, because it touches all aspects of our society and, in addition, is being collected by private parties.

Thirdly, we will address the question of how we can develop policy in a field so dynamic and so rapidly changing that it is unknown where we will be on data use (in the broadest sense of the word) in the coming years. With this field as volatile as it is, any policy, even after it is established has to be flexible enough to accommodate adjustments based on experiences and new insights. In this epilogue we will outline how we can do this, based in part on what we have learned from the various analytical frameworks developed so far.

Combining insights from different literatures, we will provide an example of a more open policy formulation, in the hopes of inspiring policymakers with it. We conclude this epilogue with a few general suggestions for policy development in an uncertain setting, such as the policy area of government data use.

2. The framework: structuring the various suggestions

The essays in this collection present a large number of exciting suggestions and recommendations for policy development in the area of government data use. In preparatory process leading to these various essays, the Ministry of the Interior and Kingdom Relations, together with stakeholders, has identified various bottlenecks about current data use by governmental bodies. This has resulted in a memorandum of issues governments are struggling with in connection to policy development. These problem areas are (BZK 2019):

1. The conflict between privacy and technology;
2. The conflict between companies and individuals, and between the individual and the collective, public interest.
3. The conflict between transparency and the extent to which data use in the public domain can be motivated and explained;
4. The conflict between the indicated purpose of data collection (in accordance with the conditions under the GDPR) and data reuse;
5. The conflicting interests in the partnership between public and private parties;
6. Finally, the question what role government should play within the ‘smart society’.

These are the points around which a new policy should ideally be oriented. As such, the various suggestions in the essays should offer solutions for these problems. The general question is whether that is indeed the case; and to answer this question, we will focus on a general framework, rather than specific actions.

A first and often-made step is to cluster the suggestions around the individual problems that have been identified. This provides an overview of whether each problem can be linked to a recommendation. That is important, but does not answer the question of whether we have a reasonable and sufficient overview of all possible ways to address these problems. There may be relevant action that is not included in the essays. A mapping by problem area therefore does not necessarily mean a breakdown of all measures and actions that are relevant for developing policy.
There are also other ways of mapping recommendations, for example, based on typologies of potential measures or by the impact of policy on society. Another mapping strategy is based on the various levels of action on which policy is based: this is a somewhat adapted version of the Kiser and Ostrom (1982) classification of ‘levels of action’. These levels differ, but they all play a role in the process of implementation. The mapping makes clear that actions being proposed in the context of policy can be carried out at various levels, but also emphasizes that the consequences of these actions, which are part of policy, are the result of choices that are made at each of these levels.

The levels of action are:
1. **The constitutional level**: the selection of basic rules that serve as the foundation for a social order (in the case of data use, this could be basic rules or fundamental rights relating to data use). Some may argue that there should also be a meta-level identified here; this corresponds somewhat to the distinction that Custers makes between legal and ethical considerations in data use.
2. **The political level**: rules governing decision-making within a political system. How are the decisions actually made? In other words, how do we deal with certain issues procedurally, for example, who will address them, who decides, and how are proposals for policy prepared?
3. **The organizational level**: rules about how governmental organizations and other relevant organizations are to act, including how organizations may cooperate. At this level, also administrative relationships are relevant, since any regulation of government data use requires coordination and consultation with local and provincial authorities, and national government.
4. **The operational level**: how in practice officials and other stakeholders actually handle government data use. This brings us to the area of actual behaviour and real-world actions. This entails many complexities with regard to, for example, the data itself, the technology and algorithms, accountability on their use, and the value considerations in implementation.

When we group the various suggestions and recommendations under this system, a number of interesting points emerges.

Firstly, we note that the policy recommendations from the essays span all levels of action. While individual essays have varying levels of emphasis on one or more of these levels, looking at them combined, there is a certain balance. It is also interesting that the recommendations vary in their degree of abstraction: while some are extremely concrete, others are much less so. This presumably has to do with the fact that the policy area of government data use is particularly dynamic, while at the same time there is still comparatively little experience, which could serve as an anchor point for policy development.

The search for some point of departure is most prominent at the constitutional level. To make sure that governments use data in a proper manner and connect this with public values, the essays strive to identify basic principles for public data use. To do so, they sometimes fall back on general principles developed for other forms of government action (see, for example, the Custers essay), or they identify multiple and sometimes comparable values that can and should be important to governmental authorities. This is an exciting exercise that makes clear that further development is needed. In this continuing search, for which experimentation is important (see the Kortuem essay), values also play a role. In their essay Engelbert and Van Zoonen point at these process values and emphasize the importance of SHARED values.

To make sure that governments use data in a proper manner and connect this with public values, the essays strive to identify basic principles for public data use.

At the political level, the recommendations refer fairly frequently to citizens who need to be involved more in the decision-making process on data use (see, for example, the essays by IJsselstein and Custers). The hope is that involvement not only reduces the information disadvantage of citizens, but also offers the possibility of creating trust and perhaps stimulates even a more proactive attitude in discussing these issues.
For this policy area, a broader engagement of the public is important. The question is whether all of these hoped-for effects will arise. Many other experiments with public participation have revealed that not all members of the public participate in an equal way, which leads to inequality in the long run. Furthermore, individuals may also have rather different views on data use, which does not facilitate creating clarity around the desired course of action.

The essays also contain policy recommendations at the organizational level where it comes to the interaction between government and private parties. Because data is collected and analysed by both private and public parties, this interaction is unavoidable. The essays also make recommendations about establishing new organizations with a task in this area. This opens up an interesting discussion about how these new tasks and new authorities line up with those of other and already existing organizations. What also stands out is that the partnership between government agencies and the impact of more data use within government receive less attention in the essays, despite the extreme importance of this aspect.

The recommendations at the operational level focus on training and familiarizing government officials with new technologies, for example in the form of pilot projects that are currently already being performed by local authorities.

Next to the fact that all levels of action are present in the policy recommendations, there is a second observation: the analyses still devote limited attention to the interaction between the four levels. This, too, has to be an important consideration in policy development, because measures at level 1 will have an impact on measures at level 2, which in turn have an impact on measures at level 3, and so on and so forth. It would be a good idea to approach the recommendations more from this integrated and perhaps holistic perspective. In other words, the selection of general principles at the constitutional level has consequences for recommendations at the levels beneath it and vice-versa, including the organizational and operational levels.

The interconnection and interaction between the various levels of action must, of course, play a role in the development of policy, and the links between the various levels must be understood. In the essays the interaction between levels is not always explicit, even though a proposal for a policy measure at one level will often imply choices at another level as well. While the essays do appeal to certain values, in many cases they are actually referring to a trade-off between several values and the question of who needs to make a decision. The proposed measures often entail an outcome (sometimes implicit) of a choice concerning such a trade-off. Deciphering which (often conflicting) values are at stake can help make clear how, for example, selected measures at an operational level have a direct connection with some (and often still implicit) choice at the constitutional or political level.

One such example is the ‘data kill switch’ described in Veltman and Kranenburg’s essay. While a kill switch could be seen as a decidedly operational measure, its underlying values are at the constitutional level: ownership and control over the individual’s own data as a right, in line with the right to be forgotten as established in the GDPR. A possibly competing value is, for example, the effectiveness of services or risk assessments that could be negatively impacted if everyone has an ‘opt out’ for their data. If the idea of a kill switch were to be broadly applied, it would make some data applications effectively impossible. For example, giving anyone the option to make their own data unavailable for data-driven fraud detection would render fraud detection pointless.
It is interesting to note that fraud detection also makes an appeal to constitutional and political values: allowing fraud to exist is unjust, and in many cases even dangerous.

In some cases, the public itself has been known to call for data-driven fraud detection. Seen in this way, the choice for a specific measure suddenly becomes an expression of the fundamental tension between values, like justice and safety versus self-determination (over one's own data). Additionally, it is conceivable that individuals, through their input into the political process, could embrace both principles, which would then demand a clear and careful consideration of any political decision (particularly if binding) before making it. While this example has obvious weaknesses, it does illustrate that it is useful to think about the interaction between the several levels of action. In the discussion on data use, it is important to keep in mind that behind every concrete policy measure, there is an implicit weighing of values. This also means that we must continue to ask the question of how concrete measures impact other levels that at first glance might not seem relevant to the measures at hand.

One final note is that the typology of Kiser and Ostrom (1982), which we have used here, devotes no attention to one important consideration in policy development in any new and dynamic field, namely: how to develop policy in an uncertain environment so that we can continue to learn from past experience.

3. Data as policy domain

When looking at data use as a policy issue, it is not obvious what the concept exactly is or represents. The essays paint a picture of a very broad and multifaceted concept, that any policy must differentiate between. Loosely based on the essays, we identify four broad areas around which policy can be oriented.

Firstly, there are the general frameworks for data use. How do we update existing laws and regulations for the digital age? How do we make sure that these are adaptable to future innovations? Who actually owns data? How to safeguard rights and principles like non-discrimination, transparency and privacy? The government can attempt to define frameworks for answering each of these questions.

However, ‘data’ is such a generic concept that perhaps the better question is where governments have a role, and when, where and what aspects of data use require policy.

Secondly, policy can be oriented towards the use of data in/by society. This can be broken down into various aspects. Data offers opportunities for innovation, improving efficiency and new economic activities. Creating these opportunities can be a policy objective. Data also represents challenges and threats to public and democratic values, such accessibility, inclusion and many other values. Policy can help, but is not necessarily effective. Complicating factors here are that tech companies generally operate internationally, that the distribution of roles and responsibilities is sometimes ambiguous, and that jurisdiction does not always correspond to how and where data are saved and used.

Policy can also be oriented towards the use of data in/by governmental authorities. The government produces and uses data. The complexity here is very high in terms of resources, applications and potential applications, interests, use, and the types of government organizations and departments involved. Fragmentation can be seen in data systems and provenance, and the presence of functional ‘silos’ that are difficult to connect or integrate. This calls for data governance that addresses sharing, management and use of data. Important aspects here are monitoring and supervision, the proper use and reuse of data as a resource, transparency, accountability and democratic control. Also relevant are good stewardship, and steering and control of third parties involved in the collection, processing or use of data. Depending on the situation, this could be an organizational matter and/or a policy matter. The diversity of use calls for policies that goes far enough to offer flexibility without going so far that it becomes an effectively empty promise.

Fourthly, policy can be geared towards the building blocks that facilitate the uses mentioned here above. Given the importance of data for society and government, infrastructural aspects also have to be identified. Data can be seen as a public good, whereby its availability and use requires (meta) standards, a technical infrastructure, supervision and disclosure. In the United Kingdom, the National Infrastructure Commission has established a development path towards the creation of a digital model (“digital twin”) for the entire national infrastructure.
Particularly when third parties will be allowed to access this resource, control of a vast number of facilities will be required in authorization, authentication and secure data exchange (to name just a few areas). In fact, quite a lot has already been done in this area, often at operational level.

4. Adaptive policy development and learning: how to deal with uncertainty?

The development of policy for a new and dynamic phenomena calls for a different approach than the one to more static problems (i.e., the ‘classical method’). In essence, as described by Lindblom (1959), the amount of information on the policy problem defines the potential pathways of policymaking. Lindblom identifies two approaches to policymaking: the *rational-comprehensive method* and the *incremental method*. While the rational-comprehensive method proposes a policy intervention on the basis of a detailed, and generally exhaustive, analysis of the policy problem, the incremental method is based on smaller steps: trying out an approach, analysing the results thereof in terms of reducing the problem, and adjusting the approach in light of this experience. The incremental method produces a ‘less ambitious’ blueprint of a policy, but it revolves around learning about the policy problem and how to tackle it.

The incremental method offers advantages in situations of uncertainty and ambiguity concerning the policy problem (when we do not know all possible interventions and their effects, and where even for known effects the likelihood of their occurrence is still guesswork). Some of the essays about data use by the government do emphasize this point by referring to the experimental nature of data use. Because the technology continues to develop, it is not possible to see what can be done ‘tomorrow’ and what the effects of acting will be. At the same time, it is important for the government to proactively develop policy that already streamlines data use now to avoid undesirable outcomes. The current situation can be accurately described as rampant growth of initiatives and applications, that are spread out across various layers of government and among various organizations. Current attempts at policy are often reactive in nature, because they are driven primarily by responses to developments that have either led to problems already or are generally seen as undesirable. These problems, as the essays also demonstrate, are mainly clustered around certain areas like privacy, inequality, discrimination and the increasing dominance of private parties in the access to and use of the data that are important for government functions.

A number of options for a more dynamic approach to policy have been developed and tested over time. We discuss a number of significant examples of models that could be relevant for policy on data use. Here we primarily consider the degree to which lessons can be drawn from previous experiences with policy implementation.

The first group of models contains the more static policy models, which exemplify the fact that policy has a limited shelf life. In this group, it is known that there is a changing environment, but government is unable or unwilling to fully include the consequences in the policy design. One example is the frequently employed ‘sunset clause’ that limits the duration of policy so that a new design can then be chosen. A less drastic example is a clause that mandates a re-evaluation after a certain period of time (say, five years).
A great deal of European legislation is based on this model. In such cases, policy is still static in terms of structure, but the hope is that after a reasonable period of time improvements can be proposed. These models are less attractive for policy on data use because they offer little stimulus for learning from the consequences of technological change.

The second group are models that emphasize uncertainty and already take into account various options that have been identified in advance in the policy-making process. One example is ‘adaptive policy planning’, which adds a number of significant elements to the policy (see Van der Plas et al., 2013; Haasnoot et al., 2013). In this model, the most important uncertainties of the proposed policy measures must be inventoried, after which policymakers consider potential corrections in the form of flanking policies. The implementation is monitored continuously so that additional measures can be applied when and where necessary. Adaptive models are a major step forward from static models because they require a consideration of unexpected and unintended effects. At the same time, these models offer little to go on if the effects are not yet sufficiently clear, and even less when it comes to proposing approaches to reducing these effects. These models are therefore less appropriate for making policy on data use because there it is important to learn from the developments that will continue to emerge in the coming years.

A third group of models emphasizes learning by means of identifying best practices by exchanging experiences. Those experiences can be used for further refining the policy. One known example of this type is the Open Method of Coordination (OMC) used within the EU. This method attempts to learn from the experiences with the implementation of policy ambitions by voluntary sharing and exchanging them. Elements of this method can be seen in the process launched by the Ministry aimed at sharing the experiences and problems encountered in data use. Application of this model shows that although sharing experiences is important, it still is a long way to actually formulating new or different policies. OMC remains too unstructured on this point to do this successfully.

A fourth group of models for policymaking is based on further policy development within a predefined framework. One familiar example is the Lamfalussy method, developed for technically complex financial issues within the EU. This method features a number of layers through which the policy is formulated in a step-by-step process. It is the legislator that defines the general policy framework, including the primary assumptions in the implementation, while the stakeholders work out the details. Their proposals must then be approved by a commission of member-state representatives (a form of delegated legislation within the EU context). The interesting aspect here is the involvement of experts in a multi-layer model of policy development. This allows new insights to be incorporated into the policy, which is important particularly for more complex problems. The flip side is its reduced political (‘democratic’) grip on the lower levels of the process.

It is important that the learning process be a continuous one, in which the policy’s implementation can be continually adjusted in response to current experiences.

These models each emphasize different aspects of a process used to arrive at a successful policy. All of them have a dynamic component, in the sense that it is understood that policy needs to incorporate new experiences. This is central to the concept of adaptive policy (Steunenberg 2018), which emphasizes learning in policy. This learning can take place in various ways, from learning in hindsight (when working with evaluation clauses) to learning during the process (as in the OMC). For adaptive policy, it is important that the learning process be a continuous one, in which the policy’s implementation can be continually adjusted in response to current experiences. For the design of policy on data use by the government, in which the desire is to formulate a policy now, the challenge is in proposing a framework that will make it possible to keep adapting this policy to insights that we do not yet have.
5. A proposal for adaptive policy

From an analytical standpoint, creating a policy structure that is adaptive (able to adapt to rapidly developing technology and usage), the objective should be to focus on learning and experiences gained during implementation. For this, a balance must be sought between the centrally selected assumptions or principles and the locally available experience and knowledge: between top-down formulated principles on the one hand, and bottom-up available knowledge of problem areas, potential policy interventions and their effects in society.

In the case of data use, local governments are important in striking this balance due to the policy experiments that they are carrying out. At the implementation level, there are of course numerous considerations to be made between competing values, diverging perspectives and wide-ranging interests of relevant actors. Expertise in technology must be paired with domain expertise and policy expertise. Private parties also have a role to play here. For productive policy lessons, the experiences of local authorities and private parties are therefore essential to further the development of policy. Adaptive policy in this context demands repeated dialogue between stakeholders and experts, based on the slogan ‘making progress rather than solutions’ (Head, 2017), keeping in mind that the policy must protect and preserve public values.

As already stated, a policy on government data use must offer sufficient room to experiment and to learn, and the resulting knowledge and experience must be the input for further adjustment and development of the policy. We therefore propose a polycentric co-regulation approach. This term, as we use it, represents a distillation of three analytical frameworks. Firstly, polycentric decision-making must reduce the ‘information asymmetry’ by bringing together knowledge that is dispersed across society as a whole (Dorsch and Flachsland 2017). Secondly, co-regulation is the process of involving a broader group of stakeholders and decision-makers in the drafting, implementation and further development of the policy. This may relate to both the regulations and the development of the standards that will play a role in the regulation/monitoring. Thirdly, it is a cyclical approach based on learning through the ongoing confrontation of the outcomes of previous choices with new insights.

The polycentric perspective is compatible with policy domains that involve very complex problems (or even ‘wicked problems’) in which policymakers must manage a vast array of goals, interests, sectors and levels relevant to the policy and its implementation. This perspective offers room for context-specificity and for experimentation and learning at the local level. It proposes a framework regulation in combination with self-organization in a specific context.

The polycentric perspective offers room for context-specificity and for experimentation and learning at the local level.

Co-regulation addresses the problem that no individual party has all the knowledge needed to solve complex, multivalent and dynamic problems, and that there is a lack of oversight and grip on the tools needed to arrive at an effective policy (Finck, 2017). Given the international nature of the IT sector, self-regulation or co-regulation would appear to be very useful here. The arrangements between the municipality of Amsterdam and Airbnb on the tourist tax and the 60-day period are an example of such an arrangement. This model has been referred to as ‘regulated self-regulation’. At the same time, the Airbnb example also shows that private parties have very little impetus to opt for such an approach and that monitoring is a challenge. Consequently, regulation and policy with regard to the private sector proves to be important after all. A prominent example using both elements is the ‘really responsive regulation’ concept proposed by Baldwin and Black (Baldwin and Black, 2008; Black and Baldwin, 2010). This concept implies a strategy in which a broad diversity of instruments (including legislation and regulation) is applied flexibly to allow space for various factors, including behaviour, attitude and culture of the company. Likewise, the institutional environment, the interaction between instruments, changing priorities and objectives, and the performance of resources can also be taken into consideration (Baldwin and Black, 2008).
With respect to the approach to be selected within a polycentric perspective, there remains an inherent tension between, on the one hand, offering ‘local’ flexibility or responsibility, and on the other hand, the objective of achieving more coherent regulation and standardization of technology within the context of the proposed policy. The tension arises from the desire to develop regulation that can adapt to the context and various government levels, while at the same time allowing that same regulation to serve as a broad basis for decision-making processes. Complex and self-policing policy is sometimes difficult to achieve, while too narrow and too inflexible policy runs the risk of not doing justice to all the nuances inherent to data use in the public sector. A focus on communication, coordination and negotiation between public, private and societal stakeholders is needed to navigate the course between these two extremes and to arrive at regulation that is compatible with the reality in which it must function.

In the academic literature, this is described as a bidirectional flow of bottom-up and top-down communication. The latter comprises the communication of an implementation strategy and implementation objectives, together with the procedures and resources that come along with the implementation. Bottom-up communication comprises the engagement of stakeholders in considerations, objections and intended utility of new or adjusted regulation, and in the assessment and cataloguing of the needs with regard to its implementation. A point of attention here is the risk of incompatibility with existing procedures and legislation. There are plenty of examples, particularly with bottom-up communication, that demonstrate that it is important to keep multiple communication channels open without allowing the feedback received via these channels to become fragmented. Stakeholders have various methods and needs in communication, and these can also change with changing circumstances. For example, bottom-up communication sometimes comes via the feedback from umbrella organizations (like the Association of Netherlands Municipalities), while other important input can come from conferences or meetings. Teams of stakeholders working in a specific context can also make recommendations and identify challenges from the perspective of their local roles.

This polycentric perspective is one step removed from a reactive regulatory approach in which the government only responds to compliance with regulation, instead of proposing a flexible approach in which ongoing communication and feedback is organized to facilitate adjustment to specific contexts and places.

The cyclical approach draws upon forms of policy learning that have already been put forward in more classical public administration studies. The incrementalism promoted by Lindblom was already based on combinations of doing and reflecting so that knowledge about how policy currently works could be used in subsequent steps. This idea of short learning cycles is reflected in various different approaches and concepts, including in the well-known ‘plan-do-check-act’ cycle. It is, however, important that the ‘check’ is handled appropriately in order to be sure of what the policy effects are and the extent to which these effects are the result of the policy. Here, scientific methods and insight (e.g. an evidence-based approach) may play an important role.

The combination of polycentric decision-making, co-regulation and a cyclical approach means that more stakeholders can participate, while at the same time offering the flexibility to develop technical, context-specific and location-specific solutions within a broader set of rules.

6. How can we apply adaptive policy?

To illustrate our polycentric, co-regulation approach for adaptive policy, we will elaborate an example of how a policy on government data use may look like. In line with our proposed approach, this example is a combination of policy and process, because it must allow for the possibility to further develop and adapt the policy on the basis of new experiences.

We first present our example, which we then clarify:

1. The legislature proposes, where possible or desired, a general substantive policy framework including central principles. These are principles that are identified in the various essays, such as privacy, equality, non-discrimination and proportionality.
2. The legislator also defines a *procedural policy framework* on how the policy should develop over time. This is based on *working groups* in which specific subjects or themes will be worked on further. The responsibility for this development may be assigned to, for example, the responsible minister. (There are also other conceivable models about political responsibility, but in this example, we assume a limited political distance in the placement of responsibilities for this area.)

3. The working groups consist of different stakeholders, each with an involvement in a given subject area; here it is important to engage the private sector. As the essays show, cooperation with private parties is essential for the success of any data use by government. The individual parties from the public sector (including local authorities) and the private sector contribute relevant knowledge on the working group’s subject area.

4. The working groups *formulate the rules with regard to the subject area or problem*, which will be adopted by the responsible minister.

5. The working groups *monitor* the implementation of the rules defined in light of ongoing developments (including compliance and enforcement) and new knowledge (evidence-based).

6. New proposals are evaluated based on available evidence on their functionality and effectiveness.

7. Monitoring becomes the basis for *short-cycle* adjustments of the policy, for example, annually or biannually. The minister approves these new adjustments. This process produces the flexibility that is so important in this subject area. (Obviously, it is important to realize that this means that the guarantees of legal certainty wane over time, but this is the price of flexibility.)

8. Over a *longer (political) cycle* (for example, once every five years), the substantive policy framework defined under point 1 is discussed in Parliament. This may result in larger political adjustments.

There are number of comments we would like to make concerning this proposal.

First, our example assumes learning by working with cycles of discovery, testing, gaining experience and adjusting; these cycles are essential for policy learning. This applies for the short cycles at the level of the working group of stakeholders, which need to arrive at more concrete forms of implementation. It also applies to the cycle of adjustment of the regulatory framework, which involves the minister. Finally, it also applies to the political cycle in which the policy is discussed in Parliament, and, in general terms, can be adjusted on certain points. Working with these learning cycles at various levels offers the potential to incorporate new experiences into the existing policy, along with the possibility to address, in due course, unexpected effects or problems that are not yet now adequately understood, but which may arise at a future point in time.

Second, our example links the various action levels from the preceding section to the policy framework. The working groups further develop the structure and implementation of policy within the general framework of standards (which, it should be noted, can be the subject of discussion in the policy process as well). This makes the connection in the design between the levels of action we identify here.

Third, the example assumes an organization of policy development that places the political responsibility with the national government, and the minister specifically. With respect to the role of the national government, an important observation is that the development of a shared framework at the national level brings with it a number of significant advantages, such as advantages of scale and guarantees of legal equality and legal certainty. At the same time, within this framework, we must maintain the freedom to experiment and learn from various stakeholders, including local authorities. This is why it is important to have several levels of government represented in the working groups alongside the private sector.
7. Conclusion

Data is a complex and frequently chaotic area in which to make policy. By definition, there will be an asymmetry of information with respect to the many forms of collecting, using and analysing data. This calls for two elements in developing policy: first, a certain degree of standardization and coherence is needed for the definition of individual policy concepts, while at the same time these concepts need to be linked to a specific context. Second, the involvement of a broad group of stakeholders must be solicited to provide ongoing input and feedback through the appropriate communications channels. This is a cyclical process, which allows for learning. All this demands a broad perspective on the process, without losing sight of the specific implementation context.

More generally, it is important to consider, at regular intervals, what exactly should be permitted in terms of policy in this domain – what exactly are the responsibilities of the government and what is the purpose of the policy? For data use in, for and by government, answering these questions might be more straightforward than for applications in which data from public sources (open or privileged) are combined with data from other sources or used in private applications. Yet another case occurs when fully private applications are used in the public domain. To what extent can we design a policy that can handle the complexity that comes from the pervasive and diverse nature of data we generate through virtually everything we do? At the very least, this pervasiveness and generic character of data should in no way be a reason to do nothing or little, but to make it clear to everyone what policy can and cannot be about.

A related observation is that due to the complexity of the subject, ambiguity is essentially unavoidable, and this is also inherent to the various levels identified in this epilogue. Intuitively, we may very quickly develop an idea about the level at which various measures have their effect. It is important to continue questioning these presumptions. A conflict between operational tools and democratic values seems, at first, to be an unequal struggle. However, much of our thinking and discussions about data are built on a whole range of value trade-offs and implicit choices. The recognition and explicit statement of these choices and the relevant values concerned, can help us move beyond overly generalized main design principles and the direction we should be going.

To facilitate adaptive policy based on our polycentric co-regulation approach, the formation of working groups (points 3 and 4 of our example) and formulating and discussing possible rules are a possible starting point. This is along the same lines as the decentralized approach the Ministry of the Interior and Kingdom Affairs selected in drafting its problem statement on government data use. This approach does demand a clear procedural policy framework in which the tasks and responsibilities are assigned (point 2). This, in turn, provides input for the substantive policy framework (point 1), insofar as it is possible to establish this at this stage. Through the other elements in our approach, this framework can then ‘grow’ into a more coherent package based on the problems as well as the potential solutions that we will encounter going forward.

To conclude, in this epilogue we outline the challenge of proposing policy oriented towards more learning and in which the insights that we do not yet have can be added to its framework. This requires a policy that is evolving and stable at the same time, that is, a policy able to cope with the dynamic area of technological development while providing a larger framework in which these changes take place. This also requires policy that stimulates policy learning, and therefore encourages policy experimentation, with the results then being incorporated into the policy framework. All in all, a difficult but exciting challenge!
References

1 See also Ostrom (2004:58), who discusses these levels and related mappings. Each of these mappings has a particular focus appropriate to the questions one would like to answer. In our case, we add the organizational level as part of the classification, because this is an important element within administrative systems.

- Dorsch, M.J. and C. Flachsland (2017) A Polycentric Approach to Global Climate Governance, Global Environmental Politics 17: 45–64. doi/10.1162/GLEP_a_00400.
Proposals for policy measures and activities

The policy recommendations set out in the essays laid the foundations for the following policy measures and activities, which were subsequently formulated in accordance with a scientific framework developed by Leiden University’s Institute of Public Administration. At the time of writing, the policy measures and activities have been placed on a longlist for the preparation of the Data Agenda Government 2020. The recommended measures and activities will be evaluated by the Data Agenda Government sounding board and steering committee.

Policy proposals

1. Investigate of whether the Integrated Impact Assessment Framework for policy and legislation (the IAK) is up to date concerning the issue of digitization in the public sphere.

Up-to-date assessment frameworks are an essential precondition for the active involvement of stakeholders (including citizens) in deliberations and the provision of information to all stakeholders. Useful instruments in this regard include the Code of Good Digital Administration, the Corporate Social Responsibility Tool Box or the Social Dialogue programme. Digitization has changed the range of stakeholders in the public space, especially in relation to ethical aspects. In the past, tech companies were solely digital entities, but they have since moved into the public space. As a result, the ethical dilemmas of digitization seem to transcend the GDPR and other laws. Readjustment of the IAK can provide concrete reference points whilst maintaining sufficient flexibility at the local level to ensure solutions are found to specific dilemmas. By extension, there is a lack of assessment frameworks at the executive (municipal) level due to the digitization of public works such as bridges and tunnels. As a result, the IAK should be evaluated to ensure it is up to date.

2. Organize a social debate on the issue of AI involving collaboration between researchers, developers, policymakers, citizens, etc.

Subjects that could be discussed in this regard include:
- Development of an Algorithm Reporting Framework (as well as a hotline)
- Establishment of a National Algorithm Register, linked to an Algorithm Forum
- Formulation of an AI Skills Agenda for educators, politicians and policymakers
- Examination of the impact of digitization of the public space on citizens’ fundamental rights
- Examination of the opportunities for certification of AI algorithms (quality mark for data safety)

3. A. Investigate how the public debate concerning data use in the public space can be incorporated into the process of innovation (during policy development and/or the execution of public duties) from the very start.

This relates to issues such as reinforcement of local governance and the use of design principles (such as user-centred and value-sensitive design) at the start of the process. Periodic evaluation featuring citizen involvement is clearly an essential aspect for the innovation process.

Such evaluations could be carried out as part of a collaboration between central and local government bodies, private parties and the academic community. It is vital that theory is implemented into practice, a process that will be conducted in living labs which offer testing grounds for experimentation. It is worth investigating whether this can be done in conjunction with existing initiatives or planned new initiatives. Living labs, with their case-specific settings and involvement of stakeholders with a wide range of interests, are particularly well-suited to this kind of interconnection.
3. **B. Organize a platform that gives complete insight into every aspect of the pilots.**

   The organization of this kind of platform is compatible with measure 3A. The purpose of this platform would be fourfold:
   - An investigative role, examining issues such as ethics (e.g. codes of conduct governing responsible processing of personal data in the public space) and authorization of government bodies at various levels to use and reuse data from one or more sources.
   - An implementing role, e.g. via application of the Responsible Innovation Tool Box, implementation of codes of conduct and formulation of the proposed Digital Environmental Strategies (analogous to the existing Environmental Strategies relating to physical public space at the local level).
   - A facilitative role with regard to knowledge sharing (e.g. via scenarios for collaborative pilots).
   - An advisory role focusing on standards and values in the digital society.

**Actions**

1. **Conduct research into technical and/or organizational solutions that guarantee the anonymity of users and security in the public space.**

   This research could be conducted in collaboration with academia, the business sector and last but not least social organizations (e.g. the Bits of Freedom Foundation, Waag Technology & Society, etc.).

2. **Conduct research into a generic trust infrastructure within the public space.**

   As well as inventorying and evaluating fundamental public digital infrastructure, opportunities concerning the establishment of a national, independent and open digital trust infrastructure for identification, authentication and authorization of personal data, including the requisite governance thereof should be investigated.

3. **Conduct a feasibility study into the registration of sensors.** At the moment, we have observed a variety of initiatives: would it be possible to scale these up?

   A variety of sensor-registration initiatives already exist at the local level. A feasibility study into the possibilities of amalgamating these initiatives, (federally) centralizing them under the authority of a single private, public or public-private party, or maintaining their decentralized status could help paint a clear and comprehensive picture of the opportunities available. In such cases, the key focus would be the possibility of establishing a national legislative framework for the registration of sensors (as mandated for cables and ducts by the Exchange of Information (Overground and Underground Networks) Act (WIBON)).

   This could be organized in collaboration with the Ministry of the Interior and Kingdom Relations, government bodies at various levels and private parties.

4. **Examine possibilities for expansion and/or scaling up the Model Agreement on the use of data collected in the public space (once it is formulated and implemented by the Ministry of the Interior and Kingdom Relations and the Association of Netherlands Municipalities).**

   Expansion could include the use of the Model Agreement by all 355 Dutch municipalities, while scaling up could involve the Model Agreement being used by other government bodies or central government for the purposes of contracts with commercial parties. In addition to model agreements, the scope of this kind of process could also include aspects such as standardization, certification or validation as well as associated standards (creation of ‘generally binding provisions’).
One possible option is to add a section to the IAK on responsible experimentation areas ("legal sandboxes") for the purposes of living labs.

For this purpose, a number of new initiatives have been set up at both the national (within the National Academy for Government Digitization) and municipal levels, which can be interconnected in myriad ways.
Publication information

This publication came about thanks to the unconditional dedication of the ‘Appropriate use of data in public space’ supervisory committee, which was made up of the following persons:

Prof. Frans Feldberg (Vrije Universiteit Amsterdam)
Ran Haase (Municipality of Eindhoven)
Professor Emeritus Peter van Hoesel (Department of Public Administration, Erasmus University Rotterdam) (Chair)
Marcel Hopman (Ministry of the Interior and Kingdom Relations)
Prof. Maurits Kaptein (JADS Institute)
Klaas Kloosterman (Municipality of Groningen)
Prof. Sofia Ranchordas (University of Groningen)
Prof. Arjan van Timmeren (AMS Institute)
Theo Veltman (Municipality of Amsterdam)
Jeanine Vosselman (Municipality of Groningen)
Aantink Yeh (VNG Realisatie)

www.nldigitalgovernment.nl/appropriate-data-use
www.nldigitalgovernment.nl
October 2019