

Ministerie van Binnenlandse Zaken en  
Koninkrijksrelaties

Aan de Voorzitter van de Tweede Kamer der Staten-  
Generaal  
Postbus 20018  
2500 EA Den Haag

[www.government.nl](http://www.government.nl)  
[www.facebook.com/minbzk](https://www.facebook.com/minbzk)  
[www.twitter.com/minbzk](https://www.twitter.com/minbzk)

Reference  
2019-0000518319  
Your reference

Date 8 October 2019

Re Beleidsbrief AI, publieke waarden en mensenrechten

Jointly on behalf of the Dutch State Secretary of the Interior and Kingdom Relations, the Minister of Justice and Security, the Minister for Legal Protection and the State Secretary of Economic Affairs and Climate Policy, I am issuing this policy brief on AI, public values and human rights. The brief provides an overview of the opportunities and risks of AI for public values that are based on human rights. It also describes existing and future policies to address any risks to these fundamental public values.

Along with this brief, the Dutch Strategic Action Plan for Artificial Intelligence (SAPAI) and the Dutch Safeguards against risks of government data analysis brief have been presented to the Dutch Lower House. The three briefs focus on different parts of the broad issue regarding utilisation of opportunities and addressing risks of AI. SAPAI comprises this Cabinet's comprehensive approach and includes policies to utilise the societal and economic opportunities that AI presents, while safeguarding the public interests. In track 3, SAPAI briefly examines the effects AI has on public values. Since the effects of AI on public values and human rights are complex and potentially significant, our Cabinet has chosen to give further attention in this brief to relevant policy-making. The same applies to the Safeguards against risks of government data analysis brief, which specifically examines possible safeguards against risks of using algorithms and data analyses by the Dutch government.

The present brief is building upon our Cabinet's reaction to the report by Utrecht University on algorithms and fundamental rights<sup>1</sup> and has come into being on the basis of input provided by ministries, government bodies and scientists. The brief also incorporates the commitment included in the Cabinet reaction to the report 'Urgent Upgrade. Protect public values in our digitized society' by the Dutch Rathenau Instituut to further develop policies on this subject.<sup>2</sup> Likewise, this brief sufficiently develops the motion tabled by the Dutch Members of Parliament Van Dam and Van der Molen regarding digitisation and public values, which was

---

<sup>1</sup> Parliamentary papers II, 2018/19, 26 643, no. 601.

<sup>2</sup> Parliamentary papers I, 2017/18, CVIII, S.

passed on 6 June 2018.<sup>3</sup> The proposals to enhance coherence of policy-making and of supervision and control mechanisms in this brief meet in part the initiative note tabled by the Dutch Member of Parliament Middendorp on 29 May<sup>4</sup>, the Verhoeven-Van der Molen motion, passed on 29 May 2019<sup>5</sup> and the Middendorp-Drost motion, passed on 20 June 2019.<sup>6</sup>

AI is a crosscutting technology that touches on the responsibilities of all ministries. The Dutch Minister of the Interior and Kingdom Relations is the coordinating minister for the issue of 'opportunities and risks of AI for the protection of fundamental rights', while every minister or state secretary is responsible for developing policies with respect to opportunities and risks of AI for the protection of fundamental rights in their own policy area, including the necessary legislation.<sup>7</sup> Therefore, this brief names policies that are being developed by several Dutch ministries.

### *1. Human rights as a starting point for AI policies*

The Netherlands operates a high protection level of human rights, based on a comprehensive system of regulations, provisions, institutions and accountability procedures.<sup>8</sup> This safeguards that human rights are not solely protected and promoted on paper, but also in practice. Subsequently, this (also) ensures that people enjoy living in the Netherlands, we have a vital society, businesses are happy to settle here, and our democratic constitutional state can function properly. Through its human rights culture, the Netherlands contributes to the functioning of the European Union as a community of values. Our Cabinet attaches great importance to protecting human rights, also at European and international level.

Infringements of human rights by AI must be prevented. However, not all deployment of AI requires a role by the Dutch government. Policy-making focuses on areas where AI applications have an evident impact on people and/or society as a whole. In such cases, our Cabinet dedicates itself to ensuring that AI applications respect and, where possible, even enhance human rights. Therefore, the Dutch Cabinet is in favour of a human-oriented approach to AI, internationally also known as a human centred approach, in which respect for public values inspired by human rights forms the basis for the objective, design and use of AI applications.

---

<sup>3</sup> Parliamentary papers II, 2017/2018, 32 761, no. 120.

<sup>4</sup> Parliamentary papers II, 2018/2019, 35 212, no. 2.

<sup>5</sup> Parliamentary papers II, 2018/2019, 26 643, no. 610.

<sup>6</sup> Parliamentary papers II, 2018/2019, 35200, VII no. 14.

<sup>7</sup> Specifically regarding the aspects of 'transparency, testability and legal protection' the responsibilities have been further developed in the letter by the Minister for Legal protection about safeguarding against risks of data analyses by the government, which has been presented to the Lower House along with this brief.

<sup>8</sup> Netherlands Institute for Human Rights. (2018). Human Rights in the Netherlands 2017 – Annual report (in Dutch), <https://mensenrechten.nl/nl/publicatie/38613>, and parliamentary papers II, 2018-2019, 33 826, no. 25. Ministry of the Interior and Kingdom Affairs. (2013). National action plan human rights (in Dutch), <https://www.rijksoverheid.nl/documenten/publicaties/2013/12/10/nationaal-actieplan-mensenrechten>.

Meanwhile, we are confronted with dilemmas that AI entails regarding the safeguarding of public values and human rights. For instance, when it is necessary to weigh the protection of privacy and the protection of security. Or the protection of privacy and fraud prevention. When formulating policies and legislation, these situations require meticulously and transparently weighing of human rights and other public interests.

## 2. AI technology, opportunities and risks and existing policies<sup>9</sup>

AI refers to systems that demonstrate intelligent behaviour through analysing their environment, interpreting the collected data and consequently determining which action is best to achieve specific objectives.<sup>10</sup> A rapidly emerging form of AI is narrow AI. This comprises systems that focus on the execution of a single task (such as image or speech recognition). Examples are virtual assistants (such as Siri), personal recommendation systems (as deployed by Netflix) and pattern recognition in images (as deployed by medics for diagnosis purposes). However, the development of the more advanced AI type of Artificial General Intelligence systems is yet at an early stage. As yet, these independently thinking, and independently reasoning systems only exist in theory. Moreover, the expectations that experts have regarding whether and at which moment this type of intelligence will be realised differ widely. Therefore, this brief focuses on opportunities, risks and policies with respect to narrow AI systems.

Literature reveals that AI offers great opportunities to society. In SAPAI, ample attention is paid to the contributions AI can make to addressing societal challenges, for instance with regard to more effective healthcare, climate regulation and better security. Apart from that, AI provides several opportunities to enhance specific human rights. For instance, the ability of AI systems to personalise information, making it more relevant to users can reinforce the right to information. Also, AI can foster non-discrimination, for instance when it is deployed to eliminate bias from a selection process. Furthermore, AI applications can enhance individual autonomy. For instance, there are numerous AI-based apps that help people make a medical diagnosis independently.

Apart from opportunities, AI developments also entail potential risks. The table below lists the risks that feature most dominantly in the literature.

Public values	Description	Risk
Non-discrimination	People must be treated equally in similar circumstances and may not be excluded unjustly on the basis of certain characteristics.	<ul style="list-style-type: none"> <li>• Bias in underlying data, resulting in pattern discrimination</li> <li>• Bias in an algorithm, resulting in pattern discrimination</li> <li>• Error margins resulting in incorrect classification</li> </ul>
Privacy	People must be allowed to be 'themselves' without fear and do as they please without interference from third parties.	<ul style="list-style-type: none"> <li>• Large numbers of data are needed for good outcomes of AI systems</li> </ul>

<sup>9</sup> This paragraph is a summary of the annex.

<sup>10</sup> On the basis of parliamentary papers II, 2017-2018, 22 112, no. 2578, p.2, adjusted to present-day understanding.

		<ul style="list-style-type: none"> <li>• Sensitive data that is generated by AI systems</li> </ul>
Freedom of expression	Everybody has the right to voice convictions, feelings and opinions and share them with others. This also includes the right to access to (balanced) information.	<ul style="list-style-type: none"> <li>• Limited access to and pluralism of information</li> <li>• Inaccurate algorithms that delete content too soon</li> </ul>
Human dignity	Simply 'being' human is associated with a certain dignity that safeguards a level of protection in relation to the government and third parties. An important part of this is human contact.	<ul style="list-style-type: none"> <li>• Decrease of interhuman (and thus the quality of) contact when AI takes over interaction</li> </ul>
Personal autonomy	People must be entitled to choose freely and largely decide for themselves how they arrange their lives.	<ul style="list-style-type: none"> <li>• Imperceptible influencing by directing AI</li> </ul>
Right to a fair trial	Everybody must have access to justice; to information, advice, guidance in negotiation, legal assistance and the possibility of a ruling by a neutral (judicial) authority.	<ul style="list-style-type: none"> <li>• Non-transparency of algorithms as a result of which persons have more trouble pursuing their rights</li> </ul>

There are already many policies in place to address these risks. The annex to this brief provides a comprehensive overview of existing policies. Various bodies are working on standards to guarantee the quality of AI systems – for instance, with regard to bias and error margins.<sup>11</sup> Thus, the Dutch Minister for Legal Protection simultaneously issues a brief on potential legal safeguards against the risks of data analyses by the Dutch government and guidelines for the application of algorithms by authorities, which partly addresses the possibility to include quality assurance standards in legislation and directives. Various Privacy-by-design concepts are available for the development and deployment of AI systems. Also, our Cabinet is supporting research into AI systems that need less data to yield a high-quality outcome. Furthermore, the Dutch Ministry of Justice and Security has commissioned a study by Tilburg University of measures restricting risks of facial recognition for privacy, in which AI can likewise play a role.

Although AI-based news personalisation does not yet have a large negative impact in the Netherlands, the Dutch Cabinet is investing in media wisdom and public campaigns to make the public conscious of the role that personalisation may play in the news provision. With respect to combatting illegal content with the help of AI, the Netherlands uses a Notice-And-Take-Down procedure, taking measures to avoid that legal content is unintentionally removed. With respect to human dignity, our Cabinet has stated in its Digital Government Agenda that meaningful contact – which means not only digital contact – is the starting point of government policy. Regarding the contact between citizens and businesses, our Cabinet has asked the Netherlands Scientific Council for Government Policy (WRR) to advise on the impact of AI on human contact. Moreover, research reveals that the Dutch government is currently not deploying AI to influence behaviour. Businesses, however, do so more and more often. Therefore, our Cabinet intends to commission research into the effects thereof. With respect to the right to a fair trial and the non-transparency of algorithms, there are several policy initiatives, such as the aforementioned guidelines for the application of algorithms by

<sup>11</sup> Moreover, an inventory recently took place regarding standards for data exchanges between government. The Cabinet is preparing a response to this.

governments and a Transparency Lab where forms of algorithm transparency are experimented on. In response to the question of the Dutch Member of Parliament Futselaar during the plenary debate on 10 September 2019, in relation to the explainability standard, attention will be paid to the extent to and cases in which citizens will be given insight into decision-making rules of the algorithm.

Apart from the above policies, which are mostly deployed to safeguard a specific public value, there are more general policies to safeguard public values and human rights with respect to AI developments. These relate to areas including enhancement of understanding and awareness among citizens through dialogue, promoting forms of self-regulation such as codes of conduct and in the area of setting the agenda and negotiating internationally.

### *3. Additional measures*

The previous paragraph makes clear that our Cabinet has a clear idea of the opportunities and risks and is actively pursuing policies on them. Still, these can be further developed in a number of respects, particularly where the coordination of safeguarding public values and human rights is concerned. My ministry has an important responsibility in that respect. Policies can gain effectivity when they are developed more coherently. Also, the concretion of the so far abstract human centred AI concepts may help various authorities to improve the inclusion of rights in the development of AI applications. Moreover, monitoring and control mechanisms may be enhanced and internationally points of view may be more strongly propagated.

#### Greater policy coherence

In order to make use of the opportunities AI offers and also be better prepared for the risks of AI applications, attention must be paid to the organisation of safeguarding and enhancing public values. As mentioned before, many measures are already being taken to address the risks that AI poses for public values. At the same time, the Cabinet is observing that the coherence between policy initiatives leaves room for improvement, as they occasionally overlap or, conversely, are fragmented. Authorities can operate more effectively when knowledge is shared, and policies are coordinated more intensely. That is why, in the coming year, my ministry will set up a government collaboration platform with regard to the subject of AI and public values. Through this platform, knowledge will be exchanged, policy alignment facilitated, connections made with science, enabling participants to work towards joint research programming.

#### Concretion human centred AI to system concepts

As stated before, the Dutch Cabinet is in favour of a human-oriented approach to AI, which is in line with the human centred AI approach prevailing in Europe.<sup>12</sup> However, current human centred AI concepts are very abstract and general. This makes it hard for developers of AI applications to include concrete public values and human rights in their designs. To enable them to do so, it must be clear how specific values and rights can be translated into an AI design. In concrete cases,

---

<sup>12</sup> <https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence>

my ministry will consider how public values and human rights can be operationalised into system principles and make them available to authorities and businesses, starting with the area of non-discrimination.

#### Reinforcing control mechanisms

The previous paragraph makes clear that various parties are working on the improvement of control mechanisms on AI (such as standards and audits).<sup>13</sup> It is still early days to judge whether these control mechanisms are effective and desirable, let alone to already make them obligatory. Monitoring is one of the control mechanisms regarding proper use of algorithms. A wide range of regulators are involved in monitoring algorithms. The Dutch State Secretary of the Interior and Kingdom Affairs and the Minister for Legal Protection will jointly examine whether regulators have sufficient tools to monitor algorithms and whether there are any blind corners in the regulatory landscape. Also, the State Secretary of the Interior and Kingdom Affairs and the Minister for Legal Protection aim to further stimulate the start that has already been made to let regulators enter a partnership where they can learn from each other's expertise in the area of algorithms and AI.

#### International propagation of the importance of public values in human centred AI

A last point concerns the international propagation of the importance of public values and human rights as well as best practices in the area of human centred AI. Many countries are investing heavily in AI. For instance, China has the ambition to be world leader in AI innovation by 2030 and the US has stated it aims to maintain its leadership position in AI innovation. Together with other European countries, the Netherlands will propagate more emphatically that AI must reinforce rather than weaken public values and human rights and present relevant best practices. Just like other value-driven technologies (for instance in the area of wind and solar energy, but also in agriculture) human centred AI can become an essential export product. As a result, the Netherlands could be among the forerunners in the area of human centred AI.

The Minister of the Interior and Kingdom Affairs,

drs. K.H. Ollongren

---

<sup>13</sup> For further examples, please to see page 8.

## **Annex to policy brief AI , public values and human rights**

### **Artificial intelligence**

AI refers to systems exhibiting intelligent behaviour by analysing their environment, interpreting the collected data and deciding on the basis of that which action is best suited to specific goals.<sup>14</sup> By and large, two types of AI systems are distinguished, namely: 'rule-based' and 'machine learning'. Rule-based systems arrive at decisions on the basis of previously defined rules and do not learn from the data they process. Machine learning systems, however, do 'learn' rules by deducing patterns from data.<sup>15</sup> Particularly in the field of machine learning great strides have been made in the last years because of new technological insights, the enhanced processing capacity of computers and the increased availability of data.

Machine learning systems can in turn be subdivided into narrow AI and Artificial General Intelligence systems. Whereas narrow AI systems focus on performing a single task (such as image recognition), Artificial General Intelligence systems are capable of performing general tasks. However, the development of Artificial General Intelligence systems is still in its infancy. As yet, these independently thinking, independently reasoning, independently learning and possibly even self-conscious systems only exist in theory. Moreover, the expectations that experts have regarding whether and at which moment this type of intelligence will be realised differ widely.<sup>16</sup>

For more and more narrow AI systems, the technology is largely ready to use. The performance of these systems approach, or even exceed, that of man. In the last few years, the number of narrow AI applications has mushroomed. Examples are virtual assistants (oral replies to questions, such as Siri and Cortana), personal recommendation systems (as deployed by Netflix and Amazon) and pattern recognition in images (as deployed by medics for diagnosis purposes).

### **Opportunities, risks and current policies per public value**

This paragraph names AI opportunities and risks<sup>17</sup> for human rights-based public valued and describes existing policies per public value.

---

<sup>14</sup> Based on parliamentary papers II, 2017-2018, 22 112, no. 2578, p.2, adjusted to present-day understanding.

<sup>15</sup> Here, learning concerns calculating weights in a mathematical formula and is also called training.

<sup>16</sup> Müller, V. & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In *Fundamental Issues of Artificial Intelligence* (pp. 553-571). Berlin: Springer.

<sup>17</sup> Based on literature and reports such as: Vetz, M.J., Gerards, J.H. and Nehmelman, R. (2018), *Algoritmes en grondrechten*, Universiteit Utrecht. Kool, L., J. Timmer, L. Royakkers and R. van Est, *Urgent Upgrade. Protect public values in our digitized society*, Rathenau Instituut 2017, Van Est, R. & J.B.A. Gerritsen, with the assistance of L. Kool, *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality – Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)*, The Hague: Rathenau Instituut 2017, Raso, Filippo and Hilligoss, Hannah and Krishnamurthy, Vivek and Bavitz, Christopher and Kim, Levin Yerin, *Artificial Intelligence & Human Rights: Opportunities & Risks* (September 25, 2018). Berkman Klein Center Research Publication No. 2018-6. General, social opportunities of AI are mentioned in, among others: Jha, K. et al. (2019). A

### *Non-discrimination*

Non-discrimination prohibits unequal treatment of people in similar circumstances. Unequal treatment is particularly problematic if the one (individual/group or case) is disadvantaged in relation to the other without due justification.

### Opportunities and risks

AI offers opportunities for combatting discrimination. For instance, there are AI applications to reduce bias in selection processes (such as instruments that warn employers of potentially biased language in job descriptions). However, AI applications also entail non-discrimination risks, particularly in case there is bias in underlying data, bias in the algorithm or with incorrect classification.<sup>18</sup> Data bias applies when (unintentional) discriminating patterns occur. When an algorithm for staff recruitment is trained with a dataset in which (considerably) more men have been promoted, this may, for instance, result in the algorithm underestimating the chances of suitability for women. Algorithm bias applies when experts' (subconscious) bias is translated into the algorithm design or in the choice of variables. Here, too, the bias may result in unjustified disadvantage of groups. Moreover, classifying AI systems by definition have an error margin. This is because these systems apply observations about groups to the individual. As a result, there are always cases that – according to the model – come under a certain category, but actually do not belong there. This means that people can be misclassified; for instance unjustly as creditworthy (false positive), or unjustly *not* as creditworthy (false negative).

### Existing policy

Various authorities are working on standards to safeguard the quality of AI systems, also with regard to bias and error margins. For instance, the international Institute of Electrical and Electronics Engineers (IEEE) is working on standards to prevent undesirable bias in data and algorithms.<sup>19</sup> Moreover, the Dutch standardisation institute NEN founded a standardisation committee Artificial Intelligence and Big Data at the end of 2018. Through this committee, NEN aims to exert influence on ISO standardisation, among others with respect to limiting bias in AI systems, risk management in AI, reliability and robustness of AI.<sup>20</sup> The Dutch Ministry of Justice and Security is simultaneously issuing a letter about possible legal safeguards from risks of data analyses by the Dutch government and guidelines for the application of algorithms by governments. This letter also addresses the possibility to include quality assurance standards in legislation and directives. Under the direction of my ministry, work is being carried out on ethical

---

comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*. 2 (June). 1-12. Topol. E. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 25. 44-56. Zhang et al. (2016). A study on key technologies of unmanned driving. *CAAI Transactions on Intelligence Technology*. 1 (1). 4-13.

<sup>18</sup> Regarding these risks, it should be noted that people are not exempt from bias either and may sometimes (subconsciously) discriminate as a result. One important risk, however, is that AI systems can replicate and automate this.

<sup>19</sup> <https://standards.ieee.org/project/7003.html>

<sup>20</sup> <https://www.iso.org/committee/6794475.html>



data principles for the government, which include discrimination prevention.<sup>21</sup> These data principles are tested in the Transparency Lab initiated by my ministry, along with other principles and guidelines. Regional authorities likewise seek to increase their control on AI systems. For instance, the municipality of Amsterdam is examining whether AI system quality can be screened.<sup>22</sup> The Dutch Inspectorate SZW commissioned a study of automated systems used in recruitment and selection procedures, the attendant risks and possibilities for monitoring. Next to authorities, businesses are working on prevention of bias and error margins. For instance, IBM developed the AI Fairness 360 toolkit to detect bias in datasets and algorithms.<sup>23</sup>

#### *Privacy and data protection*

In the Netherlands, everybody is entitled to a private life and to respect for privacy. Individuals can be uninhibitedly themselves and do as they please, without interference from third parties. For AI systems, the protection of personal data is particularly important, as AI systems often depend on data about persons to run.

#### Opportunities and risks

AI applications can contribute to the enhancement of privacy. For instance, when they are deployed to evaluate corporate privacy statements on the extent to which they comply with the General Data Protection Regulation (GDPR). Such a programme can help businesses improve their privacy policy, can support people who are standing up for their privacy interests and can inspire institutions to take action.<sup>24</sup> However, in the area of privacy risks are also in evidence, as AI systems often use large quantities of data to generate answers, predictions and patterns. The use of more data potentially yields better outcomes and new insights. This may urge organisations deploying AI systems to collect as much data as possible, resulting in increased invasions of privacy. Moreover, AI systems can chart highly personal characteristics of individuals on the basis van combinations of (non-sensitive) data. For instance, scientists could predict, among other things, sexual orientation and political preferences on the basis van Facebook-likes.<sup>25</sup> Some AI systems are capable of recognising emotions.<sup>26</sup> Such highly sensitive data – which

<sup>21</sup> Parliamentary papers II, 2018/19, 26 643, 597.

<sup>22</sup> <https://fd.nl/ondernemen/1291305/amsterdam-wil-eerlijke-computers-in-de-stad#>

<sup>23</sup> <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/>

<sup>24</sup> See: CLAUDETTE meets GDPR Automating the Evaluation of Privacy Policies using Artificial Intelligence 2018, <http://www.claudette.eu/gdpr/#>, and Hamza Harkous c.s., Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning, <https://www.wired.com/story/polisis-ai-reads-privacy-policies-so-you-dont-have-to/>

<sup>25</sup> Kosinski, M., Stillwell, D. and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior, Proceedings of the National Academy of Sciences of the United States of America, 110 (15) 5802-5805.

<sup>26</sup> Cambria, Erik (March 2016). "Affective Computing and Sentiment Analysis". IEEE Intelligent Systems. 31 (2): 102–107, Sharef, Nurfadhlina Mohd; Zin, Harnani Mat; Nadali, Samaneh (1 March 2016). "Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data". Journal of Computer Science. 12 (3): 153–168.

were not knowable for third parties before – may become knowable. This, too, results in invasions of privacy.

#### Existing policy

For the development and deployment of AI systems several Privacy by Design principles, frameworks and manuals are in place. For instance, the Dutch Centre for Information Security and Privacy Protection developed a manual to help organisations to apply Privacy by Design principles.<sup>27</sup> Meanwhile, there is a growing number of technical possibilities to process data without invading on privacy, for instance through the application of secure multi-party computation, and self-sovereign identity. Research also reveals that future AI systems may need less data to arrive at a high-quality outcome.<sup>28</sup> The Dutch Cabinet supports this type of research. Facial recognition is an AI application that requires specific attention, as citizens make more and more use of it, through apps and social media among other uses. The aforementioned study that is being performed by Tilburg University, will also address measures to limit risks and is expected to be ready by the end of this year.<sup>29</sup>

The use of personal data for AI systems also confirms the necessity to have digital security in place. After all, organisations working with personal data may be subject to hacker attacks by cybercriminals or State actors. To counter this risk, increasing digital resilience must be aimed for, apart from combatting the threat. To that end, the Dutch Cabinet has announced measures in the Dutch Cyber Security Agenda and, as announced in the policy response to the Cyber Security Assessment Netherlands 2019, is targeting its efforts at further enhancement of the digital resilience of vital infrastructure and central Dutch government.<sup>30</sup>

#### *Freedom of expression*

Everybody has the right to voice convictions, feelings and opinions and share them with others. This also includes the right to access to information.
--

#### Opportunities and risks

Through the use of AI, news websites, search engines and recommendation systems personalise information. This helps users find, share and assess relevant information, which may further the right to information and the freedom of expression. However, this entails risks too. The use of AI systems in supplying information may result in individuals receiving a limited information offer. For instance, search engines tend to put popular sources, English-language websites and commercial information sources in a more dominant (higher) position in search results.<sup>31</sup> Personalisation may also cause users to be shown mainly

<sup>27</sup> <https://www.rijksoverheid.nl/documenten/rapporten/2018/02/19/handleiding-privacy-by-design>

<sup>28</sup> <https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more>

<sup>29</sup> Parliamentary papers II, 2018-2019, 34 926, no. 8.

<sup>30</sup> Netherlands Cyber Security Agenda (Parliamentary paper 26 643, no. 536), Cyber Security Assessment Netherlands 2019 and NCSA Progress Report (Parliamentary paper 26 643, no. 614).

<sup>31</sup> T. Gillespie, 'The Relevance of Algorithms', in: T. Gillespie, P. J. Boczkowski & K. A. Foot (ed.), *Media technologies: Essays on communication, materiality, and society*, Cambridge: MIT Press 2014, p. 167-194.

information that fits in with their views and thus reinforce those views (*confirmation bias, filter bubble*). To a limited extent, AI systems can be used to remove criminal and unlawful content. AI can play an important role in detecting child porn; for the detection of terrorism, discrimination and unlawful content – where the context in which the content is presented plays a particularly important role – human intervention is still necessary. In those cases, algorithms can lead to alerts, which can subsequently be assessed by a human assessor. The Cabinet supports initiatives to develop algorithms that can pack more nuance.

#### Existing policy

The effects of AI on the freedom of expression were already acknowledged by the former Dutch Cabinet.<sup>32</sup> Earlier on, that Cabinet described the increasing media personalisation (among others, through the use of AI) and the increasing share of the population consuming news through social media. However, a recent study<sup>33</sup> reveal that as yet no large negative impact of news personalisation is evident in the Netherlands, partly because of its strong media landscape and pluriform news consumption. The risk of online ‘filter bubbles’ in which people are one-sidedly informed is relatively small. To further minimise risks, the Dutch Cabinet is investing in, among others, media wisdom and public campaigns (such as [www.blijfkritisch.nl](http://www.blijfkritisch.nl)) to make the public aware of the role personalisation can play in the news offer it gets to see. In May 2019, the Dutch Council for Public Administration issued a Dutch-language advisory report, the title of which translates as ‘Seeking for truth’. The report describes the influence of digitisation on democracy from a truth-finding perspective. My ministry is preparing a response to this report, due before the end of the year. As regards combatting illegal online content, the Netherlands has adopted a ‘Notice-And-Take-Down’ procedure. Through this procedure, citizens and businesses are encouraged and/or obligated to remove illegal online content from their platforms as soon as possible. Although the removal of illegal content by private parties may be acceptable when this is necessary to protect key public interests, it is important that safeguards are provided. The European Commission has recommended measures to prevent unintended removal of legal content.<sup>34</sup> For instance, a content provider should be notified of removal and be able to contest removal. Moreover, businesses must build in effective and appropriate safeguards to prevent improper curtailing of the freedom of expression as much as possible.

#### *Human dignity*

Simply ‘being’ human is associated with a certain dignity for which a certain level of protection in relation to the government and third parties must be safeguarded. Human dignity is mainly applied in the context of physical and psychological

<sup>32</sup> Parliamentary papers II, 2016/17, 32 827, no. 116.

<sup>33</sup> See, among others, Rathenau, 2018, ‘Digitalisering van het nieuws; online nieuwsgedrag, desinformatie en personalisatie’ by the Rathenau Instituut and the Dutch Media Authority, annex 3 to parliamentary papers II 2017-2018, 32827, no. 127.

<sup>34</sup> Parliamentary papers II, 2017/18, 22112, nr. 2420.

integrity, individual autonomy, access to court, material living conditions and equality.

#### Opportunities and risks

AI applications can contribute to human dignity, for instance when they take over dangerous or humiliating work. On the other hand, AI can also result in risks when it substitutes interhuman contact. For instance, personal digital assistants and chat bots having contact with people beg the question whether it should always be clear for people that they are communicating with a person or a machine. AI systems that are deployed for staff recruitment beg the question to what extent applicants are entitled to human contact. In such cases, the question arises how human dignity can be safeguarded.

#### Existing policy

In his agenda 'NL DIGIbeter',<sup>35</sup> the Dutch State Secretary of the Interior and Kingdom Affairs states that meaningful contact with the government is the starting point for digital government policy. The aim is for government information and services to become more accessible, better understandable, intended for everyone and more personal – and not only digital.<sup>36</sup> As regards AI developments and meaningful contact between citizens and businesses, the Dutch Cabinet has asked the Netherlands Scientific Council for Government Policy (WRR) to advise, among others, on the impact of AI on human contact. The WRR study started in the autumn of 2018 and is expected to be completed in 2020.

#### *Personal autonomy*

Personal autonomy means that a person is free to make choices and can decide largely independently how he or she arranges life. In a constitutional state, the individual is to have room to consider for himself; where he lives, how he lives, which faith he does or does not adhere to, which education or training he chooses, which job he accepts, etc.

#### Opportunities and risks

AI applications can enhance autonomy. For instance, AI-based digital coaches help people improve their quality of life, and AI-based apps are available that help people make a medical diagnosis independently. However, there are also AI applications that restrict people's autonomy by imperceptibly influencing them. This is particularly the case when systems present persons with restricted or manipulative information and/or choices, or (whether this is noticeable or not) nudge them in the direction of specific preferences, choices and behaviours.

#### Existing policy

<sup>35</sup> Parliamentary papers II, 2017/2018, 26 643, no. 549.

<sup>36</sup> Parliamentary papers II, 2018/19, 26 643, no. 583.

At the moment, the Dutch government does not deploy AI to influence behaviour.<sup>37</sup> There are, however, clear starting points in place for influencing behaviour as a government instrument. In earlier briefs, our Cabinet has indicated that with respect to influencing behaviour (such as promoting healthier life styles), the government must always take the normative considerations and the aspects of rule of law into account (legality, proportionality and sound administration).<sup>38</sup> Also, the government must be transparent regarding the policy targets of influencing behaviour. Businesses do deploy AI more and more to influence behaviour.<sup>39</sup> The Dutch Cabinet is aware of this development and of the corresponding risks, and it intends to study the effects (particularly those of persuasive technologies) on autonomy.

#### *Procedural rights*

For a constitutional state to function properly, everybody needs to have access to justice. People must have access to information, advice, guidance in negotiation, legal assistance and the possibility of a ruling by a neutral (judicial) authority.

#### Opportunities and risks

In the legal process, the use of AI offers opportunities to increase efficiency.<sup>40</sup> For instance, the Dutch Public Prosecution Department is carrying out an experiment to analyse jurisprudence by means of AI in order to prepare a public prosecutor for a case or court session (the project name translates as 'Jurisprudence robot'). On the other hand, AI applications can also entail risks to procedural rights. Algorithms that are applied in AI systems can be complex. As a result, their explanation, or the control of their functioning is difficult, or sometimes not possible at all. Moreover, in some cases, algorithms are kept secret for commercial reasons. That makes it harder for citizens to object against the functioning or outcome of these AI systems.

#### Existing policy

With regard to policy, there are several initiatives.<sup>41</sup> The aforementioned brief on potential legal safeguards against the risks of data analyses by the Dutch government and guidelines for the application of algorithms by authorities issued by the Dutch Minister for Legal Protection is also relevant in this context. This brief provides guidelines for transparency of algorithms for authorities. The Transparency Lab set up by my ministry will assess the guidelines and will subsequently operationalise them further in an online application that supports the implementing user. Moreover, my ministry is working on a code for good

---

<sup>37</sup> The Dutch-language TNO study, the title of which translates as 'Quick-scan AI in public services' reveals that nearly all studied government applications of AI are in predictive analytics.

<sup>38</sup> Parliamentary papers II, 2014/15, 34 000 XIII, no. 140.

<sup>39</sup> Kool, L., J. Timmer, L. Royakkers and R. van Est, Urgent Upgrade. Protect public values in our digitized society. The Hague, Rathenau Instituut 2017.

<sup>40</sup> Parliamentary papers I, 2018/19, 34775-VI, no. AH, The basic principle of the Dutch Ministry of Justice and Security regarding AI in jurisdiction is that experiments are taking place incrementally, as yet outside the context of real, pending court cases or disputes.

<sup>41</sup> The Dutch-language letter on AI and algorithms in jurisdiction elaborates in more detail legislation and policy with regard to AI in judicial procedures.

digital governance ('Code Goed Digitaal Openbaar Bestuur'), which, among others, will enter into the transparency principle. This Code will be included into the existing code for good governance ('Code Goed Openbaar Bestuur').

### **General policy to safeguard public values and human rights**

Apart from the policies per public value described above, more general policies have been put in place to safeguard public values and human rights in the context of AI developments. This paragraph will list various policies.

#### *Dialogue and media wisdom*

With regard to technological developments it is important that citizens know what those development signify and what they can do to make use of opportunities and address risks to rights. That is why our Cabinet is launching awareness campaigns. One example is 'www.blijfkritisch.nl' (which translates as 'stay critical') that alerts citizens to (AI-generated) online disinformation.<sup>42</sup> Also, our Cabinet is organising dialogues with citizens about AI developments. During the Dutch Digital Conference<sup>43</sup>, for instance, there was a citizens' dialogue about the opportunities and risks AI poses for fundamental rights. Likewise, the Cabinet is developing educational material for specific target groups. My ministry issued a special edition of the Dutch children's magazine Donald Duck about AI and fundamental rights. The Dutch Ministry of Education, Culture and Science is generally pursuing policies to enhance media wisdom among young people. In March last, this ministry launched the digitisation agenda for primary and secondary education, which addresses various aspects of digital literacy (a combination of media wisdom, basic ICT skills, computational thinking and information skills).<sup>44</sup> This forms part of a curriculum revision that education professionals are currently working on in the Dutch-language website [www.curriculum.nu](http://www.curriculum.nu)

#### *Self-regulation and conduct codes*

Our Cabinet expects the business sector to comply with the UN Guiding Principles on Business and Human Rights.<sup>45</sup> Among other things, this means that, when designing or purchasing AI applications, businesses must carefully consider possible negative effects on human rights and take measures to prevent or mend those effects.

One instrument that may be helpful is self-regulation. Our Cabinet is very much in favour of businesses acting on their own responsibility in this regard. Self-regulation is possible thanks to instruments such as codes of conduct, certification and binding operating rules. In recent years, various initiatives have emerged on both national and international level. For one thing, during the Dutch Digital Conference, the Netherlands presented the Ethical AI Code.<sup>46</sup> This code is based on the Guidelines for Trustworthy AI developed by the European High Level

<sup>42</sup> <https://www.rijksoverheid.nl/onderwerpen/desinformatie-nepnieuws>

<sup>43</sup> <https://www.nederlanddigitaal.nl/conferentie-nederland-digitaal>

<sup>44</sup> Parliamentary papers II, 2015/16, 33 009, no. 13.

<sup>45</sup> [https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr\\_eN.pdf](https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf)

<sup>46</sup> <https://www.nederlandict.nl/ethischecodeai/>

Expert Group On Artificial Intelligence (HLEG).<sup>47</sup> Among others, it contains principles for safeguarding public values and being transparent regarding the use and the functioning of AI. Currently, the application of the European ethical guidelines is in its piloting phase, and the results of this pilot will be presented early in 2020.<sup>48</sup> The Dutch Cabinet continues to monitor the developments closely to see if further steps need to be taken to safeguard respect for human rights by the private sector.

#### *International agenda-setting and negotiation*

The cross-border character of AI developments induces an international approach. In the international playing field, our Cabinet commits to sustainable and effective safeguarding of human rights. For instance, the Netherlands actively dedicates itself in various bodies within the Council of Europe to the development of policy instruments that aim to safeguard human rights in the context of ongoing digitisation. At a European level, the Netherlands is also collaborating on activities flowing from the European Commission's AI strategy set down in the 'Coordinated Action Plan'.<sup>49</sup> This plan starts from a human-centred AI development pivoting around values that are important to people. When presenting the aforementioned ethical guidelines, the European Commission made an additional announcement.<sup>50</sup> For the 2019-2024 period, government leaders have indicated that the EU is to develop policies that incorporates our societal values, promotes inclusiveness and continues to be consistent with our way of life. Where AI and security are concerned, the Dutch Cabinet underlines that multilateral consultation prevails. Particularly the application of AI in weapon systems and the resulting ethical issues, is an important point of international negotiation. Since 2013, it is on the agenda of the United Nations Convention on Certain Conventional Weapons. In this point, the Netherlands advocates that the development and application of weapon systems should be under meaningful human control at all times and that its application can only take place in accordance with international law of war.

---

<sup>47</sup> The HLEG has as a general objective to support the implementation of the European strategy on AI. This also includes the development of ethical, legal and societal issues with regard to AI.

<sup>48</sup> <https://ec.europa.eu/futurium/en/register-piloting-process>

<sup>49</sup> <https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence>

<sup>50</sup> Parliamentary papers II, 2018/19, 22 112, no. 2799.