SWP Comment
NO. 23 MAY 2018

The Global Debate on the Future of Artificial Intelligence

The Need for International Regulation and Opportunities for German Foreign Policy

Marcel Dickow and Daniel Jacob

With the current developments in the field of artificial intelligence, the process of digitalisation has reached a new stage. Artificial intelligence makes it possible to analyse the large amounts of data collected today in completely new ways. Companies and countries are spending considerable resources to take advantage of these analytical possibilities. However, artificial intelligence is also dependent on the quality of the underlying data; it is completely unsuited for many tasks and has, so far, largely escaped human control. Germany should therefore use its influence in international forums to regulate the use of artificial intelligence in politically sensitive areas. In addition, the Federal Government should carefully examine on what data basis, for what purposes, and under what conditions artificial intelligence can make a contribution to the planning of foreign policy strategy.

Artificial intelligence (AI) is entering more and more areas of our lives. Companies use AI to create profiles of their customers and evaluate applications. In the medical field, progress is anticipated in both research and therapy. States, too, are increasingly relying on AI: In the context of "predictive policing", they aim to enable police and intelligence services to identify signs of crimes even before they are committed. Autonomous, AI-based weapons systems are supposed to make new forms of warfare possible. In the near future, it is also to be expected that states will use AI systems for the formation of their foreign policy strategies, for example through the real-time evaluation of economic data from other states.

The information processing by AI should make it possible in the future to separate the wheat from the chaff in the multitude of data. The actual decisions will probably remain with human beings. For now. When it comes to speed and analysis capacity, human beings are increasingly falling behind compared to machines. Leaving the decision to the "perfect" machine is a tempting idea for time-critical applications, for example. However, as experiences with autonomous driving show, in the end — as with any new technology — human beings will need to intervene in a regulatory man-

Stiftung Wissenschaft und Politik
German Institute for International and Security Affairs

ner in order to avoid unwanted risks and define responsibilities.

A particular challenge is that AI, as a technology, is not only subject to the regulatory efforts of politics. To the extent that public institutions use AI systems, AI will actually influence the political process itself. Legislative institutions, supervisory authorities, and international bodies therefore face the challenge of having to regulate portions of their own analysis and decision-making instruments.

Artificial Intelligence

The term "artificial intelligence" has developed into a collective term for a series of computer-based methods. Many of today's AI applications are based on data-driven, machine-learning processes. These methods require large amounts of pre-structured data that are used to teach algorithms. Machines thus learn to classify, for example to distinguish dogs from cats in photographs. The amount, quality, and representativeness of the training data are decisive factors for the informative value of the results.

Strong and Weak AI

In certain cases, data-driven AI systems can deliver better classification results than humans. However, there remain special solutions for particular problems that have not yet allowed for generalisations. Even the recognition of wolves overwhelms a system that has been trained with photos of dogs and cats. Due to these inherent limitations, the AI systems commonly used today are also referred to as "weak" AI systems. In the informatics and philosophy sphere, there are voices repeatedly emphasising that, in the future, machines could also be equipped with more highly developed human characteristics, for example with the ability to think conceptually. For this "strong" AI system, however, no implementation concepts are yet available.

How Machines Learn

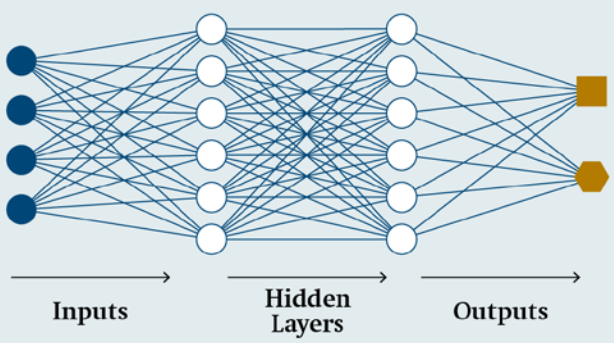
One of the dominant methods of machine learning today is so-called deep learning with neural networks (deep neural networks). These networks work with several layers of branching points. The individual layers each capture only simple concepts, but their combination allows for complex analyses. Image recognition, for example, captures a number of characteristics that are not very meaningful in themselves, but which, in combination, make it possible to distinguish between dogs and cats, for example.

Deep neural networks have a large number of such hidden layers in which probability values are formed during training (see Figure 1). To be sure, input and output are still in a defined mathematical relationship to each other. However, this ratio is not based on logical conclusions, as is customary for humans, but on statistical calculations. Such nested classification methods can intuitively derive simple rules for humans from observed behaviours that are difficult to programme using classical methods. However, this comes at a price: The analysis results of such deep neural networks cannot be validated without (time-consuming) mathematical methods.

So-called reinforcement learning makes use of the structure of deep neural networks. Algorithms are programmed to develop strategies to solve problems by

Figure 1

Schematic representation of the structure of a neural network



Source: Own presentation.

maximising success (indicated by rewards). The system is given an abstract goal that it achieves through a process of self-learning, that is, "trial and error". Reinforcement learning is now an important factor in complex control tasks, for example in the control of robotic systems. However, the inherent uncertainty as to how the AI system solves the task considerably limits the areas of application.

Research on AI

Significant progress has been made with AI since the early 2010s. This was made possible by the increasing availability of structured data on the internet and the rise in the computing power and storage capacity of modern computers. The networking of more and more devices in the context of the "Internet of Things" will further increase the availability of usable data.

The development of AI is mainly driven by large internet companies. They invest considerable sums and increasingly compete for qualified personnel. For this reason, some of the leading companies in the United States publish selected research results and make individual software components available to the scientific community as "open source".

In the meantime, many countries have jumped on the bandwagon and launched state funding and development programmes. China, in particular, has announced substantial investments and has declared its intention to catch up with the technological lead of the United States. Although Russia clearly falls short in comparison, it has also clearly articulated its ambitions in this field. The announcement of a Franco-German AI centre can also be seen as a reaction to this technological competition. So far, however, the extent of government involvement in this area is difficult to assess: Even in democratic states, there is a lack of reliable figures that could provide information on the actual volume of investment. In countries such as China and the United States, things are even more difficult because their cooperation with companies is largely out of the public eye, especially when it comes to the use of AI by militaries.

The Need for International Regulation

The specific strength of AI systems is also the reason why they require regulation: Across the various applications, AI is characterised by being able to evaluate data much faster than humans, and with an increasing degree of independence. This ability holds the promise of making better-informed decisions. As a result, however, the responsibility for far-reaching decisions is transferred to computer systems, the proceedings of which are still difficult for humans to understand. AI is therefore sometimes referred to as the "black box".

This does not have to be problematic in every case. However, there is a need for regulation wherever AI systems are used in politically sensitive areas. The different norms of international law provide orientation for the level of international regulation. If Germany wants to maintain the efficacy of these standards, it is necessary to adapt them to the special challenges of AI systems.

AI and State Coercion

In the context of a state's use of AI systems, human rights standards are a first important starting point. They draw attention to those applications in which the use of AI systems is combined with governmental coercion. Already today, countries such as the United States and the United Kingdom use AI systems in their police work: Within the framework of predictive policing, predictions are made as to where, when, and from whom criminal acts can be expected. AI systems are also already being used in the United States to support court rulings, for example in the determination of penalties. The culmination of these developments can be observed in China: There, the government is currently introducing a "Social Credit System", with the help of

which the behaviour of citizens in almost all areas of life is to be recorded and automatically translated into a rating that is reinforced by sanctions.

It is still largely unclear whether and, if so, how a state's use of AI systems can be reconciled with the requirements of human rights. This question arises in particular for the right to equal treatment and due process (Article 14 of the International Covenant on Civil and Political Rights [ICCPR]) and the right to privacy (Article 17 ICCPR). An EU directive on the protection of personal data in the context of "the prevention, investigation, detection or prosecution of criminal offences or the execution of sentences" (EU 2016/680) provides initial guidance for these matters.

The use of AI systems by militaries falls primarily within the domain of international humanitarian law. Many countries are already investing heavily in the development of (partially) autonomous weapons systems. Even the partial transfer of decisions concerning the use of violence to AI systems, however, would fundamentally call into question the protective mechanisms of international humanitarian law. In response to these developments, within the framework of the Convention on Certain Conventional Weapons (CCW), a group of governmental experts was formed to develop proposals on how autonomous weapons systems could be regulated.

AI and the Power of Companies

There is also a need for regulation with regards to the use of AI systems by private companies. Again, the human rights norms of international law provide orientation: States are also responsible for protecting the relevant legal interests from violations by third parties. Because most of the companies in this field operate transnationally, national measures must be supplemented by appropriate forms of international regulation.

Private companies such as Amazon, Google, and Baidu use AI systems to create extensive profiles of their (potential) customers. These profiles are often used for the comparatively harmless personalisation of advertising and the provision of digital services, such as the display of personalised messages. However, personal profiles can also form the basis for decisions with far-reaching consequences: For example, banks are increasingly using AI systems to assess the creditworthiness of their customers. Companies such as LinkedIn also use such systems to evaluate the profiles of job seekers. Finally, AI systems are an essential element in the current development of autonomous transport systems by companies such as Tesla, Google, and Baidu.

Although this form of corporate use of AI systems does not involve the coercive power of the state, it has a significant impact on the lives of an increasing number of people. It is the responsibility of states to regulate this use in such a way that it complies with human rights requirements, such as the prohibition of discrimination (Article 26 ICCPR) and the right to privacy (Article 17 ICCPR). Remarkably, some of the most important companies themselves are calling for a societal debate on the regulation of AI systems. Leading representatives of companies such as Deepmind, Apple, and IBM have participated in the development of the "Asilomar AI Principles". The principles, conceived in Asilomar, California, in 2017, formulate social requirements for the future development of AI systems. They also point to the need to create an appropriate legal framework for the use of AI.

Regulatory Goals

In order to meet the regulatory requirements outlined above, specific solutions are required that take account of the technical and political particularities of each case of application. Despite the diversity of applications, however, some overarching regulatory objectives can be identified.

"Garbage In, Garbage Out"

The general objective is to avoid "false" results when using AI systems to support

decisions made by machine, to critically examine the informative value of the results, and to limit the negative effects of false analyses. As described, the performance of today's AI systems is based on the inductive evaluation of data. The quantity and quality of the available data therefore directly determine their performance. Not surprisingly, there have recently been a number of reports about the risk of AI discrimination. The central problem was usually that AI systems reproduce existing discrimination in the data sets. For example, if historical data reflects that men have held more senior positions than women, an AI system may erroneously be led to the conclusion that men are generally better suited for such positions.

Data selection therefore plays a central role when using AI systems. However, the systems themselves cannot provide any information about the quality of the data on which they are based. On the contrary, even if the data basis is insufficient, they will always produce a result, even if it is not reliable. Another complication is that more subtle forms of data-based discrimination are difficult to identify. It may still be comparatively easy to block an AI system that evaluates applicants from accessing information on their gender. However, this does not exclude the possibility that other factors correlating with gender may be included in the analysis. The less obvious this correlation is, the more difficult it becomes to identify it from the results alone.

It is therefore a central responsibility of the developers and users of AI systems to carefully examine whether the available data are appropriate for the respective purposes of the analysis. However, because this thoroughness cannot always be assumed, regulatory intervention is required in politically sensitive areas in three regards.

Transparency

As an elementary prerequisite for a critical examination of the effects of AI systems, a sufficient degree of transparency must first be established. To this end, it is essential that public and private institutions provide information on where — and for what purposes — they are using AI systems. On this basis, affected persons must be able to understand how — and on the basis of which data — AI systems affect central aspects of their lives. Political decision-makers must also be informed in an understandable way about how public bodies and private actors are using AI systems.

Defining this transparency requirement more precisely is a technical and legal challenge. Technically, it is by no means trivial, but nevertheless possible, to make such information available in an understandable way. AI systems should be designed to give information on the critical analysis parameters, to disclose how representative the training data is for the real data applications, and to indicate the error rates associated with their analyses. Legally, it must be clarified how to reconcile access to meaningful information with data protection and the right of companies to protect their intellectual property. Last but not least, it is possible that the persons affected themselves, or interested third parties, may use such information to manipulate the databases of AI systems in a targeted manner.

The first steps towards solving this complex problem are offered by the European General Data Protection Regulation (GDPR), which will become binding in May of this year. The GDPR explicitly regulates the information and appeal rights of data subjects in connection with automated data analyses (Article 22 GDPR). Also of interest in this context are the efforts of the city council of New York City to make the municipal use of AI systems more transparent. At the end of 2017, the City Council decided to establish an expert group to investigate the use of automated decision systems in New York's public administration. The group is due to start work in the next few months.

Human Control

The attraction of AI lies in delegating activities previously carried out by humans

to machines. In politically sensitive areas, however, it is necessary to examine the areas in which the operation of AI systems should be subject to particular human control.

The classification methods of modern AI systems are often used to reveal previously hidden statistical correlations. However, such data correlation can neither presuppose nor prove causality. It is therefore essential to subject AI-based decisions to human plausibility checks. This requires the transparency of the data and algorithms already described. In addition, sufficient time must be available for such a plausibility check and to be able to reconstruct the procedures of AI systems. In fact, in many cases, it is also legally sensible to demand such a comprehensive plausibility check. In cases where decisions are made with far-reaching consequences for those affected, human control cannot be limited to operating a "yes/no" dialogue box on the screen. Rather, it is necessary that the users of such systems actively deal with the system and application parameters.

A fictitious example illustrates the challenge: A company offers AI-based software that analyses satellite images and provides proposals for deployment planning during a German military mission abroad. In a first step, the data basis used for training the system must be checked for representativeness and applicability. This testing should not be left to the company offering the software. In such a politically sensitive area, this should instead be done by an independent body.

In a second step, the local commanders must confirm the results of the expert system in each individual case — or correct them, if necessary — after they have subjected the system to a documented inspection process. This requires that they have the necessary professional competence. It remains a question as to how this competence can be developed and maintained when situation assessments in the future will increasingly be based on expert systems.

Clear Responsibilities

The example of the use of AI in the planning of military missions also points to the need to define clear responsibilities for the use and certification of AI-based systems. International humanitarian law, for example, requires members of armed forces to carry out a legal and moral assessment prior to the use of force. Among other things, it must be determined whether the chosen military means are appropriate and necessary. According to the prevailing understanding of international law, human beings must make such considerations and must not leave them to machines. The proposal to grant AI systems a limited status as legal entities and to make them liable in this way is therefore not convincing.

Instead, the aim should be to clearly state the responsibilities of human beings for the use of AI systems. With regards to the various military and civilian applications, it is also necessary to clarify the responsibilities of the manufacturers of AI systems.

Artificial Intelligence in Foreign Policy Planning

It is to be assumed that AI systems will also change foreign policy planning and decision-making in the near future. The enormously advanced possibilities of data processing by AI make it possible to offer forecasts on the consequences of specific foreign policy decisions and to record their effects in real time. In a way, this can be seen as an extension of traditional intelligence analysis. However, automated analyses by AI systems now make it possible to process larger amounts of data and to link data of different types and origins. For example, such systems are expected to be used in the future for early crisis detection, but also in the event of pandemics or the analysis of global migration flows. For precisely this purpose, the Federal Foreign Office (AA) is currently setting up a data-based analysis system in the Department for Crisis Pre-

vention, Stabilisation, Post-conflict Care and Humanitarian Aid. According to the AA, the system will initially evaluate publicly available data on social, economic, and political developments. Beyond analysis, states will also try to use AI to strengthen their offensive and defensive cybersecurity capacities.

These applications of AI systems have the potential to improve the information base for foreign policy decisions. But here, too, support from AI systems can only be as good as the underlying data allow. In addition, it must always be questioned which analyses AI systems are suited for, and which analyses are not. Even the best AI cannot replace political judgment. Finally, the use of such systems must be transparent — at least for decision-makers in parliament and government — and must always be subject to adequate human control. However, this is made more difficult by the fact that foreign policy decision-making has traditionally not been open to public scrutiny; this is true in almost all countries.

Tasks for German Foreign Policy

The development and application of AI are still in their infancy worldwide. This is a political opportunity, as it provides an occasion to actively shape the future technological and political development. The abovementioned regulatory objectives provide a normative guideline for this. However, they must be further developed and specified with a view to the particularities of specific applications.

Clarifications of the associated ethical, legal, and political questions are a core task of the Bundestag as a centre of the democratic public. A comprehensive debate is needed on how, and under which conditions, AI is to be used in which areas of society. A debate must also be held about which forms of national and, above all, international regulation are suitable for effectively implementing the corresponding political requirements. To this end, it is essential that the relevant considerations and processes within the framework of the EU and the United Nations be taken into account in the German debate. It was recently announced that the European Commission not only wants to expand research funding, but that it is also working on guidelines for dealing with the legal and ethical challenges of AI. The International Telecommunications Union is also increasingly addressing this issue. At the United Nations, negotiations on the use of autonomous weapons systems continue within the framework of the CCW.

There is thus a considerable need for clarification. In this respect, it seems sensible to broaden the scope of the debate. The "Data Ethics Commission" envisaged in the current coalition agreement, for example, could offer an opportunity to supplement the necessary parliamentary debate with contributions from science, business, and civil society.

In addition to these fundamental clarifications, however, it is also necessary to be able to act today on the foreign policy issues relating to the use of AI. There is international recognition that, with the German Research Center for Artificial Intelligence (DFKI), Germany is well-positioned in the field of basic research. As in other areas of digitalisation, however, a unified foreign policy approach is made more difficult by the fact that the responsibility for AI matters is spread across a number of departments. The responsibility for research funding lies with the Federal Ministry of Education and Research, whereas the Federal Ministry of Justice is responsible for legal issues relating to the application of AI, and the Federal Ministry of Labour and Social Affairs is responsible for the effects on the labour market. The Federal Ministry of the Interior, the AA, and the intelligence services are primarily concerned with questions concerning the operational use of AI. Responsibility for the relevant debates in the various international forums is divided between the AA, the Federal Ministry for Economic Affairs and Energy, and the Chancellor's Office.

This broad institutional spread is in some ways appropriate to the thematic breadth,

because AI does indeed affect a variety of policy areas. However, if Germany wishes to exert its foreign policy influence on the future development of AI, stronger coordination is needed. The creation of the position of Minister of State for Digitalisation in the Chancellor's Office, for example, is a promising measure that could be the starting point for an AI group of departmental experts based in the Chancellor's Office.

The online version of this publication contains functioning links to other SWP texts and other relevant sources.

SWP Comments are subject to internal peer review, fact-checking and copy-editing. For further information on our quality control procedures, please visit the SWP website: https://www.swp-berlin.org/en/about-swp/quality-management-for-swp-publications/

SWP
Stiftung Wissenschaft und Politik
German Institute for International and Security Affairs

Ludwigkirchplatz 3—4
10719 Berlin
Telephone +49 30 880 07-0
Fax +49 30 880 07-100
www.swp-berlin.org
swp@swp-berlin.org

ISSN 1861-1761

(English version of SWP-Aktuell 24/2018)

Dr. Marcel Dickow is Head of the International Security Division at SWP.
Dr. Daniel Jacob is an Associate in the Global Issues Division at SWP.