Hertie School

Policy Brief

# Data science and AI in government

## Why public sector organisations need in-house data science and artificial intelligence expertise

24 January 2022

#DigitalGovernment
#DataScience
#AI
#MachineLearning
#DigitalLeadership
#DigitalTransformation

**Maximilian Kupi**, Centre for Digital Governance
**Prof. Slava Jankin**, Data Science Lab
**Prof. Gerhard Hammerschmid**, Centre for Digital Governance

An essential part of the digital transformation in the public sector is the application of data science and artificial intelligence. These technologies will enable the public sector to become more efficient, responsive, prescient, sustainable as well as fairer by, for example, helping to detect and predict important trends, simulating and evaluating policy alternatives, and personalising or automating the implementation of policies. Yet governments are reticent about using these applications, in no small part because they lack in-house data science and AI capacities. To overcome their dependence on outside expertise and build up their own data science and AI capacities, governments are advised to:

- Adapt recruitment practices and improve job attractivity for experts.
- Build communities of practice and centres of excellence.
- Collaborate with external experts and research institutions.
- Strengthen interdisciplinary and intersectoral networks.
- Hold government-sponsored competitions and hackathons.
- Centralise capacities but continue to expand the base.

# 1 Data science and artificial intelligence enable better public governance

Around the world, digital transformation is currently the most comprehensive reform movement in government. This transformation is much more than using laptops, tablets and smartphones in government offices or offering governmental services online. Driven by novel technologies, digital transformation is about the systematic collection, evaluation, linking and application of data, not only to allow governments to perform tasks more efficiently but also more smartly – to enable the more "intelligent" management of public services. Data science and artificial intelligence (AI) are helping to make this transformation a reality.

AI in government refers to a vast array of technologies but most practical applications focus on using various forms of machine learning, which train models to learn from data. Data science, in turn, links machine learning to databases and distributed computing to build scalable and robust public governance solutions. These solutions are designed to perform or augment tasks, better inform decisions and accomplish objectives that have traditionally required human intelligence, such as planning, reasoning using partial or uncertain information and learning. Data science-driven public governance solutions are guided by the legal and ethical considerations for algorithms and data. They contribute to policy- and decision-making through efficient prediction and attribution. They are embedded in the realm of the possible by considering organisational and political context.

Data science and AI are already contributing to substantially improve the following areas of government:

**Regulatory research, analysis and monitoring**
Data science and AI can help governments with all tasks related to collecting, monitoring and analysing data to augment public officials' decision-making capabilities and make them more evidence-based. The most recent and prominent application in this field is the analysis of policy measures for the containment of the Covid-19 pandemic, while another is the monitoring of traffic flows to better inform city transport planning.

**Public services and engagement**
Another big application area for data science and AI technologies in government are tasks that support the direct provision of services or facilitate communication with and participation of citizens in policy development. Examples are the use of chatbots and recommendation systems, the installation of intelligent platforms for participatory democracy, the provision of matchmaking services, for instance, to help the unemployed find suitable jobs or the transcription of important verbally transmitted information for people with hearing difficulties.

**Adjudication**
Data science and AI can also assist government officials in providing benefits to or adjudicating the rights of citizens. For example, artificial intelligence-based tools are used to classify trademark applications and assign design codes or automatically verify that submitted tax declarations are complete, freeing up time for civil servants to conduct more substantive work in the later stages of these processes such as personal communication with citizens.

**Enforcement**
All tasks related to identifying or prioritising targets of agency enforcement actions are another area that can benefit extensively from the use of AI and data science. Possible use

cases range from detecting fraud to monitoring social media behaviour or identifying the licence plate numbers of illegally parked cars.

**Internal management**
The final application area for AI and data science is to help agencies manage internal organisation and resources, such as human resources, procurement, ICT systems or other utilities. Here, the possible use cases include tools to optimise energy consumption, assist users with searching for digitalised documents or support the computer vision-based transformation of handwritten documents into a digital format.

Recognising this potential of data science and AI, Germany's previous federal government released a Data Strategy and updated its AI Strategy, tapping into the possibilities of these technologies and methods to better fulfil governmental tasks. Backed by 240 million euros in funding, all federal ministries are now to install chief data scientists and establish their own data labs to foster data-driven policy-making. In the spirit of a "learning, technology-promoting state," the coalition agreement of the new government also acknowledges the potential of data science and AI to foster innovation and improve government work. While this is definitely a move in the right direction, the German government, like most other governments, is at the very beginning in utilising the full potential of data science and AI for public policy. The next section focuses on the crucial aspect of building internal data science and AI expertise to fully benefit from embedding these technologies in government. Data science and AI competencies can be viewed as a specialised subset of the broader digital skillset.

# 2 Why applying data science and AI in government requires in-house capacity

As governments seek to rely on data science and AI for policy design and implementation as well as making their work more efficient, a fundamental challenge is how to generate the necessary methodological and technical capacity to identify, develop and deploy such solutions in a responsible way. At its core, this challenge boils down to the "make-or-buy" decision: An agency can self-produce goods and services needed to perform government tasks by hiring personnel and building its own infrastructure, or it can procure them from other, often private sector institutions. While specialised solution providers have greater expertise and lower production costs in theory, in practice solely relying on external data science and AI capacities has considerable downsides. These disadvantages can be attributed primarily to the following five dimensions:

**Guaranteeing data privacy and security**
First and foremost, the data that can potentially be used for public sector AI or data science applications is fundamentally different from data in the private sector: Unlike customers, citizens don't have a choice whether to share their data with their state. Also, most of the data that governments have on their residents are quite sensitive. Hence, many government agencies have strict internal guidelines on the use and sharing of data, which on top of more generally applying regulations like the General Data Protection Regulation, make it particularly hard for governments to collaborate with external suppliers to provide AI or data science-based solutions.

**Ensuring compliance and accountability**
The public sector's double-bind with respect to AI governance – regulating algorithms while also governing by algorithms – makes the application of AI and data science in policy-making a particularly delicate task. Furthermore, cases where automated systems adversely violate the rights of citizens, like in the case of the Dutch automated surveillance system for detecting welfare fraud, can seriously harm trust in the system and with it the government.

This ultimately can pose a threat to democracy. Much is at stake when developing AI-based solutions that are meant to faithfully encode legal and policy choices and ensure compliance with international regulatory frameworks, such as the one currently being developed by the European Union. Therefore, leaving the development of these solutions to private sector software engineers, who lack the necessary legal training and experience, risks leading to potentially detrimental consequences.

**Incorporating domain, political and organisational expertise**

In the honeymoon phase of (big) data science, the solution to most problems was seen in the capacity to analyse as large amounts of data as possible; the algorithms would ultimately make sense of it. Reality, however, has proven this conviction wrong: In the three-stage process that leads from data to insights to decisions, the final step has proven to be the most critical in practice. Due to the considerable complexity of moving from insights to decisions, where ethical considerations often must be met and creative combinations of several alternatives undertaken, humans are still considered superior to machines in this step. Furthermore, to make useful and valid (causal or attributional) claims based on data, analysts need to be able to make the right assumptions that typically are not directly testable and thus require domain expertise. In turn, effectively communicating these insights to decision-makers requires an in-depth understanding of political and organisational contexts. This level of domain, political and organisational expertise is rarely possessed by government outsiders.

**Assuring integrative and iterative development**

Developing useable digital solutions and, in particular, more complex ones such as those based on data science or AI requires an iterative development process with continuous integration of user and stakeholder feedback. To do this, the development engineer must know the respective agency's bureaucratic realities and enjoy a high level of trust among potentially sceptical key stakeholders. Furthermore, automated solutions need to constantly adapt to changing environments and practices to stay effective. For example, algorithmic enforcement tools need to be continuously updated to combat new modes of wrongdoing unearthed by agency staff and avoid an undue focus on past forms of misconduct. Technical expertise embedded in the respective government agency facilitates this type of integrative, iterative and continuous development process.

**Making informed procurement and collaboration decisions**

Undoubtedly, no government will possess the necessary resources to develop all data science or AI based solutions in-house. Nevertheless, a certain level of embedded expertise in these areas is required to know what is technically possible and feasible and make informed judgements about the quality of contractor-provided solutions. Moreover, staff familiar with the respective methods and technologies is more likely to have access to the network of external experts and institutions, such as research institutes or universities, to advise the government in the process. This, in turn, also promotes successful collaborations with non-commercial sources of technical capacity. By gaining access to external talent and expertise while simultaneously maintaining control and monitoring quality, such collaborations present a promising middle ground for the "make-or-buy" decision.

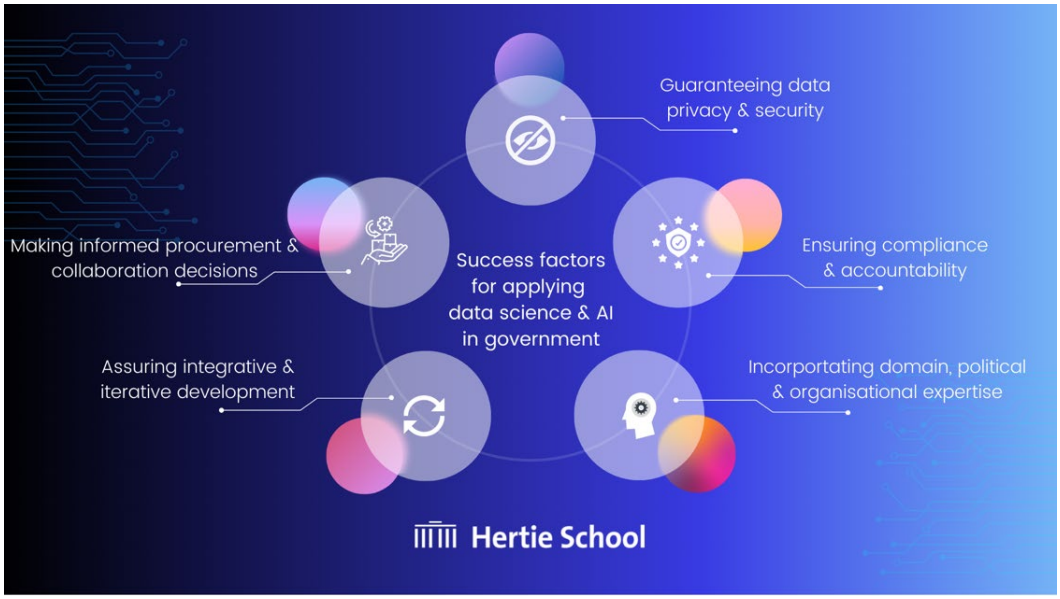


*Figure 1: Success factors for the application of data science & AI in government.*

In sum, this policy brief argues that for governments to seize the full potential of data science and AI for public policy, building up a certain level of internal methodological and technological capacity is indispensable. Incorporating data science and AI capabilities into respective public sector agencies will enable the development of governance tools that are better tailored to specific governance tasks, fully compliant with laws and regulations, and better aligned with complex organisational dynamics. This strategy has already proven to be useful in the US context, where a recent study found that over 50% of AI tools were being developed in-house by the respective government agencies. The following section formulates policy recommendations for building such capacity in government.

# 3 Recommendations for building data science and AI capacity in government

To understand how to build data science and AI capacity in government, it is first necessary to identify what concrete competencies these should include. For AI, the skillset consists of knowledge of machine learning and deep learning algorithms, including computer vision and natural language processing (NLP). For data science, the skillset consists of methods for dealing with big data, working with databases and distributed computing systems. Creating specific data science-based solutions in government (solution development) requires familiarity with causal analysis and decision theory, knowledge of human-centred design and agile management as well as of legal and ethical frameworks. Familiarity with the cycle of machine learning development, operation, deployment and maintenance ensures implementation of cost-effective and sustainable solutions. Finally, knowledge of governance and policy-making processes allows anchoring these competencies in the realm of the possible in the government context.

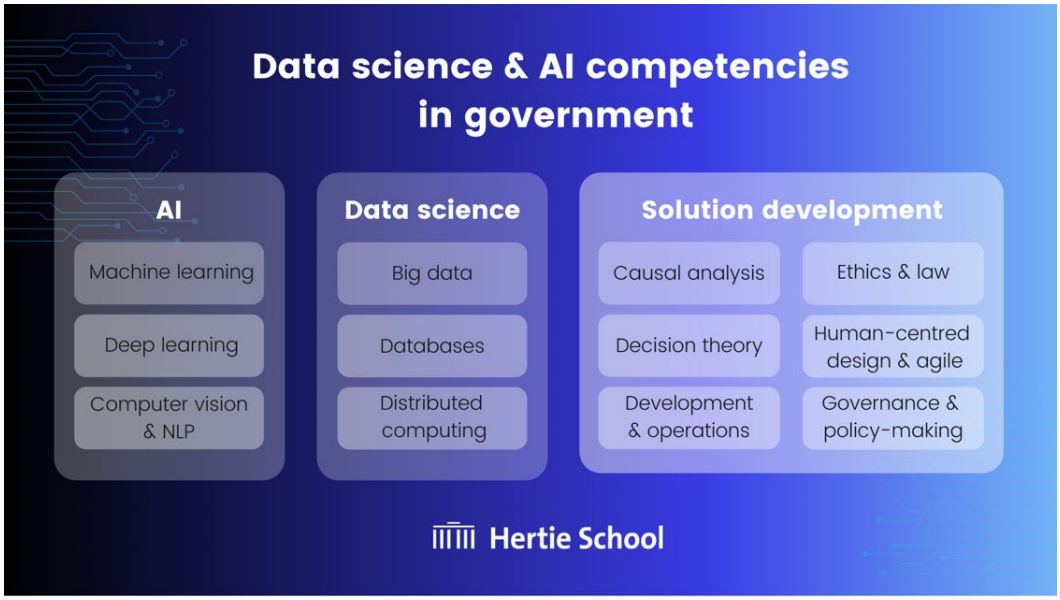


*Figure 2: Data science and AI competencies for building in-house capacities in government.*

**The following list of policy recommendations are suggestions for building and sustaining data science and AI capacity in government based on international best practices:**

**Adapt recruitment practices and improve job attractiveness**

Data science and AI specialists oftentimes do not follow the classical education and professional paths of public sector officials. Allowing for lateral entries into government, as done in Scandinavia, the US, UK or India, or entries for only a limited amount of time – as done by the Work4Germany and Tech4Germany fellowship programs – thus can be advisable to tap into this talent pool. Adapting the job classification scheme to include data science and AI-related job categories, as the US government has done, can also be useful. Furthermore, increasing the salaries and career prospects to better compete with comparable private sector job placements has proven a helpful step in the UK context. The UK even has set up a Digital, Data and Technology Profession Capability Framework, which is meant to support HR work in this field. Also, communication campaigns that raise awareness of the motivating, often socially relevant, challenges that working for the public sector can offer as well as the related benefits such as job stability and work-life balance are a further means to attract more talent. The German Foreign Intelligence Service's campaign #followtheglitchkarnickel as well as the German government's recent introduction of special allowances and increase in the salary grades of IT specialists are initial promising approaches in this area that need to be further developed and expanded for data science and AI experts. Ultimately, it is also highly advisable to invest in young talent and early career professionals working and studying at the intersection of data science and policy-making, who can offer fresh perspectives, the ability for growth and youthful enthusiasm in new technology that can steward new developments and tackle complex AI challenges.

**Establish communities of practice and centres of excellence**

For government internal data scientists and AI engineers to better learn from each other and exchange best practices, establishing communities of practice can be useful. Organised to cross boundaries within and across agencies, these communities can also help to gain legitimacy vis-à-vis the stakeholders, raise awareness for the potential of data science and AI, through initiatives such as jointly-produced (blog) articles, or foster collaboration amongst different agencies. Centres of excellence, on the other hand, can help to stay on the cutting edge of technology by providing leadership, best practices, research, support and training. Again, the UK government is a leading example in this area: It has not only built a

cross-departmental data science community with specialised sub-groups on topics such as text analysis, reproducible analytical pipelines and data ethics but also a plethora of other communities relevant for the digital transformation of government. Most inspiringly, all this doesn't happen behind closed doors but with a constant open communication, for example, on their own data in government blog or through a public government data science festival, so that others inside and outside the British Government can take notice and learn from their experience. Finally, the UK government's Data Science Campus centre of excellence works at the frontier of data science and AI to deliver a diverse set of projects for stakeholders across government and beyond.

**Collaborate with external experts and research institutions**

Collaborating with non-profit and educational institutions, such as universities and think tanks, can be an effective approach to bring external knowledge and expertise into a specific project context while maintaining control and monitoring quality. In addition, it may prove to be a viable way to attract young talent from these institutions or even build joint, tailor-made educational offers. Lastly, exchanges with external experts can help data science and AI practitioners inside government to stay up-to-date with the most recent technological developments. Best practice examples in this area come from the US context. The Intergovernmental Personnel Act Mobility Program, for example, provides for the temporary assignment of personnel between government agencies and universities or research centres for the US federal government. Furthermore, the Food and Drug Administration has established academic partnerships including an "Entrepreneur in Residence" programme as part of its Digital Health Innovation Action Plan or a Memorandum of Understanding with the Massachusetts Institute of Technology on collaborations in the areas of machine learning and data science. Also, the United States Postal Service has been cooperating with the University of Michigan to develop an automated mail delivery system, and the Environmental Protection Agency has been working with Stanford University as well as the University of Chicago to develop environmental enforcement tools.

**Strengthen interdisciplinary and intersectoral networks**

Networks of actors from public administration, business, politics, or research interested in topics related to the digital transformation in government and in particular data science and AI are a good way to foster intersectoral exchange and learning as well as open new areas for collaboration. On the other hand, it can also prove useful to build a public sector internal network as a "safe space" for officials to collect internal administrative challenges and experiences and adapt them to other departments. In the German context, NExT (public sector only) and NEGZ/N3GZ (intersectoral) are very promising examples of such networking initiatives and show that a lot is already going on. Looking at other similar networks, there is the public sector internal Dutch Government's Innovation Community that not only focuses on digitalisation related topics but also aims to connect various "islands" of innovation within the Dutch government and support cross department innovation communities. This example clearly shows the link between the establishment of networks and communities of practice which were highlighted above. However, while communities of practice are often perceived as more formal and professionalised, it is the voluntary and more open nature of networks that can help create unanticipated but highly profitable connections between actors.

**Hold government-sponsored competitions and hackathons**

Data science competitions or AI hackathons can help leverage public talent to generate and prototype ideas around declared government priorities. Often with prize money attached, these events are a potentially valuable source of innovation and an increasingly prevalent part of the capacity-building landscape. While the solutions generated certainly do not substitute for comprehensive AI strategies or automation endeavours, they can also help to raise awareness of data science and AI-related tasks in government and thus attract talent to the public sector. Unlike traditional programming hackathons, in which usually only a

problem is posed, hackathons focused on data science or AI often include governments providing the necessary data for participants to grapple with. The US government has developed a dedicated platform for all public competitions. These events are also becoming more widespread in the German context. Two recent national examples are #WirVersusVirus and UpdateDeutschland. The former is a hackathon for creative solutions that could help addressing the Covid-19 pandemic. The latter had a broader focus on "ideas with social impact" and was designed as a kind of match-making event between those with an idea and implementation partners. Building on these initiatives while strengthening a focus on data science and AI challenges thus is highly advisable.

**Centralise capacities but continue to expand the base**

When it comes to the organisation of digital capacities in government, many countries are inspired by the renowned example of UK's Government Digital Service (GDS) unit and strive to install similar centralised teams. Germany is no exception to that trend. The recently established DigitalService4Germany team is meant to increase the state's "digital capability to act" by bringing together small agile teams with government agencies to quickly develop user-centred digital solutions. This strategy can help create synergies and avoid reinventing the wheel for every project. However, when it comes to data science and AI-powered solutions, it is likely that the most transformative solutions will emerge from experiments on the front lines. This means, they are discovered by users themselves in contexts far removed from central units. Hence, building centralised data science and AI teams should not serve as an excuse to neglect the expansion of these capacities across government. After all, even in the UK, only one in 14 digital, data and technology experts in government is employed in the central GDS unit. To lead this cross-government community of professionals and deliver transformation at scale through strategy, standards, assurance mechanisms and capability development, the UK has only recently established Central Digital and Data Office (CDDO). Together with other initiatives such as the Data Science Accelerator, a capability-building programme for public sector analysts, this new office thus is an excellent example of how to sustainably expand the base of data science and AI professionals across government.

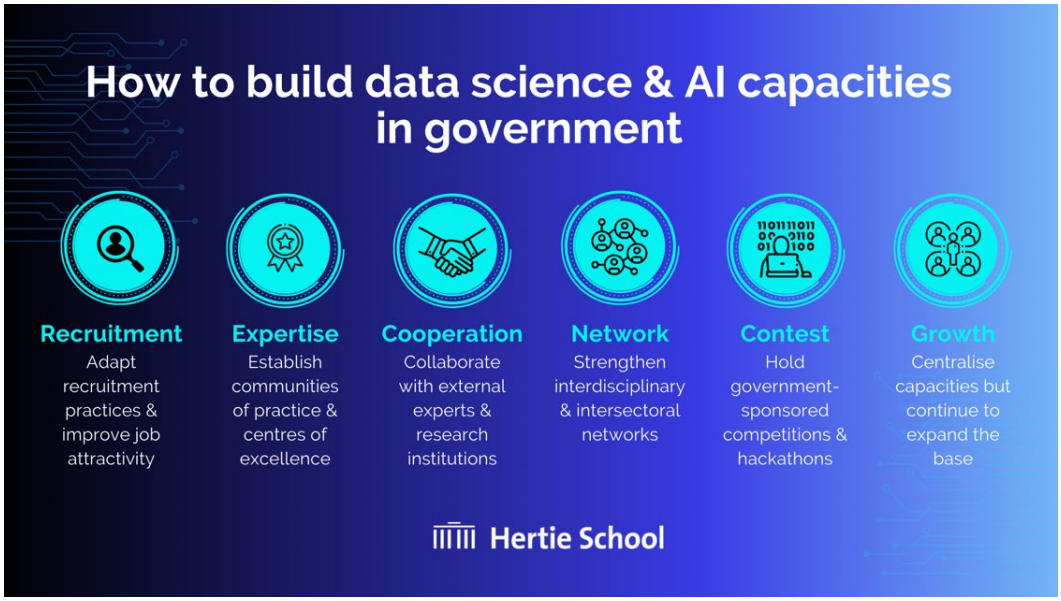


*Figure 3: The building blocks to successfully build data science and AI capacity in government.*

# 4 Conclusion

This policy brief argues that data science and AI can offer far-reaching benefits for the digital transformation of government. It also shows that an over-reliance on external expertise in the development of data science and AI-based solutions prevents governments from realising their full potential and can even be a significant source of risk. For this reason, we present six policy recommendations for building internal data science and AI capacity in government. Most of these recommendations also contribute to strengthening more general digital government skills. While the German government is on the right path, it needs to accelerate its efforts and focus on building future-proof internal data science and AI capacities.

# 5 Further Recommended Sources

Overview of the use and impact of AI in public services in the EU

Study on the usage of AI in German public administration

Report about the usage of AI in US federal administrative agencies

Essay uncovering the particularities of data-driven decisions in policy-making

This publication reflects only the personal views of the authors. Publication under Creative Commons license CC BY-ND.
Reprinting and other distribution – including excerpts – permitted only with acknowledgment of source • original version

© Data Science Lab & Centre for Digital Governance – Hertie School, Berlin 2022

Friedrichstraße 194
D – 10117 Berlin
Tel.: +49 (0)30 259219-0

Online: www.hertie-school.org
E-Mail: info@hertie-school.org
Twitter: @thehertieschool