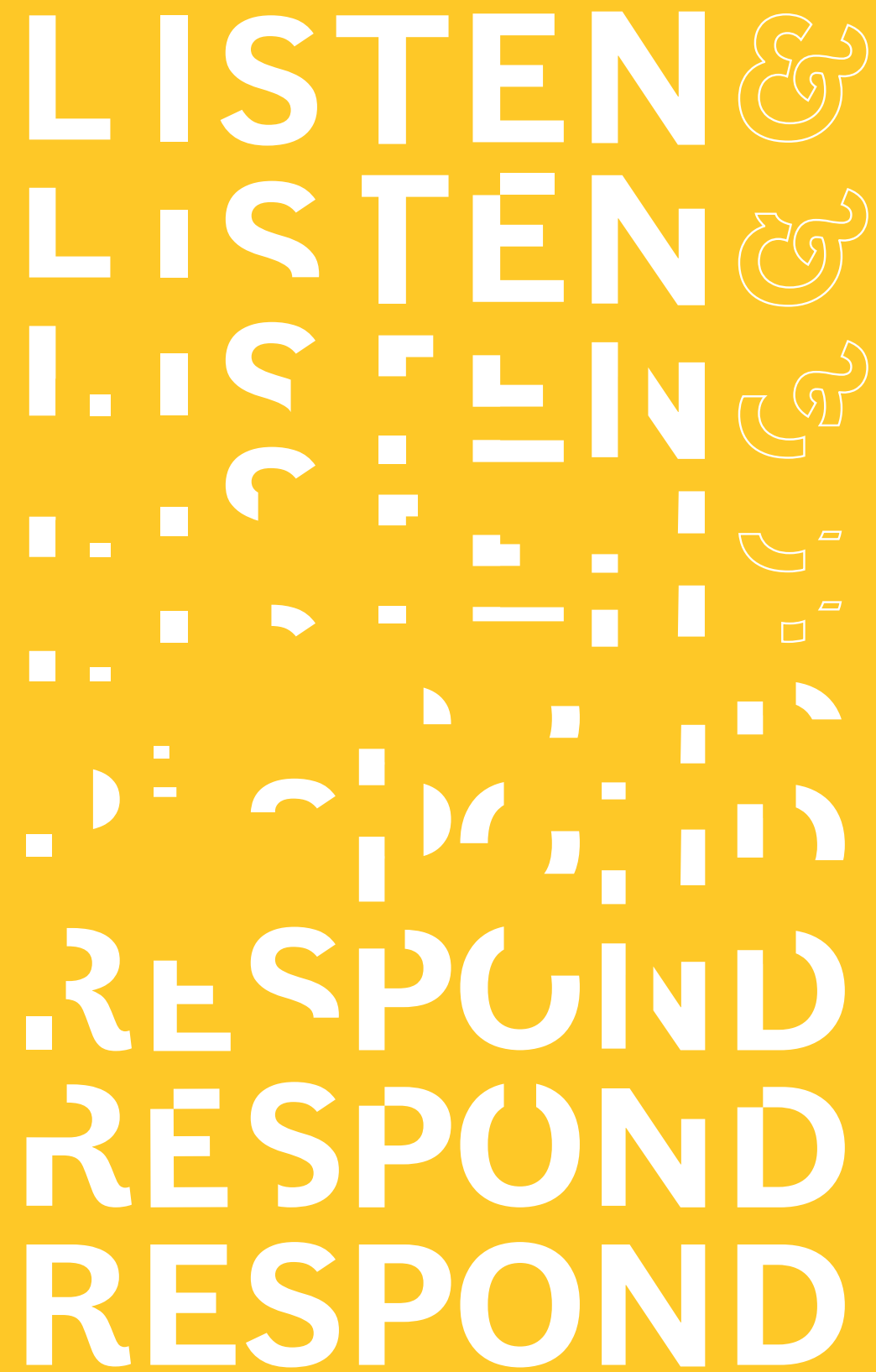
Centre for Public Impact
A BCG FOUNDATION

LEGITIMACY ARTIFICIAL INTELLIGENCE

How to make AI work in government and for people

ARTIFICIAL INTELLIGENCE in government

Centre for Public Impact



Public sector organisations across the world are looking to use artificial intelligence and, in particular, machine learning to improve policy and service delivery. These technologies are mainly used to automate defined, repeatable tasks, augment decision-making, and enhance understanding in policymaking and public service delivery. This is already having an impact on aspects of government that are often taken for granted. AI is playing a number of roles in the public sector: from managing traffic flows and maintaining public transport to helping police services to manage their data and citizens to communicate with local government.


AUTHENTICITY
AUTHENTICITY
AUTHENTICITY

As the use of AI expands into more sensitive and contentious domains, citizens are beginning to worry.

Governments are grappling with the question of how to accelerate the use of AI into these sensitive areas and take people with them. A 2018 survey found that citizens gave a lukewarm reception to governments' use of AI, and tended only to accept its application for specific purposes, such as administrative tasks or traffic management.[1] For many governments, this leaves AI at a standstill.

There is a lot of hype around AI and how it can change the world. At CPI, we take a more realistic approach. In 2017, we published our report, *Destination unknown: Exploring the impact of Artificial Intelligence on government*, in which we looked at the possibilities that AI offers the public sector.

Building on this and our 2018 report, *Finding a more human government* and the legitimacy behaviours that we outline in it (work together with people towards a shared vision, bring empathy into government, build an authentic connection, enable the public to scrutinise government, value citizens' voices and respond to them), we are now looking at how to unlock the potential of AI in government, and the role that legitimacy plays.

For AI to fulfil its potential as a tool for government and citizens, it needs *legitimacy*.

By legitimacy, we mean the reservoir of support government requires to achieve public impact.

People today are asking a number of questions about AI: How will it fit into everyday public sector tasks? If we want to go further, how do we connect AI to existing accountability and decision-making structures, and do we need to create entirely new mechanisms? How much decision-making autonomy do we want to concede to AI tools, and how is this best overseen?

What are its limitations and side-effects, and how do we address or mitigate them?

Only by starting to use AI projects in government will we be able to answer the tricky question of how to ensure legitimacy. By starting small and learning from experiences of working with AI, we can advance with confidence and, when the time comes, accelerate and move on to more sensitive areas.

We need to be realistic about the potential of AI – it is not a panacea for all the world's problems.

The AI narrative often ends with machines taking over human roles, jobs, and even the world. Examples of AI making major mistakes or errors are routinely publicised: people being misidentified for crimes they have not committed, wrongly classified based on their facial features; accidents involving driverless cars; or welfare support incorrectly refused, despite qualifying disabilities. AI will not magically solve all our problems – it is currently wholly unsuited to some policy areas and decisions. Instead, we should concentrate closely on the things it *can* help us with.

The existing narratives only hinder the use of AI in government. If we are not to accept the limitations and imperfections of AI now and talk about them openly, the technology will fail our citizens. Similarly, if we do not consider the quality of the policies we are asking AI to support, AI will take the blame for deeper, human decisions about how society should be run. Either way, legitimacy will not be achieved and AI will prove unable to help people in the public sector do their jobs more effectively.

Now that is out of the way, we can focus on what is real and possible. What we set out here can help you get there.

1 **See: 2018 BCG Digital Government Benchmarks: What Citizens Think About Governments' Use of AI**


EMPATHY
EMPATHY
EMPATHY
EMPATHY
EMPATHY

At CPI, we are excited about the potential of AI in government to improve outcomes for citizens. Even against a backdrop of mistrust and uncertainty, legitimacy is an achievable aim.

We are excited about what AI can do to help us move forward, and we believe that, in support of the right policies and deployed with care, it can bring about better outcomes for everyone. And as we said in our 2017 report, the key to unlocking its potential is seeing AI through the lens of legitimacy – bringing people on board and involving them in this exciting journey. This is not just the right and proper thing to do – it will help AI deployments solve the real, challenging and complex tasks and issues of our time.

Governments, and especially AI optimists, should take note of the story of big public sector IT projects: notorious and repeated mistakes led to disastrous outcomes for citizens, while excessive costs and delays eroded trust in government's ability to manage anything involving technology. Those thinking about introducing AI do so against a backdrop of mistrust of new technology and of governments more generally. Governments that deploy AI systems without **taking their legitimacy seriously** may well encounter resistance that scuppers the potential benefits of these tools, and are likely to find even adequate levels of trust difficult to regain. If you want to help make AI a reality, your mindset must shift from directing policy *at* people, to shaping it *with* people and being open about it in public.

Our legitimacy behaviours, researched internationally with governments and citizens, apply here and will help you. You will need to think about building AI into systems with **empathy, ensure diversity of experiences and of the people and perspectives involved, cocreate a vision, ensure authenticity in how you communicate, and allow and invite plenty of scrutiny and allow for redress, too.**

This guide is intended for anyone working in government and public services who formulates policy or deploys practices involving - or potentially involving - AI. We want to help you, the realist, understand how to proceed along a path towards AI in the public sector that has legitimacy, the key but often missing foundation of all successful policies and initiatives in government – but especially so when it comes to AI.

The **first part** of this guide provides a *five-point action plan* that applies to any initiative within government that might benefit from AI.

The **second part** provides an adaptable blueprint for an institutional landscape that supports the five action points, and enables legitimate public sector AI to take root.

All of these action points, practices, and important pieces of advice should be seen as more specific examples of how to build citizens' trust in government institutions and a belief that they will act ethically, rigorously, responsibly, and with careful consideration of all the interests at play. The action points cannot apply in a vacuum, but can support broader practices of good and legitimate policymaking when such policies also involve AI systems. There is much to do, but we can do it.

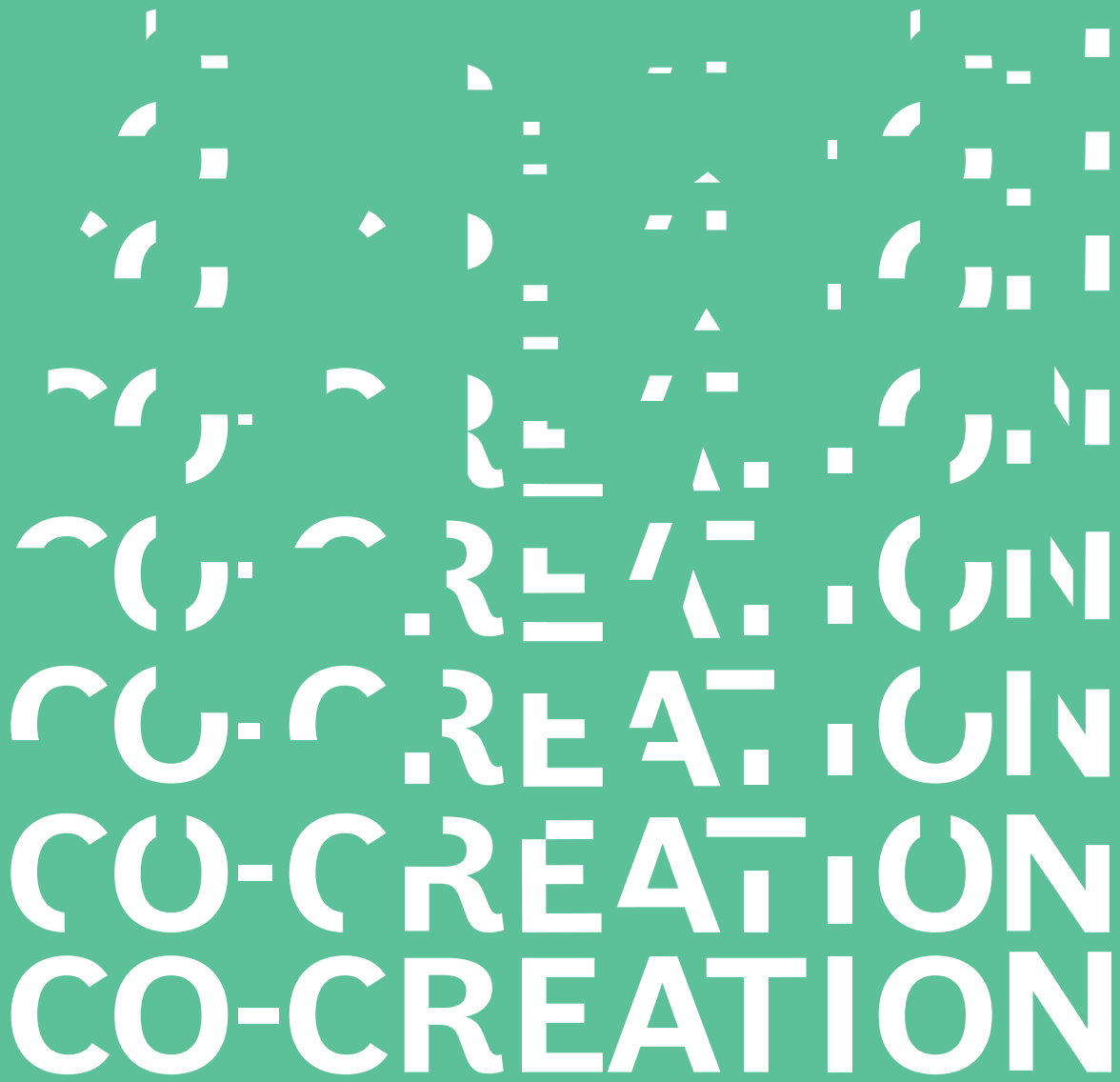
Action plan for developing legitimate public sector AI

The following five action points should motivate governments to act deploy AI systems effectively and with the legitimacy needed for their success and sustainability.

1. Empathise with the real needs of end-users and build your objectives around them

Understanding needs

Real, on-the-ground needs, including the pressures and strains on people's day-to-day lives, are generally different from those imagined by high-level decision-makers or AI vendors. As with all policymaking and public service design, government AI interventions should be shaped by a constantly evolving understanding of the needs of the end users, who might be frontline workers or citizens themselves. Governments should identify relevant end-user groups, and the designers and project leads of AI systems should apply techniques such as interviews, focus groups, and shadowing to elicit information from them,


CO-CREATION
CO-CREATION
CO-CREATION
CO-CREATION

prioritising listening closely and developing **empathy** skills. In short, all the tools and techniques deployed in "traditional" policymaking and service design are just as important, if not more so, when designing AI into these areas. **Building authentic connections** with the people who are going to use the new systems and services that emerge from the use of AI will ensure that they address real, "felt" needs and that they boost trust and support for any AI piloting processes, deployment or scaling-up.

Setting objectives

AI systems cannot function without **clear, formalised objectives**, which must be based on real requirements. The machine learning systems that power much of AI need a yardstick to tell them when they have "learned" and what good learning "looks like"; they cannot figure it out for themselves. Consider food safety agencies using AI to help flag potential breaches. One agency might want to automatically flag *more* cases than their current risk assessment process does; another might be happy with the number of cases it currently flags, but wants them flagged *earlier*; a third might want to detect *certain types* of breaches only. The system needs to know which goal to focus on, potentially at the expense of the others.

In addition to priorities, governments should set up concrete **measures of success**. One group's success measures might be to detect 10% more cases, another's to ensure cases that would normally have been detected after a year are detected after only six months. Clear success objectives will allow designers to decide how to weight other considerations, like interpretability, explainability or reliability, some of which have to be traded off against raw accuracy. Similarly, there will be types of failure that need to be avoided, such as discriminating against different groups or not representing citizens and their identities faithfully and inclusively. These measures should be defined and assessed in much the same way as successes, and treated with the same care.

Determining objectives and success measures can really only be done after deep, inclusive discussion. AI systems are no silver bullet, and it might not be possible to include every user need. This process often requires some tough prioritisation and the addressing of tricky, value-laden tradeoffs. Should a system focus on avoiding false negatives, thus catching all cases, or on avoiding false positives, thereby reducing the number of incorrect leads? By determining the approach early on – and being open about it – governments can lay a strong foundation for legitimacy.

2. Focus on specific and doable **tasks**

AI systems are tools, not workers. They will not usually replace an entire job, but instead reshape roles by replacing existing tools or tasks. Introducing AI systems to take care of certain tasks can free up staff for more creative, problem-solving, people-facing work that improves service delivery. To realise this goal, governments must collaborate in building a **shared vision** of an AI-augmented workplace with the public sector professionals who are using these technologies.

Individual AI applications work best with narrow, well-defined tasks, such as deciding where to direct a particular citizen query. A single system can be technically restricted to performing one specific function, such as image recognition or text classification, and so should be visualised as a building block. Practitioners should identify these tasks or building blocks that meet specific needs, and then decide which of them are the preserve of people and those in which AI might usefully assist. These tasks can then be examined for feasibility, such as whether high-quality training data exists. Working through this process step by step, block by block, will enable project leaders to envisage how an entire service delivery or policy process might be improved.

Involving users

Involving frontline workers as codesigners from the very beginning of the process will help to identify the building block tasks effectively. For example, using open and participatory workshops to connect with frontline workers and discover which aspects of their jobs they would like to spend more time on, and those where AI might usefully step in. By building active listening and understanding of user needs into the heart of design and focusing on using AI to alleviate or eliminate frustrating or repetitive tasks, the process not only gains legitimacy with the workforce, it becomes easier to discover new efficiencies and increase motivation in the workplace.

Road testing design for practical use

These well-defined building block tasks can then be tested to see how amenable they are to AI involvement and public acceptance. Does good data exist? Are existing systems compatible? Is the accuracy or performance rate acceptable given the risks of failure? Some tasks might be well suited to AI but unrealistic in the context of current limitations in data, processes or technology. In many cases, the tasks will be of such sensitivity that tools should never


SCRUTINY
SCRUTINY
SCRUTINY
SCRUTINY
SCRUTINY
SCRUTINY

carry them out alone, but should always be used with meaningful and critical human oversight. The boundary line between acceptable and unacceptable uses of AI is not always apparent or firm, and should be assessed in an open and inclusive way.

Being aware of this can trigger a process of improvement, for example in collecting new or better data or in overhauling legacy systems, making AI interventions possible further down the line.

Working with AI experts

AI experts play a vital role in this process, helping public servants to understand the existing capabilities of AI. They can help navigate the landscape of what AI is currently capable of, and what it is not. For example, AI might work well in triaging citizens' queries at an early stage of delivering a particular public service (by using chatbots, for example); at predicting the kind of information users might need when entering data; or at assessing the risk of some future incident, such as tax fraud. Yet there are some tasks such systems would struggle with, such as summarising the information in documents or analysing complex messages – both of which require a lot of contextual information – or detecting very rare and transitory events, as they simply do not have enough examples to learn from.

Example: taxation

In some countries' taxation sectors, it has become common practice for analytics divisions to create an in-house prototype of a working analytic tool, sometimes involving AI, before asking a vendor in to take it, scale it and replicate it. This has allowed organisations to cut costs, as vendors are aware they are working with a well-informed client, and they can deal efficiently with many of the issues around scoping tasks and problems before engaging expensive contractors.

3. Build AI literacy in the organisation and the public

The opportunities and challenges of AI require new skills to be developed. The distributed nature of these challenges means that skills cannot be "bolted on to" an organisation after the fact but need to be present, and built up, in employees who have both strategic and operational functions. We call this *AI literacy*, and it has several complementary components.

One aspect of AI literacy is the ability to **spot potential** for the better use of technology. Some of the largest AI opportunities are invisible when viewed top-down, but much more evident from the perspective of on-the-ground workers. Yet the bottom-up identification of opportunities for effectiveness or efficiency through AI is impossible unless individuals have a grasp of the current state of technology. To be able to spot opportunities for potential collaboration with, or implementation of, AI systems does not mean that civil servants must become programmers. Instead, they must possess a basic understanding of the realistic potential and possibilities of cutting-edge technologies. This might be achieved through, for example, exposure to AI in presentations, workshops, and hands-on sessions. Through **authentic connections** and clear feedback loops with those in strategic roles, operational decision-makers must feel that their suggestions will be taken seriously.

The second aspect of AI literacy involves enhancing **AI collaboration skills**. High-stakes algorithmic support systems need oversight and interpretation from frontline staff as well as from the system-builders and project owners. These frontline workers need to be able to work with decision support systems and constantly assess whether the systems can be trusted, particularly in dealing with borderline cases. Systems for decision-support should give front-line workers indications of the reasons for their behaviours. These workers in turn should be trained such that they know when to rely on their outputs, and when to challenge them or subject them to closer scrutiny.

AI literacy is not just an internal matter, it is one that affects citizens, too. Education and debate need to be fostered in all demographics and age groups that might be affected. This is a challenging task, but governments can start by engaging openly and frankly with the media, educational organisations, cultural institutions, libraries and civil society groups. Doing so will not just help governments make more legitimate use of these systems, but may spur investment and innovation in other sectors as well.

4. Keep maintaining and improving AI systems

Legitimate, trustworthy AI systems must minimise the risk of avoidable failure. An important part of doing this is to understand that deployed AI systems do not exist in a vacuum. The data that drives them is influenced by their users, the phenomena they are modelling or managing, and even other AI systems working in related policy areas. This can cause unexpected effects: the same features that make AI systems so appealing – their speed, ability to integrate many data sources, and potential to

change rapidly – also create challenges. These systems can be fragile, heavily interdependent, and prone to fail in unexpected ways. If these risks are not anticipated, understood and managed, they can undermine legitimacy – not only of the AI system in question but of these technologies more generally.

Changes that might affect AI are very common in the public sector. When data is collected from a particular service, it changes when the service changes. Even the slightest nuanced tweak can make a difference. If a food safety agency decides to change its spot-check approach, tomorrow's AI system will be trained on very differently distributed data, which may entirely change the way it makes predictions. Legislation and enforcement patterns are constantly modified. The nature of policy problems, particularly the stubborn or complex ones, is also fluid. Identifying and coping with these changes requires ongoing communication between technical teams and those to whom the change is more visible, such as operational workers or domain experts. When technical teams work in silos or are not properly integrated, these connections become difficult to make and maintain. Changes in roles, rotas, data collection methods, or interventions may have an impact on the AI systems, and so parts of government that were not used to communicating with each other may have to change their practices.

AI systems may also be vulnerable to attacks that could compromise their functioning, and efforts should be made to test and secure them as appropriate. This is likely to require close collaboration with investigation and cybersecurity professionals, such as national CERTs (computer emergency response teams), as well as international collaboration with governments facing related challenges and deploying similar systems.

The maintenance of an AI system in the face of change must be taken seriously. It cannot be assumed that technical teams can simply build an AI system and, once it is finished, move onto another system. Maintaining software, including amending it in response to social change, takes time, resource and effort, which must be properly considered at the onset of a project. Researchers call this "technical debt" – the ongoing maintenance cost of what has been built – but all too often its importance is underestimated in practice. The lack of maintenance and its consequences may have a profound effect on the legitimacy of AI.

Example: child protection

In Allegheny County, Pennsylvania, the County Children and Youth Services collect and implement staff feedback to keep adapting their predictive risk modelling tool. This tool helps them make decisions on whether or not to follow up a child abuse allegation. Taking staff experiences seriously ensures a smooth, user-friendly process and improves the way staff work with the tool.

5. Design for and embrace extended **scrutiny**

When the stakes are high and the areas where the AI system is deployed are value-laden – i.e. when there are "winners" and "losers" from a policy – governments must consistently involve and listen to **the voices of people** and civil society groups to ensure that the legitimacy of the AI system, the policy, and the wider organisation are not devalued in the eyes of the public.

As far as possible, governments should be resolutely open about AI systems, making them either partly or fully available online to anyone, or privately available to civil society groups and researchers within controlled environments. Methods for achieving this are described later in this guide. By engaging people early on and listening carefully to critiques, governments can avoid wider reputational damage (for example, through negative reporting in the media) which may be harmful, even if due diligence and care are taken with on-the-ground delivery.

One practical method for delivering this step is **extended peer review**, an approach developed for contentious types of research. When systems are being externally **scrutinised** for their statistical and technical soundness, they can be similarly externally reviewed for issues such as equity, transparency, robustness and proportionality. This would draw upon expertise in universities, civil society, other government departments, or other governments. A lack of public or societal legitimacy in an AI system is as problematic as a technical failure in the code, if not more so, and seeing it as an area for review allows it to be treated seriously from the start.

Implementing an action plan for delivering legitimate AI in government can only succeed if the right institutional conditions exist. This environment will not be the same for every government, but some successful approaches are already emerging that might be copied, adapted and built upon in other government contexts.

Creating the right environment - an institutional strategy

Coordinating AI initiatives to benefit from innovation and expertise

A number of governments are experimenting with setting up coordinating AI agencies, some of which have specific focuses on AI in policymaking, service delivery and regulation. In the UK, there is the recently established Centre for Data Ethics and Innovation, while France has its *Conseil National du Numérique*. There are also related non-departmental or non-governmental bodies, such as INRIA, the Alan Turing Institute and the Open Data Institute. These agencies can serve as a convening point for contacts within different departments and agencies, as well as taking account of relevant expertise and opinions in making strategic decisions across the public sector. In situations where governments cannot set up a new agency, they should consider establishing more or less formal networks between departments and functions, and setting aside resource to allow individuals to meet, learn and build a community and **a shared vision**.

More specifically, they can build and retain institutional knowledge and expertise that allows a wide variety of other goals and tools to become possible. AI expertise is expensive and hard to train with a public sector flavour (and hard to retain when you do). An agency that focuses on making an interesting, challenging and well-remunerated career path for AI practitioners can "lend out" expertise to different departments and regulators as necessary, and consolidate the learning from this lending for future projects. Given the cost of poorly scoped projects, bad vendor selection, or unwise technical lock-in to legacy systems, investing in systemic and coordinated learning, innovation and knowledge-sharing is a way to ensure cost-effective outcomes.

Training individuals in public sector machine learning

The problems that high-stakes AI systems in the public sector seek to solve are only distantly related to the common, lower-stakes applications by firms who are, say, analysing supply chains or delivering advertising. Government is a monopoly actor in many areas, and many of the most vulnerable individuals in society rely on it for protection and support. Consequently, governments must develop specific competences and training around public sector AI. While some aspects of this enterprise may be inspired by practice in the private sector, other aspects will differ significantly. Not only does the public sector routinely make decisions with huge impacts on individuals' lives, citizens also cannot pick and choose their provider of core infrastructure or services.

To give AI developers in the public sector the context they need for effective system design and oversight, governments should create and maintain a handbook and training programme to bring them up to speed. As the key 'living document' in this area, this handbook would explain how AI development works with national legislation, such as equality, freedom of information, or data protection laws, and the processes by which systems are built, tested, shared and documented. *[Example: Government Data Ethics Framework and the Aqua Book, UK]*. Some of these individuals should mentor newer employees or contractors, both to impart knowledge and to acquire cutting-edge technical expertise from outside sources. Communications skills concerning AI systems are important, both for communicating internally with colleagues and externally with citizens, and they should be included in these programmes.

For other employees at all levels, governments should integrate AI into training days and programmes, even at a light or minor level. Not all individuals need AI literacy to enable them to spot opportunities, so open calls should be put out to give interested individuals the ability to develop skills in this area if they want to. Individuals in particularly strategic roles could be offered university training courses or sabbatical programmes as part-time opportunities,

especially where practical project components allow individuals to work hand-in-hand with academics in building prototypes or pilots of AI systems for public sector use.

Example: policing

In Durham Constabulary in the UK, the HART machine learning assessment model attempts to detect moderate risk individuals who have been arrested but not yet charged, and offers them an alternative, rehabilitative programme instead of prosecution. It was built after the constabulary sent its head of criminal justice to undertake a two-year master's programme at Cambridge University, which involved analytics and AI in policing.

Building infrastructures for transparency and audit

To enable extended scrutiny, governments should build on the success of Open Data platforms by establishing *model repositories*, consisting of a list of AI tools deployed within government. For each tool there will be an analysis of its performance and legitimacy measures (such as consultation or public engagement reports), together with copies of the code – where applicable – or information about the processes by which the code can be **scrutinised**. In many cases, some or all of the code can be released openly (e.g. the model structure, weights, and so on), but in other cases this might present a security or privacy risk.

Where such risks exist, signposting other means of access – potentially in closed environments – still enables third parties such as civil society organisations to determine whether systems are discriminatory or are undesirable or illegal for other reasons. As happens in cybersecurity research, a responsible disclosure policy can give government a chance to respond to certain issues (e.g. by fixing them) before they are publicly released. While the above helps to foster trust in the entire system, there are cases where individuals might want to investigate or be aware of the functioning of the AI systems themselves. Potentially affected citizens or communities should have access to the reasoning of particularly influential systems and the kinds of data that go into them, the types of logic they use, and how they might be affected. This requires good engagement with human-computer interaction and user experience research to make sure the explanations are accessible, usable and actionable by a range of citizens, including the most vulnerable, who might be the ones most affected by imperfect automated decisions.

Prioritise interoperable, programmable and open technologies during procurement

Governments should impose minimum standards on interoperability and readiness for all new digital public sector tools, and they should be able to control all newly-procured systems (e.g. through APIs) in such a way that they can be automated. They should consult data scientists and AI developers about any large investment in new infrastructure to ensure that it is compatible with existing AI workflows. Governments should build in-house capacity for creating working prototypes, so they can engage vendors on an equal footing. Procurement strategies can also help foster the broader development of the AI industry, and governments should consider how SMEs can best respond to procurement requests *[Example: G-Cloud, UK; equivalents in other countries]*.

Commission challenge-led research for public sector AI

To build capacity in public sector AI, governments should create challenge-based research programmes, which can assume a number of different forms. Where project areas are well defined, research calls should highlight specific areas of government need and invite consortia of academics and industry partners to respond with innovative, sustainable solutions *[e.g. GovTech Catalyst, UK]*. Where problems are less well defined, it may be more appropriate to bring in researchers, for example as part of their doctoral training in AI, to collaborate with government departments in scoping the problems and providing new knowledge. Governments need to consider the different ways in which research can be deployed or developed in-house, and to ensure that existing processes and timescales for technology procurement or deployment do not accidentally exclude innovative pilots or experimentation. International collaboration can also be key here, as many countries face parallel challenges – their resources can be combined to understand them from different angles and their results can be shared.

Useful resources for further reading

- Case Studies accompanying this document:
    - Artificial intelligence in taxation
    - Durham Constabulary's AI decision aid for custody officers
    - Machine reading in government
    - The Allegheny County Family Screening Tool

- 2018 BCG Digital Government Benchmarks: What Citizens Think About Governments' Use of AI
- Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, April 2018, AI Now Institute.
- Data management and use: Governance in the 21st Century, 2017, The Royal Society and the British Academy
- Destination unknown: Exploring the impact of Artificial Intelligence on Government, 2017, The Centre for Public Impact
- Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making, Veale et al, 2018, *ACM CHI 2018*.
- Finding a more human government, 2018, The Centre for Public Impact
- UK Government Data Ethics Framework, Department for Digital, Culture, Media and Sport, 2018

We would like to express our gratitude to Michael Veale, academic and CPI fellow, for his help and advice in preparing this document.

To find out more about CPI's work on AI and legitimacy, contact Danny Buerkli (danny@centreforpublicimpact.org).

Centre for
Public Impact
A BCG FOUNDATION

Join the conversation

ai@centreforpublicimpact.org
Follow us @CPI_foundation: #FindingLegitimacy
#AILegitimacy

October 2018