K4D | Knowledge, evidence and learning for development

Automation of government processes

Siân Herbert
University of Birmingham
6 September 2019

Questions

What evidence tells us about the lessons learned from automating government processes, particularly in middle income countries or fragile and conflict affected environments? When possible focus on what impact the automation had?

Contents

1. Summary

This rapid literature review examines the impact of, and lessons from, automating government processes in middle-income countries (MICs) and fragile and conflict affected environments (FCAEs).

Key findings

Automating government processes involves data collection and digitisation via information and communication technologies (ICTs), artificial intelligence (AI), and sometimes also machine learning (ML) (Paul, Jolley & Anthony, 2018, p.6). Some common examples of ML in international development include: strengthening early warning systems; situational awareness; supplementing development data; point-of-service diagnostics; market segmentation; and customer and citizen service interfaces (Paul, et al., 2018).

Impacts

- Automation can improve the efficiency, quality and coverage of service delivery – e.g. automation of a medicine inventory management system in Pakistan led to control over theft, more transparency, better monitoring and evaluation, and more efficient service delivery in a pilot initiative (DFID, 2018).

- Automation of aspects of public sector staff recruitment, performance review, management, and monitoring can address nepotistic practices, can lead to efficiency savings on the salary bill, can improve staff and institutional performance, and can increase transparency and trust in institutions, among other impacts. E.g. China's staff performance reform in the State Administration of Taxation is credited with enabling them to make the switch from sales tax to value-added tax in record time, and winning the broad support of tax agency staff. It also contributed to increases in the government's tax revenues; reforms of state and local tax collection; implementation of preferential tax policies; increased participation in international tax cooperation; and a clearer understanding of institutional and individual responsibilities (World Bank, 2018).

- ML can generate more precise recommendations, identify new important data factors, and can quantify the relative importance of different factors (Paul, et al., 2018).

- ML algorithms can be used to make predictions, and can help identify emerging problems more quickly than traditional, human methods of analysis as they tend to combine a wider number of data sources than human methods (Paul, et al., 2018).

Lessons and challenges

- Present ML and AI systems are unable to recognise if the decision made agrees with the context, like human analysis can, and thus ML can lead to suboptimal decisions (Paul, et. al., 2018). So while ML and AI systems can be statistically very good, they can fail for individual cases. This raises important dilemmas in regards to accountability.

- ML systems include decisions that may, or may not, have been explicitly made by the developers. These decisions tend to be invisible to those affected by the system, those using it, and can be even for those who developed it. This raises many concerns, including around equality of treatment, the fairness of outcomes, and accountability.

There is a need for careful monitoring of the automated processes to ensure that the decisions do not unfairly disadvantage people.

- ML tools are built by people and thus contain the choices, expertise, insights, and biases (to a greater or lesser extent) of those people at each stage of the automation (Paul et al., 2018). Thus these systems can overtly, or inadvertently, lead to ineffective, unfair or exclusionary outcomes. ML models should be analysed carefully for the points at which bias may enter into their programming before they are implemented.

- The quantity and quality of data impacts whether a model will work and for who, and data can be scarce and difficult to gather, especially where there is less use of digital devices.

- Automation requires effective ICT infrastructure – which should either already be in place, or should be built as part of the project. Conflict can significantly shape the development of the ICT sector.

- It is important that AI generated information is used as complementary information alongside more traditional information and methods for decision-making, and alongside context analysis and expertise. E.g. automated climate and crop information provided to smallholders in Colombia was combined with workshops to teach about probability, uncertainty, and interpreting forecasts (Paul, et al., 2018).

- Reform plans should be designed with real time monitoring and evaluation built in (E.g. this was critical to the success of China's performance management reforms in the State Administration of Taxation) (World Bank, 2018).

- Successful reforms require changing the institutional culture, beyond technical fixes. While technology was an enabler in the case study of China's state tax reform agenda and Indonesia's automated recruitment reform agenda, in both countries the reformers had to demonstrate the benefits to the public and those affected to win over those opposed to it (World Bank, 2018).

- Mobile call detail records (CDRs) are a rich and important source of data. Yet as they include sensitive data they can be hard to obtain, and using it comes with significant legal, privacy and ethical challenges (Paul, et al., 2018).

- As ML and AI become more common, there is a need to understand when they offer a suitable solution to the challenge at hand, and whether and how they are effective, inclusive, and fair. Factors for success include: The inputs and outputs are well-defined; there is clear feedback and definable goals; datasets are large and diverse; the phenomena being modelled is fairly stable; there is no need for detailed explanations; there is no requirement for background or common sense knowledge; and when there can be tolerance for error (Paul et. al., 2018, p.34).

Literature base

Automation in international development is relatively new, and many projects are still in initial stages (Paul, et. al., 2018). There is a small amount of very good practitioner literature on this issue with useful case studies. Most of the literature and case studies found during this rapid review focus on the automation of service delivery functions (e.g. in health). However there may be more literature that could be found with more time as the term "automation" tends to bring up literature on the automation of work, rather than government processes.

2. Automating government processes – definitions and examples

Automating government processes involves data collection and digitisation via information and communication technologies (ICTs), artificial intelligence (AI), and sometimes also machine learning (ML) (Paul, Jolley & Anthony, 2018, p.6). This work comes under the field of electronic governance (e-governance). Paul et. al. (2018) explain that ML, a sub-field of AI, is "a set of methods for getting computers to recognise patterns in data and use these patterns to make future predictions" (Paul, et. al., 2018, p.6). While AI "uses computers for automated decision-making that is meant to mimic human-like intelligence". These decisions might be implemented directly (e.g., in robotics), or by a human decision-maker (e.g., product recommendations with online shopping) (Paul, et. al., 2018, p.6). AI typically relies on machine-learning algorithms to "translate data into usable predictions".[1] And digitisation is the conversion of data into a digital form so that it can be processed by a computer. These tools allow computers to automate decisions and to make data-derived predictions (Paul, et. al., 2018).

Examples

Some common examples of ML in international development include (Paul, et al., 2018):

Strengthening early warning systems – ML has been used and tested out for early warning related to nutrition, conflict, and food security (Paul, et al., 2018, p.18). And they could be used for early warnings of political instability, crop pest infestations, or commodity price shocks (Paul, et al., 2018, p.18). Examples include: the USAID-funded – Grillo – provides real-time warnings about earthquakes through monitoring ground motion data (Paul, et al., 2018, p.18). And the HealthMap initiative combines expert data (e.g. clinicians' reports) and informal sources (e.g. news reports) to generate a real time global map of emerging disease threats.

Situational awareness – ML analysis of satellite images has been used to identify human rights violations, wildlife trafficking, and deforestation (Paul, et al., 2018, p.18). ML analysis of social media data has been used to support analysis of infectious disease surveillance, pharmacovigilance (tracking the safety of medications), early warning signs of "lone-wolf" terrorism, and analysis of online advertisements might be used to detect human trafficking (Paul, et al., 2018, p.18).

Supplementing development data – ML has shown the potential to fill data gaps by drawing on satellite imagery of physical features to help predict poverty levels, population density, and basic infrastructure (Paul, et al., 2018, p.20). Mobile phone data has been used to map climate-driven migration and population displacement (Paul, et al., 2018, p.20).

Point-of-service diagnostics - Computer vision has also been used for diagnosis of symptoms, management and prevention techniques of human diseases (e.g. malaria, hookworms and schistosomiasis), and plant diseases, pests, and nutrient deficiencies (Paul, et al., 2018, p.21).

---

[11] For development practitioners that want to learn about the technical and practical aspects of ML and AI, see the Paul, et. al. (2018) paper.

Market segmentation – ML algorithms - decision trees - have been used to precisely target interventions – e.g. in health (promoting medical male circumcision), and financial inclusion (Paul, et al., 2018, p.21).

Customer and citizen service interfaces - Conversational interfaces (e.g. chatbots) have been used to field and address consumer complaints, to provide automated assistance, financial advice and financial education, and to provide mental health counselling more efficiently by strengthening customer communications, improving response time, and reducing the workload of employees (Paul, et al., 2018, p.21).

See: **Figure 1: Illustrative sectoral ML applications[2]**, Paul, et al. (2018: 23), https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf. It provides illustrative examples of these uses in the international development sector.

3. Examples and impacts

Improving the efficiency, quality and coverage of service delivery

Digitisation and e-governance can directly and indirectly lead to gains in the efficiency and quality of public services by: lowering the cost of delivery (e.g. by reducing rent-seeking opportunities); improving the quality and coverage of services through new technologies (e.g. tele-medicine, drones); strengthening feedback loops between users, service providers and monitoring agents (e.g. through mobile phone feedback, or more accessible and understandable government data); helping citizens connect with each other, fostering collective action, which can incentivise governments to improve the quality or coverage of services; enhancing citizen's participation in local affairs; and promoting participatory, transparent, responsive and inclusive democracy to enhance grassroots development (Vester Haldrup, 2018a; Ojo, 2014). Algorithms can work more quickly than people, e.g. allowing larger regions to be mapped more efficiently through analysis of images (Paul, et. al., 2018)

Automation of a medicine inventory management system in Pakistan led to control over theft, more transparency, better monitoring and evaluation, and more efficient service delivery in a pilot initiative. The Teemardar – Medicine Inventory Management (TM) is part of the DFID-funded District Delivery Challenge Fund in Pakistan. In addition to the impacts mentioned, the automation aimed to: improve medicine availability; maintain a visible inventory; create data for informed procurement and distribution of medicine at a district level; integrate with the work of district planners on evidence based needs assessment of medicine requirements; and track the exact uses of medicine and thus develop disease patterning in the district to inform planning and budgeting (DFID, 2018, p.33). Another impact is that the Electronic Record Management has improved patient records, disease records and patient disease history (DFID, 2018).

Another example of automating health services is carried out by the Ugandan social enterprise - The Medical Concierge Group (TMCG) – which provides free access to health care professionals and health information via communication technology platforms (van Niekerk, et al., 2017, p.66).

---

[2] See the original document for links to further information on the ML application

One of its functions is to provide strategic patient engagement initiatives over mobile technology, with automated reminder systems, post-hospital follow-up, and satisfaction surveys (van Niekerk, et al., 2017, p.66). Van Niekerk, et al. (2017, p.66) find that the TMCG's use of technology increases access to health information, provides health-care services to low-income communities, and also reduces the burden on public health facilities.

An example of automation in the justice sector is the Azerbaijani approach to reducing civil court backlogs. Previously, Azerbaijani court backlogs were mainly emerging from relatively simple civil cases (e.g. claims for unpaid bills), thus the government partnered with the private sector to use an automated system to streamline the handling of uncontested cases (World Bank, 2018). This then freed up the judges' time for more important cases (World Bank, 2018). It is identified by the World Bank (2018) as a key innovation in the Azerbaijani governance reform agenda.

Another initiative in the health sector is the digital platform for malaria surveillance, run by the Indian Centre for Development of Advanced Computing (C-DAC) (a government agency of the Indian Ministry of Communications and Information Technology). Through the Mobile-based Surveillance Quest using IT (MoSQuIT), the previously manual malaria surveillance process was automated and streamlined. van Niekerk, et al. (2017, p.42) highlight that the MoSQuIT case "demonstrates how disease surveillance efforts can be strengthened through a streamlined technology platform that gathers data from different care providers along the patient care continuum. Integrated data collection coupled with real-time analytics can help detect disease outbreaks and trigger a quicker health systems response. In addition, it enhances transparency, communication and trust between different care providers."

Another example of an initiative in the health sector, which includes some automation, alongside digitisation and general improvement of ICTs, is the Tele Medicine/Diagnosis (TMD), which is also part of the DFID-funded District Delivery Challenge Fund in Pakistan. The initiative equipped remote Basic Health Units with key diagnostic facilities and technological connections to a remote doctor at the district hospital (DFID, 2018; Vester Haldrup, 2018a). The equipment can be solar powered and uses a simple radio signal (GPRS) to connect to the district hospital (DFID, 2018). After the patients are tested, the consultant receives the results immediately, then the consultant and patient engage in real time through a video connection. The pilot's success has meant that the government has started to adopt and scale up the intervention (DFID, 2018).

Improving public sector staff performance, monitoring, and recruitment

Automation of aspects of public sector staff recruitment, performance review, management, and monitoring can address nepotistic practices, can lead to efficiency savings on the salary bill, can improve staff and institutional performance, and can increase transparency and trust in institutions, among other impacts.

China's staff performance reform in the State Administration of Taxation

China transformed its State Administration of Taxation (SAT), and staff performance, through introducing an automated Performance Management Reform system. Prior to the 2013 reform, "senior management could not rely on staff to execute tax policies in a timely manner, and taxpayers suffered with poor service" (World Bank, 2018, p.74-75). While post-reform, "the Chinese authorities credit the [new automated staff performance] system with enabling them to make the switch from sales tax to value-added tax (VAT) in record time, while also winning the

broad support of tax agency staff" (World Bank, 2018, p.74-75). It also contributed to: increasing the government's tax revenues; reforms of state and local tax collection; implementation of preferential tax policies; increased participation in international tax cooperation; and a clearer understanding of tax bureaus and workers' responsibilities, with their tasks tied to clear goals (World Bank, 2018). "Surveys indicated that taxpayer satisfaction in their local tax bureaus increased after the performance management reform (World Bank, 2018, p.74-75). And, "according to the World Bank's Doing Business Project, the average time to prepare and pay taxes decreased from 261 hours in 2015, to 207 hours in 2017" (World Bank, 2018, p.74-75).

Lessons from this reform project are included in the section below, though the long-term outcomes and lessons remain to be seen (World Bank, 2018, p.75). For further details on the background, design and implementation of the performance system in China, see the World Bank (2018, p.76-80) case study.

Automating public sector employment tests to reform civil service recruitment in Indonesia

The Indonesian authorities' introduction of an automated recruitment system – a computer-assisted test – is credited with bringing "more transparency and credibility to the recruitment process" and reducing "opportunities for collusion and nepotism" (World Bank, 2018, p.73-75). It has now become the "de facto standard for more than 62 ministries and agencies at the national level" (World Bank, 2018, p.81-85). It was brought in to "address long-standing perceptions that civil service recruitment was corrupt, with payments being made in exchange for public sector jobs" (World Bank, 2018, p.81-85).

Lessons from this reform project are included in the section below. For further details on the background, design and implementation of the computer-assisted tests in Indonesia, see the World Bank (2018, p.81-85) case study.

Generating more precise recommendations

ML can generate more precise recommendations, identify new important data factors, and can quantify the relative importance of different factors (Paul, et al., 2018, p.18). E.g. The South African based organisation - Harambee – uses ML in its youth employment work (Paul, et al., 2018, p.18). Harambee looks to ML tools to better analyse the data it has collected from its clients over seven years of work, especially to fill in gaps in knowledge areas such as: the features of a candidate that best predict success in certain types of jobs, and the maximum transport distance for candidates to make the job sustainable (Paul, et al., 2018, p.31).

Spot emerging problems more quickly

ML algorithms can be used to make predictions, and can help identify emerging problems more quickly than traditional, human methods of analysis as they tend to combine a wider number of data sources than human methods (Paul, et al., 2018). These data sources can include large image databases and geospatial data, as well as text-based reports (Paul, et al., 2018). The ability of ML methods to quickly filter through large image databases gives the potential for ML "to spot weaker, harder-to-define signals that might otherwise have been missed" (Paul, et al., 2018, p.18). These features make ML useful in early-warning systems (Paul, et al., 2018).

4. Lessons and challenges

While automation has a "tremendous potential" for contributing to better development outcomes, public accountability, and public service delivery, there are also significant challenges, and projects often fall short of their aims, or do not get implemented at all (Vester Haldrup, 2018a; Paul, et. al., 2018).

ML leading to suboptimal decisions

Present ML and AI systems are unable to recognise if the decision made agrees with the context (Paul, et. al., 2018). This is because humans absorb and process data (especially visual data) in the context of the environment, while artificial intelligence does not (Paul et. at., 2018). "If the interpretation of the data (for example identifying an image) doesn't fit the context of the situation, the human can recognize that something is not quite right" (Paul, et. al., 2018).

ML and AI systems can be statistically very good, but can fail for individual cases. This is because humans have the ability to learn from very small data sets, while ML and AI systems typically require very large data sets (sometimes to learn, thousands of images or words are required). This means that "many ML and AI systems may provide wrong or inappropriate answers if used in a context different from their training environment" (Paul, et. al., 2018).

Therefore, ML decisions may make better overall decisions for efficiency, but these may lead to suboptimal outcomes on an individual basis. This raises important dilemmas in regards to accountability – who is responsible when individuals are disadvantaged by a decision made through AI or ML? E.g. in South Africa, by incorporating transport route data into profiles for employment opportunities, employment candidates would only selected if located within reasonable travel distance of the job which would be better for job sustainability (Paul, et al., 2018, p.31). However, it would means that those living far from economic centres and transport links would rarely be called for jobs (Paul, et al., 2018, p.31). Another dilemma raised in this example is whether the employment organisation should use the information it gathered to advocate the government for better transport routes and urban planning (Paul, et al., 2018, p.31).

Understanding and justifying the decisions behind algorithms

ML systems include decisions that may, or may not, have been explicitly made by the developers. These decisions tend to be invisible to those affected by the system, those using it, and can be even for those who developed it. This opaqueness can stem from intentional decisions around security concerns and competitive advantage; or from technical illiteracy, or the sheer complexity of the model which may include thousands of inputs (Paul, et al., 2018, p.31). This opaqueness also means that people cannot easily understand the process by which decisions are made which raises many concerns, including around equality of treatment, the fairness of outcomes, and accountability.

Automating data may lead to decisions that lead to more optimal outcomes overall, but that are not neutral towards the people involved. E.g. data gathered by the South African employment organisation Harambee found that larger family size is a relatively strong predictor of a candidate's ability to find a job, however, it could not explain why (Paul, et al., 2018). "Deciding how to act on these insights raises important, value-laden questions. Should those with larger families be ranked lower than those with smaller families by Harambee's system because they have a better chance of finding work without Harambee's intervention? Or should Harambee's

process remain neutral to family size? These decisions involve value judgements that a ML algorithm, if simply optimized for efficiency, might gloss over without deliberation" (Paul, et al., 2018, p.33).

This highlights the need for careful monitoring of the automated processes to ensure that the decisions do not unfairly disadvantage people. Paul et al. (2018) note that this monitoring should include those involved in the business side, and those involved in the technological of the automation, to ensure the tools supports the overarching values and aims of the initiative.

Bias in models

ML tools are built by people and thus contain the choices, expertise, insights, and biases (to a greater or lesser extent) of those people at each stage of the automation (e.g. design, implementation, monitoring and evaluation, etc) (Paul et al., 2018). "ML-enabled decision systems are not merely a technological tool, but part of a socio-technical system — a system in which technologies shape and are shaped by people, organizations, and policies" (Paul et al., 2018, p.44). Thus, these systems can overtly, or inadvertently, lead to ineffective, unfair or exclusionary outcomes, especially for discriminated groups (e.g. women, ethnic minorities, etc) (Paul et al., 2018, p.44; Vester Haldrup, 2018a). Patronage-based bureaucracies may particularly resist e-government initiatives that aim to reduce reduce discretion and rent-seeking (Vester Haldrup, 2018a). Paul et al (2018) advise that the ML models should be analysed carefully for the points at which bias may enter into their programming before they are implemented.

ML tools in developed countries "have sometimes been found to automate racial profiling, to foster surveillance, and to perpetuate racial stereotypes. Algorithms may be used, either intentionally or unintentionally, in ways that result in disparate or unfair outcomes between minority and majority populations" (Pau, et al., 2018). These shortcomings can occur across contexts, and especially those with histories of ethnic conflict or inequality. It is important to recognise that these technologies can do harm, and to commit to addressing these harms (Paul, et. al., 2018). E.g. in South Sudan there has been mixed positive results from crowdsourced and ICT-enabled early warning mechanisms, "including deliberate misreporting of incidents in order to secure intervention or advantage" (Kelly & Souter, 2014, p.18).

In a book on automation across the USA, Eubanks (2018) warns of how automating processes are not neutral and are in fact automating inequality, especially as they come at a time when economic inequality is growing and when programmes serving the poor are unpopular (at least in developed countries). Eubank (2018, p.9-10) highlights that "the cheerleaders of the new data regime rarely acknowledge the impacts of digital decision-making on poor and working class people" and how they can be forces for "control, manipulation and punishment".

Automation relies on good data

The quantity and quality of data impacts whether a model will work and for who (Paul, et al., 2018, p.44). Data in MICs and FCAEs can be scarce and difficult to gather as there is less use of digital devices (therefore there is less readily available data), as connectivity tends to be slower and more expensive, and as contexts can change rapidly (particularly in FCAEs) (Paul, et al., 2018, p.44). Data "proxies are imperfect stand-ins for the values we actually want to measure, and they can introduce distortions (Paul, et al., 2018). Common data sources include: household surveys, mobile phone metadata, satellite imagery, social media, and e-health records (Paul, et

al., 2018). Paul et. al. (2018, p. 74-79) recommend: strengthening local technical capacity; strengthening relevant governance structures; ensuring responsible data practices; ensuring responsible, shared learning; bringing diverse perspectives into model building; designing for model interpretability; evaluating the model for fairness; integrating the model into practice; and ensuring local feedback mechanisms.

Automation relies on good ICT infrastructure

Digitisation and automation requires effective ICT infrastructure – which should either already be in place, or should be built as part of the project (Vester Haldrup, 2018a). E.g. prior to China's successful reform of its State Administration of Taxation through an automated performance management system, it already had a strong IT platform and a strict hierarchical structure in place (World Bank, 2018). This meant that the performance management system could be introduced quickly, with automatic collection of performance data, and that the top management could oversee the performance management of the whole administration in real time (World Bank, 2018). "Developing quantifiable measures and tracking them for an institution [of that size]… would not have been possible without the technology to provide the foundation" (World Bank, 2018, p.78).

Conflict can significantly shape the development of the ICT sector – e.g.: communications infrastructure is often destroyed during violent conflict (e.g. in Liberia, Timor-Leste, and Somalia) (Kelly & Souter, 2014, p.12); networks are unlikely to reach into dangerous or rebel-controlled areas; and private investment, policy and regulatory reform, and economic activity (including around ICTs) are often stymied by conflict (Kelly & Souter, 2014, p.12). Thus the ICT sectors of conflict and post-conflict countries are often underdeveloped compared to other countries (Kelly & Souter, 2014). The end of conflict presents a window of opportunity to reshape ICT infrastructure and institutions, and marks a period of opening for investors (Kelly & Souter, 2014, p.12). Yet the investment climate may still be one of high risk due to the risks of the re-emergence of conflict.

Designing automation in-line with context analysis from experts and locals

It is important that AI generated information is used as complementary information alongside the more traditional information and methods for decision-making that humans usually use (Paul, et. al., 2018). E.g. Colombian smallholders have been assisted by data-driven agronomy and ML by the International Center for Tropical Agriculture (CIAT) through the provision of climate and crop information (e.g. local climate forecasts and the identification of methods to maximize the yield of crops according to soil type) (Paul, et al., 2018, p.26-29). "To help farmers engage as full partners, CIAT supported workshops to teach about probability, uncertainty, and how to interpret seasonal forecasts" (Paul, et al., 2018, p.26-29). It also supports local committees to issue a jointly-authored bulletin with seasonal forecasts and recommendation, and shares the findings through a WhatsApp group that includes technicians and farmers (Paul, et al., 2018, p.26-29). Any automated reform plans should be designed in line with the specific context – e.g. public and private organisations are profoundly different in terms of their purpose, culture, and operating context (Vester Haldrup, 2018a). Public organisations face specific challenges in regards to coordination, implementation, and scale up due to the multiplicity of organisation entities within (e.g. departments and ministries) (Vester Haldrup, 2018a).

Ensuring buy-in for the reforms from all stakeholders

Successful automation of government processes requires that all stakeholders buy-in and support the reform agenda. E.g. in China's staff performance reforms in the State Administration of Taxation, change management, especially the political leadership at the head of the agency, was essential to the reform's success (World Bank, 2018). To do this the tax agency leadership progressively built support from stakeholders across the organisation throughout the course of the project roll-out (World Bank, 2018). In Pakistan, the District Delivery Challenge Fund – which manages the two reform examples provided above (Teemardar – Medicine Inventory Management (TM) and Tele Medicine/Diagnosis (TMD)) aimed at improving service delivery and was effective because its managers worked closely with the government at every stage of its operations (Coffey, n.d.).

Designing reform plans with real time monitoring and evaluation

China transformed its State Administration of Taxation (SAT) through introducing an automated performance management system (World Bank, 2018). The three-year change management plan affected more than 800,000 staff, about 10% of China's entire civil service (World Bank, 2018). The SAT's performance plan included "quantitative and qualitative indicators that cascaded down from national level to bureau level to individual, with in-year performance monitored through records of completed work and automatically-generated computer data" (World Bank, 2018, p.76-80).

The pilot initiative started in 2014 (affecting a quarter of the staff), was rolled out nationwide by 2015, and was reviewed in 2016 to improve effectiveness (World Bank, 2018). The approach was highly organised, with structured committees and monitoring and evaluation at each stage, e.g. elements of the plan included: setting up a new Performance Management Office to oversee the performance management system; transferring officials with knowledge or experience in performance management to the new office; creating a team of consultants to advice on the new system; recruiting top performance management experts from universities around China; training the new team in performance management (including trips abroad to learn from other countries); building support for the initiative from the top down; making managers at each level responsible for passing on information to the level below; creating a group (the Leading Group for Performance Management) with representatives from every level; setting up appraisal committees responsible for reviewing and adjudicating any issues; creating a national performance plan based on strategic objectives and direction from the central government, from which individual bureaus formed their own plans and performance indicators; establishing a system to monitor progress and identify and correct any problems in real time; carrying out assessments on the completion of performance indicators, and submitting performance reports to headquarters; making adjustments to the information system to make the interface more user-friendly and add additional functions; and setting a clearly defined roadmap, timetable, task statement, and responsibilities for all departments at all levels (World Bank, 2018, p.76-80).

The real-time monitoring and adjustments to the reform agenda was critical to ensuring its smooth roll-out in the short timeframe. E.g. "analysis of the first full year of the system found that bureaus were creating too many performance indicators, and not enough of them were quantifiable. After input from the performance management office... [they] drastically reduced noncritical indicators and focused more on key tasks [and] significantly increased the quantitative indicators" (World Bank, 2018, p.80).

Successful reforms require changing the institutional culture, beyond technical fixes

Yet, while technology was an enabler in the case study of China's state tax reform agenda and Indonesia's automated recruitment reform agenda, improvements to performance also needed institutional cultural change (World Bank, 2018, p.75). E.g. both China and Indonesia experienced opposition to the reforms initially, and they had to demonstrate the benefits of the reforms to internal stakeholders and to the public to garner the support needed for success (World Bank, 2018, p.75).

Privacy concerns with data gathering, use and data protection

Mobile call detail records (CDRs) are a rich and important source of data that is becoming increasingly used, including in ML systems (Paul, et al., 2018). The CDR metadata includes information about caller location, the time and date of the call, and the number dialled, but not the content of calls or messages (Paul, et al., 2018). "Metadata analysis can be powerful because it helps us bypass irrelevant detail (in this case, the contents of calls and texts) in favour of higher-level insights about where, when, and with whom people are communicating" (Paul, et al., 2018, p.20). Yet this also means that CDRs "are among the most informative (and sensitive) large-scale datasets on human behaviour" (Paul, et al., 2018, p.20). CDRs can cost less than surveys, however the sensitivity of this data can make it difficult to obtain and comes with significant legal, privacy and ethical challenges (Paul, et al., 2018).

Knowing when automation offers a suitable solution to the challenge at hand

As ML and AI become more common, there is a need to understand when they offer a suitable solution to the challenge at hand, and whether and how they are effective, inclusive, and fair. Paul et. al. (2018, p.34) find that ML works best when:

- **The inputs and outputs are well-defined** - ML is easier when the inputs and expected outputs are clear and unambiguous (e.g. rainfall levels) rather than subjective (e.g. quality of governance).
- **There is clear feedback and definable goals** - "If a model's predictions can be tested against something in the real world, then deficiencies can be identified and corrected. In some cases (e.g., estimating the risk of rare events) it is difficult to know whether a model is truly accurate".
- **Datasets are large and diverse** - "In general, algorithms will be more accurate and less biased if training data are larger and more diverse".
- **The phenomena being modelled is fairly stable** - ML predictions are based on training data, extrapolated from the past. If the phenomena changes quickly, ML will need new training data to keep up.
- **There is no need for detailed explanations** – "While explainable ML is an active area of research, the most-accurate models are still often the most opaque. In situations where there is a compelling need for explainability, it may require sacrificing some degree of model accuracy in order to retain interpretability. When sufficient accuracy cannot be achieved without compromising explainability, ML may not be a good option".

- **There is no requirement for background or common sense knowledge -** "ML researchers frequently cite Andrew Ng's 'one-second rule' — a task is best-suited for automation if a normal person could do it with less than one second of thinking. For example, we recognize the face or voice of a familiar person immediately, without much conscious thinking. By contrast, evaluating the logic of a written argument takes more cognitive effort, and is likely to rely on information from outside the text. Many "one-second" tasks remain un-automatable, because they still rely heavily on people making common sense judgments.
- **When there can be tolerance for error** – "All decision systems make mistakes, and decisions made by machines can be just as fallible as those made by people. Relying on machines to make decisions requires honestly assessing the expected rates at which machine outputs will be incorrect — and whether those rates are acceptable. Automation may sometimes require tolerating more errors in order to reduce costs or achieve greater scale".

Vester Haldrup (2018b) proposes a six-step decision-making process for when and how to use digital technology to improve public service delivery:

1. clarify the problem, evaluate solutions
2. Gauge interest in innovation
3. Determine which services to digitise
4. Account for existing institutions
5. Analyse political economy constraints
6. Consider technical feasibility

5. References

Coffey (n.d.) Working collaboratively with government to make a governance-focused challenge fund succeed. Blog. Coffey. https://coffey.com.au/en/ingenuity-coffey/working-collaboratively-with-government-to-make-a-governance-focused-challenge-fund-succeed/

DFID (2018) Project Completion Review. DFID Pakistan: Sub-National Governance Programme (SNG) DFID http://iati.dfid.gov.uk/iati_documents/42143296.odt

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press. https://www.amazon.com/Automating-Inequality-High-Tech-Profile-Police/dp/1250074312

Kelly, T. & Souter, D. (2014) The Role of Information and Communication Technologies in Postconflict Reconstruction. World Bank. https://www.infodev.org/infodev-files/final_report_10_feb.pdf

Ojo, J. S. (2014). E-governance: An imperative for sustainable grass root development in Nigeria. Journal of Public Administration and Policy Research, 6(4), 77-89. http://www.academicjournals.org/app/webroot/article/article1413993734_OJO.pdf

Paul, A. Jolley, C. & Anthony, A. (2018) Reflecting the Past, Shaping the Future: Making AI Work for International Development. USAID. https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf

van Niekerk L, Chater R, Naydenova E, Lim J, Chamas L, Manderson L, Gilson L, Peeling R, Hartigan P, Bonnici F. (2017) Social innovation in health: case studies and lessons learned from low- and middle-income countries. WHO. ISBN 978-92-4-151309-8. https://apps.who.int/iris/bitstream/handle/10665/259187/9789241513098-eng.pdf

Vester Haldrup, S. (2018a) Digitising public service delivery: opportunities and limitations. Oxford Policy Management (OPM) https://www.opml.co.uk/blog/digitising-public-service-delivery-opportunities-and-limitations

Vester Haldrup, S. (2018b) When and how to use digital tech to improve public service delivery. OPM. https://www.opml.co.uk/blog/when-and-how-to-use-digital-tech-to-improve-public-service-delivery

World Bank (2018) Improving public sector performance through innovation and inter-agency coordination. World Bank http://documents.worldbank.org/curated/en/833041539871513644/pdf/131020-WP-P163620-WorldBankGlobalReport-PUBLIC.pdf

Acknowledgements

Suggested citation

About this report