Decision-making in the Age of the Algorithm

Three key principles to help public sector organisations make the most of AI tools

Thea Snow

November 2019

nesta

This report was written by **Thea Snow**, Senior Programme Manager, Government Innovation.

Acknowledgments

A big thank you to the following people and organisations for their contributions to this report:

Alexander Babuta – Research Fellow, National Security and Resilience, Royal United Services Institute for Defence and Security Studies

Daniel Berliner – Associate Professor of Political Science and Public Policy in the Department of Government at the London School of Economics

Hannah Celia – Director, Xantura

Heeral Dave – Ernst & Young UK&I Advisory

Davin Parrot – Principle Data Scientist, West Midlands Police

Wajid Shafiq – CEO, Xantura

Shu Fei Wong – Ernst & Young UK&I Advisory

Lisa Barclay, **Theo Bass**, **Aleks Berditchevskaia**, **Vicky Clayton**, **Carrie Deacon**, **Hessy Elliot**, **Joel Klinger**, **Sinead MacManus**, **Helen Mthiyane**, **Rosalyn Old**, **Jack Orlick**, **Vicki Sellick**, **Tom Symons**, **Kyle Usher** – Nesta

About Nesta

Nesta is an innovation foundation. For us, innovation means turning bold ideas into reality and changing lives for the better.

We use our expertise, skills and funding in areas where there are big challenges facing society.

Nesta is based in the UK and supported by a financial endowment. We work with partners around the globe to bring bold ideas to life to change the world for good.

www.nesta.org.uk

If you'd like this publication in an alternative format such as Braille or large print, please contact us at: **information@nesta.org.uk**

nesta

Design: **Green Doe Graphic Design Ltd**

Decision-making in the Age of the Algorithm

Three key principles to help public sector organisations make the most of AI tools

Executive summary

Frontline practitioners in the public sector – from social workers to police to custody officers – make important decisions every day about people's lives. Operating in the context of a sector grappling with how to manage rising demand, coupled with diminishing resources, frontline practitioners are being asked to make very important decisions quickly and with limited information.

Recognising this, public sector organisations are turning to new technologies to support practitioner decision-making. In particular, predictive analytics tools, which use machine learning algorithms to discover patterns in data and make predictions, are being introduced to support more timely and accurate decision-making.

While statistical tools have existed for many decades to support frontline decision-making, predictive analytics tools are novel in a number of ways. Unlike traditional statistical tools, machine learning tools incorporate qualitative data, such as casenotes, into their analysis. In addition, predictive analytics tools are dynamic, constantly learning and adjusting their predictions based on new information.

There is no question that the introduction of predictive analytics tools into complex fields like child protection or policing carries significant risks – most notably, risks around algorithmic bias and an absence of firmly established governance and legal frameworks.

For this reason, this report focuses on a different issue, one that attracts far less attention but is equally important – the issue of human-machine interaction. How people are working with tools is significant because, simply put, for predictive analytics tools to be effective, frontline practitioners need to use them well.

This report suggests that the optimal way for practitioners to work with predictive analytics tools is to use the insights from the tool together with their own professional judgment, an approach I have called 'artificing'. This approach encourages a constructive friction between human and machine intelligence and means that practitioners' decisions are informed by the depth of human insight plus the breadth of big data analytics.

This report, though, is more than just a think-piece. Based on insights drawn from an extensive literature review, interviews with frontline practitioners, and discussions with experts across a range of fields, this report identifies three key principles which play a significant role in supporting a constructive human-machine relationship: context, understanding, and agency.

Drawing on these principles, this report offers a practical guide, together with a summary checklist, to support public sector organisations introducing predictive analytics tools to do so in a way that means they will be embraced and used by frontline practitioners but also questioned, scrutinised and challenged as they should be.

Given the nascent nature of this field, this guide does not purport to be comprehensive or complete; rather, it aims to catalyse new ways of thinking and new discussions – it is designed to be the beginning of a conversation, not the end.

Ultimately, it aims to encourage public sector organisations to think about how humans feel about algorithmic tools – what they're fearful of, what they're excited about, what they don't understand. Approaching the deployment of these tools in a way that is mindful of the people being asked to work with them offers the best chance of tools being used in a way that combines the best of both human and machine intelligence.

Background

Every day, frontline practitioners working in the public sector make important decisions about people's lives. Social workers decide which families need extra support. Police make decisions about which streets to patrol and who to arrest. Bail officers determine who should be granted and denied bail.

These decisions are made under what has been called 'conditions of uncertainty'.[1] In practical terms, what this means is that frontline practitioners are making fast decisions with limited information, often relying on their intuition to guide their decision-making.[2]

In addition to working with imperfect information, public sector organisations are having their budgets cut whilst also dealing with growing demand. These pressures mean that frontline practitioners are being asked to work faster and allocate resources in a more targeted and efficient way; in other words, to do more with less.

In recognition of this, public sector organisations are turning to new technologies to support practitioner decision-making. In particular, predictive analytics tools are being introduced across a broad range of fields as a mechanism to support more timely and accurate decision-making.[3]

What is predictive analytics?

Predictive analytics refers to the application of machine learning algorithms to mine data, create models, and analyse existing data to discover patterns and make predictions.[4]

While actuarial tools have been around for decades in certain fields (for example assessing criminality or credit risk),[5] the growing ease with which we can capture, store, and process vast quantities of data means that these tools are beginning to extend their reach into new fields.

In addition, predictive analytics differ from traditional statistical methods in a number of ways. Firstly, recent advancements in fields such as Natural Language Processing mean it is possible for these new tools to incorporate qualitative data, such as casenotes, into their analysis, something traditional actuarial tools could not do. A second – and probably more important – difference is that while traditional actuarial tools use static and historical risk factors to inform their analysis, predictive analytics tools use dynamic data and adjust their models continuously to reflect the new data that they are being given to work with. This is why they're called 'learning models'.[6]

A number of examples of predictive analytics tools are set out below.

Predictive Analytics in Healthcare

In Australia, the CSIRO[7] has developed the Patient Admission Prediction Tool (PAPT),[8] which predicts emergency department patient arrivals, their medical urgency, admissions and likely discharge times.

PAPT provides a predictive picture of patient movement through the hospital. Patient load can be calculated with precision on a daily, weekly or monthly basis and demand can be forecast up to six months in advance.

The objective of PAPT is to enable hospitals to improve resource allocation efficiency, reduce waiting times, and increase timely access to emergency care.

Automated Traffic Control

The City of Pittsburgh collaborated with Rapid Flow Technologies to develop SURTRAC (Scalable Urban Traffic Control), an automated traffic optimization and control software.[9] SURTRAC uses traffic camera data and radar data, together with a dynamic programming search algorithm, to find the optimal number and sequence of vehicles on the road and determine how long each green light should last based on that order.[10]

City traffic control departments in Pittsburgh can use SURTRAC to manage traffic flows through several intersections and use AI to optimize the traffic systems toward reduced travel times, reduced number of traffic stops, and reduced wait times.

Predicting Fires

In New Orleans, the Fire Department worked with the Office of Performance and Accountability to develop a predictive analytics tool that could identify fire risk.[11]

Drawing on Census Bureau's American Housing Survey (AHS), the city was able to identify variables that could predict the likely presence of a smoke alarm. It then used this information to target the distribution of free smoke detectors to houses in areas that were identified as least likely to have one.

Predictive risk modelling is a subtype of predictive analytics. Predictive risk modelling uses machine learning to identify the likelihood that an individual will experience a specific outcome or event.[12]

Again, a number of illustrative examples are set out below:

Predictive Risk Modelling in Children's Services

The Allegheny Family Screening Tool[13] is a predictive risk model designed to improve decision-making in Allegheny County's child welfare system.

Designed by a team from Auckland University of Technology, the tool draws on hundreds of data elements to predict the likelihood that a child referred to children's services will later experience a foster care placement.

The tool generates risk scores for social workers to refer to when making decisions about at-risk children, and is designed to support social workers who are deciding whether to escalate or de-escalate a child's case.

Predictive Risk Modelling in the Justice System

The COMPAS tool,[14] which has been applied across various jurisdictions in the United States, uses an algorithm to assess an offender's potential recidivism risk. The variables which inform the tools' analysis have been kept private by the tool designers.

The tool produces a risk score, which is then used by judges to inform decisions around bail and sentencing. COMPAS is also used more broadly by justice agencies to inform decisions regarding the placement, supervision and case management of offenders.

How to make the most of predictive analytics tools

For predictive analytics tools to be effective, frontline practitioners need to use the tools well. While this may sound fairly self-evident, the question of how practitioners are using tools – referred to as the 'human-machine interaction' – is often overlooked. It is notably absent from much of the guidance material produced to support the introduction of AI tools into the public sector.

Interactions between humans and predictive analytics tools are complex. Some frontline practitioners resent the introduction of these tools, and distrust them, while others feel completely intimidated. All of these feelings serve to undermine constructive human-machine interactions, and mean that frontline practitioners often don't use predictive analytics tools well.

But what does 'using these tools well' even mean? And what needs to happen to encourage practitioners to use the tool in the optimal way? The purpose of this guide is to answer both of these questions.

This guide has been designed to:

1. Describe the optimal way for frontline practitioners to use predictive analytics tools, and explain why this approach is preferred.
2. Offer public sector organisations practical guidance to encourage practitioners to adopt this approach.

Human-machine interaction – the optimal approach

The optimal way for frontline practitioners to use predictive analytics tools is to combine the tool's insights with their own professional intuition.[15] I have called this approach 'artificing'.*

A practitioner can be seen to be artificing when they take the information provided by the tool into consideration, combine that with their own judgment, and then come to a decision based on a synthesis of both inputs.

Artificing as an approach is encouraged because, while predictive analytics tools are very good at certain things – for example processing and synthesizing vast quantities of data and spotting patterns – they also have significant limitations.[16] In particular, they are:

- Not good at predicting rare events.[17]
- Often trained on incomplete data.[18]
- Often trained on biased data, resulting in discriminatory tools.[19]

*The term 'artificing' (a term I have created) is an extension of the concept of 'satisficing', which was introduced to decision-theory by Herbert Simon in the 1950s. It offers a single word to describe the optimal way for practitioners to work with predictive analytics tools - a combination of rational analysis together with expert intuition. A more detailed explanation is available in this blog. www.nesta.org.uk/blog/human-vs-machine-why-we-need-be-thinking-more-about-how-humans-are-interacting-ai-tools/

These shortcomings mean that active human involvement in decision-making is essential; humans must challenge, scrutinise and question the guidance that the tool is offering.

In addition to encouraging humans to act as a check-and-balance, artificing is recommended as the preferred approach because it reserves a place for some uniquely human skills in frontline decision-making. A machine can never know how it feels to be another person.[20] Humans bring empathy, insight and practice-knowledge to the decision-making process.

Artificing recognises that while algorithms add breadth to decision-making, humans add a critical depth.

Human-machine interaction in practice

Despite there being a preferred way to use the tool, research has revealed that not all practitioners are artificing. Interviews with frontline practitioners revealed the following:[21]

1. More than one-third of practitioners were ignoring the tool. This is known as algorithm aversion.[22]
2. Despite there being a fear in the media and commentary that practitioners will defer to the advice of tool (known as automation bias),[23] this was very uncommon. A key reason for this appeared to be that deference was being explicitly and very strongly discouraged in tool training. However, a number of practitioners expressed fears that deference will likely become an issue as the tool becomes more embedded in practice and the novelty (and therefore caution) of using the tool wears off.
3. Many practitioners draw on both professional judgment and unconscious bias[24] to inform the intuitive element of their decision-making. This is significant because it challenges claims that algorithmic tools will bring *"scientific order and consistency to… decision-making practice"*,[25] and highlights that bias endures as a feature of human decision-making, despite the introduction of predictive analytics tools.

The key question to be answered here is: what needs to happen to support more frontline practitioners to artifice? Why are some practitioners artificing, while others are ignoring the tool? And what can we do to support more practitioners to use the tool as it was designed to be used?

Supporting a productive human-machine interaction

Interview[26] and workshop findings[27] revealed three key principles which appear to play a significant role in shaping how humans interact with predictive analytics tools:

Context

Introducing the tool with awareness and sensitivity to the broader context in which practitioners are operating increases the chances that the tool will be embraced by practitioners.

Understanding

Building understanding of the tool means practitioners are more likely to incorporate its advice into their decision-making.

Agency

Introducing the tool in a way that respects and preserves practitioners' agency encourages artificing.

If these principles are respected in the introduction of predictive analytics tools, public sector organisations will be much more likely to see the tools embraced and used wisely.

The remainder of this guide is dedicated to helping public sector organisations translate these principles into practice.

The intention of this guide is to offer something other than another think piece; it aims to convert thinking into action, offering a practical tool which outlines both the barriers facing public sector organisations looking to introduce predictive analytics tools to do so in a way which supports a constructive human-machine interaction, as well as the opportunities to address those barriers.

In addition to the guide, a summary checklist has been created, which is designed to be printed, shared and used as a guide and prompt throughout the process of introducing predictive analytics tools into the public sector.

How to use this guide

This guide will work most effectively if it is:

- Adapted to suit the context in which it is being used
- Used at all levels of the organisation – from the executive team, to the frontline practitioners themselves
- Used before, during and after the introduction of the tool; it is as much a document designed to trigger thought and conversation as it is one to offer guidance
- Treated as a dynamic, rather than static, document. The use of predictive analytics in the public sector is an emerging field, and no doubt new barriers and opportunities will emerge. Public sector agencies are invited to add and modify the barriers and opportunities as new findings emerge. To encourage this, a number of blank spaces have been left for you to add your own ideas against each opportunity.

What this guide is not

There are many significant challenges and risks associated with introducing predictive analytics tools into the public sector. Media,[28] experts[29] and academics[30] have raised concerns about the considerable risks posed by algorithmic bias.[31] Other common criticisms include concerns that:

- Poor quality data will create poor quality tools (garbage in = garbage out)
- The challenges of data-sharing across public sector agencies means that critical datasets will likely be omitted from the tool's inputs
- An absence of firmly established governance and legal frameworks[32] means it is very difficult to enforce the accountability and transparency of predictive analytics tools.[33]

Though these concerns are of critical importance, the conversation and debate around them are already well-established in these domains. For example, a recent AI Ethics Guidelines Global Inventory lists over 80 different ethical frameworks.[34] There are also a considerable number of toolkits and resources available which have been designed to support the development of responsible and high-quality algorithmic tools.[35] The Resources section of this report points to some helpful guides to refer to if there are issues or concerns around tool design which are not yet resolved.

On this basis, this guide is deliberately narrow and specific, focusing on a topic which receives far less attention; namely, how to introduce predictive analytics into workplaces in a way which means they will be embraced and used by frontline practitioners, but also questioned, scrutinised and challenged as they should be.

PRINCIPLE 1: **Context**

Introducing the tool with awareness of the broader context in which practitioners are operating means the tool will be introduced with greater sensitivity to local conditions, an awareness of how practitioners are working, and the systems they are already working with. Accounting for these contextual factors increases the chances of the tool being embraced by practitioners.

Checklist:

1. Get the foundations in order
2. Keep it simple
3. Create conditions for success
4. Human bias persists

1. Context: **Get the foundations in order**

Barriers

The success of these tools relies on having the appropriate IT infrastructure in place to support the tool working consistently and efficiently.

Interviews revealed that technical problems were inhibiting practitioner use of the tool. One practitioner who was travelling to schools in rural England explained that she was often unable to use the tool because of poor connectivity. More generally, poor wifi, old computers and programme glitches were also highlighted as a common challenge, which contributed to practitioners ignoring the tool.

Opportunities

- Don't assume that legacy systems will be able to support new tools – do the research and user-testing needed to ensure IT systems are able to support the tool.
- Test the tool at the sites where it will be used – schools, hospitals, regional locations, etc. Limiting testing to sites with good connectivity will conceal potential challenges, which are important to unearth.

2. Context: **Keep it simple**

Challenges

Frontline practitioners are generally time poor and working in high pressure environments where the stakes are high. They want to avoid any additional mental load and will resist tools that add extra steps and/or complexity to their everyday practices and processes.

In addition, as part of field research, frontline practitioners described experiencing 'tool fatigue' – they are tired of being asked to work with a multitude of tools. The last tool to be introduced is at greatest risk of being ignored, particularly if it is added in without any rollback of some of the other tools being used.

Opportunities

- Make the tool as frictionless as possible.* For example, avoid asking practitioners to access a different screen/database/system to view the tool. Keep the interface simple, attractive and intuitive. Use simple language and short sentences.
- Engage with practitioners as part of the tool design process – adopt a 'user-centred design approach' (see Resources for tools to support this approach). Practitioner input should inform what a frictionless tool looks like and how it fits most comfortably within their practice.**
- When introducing new tools, try to replace one with another, rather than just adding more into the mix.

*For practical guidance on building a frictionless tool see https://www.webdesignerdepot.com/2018/05/frictionless-ux-how-to-create-smooth-user-flows/

**A user-centred design approach is also discussed as a discrete element below at page 28.

3. Context: **Invest in action, not just insight**

Barriers

There is a fear that these tools will lead to the identification of more at-risk people, without a commensurate investment being made in the resources, services and programmes needed to support them.

Eileen Munro, who has written extensively on the use of predictive analytics in child protection, poses two very important threshold questions which must be considered as part of any decision to introduce predictive analytics tools:[36]

1. Do we have effective methods to deal with identified needs?
2. Do we have sufficient resources to address the identified needs?

If both of these questions cannot be answered in the affirmative, frontline practitioners will likely be reluctant to use the tool; there is a real anxiety around not being able to adequately support the people being identified as high risk by the tool because services are already so stretched.

Opportunities

- Spend time thinking about the programmes and processes that are in place to support those identified by the tool as being at risk – frontline workers need to know they have the resources and tools available to them to effectively respond to whatever risk is flagged.
- When designing processes to gather practitioner feedback on the tool, include a question about whether practitioners feel the necessary programmes and resources exist to support them to act on the tool's output.

4. Context: **Human bias persists**

Barriers

Human bias in decision-making endures, despite the introduction of predictive analytics tools. Interviews revealed that when practitioners artifice they will, in the vast majority of cases, draw on elements of professional intuition, together with unconscious bias, to inform the intuitive element of the decision-making process.

Interviews also revealed that practitioners are more likely to artifice when the tool confirms their pre-existing views and are more likely to ignore the tool when it doesn't conform to their preconceptions. This is known as confirmation bias: the phenomenon whereby people only look for evidence that confirms what they already think.[37]

This highlights the fact that AI tools are not a silver bullet. Simply giving practitioners more objective information to work with does not mean irrational biases disappear. As such, efforts to cultivate professional expertise and reduce bias in decision-making remain of the utmost importance, even following the introduction of predictive analytics tools.

Opportunities

- Investing time and energy in developing practitioner expertise will likely reduce the extent to which practitioners rely on unconscious bias. To build expertise, practitioners need:[38]
  - Opportunities to reflect on their practice
  - Feedback on their decision making.*
- Continue (or start) to invest in unconscious bias training (while there is debate about the efficacy of this training, the Resources section of the report includes suggestions and articles which address this point).
- During training, make clear that it is not appropriate to ignore the tool simply because the information being presented sits in tension with the practitioner's own assessment. And, conversely, that it is not appropriate to rely too heavily on the tool simply because information supports the practitioner's views.

*Feedback on frontline practitioner decision-making has traditionally been irregular and of a low quality. This makes it difficult for frontline workers to cultivate skilled intuition, because developing expertise relies on receiving high quality feedback on decisions, and learning from mistakes. Predictive analytics tools may hold at least part of the answer. Predictive analytics tools record outcome data for every case. This data can be used to offer feedback to practitioners on the effectiveness (or otherwise) of their decisions. Having this data available should make it easier for public sector bodies to provide practitioners with regular and high quality feedback, which is critical to cultivating expertise.

PRINCIPLE 2: **Understanding**

Building understanding of the tool leads to greater trust, meaning practitioners are more likely to incorporate its advice into their decision-making. At the same time, understanding the tool also means understanding its limitations, which minimises the likelihood that practitioners will simply defer to it.

Checklist:

5. Show me the data
6. Demonstrate value
7. Feedback loops

5. Understanding: **Show me the data**

Barriers

Practitioners who have a good grasp of the tool's data sources, and who can verify that data is sufficiently current, appear to artifice. In contrast, practitioners who do not understand where data is sourced from, or how up to date it is, are more likely to ignore the tool.

Understanding data sources means understanding both where the data is coming from – in other words, which agencies' data is being used to train the tool – as well as the type of data that the tool draws on to inform its analysis. Interviews revealed a common misapprehension amongst frontline practitioners that predictive analytics tools draw only on quantitative data. In fact, many of these tools use natural language processing techniques to incorporate data from case notes and other qualitative sources. The tool's perceived failure to incorporate qualitative data was associated with a reluctance to consider the tool's advice.

While the tool interface (what it looks like) has a significant role to play in supporting better practitioner understanding, interviews revealed that skill also appears to play a significant role. Practitioners who possess core data literacy skills are more likely to work productively with new technologies than those who do not, for example.

Opportunities

- Design the tool in a way that offers practitioners an option to view data sources and data currency. A simple way to achieve this would be a drop down menu which shows (1) which agencies the data is sourced from (2) the sources of the data (e.g. casenotes) (3) when the data was last updated.
- Establish processes to allow practitioners to request the inclusion of additional datasets.
- As part of practitioner training, take time to explain that the tool has the capacity to draw on both qualitative and quantitative data sources. Explore and explain to practitioners where the tool gets its information from and how it makes decisions; compare this to how practitioners source their information and make decisions, highlighting similarities and differences.
- Invest in data literacy training for practitioners. This training could be developed and run by in-house teams, procured from external providers, or practitioners could be given time to complete one of the many online courses that exist (see Resources for suggestions).

6. Understanding: **Demonstrate value**

Barriers

While some practitioners are able to see the value a tool offers to support better decision-making, many cannot. It cannot be assumed that the tool's value will be immediately obvious to practitioners.

Practitioners will be reluctant to use a tool if they do not appreciate its value-add.

It is important to highlight and explain why it is that machines do some things better than humans, while humans do other things better than machines; this is why artificing is being promoted as the optimal way to use these tools.

For this message to really resonate, it is important for the tool's value to be promoted by leadership teams, managers and peers within the organisation.

Opportunities

- Take time to explain what predictive analytics tools are, how they are different to other tools, and how they will support better decision making by identifying risks that may otherwise have been missed, enabling a more preventative, rather than reactive, practice approach. The key points to be communicated are that these tools:
  1. Are dynamic, constantly learning and adjusting their output to reflect the new information being fed into the tool
  2. Have been proven, in certain fields, to be more accurate at predicting risk than humans[39] (be careful here – we don't want to push people towards automation bias!)
  3. Have enormous speed and processing power – far beyond the capacity of humans
  4. Are able to spot patterns and draw connections that humans can't
  5. Work to break down unhelpful silos by drawing on and synthesizing cross-agency information.
- Don't just talk to these points. To make them real, use exercises, simple analogies and stories. For example:
  - Use analogies to demonstrate the valuable role that many machines already play in our lives every day, and which we readily accept as a society. For example, there would be near universal acceptance that a computer will add the purchases we make at a supermarket more quickly and accurately than a clerk could.[40] This highlights that there are some tasks which machines are better at than humans
  - Use storytelling to show the impact tools like this can have. Use case studies to show real examples of people at risk who have been identified and helped as a result. Try to create an emotional connection between the practitioner and the tool to build the value proposition
  - Try the following exercise: print 30 different pages of information (fictional) about a particular individual. Include a variety of documents in different formats – reports, spreadsheets, etc. Ask practitioners to synthesize that information into a one-page briefing within ten minutes, focussing on key risk factors. Follow with a reflective discussion.
- The tool's value proposition must be communicated by managers and the leadership team, not an external consultant or tool developer. Hearing the value proposition from someone internal, who understands the specific context and knows the practitioners, is much more meaningful and likely to resonate.
- Set up a 'champions group' comprised of staff from different teams and of different levels of seniority. This group should be tasked with championing the tool amongst their peers.
- In promoting the strengths of these tools, it is also important to remind practitioners that they are not infallible. For example, as mentioned above, tools are bad at predicting sudden and rare events: humans are much better at this. For this reason, it should be emphasized that these tools require a human in the loop[41] to act as a check-and-balance.

7. Understanding: **Feedback loops**

Barriers

Feedback loops about the tool's performance are critical to building trust in the tool. Practitioners are reluctant to use a tool without understanding how the tool is performing.

However, the delivery and framing of this feedback must be managed very carefully as people are intolerant of machines making mistakes.[42] Care needs to be taken to explain that the tool is not right 100 per cent of the time – but neither are humans.

Opportunities

- Set up processes to share formal evaluations of the tool's accuracy with practitioners. Feedback about performance would be better shared face-to-face, rather than in written form, to allow practitioners to ask questions, clarify points of confusion, and offer the opportunity for a deeper discussion.
- The accuracy of predictive analytics tools should improve over time. These tools are built on algorithmic 'learning models', which mean that individual case-outcomes are fed back into the tool to improve predictive accuracy as time goes on. These improvements will not happen overnight, but it is important to share the tool's evolution with the practitioners who are working with it. Tool designers should be invited to present to practitioners either once or twice a year on the tool's evolution. This should be a chance for designers to either celebrate the tool's improvements or answer important questions if it is not improving as it should be.
- It is important not to rush rollout. Organisations should wait until the tool is reliable and has sufficient predictive accuracy before introducing it to the workplace. This is clearly desirable simply from a best-practice perspective, but also because introducing a tool which is seen to be 'buggy' or not accurate enough will likely lead to algorithm aversion.

PRINCIPLE 3: **Agency**

Introducing the tool in a way which respects and preserves practitioners' agency works against deference by reinforcing the important role that practitioners continue to play even once a predictive analytics tool has been introduced. This also guards against aversion by bestowing in practitioners a sense of empowerment, making them less likely to feel displaced, resentful of the tool, and prone to ignoring it.

Checklist:

8. Co-design and iteration
9. Navigating complexity
10. Encourage adaptation
11. Discourage deference… but not too much!

8. Agency: **Co-design and iteration**

Barriers

AI tools are highly technical and complex, meaning it often feels easier to 'leave it to the experts' and not involve frontline practitioners lacking in technical expertise in the design of the tool. However, involving users as part of the design process will increase the chances of the tool being adopted and used effectively.[43] This is because a user-centred design approach supports the development of a tool which meets user needs and reflects the unique operating environment and context. It also builds a sense of trust and sense of ownership in the tool.[44]

Involving practitioners as a one-off is unlikely to be successful; rather, what is needed is an ongoing dialogue and channel for practitioner feedback. It is also critical that this feedback is seen to be acted on – the design and application of the tool should iterate based on user feedback. This is particularly important in this context because research shows that people are more receptive to working with algorithmic tools when they are able to provide feedback.

Opportunities

- Involve practitioners not in the technical tool design, but in the design of the tool interface and in how the tool is used as part of their practice (see Resources for further recommendations on human-centred design). Questions to think about include:
  - What does a useful tool look and feel like for practitioners? User-testing and iteration is critical to support the development of a tool that feels intuitive and easy to use
  - Do practitioners want to use the tool as part of team meetings or individual practice?
  - At what point in the decision-making process do practitioners want to use the tool? In some cases, practitioners had decided they were most comfortable using the tool after they had formed an initial judgment about a given case; in other cases, practitioners were using the tool from the outset of the decision-making process.
- Use the champions group as a source of ongoing feedback about the tool. In addition, design processes to at least semi-regularly gather feedback from all practitioners who are using it – these may be surveys, interviews, or whatever else works for a given context.
- Processes should be put in place to allow practitioners to offer real-time feedback when they disagree with the information being provided by the tool. If the practitioners' criticism is deemed to be sound, this feedback can be used as training data.
- Be sure to communicate how practitioner feedback is being acted upon. Be prepared to iterate on the design and approach to using the tool over a number of years (or maybe forever!).
- Given that evidence suggests practitioners are likely to reject a tool if they see it make too many mistakes[45] or if they have repeated negative experiences, it is worth considering doing initial user-testing with volunteers rather than the practitioners who will actually be working with it. Professional bodies such as Frontline could be a source of volunteers in the field of children's services, while an organisation such as Citizens in Policing could be a source of volunteers for the police force.
- While introducing new tools into existing systems and practices may work, organisations also need to be open to changing work practices, norms and even workplace cultures in order to encourage full integration of the tool.[46] For example, practitioners may prefer using the tool as part of a team rather than individually, leading to more team-based decision-making. These evolutions may well be an unforeseen benefit.

9. Agency: **Encourage adaptation**

Barriers

There is no one 'right way' to artifice: interviews revealed that artificing can look different in different contexts.

In some places, tools are being used to support team-based decisions, meaning that artificing takes place in a team-based environment. In others, tools are used to support individual decision-making.

Artificing also varied at the individual level, with people adapting how they worked with the tool to suit their working style and approach. This variability is important for three reasons:

1. Artificing is promoted as the preferred way to use predictive analytics tools because it is important that uniquely human skills, such as empathy, continue to inform the decision-making process. Allowing an element of flexibility in how people are using the tool is essential to enabling this 'humanness' to express itself. In contrast, an overly prescriptive approach is likely to undermine practitioners' ability to express the very humanness that is seen to be an important counterbalance to the tool.
2. Accepting flexibility in how the tool is used provides practitioners with a sense of agency over the tool, which encourages a more constructive human-tool interaction. Tools which are seen to take away practitioner discretion are far less popular, and therefore less likely to be used, than those which preserve practitioner agency.
3. As mentioned above, the predictive capacity of AI tools vary from case to case. AI tools are not nearly as good at predicting rare events as they are common ones, simply because there is much less data available to train them on.[47] For this reason, it is likely to be appropriate to rely on the tool more in some cases than in others. Avoiding a tight definition of artificing gives practitioners more licence to determine the extent to which they rely on a tool in any given case.

Opportunities

- Offer different ideas around what it means and looks like to artifice. Practitioners who were artificing used a range of expressions to describe their approach. They described using the tool:
  - *"As part of a jigsaw"*
  - *"To build a richer picture"*
  - *"To look more deeply"*
  - *"Together with professional curiosity"*
  - *"To trigger earlier conversations"*
  - *"To look at things through a different lens"*
  - *"To reassess"*
  - *"Like the advice of an expert colleague"*
  - *"As an additional source of information"*
  - *"As part of a broader suite of tools".*

All of these approaches should be endorsed as a valid approach to using the tool.

- Consider setting up peer-to-peer support groups, which would offer practitioners working with the tool the opportunity to share how they are working, discuss what works well and what is frustrating, and offer informal support and advice to each other. The idea is that practitioners could share and learn from each other about the different approaches to working with predictive analytics tools.

10. Agency: **The art of artificing**

Barriers

Artificing is no simple task. How do frontline practitioners know how much to rely on a tool? When should they ignore a tool's advice? When should they change their mind about a case based on a tool's insights?

Practitioners need to be supported, given that artificing isn't straightforward – there is no 'right way' to artifice, and trade-offs are necessary.

Opportunities

- As part of training, take people through scenarios where the tool is in tension with their own assessment. Get them to think and talk through their reasoning.

  Example (in a children's services context):

  "You are reviewing the case of a ten-year-old girl. There are many factors in the referral report that you consider to be concerning. You decide that there are strong grounds for referring the case for further investigation. You look at the tool and the risk score is a four (out of 20). Can you talk me through what you would do, and why?"

- Incorporate discussions about the challenges of using the tool into existing supervision meetings. Often, practitioners find the process of talking through their thinking and reasoning to be very helpful.
- Again, peer-to-peer support groups could be used to allow practitioners to share any challenges they are facing in determining when, and to what extent, to rely on the information being provided by the tool, as well as any practices that they have found to be particularly helpful.

11. Agency: **Discourage deference… but not too much**

Barriers

A delicate balance needs to be struck: discouraging practitioner deference to the tool, but not to the point that they feel nervous to use it at all. Interviews indicated that public sector organisations were emphasizing the need to avoid deference so strongly that some frontline practitioners were interpreting this to mean that they shouldn't rely on the tool at all to inform their decision-making.

So while discouraging automation bias is key, it is also important not to push this message too hard.

Frontline practitioners must understand that while the tool should not be the decision-maker, it should certainly be a factor that they consider in making their decisions.

Opportunities

- To avoid an over-zealous approach to discouraging deference, take time to explain that the optimal way to use the tool is a combination of human and machine intelligence. Use easy-to-grasp language. Examples of language used by practitioners during interviews include:
    - The tool is *"not a deal-breaker"*
    - The tool is *"not gospel"*
    - *"It's about using the data but putting the human factor in as well"*.

  The phrases above are likely to discourage deference without leading to aversion, whereas a sentence like *"you should never use this tool to make a decision"* might frighten a practitioner into ignoring the tool.

  Again, for this message to have most resonance, it should be delivered by managers and the leadership team.

- Despite this, it is also important to make the risks of deference clear. Again, storytelling might be useful here. A fantastic story is shared by Hannah Fry in her book Hello World:[48]

  Stanislav Petrov was a Russian military officer in charge of monitoring the nuclear early warning system protecting Soviet airspace. His job was to alert his superiors immediately if the computer indicated any sign of an American attack.

  Petrov was on duty on 26 September 1983 when, shortly after midnight, the sirens began to howl. This was the alert that everyone dreaded. Soviet satellites had detected an enemy missile headed for Russian territory. This was the depths of the Cold War, so a strike was certainly plausible, but something gave Petrov pause. He wasn't sure he trusted the algorithm. It had only detected five missiles, which seemed like an illogically small opening salvo for an American attack.

  Petrov froze in his chair. It was down to him: report the alert, and send the world into almost certain nuclear war; or wait, ignoring protocol, knowing that with every second that passed his country's leaders had less time to launch a counter-strike.

  Fortunately for all of us, Petrov chose the latter. He had no way of knowing for sure that the alarm had sounded in error, but after 23 minutes – which must have felt like an eternity at the time – when it was clear that no nuclear missiles had landed on Russian soil, he finally knew that he had been correct. The algorithm had made a mistake.

- In order for practitioners to have the confidence to push back on a predictive analytics tool, a culture of openness must exist. A workplace culture must be cultivated in which frontline practitioners are encouraged to challenge decisions – whether that be the decisions of other practitioners, their supervisors, or a predictive analytics tool.
- Consider setting up a specific process for instances where practitioners disagree with the information provided by the tool – for example requiring peer review for all of these cases.

Conclusion

This guide, and the accompanying checklist, has been created to support public sector bodies introducing predictive analytics tools to do so in a way which supports practitioners to use the tool as it was designed to be used – together with their expert judgment – and avoid the situation where practitioners ignore, or defer to, the tool.

To do this, public sector agencies will need to think deeply about how to introduce tools in a way which:

- Is sensitive to local context
- Invests in building practitioner understanding
- Respects and preserves practitioner agency.

The assumption underpinning this guide is that public sector bodies will be working to ensure that the tool being deployed is ethical, high quality and transparent. The Resources section below points to some of the frameworks that can guide the development and/or procurement of ethical predictive analytics tools.

Another core assumption underpinning this guide is that the use of predictive analytics tools in the public sector is inevitable and will continue to grow.

In focussing on what needs to happen to successfully support a constructive relationship between practitioners and the tools they are being asked to work with, this guide plugs a gap within the existing landscape of tools, guidance notes and policy papers.

Given the nascent nature of these tools, rather than attempting to be a comprehensive document, this guide is designed to be a conversation starter; a nudge to encourage public sector bodies to start thinking about these tools in a human-centered way.

Ultimately, this report aims to encourage public sector organisations to think about how humans feel about these tools: what they're fearful of, what they're excited about, and what they don't understand.

Approaching the deployment of these tools in a way that is mindful of the people who are being asked to work with them offers the greatest chance to combine the best of human and machine intelligence, bringing greater breadth and depth to the decisions of frontline workers – decisions that can and do change the course of people's lives.

Summary checklist

The artificing checklist: Guidance to support public sector organisations make the most of AI tools

(Write more 'opportunities' on the dotted lines)

Context

Element 1: Get the foundations in order

Barriers

The success of these tools relies on the appropriate IT infrastructure being in place. Interviews revealed that technical problems were inhibiting practitioner use of the tool.

Opportunities

- Don't assume that legacy systems will be able to support new tools - do the research and user-testing needed to ensure that IT systems are able to support them.
- Test the tool at the sites where it will be used – schools, hospitals, regional locations. Limiting testing to sites with good connectivity will conceal potential challenges, which are important to unearth.
- ……………………………………………
- ……………………………………………
- ……………………………………………
- ……………………………………………

Element 2: Keep it simple

Barriers

Frontline practitioners are generally time poor and working in high pressure environments where the stakes are high. They want to avoid any additional mental load and will resist tools which add extra steps and/or complexity to their every-day practices and processes.

In addition, frontline practitioners are experiencing 'tool fatigue' – they are tired of being asked to work with a multitude of tools.

Opportunities

- Make the tool as frictionless as possible (see full guide for further detail).
- Engage with practitioners as part of the tool design process – adopt a 'user-centred design approach' (see Resources for tools to support this). Practitioner input should inform what a frictionless tool looks like and how it fits most comfortably within their practice.
- When introducing new tools, try to replace one with another, rather than just adding more into the mix.
- ……………………………………………
- ……………………………………………
- ……………………………………………

Element 3: Invest in action, not just insight

Barriers

There is a fear that these tools will lead to the identification of more at-risk people without a commensurate investment being made in the resources, services and programmes needed to support them.

Two threshold questions must be considered as part of any decision to introduce predictive analytics tools:

1. Do we have effective methods to deal with identified needs?
2. Do we have sufficient resources to address the identified needs?

If both of these questions cannot be answered in the affirmative, frontline practitioners will likely be reluctant to use the tool.

Opportunities

- Spend time thinking about the programmes and processes that are in place to support those identified by the tool as being at risk – frontline workers need to know they have the resources and tools available to them to effectively respond to whatever risk is flagged.
- When designing processes to gather practitioner feedback on the tool, include a question about whether practitioners feel that the necessary programmes and resources exist to support them to act on the tool's output.
- ……………………………………………
- ……………………………………………

Element 4: Human bias persists

Barriers

Human bias in decision-making endures, despite the introduction of AI tools.

AI tools are not a silver bullet. Simply giving practitioners more objective information to work with does not mean irrational biases disappear.

As such, efforts to cultivate professional expertise and reduce bias in decision-making remain of the utmost importance, even following the introduction of predictive analytics tools.

Opportunities

- Invest in practices which support practitioners to cultivate expertise. This means offering opportunities to reflect on their practice, and receive feedback on their decision making (the data collected by predictive analytics tools actually offers new insights which can be used as the basis for practitioner feedback. See full guide for further detail).
- Continue (or start) to invest in unconscious bias training (while there is debate about the efficacy of this training, the Resources section of the report includes suggestions and articles which address this point).
- During training, make clear that it is not appropriate to ignore the tool simply because the information being presented sits in tension with the practitioner's own assessment. And, conversely, that it is not appropriate to rely too heavily on the tool simply because information supports the practitioner's views.
- ……………………………………………
- ……………………………………………

Understanding

Element 5: **Show me the data**

Barriers

Practitioners are reluctant to use tools without understanding which data sources inform the tools' analysis and without having some way of verifying data currency.

While the tool interface (what the tool looks like to practitioners) has a significant role to play in supporting better understanding of the tool, interviews revealed that skill also appears to play a significant role. Practitioners who possess certain data literacy skills are more likely to work productively with new technologies than those who do not.

Opportunities

- Design tools in a way which enables practitioners to view data sources and data currency. For example, a drop down menu which shows (1) which agencies the data is sourced from (2) the sources of the data (e.g. casenotes) (3) when the data was last updated.
- Set up processes to allow practitioners to request the inclusion of additional datasets.
- Talk through data-sources as part of training.
- Invest in data literacy training for practitioners (see Resources for suggestions).
- ....
- ....
- ....

Element 6: **Demonstrating value**

Barriers

While some practitioners are able to see the value that the tool offers to support better decision-making, many cannot. It cannot be assumed that the tool's value will be immediately obvious to practitioners.

Practitioners will be reluctant to use a tool if they cannot see its value-add.

Opportunities

- Explain what predictive analytics tools are and how they support better decision-making (see full guide for more detailed guidance).
- Use stories, analogies and exercises to make the value-proposition clear and relatable (see full guide for further suggestions).
- Managers and the leadership team should lead conversations about the tool's value; not consultants or tool developers.
- Set up a 'champions group' comprised of staff of different levels of seniority, and from different teams. This group should be tasked with championing the tool amongst their peers.
- In promoting the strengths of these tools, it is also important to remind practitioners that they are not infallible. Emphasize that these tools require a human in the loop to act as a check-and balance.
- ....
- ....
- ....

Element 7: **Feedback loops**

Barriers

Feedback loops about the tool's performance are critical to building trust. However, the delivery and framing of this feedback must be managed very carefully because people are very intolerant of machines making mistakes.

Care needs to be taken to explain that the tool is not right 100 per cent of the time – but neither are humans.

Opportunities

- Set up processes to share any formal evaluations of the tool's accuracy. Feedback about performance would be better shared face-to-face, rather than in written form, to allow practitioners to ask questions, clarify points of confusion, and offer the opportunity for a deeper discussion around the tool's performance.
- Tool designers should be invited to present to practitioners either once or twice a year on the tool's evolution. This should be a chance for designers to either celebrate the tool's improvements or answer important questions if the tool is not improving as it should be.
- It is important not to rush the rollout of the tool. Organisations should wait until the tool is reliable and has sufficient predictive accuracy before introducing it to the workplace.
- ....
- ....
- ....

Agency

Element 8: Co-design and iteration

Barriers

AI tools are highly technical and complex, meaning that it often feels easier to 'leave it to the experts' and not involve frontline practitioners, who lack technical expertise, in the design of the tool. However, involving users as part of the design process will increase the chances of the tool being adopted and used effectively.

Involving practitioners as a one-off is unlikely to be successful; rather, what is needed is an ongoing dialogue and channel for practitioner feedback. It is also critical that this feedback is seen to be acted on.

Opportunities

- Involve practitioners in design of user-interface and in how the tool is used as part of their practice (see full guide for some key questions to consider).
- Use the champions group as a source of ongoing feedback about the tool. In addition, design processes to at least semi-regularly gather feedback on the tool from all practitioners who are using it.
- Establish processes to allow practitioners to offer real-time feedback when they disagree with the information being provided by the tool. If criticism is deemed to be sound, this feedback can be used as training data for the tool.
- Be sure to communicate how feedback is being acted upon. Be prepared to iterate on the design and approach to using the tool over a number of years (or maybe forever!).
- Consider doing initial user-testing with volunteers, rather than the practitioners who will actually be working with the tool, to avoid it being rejected as a result of being seen to make too many mistakes.
- Be open to changing work practices in order to use the tool to best effect.
- ....
- ....

Element 9: Encourage adaptation

Barriers

There is no one 'right way' to artifice. Interviews revealed that artificing can look different in contexts.

Encouraging flexibility in how practitioners are working with the tool is important for three reasons:

1. An overly prescriptive approach is likely to undermine practitioners' ability to express the very humanness that is a counterbalance to the tool
2. Allowing flexibility offers practitioners a sense of agency, which means they are less likely to resent and reject the tool
3. The tool's predictive capacity varies, meaning that it is appropriate to rely on the tool more in some cases than in others.

Opportunities

- Offer different ideas around what it means and looks like to artifice (see full guide for suggestions) and endorse all as being valid ways of using the tool.
- Consider setting up peer-to-peer learning groups to offer practitioners working with the tool the opportunity to share how they are working with the tool, what works well, what is frustrating, and offer informal support and advice to each other.
- ....
- ....
- ....

Element 10: The art of artificing

Barriers

Artificing is no simple task.

Practitioners need to be supported to artifice, given that artificing isn't straightforward – there is no 'right way' to artifice, and trade-offs are necessary.

Opportunities

- Use scenarios in training to support practitioners to practice navigating more complex cases (see full guide for an example).
- Incorporate discussions about the challenges of using the tool into existing supervision meetings.
- Again, peer-to-peer support groups could be used to allow practitioners to share any challenges they are facing in determining when, and to what extent, to rely on the information being provided by the tool, as well as any practices that they have found to be particularly helpful.
- ....
- ....
- ....
- ....

Element 11: Discourage deference... but not too much!

Barriers

A delicate balance needs to be struck between discouraging practitioner deference to the tool, but not to the point that practitioners feel nervous to use the tool at all.

Frontline practitioners must understand that while the tool should not be the decision-maker, it should certainly be a factor that they consider in making their decisions.

Opportunities

- To avoid an over-zealous approach to discouraging deference when introducing the tool, take time to explain that the optimal way to use the tool is a combination of human and machine intelligence (see full guide for specific suggestions).
- The risks of deference and the powerful check and balance that humans offer should be made clear. Again, storytelling might be useful here (see full guide for an example).
- Cultivate a workplace culture where frontline practitioners have the confidence to challenge decisions.
- Consider setting up a specific process for instances where practitioners disagree with the information provided by the tool. For example, requiring peer review of all of these cases.
- ....
- ....
- ....

Suggested resources

Data Literacy

Delivering Digital Skills: A guide to preparing the workforce for an inclusive digital economy (Nesta)

Data and Digital Directory: 100 places for public servants to learn digital skills for free (Apolitical)

The School of Data: Free courses on the fundamentals of data, as well as how to extract, map and clean it

The Open Data Institute: Paid and free classes in data essentials

General Assembly: A number of free courses in data analysis

Co-design/Co-production

Co-production catalogue: Designed to help practitioners learn about co-production practice (Nesta)

Human-Centred Service Design course: Five-week online course (IDEO, paid)

Crash Course in Human-Centred Design for Policymakers: How to find out what your users want (Apolitical)

Designing for Public Services: A collection of practical tools and methods for using design in public services (Nesta and IDEO)

Introduction to Human-Centered Design: Learn to use human-centered design for social innovation (IDEO.org)

The Field Guide to Human-Centered Design: A step-by-step guide that will get you solving problems like a designer (IDEO.org)

18F: Collection of Human-Centred Design Tools: A collection of methods to bring human-centred design into your project (18F Methods)

Co-production knowledge base: One-stop-shop for resources about all aspects of co-production (Co-Production Network for Wales)

Unconscious bias

Unconscious bias – from awareness to action: Three-week course (EdX)

Unconscious bias training: 30 minute masterclass on how to mitigate biases (Declic)

Nine online classes for Managers who care about diversity and inclusion: A range of free courses, including some specifically focussed on unconscious bias (via The Muse)

Cognitive bias cheat sheet: An article summarising and explaining the most common cognitive biases (Better Humans)

Don't give up on Unconscious Bias Training: Make it Better: an article addressing skepticism around unconscious bias training (Harvard Business Review)

Unconscious bias training: An assessment of the evidence for effectiveness: article which includes practical advice about how to make Unconscious Bias Training most effective (Equality and Human Rights Commission)

How to Minimize Your Biases When Making Decisions: Article about minimising bias in decision-making (Harvard Business Review)

Ethical AI tool design

AI Ethics Guidelines Global Inventory: An inventory of more than 80 international AI Ethics Guidelines (Algorithm Watch)

A guide to using artificial intelligence in the public sector: Guidance on building and using artificial intelligence in the public sector (GOV.UK)

Planning and preparing for artificial intelligence implementation: Guidance to help you plan and prepare for implementing artificial intelligence (GOV.UK)

Ethics and Algorithms Toolkit: A risk management framework for government (GovEx, the City and County of San Francisco, Harvard DataSmart, and Data Community DC)

Big data, artificial intelligence, machine learning and data protection: This discussion paper looks at the implications of big data, AI and machine learning for data protection, and explains the ICO's views on these (Information Commissioner's Office)

OECD Principles on AI

Data Ethics Framework: Guides the design of appropriate data use in government and the wider public sector (GOV.UK)

Data ethics canvas: A guide for anyone who collects, shares or uses data (ODI)

The Technology Code of Practice: A set of criteria to help government design, build and buy technology (GOV.UK)

The Data Nutrition project: Empowering data scientists and policymakers with practical tools to improve AI outcomes (MIT)

Endnotes

1. Kahneman, D. & Klein, G. (2009) Conditions for intuitive expertise: A failure to disagree. American Psychologist. 64 (6), 515–526.
2. Kirkman, E. & Melrose, K. (2014) Clinical Judgement and Decision-Making in Children's Social Work: An analysis of the 'front door' system. (DfE Research Report 323).
3. Shafiq, W. (2016) 'Algorithm guided decision making in the public sector', available at https://www.xantura.com/points-of-view/alogorithm-guided-decision-making-public-sector
4. Cuccaro-Alamin, S. et al. (2017) Risk assessment and decision making in child protective services: Predictive risk modeling in context. Children and Youth Services Review. 7 (9), 291–298.
5. Courtland, R (2018) 'Bias detectives: the researchers striving to make algorithms fair.' Nature. Available at https://www.nature.com/articles/d41586-018-05469-3.
6. Cuccaro-Alamin, S. et al. (2017) Risk assessment and decision making in child protective services: Predictive risk modeling in context. Children and Youth Services Review. 7 (9), 291–298.
7. See: https://www.csiro.au/
8. See: https://www.csiro.au/en/About/Our-impact/Our-impact-in-action/Health/Patient-care
9. Snow, J. (2017) 'This AI traffic system in Pittsburgh has reduced travel time by 25%'. Smart Cities Dive. See: https://www.smartcitiesdive.com/news/this-ai-traffic-system-in-pittsburgh-has-reduced-travel-time-by-25/447494/
10. Torres, N (2017) 'How Data and Cities Can Shape the Future of Mobility.' Data Smart City Solutions. See: https://datasmart.ash.harvard.edu/news/article/how-data-and-cities-can-shape-the-future-of-mobility-1182
11. Hillenbrand, K (2016) 'New Orleans Develops Data-Intensive Predictive Fire Risk Model'. Government Technology. See: https://www.govtech.com/data/New-Orleans-Develops-Data-Intensive-Predictive-Fire-Risk-Model.html
12. Cuccaro-Alamin, S. et al. (2017) Risk assessment and decision making in child protective services: Predictive risk modeling in context. Children and Youth Services Review. 7 (9), 291–298.
13. See: https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx
14. See: http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf
15. Kahneman, D. & Klein, G. (2009) Conditions for intuitive expertise: A failure to disagree. American Psychologist. 64 (6), 515–526.
16. Shlonsky, A. & Wagner, D. (2005) The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. Children and Youth Services Review. 27 (4), 409–427.
17. Pryce, J. et al. (2018) Using Artificial Intelligence, Machine Learning, and Predictive Analytics in Decision-Making. Available from: https://pdfs.semanticscholar.org/d8b6/b80d41d34c460ddfb351ca95ff2c2965ead4.pdf
18. Munro, E (2019), Predictive Analytics in Child Protection, available at https://www.researchgate.net/publication/332528200_Predictive_analytics_in_child_protection
19. Munro, E (2019), Predictive Analytics in Child Protection, available at https://www.researchgate.net/publication/332528200_Predictive_analytics_in_child_protection
20. Munro, E. (2008) Effective child protection. 2nd ed. Los Angeles: Sage Publications.
21. The author interviewed 13 practitioners who had commenced – or were about to commence – using predictive analytics tools in the field of children's services across four sites; three in the United Kingdom and one in the United States. Interviews were agreed on the basis of anonymity.
22. Dietvorst, B. J. et al. (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General. 144 (1), 114–126.
23. Skitka, L. J. et al. (2000) Automation Bias and Errors: Are Crews Better Than Individuals? The International Journal of Aviation Psychology. 10 (1), 85–97.
24. Kahneman, D. & Klein, G. (2009) Conditions for intuitive expertise: A failure to disagree. American Psychologist. 64 (6), 515–526.
25. Parada, H. et al. (2007) Negotiating 'Professional Agency: Social Work and Decision-Making within the Ontario Child Welfare System. The Journal of Sociology & Social Welfare. 34 (4), 35–56.
26. This included the original thirteen interviews with frontline practitioners working in children's services, together with addition interviews with frontline practitioners working with predictive analytics tools across a range of sectors including policing, housing and children's services. In addition, this research was informed by more informal conversations with tool designers, academics and experts in the field of predictive analytics.
27. Workshop convened by The Greater London Authority and Nesta about the use of AI in the public sector (June 2019).

28. See, e.g. https://www.theguardian.com/society/2018/sep/16/child-abuse-algorithms-from-science-fiction-to-cost-cutting-realityr; and https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit; and https://psmag.com/social-justice/does-big-data-belong-in-courtrooms
29. Eubanks, V. (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York. St Martin's Press, Inc; Angwin, J. et al. (2016) Machine Bias. ProPublica. Available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
30. Crawford, K. 'The Trouble with Bias', (2017) Neural Information Processing Systems Conference Keynote Speech, available at https://www.youtube.com/watch?v=fMym_BKWQzk
31. Babuta, A. and Oswald, M. - Data Analytics and Algorithmic Bias in policing Briefing Papers, 16 September 2019. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/831750/RUSI_Report_-_Algorithms_and_Bias_in_Policing.pdf
32. Munro, E (2019), Predictive Analytics in Child Protection, available at https://www.researchgate.net/publication/332528200_Predictive_analytics_in_child_protection
33. Bird, S et al. (2016) 'Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI', Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT-ML), New York University. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2846909
34. See: https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/. Note that the authors of the Global Inventory crucially highlight that "the absence of internal enforcement or governance mechanism shows that there is a lot of virtue signalling going on and few efforts of enforcement." It is crucial that organisations commit to both adopting and enforcing an ethical approach to designing and implementing AI tools.
35. See, e.g. http://ethicstoolkit.ai. See 'Resources' for further examples.
36. Munro, E (2019), Predictive Analytics in Child Protection, available at https://www.researchgate.net/publication/332528200_Predictive_analytics_in_child_protection
37. Kirkman, E. & Melrose, K. (2014) Clinical Judgement and Decision-Making in Children's Social Work: An analysis of the 'front door' system. (DfE Research Report 323).
38. Klein, G. (1999) Sources of Power: How People Make Decisions. MIT Press, Massachusetts.
39. Grove et al.'s meta-analysis of 136 studies found that algorithmic tools demonstrate higher levels of accuracy and consistency than humans in predicting the likelihood of a specified event occurring. See Grove, W. M. et al. (2000) Clinical versus mechanical prediction: A meta-analysis. Psychological Assessment. 12 (1), 19–30.
40. Fry, H. (2018) Hello world: being human in the age of algorithms. First edition. New York, NY: W.W. Norton & Company.
41. Elish, M. C. (2019) Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. Engaging Science, Technology and Society (pre-print). Available at SSRN: https://ssrn.com/abstract=2757236 or http://dx.doi.org/10.2139/ssrn.2757236
42. Dietvorst, B. J. et al. (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General. 144 (1), 114–126.
43. See: https://www.nesta.org.uk/report/co-production-catalogue/
44. Dietvorst, B. J. et al. (2018) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. Management Science. 64 (3), 1155–1170.
45. Dietvorst, B. J. et al. (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General. 144 (1), 114–126.
46. Elish, M and Mateescu, A. 'AI in Context: The Labor of Integrating New Technologies.' Data and Society. Available at: https://datasociety.net/wp-content/uploads/2019/01/DataandSociety_AIinContext.pdf
47. Pryce, J. et al. (2018) Using Artificial Intelligence, Machine Learning, and Predictive Analytics in Decision-Making. Available from: https://pdfs.semanticscholar.org/d8b6/b80d41d34c460ddfb351ca95ff2c2965ead4.pd; Church, C. E. & Fairchild, A. J. (2017) In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics. Juvenile and Family Court Journal. 68 (1), 67–81.
48. Fry, H. (2018) Hello world: being human in the age of algorithms. First edition. New York, NY: W.W. Norton & Company.

58 Victoria Embankment
London EC4Y 0DS

+44 (0)20 7438 2500

information@nesta.org.uk

@nesta_uk

www.facebook.com/nesta.uk

www.nesta.org.uk