How do public sector values get into public sector machine learning systems, if at all?

Big Data: New challenges for law and ethics
University of Ljubljana, 22 May 2017

Michael Veale
Department of Science, Technology, Engineering & Public Policy
University College London
m.veale@ucl.ac.uk / @mikarv

Current applications of public sector ML

anticipating

crime hotspots
abusive households
food safety breaches
'solvability' of crimes
firm insolvency

detecting

fraudulent tax returns
incorrectly coded crime records
mobile homes for address registers
changes in stats between censuses

STEaPP
Applied in Focus. Global in Reach.

Are there public sector values?

regardless of politics,
the public sector
should

be
- fair and equitable
- accountable
- reliable
- usable
- legal
- effective
- dialogue
- innovative

navigate
- openness–secrecy
- advocacy–neutrality
- competition–cooperation

Jørgensen, T. B. & Bozeman, B. (2007). Public values an inventory. Administration & Society, 39(3), 354–381. doi:10.1177/0095399707300703

Discovering digital discretion

street level bureaucrats

where does discretion **(to be ethical or unethical)** go when you digitise?

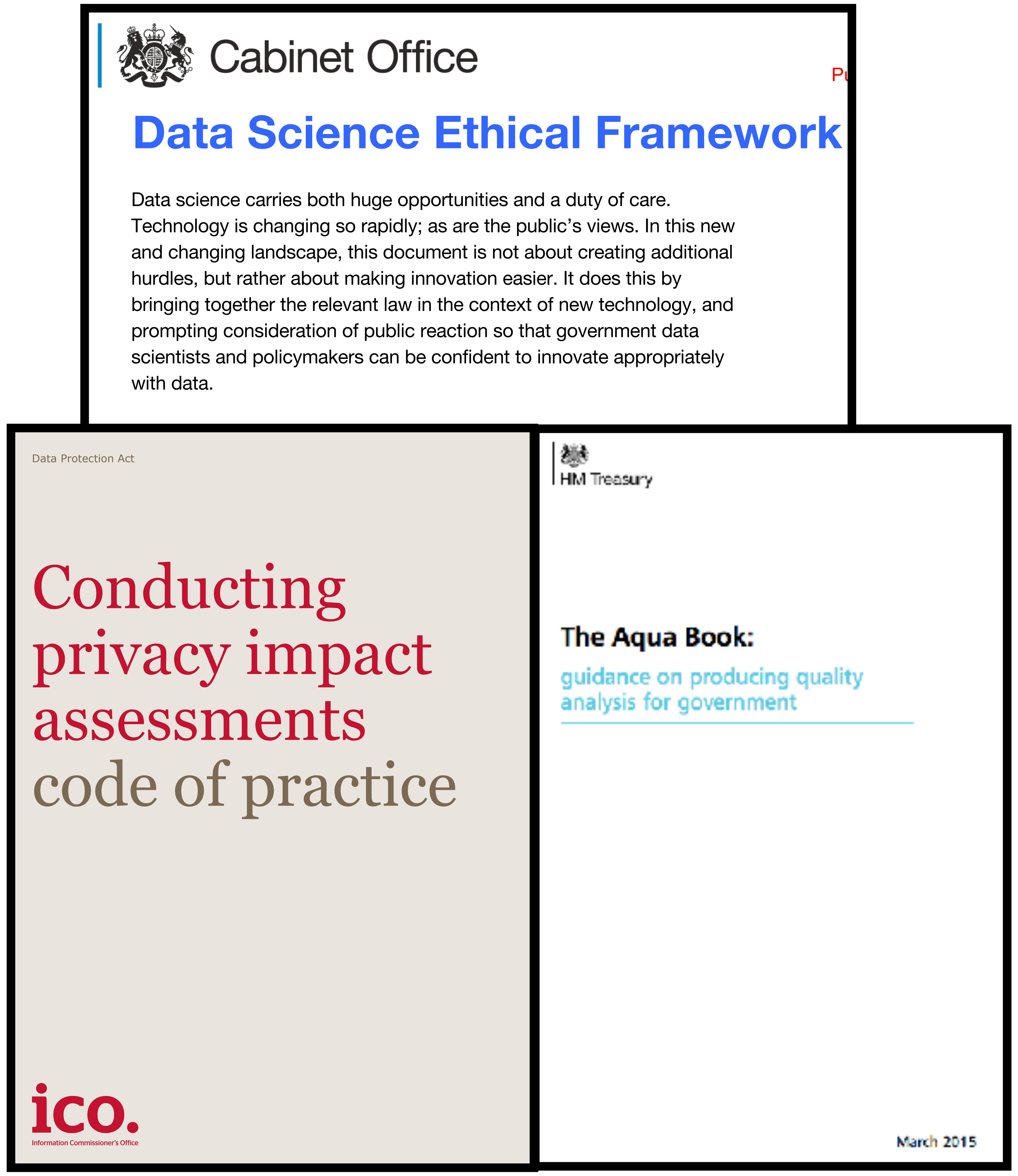
system-level bureaucrats

‘screen-level’ bureaucrats

Flickr: cabinetoffice (L) [CC BY-NC-ND 2.0] // garryknight (TR), seattlemunicipalarchives (BR) [CC BY 2.0]
See Bovens and Zouridis (2002) From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control 10.1111/0033-3352.00168

STE@PP
Applied in Focus. Global in Reach.

Formal or less formal knowledge exchange


Cabinet Office

Data Science Ethical Framework

Data science carries both huge opportunities and a duty of care. Technology is changing so rapidly; as are the public's views. In this new and changing landscape, this document is not about creating additional hurdles, but rather about making innovation easier. It does this by bringing together the relevant law in the context of new technology, and prompting consideration of public reaction so that government data scientists and policymakers can be confident to innovate appropriately with data.

Data Protection Act

Conducting privacy impact assessments code of practice

ico.

Information Commissioner's Office

HM Treasury

The Aqua Book: guidance on producing quality analysis for government

March 2015

formal documents

sectors

at work

or after work

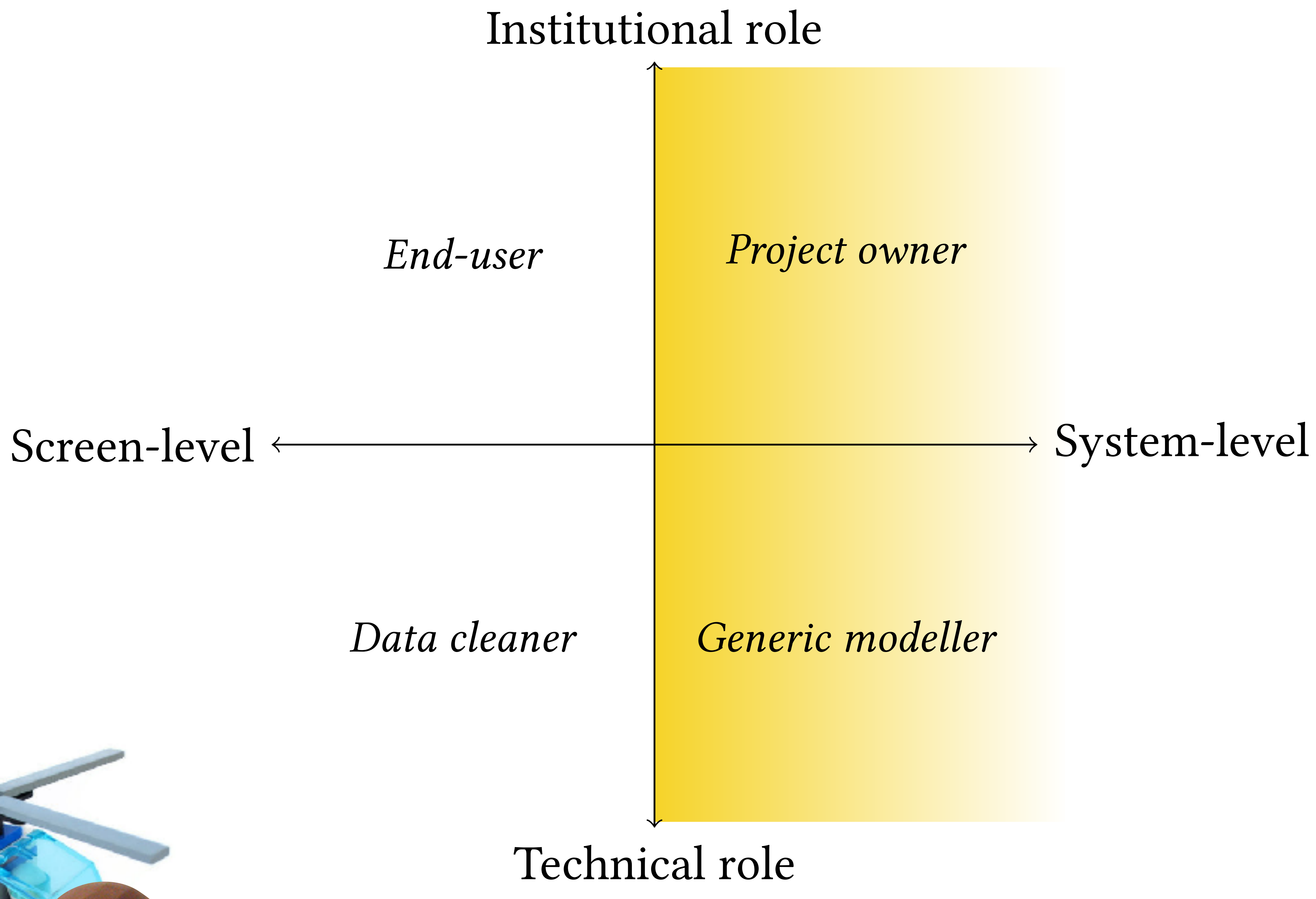
Machine learning as a (social) process

Study background

Shocking lack of empirical data on how machine learning ethics are deployed in practice.

Interviewed:
- 27 individuals (just over 1/5 women)
- working in and with the design and deployment of public sector machine learning systems
- across 5 countries
- about barriers, opportunities, ethical challenges.

Sectors include: police, tax, child protection, fire, council services, interior affairs, prison management, emergency helicopter support



Focus quadrants

five core takeaways

Transparency brings mixed blessings

>> Social factors interact with enhanced interpretability.

Worries of gaming from external actors…

Reliability of a system in the big bad world.

In compliance models we don't give many details. We might say we are interested in sectors or size, and perhaps share the weights with one or two key people. […] We're primarily concerned that if the model weights were public, their usefulness might diminish.

Analytics lead at a tax agency

… but also from internal ones.

System level designers wanting to minimise discretion.

To explain models we talk about the target parameter and the population, rather than explanations of individuals. The target parameter is what we are trying to find — the development of debts, bankruptcy in six months. The target population is what we are looking for: for example, businesses with minor problems. We only give the auditors [these], not an individual risk profile or risk indicators […] in case they investigate according to them.

Analyst at a national tax agency

Sometimes organisations want transparency…

Even if you are dilligent, your word might not be enough.

"[Neural networks] might give us a small uplift [in performance], but … our [internal] customers … look at models that aren't transparent with … suspicion. If they aren't confident they know what a model is doing, they get wary … [and] concerned about accidentally picking up protected characteristics."

A modeller at a tax agency

… and sometimes they don't

The real world can be a problematic place

"We will surely find things that are uncomfortable, unpleasant, even shocking, and we'll have to face up to those and be happy we discovered them. This is is realistically likely to be what [the local government] is scared of: *Aw, shucks*! *What will this algorithm unearth?!*"

An NGO analyst in a predictive child abuse project

five core takeaways

Transparency brings mixed blessings

External information and advice important for ex ante governance

>> Practical and governance constraints limit in-data ethical analyses

External information can shine light on fairness issues

Important for ex ante governance

Whether a child is deaf or disabled is empirically linked to abuse, according to [NGO] research. But of course [local governments] are also aware they don’t want parents singled out as potential abusers simply because they have a disabled child.

Police project lead in predictive child abuse project

Transparency brings mixed blessings

External information and advice important for ex ante governance

Organisational routines for humans-in-the-loop can be rich

>> People organise to augment machine learning systems

Adding humans to to maps I
through meeting and discussion

We also have weekly meeting with all the officers, leadership, management, patrol and so on, and the intelligence officers are the core of this meeting. There, he or she presents what they think is going on in this map, and what should or could be done about it.

Former police lead on a national predictive policing project

Adding the human to maps II

Qualitative routines to augment machine learning systems.

We ask intelligence officers, to look at […] the [predictive] maps. […] They […] file or read […] local reports [… so often] know something about particular burglars or say a high risk building is no longer so because they local government just arranged all the locks to be changed.

Police lead on a national predictive policing project

five core takeaways

Transparency brings mixed blessings

External information and advice important for ex ante governance

Organisational routines for humans-in-the-loop can be rich

Dynamic primary purposes for data might undermine modelling

>> Data wasn't and isn't collected with modelling in mind

‘Secondary uses of data’ are not free lunches

Their sampling logics might undermine model-making

“The intelligence department got a tip-off and looked into cases of human trafficking at car washes. But now when we try to model human trafficking we only see car washes being predicted — they suddenly seem very high risk. We’ve essentially produced models that tell us where car washes are. This kind of loop is hard to explain to those higher up.

A modeller at a police department

Individuals can matter

and coping mechanisms can be contentious

There's one woman who calls in whenever her kid is out after 10pm. She calls back about 30 mins or so later to say that everything is fine […] But then it looks like in the model that kids always go missing at 10pm […] In the end I had to manually remove her from the model to remove the spurious pattern.

A modeller at a police department

five core takeaways

Transparency brings mixed blessings

External information and advice important for ex ante governance

Organisational routines for humans-in-the-loop can be rich

Dynamic primary purposes for data might undermine modelling

Feedback loops matter often in public sectors, private less so

>> Public sector decisions are consequential, and change the future

Feedback loops should not be underemphasised

Public decisions are often more impactful than private ones

The highest probability assessments are on the mark, but actual deployment causes displacement, dispersion and diffusion, and that throws the algorithm into a loop [...] as you deploy resources, displacement and dispersal goes through the roof [...] In the first four weeks of trialling it out, the probability of being correct just tanked

Head of analytics at a city police department

What are you modelling?

Variable names mask difficult, and discrminatory, processes

“Race is very predictive of reoffending … [but] we don’t include race in our predictive models, [because] conviction is [only] a proxy variable for offending … you can get into cycles looking at certain races which might have a higher chance of being convicted … you’re building systems not based on the outcome, but on proxy outcomes.”

A modeller at a justice ministry

five core takeaways

(+ their importance)

Transparency brings mixed blessings

>> Need to consider and study how/if people game and understand models

External information and advice important for ex ante governance

>> Science advice and data science advice

>> Paper soon (Veale and Binns, *under review*).

Organisational routines for humans-in-the-loop can be rich

>> How do they emerge and evolve? Are some better than others?

Dynamic primary purposes for data might undermine modelling

>> How to bridge ML assumptions and organisational reality?

Feedback loops matter often in public sectors, private less so

>> Largely ignored in literature, potentially invalidates methods in practice.

>> Can we elicit and simulate them as part of ML testing?

> fin
> |
Q?
@mikarv
STE[a]PP
Applied in Focus. Global in Reach.
UCL ENGINEERING
Change the world
EPSRC
Engineering and Physical Sciences
Research Council