DATA FOR POLICY REPORT

Unleashing collective intelligence for public decision-making: the *Data for Policy* community

Zeynep Engin[1,2] , Emily Gardner[1,2], Andrew Hyde[3], Stefaan Verhulst[4] and Jon Crowcroft[5,6]

[1]Department for Computer Science, University College London (UCL), Data for Policy CIC, London, UK
[2]Institute of Child Health, University College London (UCL), University College London, London, UK
[3]Cambridge University Press, Cambridge, UK
[4]The GovLab, New York University, New York, NY, USA
[5]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
[6]The Alan Turing Institute, London, UK
Corresponding author: Zeynep Engin; Email: z.engin@dataforpolicy.org

Abstract

Since its establishment in 2014, *Data for Policy* (https://dataforpolicy.org) has emerged as a prominent global community promoting interdisciplinary research and cross-sector collaborations in the realm of data-driven innovation for governance and policymaking. This report presents an overview of the community's evolution from 2014 to 2023 and introduces its six-area framework, which provides a comprehensive mapping of the data for policy research landscape. The framework is based on extensive consultations with key stakeholders involved in the international committees of the annual Data for Policy conference series and the open-access journal *Data & Policy* (https://www.cambridge.org/core/journals/data-and-policy), published by Cambridge University Press. By presenting this inclusive framework, along with the guiding principles and future outlook for the community, this report serves as a vital foundation for continued research and innovation in the field of *data for policy.*

Policy Significance Statement

This report showcases the progressive development of the *Data for Policy* global community of interest, a diverse group of stakeholders dedicated to exploring the possibilities and implications of data science, AI, and related technologies in governance and policymaking. The report presents a comprehensive framework comprising six interdisciplinary and cross-sector areas that have emerged as key focal points in *data for policy* research. The community's primary emphasis is on translational methodologies, tools, and practices that bridge the gap between data and policy decisions. By harnessing the growing complexity of diverse data sources and formats, the community strives to extract valuable insights while acknowledging the inherent uncertainties and imperfections in the process. This report highlights the profound significance of data-driven approaches in informing and shaping effective policy outcomes within a constantly evolving landscape.

1. Introduction

The rapid adoption of digital technologies, accelerated by the COVID-19 pandemic, has brought about profound changes in our daily lives, work, and social interactions. With the advent of algorithmic

decision-making and advanced language models like ChatGPT, the pace of this transformation has created an urgent need for informed debates and rigorous research across sectors and stakeholders. Successfully navigating the complexities of digital transformation demands a commitment to prioritizing innovation, fostering collaboration, and embracing creative problem-solving, all directed toward crafting a future that is more equitable, ethical, and sustainable. Since 2014, *Data for Policy* conferences have played a pivotal role in stimulating relevant research and facilitating these crucial discussions. The launch of the *Data & Policy* journal in 2019 further supports this endeavor (Verhulst et al., 2019). This report describes our perspective on *data for policy* as a distinct field of research, exploring the intersection between increasingly complex governance processes and the growing abundance of data, computational power, and analytical methods.

The contemporary landscape presents new opportunities to enhance human decision-making through the generation of novel information and knowledge, thanks to the proliferation of data and associated technologies (Engin and Treleaven, 2019). Global challenges, ranging from environmental sustainability to bias and displacement, demand fresh perspectives, and innovative approaches (Verhulst, 2021). The ongoing digital revolution prompts a reevaluation of human values and ethical concerns, exemplified by the evolving interactions between humans and generative AI technologies like Large Language Models (LLMs) and Foundation Models (Ge et al., 2023), as well as the convergence of physical and virtual realms and the emergence of potential "metaverse" futures (Robertson and Peters, 2021). In its 9-year existence, the *Data for Policy* community of researchers, practitioners, and private stakeholders has solidified *data for policy* as a distinct and valuable field of study. This report seeks to delve deeper into this field by articulating a comprehensive vision of the landscape and mapping the breadth and depth of its interactions. Grounded in the community that has evolved around the *Data for Policy* conferences and the authors' original research, this vision aims to further develop and enrich the field and its community.

This report offers a comprehensive overview of the Data for Policy global community of interest, tracing its evolution from 2014 to 2023 and outlining its vision for data-driven policy research. By introducing a six-area framework, the report provides a structured organization of the diverse contributions made to the field. Additionally, it highlights the guiding principles that have shaped the community's development and offers insights into its future outlook.

2. The *Data for Policy* community: a brief history of development

While there is an abundance of data and significant advancements in methodologies and technologies, decision-making processes often fail to fully leverage these resources in the public interest or to address stakeholder needs effectively (Head, 2008; Wesselink et al., 2014). This gap may be attributed to various factors, such as disparities in data access, inconsistencies in technological implementation, and divergent priorities and capabilities among different decision-making entities within democratic systems. In this context, emerging technologies and data hold immense potential to mitigate the inherent complexity and uncertainty in policy domains. The scientific and practitioner communities are presented with an opportunity to harness these advancements to address entrenched issues in decision-making, such as systemic inequalities, thereby contributing to a more sustainable future. The inception of the *Data for Policy* community in October 2014 was driven by this imperative. The name, *Data for Policy*, was chosen to reflect the community's inclusive approach, encompassing diverse disciplines, sectors, and the general public as the ultimate beneficiaries. The name also underscores the community's focus on transdisciplinary methodologies, tools, and practices that bridge data and policies, enabling valuable insights to be extracted from the growing complexity of available data sources and formats while accounting for uncertainties and imperfections.

The Data for Policy community (see Figure 1) has played a significant role in capturing, reflecting, and framing the evolution of the *data for policy* research field organically through its conference series complemented with its multi-model dissemination channels. The inaugural conference was held in 2015 at University of Cambridge with the theme of "Policy-making in the Big Data Era: Opportunities and Challenges." It was immediately a premier platform explicitly focusing on innovations in the public
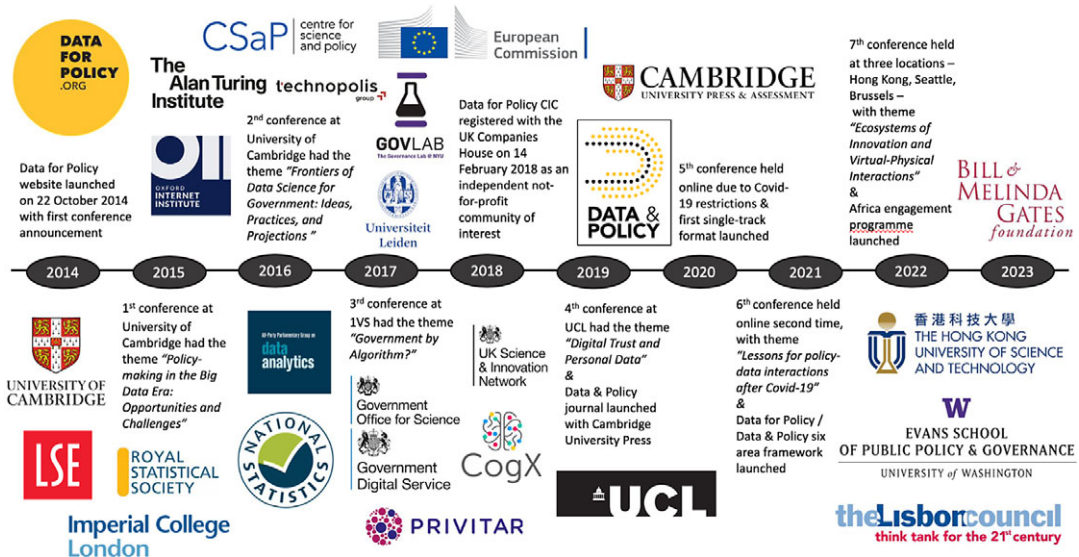
Figure 1. *Brief timeline of the Data for Policy community evolution and the growth of its support network.*

sector under the so called "data revolution" (United Nations Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014). The call attracted wide interest from research, government, industry, third sector and intergovernmental stakeholders attempting to bridge the gap between diverse datasets, computing capacity and methods and the pressing public sector challenges. The conference was also at the forefront of discussion of security and privacy issues, as well as the ethical and legal concerns in this space.

Subsequent conferences brought more structure and depth to the conversation. The 2016 edition embraced a holistic data science framing, exploring frontier ideas, practices, and projections beyond the notion of big data as the sole source of knowledge generation (Anderson, 2008). In 2017, the conference centered around the question of "Government by Algorithm?"—a theme that has since evolved into a distinct area of research. This edition particularly challenged the hierarchical structures of governments as new concepts and tools from technologies including artificial intelligence (AI), Blockchain, and Internet of Things were enabling automation of the public sector, with the unsettling possibility of large-scale job losses (Anderson et al., 2017a,b; Edwards et al., 2017; Henman, 2017; Sel et al., 2017; Engin et al., 2019).

After a break in 2018, the international conference series resumed in 2019 highlighting "Digital Trust and Personal Data" as its central theme, capturing growing worldwide concerns and interest in the topic following the EU's introduction of the General Data Protection Regulation (GDPR) (Engin, 2018). It brought up issues emerging from the automation of decision-making processes with direct impact on human lives (e.g., recruitment, criminal sentencing, loans, and insurance), as well from the mass surveillance, and manipulation of voter behavior. The hype around Bitcoin at the time was a harbinger of the potential environmental cost of such highly advanced computational processes (de Vries, 2018; Mian et al., 2020). There was also a proactive effort to balance these concerns with the potential cost of not using—or the slow uptake of—data science technologies in the public sector. The 2019 conference was also marked by the launch of the open-access *Data & Policy Journal* (Verhulst et al., 2019) as the primary peer-reviewed publication outlet for the community, building on from the open-access repository on the community's profile on the Zenodo platform.

The discussions in 2020 and 2021 were significantly influenced by the global disruptions caused by the COVID-19 pandemic, causing a noticeable shift from a methodology and process-oriented conversation to a more societal challenge-focused discourse. Early responses to the pandemic were presented at the two virtual conferences in the series, and there was an expansion to the potential of this field to support the

delivery of the UN's Sustainable Development Goals (Guerrero and Castañeda, 2020). The community's return to in-person/hybrid conferences was through the introduction of a regional conference model, piloted in 2022 at three locations—Hong Kong, Seattle, and Brussels. The highlight theme "Ecosystems of innovations and virtual-physical interactions" was chosen to reflect the emerging diversity of approaches in the *data for policy* space across the globe, particularly reflecting discussions in Asia, United States, and Europe. The 2022 edition also saw the strategic engagement of the community with the Bill & Melinda Gates Foundation to start a new expansion programme to proactively engage and facilitate participation from regions underrepresented in the global *data for policy* dialogue.

As of June 2023, the *Data for Policy* conference series has received 650 submissions from 1530 authors, representing 73 countries (Figure 2). This substantial participation has established Data for Policy as a platform for cross-sectoral and interdisciplinary work, attracting authors from diverse backgrounds and locations worldwide. The community is characterized by its shared interest in data-driven transformations in the public sector and governance.

3. An interdisciplinary and sector-agnostic framework to map the *data for policy research* landscape

Data for Policy community has grown organically since its establishment in 2014, both in terms of its geographical and topical coverage. The highly abstract set of intuitive questions posed for the first conference in 2015 immediately captured the imagination across sectors and disciplines, evolving and converging over time into the more detailed sector-agnostic six-area framework we now use to organize our community contributions (Figure 3). Each conference call has come as a new iteration of how our community envisions the framing of the *data for policy* research space, based on discussions and feedback from previous conference attendees and committee consultations that took place before the announcement of any new edition. The learning process continues, and the six-area framework we present in this section will evolve to capture the dynamic nature of the field. The framework aims to enable sufficient flexibility to streamline the different disciplinary perspectives, methodologies, and cross-sector thinking that define this sector-agnostic research field. Typically, each of the six areas would complement each other when discussing any socioeconomic or policy problem but remain distinct in their methodological
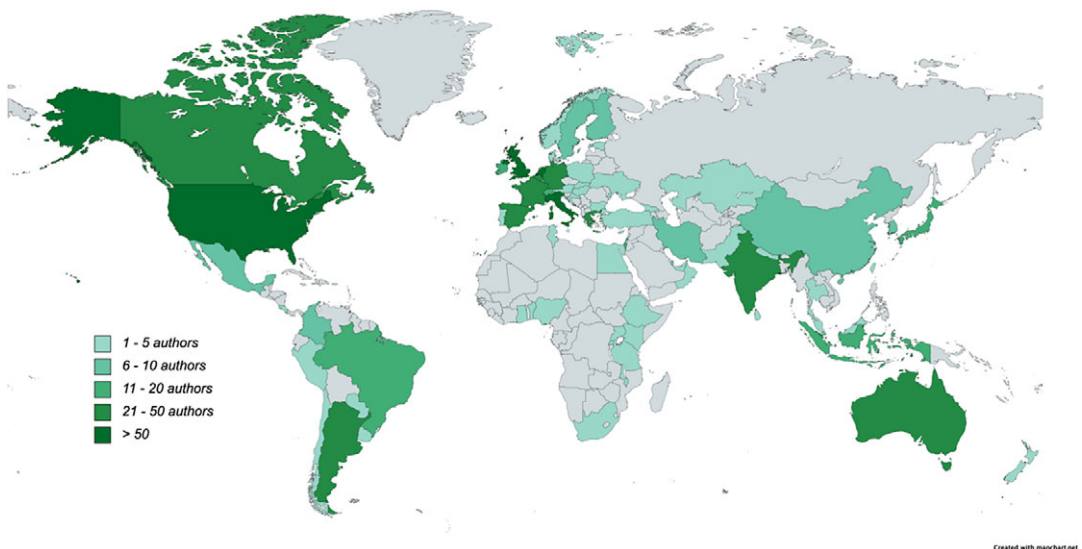

Figure 2. Countries of the 1530 authors, who made submissions to the seven Data for Policy conferences until December 2022.
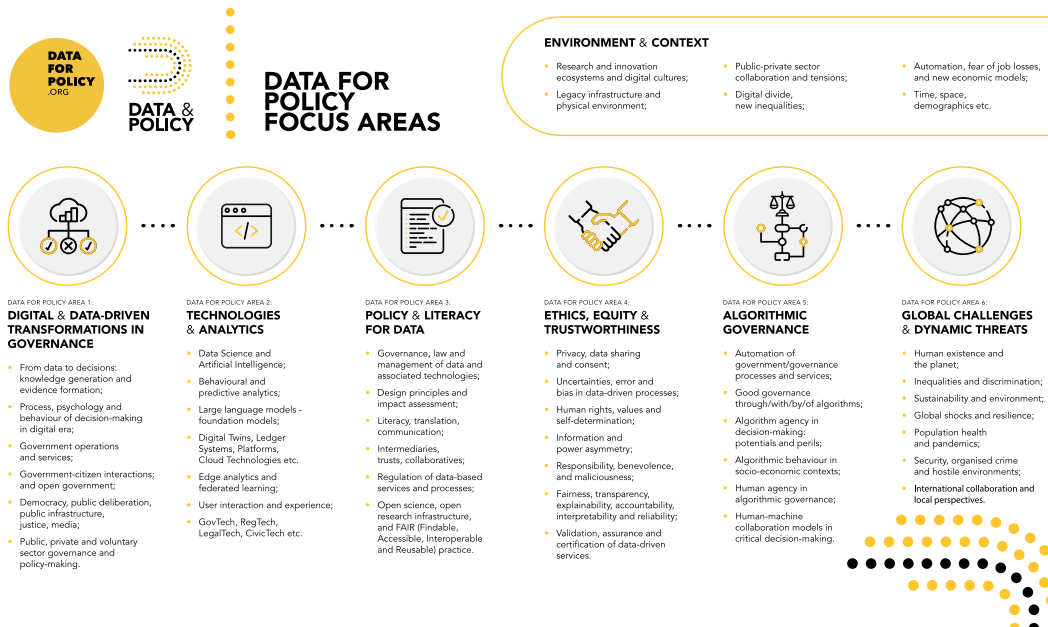
Figure 3. *Data for policy six-area framework to organize community contributions across conference and journal activities.*

approaches and the researcher's starting points. The framework is applied with some flexibility in conference tracks and journal submissions—we recognize that authors may think their research fits into more than one area. It is not simply a tool for categorizing content: we also aim, through the committees of area editors that we have appointed, to use the area structure to stimulate interdisciplinary thinking and new contributions. In addition to the six core areas, we also include further discussion on the environment and context of data-policy interactions.

3.1. Data for policy focus area 1: digital and data-driven transformations in governance

The vision of *Data for Policy* originates from extensive inquiries into the philosophy, ideation, formulation, and implementation of new approaches that lead to paradigm shifts, innovation, and efficiency gains in collective and critical decision-making processes (O'Reilly, 2011; Margetts and Dunleavy, 2013; Pentland, 2013; Scholl and Scholl, 2014; Engin et al., 2019). This typically covers a vast range of topics—from the foundational aspects of democratic systems and institutions to more practical applications such as citizen service delivery, public infrastructure monitoring, and support for civil servants (Engin et al., 2019). Reimagining governance requires transcending existing human and institutional decision-making processes and policy cycles, and establishing connections between top-down, theory-driven, and value-based thinking and bottom-up, data-driven, and emergent approaches to generate new knowledge for critical decision-making. However, the challenge lies in navigating the existing infrastructure, established communities, and policies while adapting to the new forms of engagement in sociotechnical and virtual-physical environments. Three larger areas particularly emerge from this perspective: (1) public participation and collective intelligence; (2) transformation and reimagination of relationships and organizations; and (3) openness. The stakeholder domain covers a broad spectrum of society, including individuals and organizations across the public, private, and voluntary sectors at local, national, and international scales. This raises novel questions regarding government–private sector–individual citizen interactions, data and information access, and their implications for democracy, citizen rights, and public services. An example contribution to our community that fits this area is the work by Boniface et al.

(2022) on the Social Data Foundation model, which combines governance with data trust services to allow citizens, service providers, and researchers to work together to transform systems.

The next three areas focus more specifically on the current methodologies, strategies, and concerns that shape data-driven transformations in governance and policy.

3.2. Data for policy focus area 2: technologies and analytics

Data for Policy research builds on both established and new data sources in all forms, coming from personal, proprietary, administrative and public sources. Similarly, both established theory-driven methods (e.g., qualitative methods and traditional/computational statistics) and the new data-driven methods (e.g., machine learning) are valuable tools for extracting useful and complementary insights from real-time and historical data. The Data for Policy community emphasizes the following technologies and analytics as primarily underpinning the current digital government and governance transformations:

- *Data science and AI* are arguably the two of the most powerful technologies of our time with potential to revolutionize the way governments operate, if developed and deployed responsibly. The ability to analyze large amounts of data beyond human comprehension and making any insights available instantly to support critical decision-making processes offers unprecedented potential in public sector operations, as well as governance decisions more broadly. It can inform evidence-based policymaking, enhance service delivery, optimize resource allocation, and address societal challenges more proactively.
- *Behavioral and predictive analytics* utilize data patterns and trends to analyze user/citizen behavior and forecast needs, which may inform the optimization of service delivery. This approach has the capacity for personalization and efficacy in public services. This type of capacity clearly also holds potential harms at scale. For example, discussions around China's Social Credit System provide rich understanding of potential use cases and harms that can emerge from government utilization of such capabilities.
- *LLMs and foundation models* received massive public recognition particularly with the OpenAI's introduction of ChatGPT in late 2022. The technology essentially enabled a new interaction model between any citizen and the cutting-edge AI models. The potential impact of such technologies in governance and policymaking is enormous—ranging from processing and analyzing large volumes of information and drafting policy documents, to exploring different scenarios and handling citizen inquiries in an efficient and personalized manner. They have also sparked high profile public debate with calls from technology leaders to immediately pause giant AI experiments (Future of Life Institute, 2023).
- *Digital twins, ledger systems, platforms, cloud technologies, and so forth* provide the essential infrastructure for data storage, analysis and sharing (Bennett et al., 2023), safe and secure transactions (Singh and Michels, 2018), collaboration environments (e.g., online platforms and service nodes), and scalable computing resources—hence enabling seamless and efficient operations in the digital realm.
- *Edge analytics and federated learning* are technologies that process data closer to the source, helping improve the performance and preserving privacy and security while harnessing the collective intelligence of distributed systems. For example, smart city or public health infrastructures can be empowered to maximize real-time insight generation from a diversity of data sources to localize and personalize services.
- *User interaction and experience* interfaces facilitate citizen engagement and access to services, voice opinions and participate in public decision-making processes. Mobile devices, location-based technologies, and biometrics provide new forms of human interactions with the latest technologies, government services, and physical environment. Autonomous systems with attributed agency in human social interactions, such as ChatGPT, caretaker or delivery robots, and virtual reality technologies also underpin new types of human–machine ecosystems.

- *GovTech, RegTech, LegalTech, CivicTech, and so forth* encompass a range of technologies tailored for specific government, regulatory, legal, and civic contexts. These technologies support digital government and governance transformations by streamlining administrative processes, ensuring regulatory compliance, facilitating legal services, and promoting citizen engagement and participation in governance.

These technologies collectively provide previously unimaginable capacity to transform governance operations and policymaking, as well as new kinds of ethical and legal challenges (e.g., the risks of a surveillance society). Furthermore, expertise on these technologies and their applications in government/policy domains are scattered across different disciplines and sectors (Hon et al., 2015). The *Data for Policy* community aims to bring together this diversity to lay foundations for a transdisciplinary outlook in this space, unifying different methods of knowledge creation to address the most pressing issues of our current era, through the so-called "digital revolution." An example contribution to our community that fits this area is by Simonofski et al. (2022) on the use of advanced analytics (data mining, big data analysis, or AI) to detect fraud and the complexity of balancing these with legal requirements.

3.3. Data for policy focus area 3: policy and literacy for data

Beyond the opportunities in government, governance and policy, and the complications within technology and analytics environments, the third main area of interest for *data for policy research* deals with the policy, governance, and management issues involved in development and implementation of data-driven solutions. The governance models and frameworks for data and associated technologies are developed across the globe with variances according to local context and value judgments, as well as public literacy and acceptance. Europe's GDPR has set a global gold standard in data practices (Voigt and von dem Bussche, 2017) despite its limitations (Gal and Aviv, 2020). Laws and regulations related to technologies such as AI and Blockchain are increasingly established through both reinterpretation of existing laws and introduction of new laws and regulations (e.g., EU AI Act—2021). Data intermediaries in the form of trusts or collaboratives, or concepts such as data banking or "trust-less" peer-to-peer data infrastructures (e.g., distributed ledger) are important areas in development as potential solutions to a range of problems associated with sensitive/personal data sharing and public–private collaborations (Boniface et al., 2022; Wong et al., 2022). From a management perspective, data supply chains, ownership and provenance, sharing and linkage, and curation and expiration are of primary concern. Closely overlapping with this, effective data and technology governance requires high-quality meta-data to be made openly available with common standards to ensure interoperability at all scales. An example that fits this area is the work by Wong et al. (2022) on a data protection-focused data commons to encourage policymakers to reconsider balances of power between data subjects, data controllers, and data protection stakeholders.

3.4. Data for policy focus area 4: ethics, equity, and trustworthiness

With the rapid adoption of LLMs in particular, several high-profile experts and public figures cautioned in a May 2023 statement that mitigating AI's extinction-level risk should be a global priority, on par with societal-scale risks such as pandemics and nuclear war (Centre for AI Safety, 2023). Conversely, many others, such as Bender et al. (2021), highlight more immediate concerns, including the concentration of power, the perpetuation of systemic oppression, and damage to both the information ecosystem and the natural environment. Opacity and complexity of data-driven insights, together with growing artificial *agency* in social and policy contexts, are arguably the main distinguishing characteristics of the ongoing data science and AI revolution. In comparison to other general purpose technologies and disruptive innovations which drastically altered preexisting economic and social structures (such as the steam engine, electricity, computers, or the internet), disruption caused by data-driven technologies, particularly AI, goes beyond process change to impact human intelligence, knowledge generation methods, and cognitive capabilities. It is therefore inevitable that the public and policymakers are increasingly

concerned about potential unintended consequences, malicious actors in the system, and possible loss of human control in this space. Increasingly, the concern is that the new data-driven technologies essentially replicate and exacerbate both known and unknown problems of human decision-making processes, such as systemic inequalities, amplification of power imbalance, the widespread exploitation of human labor, and the entrenching of power imbalance (Pasquale, 2015; O'Neill, 2016; Russell, 2019; Zuboff, 2019). It is also becoming a computational problem to address such topics, which traditionally fall into the remit of social science and humanities. *Algorithmic fairness* has emerged as a particularly interesting subfield of machine learning, striving to computationally formulate, analyze and embed "fair" behavior in data-driven learning processes (Barocas et al., 2019; Wachter et al., 2021). Other emerging areas include algorithm transparency, explainability and interpretability (Weller, 2017), trustworthiness (Spiegelhalter, 2020), interoperability (Brown and Korff, 2022), accountability (Binns, 2018), and contestability (Lyons et al., 2021). An example contribution to our community in this area is the article by Biddle et al. (2022) that aim to help policymakers and other stakeholders understand how data trust and attitudes on data privacy have changed throughout the Covid-19 pandemic, and which demographic groups have lower levels of trust.

Beyond the above sector-agnostic classification of the Data for Policy research landscape, we have established the following two areas of primary interest.

3.5. Data for policy focus area 5: algorithmic governance

Governance operations are becoming increasingly dependent on algorithm assistance, from democratic processes, everyday operations and public service delivery to critical life decisions at important junctures, such as credit applications, hiring processes and criminal sentencing. Algorithmic governance (AG) is an emerging field of research which systematically explores the design of social and economic processes through the use of algorithmic systems (Aneesh, 2009; O'Reilly, 2013; Danaher et al., 2017; Athey, 2017; Varian, 2018; Yeung, 2018; Katzenbach and Ulbricht, 2019; Gritsenko and Wood, 2020). While *techno-utopian* perspectives advocate for algorithmic capacity to overcome the systemic imperfections of politics (e.g., bias) and faulty forms of knowledge (O'Reilly, 2013; Bhushan, 2014; Kitchin, 2014), *techno-dystopian* perspectives stress the algorithmic potential to amplify existing forms of power and control by unaccountable actors (Pasquale, 2015; O'Neill, 2016; Zuboff, 2019). AG is also often confused with a growing body of research around the control of algorithmic behavior, which is better termed as governance of algorithms (GA) and better suited for exploration as part of the *"Policy & Literacy for Data"* focus area in our framework. Although AG and GA are closely related, we distinguish AG by its emphasis on the algorithmic role and agency in critical decision-making tasks, along with existing networks, hierarchies, and bureaucracies of governance. The current global debate on AG (and also GA) is largely dominated by legal and policy perspectives with limited understanding of the technology development and the landscape of engineering solutions offered in this space. Also, although there are substantial technical research communities emerging in areas very relevant to AG—such as algorithmic fairness, transparency, interpretability, accountability and adaptability—these efforts do not directly address the core governance problems and are not formulated as such. They are also predominantly reactive in nature, addressing emerging "issues" from the growing use of algorithms in society, and are limited to solutions offered in specific problem contexts and theoretical configurations (Kroll, 2015; Gebru et al., 2017; Tsvetkova et al., 2017; Narayanan, 2018; Chollet, 2019; Russell, 2019; Weller, 2019; Friedler et al., 2021). There is, therefore, an urgent need for the unity of intellectual frameworks beyond disciplinary perspectives and research practices to offer holistic thought leadership in this space.

3.6. Data for policy focus area 6: global challenges and dynamic threats

The Covid-19 pandemic was yet another reminder of a crucial message: the most pressing challenges of our time transcend national boundaries, necessitating global collaboration that goes beyond the interests of any single nation. In addressing these challenges, data and associated technologies have played a

central role. They have provided real-time evidence to policymakers, localized advice to individuals, facilitated the rapid development of vaccines (Latif et al., 2020; Oliver et al., 2020; Benjamins et al., 2022), and underpinned economic and social responses against the pandemic's disruption. Looking beyond the pandemic, data-driven innovation, especially with the aid of AI and when deployed responsibly, holds significant potential in tackling other major challenges. These encompass ensuring the sustainability of human existence on the planet, addressing inequalities and discrimination, promoting environmental sustainability, building resilience against global shocks, safeguarding population health from pandemics, countering security threats, and navigating hostile environments. The Data & Policy community recognizes the gravity of these challenges and actively contributes to addressing them. For instance, the Data & Policy community, through initiatives like the Big Data for Migration Alliance's Data & Policy special collection, seeks to facilitate responsible data innovation and collaboration. This collection aims to inform humanitarian assistance and policy responses by considering the implications of data-driven approaches for the complex dynamics of migration. Moreover, the Data for Policy community actively engages in international collaboration, acknowledging the importance of cross-border partnerships and knowledge exchange in tackling global challenges. By harnessing the power of data and advancing data-driven methodologies, the Data for Policy community is committed to driving progress in these critical areas. It promotes the responsible use of data to inform policy decisions, develop innovative solutions, and foster collaboration among diverse stakeholders. Through its initiatives, the community strives to contribute to building a more sustainable, inclusive, and resilient future, where global challenges are met with collective action and data-driven insights.

Beyond the six focus areas described, *data for policy research* is significantly shaped by broader, fundamental factors grouped under the *environment and context of data for policy research* heading. This encompasses a variety of challenges, including diverse research and innovation cultures, limitations of legacy infrastructure, and physical environmental constraints. It also considers the dynamics between public and private sectors, evolving collaboration models, the emergence of new inequalities alongside a widening digital divide, societal concerns such as job loss fears tied to economic shifts (like those seen with the sharing economy), and demographic constraints. These factors underscore the need for ongoing, in-depth discussions that extend beyond the scope of this paper.

4. Outlook for the future: driving global debate, inspiring local solutions

Building upon the foundation of our six-area framework, the *Data for Policy* community is dedicated to generating practical insights that enhance the process of critical decision-making, adopting a forward-thinking perspective that embraces data-driven technologies and methodologies. Our goal is to shape a more informed, equitable, and sustainable future for all by harnessing the transformative potential of data, AI, and associated technologies. While we focus on the ultimate goals, we also place significant emphasis on the "how" question, which encompasses the values, principles, and processes guiding our journey toward achieving these goals. Our approach to shaping the data for policy discourse revolves around three main axes (Figure 4):

4.1. Advanced computation versus policy and practice

This axis delves into the interplay between advanced computation in data science and practical policy implementation. Advanced computation, grounded in computer science, statistics, and mathematics, contrasts with policy practice, which tackles socioeconomic and political issues through social sciences and humanities. The Data for Policy community champions a transdisciplinary approach, integrating diverse insights into pragmatic solutions for societal challenges. Our initiatives include:

- Strategic landscape reviews to monitor the evolution of the data for policy research and practice to support all stakeholders in this space;
- Promoting "translational articles" that bridge research and practice;
- Incorporating multiple perspectives in our committee and peer-review processes;

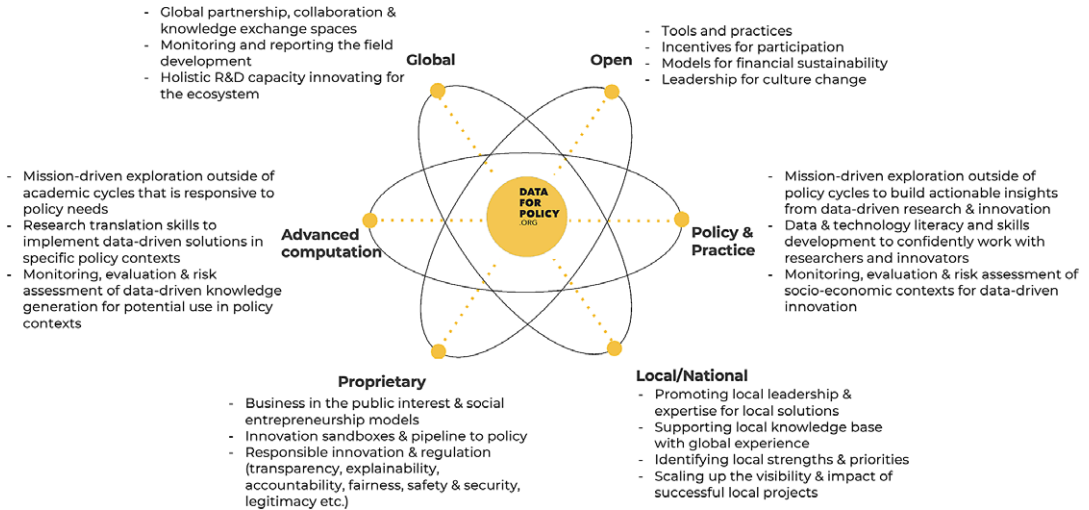## Realising Data for Policy Community Vision and Values

- Global partnership, collaboration & knowledge exchange spaces
- Monitoring and reporting the field development
- Holistic R&D capacity innovating for the ecosystem

Global

Open
- Tools and practices
- Incentives for participation
- Models for financial sustainability
- Leadership for culture change

- Mission-driven exploration outside of academic cycles that is responsive to policy needs
- Research translation skills to implement data-driven solutions in specific policy contexts
- Monitoring, evaluation & risk assessment of data-driven knowledge generation for potential use in policy contexts

Advanced computation

DATA FOR POLICY .ORG

Policy & Practice
- Mission-driven exploration outside of policy cycles to build actionable insights from data-driven research & innovation
- Data & technology literacy and skills development to confidently work with researchers and innovators
- Monitoring, evaluation & risk assessment of socio-economic contexts for data-driven innovation

Proprietary
- Business in the public interest & social entrepreneurship models
- Innovation sandboxes & pipeline to policy
- Responsible innovation & regulation (transparency, explainability, accountability, fairness, safety & security, legitimacy etc.)

Local/National
- Promoting local leadership & expertise for local solutions
- Supporting local knowledge base with global experience
- Identifying local strengths & priorities
- Scaling up the visibility & impact of successful local projects

Figure 4. *Realizing our vision and values.*

- Employing a multimodal dissemination strategy, including an open-access repository, the Data & Policy journal, YouTube, and social media.

Through interdisciplinary and cross-sector collaboration, we aim to evaluate and leverage data-driven methods in relevant socioeconomic contexts. Our goal is to foster dialogue, knowledge exchange, and impactful policymaking.

4.2. Local versus global

This axis in data for policy initiatives encapsulates the interplay between local/national and global perspectives. The global aspect focuses on wide-reaching challenges and solutions spanning multiple disciplines, sectors, and geographies, fostering international collaboration and resource sharing. In contrast, the local perspective prioritizes bespoke solutions attuned to specific regional needs, respecting unique social, cultural, and political contexts. Our key initiatives include:

- Regional conferences—our 2022 conference edition expanded to regional events in Hong Kong, Seattle, and Brussels, reflecting distinct digital cultures of China, the United States, and Europe. Plans to extend to more locations, especially in the Global South, are underway.
- Our Sub-Saharan Africa Engagement Programme, in collaboration with the Bill & Melinda Gates Foundation, promotes a two-way knowledge exchange. It supports the SSA data and AI ecosystem to address local challenges while integrating regional insights into the global discourse, particularly in agriculture, food security, and climate change adaptation.

These initiatives aim to balance global collaboration with local action, ensuring contextually relevant and sustainable *data for policy* practices and outcomes.

4.3. Open versus proprietary

This axis explores the continuum from open, collaborative innovation to traditional, proprietary methods. Open innovation values transparency, accessibility, and equity, encouraging knowledge sharing and

co-creation. Conversely, proprietary innovation, with its clear financial models and control, often conflicts with open practices, raising concerns over safety, security, and ownership. Our key initiatives along this axis include:

- Ensuring all our publications and dissemination channels are available on an open-access basis, and support unfunded authors through a proactive campaign.
- Facilitating connected discussions between industry stakeholders and public interest research groups.
- Encouraging and supporting new GovTech ventures to embrace open innovation practices.

These steps reflect our commitment to blending open and proprietary innovation for ethical, inclusive data-driven policies and technology innovation.

Through our commitment to these axes and the six areas of *data for policy*, we aim to drive global debate and inspire local solutions. By facilitating constructive dialogue, promoting knowledge exchange, and enabling effective and impactful policy decisions and practices, we envision a future where data-driven technologies and methodologies contribute to a more informed, equitable, and sustainable world for all.

5. Summary remarks

This inaugural paper in the *Data & Policy Reports* category serves as a reference point for the formation, development, and key milestones of the Data for Policy global community since its inception in 2014. It provides an overview of the data for policy research landscape from a sector-agnostic perspective, with a strong emphasis on achieving interdisciplinarity and transdisciplinarity in research and cross-sector thought leadership. As we move forward, we plan to develop a series of landscape reviews that offer a holistic outlook on each of the six areas presented, as well as other key cross-cutting elements that shape this field, including genuine inclusion of all geographic regions, demographic segments, and communities in the global debate.

As of June 2023, the Data for Policy community has taken the following practical steps to realize its vision leveraging available resources:

- Establishing an inclusive global collaboration and knowledge exchange infrastructure centered around the premier Data for Policy conference series and the open-access Data & Policy journal published by Cambridge University Press.
- Hosting seven international conferences since 2015, with the 2022 edition expanding to three continents. Each conference has brought together over 200 delegates, featured more than 100 presentations, and spanned over 2 days.
- Involving 1530 authors from 73 countries, who have contributed valuable knowledge and insights to the community.
- Experimenting with different dissemination methods to adapt to evolving knowledge presentation formats. This includes publishing articles in the Data & Policy journal, utilizing the Zenodo open access repository for pre-prints, and producing video content on YouTube.
- Initiating a proactive "inclusion program" with a regional focus on underrepresented areas in the global data for policy discourse, starting with Sub-Saharan Africa.

The community's efforts demonstrate its commitment to fostering collaboration, sharing knowledge, and addressing the needs of diverse regions and stakeholders in the field of *data for policy*. Moving forward, the community aims to formalize its collaboration infrastructure by establishing a transparent governance structure that prioritizes inclusivity and sustainability. It seeks to maintain its strategic independence to enable a healthy global debate while ensuring that the voices and perspectives of all stakeholders are

valued and represented. In doing so, the community strives to create an environment conducive to the advancement of data-driven policy discussions and initiatives.

References

Anderson C (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired* (accessed 23 June 2008).

Anderson CL, **Biscaye PE**, **Hayes AL**, **Klawitter MM and Travis WR** (2017a) Public policy and the promise of digital credit for financial inclusion. In *Data for Policy 2017 Conference: Government by Algorithm?* https://doi.org/10.5281/zenodo.884186.

Anderson CL, **Biscaye PE and Reynolds TW** (2017b) National ID programs: A multi-country review and analysis of policy and practical challenges. In *Data for Policy 2017 Conference: Government by Algorithm?* https://doi.org/10.5281/zenodo.884177.

Aneesh A (2009) Global labor: Algocratic modes of organization. *Sociological Theory 27*(4), 347–370. https://doi.org/10.1111/j.1467-9558.2009.01352.x.

Athey S (2017) Beyond prediction: Using big data for policy problems. *Science 355*(6324), 483–485. https://doi.org/10.1126/science.aal4321.

Barocas S, **Hardt M and Narayanan A** (2019) Fairness and Machine Learning. Available at https://fairmlbook.org/.

Bender E, **Gebru T**, **McMillan-Major A and Shmitchell S** (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Canada: ACM Digital Library, pp. 610–623. https://doi.org/10.1145/3442188.3445922.

Benjamins R, **Vos J and Verhulst S** (2022) Mobile big data in the fight against COVID-19. *Data & Policy 4*, E9. https://doi.org/10.1017/dap.2021.39.

Bennett H, **Birkin M**, **Ding J**, **Duncan A and Engin Z** (2023) "Towards ecosystems of connected digital twins to address global challenges." London: The Alan Turing Institute, May 18, 2023. Available at https://www.turing.ac.uk/news/publications/towards-ecosystems-connected-digital-twins-address-global-challenges; https://doi.org/10.5281/zenodo.7840266.

Bhushan A (2014) Fast data, slow policy: Making the most of disruptive innovation. *SAIS Review of International Affairs. 34*(1), 93–107.

Biddle N, **Edwards B**, **Gray M**, **Hiscox M**, **McEachern S and Sollis K** (2022) Data trust and data privacy in the COVID-19 period. *Data & Policy 4*, E1. https://doi.org/10.1017/dap.2021.36.

Binns R (2018) Algorithmic accountability and public reason. *Philosophy and Technology 31*, 543–556. https://doi.org/10.1007/s13347-017-0263-5.

Boniface M, **Carmichael L**, **Hall W**, **Pickering B**, **Stalla-Bourdillon S and Taylor S** (2022) The social data foundation model: Facilitating health and social care transformation through datatrust services. *Data & Policy 4*, E6. https://doi.org/10.1017/dap.2022.1.

Brown I and Korff D (2022) Interoperability, Data Protection and Digital Competition. Available at https://www.ianbrown.tech/2022/03/22/generative-interoperability-building-online-public-and-civic-spaces/.

Centre for AI Safety (2023) *Statement on AI Risk.* May 30, 2023. Available at https://www.safe.ai/statement-on-ai-risk (accessed 7 June 2023).

Chollet F (2019) On the Measure of Intelligence. Preprint, https://arxiv.org/abs/1911.01547.

Danaher J, **Hogan MJ**, **Noone C**, **Kennedy R**, **Behan A**, **De Paor A**, **Felzmann H**, **Haklay M**, **Khoo S-M**, **Morison J**, **Murphy MH**, **O'Brolchain N**, **Schafer B and Shankar K** (2017) Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society 2*(4). https://doi.org/10.1177/2053951717726554.

de Vries A (2018) Bitcoin's growing energy problem. *Joule 5*(2), 801–805. https://doi.org/10.1016/j.joule.2018.04.016.

Edwards L, **Mullagh L**, **Towe R**, **Nundloll, V**, **Dean C**, **Dean G**, **Simm W**, **Samreen F**, **Bassett R and Blair G** (2017) Data-Driven Decisions for Flood Risk Management. In *Data for Policy 2017 Conference: Government by Algorithm?*. London, UK: Data for Policy on the open-access Zenodo repository. https://doi.org/10.5281/zenodo.884180.

Engin Z (2018) Digital Ethics: Data, Algorithms, Interactions. Zenodo: Data for Policy. https://doi.org/10.5281/zenodo.1481153.

Engin Z and Treleaven P (2019) Algorithmic government: Automating public services and supporting civil servants in using data science technologies. *The Computer Journal 62*(3), 448–460. https://doi.org/10.1093/comjnl/bxy082.

Engin Z, **van Dijk J**, **Lan T**, **Longley PA**, **Treleaven P**, **Batty M and Penn A** (2019) Data-driven urban management: Mapping the landscape. *Journal of Urban Management 9*(2), 140–150. https://doi.org/10.1016/j.jum.2019.12.001.

Friedler SA, **Scheidegger C and Venkatasubramanian S** (2021) The (im)possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM 64*(4), 136–143. https://doi.org/10.1145/3433949.

Future of Life Institute (2023) *Pause Giant AI Experiments: An Open Letter.* March 22, 2023. Available at https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Gal MS and Aviv O (2020) The competitive effects of the GDPR. *Journal of Competition Law & Economics 16*(3), 349–391. https://doi.org/10.1093/joclec/nhaa012.

Ge Y, **Hua W**, **Ji J**, **Tan J**, **Xu S and Zhang Y** (2023) "OpenAGI: When LLM Meets Domain Experts." April 12, 2023. Preprint, arXiv:2304.04370.

Gebru T, **Krause J**, **Wang Y and Fei-Fei L** (2017) Using deep learning and Google street view to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences - PNAS 114*(50), 13108–13113. https://doi.org/10.1073/pnas.1700035114.

Gritsenko D and Wood M (2020) Algorithmic governance: A modes of governance approach. *Regulation & Governance 16*(1), 45–62. https://doi.org/10.1111/rego.12367.

Guerrero OA and Castañeda G (2020) Policy priority inference: A computational framework to analyze the allocation of resources for the sustainable development goals. *Data & Policy 2*, E17. https://doi.org/10.1017/dap.2020.18.

Head BW (2008) Three lenses of evidence-based policy. *The Australian Journal of Public Administration 67*(1), 1–11. https://doi.org/10.1111/j.1467-8500.2007.00564.x.

Henman P (2017). The Computer Says 'DEBT': Towards a Critical Sociology of Algorithms and Algorithmic Governance. In *Data for Policy 2017 Conference: Government by Algorithm?*. London, UK: Data for Policy on the open-access Zenodo repository. https://doi.org/10.5281/zenodo.884117.

Hon WK, **Millard C**, **Singh J**, **Walden I and Crowcroft J** (2015) Policy, legal and regulatory implications of a Europe-only cloud. *International Journal of Law and Information Technology 24*(3), 251–278. https://doi.org/10.1093/ijlit/eaw006.

Katzenbach C and Ulbricht L (2019) Algorithmic governance. *Internet Policy Review 8*(4), 1–18.

Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. London: Sage Publications Ltd.

Kroll J (2015) *Accountable Algorithms*. Princeton, NJ: Princeton University.

Latif S, **Usman M**, **Manzoor S**, **Iqbal W**, **Qadir J**, **Tyson G**, **Castro I**, **Razi A**, **Boulos MNK**, **Weller A and Crowcroft J** (2020) Leveraging data science to combat covid-19: A comprehensive review. *IEEE Transactions on Artificial Intelligence 1*(1), 85–103. https://doi.org/10.1109/TAI.2020.3020521.

Lyons H, **Velloso E and Miller T** (2021) Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. In *Proceedings of the ACM on Human-Computer Interaction*. ACM Digital Library, Volume *5* Issue CSCW1 Article No: 106, pp. 1–25. https://doi.org/10.1145/3449180.

Margetts H and Dunleavy P (2013) The second wave of digital-era governance: A quasi-paradigm for government on the web. *Philosophical Transactions of the Royal Society A 2013*. http://doi.org/10.1098/rsta.2012.0382.

Mian I, **Twisleton D and Timm D** (2020) What is the resource footprint of a computer science department? Place, people, and pedagogy. *Data & Policy 2*, E14. https://doi.org/10.1017/dap.2020.12.

Narayanan A (2018) Translation Tutorial: 21 Fairness Definitions and their Politics. In *ACM Conference on Fairness, Accountability and Transparency.* https://www.youtube.com/watch?v=jIXIuYdnyyk.

Oliver N, **Lepri B**, **Sterly H**, **Lambiotte R**, **Deletaille S**, **De Nadai M**, **Letouzé E**, **Ali Salah A**, **Benjamins R**, **Cattuto C**, **Colizza V**, **de Cordes N**, **Fraiberger SP**, **Koebe T**, **Lehmann S**, **Murillo J**, **Pentland A**, **Pham PN**, **Pivetta F**, **Saramäki S**, **Scarpino SV**, **Tizzoni M**, **Verhulst S and Vinck P** (2020) Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances 6*(23), eabc0764.

O'Neill C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Vol. *2016*. New York: Crown Books.

O'Reilly T (2011) Government as a platform. *Innovations: Technology, Governance, Globalization 6*(1), 13–40. https://doi.org/10.1162/INOV_a_00056.

O'Reilly T (2013) Open data and algorithmic regulation. In Goldstein B with Dyson L (eds.) *Beyond Transperancy: Open Data and the Future of Civic Innovation*. San Francisco: Code for America Press.

Pasquale F (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Pentland AS (2013) The data-driven society. *Scientific American 309*(4), 78–83.

Robertson A and Peters J (2021) What is the metaverse, and do I have to care? *The Verge*. Available at https://www.theverge.com/22701104/metaverse-explained-fortnite-roblox-facebook-horizon. 4 October 2021–18 March 2022.

Russell S (2019) *Human Compatible: Artificial Intelligence and the Problem of Control*. London: Penguin Publishing Group.

Scholl JJ and Scholl MC (2014) Smart Governance: A Roadmap for Research and Practice. In *iConference 2014 Proceedings*, pp. 163–176. https://doi.org/10.9776/14060.

Sel M, **Diedrich H**, **Demeester S and Stieber H** (2017) How Smart Contracts Can Implement "Report Once". In *Data for Policy 2017 Conference: Government by Algorithm?*. London: Data for Policy on Zenodo open-access repository. https://doi.org/10.5281/zenodo.884497.

Simonofski A, **Tombal T**, **De Terwangne C**, **Willem P**, **Frenay B and Janssen M** (2022) Balancing fraud analytics with legal requirements: Governance practices and trade-offs in public administrations. *Data & Policy 4*, E14. https://doi.org/10.1017/dap.2022.6.

Singh J and Michels JD (2018) Blockchain as a Service (BaaS): Providers and Trust. In *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. London: IEEE, pp. 67–74. https://ieeexplore.ieee.org/document/8406562.

Spiegelhalter D (2020) Should we trust algorithms? *Harvard Data Science Review 2*(1). https://doi.org/10.1162/99608f92.cb91a35a.

Tsvetkova M, **Yasseri T**, **Meyer ET**, **Pickering JB**, **Engen V**, **Walland P**, **Lüders M**, **Følstad A and Bravos G** (2017) Understanding human-machine networks: A cross-disciplinary survey. *ACM Computing Surveys 50*(1), 12. https://doi.org/10.1145/3039868.

United Nations Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014) *A World That Counts*. Available at https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

Varian H (2018) Artificial Intelligence, Economics, and Industrial Organization [Report]. National Bureau of Economic Research. https://doi.org/10.3386/w24839.

Verhulst S (2021) Reusing data responsibly to achieve development goals [report]. *OECD iLibrary 2021*. https://doi.org/10.1787/ce08832f-en.

Verhulst SG, **Engin Z and Crowcroft J** (2019) Data & Policy: A new venue to study and explore policy–data interaction. *Data & Policy 1*, E1. https://doi.org/10.1017/dap.2019.2.

Voigt P and von dem Bussche A (2017) *The EU General Data Protection Regulation (GDPR)*. Cham: SpringerNature.

Wachter S, **Mittelstadt B and Russell C** (2021) Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review 41*. https://www.sciencedirect.com/science/article/abs/pii/S0267364921000406.

Weller A (2017) *Challenges for Transparency*. CoRR. https://openreview.net/forum?id=SJR9L5MQ-.

Weller A (2019) Transparency: Motivations and Challenges. In Samek W, Montavon G, Vedaldi A, Hansen LK and Müller KR (eds.) *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Vol. *11700*. Cham: SpringerNature.

Wesselink A, **Hal C and Pearce W** (2014) Evidence and policy: Discourses, meanings and practices. *Policy Sciences 47*, 339–344. https://doi.org/10.1007/s11077-014-9209-2.

Wong J, **Henderson T and Ball K** (2022) Data protection for the common good: Developing a framework for a data protection-focused data commons. *Data & Policy 4*, E3. https://doi.org/10.1017/dap.2021.40.

Yeung K (2018) Algorithmic regulation: A critical interrogation. *Regulation & Governance 12*(4), 505–523. https://doi.org/10.1111/rego.12158.

Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.