

Book Review

Jérôme Duberry (2022) *Artificial Intelligence and Democracy: Risks and Promises of AI-mediated citizen-government relations*, Edward Elgar: Cheltenham

In *Artificial Intelligence and Democracy*, Jérôme Duberry sets out to explore the use of artificial intelligence (AI) in liberal democracies. While AI is promising, especially in terms of efficiency, it is not neutral. At a time when governments are increasingly using AI in their interactions with citizens, this book offers welcome, and much needed, reflections on the impact of such technologies on democratic values. Duberry presents an extensive overview of the use of AI in citizen-government relationships, particularly focussing on participation, power structures and citizen trust.

Aside from the introduction and conclusion, the book can be roughly divided in two parts. The first two chapters provide the theoretical foundation of the book. Suitably, for a book titled *Artificial Intelligence and Democracy*, it builds not only on academic literature, but also on literature from democratic institutions, such as the AI Watch reports by the European Commission Joint Research Centre (Misuraca & van Noordt, 2020; Samoili et al., 2020), OECD policy documents (e.g., OECD, 2001, 2019) and the European Commission White Paper on Artificial intelligence (European Commission, 2020).

The first chapter introduces artificial intelligence as a concept and explores its uses in government contexts. Duberry provides a short history of AI, taking the reader from its birthplace; the 1956 Dartmouth Summer Research Project, to what it is today. The author states that we are today in a stage of Artificial Narrow Intelligence (ANI), or applied AI, which can perform dedicated tasks that would require intelligence, but it cannot mimic human behaviour (yet). It is this kind of ANI that the book focuses on. The second part of the chapter introduces the use of AI in the public sector, providing concrete examples of how AI is used by governments for different kinds of tasks, from a system used in Slovakia to create a searchable database to self-driving snowploughs in Norway. While useful for improving performance, Duberry warns us of potential risks, such as the digital divide and how AI is challenging for citizens and policy makers alike.

Duberry manages to steer clear of narrow, instrumental definitions of artificial intelligence. Instead he acknowledges the complexity of the concept and presents the reader with a nuanced view of AI as “a (1) blurry (i.e. conceptual challenges, ongoing developments and multiple applications), (2) sometimes unreliable (i.e. AI technical or adversarial vulnerabilities, data and algorithm bias), and (3) often opaque (i.e. black box phenomenon) technological agent with (4) various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions).” (p. 34).

Chapter two introduces the Multiple Streams Framework (Kingdon 2003) to examine the role of civil society in agenda-setting processes. Duberry argues that the civil society has a policy entrepreneurial role aiming to provide packages of problems and solutions to policy makers at the right moment during the agenda setting processes. The chapter also paints an optimistic picture which contradicts existing narratives that citizen participation is on the decline, instead proposing that citizen participation has been transformed. Traditional citizen participation, Duberry poses, is increasingly replaced by non-conventional forms of citizen participation such as demonstrations. The inclusion of civil society in the policy-making cycle is an important aspect of the citizen-government relations which Duberry discusses in his book. In the concluding remarks, the author reiterates the OECD recommendations to use technology in

government for three types of action: enhancing citizen information access, enabling citizen consultations, and engaging citizens in decision-making (OECD, 2001).

Chapters three to seven examine these actions more closely by exploring the use of AI in different citizen-government relations. Duberry elaborates on the use of AI for social media platforms (chapter 3) surveillance (chapter 4), influencing citizens in national political contexts (chapter 5), disinformation campaigns in international contexts (chapter 6) and civic technology (chapter 7). In each case, we are presented with a multitude of examples of specific tools, tactics and technologies that are being used in these contexts. In contrast to the first two chapters, these chapters apply a narrower definition of AI. This requires attention on the part of the reader, particularly when seemingly generalised statements are made about artificial intelligence or machine learning. Contrast, for example, the statement that “MLAs use data collected through behavioural tracking and cookies technologies to identify what content is the most relevant for the user at any given time.” (p. 81) and the statement that “(...) machine learning improves its capacity and precision with every new search (...)” (p. 210). Both statements refer to different applications of Machine Learning (social media algorithms in the first, natural language processing in the second) and are not valid for all types of algorithms using machine learning.

Chapter three introduces social media as an important platform for civil society and explains the use of Machine Learning Algorithms (MLAs) to filter, rank, and diffuse information. While acknowledging the potential benefits social media platforms have, Duberry warns us of emerging phenomena (e.g., filter bubbles and echo chambers) that obstruct citizens’ informed decision making. Chapter four shows us how the introduction of AI, especially through such platforms, has enabled a new surveillance paradigm. AI provides public sector actors and states access to large amounts of citizen data, particularly collected through social media platforms, as well as the capacity to analyse these data. Duberry questions whether such surveillance is compatible with democratic principles and describes risks to privacy and state security – after all, if the data is a gold mine for states, it is also a gold mine for cybercriminals and foreign intelligence services. The new surveillance paradigm, Duberry concludes “(...) increases the degree of uncertainty and vulnerability in the citizen-government relations.” (p. 118)

Chapter five focuses on uses of AI to influence citizens. AI-powered ‘persuasion tactics’ include e.g. use of semantic analysis of social media posts, A/B testing results, or information about citizens’ whereabouts through location tags. This information enables the targeting of individual citizens with personalised political advertising campaigns, and potentially harming citizens’ unhampered access to information. Beyond advertising, Duberry also touches upon disinformation campaigns such as the dissemination of false news, trolling efforts and social bots. These topics are further explored in chapter six, where we move from the national political context to the stage of international relations. States can (and do) weaponize disinformation to threaten the citizen-government relationship. Where no trustworthy information is available, citizens become vulnerable to manipulation and disengagement from democratic processes. To illustrate, Duberry dissects two cases: the Chinese disinformation efforts during the Covid-19 pandemic and the Russian propaganda projects about AIDS. Both campaigns aimed to challenge the social imaginary of European liberal democracies. AI is also presented as the main defence against such disinformation campaigns. AI, according to Duberry, plays a dual role in disinformation campaigns, and “(...) is at the center of this battlefield (...)” (p. 184)

Chapter seven is all about civic technology, which Duberry defines as technology that aims to increase and deepen democratic participation. While not all civic tech use AI, AI may help increase efficiency and engage a broader range of stakeholders in policy-making processes. Examples are the automatic translation of government information or the analysis of citizen responses on a participation platform. However, civic tech also holds the capacity of deepening the digital divide, as not all citizens have equal

access and digital proficiency. Also, due to the opacity and black-boxed character of many AI uses, outcomes might be difficult to explain to citizens. Duberry warns that this could decrease citizen trust and participation motivation.

Through Duberry's lively writing style, readers are teleported to a reality that was mostly hidden from sight until now. In this reality governments and other actors may use increasingly complex techniques to influence national politics. Relying on data gathered by tracking people through cookies and sentiment analysis, government actors can use targeted political advertising campaigns or distribute disinformation. Simultaneously, on an international scale, authoritarian regimes rely on bots and large-scale trolling farms to undermine citizens' faith in information and democracy. Even the appliance of AI with the aim of strengthening democracy, such as civic tech aiming to enhance participation, come with new risks related to e.g., the digital divide (access) and the opaque, black-boxed nature of AI – making it less susceptible to (public) scrutiny. Considering these risks, the author concludes with a new outlook on developing and using AI, stating that "(...) a dedicated approach to AI for the citizen-government relation is needed and in particular for citizen participation." (p. 232) Duberry calls for risk assessments to defend equality, freedom, human rights, and the notion of popular sovereignty, and warns governments not to focus exclusively on efficiency when implementing AI.

Although Duberry does acknowledge the positive potential of AI in each chapter, the reader is left with a future imaginary that is predominantly gloomy. A future where power increasingly shifts from citizens to (those creating and using) AI. Meanwhile, accountability, transparency and inclusion are not evident, potentially impacting citizen trust and participation. Duberry pays limited attention to efforts to alleviate some of these risks, such as advancements in the use of open-source code, explainable AI, new regulation, or digital literacy campaigns to recognise fake news. Particularly worrying is the fortitude around the phenomena of filter bubbles and echo chambers. These concepts are not undisputed. The current consensus amongst scholars in Media Studies and Critical Algorithm Studies is (to the extent to which these phenomena exist) that the techno-determined manipulation of social media users through MLAs is questionable, and much more nuanced than sketched here (Bruns, 2019). Although the risks Duberry mentions are significant and certainly deserve the amount of attention given in this book, a counter-narrative could make for a more nuanced view on AI in the citizen-government relationship.

This book's greatest strength lies in extensive examples it provides of AI applications, tools, techniques, and strategies that may play a role in the citizen-government relationship. Through these examples, the book offers a very useful introduction to the topic of AI in government and will be a wonderful reference for any scholar or student working on these topics. Duberry tackles big questions that are currently shaping our society, adding essential knowledge to the scholarship hoping to answer such questions while safeguarding important democratic values. Above all, however, it is my hope that governments will read this work and take Duberry's message to heart. It is high time to shift away from efficiency-focused innovation in favour of putting democratic, ethical, and public values centre stage.

Isabelle C. Fest
Utrecht School of Governance
Utrecht University
E-mail: i.c.fest@uu.nl

References

Bruns, A. (2019). Filter bubble. *Internet Policy Review*, 8(4). doi: 10.14763/2019.4.1426.

- European Commission. (2020). *White Paper on Artificial Intelligence – A European approach to excellence and trust*. Luxembourg: Publications Office of the European Union. COM(2020).
- Kingdon, J.W. (2003). *Agendas, Alternatives, and Public Policies* (3rd edn). New York: Longman.
- Misuraca, G., & Van Noordt, C. (2020). AI Watch – artificial intelligence in public services: Overview of the use and impact of AI in public services in the EU. *JRC Working Papers* (JRC120399).
- OECD. (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publications.
- OECD. (2019). *Artificial Intelligence in Society*. Paris: OECD Publishing.
- Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, EUR 30117 EN. Luxembourg: Publications Office of the European Union.