

## COMMENTARY

# ARTIFICIAL INTELLIGENCE, DATAFICATION AND EXPLORING THE MINIMUM CONTENT OF NATIONALITY

JASON TUCKER\*

## TABLE OF CONTENTS

I	Introduction.....	124
II	Datafication of Our Everyday Citizenship and Artificial Intelligence.....	125
III	Quantifying the Minimum Content of Nationality .....	126
IV	Taking the Idea Forward.....	128
V	Summary.....	129

## I INTRODUCTION

Identifying the minimum content of nationality ('MCN'), the inalienable core elements or conditions of citizenship without which one should be considered stateless,<sup>1</sup> has remained a non-starter in stateless studies. From a normative legal perspective, identifying this is seen as a near impossible task. Even attempting to do so has been described as potentially opening Pandora's Box on the debate about where we should draw the line between citizenship and statelessness.<sup>2</sup> The ensuing endless debate has the potential to stall and sidetrack critical work on statelessness.

This is problematic as the dominant narrative of how statelessness is understood, in both scholarship and practice, is as a lack of citizenship, as per the 1954 *Convention relating to the Status of Stateless Persons* ('1954 Convention') definition.<sup>3</sup> Yet, the point at which citizenship stops and statelessness starts is very unclear. While one can point to many nuanced approaches to citizenship and statelessness that recognise the polyvalence, fluidity and interrelatedness of these two social constructs,<sup>4</sup> the 1954 *Convention* and how this is interpreted, as seen in policy and practice, remains very unclear regarding the MCN. The lack of ability to reconcile the 1954 *Convention* definition of statelessness with the messiness of statelessness and citizenship manifests in the creation of various labels to try to patch over the blurry MCN, for example de facto statelessness, risk of statelessness and legal invisibility pending recognition of citizenship. These labels blur not only the definition of statelessness itself, but also states' obligations to protect those

---

\* Dr Jason Tucker is an Assistant Professor in the Global Politics of Artificial Intelligence and Health at Malmö University, Sweden. This work was partly supported by the Wallenberg AI, Autonomous Systems and Software Program — Humanity and Society ('WASP-HS'), funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

1 Gábor Gyulai, 'Should Nationality Have a "Minimum Content"? — Italian Supreme Court Passes Landmark Decision', *European Network on Statelessness* (Blog Post, 19 September 2014) <<https://www.statelessness.eu/updates/blog/should-nationality-have-minimum-content-italian-supreme-court-passes-landmark-decision>>, archived at <[perma.cc/2LWZ-6ZQL](https://perma.cc/2LWZ-6ZQL)>.

2 *ibid.*

3 *Convention relating to the Status of Stateless Persons*, opened for signature 28 September 1954, 360 UNTS 117 (entered into force 6 June 1960) ('1954 Convention').

4 See, eg, Tendayi Bloom, *Noncitizenism: Recognising Noncitizen Capabilities in a World of Citizens* (CRC Press 2017); Christian Joppke, *Citizenship and Immigration* (Polity 2010).

that fall within them.<sup>5</sup> As such, this is not simply an academic exercise to explore the space between citizenship and statelessness. Gaining a better understanding of the MCN will, as is set out here, allow us to improve the identification of statelessness and our ability to predict new cases where statelessness could arise.

We can do so if we critically engage with the concept of the MCN from an empirical and performative perspective. Situated within the context of the ever increasing datafication of even the most mundane aspects of our daily lives<sup>6</sup> and advances in artificial intelligence ('AI'), such as pattern recognition, we are at a point where we can quantify aspects of statelessness and citizenship.<sup>7</sup> These aspects may be compared in ways which were previously inconceivable.<sup>8</sup> I claim this provides us with new data and digital research methods to narrow in on the MCN. I should be clear, I am not saying that AI will be able to provide us with a universally acceptable definition of the MCN. However, it could be a valuable tool in identifying some of the core elements or conditions of citizenship by focusing on the latter's performativity,<sup>9</sup> thus allowing us to gain greater insight into the MCN and the relevant policy outcomes.

## II DATAFICATION OF OUR EVERYDAY CITIZENSHIP AND ARTIFICIAL INTELLIGENCE

The collection of data is something we have been doing for a very long time. Datafication, however, points to a unique contemporary phenomenon where human life has been quantified through digital information, with this process being undertaken for the creation of economic value.<sup>10</sup> Datafication can be understood here as 'the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis'.<sup>11</sup>

The datafication of societies and individuals has massively expanded in scale and scope in recent times.<sup>12</sup> While by no means equally impactful globally, states and non-state actors have increasingly vast amounts of data on even the most mundane, everyday social actions and interactions of large numbers of the population. In some states we have seen the normalisation of the digitalisation of

---

5 See Roger Zetter, 'More Labels, Fewer Refugees: Remaking the Refugee Label in an Era of Globalization' (2007) 20(2) *Journal of Refugee Studies* 172. While this work relates to the *Convention relating to the Status of Refugees*, opened for signature 28 July 1951, 189 UNTS 137 (entered into force 22 April 1954), clear parallels can be drawn to the process of new labels being used by those working with the *1954 Convention*.

6 Sarah Pink et al, 'Mundane Data: The Routines, Contingencies and Accomplishments of Digital Living' (2017) 4(1) *Big Data & Society* 1.

7 John Cheney-Lippold, 'Jus Algoritmi: How the National Security Agency Remade Citizenship' (2016) 10 *International Journal of Communication* 1721.

8 Arne Hintz, Lina Dencik and Karin Wahl-Jorgensen, *Digital Citizenship in a Datafied Society* (Polity Press 2019) ch 2. This is not a one way relationship whereby our lived experience of citizenship is simply quantified. Rather, datafication also facilitates new forms of governance and shapes social identities and practices.

9 Engin Isin and Evelyn Ruppert, *Being Digital Citizens* (Rowman & Littlefield Publishers 2020) 22–3.

10 Ulises A Mejias and Nick Couldry, 'Datafication' (2019) 8(4) *Internet Policy Review* 1.

11 José van Dijck, 'Datafication, Dataism and Surveillance: Big Data Between Scientific Paradigm and Ideology' (2014) 12(2) *Surveillance & Society* 197, 198, citing Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray 2013).

12 Hintz, Dencik and Wahl-Jorgensen (n 8) 54–6.

public administration,<sup>13</sup> public health,<sup>14</sup> education,<sup>15</sup> borders and migration<sup>16</sup> and citizen engagement in policy making.<sup>17</sup> Many states are increasingly relying on digital platforms to interact with their populations and using private actors to manage their relationship with their citizens.<sup>18</sup> Framed as cost effective and accessible, these platforms also collect data on their users. This has been greatly facilitated by the use of certain technology that has the capacity to quantify our daily lives, such as smart watches and mobile phones, which collect huge amounts of data on even our most mundane actions and interactions.<sup>19</sup>

While states were traditionally the ones collecting data on their populations, the economic incentives for the private sector to create and collect data have meant that the latter has moved into existing data markets and is constantly seeking to create new ones. As such, it is now the private sector that is the main proprietor of data<sup>20</sup> and, by and large, has ‘better’ data than states do.<sup>21</sup>

The availability of sufficient data is only one piece of the puzzle, as one also needs to sift through and make sense of the data; this is where AI comes into play. It is important to remember that when we talk about AI, we talk about a family of ever-evolving technologies. Regarding the vast datasets discussed in this part, AI with functions like pattern recognition is the only viable option. It allows for engagement with massive datasets in new and innovative ways, identifying patterns and making predictions of future outcomes. For example, this technology is excellent for weather forecasting. By taking in all the data of past and current weather conditions, patterns can be identified and predictions can be made for the weather to come.<sup>22</sup>

So, what does this all mean for the ability to further narrow in on the MCN? The datafication of society and the interconnected advancements in AI technology mean that there is scope for empirically grounding this puzzle.

### III QUANTIFYING THE MINIMUM CONTENT OF NATIONALITY

In terms of what better understanding the MCN could mean for academics and practitioners in stateless studies, I want to focus on two main areas that speak to both communities: identifying statelessness and predicting the occurrence of

- 
- 13 Gianluca Misuraca and Colin van Noordt, *AI Watch, Artificial Intelligence in Public Services: Overview of the Use and Impact of AI in Public Services in the EU* (Report, July 2020).
- 14 Klaus Hoeyer, Susanne Bauer and Martyn Pickersgill, ‘Datafication and Accountability in Public Health: Introduction to a Special Issue’ (2019) 49(4) *Social Studies of Science* 459.
- 15 Juliane Jarke and Andreas Breiter, ‘Editorial: The Datafication of Education’ (2019) 44(1) *Learning, Media and Technology* 1.
- 16 Matthias Leese, Simon Noori and Stephan Scheel, ‘Data Matters: The Politics and Practices of Digital Border and Migration Management’ (2022) 27(1) *Geopolitics* 5.
- 17 Hintz, Dencik and Wahl-Jorgensen (n 8) 27.
- 18 *ibid* 60.
- 19 Constantine Gidaris, ‘Surveillance Capitalism, Datafication, and Unwaged Labour: The Rise of Wearable Fitness Devices and Interactive Life Insurance’ (2019) 17(1/2) *Surveillance & Society* 132.
- 20 The global political economy of data is highly unequal, with power being concentrated in a few private sector actors or public–private partnerships in the United States of America, China and, to a lesser extent, the European Union states: See Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021) ch 6.
- 21 By this I mean the private sector often has higher levels of variety and velocity in their datasets, these being seen as broad benchmarks for ‘good’ data.
- 22 Martin G Schultz et al, ‘Can Deep Learning Beat Numerical Weather Prediction?’ (2021) 379(2194) *Philosophical Transactions of the Royal Society* 1.

statelessness. Being able to identify statelessness has always been a problem.<sup>23</sup> At an individual level it can be a hugely resource-intensive endeavour.<sup>24</sup> Doing so across large and diverse populations, where there are no grounds for group recognition, is an even greater challenge. Additionally, the reliance on states to identify statelessness is very problematic. The lack of stateless determination procedures and the highly politicised nature of statelessness in many contexts means that our knowledge of statelessness is, at best, patchy.

Using AI to analyse the vast amounts of data on a population's everyday lived experience of citizenship allows us to approach the MCN from an empirical and performative angle, rather than a normative one. With the array of data on how we act and interact as members of a population, we have the means to identify patterns and subgroups of marginalisation and discrimination within certain populations of citizens, those experiencing statelessness and those in-between. This is similar to when we look for patterns in establishing the 'operation of the law' when determining statelessness, though at a much larger, more complex scale which can be constantly updated.

Interdisciplinary teams with contextual knowledge would be able to develop systems to begin to narrow in on aspects of the MCN. They could borrow ways of acknowledging the fluidity of citizenship from citizenship studies, such as Kristian Stokke's analytical framework of the 'dimensions and stratification of citizenship'.<sup>25</sup> One could also draw on the existing research to focus on the least 'vulnerable' stateless populations or persons and the most 'vulnerable' citizens in certain regions or countries, as the MCN can be found somewhere between the two.<sup>26</sup> This would lead to identifying 'vulnerable' populations who may be considered citizens; a group of interest to many human rights and statelessness practitioners. In addition, this would allow for the academic exploration of the MCN by grounding this concept in its everyday performativity.

The digitalisation of society, driven by the economic incentives of datafication, means that the private sector has a wealth of data on many of our lives. Thus, states would not always be necessary partners in exploring the MCN. This is useful as while states may be reluctant to collect and share data on various groups in society, such as the stateless or very marginalised citizens, the private sector does not have such qualms.<sup>27</sup> These datasets are large, varied and constantly updated, and as such, contain rich pickings to feed into AI. Datasets could be bought by researchers or gifted to them by private actors to train AI models to identify the MCN.

States themselves could also capitalise on AI to identify the MCN. They could merge their national data sets, where they exist, with private actors' data to better understand statelessness and citizenship amongst their population. For example, by looking at experiences of access to healthcare, education, employment and housing, and how these vary within a population, one could identify subgroups

---

<sup>23</sup> Heather Alexander, 'The Ethics of Quantifying Statelessness' in Tendayi Bloom and Lindsey N Kingston (eds), *Statelessness, Governance, and the Problem of Citizenship* (Manchester University Press, 2021) 238.

<sup>24</sup> *ibid* 242–3.

<sup>25</sup> Kristian Stokke, 'Politics of Citizenship: Towards an Analytical Framework' (2017) 71(4) *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography* 193, 195.

<sup>26</sup> In so doing one would not have to map the entire population but could use previous research to focus in on a point where the lived experience of citizenship and statelessness 'meet' or overlap.

<sup>27</sup> See, eg, Yiquan Gu, Leonardo Madio, and Carlo Reggiani, 'Data Brokers Co-opetition' (2022) 74(3) *Oxford Economic Papers* 820, 820–39.

whose experiences are similar to those recognised as stateless persons in that state. By narrowing in on the MCN, it could also reveal how citizenship is stratified and which groups are more disadvantaged. For example, AI systems have already highlighted discriminatory practices in a range of settings.<sup>28</sup>

One must not be naïve to the dangers of AI, as the application of AI systems with deliberate or unintended algorithmic or data biases can have devastating impacts on people and communities.<sup>29</sup> Nor should one believe that a digital divide does not exist.<sup>30</sup> AI itself is neither good nor bad, but rather a reflection of our bias in the data being used and the design of the system itself. Yet, while AI can be used for targeting and persecuting certain populations,<sup>31</sup> it could also be used to identify vulnerable people and groups and ensure that they receive adequate support or protection.

Regarding the prevention of statelessness, AI could be used to predict the occurrence of statelessness and improve the implementation of pre-emptive policy interventions. Similar to the pattern recognition that allows us to know what weather is likely tomorrow or the next day, AI used to identify the MCN could also be used to monitor populations whose experiences of citizenship begin to shift towards identified ‘danger zones’. This is extremely useful as statelessness is more often than not a sociopolitical phenomenon before it is a legal one.<sup>32</sup> Once it reaches a certain level of sociopolitical discourse, it becomes entrenched and is tough to counter. With vast and constantly updated data on experiences of citizenship, as well as the ability to factor in the shifting socio-political and policy context, we could have a powerful early warning system for the risk of statelessness. This could be used to highlight where and when policy interventions should be implemented at a local, regional or national level. Even if these interventions were to fail, there also would be more time to plan and allocate resources to respond to emerging statelessness situations as they happen, thus avoiding the need for knee jerk reactions.

#### IV TAKING THE IDEA FORWARD

To achieve the above, the operationalisation of citizenship and statelessness would need to be decided upon so that the AI algorithm could be built. Doing so would not mean defining a universally accepted notion of citizenship or statelessness. Instead, as discussed, we would only have to look at those whose lived experience is around the MCN. There could be multiple variations of these algorithms focusing on aspects of citizenship and statelessness in different contexts and at different times. Indeed, given the complexity of this lived experience, multiple inputs would be essential in allowing us to further narrow in on the MCN.

---

28 Ziad Obermeyer et al, ‘Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations’ (2019) 366(6464) *Science* 447.

29 Safiya Umoja Noble, *Algorithms of Oppression* (New York University Press 2018); Carolina Criado-Perez, *Invisible Women: Exposing Data Bias in a World Designed for Men* (Chatto and Windus 2019) ch 8.

30 Sophie Lythreathis, Sanjay Kumar Singh and Abdul-Nasser El-Kassar, ‘The Digital Divide: A Review and Future Research Agenda’ (2022) 175 *Technological Forecasting and Social Change* 1.

31 House of Commons Canada, *The Human Rights Situation of Uyghurs in Xinjiang, China* (Report, March 2021) 23–4.

32 See generally J Tucker, ‘Challenging the Tyranny of Citizenship: Statelessness in Lebanon’ (PhD Thesis, University of Bath, 2014).

The next step would be the data; more specifically, the access to and the labelling of the data.<sup>33</sup> Research would have to be undertaken to see what data exists, whether it would be available from states or the private sector and for which members of the population it could be obtained. It would also need to be determined whether the data should be shared in an unlabelled format, or as a labelled dataset based on certain criteria.<sup>34</sup> When labelling the data in general, what criteria are to be used would need to be discussed, as well as how one should try to navigate bias in the data, its labelling and the development of the AI system. These questions are critical, as even with a well-designed AI system, the quality of the data matters. As the saying in computer sciences goes, ‘garbage in, garbage out’.

## V SUMMARY

AI and the vast datasets resulting from the datafication of society are being used effectively in many other fields related to statelessness, such as citizenship, migration and refugee studies. These can be drawn on as inspiration for how to create new knowledge and better inform policy decision-making about statelessness. AI will not be able to identify a clearly defined and universal MCN. Such a thing does not exist, and despite popular perceptions, technology, and AI in particular, cannot solve all our problems. However, the development of this technology, as well as the datafication of society, provides us with an opportunity to begin to empirically narrow in on this concept of the MCN. This would not simply be a normative exercise, whereby Pandora’s Box would be opened for its own sake. Rather, it could have significant benefits for research, policy and practice for a range of actors. The possibility is therefore one that we should at least entertain and begin to discuss more concretely.

---

33 By this I am referring to the work that goes into making it ready for training the artificial intelligence (‘AI’) applications. Doing so is problematic as it is very time and resource intensive, but also because we can see further bias creep into datasets. There would also need to be considerable care in designing this process regarding who would have access to the data, own it, pay for it, store it and process it. There is not scope within this paper to draw out these complexities, but turning to good practices and lessons learned in the refugee field where similar work is being undertaken would be a good starting point.

34 The difference between labelled and unlabelled data is whether it has been pre-coded or provided classifications before it is received. Unlabelled data provides the opportunity to develop, implement and better manage a coding process tailored to specific needs, as it has not been coded before it is received. However, this is a very time and labour-intensive undertaking, whereas labelled data can be used immediately to train the AI system: see Natalie Kudan, ‘The Difference Between Labeled and Unlabeled Data’, *Toloka* (online, 3 March 2023) <<https://toloka.ai/blog/labelled-data-vs-unlabelled-data>>, archived at <[perma.cc/XKR6-WPDA](https://perma.cc/XKR6-WPDA)>.