#TRANSFORMATION – THE TRANSFORMATION PROCESS – THE ROAD TO A DATA-DRIVEN (INTELLIGENCE) ORGANIZATION

BIG DATA AT THE BORDER: WHERE DO WE DRAW THE LINE?

Liv DORAK*

Abstract:
The buzz of "Big Data" permeates society. This 'new' data, characterised by its volume, variety, velocity, and veracity changes the ways we live, as an increasing number of actors from a variety of sectors in public/private spheres and civilian/military contexts seek ways to operationalise data for value. This paper seeks to offer an additional viewpoint from which we can explain some of the challenges and limitations on the road to data-driven intelligence and its integration in security and intelligence organisations. The scope of this paper is refined to analysing such challenges in the context of border security, although similar challenges might and do appear in other applications and fields. An auto ethnographic approach is adopted as a way to illustrate and remind that the data we collect, process, and store is data about people, and further, how we choose to use (or not use) this data has implications for people. The bottom line of this work is that as we pave and forge towards data driven intelligence, we must acknowledge and remember that the road both starts and ends with human intelligence.

Keywords: *data-driven intelligence, Big Data, artificial intelligence, borders, security*.

Introduction

Improvements and innovations in information and communications technologies (ICTs), in tandem with more efficient and affordable mechanisms to travel and exchange ideas, produce today's increasingly interconnected information society. The volume and velocity with which data is produced, collected, and analysed is unprecedented; this offers benefits for fields such as science, medicine

* MA student, International Master's in Security, Intelligence, and Strategic Studies (IMSISS) Programme; University of Glasgow, Dublin City University, and Charles University, email: livjdorak@gmail.com

and healthcare, business, commerce and trade, transportation, law and criminal justice, and security. In these contexts, capitalisation of Big Data can assist in supporting decision makers to make well-informed judgements and minimise uncertainty, cognizant of historical evidence, facts, and trends. For policymakers, governmental agencies, and security practitioners, Big Data can also support activities to mitigate threats to their populations' values, freedoms, rights, and lives. Intelligence agencies and departments have a critical role: they must collate, process, and interpret data and communicate their analyses with the aims of maximising benefits and opportunities and minimising pitfalls and threats. Within the Schengen and EU zones, for instance, as citizens share common values and enjoy freedom of movement, it is imperative to strike a balance that permits this mobility, but also offers protection, security, and justice. Border security, control services, and intelligence agencies thus have an important responsibility to maintain and strengthen social and physical borders in pursuit of these goals.

Data collection and sharing pertaining to citizens, foreigners, movements, and threats are critical for the protection of Member States and the collective Zone. Socio-technical systems of surveillance and border control offer spatial and temporal advantages; however, there are also drawbacks regarding their use. This paper seeks to offer an alternative perspective from which to consider data-driven intelligence by elucidating some of the challenges, limitations, and implications in the context of border security. An auto ethnographic approach is adopted as a way to illustrate and remind that the data we collect, process, and store is data about people, and further, how we choose to use (or not use) this data has implications for people. To achieve its purpose, this paper first traces the history of the Schengen Area, before discussing relevant current and upcoming instalments of information technology systems that are deployed for intelligence purposes. Subsequently, this piece comments on the benefits and opportunities Big Data, data-driven intelligence, and artificial intelligence offer for border stakeholders. Highlighted are three emergent concepts of how Big Data and information technologies have increasingly been deployed for purposes relating to surveillance, migration, and border control; these concepts are "dataveillance," "social sorting," and the "Ban-

Opticon." The paper then analyses three caveats of data-driven intelligence in border management, namely concerns of biases and fairness, a lack of context, and reliance on the past as an oracle for the future. This paper then speculates about additional obstacles of Big Data, data-driven intelligence, and artificial intelligence (AI) by delving deeper into the ramifications of the COVID-19 pandemic. It concludes with final reminders about the importance of human involvement in human-techno problem-solving approaches and paving the way forward for facets of border control, migration, surveillance, intelligence, and security.

Entering the Schengen

In the past centuries, as the world became globalised and interconnected, fear of the foreign and the unknown settled in the hearts and minds of citizens and institutions alike. In tandem, populations grew and travel became more affordable, putting strains on the abilities and resources of the state. This became apparent saliently in the contexts of law enforcement and authorities recognised a need to regain control for the safety and security of law-abiding citizens and the whole-of-society (Cole, 2002). Nevertheless, the aspiration for freedom of movement between European nations, a dream dating back to the Middle Ages ("The Schengen Agreement"), was not lost. During the 1980s, concrete discourse and groundwork officially broke for the establishment of a border-free zone; spearheaded by France and Germany, on 14 June 1985, an agreement abolishing internal borders within the termed "Schengen Area" ("The Schengen Agreement") was signed by France, Germany, Belgium, Luxembourg, and the Netherlands ("The Schengen Agreement"). Since then, the Area has expanded more than five-fold, and with it, the development of social, judicial, legal, and technological frameworks to ensure safety and security within the territories. The current area comprises 26 countries, of which 22 are also Member States of the European Union (European Parliament, n. d.). Once within the Schengen Area, travel is permitted internally from one country to another with people generally not subject to border checks (European Parliament, n. d.).

A System of Systems

Building a Europe without internal border controls implies the concern to harmonise and reinforce (European Parliament) external borders; indeed, to guarantee internal freedom of mobility for citizens, policymakers have "sought to compensate for loss of control towards third-country nationals" (TCNs) (König, 2016). Currently, largely three interconnected databases— the Schengen Information System II (SIS II), Visa Information System (VIS), and the European Dactyloscopy (EURODAC)—form the technological infrastructure that allows for storage and retrieval of information relating to border security and migration; in the coming years, these systems are expected to be further supplemented by the Entry/Exit System (EES) and the European Travel Information and Authorisation System (ETIAS) (European Commission). These systems are interoperable and complementary. SIS II contains alerts for the purpose of refusing a TCN entry or stay in the Schengen zone (Regulation (EC) No 1987/2006).

As it is the second-generation rendering of the databased ICT infrastructure, the newer system allows permitted authorities the ability to access and exchange information concerning issued alerts on individuals and objects (Directorate General for Communication, n. d.). VIS serves to improve the implementation of common visa policy and reinforce cooperation between consular parties (Regulation (EC) No 767/2008) by allowing access to information regarding visa applications, acceptances, and refusals (Regulation (EC) No 767/2008).

The system can also be used to identify and verify persons present within the Schengen Zone, and their right of stay (Directorate General for Communication). The EURODAC database compiles data on applications filed by TCNs or stateless persons for the purposes of international protection; the database also assigns responsibility for the individual to the Member State wherein the application was lodged (Regulation (EC) No 203/2013). Querying this database, authorities can verify the identity of applicants or those crossing physical borders (Regulation (EC) No 203/2013). Expected for implementation and operationalisation within in the next two years, EES and ETIAS are the latest additions to the border, migration, and security ICT arrangement.

The EES collates the identity and travel documents of non-EU nationals, registering details of their entries and exits at border crossing points (European Commission). The system will replace the manual stamping of passports and calculate the permitted duration of stay, facilitating authorities' abilities to identify and trace over-stayers (Regulation (EC) No 2017/2226). ETIAS collects information on non-EU nationals who travel under visa-free regimes for stays under 90 days (European Commission, n. d.); the system thus issues authorisations for travel as a precondition for entry into the internal territories (Regulation (EC) 2018/1240).

Collectively, these systems strive to offer benefits such as assisting in the fight against terrorism (Bux, n.d.), combatting transnational crime (Bellanova and Glouftsios, 2020), addressing irregular migration (Bellanova and Glouftsios, 2020; König, 2016), thwarting human trafficking, and offering assistance in search and rescue missions (Bux, n.d.). These infrastructures support migration and security policy and help to engender an area of freedom, security, and justice (König, 2016). In the context of budgetary pressures, resource constraints, high costs of training and maintaining human personnel (Haggerty, 2006), and the perceived unreliability of these human agents, automated systems that filter, monitor, and alert are advantageous. Furthermore, higher traveller traffic and increasing volumes of data campaign for an imperative to organise, rank, and prioritise individuals based on the threat they pose to society (Duwe and Kim, 2016). Databases, automated systems, and AI-algorithmic processes prove comparably inexpensive and efficient, reducing the ratio of human-to-technological monitoring (Haggerty, 2006). In addition, these systems offer benefits to travellers in the form of facilitating faster entry/exit speeds and greater efficiency and security (European Commission for Migration and Home Affairs).

From enhanced ICT infrastructure, systematic collections of data, ubiquitous surveillance mechanisms, and heightened uncertainty and fear, such as following attacks in Madrid in 2004 and London in 2005 (König, 2016; Bigo, 2006), three distinct concepts for protecting internal society, enhancing security, and vetting external parties emerge: "dataveillance," "social sorting," and the "Ban-Opticon."

"Dataveillance" is a term coined in 1988 to suggest that today's governing actors can more easily trace and track individuals and groups compared to authorities of the past; this ease is attributed to facilitation based in computer-based technologies, faster computational processing, larger storage capacities, and increased access to digitised information (Clarke, 1987; Galič, Timan, and Koops, 2017).

"Social sorting" describes the acquisition of personal and group data, with the intention of classifying people into predetermined categories, which allows for targeted intervention, "special treatment, suspicion, eligibility, inclusion, access," (Lyon, 2003) and more. While data categorisation and the demarcation of populations are not new phenomena in themselves, an uptake in perceived risks (Lyon, 2003) contestably justifies the deployment of more surveillance devices. Furthermore, as practices using Big Data accelerate "routine and systematic searches of data doubles," (König, 2016) human operators are increasingly at a distance both physically and psychologically. Lastly, the 'Ban-Opticon' refers to tactics of surveillance whose primary purpose is not to monitor or capture misbehaviour, but to prohibit certain "bad" (Galič, Timan, and Koops, 2017) individuals from some type of access or opportunity. The concept proposes that, particularly in times of crisis or emergency, government unease and insecurity permits the delineation between perceived hostile foreigners ("them") and benign citizens ("us"); moreover, the state of emergency rationalises "practices of exceptionalism, acts of profiling and containing foreigners, and a normative imperative of mobility." (Bigo, 2006). Commonly, these concepts tie together xenophobic tendencies, thus supporting technological and sociotechnical means and motives; the imperative to vet, detect, and monitor foreigners within or attempting to enter internal borders; and extracted data-driven intelligence to drive future mobility policy. Moreover, recent attacks in France in 2015 and 2016, Belgium in 2016, and the UK in 2017 amplify the relevancy of these concepts, their use, and their effects in practice.

An Understanding of Self

In an attempt to understand and shed light on Big Data, border security mechanisms, and intelligence services, this paper adopts auto ethnography as a methodology, in consort with examples collected from informal conversation with other (anonymised) TCNs. The rationale behind this approach is threefold. First, this paper strives to offer an alternative lens from which to view the use of a data-driven approach in border security. Recognising insider knowledge entails one is privy to degrees and types of specialised knowledge, a fresh look from the outside-looking in or a "non-expert" can pave the way for critical reflection. Further, despite lacking an intimacy with the mechanisms of border control that are achieved through professional training and working experience, this author posits that personal experiences offer a different, yet equally valid, epistemological approach and contribute both to 'non-expert' expertise and a degree of credibility. Auto ethnographic accounts and communications with others who have lived through similar experiences have particular relevance for border control and other security applications, for these techniques allow us to avoid generalising on behalf *of*, but rather allow us to speak *with* those affected (Jarvis and Lister, 2013). Finally, embracing an auto ethnographic account can illustrate themes of belonging and self-identity (Weber, 2011), while also humanising the notions of a data-driven approach to problem-solving. As will be expounded upon later, an interesting aspect of Big Data is the attempt to categorise, organise, and sort mass collections of data, which has arguably reduced human beings to their "data doubles."(Galič, Timan, and Koops, 2017). Puzzlingly, this process allows for both hyper personalisation and depersonalisation (Dunlap, 2018). For border controls, this means the individual is dually a sum of their individual data and none of their individual data.

There are limitations and caveats to such an approach. Most saliently, writing as a TCN implies the author might hold predispositions or bias regarding the use of data-driven processes for matters of border security. Secondly, this paper is limited to its collected data, herein, personal and first-hand accounts acquired

through informal conversations with other TCNs. This data collection caveat entails there is an extent to which the data can be alleged as confirmational, that is, consistent with leading judgements or conclusions. It should also be noted these accounts are not universal.

Lastly, this paper was written in the first year of the COVID-19 pandemic, recognised around the world as a time of immense uncertainty and unknowns. Data and knowledge acquired in hindsight or as people and the world adjust and learn to cope with the pandemic and its ramifications might yield differing viewpoints, conclusions, or recommendations. Weighing these considerations, however, this approach is deemed justified and appropriate as it first and foremost illuminates an alternative perspective from which to view data-driven intelligence for border services and decision making. Secondly, this author attempts to mitigate personal biases and experiences with border control by offsetting these with her diverse academic and professional background in both the social sciences and information systems and technology. These fields are at times divergent, or even at odds, and unable to find a common ground or communicate with one another. Thus, the ethnographic approach enables the author to communicate and comment on a subject of not only personal significance, but also one on which she has pluralistic subject-matter insight.

Fairly Biased?

Three sociotechnical challenges inherent to data and data-driven technological processes further underpin data-driven intelligence for the purposes of border security and control; consideration of these limitations consequently offers authorities a stronger ability to manage or minimise the negative consequences or adverse ramifications. The first concern is in regard to bias and fairness. Paradoxically, while automation intends to remove the problem of decision-making based on individual biases/prejudices by instead using a data-driven approach, this ignores problems of prejudicial, sample, or measurement data biases. Likewise, the notion of 'fairness' as a societal value is also contestable when human beings are subject to deconstruction into data

categories, and the credence of particular attributes working to their favour/detriment is indeterminate.

Concerning bias, prejudicial biases arise from social stereotypes, orthodox opinions, and their influences (Mehta, Shah, Patel, and Kanani, n.d.). The aforementioned databases store information pertaining to, *inter alia,* names and aliases; place and date of birth; sex; nationalities; "specific, objective physical characteristics not subject to change" (Regulation (EC) No 1987/2006); biometrics; passport and visa information; residence; occupation and employer; and anticipated travel dates (Regulation (EC) No 1987/2006).

Interestingly, while some systems, such as the SIS II, can include data and issue alerts regarding EU citizens or property (e.g. vehicles), the delineation between 'citizen' and 'non-citizen' is apparent; to this point, Article 30 of the Regulation establishing SIS II prescribes the erasure of data on individuals who acquire the citizenship of a Member State (Regulation (EC) No 1987/2006). Removals could imply the systems are biased to favour citizens; amplified to data-driven decision-making, the collection of information is thus incomplete and implies TCNs pose a disproportionately higher risk and threat compared to citizens with concern to matters of internal security. Randomness or irregularities in data samples can also predispose data-driven intelligence and technological tools to particular biases (Mehta, Shah, Patel, and Kanani, n.d.).

The example highlighted above is also an illustration of sample biases in border control practices; indeed, by redefining the dataset once an individual becomes a citizen, the data sample for a data-driven approach might no longer depict accurate threat levels. Finally, measurement biases occur in the parameters or the "features we wish to incorporate" (Mehta, Shah, Patel, and Kanani, n.d.) into systems (Bollier, n.d.). Measurement biases alarmingly manifest when data is employed in AI-based algorithms and systems. AI algorithms assign specific 'weights' to specific parameters in order to compute functions that explain the correlation between input values and output results (Waldorp, n.d.). This is best illustrated by way of another example; in this case, a given database stores input information regarding nationality, sex, and the type of visa application (study, work, business,

etc.). During the 'learning' stages of AI, algorithms attempt to compute a function that connects this training input data with the outcome data, which is in this instance visa application approval/denial. If the sample of training data includes more female applicants whose visa applications were rejected compared to their male counterparts, the algorithm might attribute more credence to this variable as an indicator for a future female applicant's approval or, more likely, denial.

Looking to fairness, mindful that the definition changes over time in line with societal views and priorities (Sylvester and Raff, 2018), database designers, software developers, and the AI field adopt "various methodologies that reify fairness as a social concept into fairness as satisfiable technical criterion" (Green and Hu, 2018) or define the concept by a statistical metric. However, even reifying fairness to statistical metric fails to account for "intuitively constructed data association rules that (…) are coded into the software supporting the functioning of databases used by border guards" (Bellanova and Glouftsios, 2020). This argument implies the work of programmers and database developer's influences or has consequences for future risk flags or warning indicators (Bellanova and Glouftsios, 2020) that are used to process travellers' data and inform the proper stakeholders. Here, it can also be argued in several democracies, there are increasing demands for transparency as a surrogate for fairness.

The extent to which fairness then exists is compounded by the "black-box" nature of data and AI algorithms (Goodfellow, McDaniel, and Papernot, 2018); stated differently, in databases, the human being is reduced to the particular data attributes or "parameters" they possess, such as "sex," "nationality," or "occupation." Neither the weights assigned to these parameters, nor the "thought process" and decision-making of the algorithm when processing test data is explicit. In the context of border security, for example, a lack of transparency in decision-making makes it difficult for the issuers of visas to explain the reasoning for application approval/denial and even more complicated for applicants to file an appeal. It is additionally recognised that current border control and wider surveillance and mobility measures themselves "are unequal and do not target the same people in the same way. They reinforce the advantages of some and the disadvantages of

others, even if sometimes they have contradictory and unpredictable effects" (Bigo, 2006). At border checkpoints, discrimination can emerge when individuals are singled out based on their appearance, behaviour (ETICAS, n.d.), citizenship or nationality, or many other criteria. This can result in greater suspicion and scrutiny, longer questioning, additional checks, or undue influence in denying/permitting entry/exit. Where Big Data and data-driven intelligence come into consideration, it would be further unfair and unethical to categorise and segregate individuals on the antecedent treatment of other travellers who fit the same appearance, routine, citizenship, or nationality criterion. This reinforces discrimination and unfairness under the guise of evidence-based objectivity.

Missing the Big Picture

A related, second limitation of employing Big Data, data-driven intelligence, and AI for functions of border control and security is a lack of context. While data collection procedures, intended purposes, and interpretations are subject to human bias as demonstrated above, the raw facts themselves are taken out of context, ignoring a possible degree of significance. Elements such as "female," "loop right thumb fingerprint," "Italian citizen," or "date of birth 8 August 1987" alone do not offer value; it is the wider contexts, analysis, and human explanation or judgement that offer meaning and allow the transformation of data into *intelligence*. Big Data, data-based technologies, and data mining have supported the deconstruction of human beings into data elements, the amalgamation of which forms a person's "data double" (Galič, Timan, and Koops, 2017). In an era of surveillance increasingly reliant on data and ICT infrastructures, human beings are perceived as *assemblages* (Haggerty and Ericson, 2000), or "devices hosting opaque flows of auditory, olfactory, visual, and informational stimuli" Galič, Timan, and Koops, 2017). For border matters, the development and deployment of risk assessment instruments are reliant on Big Data, AI-algorithmic capabilities, and these *assemblages* can pose serious ethical and security related concerns. Risk assessment instruments are tools that use the multiplied

product of the probability and the expected adverse effect of a societal harm (Paul, 2018) to compute a score or degree to which individuals pose a threat to society. Integrated in mobility and migration matters, pre-arrival vetting in the form of background checks, visas, permits, and risk assessment profiles are all techniques used to support decision-making on minimising threats internal society.

Increasingly common in other fields, such as criminal justice throughout the stages of pretrial, trial, sentencing, and parole (Duwe and Kim, 2016), actuarial risk assessments are presented as fair and objective. Indeed, the *raison d'être* for using statistical and actuarial methods in these contexts is to "prevent racism, sexism and other forms of discrimination that can be part of discretionary decisions made by humans" (Dekkers, van der Woude, and Koulish, 2019).

In recent times, access to larger batches of data supports improvements in AI and AI-based risk assessments. To generate these assessments, AI algorithms detect patterns between input data and outcomes by adjusting the weights of certain criterion or data categories (such as age, gender, race, nationality, criminal history) (Dressel and Farid, 2018). The result is AI algorithms develop a proposed function of causality; subsequently, when new data is subjected to this function, software seeks to detect similar patterns and compute a rank or score detailing the degree to which an individual poses risk to society. However, lacking external context, these tools can impair decision-making or have wider ethical and legal consequences. To demonstrate is an example of a TCN in possession of a short-term tourist visa during the COVID-19 lockdowns beginning in February-March 2020. While some were able to either return to their home countries or extend their right of stay in a given country due to travel restrictions and bans, this particular TCN was unable to return home or exit due to limitations in flights and the number of homebound returnees permitted entry; furthermore, with the rapid and prolonged closure of embassies and consulates, attempts to obtain consular assistance or visa extension proved futile. Database entries might demark this individual immediately as *illegal*, having overstayed their visa. In the short-term, alerts might be triggered for border and law enforcement officials. In the long-term, the TCN might, therefore, face

future barriers to mobility, extended checkpoint screening, or visa application rejections; as such databases and alerts might lack the contextual hampering effects of the external environment.

Risk profiles can also be generated for individuals who, in themselves, are not the direct subjects of inquiry for border and security concerns. With SIS II's ability to create links between people-and-people and people-and-objects, the system might detect links between a person about whom intelligence is gathered and an object for which an alert is issued (Bellanova and Glouftsios, 2020). These technologies also construct networks of relations, such as travel groups or family members; in practice, "If an alert is issued on one member of this 'network,' other members are automatically controlled too. Thus, an alert affects more than the concerned file" (König, 2016). However, these additional uses and perceived benefits of technologies can be caveated by the inherent collected (or uncollected) data, the lack of circumstantial information, or, in some cases, infringement on human rights. In the former illustration, data not collected might be that the car's owner reported its theft to the police months prior; in the latter, even if the only linkage between individuals is a day-tour group, the attributed linkage might disproportionately affect innocent parties. It can be argued this practice contravenes European fundamental human rights (European Court of Human Rights and Council of Europe, 2012) as it evokes the "surveillance of third parties who did not, for themselves, cause a reason for being surveilled" (König, 2016).

An All-Seeing Eye

A final limitation is the challenge of the extent to which past data serves as an oracle to presage the future. For risk management purposes, this is the use of Big Data and AI tools to support a form of pre-emptive governance that arguably "grants databases the 'power to predict' events" (König, 2016). Currently, the accelerated scope and speed of Big Data—through networked computing power, profiling, and data-mining (Broeders and Hampshire, 2013)—debatably enable better-informed decision-making. Complications surface from two caveats inherent to Big Data and AI and three assumptions about their

deployment in supporting decision-making. The first caveat is the temporal stationarity of data. Data is ephemeral, capturing a specific moment in time. Concerning border control, an example of this caveat is the retention of vast amounts of data in databases, which requires continual maintenance and updating (Bellanova and Glouftsios, 2020) to reflect dynamic individuals and circumstances. In addition, input errors, in the form of incorrect data entry, or a failure to input the most recently collected data, can result in particular depiction that might or might not be accurate. For instance, an applicant seeking international protection might be permitted temporary stay in a Member State and later granted permanent residence. If databases such as EURODAC are not updated, the applicant's residency status remains static and outdated, effectuating possible serious ramifications.

The use of historical data also underlines the second caveat to Big Data and, in particular, AI, which is envisioning static, past data as relevant for the present and future. Phrased differently, there is a reliance on expectations of continuity (Bennett Moses and Chan, 2018); AI models in particular seek to detect "historical correlations between features and outcomes…applying those correlations to new data under the assumption that those same correlations will apply" (Green and Hu, 2018). This issue is again prominent with the development of risk assessments. In addition to lacking contextual information, risk assessments might prioritise past, external data, opposed to current data. In an oversimplified example, a data sample includes 1000 individuals attempting to pass border checkpoints in a particular Member State, 500 of whom were permitted entry and 500 of whom were denied entry. Of the 500 individuals denied, 400 were of a certain sex and nationality.

If Big Data and AI-driven technologies undertake similar border filtration mechanisms in the future, when an individual of that same sex and nationality attempts to cross the border, the technologies might conflate the person's correlations (sex and nationality) with historical outcomes (denied entry) rather than analysing the individual's particular current data and circumstances. Also relevant to risk profiles and actuarial tools that aim to assess individuals' risk level, such profiles are underscored by "assumptions concerning the possible

future, or more exactly the belief that the intelligence services have a grammar of '*futur anterieur*'" (Bigo, 2006) that technological profiling equates to foreseeing the future clearly (Bigo, 2006). In another simplified example, a set of data might indicate for the past five years, migration has increased two-fold at a particular checkpoint; this data alone might urge decision-makers to allocate additional financial and personnel resources to assist. However, as witnessed primely in the first six months of the COVID-19 pandemic's ramifications on halts in mobility and restrictions on movement, the present was not characteristic of the past; further, it might be dangerous then to assume because of those drops in border traffic, a trend of lower mobility will continue, and thus those financial and/or personnel resources are no longer necessary.

This last example highlights again not only the challenges where Big Data lacks contextual insight, but also where there is undue reliance on the past as foretelling for the future. Thus, Big Data and AI-algorithms are invariant to permutations to ground truths. Three salient assumptions further underscore these tools and their use for predictive purposes; first, that it is possible to use technologies to predict crime (van Brakel and De Hert, 2011). Second, by anticipating "the likelihood of an individual with a particular profile experiencing a negative outcome, interventions targeting specific issues can be put in place ahead of time" (Ting, Chu, Zheng, and Chng, 2018).

Lastly, that terrorism, crime, and other security concerns can be reduced by intervening, screening, or barring individuals from mobility opportunities, opposed to resolving wider social conditions or environments that might enable or foster such behaviour. A principle borrowed from risk management underscores this final assumption: the system is only as strong as its weakest link. Given enough time, resources, and motivation or intent, an adversary will find gaps or loopholes in security in order to find a way in. For border security, this implies that while strict rules and procedures, state-of-the-art technologies and systems, and severe legal consequences might deter a degree of terrorist or trans-border criminal activity, striving for total elimination of security threats is futile.

A State of Emergency

Alluded to and illustrated throughout some arguments thus far, the COVID-19 pandemic serves as a notable case study. The pandemic exposed several of the aforesaid security concepts and impediments to data-driven intelligence for border management. At the time of this writing, it has been approximately one year since the first virus cases appeared; the below documents some of the challenges of the pandemic and actions taken in light of it from the start, in February-March 2020 through to the present moment. It is thus written as dually a reflection and ongoing account.

While the pandemic has ramifications around the world with many rippling effects across borders, the impacts are not equal; populations were disproportionately affected in innumerable, different ways. As this paper focuses on issues of border control, it looks at some of the consequences to relevant parties. In the wake of security threats, such as terrorist attacks, epidemics and pandemics, or other extreme circumstances, one line of thought contends "The first move of any government that considers its survival threatened is to close its borders and detain foreigners. This is not new" (Bigo, 2006).

While the unprecedented circumstances posed by the rapidly spreading COVID-19 virus did not necessarily include detainment at the onset, some foreigners faced increased social prejudice and stigmatisation, alienation from within the region of their location, and in some instances, abandonment by their own national governments. The costs, financial and otherwise, to travel home soared exponentially; altogether, exorbitant transportation prices, increasing demands for a seat, dwindling numbers of transportation options available, and health and safety risks made (and continue to make) it difficult for some to return to a country of permitted residence, while for others, near, or entirely, impossible.

Following the systematic and prolonged closures of borders, foreign consulates, embassies, and citizen services, and interior ministries and agencies, some TCNs in particular found themselves with expiring visa-free regimes, visas, and permits. While in the early months some countries offered options for extending the right of stay

due to the extenuating circumstances, for some TCNs these communications were unknown or unclear, or governed by additional sets of rules, procedures, and exceptions of which they were not well-informed. Some applicants with pending applications for international protection or asylum suddenly found themselves stuck in dangerous situations and unable to move to safer environments. When attempting to travel home or transit to other countries where they were permitted legal stay, some faced discrimination at checkpoints or refusals for entry. Real-world manifestations of "social sorting" and the Ban-opticon were apparent in the categorisation of individuals based on factors such as nationality, citizenship, occupation, and purpose of stay, coupled with blanketed bans of certain populations and increased "monitoring and tracking of individuals or groups" (Galič, Timan, and Koops, 2017). In addition to these difficulties, there are increasing challenges for the current and future database systems, border control authorities, law enforcement, and policymakers. Current datasets are reliant on historical (i.e. pre-pandemic) data and lack the contextual affordances of the exceptions and changes to areas of migration, law, criminal justice, and security. In this sense, it can be put forth the explanatory value of data and data-driven intelligence that includes (or does not include) human intelligence, subjectivity, and judgement must be duly taken into consideration in current and future action and policy.

Concluding Remarks

This paper has attempted to discuss some of the underlying challenges inherent to Big Data, data-driven intelligence, and artificial intelligence employed in the context of border security and intelligence. In doing so, it aimed to raise awareness of the limitations of data bias and (un)fairness, a lack of context, and dependency on the past for future prediction and action. It is pertinent to also account for potential consequences; indeed, these might be irreconcilable when used to inform decision makers on matters that have profound ramifications for individual human lives and the whole-of-society. This is not to say that there are no benefits to Big Data and its use in intelligence applications. On the contrary, there are numerable benefits that can be realised by

employing Big Data and AI on the road toward improving intelligence capabilities and organisations. These benefits can also be reaped and enjoyed by external stakeholders, including travellers, border guards, law enforcement authorities, consular officials, and strategic organisational decision makers and policymakers.

However, in the transformation towards data-driven intelligence, it is in society's best interest to be mindful of pitfalls in data and technology, which are often not afforded careful thought and well-rounded critique. While seen to be increasingly removed in current times, it is important not to eclipse the role of human intelligence in many of the aforementioned applications. Indeed, decision-making, especially when the stakes are high for individual lives and internal and external society, remains a human endeavour. The human-techno nexus should be a relationship where Big Data, AI, and other technologies perform supporting functionalities rather than a driving role. As well summarised, "The real world can be complex and many situations need careful consideration and the weighing of possibilities by human beings because 'the nature of service provision calls for human judgement that cannot be programmed and for which machines cannot substitute" (Dekkers, van der Woude, and Koulish, 2019).

References:

1. Bellanova, R. and Glouftsios, G. (2020). Controlling the Schengen Information System (SIS II): The Infrastructural Politics of Fragility and Maintenance. *Geopolitics*. https://doi.org/10.1080/14650045.2020.1830765

2. Bennett Moses, L. and Chan, J. (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society,* 28(7), pp. 806-822. https://10.1080/10439463.2016.1253695

3. Bigo, D. (2006). Security, exception, ban and surveillance, In D. Lyon (Ed.), *Theorising surveillance: The panopticon and beyond.* Willan Publishing, pp. 46-69.

4. Bollier, D. (n.d.) *The Promise and Peril of Big Data.* The Aspen Institute. Accessed 8 January 2021 at: https://www.aspeninstitute.org/publications/promise-peril-big-data/

5. Van Brakel, R. and De Hert, P. (2011). Policing, surveillance and law in a pre-crime society: understanding the consequences of technology based strategies. *Journal of Police Studies*, 20(3), pp. 163-192.

6. Broeders, D. and Hampshire, J. (2013). Dreaming of Seamless Borders: ICTs and the Pre-Emptive Governance of Mobility in Europe. *Journal of Ethnic and Migration Studies*, 39(8), pp. 1201-1218. https://10.1080/1369183X.2013.787512

7. Bux, U. (n.d.). *Management of External Borders.* [Factsheet]. European Parliament. https://www.europarl.europa.eu/ftu/pdf/en/FTU_4.2.4.pdf

8. Clarke, R. A. (1987). Information technology and dataveillance. *Communications of the ACM*. https://doi.org/10.1145/42411.42413

9. Cole, S. A. (2002). *Suspect Identities: A History of Fingerprinting and Criminal Identification,* Harvard University Press.

10. Directorate-General for Communication. (n.d.). *The EU explained: Borders and security.* [Brochure]. https://ec.europa.eu/home-affairs/sites/homeaffairs/files/e-library/docs/brochure-borders-and-security/brochure_borders_and_security_en.pdf

11. Dekkers, T., van der Woude, M., and Koulish, R. (2019). Objectivity and accountability in migration control using risk assessment tools. *European Journal of Criminology*, 16(2). pp. 237-254. https://10.1177/1477370818771831

12. Dressel, J. and Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances,* 4, p. 1-5.

13. Dunlap, C. J. Jr. (2018). The Hyper-Personalization of War: Cyber, Big Data, and the Changing Face of Conflict. *Georgetown Journal of International Affairs*, pp. 108-118.

14. Duwe, G. and Kim, K. (2016). Sacrificing Accuracy for Transparency in Recidivism Risk Assessment: The Impact of Classification Method on Predictive Performance. *Corrections* 1(3), pp. 155-176. https://10.1080/23774657.2016.1178083

15. ETICAS Foundation. *Big Data at the Border. ETICAS Foundation.* https://eticasfoundation.org/migration/big-data-border-report/

16. European Commission. (n.d.). EU Information Systems: Security and Borders. [Factsheet]. https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/policies/european-agenda-security/20190416_agenda_security-factsheet-eu-information-systems-security-borders_en.pdf

17. European Commission for Migration and Home Affairs. (n.d.)a. "European Travel Information and Authorisation System (ETIAS), European

Commission, accessed 2 January 2021 at: https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias_en

18. European Commission for Migration and Home Affairs. (n.d.)b. "Commission Staff Working Document: Executive Summary of the Impact Assessment. Accompanying the document Proposal for a regulation of the European Parliament and of the Council establishing an Entry/Exit System (EES) to register entry and exit data and refusal of entry data of third country nationals crossing the external borders of the Member States of the European Union and determining the conditions for access to the EES for law enforcement purposes and amending Regulation (EC) No 767/2008 and Regulation (EU) No 1077/2011 and Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) 2016/xxx as regards the use of the Entry/Exit System (EES)," European Commission, accessed 2 January 2021 at: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/policies/securing-eu-borders/legal-documents/docs/20160406/smart_borders_package_-_20160406_-_impact_assessment_-_summary_en.pdf

19. European Court of Human Rights (ECHR) and Council of Europe (CoE). (2012). "Charter of Fundamental Human Rights of the European Union,' European Convention on Human Rights, as amended by Protocols Nos. 11 and 14 and as supplemented by Protocols Nos. 1, 4, 6, 7, 12, 13 and 16," accessed 12 January 2012 at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

20. European Parliament. (n.d.). "Schengen: a guide to the European border-free zone," accessed 1 January 2021 at: https://www.europarl.europa.eu/news/en/headlines/security/20190612STO54307/schengen-a-guide-to-the-european-border-free-zone

21. Galič, M., Timan, T., and Koops, B. (2017). Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation. *Philos. Technol.*, 30, pp. 9-37, https://doi.org/10.1007/s13347-016-0219-1

22. Goodfellow, I., McDaniel, P., and Papernot, N. (2018). Making Machine Learning Robust Against Adversarial Inputs. *Communications of the ACM,* 61(7). https://10.1145/3134599

23. Green, B. and Hu, L. (2018). *The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning* [Conference Presentation]. Machine Learning: The Debates Workshop at the 35th International Conference on Machine Learning, Stockholm, Sweden.

24. Haggerty, K. (2006). Tear down the walls: on demolishing the panopticon, In D. Lyon (Ed.), *Theorising surveillance: The panopticon and beyond*. Willan Publishing, pp. 23-45.

25. Haggerty, K. and Ericson, R. V. (2000) The surveillant assemblage. *British Journal of Sociology*, 51(4), pp. 605–622

26. Jarvis, L. and Lister, M. (2013). Vernacular Securities and their Study: A Qualitative Analysis and Research Agenda. *International Relations,* 27(2), https://doi.org/10.1177/0047117812460880

27. Kemshall, H. (2011). Crime and risk: Contested territory for risk theorising. *International Journal of Law, Crime and Justice* 39, pp. 2018-229. https://10.1016/j.ijlcj.2011.05.009

28. König, M. (2016). The borders, they are a-changin'! The emergence of socio-digital borders in the EU. *Internet Policy Review*, 5(1), https://doi.org/10.14763/2016.1.403

29. Lyon, D. (2003). *Surveillance as Social Sorting: Privacy, risk, and digital discrimination.* Routledge.

30. Mehta, H., Shah, S., Patel, N., and Kanani, P. (n.d.). Classification of Criminal Recidivism Using Machine Learning Techniques. *International Journal of Advanced Science and Technology,* 29(4), pp. 5110 – 5122

31. Paul, R. (2018). Risk Analysis as a Governance Tool in European Border Control. In A. Weiner, S. Bonjour, and L. Zhyznomirska (Eds.), *Handbook on the Politics of Migration in Europe.* Routledge, pp. 227-239.

32. Regulation (EC) No 603/2013. On the establishment of ·Eurodac· for the comparison of fingerprints for the effective application of Regulation (EU) No 604/2013 establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person and on requests for the comparison with Eurodac data by Member States' law enforcement authorities and Europol for law enforcement purposes, and amending Regulation (EU) No 1077/2011 establishing a European Agency for the operational management of large-scale IT systems in the area of freedom, security and justice (recast), European Parliament, Council of the European Union, accessed 4 January 2021 at: http://data.europa.eu/eli/reg/2013/603/oj

33. Regulation (EC) No 767/2008. Concerning the Visa Information System (VIS) and the exchange of data between Member States on short-stay visas (VIS Regulation), European Parliament, Council of the European Union, accessed 4 January 2021 at: http://data.europa.eu/eli/reg/2008/767/oj

34. Regulation (EC) No 1987/2006. On the establishment, operation and use of the second generation Schengen Information System (SIS II), European Parliament, Council of the European Union, accessed 4 January 2021 at: http://data.europa.eu/eli/reg/2006/1987/oj

35. Regulation (EC) No. 2017/2226. On establishing an Entry/Exit System (EES) to register entry and exit data and refusal of entry data of third-country nationals crossing the external borders of the Member States and determining the conditions for access to the EES for law enforcement purposes, and amending the Convention implementing the Schengen Agreement and Regulations (EC) No 767/2008 and (EU) No 1077/2011 European Parliament, Council of the European Union, accessed 4 January 2021 at: http://data.europa.eu/eli/reg/2017/2226/oj

36. Regulation (EU) 2018/1240 of the European Parliament and of the Council of 12 September 2018 establishing a European Travel Information and Authorisation System (ETIAS) and amending Regulations (EU) No 1077/2011, (EU) No 515/2014, (EU) 2016/399, (EU) 2016/1624 and (EU) 2017/2226. European Parliament, Council of the European Union, accessed 4 January 2021 at: http://data.europa.eu/eli/reg/2018/1240/oj

37. Sylvester, J. and Raff, E. (2018). What About Applied Fairness? *Machine Learning: The Debates (ML-D)*. arXiv:1806.05250

38. Ting, M. H., Chu, C. M., Zheng, G., Li, D., and Chng, G. S. (2018). Predicting recidivism among youth offenders: Augmenting professional judgement with machine learning algorithms. *Journal of Social Work*, 18(6), pp. 631-649, https://10.1177/1468017317743137

39. Waldorp, M. M. News Feature: What are the limits of deep learning? *Proceedings of the National Academy of Sciences,* 116(4), https:// 10.1073/pnas.1821594116

40. Weber, C. (2011). 'I am an American': Filming the Fear of Difference. Chicago University Press.

41. "The Schengen Agreement – History and the Definition," Schengen Visa Info, https://www.schengenvisainfo.com/schengen-agreement/