

Demystifying the Algorithmic Drama

Mitch Davies

Sociology, Victoria University of Wellington,

Wellington 6012, New Zealand

daviesmitc@myvuw.ac.nz

Mike Lloyd

Sociology, Victoria University of Wellington,

Wellington 6012, New Zealand

mike.lloyd@vuw.ac.nz

Daniel Neyland: *The Everyday Life of an Algorithm*. Cham 2019: Palgrave Pivot; 151 pp.

At its simplest, an algorithm is a sequence of steps that need to be followed to accomplish a task. This book contributes to the emerging field of critical data and algorithm studies, where the word 'critical' reflects the argument that via their increasing centrality in contemporary computing technology algorithms have an ability to directly spread values and affect culture. D. Neyland calls concern over algorithmic power the 'algorithmic drama', that is, the standard story that algorithms have a powerful ability to make decisions over our futures, with most people having little understanding of or control over this process. For this reason alone, the book is a welcome addition to the literature as it employs an ethnographic approach that helps demystify the opacity of algorithms.

It draws on his experiences working with various stakeholder groups during the development of an automated video security alert system focussed upon

abandoned luggage left anywhere in a transport hub (the alert system also focuses on persons who enter forbidden areas). Thus, the social context of the algorithm's development is concern over terrorism, but an interesting feature of the project under D. Neyland's ethnographic scrutiny is that it is intended to *reduce* the amount of visual video data seen by operatives in train stations and airports. The technical details of exactly how it does this are necessarily simplified in the book. D. Neyland does not claim to have expertise as a computer engineer or software designer, instead he adopts a relatively standard social science mode of doing fieldwork amongst the project's management, investors, software development team, ethics committee and end-users. This data provides input for his subsequent discussions about how the algorithm fitted in with the progress of the project, and how the algorithm was variously implemented to compose its own model of life as it unfolded in the transport hubs intended to be under its surveillance.

The introductory chapter discusses recent literature on algorithms noting concerns about the amount of power given to algorithmic systems. D. Neyland has three key research questions: How do algorithms participate in the everyday? How do algorithms compose the everyday? And how does the algorithmic become the everyday? Algorithm based systems increasingly participate in our lives, but we know little about how algorithms are produced, what they consist of, how they change, and the ways that everyday life and algorithms combine into arrangements of effects. Chapter 1 provides the reader with the initial pseudo-code for the abandoned luggage algorithm. The pseudo-code is given as a set of ordered step-by-step instructions that is easy to understand. The implementation of this algorithm into executable and accurate software by the project team observed by D. Neyland is the principle activity recorded and used for the basis of discussion in the book.

Chapter 2 describes the experimentation process that computer scientists undertake when developing an algorithm from descriptive pseudo-code into a computer programme that works correctly. Based on real-time analysis of digital video, the algorithm has to classify the movement of human-shaped and luggage-shaped objects to determine if luggage has been left unattended. The computer scientists experiment with their coding in order to find a way of doing the required

tasks with ‘elegance’. D. Neyland notes that recording this experimental process shows how algorithms may be opened up to scrutiny, revealing their instability and the frequent changes that are made to them during development phases. It is here that he first discusses how pre-existing algorithms (for object classification) were to be ‘dropped into’ the system, and that this was an overly optimistic part of the implementation plan. He also describes sitting-in on the project team’s meetings: how the computer scientists presented and discussed system architecture, technical and other problems during these meetings. Some example images are given that indicate the bounding-box techniques used to classify human and luggage shaped objects. Chapter 2 concludes with the observation that if algorithms eventually participate in everyday life then a great deal of experimental and technical work is black-boxed in the process.

Having set the scene of the project’s algorithmic development, Chapter 3 discusses a science and technology studies (STS) perspective on algorithmic accountability. From this perspective, the network of relations within which the algorithm is connected is the basis for its accountability in action. Here D. Neyland is suggesting that algorithmic accountability based solely on the transparency of its functional operation is pointless – an algorithm must be accounted for in its entire systemic setting. A problem in the form of accountability of the algorithm is described, as the algorithm provides a text based list of alerts to human surveillance operators, all of whom are trained to detect abandoned luggage by watching a continuous stream of video. This problem required the development team to back-track their algorithm design in order to provide video for the route reconstruction of an item of luggage tagged as abandoned, so that the surveillance operators were able to verify the alert. This meant that the system had to store data, and develop new data retention and deletion rules, over and above what had been anticipated. These system breaches of initial design principles were reported to the project’s ethics board by D. Neyland (who was responsible for assessing the ethics of the emerging system). He describes the constitution and functioning of the ethics board and notes that they drew attention to matters of concern that were taken back to the project team. This reporting accountability of the system intersected with the sense of

overall system accountability pursued by the ethics board (with respect to minimising the storage of video footage on privacy considerations). D. Neyland suggests that this dialogue between the different registers of algorithmic accountability is a productive method of engagement with calls for scrutiny and transparency of algorithms.

Following on from these considerations of accountability, Chapter 4 deals with how data was to be deleted by the algorithm. As the project was funded in order to create a system that would minimise video data used and stored for surveillance purposes, the maximal deletion of data was a key ethical concern for the project. As already noted in Chapter 3, the end user dissatisfaction with text-based alerts meant that more data than had been anticipated had to be provided for viewing by the operators, and stored by the system. It turned out that it was a struggle to delete the 'vast majority of data deemed irrelevant' (p. 74). Inadequate methods of deletion were tried that left some (or all) of the data retrievable by forensic processes. This meant that the development team had to implement a 'deleting machine' that expunged the unwanted data from the system, and provided an account of having done so. With the recent introduction of data privacy laws by the European Union, the project coordinators thought that development of the deletion machine may have market value in its own right. For the algorithm itself, the inclusion of deletion capabilities meant that the system could calculate from data representing the complexities of everyday life, classifying what is relevant from what is irrelevant. D. Neyland discusses this in terms of a 'calculative agency' that is capable of imposing a 'hierarchy of relevance on everyday life' (p. 78). A lengthy discussion of concepts relating to 'zero' and the production of nothing is included in the chapter. Pragmatically though, the development team tried to produce a deletion system where data is overwritten and an audit log of data that has been verifiably deleted is produced. However, this deletion system proved to be a failure, with 'orphan frames' and metadata being missed by the overwriting function and remaining readable after the blocks of data they were in had supposedly been overwritten. Hence, the deleting machine failed to produce an 'accountable nothing' as was intended.

With the project lingering in an unsuccessful state, Chapter 5 describes how the project team nonetheless developed different demonstrations of the system for various audiences, and this is where things get more sociologically complex. Demonstrations were created for project funders, academics, the ethics committee and end-users based on what the project team imagined their demands to be. The same system lent itself to the creation of various presentations that were discursively assembled and held varying degrees of integrity. The reason for this varying lack of integrity was that the system had failed to grasp the everyday, in the sense of accurately distinguishing unattended luggage and people out of place, therefore promises made to the range of stakeholders were, more or less, not able to be kept. Earlier demonstrations were reconfigured for this later phase to display features that could *possibly* be made to work, rather than predict complete project success. The author discusses STS literature on the integrity of technology demonstrations, noting that the audience may be willing to suspend disbelief in the artifice of the presentation if it is able to present a similarity between the current system and a future reality. Nevertheless, the question of integrity remained as the project team prepared presentations based on recorded footage of system responses under ideal conditions. However, the demonstration of the system to end-users came unstuck when an attempt to show real-time results in a six hour period resulted in 2654 detections of abandoned luggage, as opposed to 6 detections by the conventional system (2648 incorrect classifications by the new algorithmic system). The book shows several images where incorrect classifications of people and luggage have occurred, due to reflections and other video data artefacts that the system did not process correctly. Because of this glaring failure, the demonstration prepared for project founders (after an awkward team meeting) was concocted around idealised video footage of a detection of an abandoned luggage item in a railway station. D. Neyland expresses surprise at the level of concealment and careful scripting that went into the preparation of system demonstrations, discussing them theoretically in comparison to fake artwork.

This 'surprise' is elaborated in the final chapter which details how, despite the failure of the algorithmic system, considerable effort went into building a market

value of the algorithmic system. Technological inadequacies of the project were excluded in what is called the 'Exploitation Report', and 'imaginative and dexterous calculative work' (p. 135) went into trying to enrol investors. D. Neyland describes this in an understated manner, one consequence of which is that the reader may infer something of a blatant 'con' to be in operations, that is, the system demonstrably failed, but this did not stop marketing activities designed to bring in investors' money.

We were expecting more commentary on this aspect. We were left wondering whether D. Neyland's conclusion and its tone of surprise may partly reflect methodological limitations of his research. This is a common problem in sociological studies of STS: given the technical and expert knowledge of scientists and technologists, what exactly can ethnographers achieve by simply following people around? Although D. Neyland had ongoing insider access to the activities of the software development team, he was not one of 'them' because he has no background in computer science and software engineering. Ironically, with a similarity to the way potential investors were deceived as to the efficacy of the algorithm, he may have misunderstood the technical work of the software team, and he does not seem to reflect on this possibility in the book. Computer scientists can (and do) display an attitude of technical imperialism when interacting with their less technically knowledgeable observers, and maybe this is what D. Neyland was being dealt by the development team. The description of the project as 'built on a decade of research' (p. 124) could be a feint by the software team to D. Neyland. He faces technical difficulties in uncovering these sacred sub-parts that coincidentally form the fundamental image processing sub-system of the overall system - some technicalities remain beyond his investigative knowledge. Unfortunately, if it is just these parts that make the system fail, perhaps it is the case that the software developers could salvage their algorithmic system despite the obvious failure, thus making the pitch to investors less 'imaginative'? The author has witnessed the attachment of 'easy' outer-parts that serve the stream of video data to the critical core processing sub-algorithms (that do not work). Ultimately, the non-functionality of the processing core remains opaque to both D. Neyland and the reader, and this

maintenance of opacity by the computer scientists and his acceptance of it is an interesting phenomenon that could do with more discussion. Nevertheless, even given some repetitive elements, the entire story of the algorithm is well worth reading, at the very least showing that system development is an awkward business where it cannot simply be assumed our data and lives will be controlled from without by computing technologies. Given the role of algorithms in the development of new technology – autonomous vehicles for example – this kind of obvious failure needs documenting, and D. Neyland’s presentation of this case makes an important contribution to this field.

Wpłynęło/received 30.07.2019; poprawiono/revised 28.08.2019