Surveiller et contrôler. Towards a unified theory of AI-powered surveillance.

Final Year Project by Sebastian Dodt

Supervisor: Prof. Arshin Adib-Moghaddam

659605

21 May 2021

Word count: 9,931

Abstract

The concept of surveillance is central to the analysis of technological progress and its effects on our society. Yet the field of surveillance studies currently consists of inductive theories that apply only to a narrow subgroup of surveillance practices. When new technologies arise, contemporary surveillance theory is frequently unable to explain novel phenomena and power dynamics. I argue that surveillance scholars should return to the classical theories of Foucault and Deleuze, which are more suited to account for the changing nature of surveillance practices. Grounded in their work from the pre-digital era, I produce a taxonomy of surveillance that can account for current and future forms of surveillance. Its applicability is demonstrated in the example of recent artificial intelligence technologies that are reshaping surveillance. The facial-recognition tool by the start-up Clearview AI serves as a case study to illustrate these transformations and concludes my paper.

Table of Contents

1. Introduction 4

2. Literature Review 5

2.1. Panopticism 5

2.2. Control societies 7

2.3. Contemporary surveillance theory 8

2.4. Downfalls of modern surveillance theory 10

2.5. Artificial intelligence and machine learning 10

3. Theory of "AI-powered" surveillance 15

3.1. How machine learning reshapes surveillance 15

3.2. Return to Foucault and Deleuze 17

3.3. Taxonomy of surveillance 19

3.4. Hypotheses 23

4. Case study: Clearview AI 24

4.1. Facial Recognition and Machine Learning 24

4.2. Use and justification 26

4.3. Criticism 27

5. Conclusion 30

6. Bibliography 31

1. Introduction

Artificial intelligence-powered surveillance technology is proliferating faster and in more places than experts and the public have recognised (Feldstein 2019). Deployed in at least seventy-five countries, the implications of its use remain understudied in surveillance theory, which lacks a sustainable theoretical foundation (Powers and Ganascia 2020, 38). Pre-digital surveillance theories have frequently been discarded, while novel conceptualisations have attempted to account for the transformations of surveillance caused by technological innovation. This has resulted in a state of disarray in the field of surveillance theory. The successive proposals of new frameworks and theories, based on recent technological developments at a given time, have led to surveillance theory being highly technology-dependent. This has left social scientists ill-prepared for a leap in surveillance capabilities achieved by artificial intelligence (AI). Given the increasing pace of innovation in AI and beyond, surveillance theory requires a solid theoretical foundation for the analysis and critique of empirical phenomena. By returning to the classical theories by Michel Foucault (1977) and Gilles Deleuze (1995), this paper proposes a new theoretical basis from which analyses and critiques of current and future surveillance practices can be developed. I argue that Foucault's concept of *panopticism* and Deleuze's *society of control* provide a meaningful and sustainable theoretical basis for the study of surveillance in the age of AI. While past research has treated them as opposing concepts, AI-powered surveillance unifies the individualism of panopticism with the de-humanisation and diversification of surveillance in control societies.

This paper begins with a review of existing surveillance theories from the classical work of Bentham (1838), Foucault (1977), and Deleuze (1995) to contemporary publications by Haggerty (2000), Lyon (2014), and Zuboff (2019). The dependency of contemporary surveillance theories on transient technologies will be demonstrated by their inability to account for recent advancements in AI surveillance. This will be countered with my own theory that anchors in the classical theories of the 20th century to provide a stable and all-encompassing framework. In my taxonomy of surveillance, I will differentiate between architectural surveillance that runs hierarchically through society and infrastructural surveillance that contains a multi-centric field of actors. On a

second axis, I will differentiate between the awareness of the surveilled, which is crucial to understand the power dynamics of the respective society. From this theory, I will formulate five hypotheses that will be tested against a brief case study of a new facial-recognition technology called "Clearview AI". Analysing existing phenomena as well as potential implications of a wider dissemination of Clearview AI, I will demonstrate the applicability of my unified theory of panopticism and control society.

This paper will use a theory-based approach to analyse and critique the works of surveillance theorists. In my case study, I will use secondary research from social science and data science with attention to technical detail. This interdisciplinary perspective provides a foundation that is based on the actual inner processes of artificial intelligence and facial recognition. By distinguishing hype from reality, my analysis will offer a fruitful foundation for our understanding of surveillance grounded in technological fact.

2. Literature Review

Surveillance has been conceptualised in various ways in academic literature. For simplicity, this paper will adopt David Lyon's (2007, 13–16) definition as "any systematic, routine, and focused attention to personal details for a given purpose", which can *inter alia* be management, influence, or entitlement. Its practice has undergone alterations and transformations, but, as I will argue, historical approaches to the study of surveillance continue to represent relevant theoretical lenses. In the following, I will delineate two approaches in particular: *panopticism* by Foucault and *societies of control* by Deleuze.

2.1. Panopticism

In 1975, Michel Foucault published his work "Surveiller et punir"—commonly translated as "Discipline and Punish"—which offered a *sui generis* perspective on the social and theoretical mechanisms that guided the transformation of the Western penal system in the modern age. Generalising Jeremy Bentham's (1838) imaginary "panopticon", an architecturally optimised prison that enables the permanent monitoring of its inmates, Foucault applied these social dynamics to other institutions and society as a whole and termed this form of surveillance "panopticism".

To illustrate panopticism, Foucault describes the containment strategies by cities during the plague in the late-17th century. When cases of the 'Black Death' appeared in a town, people were obliged to stay in their houses. Every street was constantly under surveillance from syndics who locked the doors of all houses from the outside (Foucault 1977, 195).

> "This enclosed, segmented space, observed at every point, in which the individuals are inserted in a fixed place, in which the slightest movements are supervised, in which all events are recorded, in which an uninterrupted work of writing links the centre and periphery [...], in which each individual is constantly located, examined and distributed among the living beings, the sick and the dead—all this constitutes a compact model of the disciplinary mechanism." (Foucault 1977, 197)

Already before the COVID-19 pandemic, similarities between the methods of surveillance of the syndic—perpetual observation, recording of movements, location tracking, uninterrupted flows of data to a central place, classification of the individuals—and current digital surveillance methods seem hard to dismiss, and I will return to this later. In Bentham's prison and the plague-stricken town alike, surveillance is primarily characterised by the individual's belief in constant and invariable supervision. This belief makes the individual an active, yet involuntary, participant in its own surveillance as it conforms its behaviour to the expectations of the prison guard or syndic, expecting to be watched and punished for any missteps (Galič, Timan, and Koops 2017, 12). This internalisation of power is the source of the panopticon's effectiveness as a tool for discipline, as individuals do not have to be surveilled at all times. There merely needs to be a prominent system through which they *could be* surveilled at any time, without them knowing exactly when. "This invisibility [of the guard] is a guarantee of order" (Foucault 1977, 200).

Consequently, the "disciplinary society [… is] an indefinitely generalizable mechanism of panopticism" (Foucault 1977, 216), i.e., a translation of Bentham's prison to any society characterised by prominent, but invisible, hierarchical surveillance. The internalisation of surveillance causes individuals living or working in the panoptically-arranged spaces to become "docile bodies" (Spaulding 2020, 392).

2.2. Control societies

With the onset of new information and communication technologies, Foucault noticed novel methods and behaviours of institutions and societies for which panopticism could not account. This realisation provided the basis for studies beyond disciplinary societies in surveillance theory (Galič, Timan, and Koops 2017, 18). After Foucault's death, this work was developed further by Deleuze, partly in collaboration with Guattari, who termed their post-panoptic theory "society of control" (Deleuze and Guattari 1987), a term that had already been proposed by Burrough—though with slightly different connotations (Moore 2007).

Foucault's disciplinary society consisted of "hierarchy, surveillance, observation, [and] writing" (Foucault 1977, 198). While the latter three still apply to Deleuze's control society, hierarchy has been replaced with a multi-centric surveillance network in which various actors with different motives and motivations collect information of interest to them. Among them are companies aiming to control profits, as well as employees and customers—this represents a notable shift towards an increasing marketisation of surveillance that contrasts starkly with Foucault's focus on formal institutions (Deleuze 1995, 6; Foucault 1977, 183). The rise of large corporations has led to a new system of domination where power does not lie in said formal institutions, but in "ad hoc and informal networks" subject to internal decision processes (Deleuze 1995, 7; Galič, Timan, and Koops 2017, 18). The surveillance practices of these networks, which can be companies and other organisations, are invisible to the individual, resulting in a more distant and abstract form of surveillance.

The distance between the subject and object of surveillance is further increased by introducing digital technologies that target individuals through *representations* defined as numerical abstractions of human attributes. The representation, or 'data body', of an individual is more important to the surveilling actor than the physical being it represents, the 'real body'. Galič et al. (2017, 18) call this a *de-humanisation* of surveillance accompanied by a *de-individualisation* as surveillance moves its focus away from the individual to the mass from which it seeks some gain in profit or power. In contrast to Foucault, the individual is unimportant as surveillance aims to record and control the aggregate behaviour of a larger group.

Writing in a time of digital transformation as computers were becoming more widely used and the internet accessible for the public, Deleuze was able to account more directly for the changes that resulted from these technologies. Crucially, although Deleuze developed his theory years before digital surveillance as we know it today came into existence, it remains generally applicable to digital and physical surveillance alike, which demonstrates the author's remarkable foresight.

2.3. Contemporary surveillance theory

The use of the internet for surveillance purposes and political changes in response to the September 11 attacks prompted a proliferation of new surveillance theories. Some attempted to develop and adapt Foucault's panopticon and were given names that hinted at the origin of their theoretical foundation, e.g., "panopticommodity" by Lyon (2007), "participatory panopticon" by Whitaker (1999), "oligopticon" by Latour (in Kitchin and Dodge 2011, 85), and "BANopticon" by Bigo (2006). However, the surveillance theory community at large has moved away from Foucault's panopticism as it appeared less relevant in current affairs (e.g., Haggerty 2006; Murakami Wood 2007; Lyon 2008).

Deleuze's society of control has been drawn upon more heavily in contemporary surveillance theory. Using his concept of *assemblages*, a form of territorialism and ordering of bodies, Haggerty and Ericson (2000) proposed *surveillant assemblages* as a theoretical understanding of surveillance. They function as 'recording machines' of data flows and their storage for future analysis (Galič, Timan, and Koops 2017, 21). By unifying different elements of an object (e.g., a consumer and their attributes) and arranging them into a numerical structure, the concept of assemblage primarily attempts to account for the large databases that emerged in the industry in the late-1990s. Writing before tech companies such as Google had oriented their business model towards data accumulation, Haggerty and Ericson accurately foresaw a capitalist imperative to capture all data possible for potential future profit-making (Clegg 2017, 68). Focusing on market actors and their aims to govern the actions of consumers, Haggerty and Ericson attribute new purposes to surveillance, in particular, the monitoring of consumptions patterns and controlling of access to places and perks.

Later, Haggerty (2006) acknowledged that the purposes of surveillance have diversified and extended. While surveillance assemblage has lost some of its prominence in the field of surveillance theory, it aligns in part with other surveillance theories, e.g., *dataveillance* by Clarke (1988). Dataveillance was a response to new electronic forms of data storing and processing, particularly cross-referencing of databases.

Surveillant assemblage also contains features that Shoshana Zuboff (2019) integrated in her theory of *surveillance capitalism*. Although alluding to the "mutation of capitalism" that Deleuze (1995, 6) expected, surveillance capitalism can neither be placed into the Foucauldian nor the Deleuzian tradition. Instead, Zuboff (2019) offers her post-panoptic, neo-Marxist, 'all-encompassing' surveillance theory that proclaims the dawn of a radically new economic order. In this new system, the raw materials to be mined are human experiences: surveillance capitalists collect large quantities of behavioural data, which are subsequently analysed to predict internet users' future behaviour. These behavioural predictions are sold as products on "behavioural futures markets" (Zuboff 2019, 8), enabling companies to target a subset of users that are most likely to respond to their offers. Zuboff's theory can be seen as a response to the emergence of 'big data analytics' and 'machine intelligence'. Big data refers to datasets that were thought to be too big to be analysed, yet modern processing units and innovative machine learning algorithms allow companies to learn consumers' pattern of behaviour and make recommendations in real-time. As a result, online experiences become increasingly personalised and customised to the users' taste, allowing ad-selling companies such as Google (since 2015: Alphabet) to earn high profits when making the right prediction about users' behaviour.

Other surveillance theories also divert from Foucault and Deleuze. *Participatory surveillance* seeks to account for the culture of voluntary and conscious watching and being watched on social media, which, according to the authors, "can empower and not necessarily violate the user" (Albrechtslund and Dubbeld 2002, 3; Albrechtslund 2008). Koskela (2002) takes this thought further and argues that online *exhibitionism* "is not a negation of privacy but an attempt to reclaim some control over the externalisation of information" (Koskela in Dholakia and Zwick 2001). Perhaps only because of the early

time of writing did Koskela not recognise that this is merely a façade of users' control over their data (Galič, Timan, and Koops 2017, 30).

2.4. Downfalls of modern surveillance theory

Philosophy of science distinguishes between inductive versus deductive reasoning. Deductive theory refers to hypothesising a set of patterns and expectations which is subsequently tested against real-world observations. In reverse, inductive theory uses empirical observations as its starting point to develop a general theory. While both come with methodological issues, deductive theory is generally accepted as superior to inductive theory as it is more likely to be generalisable, parsimonious, and logically consistent due to the relatively greater distance between theory and observation. Panopticism and societies of control fall into this category as they have unfolded their largest explanatory value years and decades after their publication and are flexible to be applied to a significant number of cases.

In contrast, modern surveillance theory is neither comprehensive nor generalisable, except for surveillance capitalism that seeks to be all-encompassing at a "civilizational scale" (Zuboff 2015, 77). However, all post-Deleuzian theories presented above, including surveillance capitalism, can be classified as inductive theories that have serious limitations when employed in surveillance studies. After all, each theory was introduced after some technological change to provide a better conceptualisation of novel mechanisms and patterns. *Surveillant assemblages* were a response to databases, *dataveillance* to cross-referencing, *participatory surveillance* and *exhibitionism* to social media, and *surveillance capitalism* to big data and 'machine intelligence'. This has caused an 'overfitting' of theory to empirical observation; in other words, theories have become highly dependent on technological trends. As a result, surveillance theory has required frequent updating and reconceptualising to account for new practices, while former theoretical concepts have lost their applicability with technological changes.

2.5. Artificial intelligence and machine learning

As this paper will show, new technologies related to artificial intelligence are emerging as prominent tools for surveillance, forcing yet another revision of surveillance theory. In order to provide a theoretical foundation that is sustainable and of long-term

applicability, I will argue that surveillance theory must return to Foucault and Deleuze, using panopticism and society of control as its anchor points. Before presenting my theoretical framework, it is integral to first define how artificial intelligence will be understood for the purpose of this paper.

Artificial intelligence has been used to describe two different technologies. The first is that of "human-level machine intelligence" (HLMI), or "human-imitative machine intelligence" (HIMI). This is not to be confused with machine learning, to which I will return later. HLMI refers to the original definition of artificial intelligence by the famous meeting of ten scholars at Dartmouth College in summer 1956, led by John McCarthy, and describes an artificial entity that *is* as smart as a human, or, due to the impossibility of objectively measuring intelligence, at least *seems* as smart (Bostrom 2014, 5). The 'Turing Test' by the English statistician Alan Turing is considered a possible benchmark to determine whether HLMI has been achieved (Pinar Saygin, Cicekli, and Akman 2000). In this test, a person has two conversations, one with a human and one with an AI. If the person is unable to determine correctly which is which, the AI passes. AI researchers have rightfully questioned whether what *seems* intelligent also *is* intelligent. Therefore, new definitions of intelligence for humans and robots have emerged, with some defining it narrowly as "possessing common sense and an effective ability to learn, reason, and plan to meet complex information-processing challenges [in many domains]" (Bostrom 2014, 3) and others referring to more comprehensive explanations, e.g., Gardner's (1983) nine types of intelligence including not only the typical logic-mathematical skills measured in IQ tests but also interpersonal skills.

Irrespective of the definition, human-like artificial intelligence, often referred to as 'strong AI' or 'Artificial General Intelligence' (AGI) in public discourse, has not yet been achieved. Since the Dartmouth seminar in the 1950s, computer scientists, statisticians, and mathematicians have failed to develop a system that matches human intelligence in more than some narrow domains. The attempts have been rich, from high-level symbol manipulation, also known as "Good Old-Fashioned Artificial Intelligence" (GOFAI), and related *expert systems*, to Artificial Neural Networks (ANN), which I will turn to shortly (Bostrom 2014, 7–8). While this research has produced valuable outcomes for other disciplines, namely statistics, HLMI is very far from being realised. As humans

learn more about the complexities of their intelligence, researchers have predicted HLMI to be achieved increasingly distant in the future. As Bostrom (2014, 4) notes, "the expected arrival date has been receding at a rate of one year per year". Followingly, researchers do not feel closer to achieving HLMI than they felt in the 1950s.

If one read recent tabloid press reporting on artificial intelligence, on "AI apocalypse" (Fish 2018), "artificial brains" (Del Bello 2018), and "artificial superintelligence" (Marri 2018), occasionally extending to renowned news sources such as The Guardian (e.g., 2020), one could think that the development of AGI is imminent. On the one hand, this might be because some companies and scientists are trying to present their discoveries in the best possible light for marketing purposes and prestige, using the term "artificial intelligence" whenever possible. On the other hand, it is often scientists who try to dampen expectations of their developed systems, while journalists are more than complicit in forging a hype, and sometimes hysteria, around artificial intelligence—this often relates to old myths about inanimate things coming to life or robots taking over the world. As a result, the public is ill-informed about the actual state of affairs in the development of AGI, leading to an overestimation of AI's abilities and disproportionate fear.

That is not to say that we should not be worried or critical about artificial intelligence at all. However, "people are afraid about the wrong things" (Lipton in Schwartz 2018), and it requires clear differentiation and terminology with respect to current and near-future technology to enable fruitful debate and policymaking around AI. One crucial distinction must be drawn between artificial intelligence, as described first, and machine learning. Machine learning can be traced back to the idea of *cybernetics* introduced by MIT professor Norbert Wiener (1961) at a similar time as McCarthy's seminar at Dartmouth College. Instead of trying to achieve human-level machine intelligence as his Dartmouth counterpart, Wiener worked on intelligent systems for pattern recognition, operations research and more. He purposely chose the term 'cybernetics' to distinguish his own research from that of McCarthy, who worked with rather different logic frameworks at the time.

After Wiener, as digital computers allowed more complex calculations, new algorithms for statistical inference and prediction were developed. These algorithms

allowed statisticians and computer scientists to use a set of 'training data' to build a model from which to 'learn' to reason or predict on new data. These approaches were generalised under the term "machine learning" and included tools from simple linear regression, invented around 1800 by Gauss (Stigler 1981), to compositional non-linear models such as Artificial Neural Networks that are at the centre of AI research today. Artificial Neural Networks' ability to recognise more complex patterns, especially in natural language and image processing, has earned it a lot of public and academic attention, which, reinforced by loose analogies to the human brain, has only advanced the myths about them being "artificial brain[s]" (Larson 2021). In practice, neural networks are merely statistical tools with a vast number of parameters requiring a large amount of training data and processing power to optimise their predictions. The theoretical and computational framework behind them, in fact, has been around since the 1940s. *Backpropagation*, a training method by which the algorithm is rewarded if it performs well and penalised if it does not, has for instance been used to control the thrusts of the Apollo spaceships. Only due to recent progress in big data collection and computational power (enabling multiple "hidden layers"), neural networks have become sufficiently accurate in predictions and profitable for private and public sector use.

To make informed criticisms of current and near-future use of neural networks and machine learning more generally, we must recognise the abilities and limitations of this technology. It is not, and, with all likelihood, will not be a form of artificial general intelligence (Chiang 2021). Machine learning rarely performs better than any human (with sufficient time) in inference and prediction tasks. Instead, machine learning's added value comes from its efficiency as it is able to complete tedious information processing tasks much quicker than humans. In cases where time efficiency is not a factor, however, for example when driving a car, machine learning performs poorly (Greene 2020).

Machine learning is often referred to as 'weak' or 'narrow artificial intelligence', in contrast to 'strong' artificial general intelligence, but both are conflated to the simple term artificial intelligence in everyday popular discourse. Despite the significant differences, McCarthy's terminology is essentially employed for technology that, albeit

more advanced, is similar to Wiener's work and falls in the tradition of statistics and pattern recognition rather than HLMI.

This terminological inaccuracy has had severe negative effects on research and discourse. Firstly, it has led to an overestimation of current machine learning capabilities, shifting the focus of researchers and politicians away from factually more pressing issues. For instance, there have been deliberations as to whether robots have rights, while the development of non-discriminatory "fair" machine learning software has not received sufficient attention (Schwartz 2018). Secondly, making the lines between weak and strong AI seem blurred (see e.g., Bostrom 2014; Chace 2016), has caused unnecessary fear of 'intelligence explosions' and 'technological singularity' (Chiang 2021). Analogous to singularity in physics, technological singularity is the theoretical threshold in technological advancements, after which reversibility becomes impossible and an artificial intelligence takes control of all matters on earth and beyond. Most AI researchers agree that Artificial Neural Networks are not a way to achieve artificial general intelligence, regardless of the available computational power (Fjelland 2020). A system of human-like intelligence would be able to learn from much less data than the amount required by artificial neural networks. Finally, this raises the question of whether artificial *intelligence* is an appropriate term to refer to machine learning. There is little intelligence to be found in Artificial Neural Networks and other machine learning algorithms as they are simply the mathematical optimisation of some parameters over a set of data. If anything, humans perceive the output as intelligent, e.g., if the algorithm defeats human chess or Go players. However, as Donoho (2019) rightfully notes, this is mere "intelligence recycling". Existing human judgements are recycled by the algorithm and reproduced on new data, for example in the form of pre-labelled training data from humans or, in cases where an algorithm supposedly "teaches itself" (Cookson 2017), in the form of ingenious coding of the respective algorithm by computer scientists.

Due to its vagueness and low likelihood of realisation in the near future (if at all), I will set aside the issue of artificial general intelligence and only focus on machine learning. In what follows, any use of the term artificial intelligence is intended to be synonymous with machine learning.

3. Theory of "AI-powered" surveillance

> "We are in a little bit of a Manhattan Project moment for A.I. and machine learning" (Kearns in Hutson 2021).
>
> "Technological progress without an equivalent progress in human institutions can doom us." (Obama 2016 in Hiroshima)

Contemporary surveillance theory has struggled to account for new technologies, and the use of new machine learning algorithms in surveillance practices all but highlights this concern. In response, I will present an alternative framework, anchored in Foucault's and Deleuze's theories, that accounts for machine learning-based surveillance.

3.1. How machine learning reshapes surveillance

The relationship between society and technology is symbiotic. Societies give rise to new technologies which represent the structures and demands of the respective society (Deleuze 1995, 6). Reversely, technological innovation can also form new societal relations (Julius in Foucault 1977, 216). Artificial general intelligence researcher Nick Bostrom (2019) has conceptualised this relationship in his *vulnerable world hypothesis* which states that humans are continually drawing balls from an 'urn of inventions' by developing new technologies. In the urn, there are white balls representing inventions that are beneficial to humankind, grey balls representing inventions that have mixed impacts, e.g., nuclear energy, and black balls representing inventions that inevitably lead to the destruction of civilisation. Humankind has not yet drawn a black ball; however, Bostrom believes that there is at least one black ball in every civilisation's urn (Bostrom 2019, 457–58). Using this analogy, we may well ask whether narrow artificial intelligence, if used for surveillance practices, can already be a black ball to our society and democracy (see e.g. O'Neil 2017). This paper does not intend to answer this question, but to lay a theoretical foundation grounded in technological evidence which could enable analyses of this sort in the future.

Machine learning affects surveillance in data processing, data analysis and accessibility. Data processing refers to the capturing and structuring of data and has seen significant advancements since the use of machine learning. This is because, firstly, machine learning can enable the processing of data larger volumes of data, known as big data. Secondly, this data can be of a wider variety, i.e., not merely text, but also audio

(see e.g., Schlossberg 2015, 21), photo and video (see case study below). Lastly, data can be processed at greater speed thanks to optimised algorithms.

Data analysis, i.e., inference and prediction, has also seen improvements through machine learning. Accuracy in particular has been improved when making inferences and predictions from data thanks to Artificial Neural Networks finding more complex patterns in data. This has increased the overall efficiency to use digital technology in surveillance. While machine learning might not be as accurate as humans in the classification of images, videos, and the resulting predictions, it is much quicker, and its error rate is decreasing continuously as algorithms become more sophisticated. Algorithms can be paired with humans to yield better results, for instance by having a software flag cases which a human examines in a second stage. This cost-efficient form of surveillance allows for an ever-greater expansion of data sources capturing users' data from many different angles. Due to the increase in volume, variety, and velocity, there is no need to aggregate or categorise groups of individuals when using machine learning. By learning from the mass, machine learning can make inferences and predictions on the level of the individual based on their data.

The proliferation of machine learning in the private sector has increased the number of actors that are able to use and profit from engaging in surveillance. Historically, formal state institutions have been the sole or primary entity to employ surveillance, and while they do still engage in it, there is now a more complex field of interacting private and public organisations.

Contemporary surveillance theory cannot account for all transformations caused by machine learning. Panopticist theories by Lyon (2007), Whitaker (1999), Latour, and Bigo (2006) presuppose hierarchical surveillance structures, usually headed by the state. The current multi-actor field has challenged this assumption. 'Surveillant assemblage(s)' excel at capturing the variety of data collected in databases (Galič, Timan, and Koops 2017, 21), yet, as Haggerty (2006) acknowledged, they were conceptualised too narrowly concerning the purpose of surveillance to be applicable today. Surveillance capitalism, which promises to be all-encompassing yet focuses narrowly on private sector surveillance and user-tailored online advertisement, is similarly inapplicable to cases outside of the scope of its theoretical foundation.

These inconsistencies are a symptom of the technology-dependency preeminent in contemporary surveillance theory. As inductive theories, they often fail to be analytically valuable after their respective technological innovation has passed or other technologies have become more determinant.

3.2. Return to Foucault and Deleuze

It is paramount that the field of surveillance studies finds a theoretical foundation from which to develop its analysis of current and future forms of surveillance, most importantly, the advent of machine learning in this field. This paper attempts to fill this gap by presenting a comprehensive framework that facilitates the analysis of any type of digital surveillance. It rests upon the deductive theories developed by Foucault and Deleuze, which have proven to be applicable long after their publication and can flexibly fit to the newest technologies while upholding the theoretical superstructure.

Foucault and Deleuze suggested that disciplinary societies and control societies exist at different moments in time. The former were located in 18th-century Europe following the *sovereign societies* of the Middle Ages and existed until the early 20th century. The latter emerged in the West after World War II. However, as it has been shown in the literature review, scholars have successfully applied either theory to current phenomena. In what follows, I therefore employ features of both theories in my framework for contemporary analysis.

Combining Foucault's and Deleuze's work helps us to account for the three core features of machine learning-driven surveillance: abstraction, individualisation, and behavioural modification. First, in contemporary surveillance practice, information about the individual is generally obtained and stored digitally. The surveilling actor does not see the physical being as in Bentham's and Foucault's panopticon, but only an abstraction—a "representation" (Deleuze 1995) or "data double" (Haggerty and Ericson 2000, 611). Foucault's panopticism required individuals to still be seen as human subjects, yet there is no reason why his theory could not be applied to a digital setting. In fact, one of panopticism's most prominent features, the invisibility of the observer, is strongly amplified in digital surveillance through cameras and microphones (Fasman 2021).

Second, according to Foucault, in disciplinary societies, there is a "focus on individual rather than aggregated actions" (Galič, Timan, and Koops 2017, 17). This is increasingly the case as machine learning methods enable surveillance tailored to each individual. In the private sector, this is observable *inter alia* in the personalisation of online advertisement, which is not just directed at a particular subset of the population, but at each user individually based on their recorded behaviour. "At the same time, [Foucault's disciplinary society] individualizes *and* masses together" (Deleuze 1995, 5, my emphasis). In the example of online advertisement, this is represented in the superordinate motives of the advertiser, who is primarily interested in maximising their aggregated revenue and less so in the individual sale. Foucault (2009, 183) compares this to a priest who has power both over the flock as a whole and over each animal individually. The simultaneous individualisation and massing together are also observable when the state is using surveillance to keep the individual from violating some predefined interest (national security, political opposition) on the one hand, and to obtain an overall control over the population on the other. In the context of terrorism prevention, for instance, the state is trying to stop every individual terror attack while hoping to keep an overall secure and controlled environment.

Third, surveillance is about modifying the behaviour of the surveilled. This applies to present and past forms of surveillance and is a component of both Foucault's and Deleuze's theory. In the panopticon, the director of the prison can "alter their [inmate's] behaviour" through discipline (Foucault 1977, 204). This form of discipline is a type of power where power can be broadly defined as individual A getting individual B to do something that individual B would otherwise not have done. Power can be exercised visibly, e.g., when surveillance is used towards regulating access as in a control society (Best 2010), or indirectly as in Foucault's panopticon where people's behaviour is changed through "productive soul training" (Haggerty 2006) without them knowing. This foundation is crucial to understand the workings of modern surveillance. Oppressive states use surveillance to discourage dissent while companies try to influence consumers to buy their products (Lukes 2005). In both examples, the surveilling actor is modifying the behaviour of the surveilled.

Although Foucault and Deleuze have never seen the internet in its current form, their theories provide a fruitful foundation from which to develop our theory. In a first step, I have shown their applicability to *all* forms of contemporary surveillance. In the following, I will focus on the heterogeneity of surveillance and delineate along which lines surveillance may be differentiated. To this end, I propose a taxonomy of surveillance that helps to analyse and critique different forms of surveillance that come with different power relations and justifications.

3.3. Taxonomy of surveillance

		Intent	
		Controlling individual behaviour (*architectural surveillance*)	Interested in aggregated behaviour (*infrastructural surveillance*)
Awareness	Object not aware	Government espionage with limited interference, e.g., US National Security Agency	Corporate surveillance for profit, e.g., personalisation of advertisement
	Object aware	Visible state surveillance to discourage certain behaviour, e.g., "Social Credit System" in China	Possible future AI-powered society and economy, e.g., Internet of Things, virtual assistants

Figure 1. *Taxonomy of surveillance*

There are two core dimensions along which surveillance can be differentiated: the awareness of the object under surveillance and the intent of the driver of surveillance. Together, they produce a fourfold taxonomy of surveillance, as presented in Figure 1, where each cell constitutes a different conceptual type. A non-exhaustive example illustrates each type.

The awareness of the surveilled is crucial to understand the form of power that runs from subject to object. If the object is aware that it is being watched, it internalises the constant scrutiny of its action and unconsciously adapts its behaviour to conform to the expectations set. This behaviour is paradigmatic of Bentham's panoptically-arranged prison where inmates are more subservient when they *think* that they are being watched. It is irrelevant whether this is actually the case; it suffices that there always is the

possibility. Panoptic surveillance requires a set of rules that the individual fears to violate in expectation of some punishment. The rules and punishments are set explicitly or implicitly by the observer. In this case, the "effect of surveillance is greater than the sum of its component forces" (Deleuze 1995, 3), as the rules need not be constantly enforced. Instead, individuals engage in self-discipline and self-censorship, trying not to violate the rules. A popular justification of this form of surveillance is: "If you have nothing to hide, you have nothing to fear." This argument is problematic in two respects. First, the 'rules' are defined by the institution that uses surveillance which can be worrisome if it is a malign entity. Second, while an individual might not *yet* have anything to hide, it is likely to adapt its behaviour so that it will also not have to in the future.

If the individual is not aware they are being watched, power dynamics are different. Now, the institution or organisation cannot count on the individuals' self-discipline but has to actively interfere to affect the individuals' behaviour. Here, knowledge is the source of power, reflecting the Foucauldian notion of power-knowledge (Foucault 1977, 55). The interference may be visible or invisible for the individual. When it is visible, the surveilling actor gains its power through the data of the surveilled and by using compulsory power based on that data. An example of this could be law enforcement arresting someone who has been secretly surveilled or a state intelligence agency detaining someone on the suspicion of planning a crime. However, interference can also happen invisibly when individuals' data are used to nudge or manipulate them. Online advertisement is a case of invisible nudging as the individual's data is analysed to recommend products that the user is likely to buy (Zuboff 2019).

The intention and motivation behind surveillance is the second axis of my taxonomy. It can be differentiated with respect to the change in behaviour that the surveilling entity hopes to achieve—on an individual or aggregated level. These two types somewhat correlate with the distinction of *architectural surveillance* and *infrastructural surveillance* by Galič et al. (2017). The former refers to "top-down architectures of surveillance" (Galič et al. 2017, 26), more commonly associated with government institutions and agencies rather than companies. The latter denominates multi-centric networks that perform opaque forms of control and is more prominent in the private sector. Surveillance is a means to an end, namely, to cause some changes in

behaviour in the population, which produces a benefit (profit) for the entity (company) when taking the sum. This stands in contrast to architectural surveillance, which is interested in controlling and discouraging specific individual behaviour.

Surveillance is *ipso facto* an invasion of privacy. A society may choose to allow states and companies to interfere in their privacy in a trade-off for some other social good. In architectural surveillance, privacy is often competing with security. This may be physical security from violence, e.g., when a state uses surveillance to find criminals, or health security, e.g., when governments track the spread of COVID-19 through apps.[1] Invoking security is a common and, as studies show, successful justification of surveillance among governments (see for instance, Wirth, Maier, and Laumer 2019, 1345). This is problematic as humans often have a distorted view of their security threats and consequently cannot objectively balance the actual threats posed against the privacy-invading measures to counter it. One need only look at the example of terrorism which served as justification for some of the largest espionage programmes in Europe's and US history despite attacks being rare and the chance of dying from terrorism being roughly 1 in 30 million (Nowrasteh 2018). This phenomenon has been termed *securitisation* by Buzan et al. (1998). Securitising an issue means that an actor is framing it as a security threat to someone or something to garner support for more drastic countermeasures. Data collection and machine learning have proved very effective and efficient against many security threats, not least when tackling the spread of covid-19. Not using these methods would create redundancies and inefficiencies. However, as Harari (2021) poignantly remarks, "inefficiency is a feature, not a bug. You want to prevent the rise of digital dictatorship? Keep things at least a bit inefficient". While Harari may exaggerate, it is clear that careful balancing between privacy protection and security will be pivotal as machine learning algorithms become more powerful.

In infrastructural surveillance, privacy is most often balanced against comfort. This may not seem like a tough battle for privacy, but humans are often keen on taking the easiest route (Thaler and Sunstein 2008). Almost any privacy invasion in the online sphere makes life a little more comfortable: Sharing phone location data for better app

[1] This does not apply to decentralised COVID-19 apps that do not send data of individuals to a main server (Munzert et al. 2021).

services, using Alexa or Google Assistant, connecting with friends on social networks, storing photos and data in a cloud, using Gmail instead of paying for an email service. States are increasingly introducing stricter legislation to oblige companies to ask for user consent, yet this hardly tackles the issue if users indiscriminately click on "I consent". This suggests that new ways of thinking about privacy are needed.

Within this debate, one realises that there is no universally accepted definition of privacy. Inness (1992) goes as far as saying that the legal and philosophical discourse is in a state of chaos, and there has been little improvement in the last thirty years (Solove 2008). The UN Universal Declaration of Human Rights (1948) says, "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence." However, solid conceptualisations have been a point of contention. This affects the ability of privacy proponents to articulate issues.

> "As a result, we frequently lack a compelling account of what is at stake when privacy is threatened and what precisely the law must do to solve these problems." (Solove 2008, 2)

This conceptual ambiguity helps those engaging in surveillance to muddy the waters and redefine privacy in terms that do not threaten their activities. For example, when Facebook asks you to review your privacy settings, you can choose who can see your posts (Public, Friends of Friends, Friends, Only you etc.), but not which information is shared with Facebook itself. In this case, privacy is redefined to mean interpersonal protection of data within the platform, rather than between the user and the data collecting companies. Users who complete their "privacy check-up" feel good about their choices while Facebook continues collecting their data unnoticed. It is paramount to resist such redefinitions of privacy to enable the concept's use as a mode of critique against contemporary practices (Lyon 2014). As Solove (2008, 10) aptly states, privacy's value "depends upon the social importance of the activities that it facilitates" and defending it is integral to ensuring a free society.

As will be apparent in my case study, surveillance cannot always be categorised as easily in real life because the lines between the four types are blurred. The public may slowly become aware of surveillance that has previously been secret. Similarly, public and private institutions may reach across the architectural-infrastructural divide and collaborate. If seen as a spectrum without hard borders, however, I argue that my

taxonomy can account for all forms of surveillance and, hence, remedies the current technology-dependence of theories of surveillance.

3.4. Hypotheses

In summary, my framework builds on Foucault and Deleuze and finds common ground among all forms of surveillance. It differentiates between four types that differ in terms of awareness and intent and come with different power relations and justifications. To substantiate my theory, I will demonstrate its applicability to recent AI-based surveillance, which existing surveillance theory has not yet been able to account for. Facial recognition software that is interlinked with machine learning algorithms will be used as a case study and as a test for my theory's main hypotheses:

(a) Surveillance is about behavioural modification, in architectural surveillance on an individual level and in infrastructural surveillance on an aggregated level.

(b) The individual's awareness of surveillance affects power dynamics. If they are aware, surveillance causes self-discipline and self-censorship. If they are not, surveillance works through visible and invisible interferences.

(c) Digital surveillance sees humans as abstractions of their data, which we can analyse using Deleuze's concept of *representations*.

(d) Machine learning-based surveillance is faster, more accurate, more comprehensive, and individualises surveillance as in Foucault's panopticism.

(e) Architectural surveillance is justified with increased security, while infrastructural surveillance aims to be accepted by making users' lives more comfortable.

To test these hypotheses, I will describe and analyse the proliferation of AI-powered facial recognition using my taxonomy of surveillance. It will support my argument that Foucault and Deleuze's theories are indeed applicable to novel surveillance technologies and add analytical value to the study of AI surveillance.

4. Case study: Clearview AI

4.1. Facial Recognition and Machine Learning

> "Bentham dreamt of transforming [institutions] into a network of mechanisms that would be everywhere and always alert, running through society without interruption in space and time." (Foucault 1977, 209)

In a world of automated surveillance, facial recognition is at the crucial intersection of the physical world and the data sphere as it allows digital data to be linked to real-world identities. It translates and categorises events into machine-readable databases, which are the basis for any form of digital surveillance. Facial recognition technology consists of three parts: detection, analysis, and recognition. Detection refers to finding a face in a picture or video. Subsequently, the facial features are quantified (*analysis*), and a "faceprint" is created that contains unique measurements of the face. This machine-readable record consists of mathematical vectors of the human face, echoing Deleuze's description of the corporealisation of the body into numerical *representations*. These "vectors" are then compared to a database to match the respective face to a set of previously created faceprints (*recognition*). The first algorithm of this kind was developed by Woodrow W. Bledsoe in the 1960s, which required manual measurements by a human (Libby and Ehrenfeld 2021). It already matched faces accurately enough to be used by law enforcement and outperformed humans in efficiency and speed. Efficiency has remained the key gain from automated facial recognition in the past sixty years, allowing law enforcement and other institutions to search through databases of faces more quickly. However, facial recognition technology has been limited in effectiveness and proliferation until the rise of machine learning in the 2010s. In particular, deep neural networks and larger data sets allowed significant improvements of facial recognition systems in terms of accuracy, speed, and volume, as faces do not have to be manually measured or quantified or even explicitly by a programmed code. Instead, neural networks find patterns themselves in the large dataset it has been trained on (Van Noorden 2020, 357).

An excellent example of recent progress is the new facial recognition tool by Clearview AI, a New York-based company founded by the Australian entrepreneur Hoan Ton-That. By scraping the internet for all publicly available photos from social

media, YouTube, and all other public websites, Clearview AI has built the world's largest database of faces. According to New York Times reporter Kashmir Hill (2020), the database comprises more than three billion pictures which is over four times more than the FBI database (Goodwin 2019). Neural networks generally become more accurate when trained on more data, which has made Clearview AI a forerunner not only regarding the size of their database but also in the accuracy of their recognition tool. Using a 99.6 per cent confidence interval, Clearview AI claims that its error rate lies at 1.4 per cent (Chawla 2020). However, this figure is misleading as it merely states the false positive rate, meaning that users are rarely shown faces that do not match the uploaded picture. The false negative error rate, i.e., the number of faces *not* shown that would have matched the individual, is much assumed to be higher (Ibid.). Clearview has also rejected an independent accuracy test by the US National Institute of Standards and Technology that periodically reviews facial recognition algorithms (Hill 2021). Nevertheless, the performance of Clearview AI's tool is ground-breaking, and its accuracy has drawn interest from public and private institutions.

Clearview AI also exemplifies the proliferation of machine learning innovation from large institutions to smaller actors. In the 20th century, progress in computational science was spearheaded by military and academic research as well as larger companies (e.g., Naughton 2016). To some extent, this has changed with the development of the field of deep neural networks. Individuals' wit and a significant portion of luck were now as important as having strong computer processors, leading to many breakthroughs by start-ups. Leading companies in the field, such as *DeepMind*, were small but innovative start-ups before being bought by large Silicon Valley firms (Gibbs 2014). Clearview AI, which relied heavily on the technological expertise of Ton-That, is no exception. According to internal communications obtained by the New York Times, Ton-That's business co-founder Schwartz joked about academic research lagging behind: "Sounds like Caltech is a year behind you" (Hill 2020; for the Caltech article referred to, see Sheikh 2017). We may conclude that the accessibility of machine learning, which allowed Ton-That to enter the field without significant resources, has led to a multi-centric environment of innovation. Especially in the field of surveillance, where

hundreds of companies of different sizes offer facial recognition technology, this "progress" must be handled with extreme caution.

4.2. Use and justification

Clearview AI started acquiring customers in April 2018 upon successfully developing its facial-recognition tool and a corresponding mobile application to access it. It primarily offered its service to police departments and other law enforcement agencies to identify suspects. If a picture is uploaded, all available pictures of that person are returned together with a link to the website where it was found. This could be a person's Facebook page or their employer's website. The same can be done with a single tap for every person who appears on the same picture. Through a trial offer of only $2,000 per year, Clearview AI was able to demonstrate its superior accuracy, handiness, and reach which far exceeded the number of mug shots and passport pictures in police databases that its more expensive competitors used (Hill 2021). By the end of 2019, six hundred law-enforcement agencies used Clearview AI, and this number has since climbed to over 3,100, according to the company. A leaked list of customers included the FBI, the US Immigration and Customs Enforcement (ICE), Interpol, the London Metropolitan Police, as well as private organisations such as the N.B.A., Macy's, and Walmart (Mac, Haskins, and McDonald 2020). Interviewed police officers who use the app say that they have successfully identified dozens of suspects from previously unsolved cases, including a sexual abuse case and the storming of the capitol on the 6th of January 2021 (Lyons 2021).

Clearview AI acted secretly for more than a year before being uncovered in a front-page article on the New York Times (Hill 2020). Government agencies using the app did not want to tip off criminal suspects to the novel investigative methods used, while the company itself was worried about a public backlash to their product. Since its exposure, Clearview AI has started a public relations campaign attempting to justify the use of its algorithm by law enforcement. Its core argument is that the facial-recognition app increases human security by aiding police investigations. According to CEO Hoan Ton-That (2020b, 12:45), the interest in the app "is just a sign that it's such a human need to be safe". If a person unknowingly appears in the Clearview AI database, for example, because they are in the background of a public social media post, "it could lead to solving a crime and that's a good thing" (Ton-That 2020a, 4:58).

4.3. Criticism

Since being revealed by the New York Times, Clearview AI has faced strong public criticism for its practices. It currently faces ten class-action lawsuits in the United States for using individuals' pictures without their consent. However, the company argues that the pictures are public information since they have been uploaded publicly to social media or another website and are therefore protected under the First Amendment Right. The legal proceedings are likely to stretch over many years, but it is already clear that regulation of AI facial-recognition technology is insufficient. In the United States, there is no federal law regulating facial recognition. In Europe, it is currently only covered by the General Data Protection Regulation (GDPR), which has not impeded the use of Clearview AI within the European Union.

Clearview AI is also gaining public approval through its claim of improving security. The company promotes itself using cases where criminal suspects were caught with the aid of Clearview AI to show its societal benefits, which was well-received, particularly by right-wing commentators (e.g., Llenas 2020). Meanwhile, the left welcomed its use to track rioters who participated in the storming of the capitol (Harwell and Timberg 2021). In a *Nature* survey in 2020, only one-fifth of the surveyed sample group said that they felt uncomfortable with the police using facial recognition of this kind (Van Noorden 2020, 357).

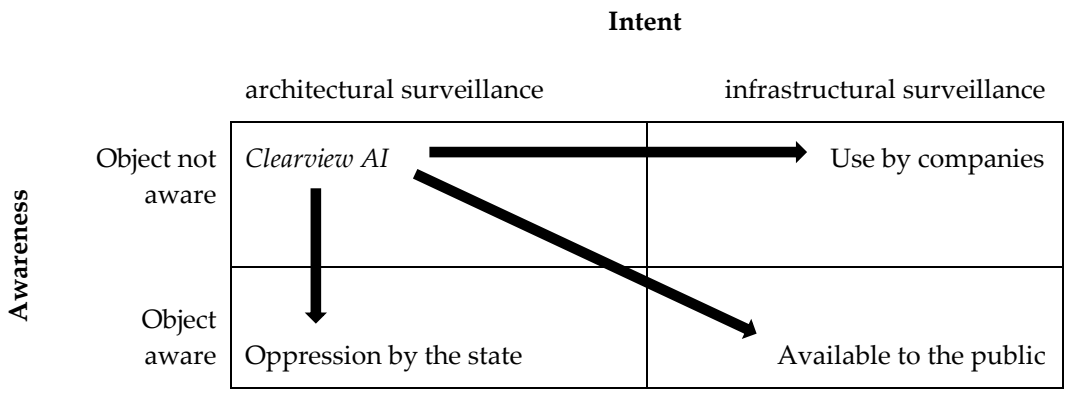

Figure 2. *Clearview AI in the taxonomy of surveillance*

However, if we look at the case study of Clearview AI from the perspective of surveillance, we can unveil crucial processes that may help us understand the risks of AI-powered facial recognition. In my taxonomy of surveillance, the example of Clearview AI may be located on the top left—architectural surveillance where the surveilled are unaware (**Figure 2**). The data is collected secretly by scraping the internet for photos and browsed without the knowledge of the subject. By providing not only the pictures but also the respective website that likely contains further personal information of the individual, organisations using Clearview AI gain immense knowledge and, by extension, power over the searched objects. This imbalance of power between the people *in* the database and the people *using* the database can, if misused, have adverse effects on people's liberties. Being able to link every person to a digital profile allows Clearview AI users to identify almost everyone they may come across—a power that may easily be abused if left unchecked. We must stay alert to securitisation attempts to justify the use of such tools, as exemplified in founder Ton-That's, who frequently appeals to people's fear of crime.

The lines between the four types of surveillance are blurred, and cases may move between them. Clearview AI—which started as a secret law-enforcement tool—was quickly located on the top left, yet it currently moves downwards and rightwards, as the surveilled become more aware and the tool is used by diverse actors with diverse intentions, respectively. The downward move on our spectrum began as the public learned about Clearview AI and its use by law enforcement. Viewing it through the lens of panopticism, it becomes clear that this may have severe consequences by affecting individuals' behaviour. A study conducted on a random sample of 255 participants in 2015 examined the behavioural effects of state surveillance in the United States (Stoycheff 2016). The participants were shown a fictional Facebook post about airstrikes by US forces on ISIS and were asked about how they would comment and share the post if they saw it on their personal Facebook feed. Before the experiment, half of the sample group was subtly reminded of governmental surveillance conducted by the National Security Agency, revealed by whistle-blower Edward Snowden in 2013 (Snowden 2019). The study found that people who were reminded of surveillance were less likely to express views that do not correspond to the mainstream narrative (Stoycheff 2016, 303).

This is evidence of the panoptic effect that causes people to alter their behaviour if they believe that they might be watched, which Noelle-Neumann (1974) also referred to as the "spiral of silence". This may be unintended from the perspective of law enforcement, but people may nevertheless change what they upload, what they do, and whom they meet. Autocratic states are already using this effect to their benefit, as demonstrated for instance by China's Social Credit System which scores the behaviour of every individual. The basis of the system is a network of 20 million facial-recognition cameras called the "Sky Net Program" (Liang et al. 2018, 6). This causes a "penetration of regulation into even the smallest details of everyday life" (Foucault 1977, 198), with the invisibility of the observer behind the cameras as a "guarantee of order" (Ibid., 200). Chinese researchers have also published papers on how machine learning can determine ethnicity, particularly Uyghur ethnicity, from a person's face (Wang et al. 2019)—potentially aiding horrific human rights abuses in the ongoing genocide in Xinjiang (Mattis 2021). Clearview AI's algorithm is, of course, far away from being used for any of such crimes. However, the Chinese case, paired with Foucault's panopticism, may serve as a warning of the behaviourally modifying effects of AI-powered surveillance.

Clearview AI is also moving rightwards on my taxonomy, i.e., towards infrastructural surveillance, as it is expanding into the private sector. Representatives have often claimed that the company will only work with law enforcement and government agencies of selected countries, yet according to leaked customer lists, it has sold the software to private companies, too (Mac, Haskins, and McDonald 2020). A proliferation of facial-recognition software in the hands of the private sector may be equally dystopian as architectural surveillance. As long as people are unaware of being surveilled, Foucault's panopticon is not applicable, but behavioural modification can be achieved differently. The possible applications of Clearview's tool are diverse, from tracking consumers' emotions on their phone or in real life, to their shopping behaviour in physical stores. By linking online with offline behaviour through facial recognition, companies gain significantly more information about their consumer allowing them to personalise advertisement and products even further. This could lead to a "marketing nirvana" (Eagleton-Pierce 2016, 104) by all-knowing companies who thereby have significant control over their customers. Here, Foucault's analogy of the priest's dual

power over the individual and the mass applies. Although the line may be blurred, the personalisation of consumerism may easily lead to manipulation and a (partial) loss of individual agency (Zuboff 2019, 330).

The release of a tool such as Clearview AI's to the public may be the most dystopian scenario. In 2012, the US Federal Trade Commission warned of an app that could identify anybody in real-time, stating it could "cause serious privacy and physical safety concerns" (US Federal Trade Commission 2012, 8). If the anonymity that we currently enjoy when moving in the public sphere ended, society would change dramatically. Contrary to Hoan Ton-That's claim, it may be *less* safe as people could be identified on the street or in a pub by any stranger with a smartphone—like a "Shazam for people" (Hill 2021). It would also alter people's behaviour dramatically through the panopticon effect, the only difference being that, this time, it is the population that watches itself. Similar services have recently appeared online, e.g., by the website *PimEyes*, which matches faces surprisingly well and, like Clearview, returns the website or profile of the matching pictures it has found (Harwell 2021).

5. Conclusion

On 21 April 2021, the European Commission presented a radically new proposal for the responsible use of AI and facial recognition (European Commission 2021). It precludes the use of "real-time remote biometrical identification" in public places, including by law enforcement, and, if passed, would represent a significant step towards privacy protection of citizens (MacCarthy and Propp 2021). However, even with strict regulation, it is inevitable that a person or company will develop a tool like Clearview AI and open it to the public. It will be up to us to resist the temptation to use such software to prevent the formation of a surveillance society. This requires awareness and education of the public as well as transparency from the government and private sector. Equally important, we must see the more profound implications of artificial intelligence in surveillance practices by other governmental or private actors. We shall be suspicious of any institution or company offering security or convenience in return for the use of our data in facial-recognition databases.

This paper has attempted to provide a theoretical basis for analysing AI-powered surveillance practices that is anchored in classical surveillance theory. I have argued that the pre-digital conceptualisations of surveillance by Foucault and Deleuze provide a well-suited foundation for the analysis of current and future surveillance practices. Presenting a taxonomy that differentiates between four forms of surveillance, my theoretical basis can accommodate all forms of surveillance in a single framework. I have demonstrated this using the example of the recently revealed AI facial-recognition tool Clearview AI. The five hypotheses laid out by my theoretical framework were shown to apply to Clearview AI and its use in different contexts. Whether or not AI-powered surveillance is the 'black ball in our society's urn of inventions' remains to be answered. This paper has provided the necessary basis to conduct such analyses, giving us a deeper understanding of the processes, power relations, and justifications behind surveillance practices.

6. Bibliography

Albrechtslund, Anders. 2008. 'Online Social Networking as Participatory Surveillance'. *First Monday* 13 (3).

Albrechtslund, Anders, and Lynsey Dubbeld. 2002. 'The Plays and Arts of Surveillance: Studying Surveillance as Entertainment'. *Surveillance & Society* 3 (2/3).

Bentham, Jeremy. 1838. 'Panopticon: Or the Inspection-House'. In *The Works of Jeremy Bentham*, edited by John Bowring, 4:27–172. Edinburgh: William Tait.

Best, Kirsty. 2010. 'Living in the Control Society: Surveillance, Users and Digital Screen Technologies'. *International Journal of Cultural Studies* 13 (1): 5–24.

Bigo, Didier. 2006. 'Security, Exception, Ban and Surveillance'. In *Theorizing Surveillance: The Panopticon and Beyond*, edited by David Lyon, 46–68. Cullompton, Devon: Willan Publishing.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

———. 2019. 'The Vulnerable World Hypothesis'. *Global Policy* 10 (4): 455–76.

Buzan, Barry, Ole Waever, and Jaap de Wilde. 1998. *Security: A New Framework for Analysis*. Boulder, Colo.: Lynne Rienner.

Chace, Calum. 2016. *The Economic Singularity: Artificial Intelligence and the Death of Capitalism*. London: Three Cs.

Chawla, Vishal. 2020. 'What Went Wrong With Clearview AI?' *Analytics India Mag*, 16 July 2020. https://analyticsindiamag.com/what-went-wrong-with-clearview-ai/.

Chiang, Ted. 2021. 'Why Computers Won't Make Themselves Smarter'. *The New Yorker*, 30 March 2021. https://www.newyorker.com/culture/annals-of-inquiry/why-computers-wont-make-themselves-smarter.

Clarke, Roger. 1988. 'Information Technology and Dataveillance'. *Communications of the ACM* 31 (5): 498–512.

Clegg, Brian. 2017. *Big Data: How the Information Revolution Is Transforming Our Lives*. Hot Science. London: Icon Books Ltd.

Cookson, Clive. 2017. 'DeepMind Computer Teaches Itself to Become World's Best Go Player'. *Financial Times*, 18 October 2017. https://www.ft.com/content/39786fe4-b3e4-11e7-aa26-bb002965bce8.

Del Bello, Lou. 2018. 'Scientists Are Closer to Making Artificial Brains That Operate Like Ours Do'. *Futurism*, 28 January 2018. https://futurism.com/artificial-brains-operate-like-humans-close.

Deleuze, Gilles. 1995. 'Postscript on the Societies of Control'. In *Negotiations, 1972-1990*, edited by Gilles Deleuze, translated by Martin Joughin, 177–82. New York: Columbia University Press.

Deleuze, Gilles, and Félix Guattari. 1987. *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press.

Dholakia, Nikhilesh, and Detlev Zwick. 2001. 'Privacy and Consumer Agency in the Information Age: Between Prying Profilers and Preening Webcams'. *Journal of Research for the Consumer* 1 (1).

Donoho, David. 2019. 'Comments on Michael Jordan's Essay "The AI Revolution Hasn't Happened Yet"'. *Harvard Data Science Review*, June.

Eagleton-Pierce, Matthew. 2016. *Neoliberalism: The Key Concepts*. Routledge Key Guides. New York, NY: Routledge, Taylor & Francis Group.

European Commission. 2021. 'Europe Fit for the Digital Age: Commission Proposes New Rules and Actions for Excellence and Trust in Artificial Intelligence'. Press Release. https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682.

Fasman, Jon. 2021. *We See It All: Liberty and Justice in an Age of Perpetual Surveillance*. New York: Public Affairs.

Feldstein, Steven. 2019. 'The Global Expansion of AI Surveillance'. *Carnegie Endowment for International Peace*, September. https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847.

Fish, Tom. 2018. 'AI to Bring "mankind to Edge of APOCALYPSE" – with Robots a Bigger Risk than NUKES'. *Daily Star*, 15 July 2018. https://www.dailystar.co.uk/news/latest-news/ai-artificial-intelligence-autonomous-weaponry-16872452.

Fjelland, Ragnar. 2020. 'Why General Artificial Intelligence Will Not Be Realized'. *Humanities and Social Sciences Communications* 7 (1): 10.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan. New York: Pantheon Books.

———. 2009. *Security, Territory, Population: Lectures at the Collège de France 1977-78*. Lectures at the Collège de France. Basingstoke: Palgrave Macmillan.

Galič, Maša, Tjerk Timan, and Bert-Jaap Koops. 2017. 'Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation'. *Philosophy & Technology* 30 (1): 9–37.

Gardner, Howard. 1983. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Gibbs, Samuel. 2014. 'Google Buys UK Artificial Intelligence Startup Deepmind for £400m'. *The Guardian*, 27 January 2014, sec. Artificial Intelligence. https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind.

Goodwin, Gretta L. 2019. 'Face Recognition Technology - Testimony before the Committee on Oversight and Reform, House of Representatives'. United States Government Accountability Office. https://www.gao.gov/assets/700/699489.pdf.

Greene, Tristan. 2020. 'A Beginner's Guide to AI: Separating the Hype from the Reality'. *Neural*, 10 September 2020. https://thenextweb.com/news/a-beginners-guide-to-ai-separating-the-hype-from-the-reality.

Haggerty, Kevin D. 2006. 'Tear Down the Walls: On Demolishing the Panopticon'. In *Theorizing Surveillance: The Panopticon and Beyond*, edited by David Lyon, 23–45. Cullompton, Devon: Willan Publishing.

Haggerty, Kevin D., and Richard V. Ericson. 2000. 'The Surveillant Assemblage'. *British Journal of Sociology* 51 (4): 605–22.

Harari, Yuval Noaḥ. 2021. 'Lessons from a Year of Covid'. *Financial Times*, 26 February 2021. https://www.ft.com/content/f1b30f2c-84aa-4595-84f2-7816796d6841.

Harwell, Drew. 2021. 'This Facial Recognition Website Can Turn Anyone into a Cop — or a Stalker'. *The Washington Post*, 14 May 2021. https://www.washingtonpost.com/technology/2021/05/14/pimeyes-facial-recognition-search-secrecy/.

Harwell, Drew, and Craig Timberg. 2021. 'How America's Surveillance Networks Helped the FBI Catch the Capitol Mob'. *The Washington Post*, 2 April 2021. https://www.washingtonpost.com/technology/2021/04/02/capitol-siege-arrests-technology-fbi-privacy/.

Hill, Kashmir. 2020. 'The Secretive Company That Might End Privacy as We Know It'. *The New York Times*, 18 January 2020. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

———. 2021. 'Your Face Is Not Your Own'. *The New York Times*, 18 March 2021. https://www.nytimes.com/interactive/2021/03/18/magazine/facial-recognition-clearview-ai.html.

Hutson, Matthew. 2021. 'Who Should Stop Unethical A.I.?' *The New Yorker*, 15 February 2021. https://www.newyorker.com/tech/annals-of-technology/who-should-stop-unethical-ai?utm_campaign=falcon&utm_source=twitter&utm_medium=social&utm_social-type=owned&utm_brand=tny&mbid=social_twitter.

Inness, Julie C. 1992. *Privacy, Intimacy, and Isolation*. 3rd ed. Oxford: Oxford University Press.

Kitchin, Rob, and Martin Dodge. 2011. *Code/Space: Software and Everyday Life*. Software Studies. Cambridge, Mass: MIT Press.

Koskela, Hille. 2002. 'Webcams, TV Shows and Mobile Phones: Empowering Exhibitionism'. *Surveillance & Society* 2 (2/3).

Larson, Erik J. 2021. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Liang, Fan, Vishnupriya Das, Nadiya Kostyuk, and Muzammil M. Hussain. 2018. 'Constructing a Data-Driven Society: China's Social Credit System as a State Surveillance Infrastructure: China's Social Credit System as State Surveillance'. *Policy & Internet* 10 (4): 415–53.

Libby, Christopher, and Jesse Ehrenfeld. 2021. 'Facial Recognition Technology in 2021: Masks, Bias, and the Future of Healthcare'. *Journal of Medical Systems* 45 (4): 39.

Llenas, Bryan. 2020. 'New Facial Recognition App Promises to Solve Crimes, but Critics Say It Means End of Privacy'. *Fox News*, 14 February 2020. https://www.foxnews.com/us/new-facial-recognition-app-promises-to-solve-crimes-critics-say-its-the-end-of-privacy.

Lukes, Steven. 2005. *Power: A Radical View*. London: Red Globe Press.

Lyon, David. 2007. *Surveillance Studies: An Overview*. Cambridge, UK: Polity.

———. 2008. 'An Electronic Panopticon? A Sociological Critique of Surveillance Theory'. *The Sociological Review* 41 (4): 653–78.

———. 2014. 'Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique'. *Big Data & Society* 1 (2): 1–13.

Lyons, Kim. 2021. 'Use of Clearview AI Facial Recognition Tech Spiked as Law Enforcement Seeks to Identify Capitol Mob'. *The Verge*, 10 January 2021. https://www.theverge.com/2021/1/10/22223349/clearview-ai-facial-recognition-law-enforcement-capitol-rioters.

Mac, Ryan, Caroline Haskins, and Logan McDonald. 2020. 'Clearview's Facial Recognition App Has Been Used By The Justice Department, ICE, Macy's, Walmart, And The NBA'. *Buzzfeed*, 27 February 2020. https://www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement.

MacCarthy, Mark, and Kenneth Propp. 2021. 'Machines Learn That Brussels Writes the Rules: The EU's New AI Regulation'. *Brookings*, May. https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/.

Marri, Shridhar. 2018. 'Can Super Intelligence and Emotional Intelligence Co-Exist?' *Forbes India*, 9 July 2018. https://www.forbesindia.com/blog/technology/can-super-intelligence-and-emotional-intelligence-co-exist/.

Mattis, Peter. 2021. 'Yes, the Atrocities in Xinjiang Constitute a Genocide'. *Foreign Policy*, April. https://foreignpolicy.com/2021/04/15/xinjiang-uyghurs-intentional-genocide-china/.

Moore, Nathan. 2007. 'Nova Law: William S. Burroughs and the Logic of Control'. *Law and Literature* 19 (3): 435–70. https://doi.org/10.1525/lal.2007.19.3.435.

Mudigere, Dheevatsa, Yuchen Hao, Jianyu Huang, Andrew Tulloch, Srinivas Sridharan, Xing Liu, Mustafa Ozdal, et al. 2021. 'High-Performance, Distributed Training of Large-Scale Deep Learning Recommendation Models'. *Facebook*, April. http://arxiv.org/abs/2104.05158.

Munzert, Simon, Peter Selb, Anita Gohdes, Lukas F. Stoetzer, and Will Lowe. 2021. 'Tracking and Promoting the Usage of a COVID-19 Contact Tracing App'. *Nature Human Behaviour* 5 (2): 247–55.

Murakami Wood, David. 2007. 'Beyond the Panopticon? Foucault and Surveillance Studies. In J. Crampton & S. Elden (Eds.), Space, Knowledge and Power: Foucault and Geography (Pp. 245–263). Aldershot: Ashgate.' In *Space,*

Knowledge and Power: Foucault and Geography, edited by Jeremy W. Crampton and Stuart Elden, 245–63. Aldershot: Ashgate.

Naughton, John. 2016. 'The Evolution of the Internet: From Military Experiment to General Purpose Technology'. *Journal of Cyber Policy* 1 (1): 5–28.

Noelle-Neumann, Elisabeth. 1974. 'The Spiral of Silence a Theory of Public Opinion'. *Journal of Communication* 24 (2): 43–51.

Nowrasteh, Alex. 2018. 'More Americans Die in Animal Attacks than in Terrorist Attacks'. *CATO Institute*, March. https://www.cato.org/blog/more-americans-die-animal-attacks-terrorist-attacks.

Obama, Barack. 2016. 'Text of President Obama's Speech in Hiroshima, Japan'. *The New York Times*, 27 May 2016. https://www.nytimes.com/2016/05/28/world/asia/text-of-president-obamas-speech-in-hiroshima-japan.html.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books.

Pinar Saygin, Ayse, Ilyas Cicekli, and Varol Akman. 2000. 'Turing Test: 50 Years Later'. *Minds and Machines* 10 (4): 463–518.

Powers, Thomas M., and Jean-Gabriel Ganascia. 2020. 'The Ethics of the Ethics of AI'. In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank A. Pasquale, and Sunit Das, 27–52. New York: Oxford University Press.

Schlossberg, Tatiana. 2015. 'City Starts Using System That Pinpoints Gunshots'. *The New York Times*, 17 March 2015, sec. A.

Schwartz, Oscar. 2018. '"The Discourse Is Unhinged": How the Media Gets AI Alarmingly Wrong'. *The Guardian*, 25 July 2018. https://www.theguardian.com/technology/2018/jul/25/ai-artificial-intelligence-social-media-bots-wrong.

Sheikh, Knvul. 2017. 'How We Save Face: Researchers Crack the Brain's Facial-Recognition Code'. *Scientific American*, June. https://www.scientificamerican.com/article/how-we-save-face-researchers-crack-the-brains-facial-recognition-code/.

Snowden, Edward J. 2019. *Permanent Record*. London: Macmillan.

Solove, Daniel J. 2008. *Understanding Privacy*. Cambridge, Mass: Harvard University Press.

Spaulding, Norman W. 2020. 'Is Human Judgement Necessary? Artificial Intelligence, Algorithmic Governance, and the Law'. In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank A. Pasquale, and Sunit Das, 375–402. New York: Oxford University Press.

Stigler, Stephen M. 1981. 'Gauss and the Invention of Least Squares'. *The Annals of Statistics* 9 (3): 465–74.

Stoycheff, Elizabeth. 2016. 'Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring'. *Journalism & Mass Communication Quarterly* 93 (2): 296–311.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

The Guardian. 2020. 'A Robot Wrote This Entire Article. Are You Scared yet, Human? GPT-3'. *The Guardian*, 8 September 2020. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

Ton-That, Hoan. 2020a. This man says he's stockpiling billions of our photos Interview by Donie O'Sullivan. CNN. https://edition.cnn.com/2020/02/10/tech/clearview-ai-ceo-hoan-ton-that/index.html.
———. 2020b. Clearview AI CEO Defends Facial Recognition Software Interview by Hari Sreenivasan. PBS. https://www.pbs.org/wnet/amanpour-and-company/video/clearview-ai-ceo-defends-facial-recognition-software/.
United Nations UDHR. 1948. 'Universal Declaration of Human Rights, G.A. Res. 217A (III), U.N. Doc A/810 at 71'. United Nations.
US Federal Trade Commission. 2012. 'Facing Facts: Best Practices for Common Uses of Facial Recognition Technologies'. US Federal Trade Commission. https://www.ftc.gov/sites/default/files/documents/reports/facing-facts-best-practices-common-uses-facial-recognition-technologies/121022facialtechrpt.pdf.
Van Noorden, Richard. 2020. 'The Ethical Questions That Haunt Facial-Recognition Research'. *Nature* 587 (7834): 354–58.
Wang, Cunrui, Qingling Zhang, Wanquan Liu, Yu Liu, and Lixin Miao. 2019. 'Expression of Concern: Facial Feature Discovery for Ethnicity Recognition'. *WIREs Data Mining Knowl Discov.* 9 (1278): 1–17.
Whitaker, Reginald. 1999. *The End of Privacy: How Total Surveillance Is Becoming a Reality*. New York: New Press.
Wiener, Norbert. 1961. *Cybernetics: Or, Control and Communication in the Animal and the Machine*. 2nd ed. Cambridge, MA: The MIT Press.
Wirth, Jakob, Christian Maier, and Sven Laumer. 2019. 'Justification of Mass Surveillance: A Quantitative Study'. *14th International Conference on Wirtschaftsinformatik*, February, 1337–51.
Zuboff, Shoshana. 2015. 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization'. *Journal of Information Technology* 30 (1): 75–89. https://doi.org/10.1057/jit.2015.5.
———. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.