Edinburgh Research Explorer

Reconciling governmental use of online targeting with democracy

Reconciling Governmental Use of Online Targeting With Democracy

KATJA ANDRIĆ, The University of Edinburgh, United Kingdom

ATOOSA KASIRZADEH, The University of Edinburgh & The Alan Turing Institute, United Kingdom

The societal and epistemological implications of online targeted advertising have been scrutinized by AI ethicists, legal scholars, and policymakers alike. However, the government's use of online targeting and its consequential socio-political ramifications remain under-explored from a critical socio-technical standpoint. This paper investigates the socio-political implications of governmental online targeting, using a case study of the UK government's application of such techniques for public policy objectives. We argue that this practice undermines democratic ideals, as it engenders three primary concerns — Transparency, Privacy, and Equality — that clash with fundamental democratic doctrines and values. To address these concerns, the paper introduces a preliminary blueprint for an AI governance framework that harmonizes governmental use of online targeting with certain democratic principles. Furthermore, we advocate for the creation of an independent, non-governmental regulatory body responsible for overseeing the process and monitoring the government's use of online targeting, a critical measure for preserving democratic values.

CCS Concepts: • **Social and professional topics** → **Government technology policy**; • **Security and privacy** → *Social aspects of security and privacy*.

Additional Key Words and Phrases: governmental online targeting, public policy, targeted advertising, democracy, democracy and AI, AI governance

1 INTRODUCTION

Online targeting, a fundamental aspect of the modern digital economy, involves customizing online products and services based on users' psychological profiles. These profiles are derived from algorithmic analysis of personal data, primarily acquired through the online monitoring and processing of data [7, 50]. Targeted ads, which have become an omnipresent feature of our online experiences, extend beyond merely displaying personalized ads for commercial items we have recently searched. As the Cambridge Analytica scandal revealed, online targeting has also permeated the political sphere, raising concerns about its broader implications [66].

In this paper, we explore the socio-political implications of the government's use of online targeting by examining a case study conducted by Collier et al. [18]. This study uncovers that, since at least 2015, the UK government and law enforcement agencies have been employing targeted online advertising in various campaigns. Examples of such campaigns include a fire safety campaign ran by the Home Office that utilised Amazon data and targeted particular citizens through their Alexa speakers, or a National Crime Agency campaign, targeting young video gamers in their

online environments with a goal of decreasing online crime activity. These campaigns aim to influence citizens' behavior by tailoring personalized online advertisements according to their psychological profiles. Prior to Collier et al.'s research [18], civil society in the UK was largely unaware of the prevalence and scope of such practices.

Apart from the work by Collier et al. [18], there is a notable gap in research regarding the socio-political implications of governments' use of online targeting.[1] Given the unique relationship between governments' use of online targeting, legitimate power, and the vast number of individuals affected by this contentious practice, there is an urgent need for interdisciplinary research to address the socio-political implications of governmental use of online targeting.

In this paper, we present a critical and philosophical analysis of the socio-political challenges posed by the UK government's use of online targeted advertising to achieve public policy objectives. We investigate the contested relationship between this practice and the fundamental principles of democracy. We explore potential strategies to reconcile the government's use of online targeting with the core values and tenets of a democratic society. We emphasize that this issue warrants significant attention in terms of designing appropriate regulatory and governance mechanisms, particularly as generative artificial intelligence (e.g., large language models) has the potential to significantly enhance online targeting by accelerating the process, reducing costs, and improving the quality of content production [77]. This underscores the need for a timely and thorough examination of the implications of such governmental use of online targeting within the context of democracy and public policy.

Thus far, the majority of philosophical literature on online targeting within the political domain has primarily concentrated on its influence on personal autonomy during voting and the negative repercussions for democracy [6, 69, 70, 83]. Nevertheless, several other essential democratic values and principles, including equality, transparency, and privacy, frequently remain underexplored in this context. This paper endeavors to examine and highlight the often-neglected impacts of the governmental use of online targeting on these values and principles, thereby offering a more comprehensive perspective on the subject. The paper will proceed as follows.

In Section 2, we define online targeting and briefly delve into its history and applications. In Section 3, we present Collier et al.'s study [18] of the UK government's use of online targeting, analyzing various past campaigns undertaken by the government. In Section 4, we highlight three key problems posed by governments utilizing online targeting: the Transparency Problem, the Privacy Problem, and the Equality Problem. We will demonstrate that each of these concerns conflicts with at least one fundamental aspect of democracy, as outlined by Lever [52]: (i) enabling citizens to be informed participants in law-making and electing representatives; (ii) upholding citizens' civil, socioeconomic, and political rights; and (iii) safeguarding citizens' equality and freedom. In Section 5, we argue that online targeting by governments can still be reconciled with democratic principles and values. Drawing on Züger and Asghari's AI governance framework [82], we advocate for the establishment of an independent institution to oversee campaign design and deployment, serving as a check on government power. We show that this solution can, in principle, effectively address the three problems raised in Section 4. The final section concludes the paper.

2 WHAT IS ONLINE TARGETING?

In today's digital world, an immense volume of personal data is accessible, encompassing not only basic details like names, email addresses, and birth dates but also extending to information derived from social media activity, sexual orientation, health records, search history, and purchasing habits. This data is collected, generated, stored, and managed

---

[1]Some discussions of similar practices can be found within the field of strategic communication. However, much of how governments employ online targeting remains implicit. For example, see [56] for a good overview of the field of strategic communication in the era of Big Data and [42] for a discussion of governmental communication during the COVID-19 pandemic, which included some online targeting.

by various entities, including commercial companies like Google, Facebook, and Amazon, as well as data brokers who specialize in aggregating and selling data to third parties [3, 21].[2]

Psychological profiles in online targeting refer to the comprehensive digital representation of an individual's personality, preferences, and behaviors, derived from their online activities [11]. These profiles are generated using data mining and machine learning algorithms that analyze various data points, such as browsing history, social media interactions, and online purchases, to infer users' interests, habits, and tendencies. The profile-building process often involves combining data from multiple sources, employing machine learning techniques to identify patterns, and categorizing individuals based on shared characteristics. These profiles are designed to forecast users' behavior and decision-making processes in various situations [50].

By constructing such detailed portraits of users, advertisers and commercial companies can tailor their content, messages, or campaigns to resonate with specific target audiences, ultimately enhancing the effectiveness of their online engagement [7] and advancing their business goals [54, 70]. This process of adapting online content to align with the psychological profiles of users is known as *online targeting*.

While online targeting is a relatively recent development, its offline counterpart has a more extensive history. The practice of influencing individual or group behavior by appealing to their psychology predates the advent of the internet or algorithm-driven profiling. A notable example, as described by Halpern [37], occurred in the 18th century when Frederick the Great successfully promoted the consumption of potatoes — a formerly unpopular and bland vegetable — in order to stave off famine. He accomplished this by establishing a guard around the royal potato fields and publicly expressing his admiration for the crop. This strategy piqued public interest and facilitated the widespread popularity of potatoes throughout Prussia. Later, in the 20th century, Edward Bernays leveraged psychological principles to influence public opinion and behavior, making him a key historical figure in the practice of manipulating individual or group behavior by appealing to their psychology [9, 10]. Today, such behavior can be examined through the lens of social and cognitive psychology, disciplines that were first combined with economics in the early 20th century to give rise to the field of behavioral economics.

Behavioral economics investigates the consequences of human cognitive limitations on decision-making within markets [59]. Classical economics traditionally portrays individuals as rational decision-makers who consistently select the optimal choice after conducting thorough cost-benefit analyses, uninfluenced by extrinsic factors or emotions. In contrast, insights from various disciplines have revealed the impact of cognitive biases, emotions, perceptions, heuristics, and social contexts on rationality [37]. Consequently, economists have come to acknowledge that the way the choice options are presented can significantly influence the choices we make.

In 2008, renowned scholars Richard H. Thaler and Cass R. Sunstein, the former having been awarded a Nobel Prize for his significant contributions to behavioral economics, introduced the concept of *choice architecture* [71]. This idea pertains to the environments in which we make decisions, encompassing the range of options available to us, the manner in which they are presented, and the entities presenting them. Thaler and Sunstein maintain that these choice architectures inherently influence the decisions we make. As a result, if we aim to guide someone towards a specific choice, we can modify their choice architecture through the application of *nudges*. As defined by Thaler and Sunstein [71, p.6], a nudge is an aspect of the choice architecture that predictably alters people's behavior without forbidding any options or significantly changing their economic incentives. This concept suggests that subtle changes in choice architecture can have a significant impact on decision-making. A prime example of this would be strategically placing

---

[2]Data brokers often operate behind the scenes, amassing information from numerous sources to create comprehensive psychological profiles of individuals, which are then traded within the data market for various purposes.

healthy food options at eye level in a grocery store, while positioning unhealthy alternatives out of sight, in order to boost sales of healthier choices.

Online targeting operates on a principle similar to that of its non-digital counterparts, such as grocery stores, wherein choice structures are manipulated to guide user behavior towards a desired outcome. What sets online targeting apart, however, is the capacity to accurately segment audiences based on their algorithmically generated profiles. Furthermore, it allows for dynamic adjustments to choice structures tailored to each individual user, taking into account their personal preferences [68]. User-facing platforms, which serve as primary channels for online targeting, are meticulously optimized through extensive user experience (UX) testing methods and the application of behavioral insights. This optimization aims to maximize user attention, resulting in heightened engagement with targeted content and the conversion of this attention into revenue or other desired behaviors [4]. This characteristic sets online targeting apart from offline targeting, as it enables the deliberate and strategic delivery of customized nudges to each user, informed by an understanding of the nudge's impact on that specific individual. As a result, these personalized nudges prove more potent, allowing those employing targeting tactics to accomplish their business objectives with greater efficiency and effectiveness.

The pervasive use of online targeting, especially in targeted advertising — which constitutes 79% of all online advertising [57] — is hardly surprising in today's digital landscape. Internet companies frequently employ Chief Behavioral Officers and Choice Architecture Engineers to shape their customers' behavior [49]. As a foundational component of the modern internet economy [7], online targeting contributes to enhanced sales of goods and services. Additionally, it enables numerous apps and websites, ranging from Facebook to The Guardian, to provide low-cost or free services by relying on revenue generated through targeted ads [45].

Online targeting techniques, which have proven effective in shaping behavior, are not only confined to the commercial realm. These methods have infiltrated the political sphere, where they have been employed in political campaigns to sway voters' decisions [22, 28]. The infrastructure of online targeting has thereby enriched the longstanding practice of political marketing, which dates back to the 1980s [80]. The tactics of online political targeting were reported during Barack Obama's groundbreaking 2008 presidential campaign which involved identifying supporters and persuading undecided voters [32, 62], but gained even greater notoriety during Donald Trump's 2016 presidential campaign [4, 15]. In the latter campaign, the Trump team enlisted the expertise of Cambridge Analytica, a British political consultancy firm. This company harvested data from an astonishing 87 million Facebook users through a seemingly innocuous personality quiz [66]. With this wealth of information, Cambridge Analytica crafted intricate psychological profiles of users and deployed emotionally manipulative, personalized advertisements to influence their voting behavior in favor of Trump [4, 15]. This striking example underscores the pervasive reach of online targeting techniques. It also highlights their potentially profound impact on political outcomes.

Online political targeting is not limited to the United States, as political parties across Europe have also utilized these techniques [83]. The UK's House of Commons Select Committee on Digital, Culture, Media and Sport has described online political targeting as a "democratic crisis" [24, p.51]. As we intend to show in this paper, this crisis has deepened, with the government now using the online targeting infrastructure for public policy purposes, which can undermine fundamental principles of democracy if used inappropriately. In the following section, we examine a case study that demonstrates how the UK government's utilization of online targeting can potentially compromise certain democratic principles.

3 GOVERNMENTAL USE OF ONLINE TARGETING

A recent study by Collier et al. [18] has uncovered a striking finding - the UK government and law enforcement agencies are utilizing online targeted advertising to influence citizens' behavior and achieve public policy objectives. This use of online targeting presents unique socio-political implications that warrant further exploration. To understand the potential consequences, it is essential to first grasp the nature of the UK government's employment of online targeted advertising.

The UK government has a history of attempting to change the behavior of its citizens in order to achieve public policy goals. Following the establishment of the Behavioural Insights Team (also known as *Nudge Unit*) in 2010, the government has been openly applying nudge theory to public policy [18]. The Team's objective is to influence people's choices by designing appropriate incentives or obstacles, thereby encouraging the desired options, all while incorporating insights from behavioral economics [38]. The Nudge Unit's inception was fueled by the growing popularity of behavioral change techniques combined with social marketing practices among UK public policy makers, who had been experimenting with the application of commercial marketing principles to promote public goods since 2004 [64]. Examples of the Team's initiatives include adjusting tobacco prices to deter individuals from purchasing it or incorporating carefully crafted tax prompts in letters to taxpayers to encourage prompt payments (e.g., "most people pay their tax on time") [38].

While influencing citizens' behaviour for the purpose of achieving public policy objectives is not a new practice for the UK government, what *is* a new practice is combining these behaviour change strategies with the infrastructures of online targeted advertising.[3] As explained in Section 2, those infrastructures involve data gathering and processing by machine learning algorithms to create profiles of individuals and groups so that messages specifically tailored to each profile can be created and delivered to those deemed susceptible. Collier et al.'s research [18] shows that government departments are combining operational data gathered and produced by state institutions and the associated systems of classification and profiling of social groups (e.g. the needs, risks and vulnerabilities of groups such as patients, immigrants and welfare recipients) with data gathered and produced by commercial companies and the internet economy (e.g. clicks, page visits, shopping habits, social media activity, and online social interactions). The hybridisation of the two categories of data and their algorithmic processing allows for both a wide and a deep insight into UK's communities and individuals.

During the past decade, the UK government ran numerous campaigns at the heart of which was online targeting. Collier et al. [18] map those into three distinct modes of operation, which we term (i) the *minimally targeted mode*, (ii) the *maximally targeted mode*, and (iii) the *outsourced mode* for ease of reference. The first two modes are delivered by government and law enforcement agencies and differ in the level of sophistication of the targeting used. In contrast, the last mode involves outsourcing the services of private sector companies but is on par with the maximally targeted mode in terms of targeting sophistication.

The first and the least sophisticated mode, i.e., the *minimally targeted mode*, amounts to simply extending the advertising scope to online spaces through online advertisement buys. As the name suggests, this mode is minimally targeted as it does not involve much audience segregation or reliance on individuals' profiles but is targeted at entire population groups. It involves actions such as running advertisements on Tiktok to reach younger audiences.

---

[3]Adapting commercial marketing technologies, such as online targeted advertising, to programs that are designed to influence the behaviour of targeted individuals for their benefit as well as wider social benefit, is what Andreasen [1] terms *social marketing*.

The second and significantly more sophisticated mode of operation, which we term the *maximally targeted mode*, leverages targeting to reach specific groups or individuals and informs the design of the run adverts. This mode employs algorithms that enhance personalization and relies on a network of government entities led by the Government Communication Service to develop nationwide behavior change strategies. In addition to implementing multi-site and single-site campaigns to change behavior through tailored messages, this mode also includes countering misinformation online and protecting the government's reputation from negative messages. One notable example cited by Collier et al. [18, p.7]. is a fire safety campaign by the Home Office, where the department utilized purchasing data from individuals who recently bought candles on Amazon and targeted them with fire safety messages through their smart speakers, such as Alexa.

The maximally targeted mode also involves leveraging the maximally targeted technologies and methods employed by law enforcement for preventive purposes. A prime example of this is the CYBER CHOICES preventative diversion program ran by the UK National Crime Agency (NCA) in collaboration with behavioral psychologists [18]. This initiative utilized Google and YouTube advertisements targeted towards UK adolescents between the ages of 14 and 20 who were identified through NCA surveillance as potentially interested in gaming. The ads would appear whenever they searched for cybercrime services and warn them of the illegal nature of purchasing such services and the consequences they could face if they did so. As a part of this campaign, NCA officers also visited the identified "targets" at their homes, discussed their online behaviour with their parents, and invited them to workshops organised and ran by NCA. The goal of those workshops was twofold. Firstly, the individuals were taught the skills required to turn their illegitimate interests into a legitimate career. Secondly, NCA used the workshops to gather data to optimise the design of further targeted ads. There is evidence that this particular campaign was successful in reducing the rate of cybercrime [17].

The third and final mode of operation, the *outsourced mode*, entails entrusting the entire process of designing, developing, and executing campaigns to private sector companies [18]. One instance is SuperSisters, a Muslim online lifestyle platform established by J-Go Media in 2015, aimed at young British Muslim girls. Although marketed as a platform for sharing and creation of empowering content, the project sparked controversy when it was revealed that it was covertly funded by a government counter-extremism arm and that the content on the website was carefully curated to counteract what the state deemed to be "overtly Islamic" [43]. Another example includes a UK Home Office-supported knife prevention campaign, targeting young Black individuals residing in London's deprived neighborhoods, created by FCB Inferno and All City Media [18]. The campaign's offline component included messages displayed on takeout boxes in fried chicken restaurants, based on police data which indicated that Black individuals commit more knife crimes and perpetuated a racist stereotype that they consume fried chicken [76]. The online component aimed at young Black males living in impoverished areas of London drew upon the same data and stereotype.

4 SOCIO-POLITICAL ISSUES AND ONLINE TARGETING'S CONTESTED RELATIONSHIP WITH DEMOCRACY

Some of the examples outlined in the previous section may cause discomfort. In this section, we will pinpoint some of the social and political issues that arise from the use of online targeted advertising by democratic governments to achieve public policy goals. Clearly formulating these issues will not only concretize the discomfort, but it will also lay the groundwork for a deliberation on the legitimacy of such practices and, if necessary, their proper form. The examination of these questions will be the focal point of the upcoming section. For the moment, our attention will be directed toward highlighting the challenges associated with this procedure.

In the context of governmental use of online targeted advertising, the primary socio-political concern is the potential for abuse of power and erosion of democratic values.[4] According to Susser [68], autonomy refers to a person's ability to make decisions based on their own personal values and beliefs. Online targeted advertising undermines autonomy by intentionally and covertly manipulating individuals through exploiting their decision-making vulnerabilities and cognitive biases [70, p.4]. This manipulation often goes unnoticed, as individuals are unaware of the influence on their decision-making. In line with this concern and as a result of governmental online targeting, individuals are left open to being molded into the ideal citizens according to the government's preferences. While we acknowledge the negative impacts on autonomy through manipulation, governmental use of online targeting raises additional and novel socio-political issues that have not yet been thoroughly analyzed. This paper will focus specifically on such socio-political concerns.

At the core of the socio-political concerns specific to governmental online targeting is the practice's contested relationship with democracy and democratic values. So far, philosophical literature tackling the relationship between online targeting and democracy has focused almost solely on the detrimental effect online targeting has on one's autonomy during voting, the manipulative nature of this practice and the dangers it poses for democratic elections [6, 69, 70, 83]. However, governmental use of online targeting for public policy goals brought to light by Collier et al.'s research [18] expands the known scope of online targeting in the political domain beyond using targeted adverts on citizens during election campaigns. This novel use of online targeted advertising undermines other, overlooked, features central to democracy besides voting. These are the ones we will tackle.

According to Lever [52], democracy comprises three key elements: (1) allowing citizens to be informed participants in decision-making processes, including the creation of laws and the election of representatives; (2) guaranteeing and preserving civil, socioeconomic, and political rights; and (3) ensuring equality and freedom for all citizens. However, the UK government's practice of using tailored online advertisements, particularly through maximally targeted and outsourced methods, violates — to a certain degree — all three of these democratic principles. We argue that this practice raises three major issues - the Transparency Problem, the Privacy Problem, and the Equality Problem - that directly challenge one or more of the key elements of democracy.

4.1 The Transparency Problem

Transparency is generally understood to be one of the central principles of democracy [65]. If citizens do not have the ability to freely access information, they cannot keep the government accountable for their actions and decisions. Moreover, they cannot make informed decisions at the ballot box or actively participate in other democratic processes, such as publicly questioning and criticising the government's decisions. The lack of transparency about the governmental use of online targeting undermines the feature (1) of democracy mentioned above: enabling its citizens to participate in the determination of laws and the election of their representatives. In this subsection, we will show that the UK government's use of online targeting suffers from a lack of transparency. We call this the Transparency Problem.

The Transparency problem can be viewed from different angels. Firstly, there is a general lack of governmental transparency about the practice of online targeted advertising. The government has never published a comprehensive list of online targeting campaigns it has created itself or outsourced from the private sector. While some scarce information is available on different governmental bodies' websites, most of the information about the campaigns comes from the

---

[4]Additional ethical issues and potential solutions surrounding the practice of targeted advertising have been extensively explored in literature, including works by Thaler and Sunstein [71], Hansen and Jespersen [39], Wilkinson [78], and Nys and Engelen [60]. It is important to note that some of these issues are not exclusive to online targeting or governmental use of it, as non-digital nudging also raises similar concerns. For the lack of space, we will not be covering the non-digital instances in this paper.

documentation submitted to various industry awards by third-party agencies hired by the government to develop and run the campaigns.[5]. From government records only, it is uncertain how many campaigns the government has run so far, who the campaigns targeted, or what the content of those campaigns was.

Apart from the former lack of transparency concerning the practice, online targeted advertising suffers from an inherent lack of widespread transparency. As Collier et al. [18] argue, targeted adverts are normally only seen and intended to be seen by those targeted, meaning that most of the population will never encounter them. In contrast, non-targeted governmental campaigns delivered either online or offline are, in principle, visible to the entire population (including the press) who can scrutinize and challenge them. The lack of extensive visibility of targeted adverts reduces the capability for broader scrutiny. Consequently, it reduces the public's ability to hold the government accountable for its actions.

Moreover, determining the effectiveness of targeted governmental advertisements is challenging since it cannot be measured by directly observable outcomes, such as sales conversion, as is the case with commercial targeting [18]. This not only makes it difficult to justify the measure, but also reduces the citizens' ability to hold the government accountable, as they must not only know what the government is doing, but also whether it is achieving satisfactory results.[6]

A final blow to transparency is delivered by the fact that machine learning algorithms underpinning the structure of targeted advertising suffer from an inherent opacity problem. This means that it is not always possible to know why and how — in non-mathematical terms — an algorithm reached a specific prediction [14, 35, 47, 48]. The implication is that citizens are often unable to obtain an explanation for why they were targeted with a specific advert. This issue is exacerbated by Collier's assertion that, even upon request, the government will not disclose targeting data, which is the data used to identify individuals for a particular advert.[7] Without sufficient information being provided to the relevant stakeholders about the algorithmic systems that make decisions about them, it is challenging to envision any meaningful discussion of the ethical concerns raised by the behavior of the system [27].

4.2 The Privacy Problem

The concept of privacy carries various interpretations [29, 34, 53, 58]. A significant definition posits privacy as the capacity to control our personal information. This control enables us to manage others' knowledge about us, thereby establishing varying degrees of intimacy with different individuals or groups [67]. Crucially, the freedom to disclose our personal details at our discretion and to chosen recipients safeguards our autonomy and dignity. In instances where our personal data is collected and analyzed, the exercise of informed consent ensures we retain control. Conversely, data collection and analysis conducted without our informed consent infringes upon our right to privacy.

The right to privacy is a fundamental human right recognised by democratic countries [16, 72]. To illustrate, in the UK, this right is protected by the *Human Rights Act 1998*. This right offers protection to citizens from undue and illegal governmental surveillance. Moreover, it fosters an environment where individuals feel secure to explore, express their beliefs, and cultivate their interests. In this subsection, we will make the argument that the UK government has, in some instances, collected data on its citizens without obtaining informed consent. This action can be seen as an infringement on the citizens' right to privacy, a concern we refer to as the Privacy Problem. This problem not only

---

[5]We confirmed this through personal communication with Ben Collier, 2022.
[6]The field of public relations also faces the challenge of determining the success of its campaigns due to the lack of readily quantifiable metrics for success. Some efforts have been made to establish a framework for tracking and measuring the impact of public relations campaigns in the field, which could be relevant to governmental online targeting. For example, see [5], [55] and [63]
[7]This was revealed to us in a personal conversation with Ben Collier in August 2022.

conflicts with Lever's second principle of democracy - the preservation of citizens' civil, socioeconomic, and political rights - but also compromises the citizens' capacity to make genuinely independent political decisions, contradicting the first principle. Let us unpack this problem.

Privacy and consent are closely interrelated. The act of consenting to data collection and analysis allows you to exert control over your personal data. Consent, viewed as a normative concept, can make an act that would otherwise be impermissible, permissible by facilitating the transfer of rights and obligations between involved parties [46]. For consent to be morally transformative, it must be informed. In the context of data, this necessitates the provision of the following: (i) an explicit description of the potential uses and restrictions of your data, (ii) a specific definition of the scope within which your data can be utilized, (iii) ample information for the consenting party to comprehend what they are agreeing to, (iv) a range of free-to-choose options, and (v) a mutually equitable treatment between both parties [46]. In the UK, the personal information of individuals is safeguarded by the *Data Protection Act 2018*, which outlines six principles of data protection that all parties responsible for using personal data must adhere to. For example, the first data protection principle requires that consent is obtained from the individual for the information to be collected and processed.[8]

If a party plans to use your data in a way that was not disclosed during the initial consent process, this original consent becomes invalid and a fresh consent is required. For example, if a party intends to use your postal code for a different purpose than the one you initially agreed to, such as determining insurance rates, they must seek your permission again [2]. Similarly, if a party intends to combine your data with another dataset (for both of which you have given individual consents), they must ask for your renewed consent. This is because the merger could generate new information that was not anticipated when you initially gave consent.

The UK government's use of targeted advertising raises concerns regarding the validity of consent due to not disclosing relevant information in both of above senses. While the government has sought permission from citizens to gather administrative data, it neglected to disclose that this data could be utilized for profiling and online targeted advertising. It is probable that the surveys used to collect such data contained a clause explicitly stating the data would not be used for marketing purposes. As a result, the consent previously acquired is flawed and fresh consent must be obtained. Additionally, the lack of transparency surrounding this practice leaves many unaware that their online activity, such as Facebook likes and Amazon purchases, could be leveraged to fulfill public policy objectives. This data, when merged with census data, generates new information for which proper consent was not sought. Consequently, this practice seems to be illegitimate.

The Privacy Problem, besides raising questions about consent validity, also impedes UK citizens' full participation in democratic processes. Here is why. The erosion of privacy harms not only those whose privacy is at stake, but also democracy itself [25, 52]. Privacy creates a safe space for individuals, facilitating the growth of their opinions, the exploration and pursuit of diverse interests, and the freedom to make decisions without fear of judgment or backlash. Being constantly monitored, with the potential of being categorized into risk groups by the government or targeted with intimidating adverts, and in extreme cases, even receiving home visits for merely searching certain terms (as seen in the CYBER CHOICES campaign), can instigate self-censorship and make people feel unsafe in their own country. This is especially true for historically disadvantaged groups, who may feel particularly unsafe and skeptical towards the government [52].

---

[8]Notably, the Act lists certain exceptions, but a full interpretation of these is beyond the scope of this paper.

Governments have long engaged specific subsets of the population through various forms of offline advertising and public policy campaigns. For instance, public health campaigns have often targeted smokers, warning them of smoking dangers via pamphlets distributed in healthcare facilities and television commercials. However, there is a distinct contrast between such campaigns and those founded on the framework of online targeting. In the case of offline campaigns, individuals are aware they are the target because they identify with the group in question. On the other hand, when it comes to online governmental targeting, individuals receive specific advertisements because they have been personally identified by the government as relevant recipients. This identification process is made possible by continuous governmental monitoring of the individual's online activities, to which they likely did not consent. Furthermore, this kind of targeting grants governments access to previously inaccessible areas — private homes. For example, whereas in the past, you might have encountered a fire safety poster while commuting to work, now, you might be greeted by a message about fire safety from Alexa when you return home because you purchased a candle on Amazon the previous week.

Online targeting by government agencies is significantly more invasive than traditional methods, largely due to the opaque nature of the data collection process. This lack of transparency can exacerbate feelings of paranoia and vulnerability. The core of the chilling effect we are discussing lies in this disparity. Persistent surveillance fosters an environment of constraint, where individual autonomy is violated [8]. For a democracy to thrive, it necessitates independent and autonomous decision-makers, a condition which becomes challenging to fulfill in the absence of privacy. Consequently, privacy appears to be a crucial component in exercising our democratic rights, including political choice [52].

4.3   The Equality Problem

The third socio-political concern central to our discussion is what we call the Equality Problem. This problem relates to the possibility that government-led targeted advertising may undermine the principles of equality and justice upheld by democracy, thereby violating the third democratic feature outlined by Lever.

In an effort to identify the target audience for a campaign, the UK government and law enforcement construct a profile of the ideal target and advertise to individuals who fit this profile. This profile is created using available data and is generated algorithmically. However, it has become increasingly clear that AI algorithms have the potential to perpetuate existing wrongful social inequalities [13, 31, 61]. This is largely due to the data used to develop these algorithms often reflecting the biases and disparities inherent in society. The algorithms look for patterns in the data, but if the data reflects wrongful inequalities, such as over-policing of Black neighborhoods, then the output of these algorithms risks perpetuating existing social hierarchies.

Consequently, this approach has the potential to exacerbate the marginalization of already disadvantaged groups. It could erroneously place these individuals into risk categories, not because of their actions, but due to systemic discrimination. This sort of bias, targeting protected characteristics like race, sex, gender, and disability, is not only unethical but also unlawful according to the authors of *Equality Act 2010*. A notable instance of this is the targeting of young Black males residing in economically disadvantaged areas of London in a knife crime prevention campaign, a topic elaborated upon in Section 3.

Algorithmic bias can lead to discrimination, but it's not the only source. As AI algorithms unearth patterns within data sets, novel forms of discrimination can arise, not necessarily linked to traditionally protected attributes [75]. People can be placed into "ad hoc" categories, like dog owners or video gamers, and subsequently face unfair treatment compared to those outside these groups. This is because the algorithm had detected a correlation between owning a

dog or playing video games and certain behaviors. However, owning a dog or playing video games may not be a normatively acceptable basis for forming a group around them if the formation occurred as a result of a spurious correlation found in the data or if the statistical correlation is insufficiently significant [75].

The use of online targeting by the government is not only leading to unfair and discriminatory outcomes for individuals, but it also poses a deeper problem with regards to equality. It creates a shift in power dynamics within society, giving the government more control over its citizens than is desirable in a democratic society [81].

The connection between knowledge and power cannot be ignored. The government's collection and analysis of individuals' data gives them a greater ability to influence and control their citizens [74]. This raises concerns about potential misuse of such power, as warned by Königs [51]. Furthermore, it raises questions about who has the authority to shape society and determine its priorities [18].

The question therefore persists: Can the practice of targeted governmental advertising ever align with democratic principles and values? The following section outlines steps towards answering this inquiry.

5 RECONCILING ONLINE TARGETING WITH DEMOCRACY

In light of our previous discussions, it is clear that online targeting can be an effective strategy for achieving certain goals. This approach allows the government to identify and engage with specific subgroups more effectively, offer tailored resources to address public issues, and optimise the use of their limited resources. However, its employment by the UK public sector has provoked serious questions about transparency, privacy, and equality. In order to align its use with democratic principles, the government must address these concerns. While we acknowledge there is no easy fix, we propose a few steps towards a potential solution.

First, we suggest requirements for the design and execution of online governmental targeting campaigns that are in alignment with democratic values. We will reference the recent AI governance framework developed by Züger and Asghari [82] as our initial guideline. Second, we propose the creation of an independent body to monitor the development and implementation of these campaigns. This institution will ensure compliance with the defined requirements and provide guidance to government officials.

The use of AI for social benefits has been gaining momentum in recent years. Numerous applications of AI are now directed towards addressing issues that impact human life and well-being [20, 73]. Considering that public policy should embody public interests [44], utilizing AI-powered online targeting to achieve these policy goals is another way in which AI can serve the public interest and contribute to societal good. This usage of AI-powered online targeting to fulfill public policy objectives falls within the ambit of an AI governance framework that prioritizes public interest. According to Züger and Asghari [82], there are five prerequisites for a system to align with the public interest: (1) public justification, (2) focus on equality, (3) inclusion of a deliberation/co-design process, (4) implementation of technical safeguards, and (5) commitment to openness for validation.

To successfully address the three problems we highlighted in the previous section, all five requirements must be met. The first three requirements are context-sensitive and their fulfillment will differ depending on the project. Requirements 4 and 5, being technical supplements, are less dependent on the specific nature of the project.

5.1 Public Justification

Züger and Asghari [82] propose that for an AI-based solution to be recognized as serving the public interest, it must possess a normative democratic justification that is widely accepted by the public. This justification should include a lucid explanation of the issue that the AI solution seeks to address and why it is superior to alternative solutions.

This stipulation is anchored in the philosophical views of public interest presented by Habermas [36], Held [40], and Bozeman [12], who assert that the public should determine what is in their best interest on an individual case basis.

In terms of the UK government's approach to online targeting, it is crucial for the government to maintain transparency. This involves disclosing any online targeting campaigns to the public, clarifying why this particular method was chosen, and detailing the operational aspects of the campaign, such as the ad content and target demographic. By doing so, the government can counteract the prevalent lack of transparency in this practice. To further enhance accountability, the government should also consistently update and provide a predictive measure of the campaign's efficacy, as well as quantifying interactions with the ads. Nevertheless, while conducting these processes, it is of utmost importance to safeguard individual identities by anonymising the data.

Finally, the justification should also provide easily understandable information about opaque algorithms. By adequately fulfilling these requirements, the Transparency Problem can be avoided.

5.2 Equality

Züger and Asghari [82] assert that serving equality (and at the very least not hurting it) must be the most important normative goal of a solution that aims to promote public interest. Thus, they argue that any public interest AI-based solution must find a way to solve the problem of algorithmic bias and not create unwanted power imbalances in society. The second requirement is also based on previous scholarship. More precisely, the work of a legal scholar Feintuck [33], who argues that something can be in the public interest only if it promotes equality of citizenship.

To satisfy this requirement, the UK government should not run any campaigns that discriminate against its citizens (whether based on some protected characteristic stemming from algorithmic bias or on spurious correlations found in the data), reproduce harmful social hierarchies or create new power imbalances. By fulfilling this requirement, the government would solve the first part of the Equality Problem - discrimination by the government. There remains the second part of the problem - unwanted power asymmetry. The proposed preliminary solution for this issue will be presented at the end of this section, where we suggest the formation of an independent institution tasked with monitoring these practices.

5.3 Deliberation/Co-Design Process

Züger and Asghari [82] argue that to determine what is in the public interest for a public at a given time, the public must be involved in the system's design through the process of deliberation. The process can take any form, from online documentation to interviews with citizens. Without public deliberation about the public's interests, the team of developers will have to assume the interests of others which can easily lead to harmful mischaracterisation, unintended consequences and public rejection of the project [82]. Therefore, they argue that those who will be affected by the system must have their say. As with the previous two requirements, Züger and Asghari root this requirement in an existing philosophical theory. According to Bozeman [12], who draws from Dewey [26], the individuals who form the public can only determine what is in the public interest through public deliberation by expressing their personal views, listening empathetically to others and reaching a compromise which benefits everybody.

Such a highly democratic approach is available to the UK government, although it is rarely followed since it requires significant resources. However, a 2021 outsourced campaign by Police Scotland called Breaking the Cycle of Fear, whose goal was to reduce violence in the most deprived areas of Glasgow and Dundee, showed why such a process is desirable, despite the required resources. During the design process, interviews were conducted with individuals from targeted areas who had managed to escape the cycle of violence. The interviews aimed to understand what life in such areas

looks like, how one gets involved with violence, and how one escapes it. Data gathered through these interviews informed the design of a targeted advert - in this case, a short movie[9]- which was assessed for credibility and realism by the interviewees.

In three months, almost 500 targeted people reached out to the Scottish Violence Reduction Unit asking for help in escaping the cycle of violence themselves. Moreover, the campaign received only positive criticism from the targeted population, who did not report feeling marginalised. Had the previously discussed knife crime reduction campaign followed a similar approach, it could have avoided relying on harmful stereotypes and further marginalising the communities they were trying to reach. Therefore, governments should follow this approach in future online targeting-based campaigns as this will significantly help solve the Equality Problem.

5.4 Technical Safeguards and Openness to Validation

Züger and Asghari [82] argue that AI-based systems need to implement technical safeguards, including data quality and system accuracy, data privacy, and safety and security. That is, the data fed into the AI systems needs to be free of bias and of high quality so that the outcomes are accurate, can be validated and serve equality. Satisfying this technical safeguard would take us closer to solving the Equality Problem. Further, data protection and privacy laws must be complied with. This includes obtaining informed consent from everyone whose data is being used, which would solve the consent part of the Privacy Problem. Finally, it must be ensured that the system is secure and robust to eliminate malfunctions, unintended functionalities and security breaches.

The next requirement Züger and Asghari [82] propose amounts to having the entire system, including the design process, open to validation of others. They note two main reasons for this requirement. Firstly, any system that impacts the public at large may cause unintended harm, regardless of the good intentions of its makers. Secondly, any system that claims to be in the public interest should follow the basic democratic norm of transparency, allowing those impacted by the system to review all decisions made by the systems' makers and the workings of the technology to ensure that its mechanisms are democratic [82]. This would help with the explainability of the system and, consequently, people's trust in the system. Thus, if the UK government satisfied this requirement, it would take another step toward solving the Transparency Problem.

5.5 Independent Institution

In theory, satisfying these five requirements could be left entirely to the government's discretion while trusting that they will behave ethically. However, there would be no way of ensuring that the government is sincere in following the guidelines without an independent organization overseeing the entire process and serving as a check on the government's power. Moreover, the government officials who currently run these campaigns do not always possess the necessary knowledge to understand the harms a campaign may cause, the technical workings of the algorithms behind the campaigns or the regulation that needs to be followed. Interviews with UK public officials reveal that online targeting campaigns, especially at a local level, are often designed without much consideration of the ethical issues they may raise, prior research or planning. For example, Collier and Wilson [19] report an insider to a UK government-led counter-radicalisation campaign describe their approach as "throwing things at a wall to see what sticks". Similarly, Wilson's conversations with an employee from the UK's Foreign, Commonwealth and Development Office reveal that online targeting of citizens identified as being at risk of turning to religious extremism is often done because it can

---

[9]The movie can be viewed here: https://www.youtube.com/watch?v=APUfXvepLQQ&t=3s

be used to show that something is being done, despite evidence that such campaigns have no positive effects or even have negative effects [79].

Thus, an independent, interdisciplinary team of experts whose purpose would be to ensure that the government is fulfilling Züger and Asghari's requirements and educate the public officials wishing to use the infrastructure of online targeting for public policy goals is needed. Such an institution would also reduce the unwanted shift in the power balance to the government's favour since the institution would serve to ensure that the government is not abusing its power. This would increase the public's trust in the system and make them less likely to fear judgement or retribution from the government for what they do in the privacy of their online spaces, thereby making them less likely to self-censor their behaviour. Therefore, an institution imagined in this way would contribute to solving both the Equality and the Privacy Problems.

In the end, as we have argued, it seems possible to reconcile governmental use of online targeting with democratic principles and values. To do so, the government must satisfy Züger and Asghari's five requirements, while being overseen by an independent organisation yet to be established.

6 CONCLUSION

This paper addresses the socio-political implications of a previously overlooked governmental practice: the use of online targeted advertising for public policy objectives. Our discussion is anchored on a particular case involving the UK government's use of this practice, as explored in the study by Collier et al. [18]. We argued that this practice, as characterized in this paper, is strikingly undemocratic and raises three major anti-democratic concerns: the Transparency Problem, the Privacy Problem, and the Equality Problem. To reconcile this practice with democratic principles, we sketch the outline of a solution: that the governmental use of online targeting should adhere to an AI governance framework, such as the one developed by Züger and Asghari [82], and be monitored by an independent organization comprising interdisciplinary experts.

In order to provide a comprehensive understanding of the topic, it is important to acknowledge the limitations of this paper. Firstly, our analysis is based on Lever's formulation of democratic principles [52]; future work should consider and test other desirable socio-political principles in relation to the use of online targeting by governments. Secondly, we focused exclusively on a single AI governance framework proposed by Züger and Asghari [82]; future research should explore alternative frameworks and assess their applicability to this context. Lastly, our investigation primarily centers on the practices of the UK government; conducting further studies to examine the use of online targeting in other democratic countries would be beneficial to better understand the generalizability of our findings and recommendations.

While this paper sheds light on the issue of governmental use of online targeting, numerous questions still remain unanswered. For example, is it appropriate for governments to employ these technologies for national security purposes when transparency may not be feasible? Should they also utilize it to counter misinformation and disinformation within online communities? Our objective in writing this paper is to contribute to the ongoing discourse and provide preliminary insights into these complex questions.

REFERENCES

[1] Alan R. Andreasen. 1994. Social Marketing: Its Definition and Domain. *Journal of Public Policy & Marketing* 13, 1 (1994), 108–114. http://www.jstor.org/stable/30000176

[2] Adam J. Andreotta, Nin Kirkham, and Marco Rizzi. 2022. AI, big data, and the future of consent. *AI & SOCIETY* 37, 4 (2022), 1715–1728. https://doi.org/10.1007/s00146-021-01262-5

[3] Gary Anthes. 2015. News: Data brokers are watching you. *Commun. ACM* 58, 1 (2015), 28–30. https://doi.org/10.1145/2686740

[4] Vian Bakir. 2020. Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting. *Frontiers in Communication* 5 (2020). https://doi.org/10.3389/fcomm.2020.00067

[5] Vian Bakir, Eric Herring, David Miller, and Piers Robinson. 2019. Organized Persuasive Communication: A new conceptual framework for research on public relations, propaganda and promotional culture. *Critical Sociology* 45, 3 (2019), 311–328. https://doi.org/10.1177/0896920518764586 arXiv:https://doi.org/10.1177/0896920518764586

[6] Anne Barnhill. 2022. *How philosophy might contribute to the practical ethics of online manipulation.* Taylor and Francis, 49–71. https://doi.org/10.4324/9781003205425-4 Publisher Copyright: © 2022 selection and editorial matter, Fleur Jongepier and Michael Klenk; individual chapters, the contributors..

[7] David Gareth Beer, Joanna Redden, Ben Williamson, and Simon Yuill. 2019. *Landscape Summary: Online Targeting: What is online targeting, what impact does it have, and how can we maximise benefits and minimise harms?* Centre for Data Ethics and Innovation.

[8] Claire Benn and Seth Lazar. 2022. What's Wrong with Automated Influence. *Canadian Journal of Philosophy* 52, 1 (2022), 125–148. https://doi.org/10.1017/can.2021.23

[9] Edward L Bernays. 1928. Manipulating public opinion: The why and the how. *American journal of sociology* 33, 6 (1928), 958–971.

[10] Edward L Bernays. 1947. The engineering of consent. *The Annals of the American Academy of Political and Social Science* 250, 1 (1947), 113–120.

[11] Sophie C Boerman, Sanne Kruikemeier, and Frederik J Zuiderveen Borgesius. 2017. Online behavioral advertising: A literature review and research agenda. *Journal of advertising* 46, 3 (2017), 363–376.

[12] Barry Bozeman. 2007. *Public Values and Public Interest: Counterbalancing Economic Individualism.* Georgetown University Press. http://www.jstor.org/stable/j.ctt2tt37c

[13] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency.* PMLR, 77–91.

[14] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society* 3, 1 (2016), 2053951715622512.

[15] Raphael Bömelburg and Oliver Gassmann. 2021. Cambridge Analytica: Magical Rise, Disastrous Fall. In *Connected Business: Create Value in a Networked Economy*, Oliver Gassmann and Fabrizio Ferrandina (Eds.). Springer International Publishing, Cham, Switzerland, 387–396. https://doi.org/10.1007/978-3-030-76897-3_28

[16] Julie E Cohen. 2013. What privacy is for. *Harvard law review* 126, 7 (2013), 1904–1933.

[17] Ben Collier, Gemma Flynn, James Stewart, and Daniel R. Thomas. 2021. *SCCJR briefing paper: Influence government: Exploring practices, ethics, and power in the use of targeted advertising by the UK state.* The Scottish Centre for Crime & Justice Research.

[18] Ben Collier, Gemma Flynn, James Stewart, and Daniel Thomas. 2022. Influence government: Exploring practices, ethics, and power in the use of targeted advertising by the UK state. *Big Data & Society* 9, 1 (2022), 20539517221078756. https://doi.org/10.1177/20539517221078756 arXiv:https://doi.org/10.1177/20539517221078756

[19] Ben Collier and Lydia Wilson. 2022. Governments Try to Fight Crime via Google Ads. *New Lines Magazine* (2022).

[20] Josh Cowls, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. 2021. A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence* 3, 2 (2021), 111–115.

[21] Matthew Crain. 2018. The limits of transparency: Data brokers and commodification. *new media & society* 20, 1 (2018), 88–104.

[22] Matthew Crain and Anthony Nadler. 2019. Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy* 9 (2019), 370–410. https://www.jstor.org/stable/10.5325/jinfopoli.9.2019.0370

[23] Data Protection Act 2018. 2018. Parliament of the United Kingdom. Available at: https://www.legislation.gov.uk/ukpga/2018/12/introduction/enacted.

[24] DCMS. 2018. Disinformation and 'fake news': Interim Report. <https://publications.parliament.uk/pa/cm201719/ cmselect/cmcumeds/363/363.pdf>.

[25] Firmin DeBrabander. 2020. *Life after Privacy: Reclaiming Democracy in a Surveillance Society.* Cambridge University Press. https://doi.org/10.1017/9781108868280

[26] John Dewey. 2012. *The Public and Its Problems: An Essay in Political Inquiry.* Penn State University Press. http://www.jstor.org/stable/10.5325/j.ctt7v1gh

[27] Nicholas Diakopoulos. 2020. 197Transparency. In *The Oxford Handbook of Ethics of AI.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.11 arXiv:https://academic.oup.com/book/0/chapter/290661457/chapter-ag-pdf/44521945/book_34287_section_290661457.ag.pdf

[28] Katharine Dommett. 2019. The Rise of Online Political Advertising. *Political Insight* 10, 4 (2019), 12–15. https://doi.org/10.1177/2041905819891366 arXiv:https://doi.org/10.1177/2041905819891366

[29] Cynthia Dwork. 2006. Differential privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33*. Springer, 1–12.

[30] Equality Act 2010. 2010. Parliament of the United Kingdom. Available at: https://www.legislation.gov.uk/ukpga/2010/15/contents.

[31] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

[32] Lisa Farman and Daniel Kreiss. 2013. A Research Agenda for Online Political Advertising: Surveying Campaign Practices, 2000-2012. *International Journal of Communication* 7 (01 2013), 2046–2066.

[33] Mike Feintuck. 2004. *'The Public Interest' in Regulation.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199269020.001.0001

[34] Rachel L Finn, David Wright, and Michael Friedewald. 2013. Seven types of privacy. *European data protection: coming of age* (2013), 3–32.

[35] Mario Günther and Atoosa Kasirzadeh. 2022. Algorithmic and human decision making: for a double standard of transparency. *AI & SOCIETY* (2022), 1–7.

[36] Jürgen Habermas. 2022. *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik: Platz 1 der Sachbuchbestenliste der WELT.* Suhrkamp Verlag.

[37] David Halpern. 2015. *Inside the Nudge Unit: How Small Changes Can Make a Big Difference.* WH Allen. https://books.google.hr/books?id=M4uSnQAACAAJ

[38] David Halpern and Michael Sanders. 2017. Nudging by government: Progress, impact, & lessons learned. *Behavioral Science & Policy* 2 (2017), 52 – 65.

[39] Pelle Guldborg Hansen and Andreas Maaløe Jespersen. 2013. Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation* 4, 1 (2013), 3–28. https://doi.org/10.1017/S1867299X00002762

[40] Virginia Held. 1970. *The Public Interest and Individual Interests.* Basic Books. https://books.google.hr/books?id=bOIuswEACAAJ

[41] Human Rights Act 1998. 1998. Parliament of the United Kingdom. Article 8, Available at: https://www.legislation.gov.uk/ukpga/1998/42/schedule/1/part/I/chapter/7.

[42] Bernadette Hyland-Wood, John Gardner, Julie Leask, and Ullrich K. H. Ecker. 2021. Toward effective government communication strategies in the era of COVID-19. *Humanities and Social Sciences Communications* 8, 1 (2021), 30. https://doi.org/10.1057/s41599-020-00701-w

[43] Nosheen Iqbal. 2019. "We acknowledge we went wrong": Lifestyle website for Muslim teens admits it should have been clearer about Home Office funding. https://www.theguardian.com/uk-news/2019/sep/15/lifestyle-website-for-muslim-teens-is-covertly-funded-by-the-home-office

[44] Paul De Jersey. 2003. Public Interest and Public Policy: Unruly Horses Alike? *Legal Ethics* 6, 1 (2003), 16–22. https://doi.org/10.1080/1460728X.2003.11424173 arXiv:https://doi.org/10.1080/1460728X.2003.11424173

[45] Ashley Johnson. 2021. Banning Targeted Ads Would Sink the Internet Economy. https://itif.org/publications/2022/01/20/banning-targeted-ads-would-sink-internet-economy

[46] Meg Leta Jones and Elizabeth Edenberg. 2020. 358359Troubleshooting AI and Consent. In *The Oxford Handbook of Ethics of AI.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.23 arXiv:https://academic.oup.com/book/0/chapter/290665619/chapter-ag-pdf/44521933/book_34287_section_290665619.ag.pdf

[47] Atoosa Kasirzadeh. 2021. Reasons, values, stakeholders: A philosophical framework for explainable artificial intelligence. *arXiv preprint arXiv:2103.00752* (2021).

[48] Atoosa Kasirzadeh and Andrew Smart. 2021. The use and misuse of counterfactuals in ethical machine learning. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 228–236.

[49] Paula Johanna Kirchhof. 2021. *The Marionette Question: What Is Yet to Be Answered about the Ethics of Online Behaviour Change?* Springer International Publishing, Cham, 35–38. https://doi.org/10.1007/978-3-030-80083-3_4

[50] Ansgar Koene, Elvira Perez, Christopher James Carter, Ramona Statache, Svenja Adolphs, Claire O'Malley, Tom Rodden, and Derek McAuley. 2015. Ethics of Personalized Information Filtering. In *Internet Science*, Thanassis Tiropanis, Athena Vakali, Laura Sartori, and Pete Burnap (Eds.). Springer International Publishing, Cham, 123–132.

[51] Peter Königs. 2022. Government Surveillance, Privacy, and Legitimacy. *Philosophy & Technology* 35, 1 (2022), 8. https://doi.org/10.1007/s13347-022-00503-9

[52] Annabelle Lever. 2006. Privacy Rights and Democracy: A Contradiction in Terms? *Contemporary Political Theory* 5, 2 (2006), 142–162. https://doi.org/10.1057/palgrave.cpt.9300187

[53] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. 2021. When Machine Learning Meets Privacy: A Survey and Outlook. *ACM Comput. Surv.* 54, 2, Article 31 (mar 2021), 36 pages. https://doi.org/10.1145/3436755

[54] Sandra C Matz, Michal Kosinski, Gideon Nave, and David J Stillwell. 2017. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the national academy of sciences* 114, 48 (2017), 12714–12719.

[55] David Michaelson and Don W. Stacks. 2011. Standardization in Public Relations Measurement and Evaluation.

[56] Nicholas Michelsen and Thomas Colley. 2019. The field of Strategic Communications Professionals: a new research agenda for International Security. *European Journal of International Security* 4, 1 (2019), 61–78. https://doi.org/10.1017/eis.2018.9

[57] Catherine Miller, Rachel Coldicutt, and Hannah Kitcher. 2018. People, Power and Technology: The 2018 Digital Understanding Report. doteveryone.org. https://doteveryone.org.uk/wp-content/uploads/2019/07/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf

[58] Adam D Moore. 2008. Defining privacy. *Journal of Social Philosophy* 39, 3 (2008), 411–428.

[59] Sendhil Mullainathan and Richard H. Thaler. 2001. Behavioral Economics. In *International Encyclopedia of the Social & Behavioral Sciences*, Neil J. Smelser and Paul B. Baltes (Eds.). Pergamon, Oxford, 1094–1100. https://doi.org/10.1016/B0-08-043076-7/02247-6

[60] Thomas RV Nys and Bart Engelen. 2017. Judging Nudging: Answering the Manipulation Objection. *Political Studies* 65, 1 (2017), 199–214. https://doi.org/10.1177/0032321716629487 arXiv:https://doi.org/10.1177/0032321716629487

[61] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453.

[62] D. Plouffe. 2010. *The Audacity to Win: The Inside Story and Lessons of Barack Obama's Historic Victory.* Thorndike Press. https://books.google.hr/books?id=9zj9QgAACAAJ

[63] Kenneth D. Plowman, Robert I. Wakefield, and Beki Winchel. 2015. Digital publics: Tracking and reaching them. *Public Relations Review* 41, 2 (2015), 272–277. https://doi.org/10.1016/j.pubrev.2014.12.007 Digital Publics.

[64] Jessica Pykett, Rhys Jones, Marcus Welsh, and Mark Whitehead. 2014. The art of choosing and the politics of social marketing. *Policy Studies* 35, 2 (2014), 97–114. https://doi.org/10.1080/01442872.2013.875141 arXiv:https://doi.org/10.1080/01442872.2013.875141

[65] B. Peter Rosendorff and John A. Doces. 2006. Democracy and Transparency. *Swiss Political Science Review* 12, 3 (October 2006), 99–112.

[66] Christophe Olivier Schneble, Bernice Simone Elger, and David Shaw. 2018. The Cambridge Analytica affair and Internet-mediated research. *EMBO reports* 19, 8 (2018), e46579. https://doi.org/10.15252/embr.201846579 arXiv:https://www.embopress.org/doi/pdf/10.15252/embr.201846579

[67] D.J. Solove. 2008. *Understanding Privacy.* Number s. 10 in Understanding privacy. Harvard University Press. https://books.google.hr/books?id=XU5-AAAAMAAJ

[68] Daniel Susser. 2019. Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 403–408. https://doi.org/10.1145/3306618.3314286

[69] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2019. Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review* 4 (2019), 1–45.

[70] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2019. Technology, autonomy, and manipulation. *Internet Policy Review* 8, 2 (2019), 1–22. https://doi.org/10.14763/2019.2.1410

[71] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press. https://books.google.hr/books?id=dSJQn8egXvUC

[72] Judith Jarvis Thomson. 1975. The right to privacy. *Philosophy & Public Affairs* (1975), 295–314.

[73] Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle CM Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, et al. 2020. AI for social good: unlocking the opportunity for positive impact. *Nature Communications* 11, 1 (2020), 2468.

[74] Carissa Véliz. 2020. *Privacy is Power: Why and How You Should Take Back Control of Your Data.* Penguin Random House.

[75] Sandra Wachter. 2021. Affinity profiling and discrimination by association in online behavioural advertising. *Berkeley Technology Law Journal* 35, 2 (2021), 369–430.

[76] Julia Webster. 2019. 'Racist, or, at Best, Unfathomably Stupid.' Britons Finding it Hard to Swallow Campaign to Curb Knife Crime With Ads on Fried Chicken Boxes. https://time.com/5656420/britain-knife-crime-racist-chicken-campaign/

[77] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks posed by Language Models. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, 214–229. https://doi.org/10.1145/3531146.3533088

[78] T. M. Wilkinson. 2013. Nudging and Manipulation. *Political Studies* 61, 2 (2013), 341–355. https://doi.org/10.1111/j.1467-9248.2012.00974.x arXiv:https://doi.org/10.1111/j.1467-9248.2012.00974.x

[79] Lydia Wilson. 2021. Gone to Waste: the 'CVE' Industry After 9/11. *New Lines Magazine* (2021).

[80] Dominic Wring. 1997. Reconciling marketing with political science: Theories of political marketing. *Journal of Marketing Management* 13, 7 (1997), 651–663. https://doi.org/10.1080/0267257X.1997.9964502 arXiv:https://doi.org/10.1080/0267257X.1997.9964502

[81] Shoshana Zuboff. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of information technology* 30, 1 (2015), 75–89.

[82] Theresa Züger and Hadi Asghari. 2022. AI for the public. How public interest theory shifts the discourse on AI. *AI & SOCIETY* (2022). https://doi.org/10.1007/s00146-022-01480-5

[83] Frederik J. Zuiderveen Borgesius, Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, and Claes de Vreese. 2018. Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review* (Feb 2018). https://doi.org/10.18352/ulr.420