March 2020

Instrumentalization Theory: An Analytical Heuristic for a Heightened Social Awareness of Machine Learning Algorithms in Social Media

Andrew R. Miller
University of South Florida

Instrumentalization Theory: An Analytical Heuristic for a Heightened Social Awareness of Machine-Learning Algorithms in Social Media

by

Andrew R. Miller

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Rhetoric and Composition
Department of English
College of Arts and Sciences
University of South Florida

Major Professor: Carl Herndl, Ph.D.
Kristen Allukian Ph.d.
John Skvoretz Ph.d.

Date of Approval:
March 10th, 2020

Keywords: digital technology, rhetoric, science and technology policy, socio-technical opacity, socio-reduction, disinformation architecture, technical code, ethical design implementation

Table of Contents

List of Figures

Abstract

New innovations in information management and communication technologies have produced technological assemblages which have radically altered the way people socialize and interact with the world. The most significant and ubiquitous of these technologies is what is colloquially referred to as 'machine learning.' Like most, if not all, technologies, machine learning models are neither wholly good nor bad. Their functional ethics are largely determined by the context in which they are employed. However, their ubiquity demands that we develop a heightened social consciousness of the way machine learning simultaneous constrains, manipulates and democratizes social processes. In order to develop better social understanding of technologies that incorporate machine learning, we must clarify how and why corporate engineers and executives scale and implement machine learning into their respective applications and services. Unfortunately, high-level calculus and computer science obscure this situation and make formulating a critical space for humanist theorists and Science and Technology policymakers an exhaustive discursive endeavor. The absence of a well understood discourse on the manner in which machine-learning algorithms are implemented represents a kind of socio-technical opacity, which obscures technological processes for contextualized corporate, design and user-motivated ethics. In order to address this problem, I propose to analyze the primary machine-learning algorithm models which organize and rank the information presented on social media newsfeed. An analysis that clarifies the function of machine learning algorithms can

promote academic research and provide the impetus for Science and Technology policy incentives. Finally, this sort of analysis suggests the need for a regulatory agency for machine learning algorithms prior to their implementation into public production site environments (i.e. social media)

Introduction

In the last decade and a half social networking technology has risen to global prominence. Applications and websites that enable users to create and share content as a means of social networking have created a new means of communication, but most importantly a new digital space in which communication and ethics can be generated and managed. Already there are calls for Facebook Inc. to develop machine learning models to distinguish between political fact and lie in digital journalism. Done without adequate ethical consideration, this could result in unforeseen ethical and social implications. This is but one of many emerging discursive issues surrounding machine-learning in social media. One of the profoundly wonderful characteristics of the modern era is its ability to quickly build new technologies to grapple with modern issues of information management overload and sustainability. But a key disadvantage of such an exponential rate of technological innovation is that the humanist theories which should be associated with their ethical implementation and use lag behind developmentally. This limits public and academic discourse and prevents rhetoric and Science and Technology scholars from properly evaluating the communicative ethics of new digital technologies. A key meta-physical characteristic of present technological usage is reductionism, described by philosophers as the manner in which objects are stripped of their natural connections so that new, implicit, social and technical connections emerge in their place. Instrumentalization theory, first postulated by Andrew Feenberg in 'A Critical Theory of Technology' (1991), provides a dual-level (primary and secondary) analytical heuristic to probe the implications of social processes that are socio-technically reduced in a digital space which is algorithmically designed, determined and managed. Feenberg's instrumentalization theory stresses the notion that, like any tool, technology embodies the ethics and ontological nature of the individuals who design and create them. As such, his theory exposes a deep exigency for a synthesized rationality towards the democratic and ethical implementation and critique of new technological systems. The emergence of ubiquitous social media which reduce social interactions provides a tangible space to use Feenberg's dual level (primary, secondary) instrumentalization theory to develop a humanist approach to technological critique and design.

This first section of this paper examines the manner in which instrumentalization theory operates as a critical theory of technology and offers a useful analytical heuristic for examining the ethics of machine learning algorithms. Andrew Feenberg, and his theoretical forefathers such as Herbert Marcuse, Martin Heidegger and Jurgen Habermas, provide a theoretical space in which to discuss issues of technological ethics, purpose and being. A grounding in critical theory is necessary in order to understand instrumentalization theory and the applicability of its dual-level process to a complex technology such as machine learning. Furthermore, understanding the theories which define technological ethics is also essential in order to apply them to specific machine learning processes.

Section 2 focuses on issues related to the ethics of machine learning, more specifically on the perceived functional ethics of the technology as they exist in their leveraging of multiple public services on Facebook and other select case studies, such as Virginia Eubanks examination of the social repercussions of automated decision-making technologies on low-income households across the United States, Obermeyer et. al research detailing the algorithmic targeting and marginalization of Black insurance-seekers in the United Kingdom . The section also discusses the manner in which our understanding of this technical ethic is obscured by the prominence of socio-technical opacity. Socio-technical opacity, as explained by machine learning researchers and engineers from the public and private sectors, arises out of public lack of understanding of technical processes, adherence to experts and technical hurdles which prevent even machine learning engineers from understanding exactly how ML decisions occur. The primary effect of socio-technical opacity in machine learning is that it serves to obfuscate the importation of algorithmic bias from machine learning engineers and the context and nature of the available data which the algorithm can utilize. It is essential to firmly establish the functional ethics of machine learning, prior to the formation of a solution to the harmful aspects of their public implementation.

The third section of the paper spends time overlaying Feenberg's instrumentalization theory and its primary and secondary analytical levels with specific machine learning algorithm models utilized by Facebook Inc. to leverage important public applications and services. The analysis aims to provide a means of establishing the general technical code of Facebook's utilization of machine learning in public production sites, and the manner in which this technical code presents an exigency for public policy related to the responsible and considerate implementation of machine learning models in society. The paper concludes with a discussion of the specific socio-technical ethics that American science and technology policy should consider in order to ensure that machine learning algorithm models that are implemented by multi-national tech-corporations are done so in a way that is codified, regulated and more in line with humanist concerns of wellbeing and dignity, over those of technical expediency and control.

Research Question: How can instrumentalization theory provide a theoretical basis for the realization of a heightened social-technical consciousness of the beneficial and reductive characteristics of machine learning models on Facebook's Newsfeed, and what are the implications of the need of such a realization for Science and Technology Policy?

Section 1: Instrumentalization Theory as an Analytical Heuristic

Andrew Feenberg is deeply engrained within Critical Theory's critique on the philosophy of technology. In *Critical Theory of Technology*, Feenberg lays out the various figures from the Frankfurt School who influenced his outlook on the state of technological critique today (namely Martin Heidegger's *Question Concerning Technology*, Adorno & Horkheimer's *Dialectic of Enlightenment*, Marcuse's *1-Dimensional Man*, and Habermas's *Towards a Rational Society*). The critique of technology is a key feature of Critical Theory dating back to its founders Adorno and Horkheimer's disparaging notion that any form of technical development was "a substrate of domination" (6). Over time many rhetorical and philosophical luminaries have added their perspectives to the lexicon, but none more prolific or integral to present understandings of technology as Herbert Marcuse and Jurgen Habermas. Constructivism and Marcuse's reductionist theory of technology posited that technology embodied a specific cultural incentive for the alteration and domination of nature. For Marcuse technology was not a single device, or technic as he would say, but "was a social process in which technics proper (that is, the technical apparatus of industry, transportation, communication) is but a partial factor…[humans] are themselves an integral part and factor of technology, not only as the men who invent or attend to machinery but also as the social groups which direct its application and utilization" (Marcuse 65). As such, it operates as a broad ranging criticism of technology's allegiance to capitalist industry and culture as the main generator of its' authoritarian and reductive characteristics.

Martin Heidegger's substantivist critique of modernity is also a key influence behind Feenberg's theory of instrumentalization. Heidegger, like Marcuse, was skeptical of technology and the implications of technological usage. Unlike Marcuse however, Heidegger believed that technology was not only a phenomenon of power but also that of enframement. In Heidegger's view what is most important about technology was not its inherent technical utility but its essence. Heidegger uses the term *poises*, or 'revealing', to describe the process by which technology presents new social assemblages or truths. Instead of attempting to gauge a technology's technical application from its real-world applications, it was more important to achieve a revealing of the essence of technology's being. Heidegger calls this essence of technology gestell, or enframing. "Enframing means the gathering together of that setting-upon which sets upon man i.e., challenges him forth, to reveal the real, in the mode of ordering, as standing reserve" (Heidegger 10). When we view technology as enframing, we come to understand that technology is a force by which things are made open, and ultimately by which we come to understand and make sense of the world around us.

"Revealing that holds sway through modern technology […] is […] a challenging […] which puts to nature the unreasonable demand that it supply energy which can be extracted and stored" (Heidegger, 6). Unlike the revealing of early rudimentary technologies, such as the fishing rod or spade, modern technology does not reveal in the same manner as poiesis. The primary objective difference between older forms of technology versus those presently is that our present technology is based on modern physics as an exact science. On an essential level, however, the difference in the revealing that occurs in modern technological usage challenges the natural world to supply it with energy which can be stored for later retrieval and use.

As modern technology reveals nature is reduced to a standing reserve of potential energy and material for human consumption. However, as the process of revealing is never ending, according to Heidegger, the challenging put forth to nature by contemporary revealing is also put forward unto us. We are challenged as users of technology to envision nature as an untapped resource, awaiting conversion into standing reserve. By virtue of this technologically dominated relationship between humanity and nature, we too are situated as standing reserve in society. Heidegger notes that modern technology challenges us to view nature as an object of research to reveal, or simply something to order into a standing reserve, which occurs through enframing. Heidegger's concept of enframing is an integral ontological process that Feenberg incorporates into his instrumentalization theory. When we are engaged in an enframed relationship with technology and reality, we are also engaging in what Heidegger defines as a process of disclosure. "According to Heidegger, our nature as human beings is to be world disclosers. This means that we open, by means of our equipment and coordinated practices, coherent, distinct contexts or worlds in which we perceive, act, and think" (White and Searle, 131). In other words, new forms of technology disclose new horizons of thinking, new assemblages for interacting with the world, each other and the self. If the world disclosed to us through machine learning algorithms on social media is an enframed world, what elements of our natural socialization processes are currently being converted into standing reserve? Andrew Feenberg's instrumentalization theory provides a framework in which to examine how ubiquitous digital technologies, such as machine learning algorithms, disclose, reveal and enframe human beings into a new relationship with digital technology.

The Dual Ethics of Technology

As Feenberg writes, "the debate between Marcuse and Habermas over technology marked a significant turning point in the history of the Frankfurt School. After the 1960's Habermas's influence grew as Marcuse's declined and Critical Theory adopted a far less utopian stance" (Feenberg 45). It is here that Habermas, a late-coming philosophical member of Marcuse's same Frankfurt School, provides Critical Theory with a counterpoint to constructivist thought, the view that technology holds an inherently reductionist or enframing characteristic. In Habermas's view, the technical control of nature is a genuine species-wide interest for humans, an interest with no ties to any singular cultural or economic feature. When Feenberg views the state of the debate he observes that "while much of Habermas's argument remains persuasive, his defense of modernity now seems to concede far too much to claims of autonomous technology" (Feenberg 45). In Habermasian communicative rationality theory, technology operates not as an artificial technic subject to the perpetuation of ideological control, as Marcuse would attest, but rather an ideology. According to Habermas, through a reduction of questions of what a good, well-lived life is to technical concerns for experts, contemporary elites eliminate the need for a public democratic discourse of values, thereby depoliticizing them. In this view, technology only operates as a veil to mask the value-laden nature of government decision making (Habermas 83).

Marcuse, Heidegger, Habermas and Marx all contributed to Feenberg's understanding and valuing of positivistic and socially constraining features of technology. Feenberg defines a good society as one which "enlarges the personal freedom of its members while enabling them to participate effectively in a widening range of public activities. At the highest level, public life involves choices about what it means to be human. Today these choices are increasingly mediated by technical decisions" ("Transforming Technology: A Critical Theory Revisited" 3). As such, the design of technology is an ontological decision fraught with political consequences. Traditional accounts of technology, determinist and instrumentalist, highlight efficiency as the principle of selection which determines a successful or failed technology. In the formation of his instrumentalization theory, Feenberg argues that the intervention of interests and subjective ideologies into technological design does not reduce efficiency but rather biases its' achievement according to a broader social program, which he refers to as a 'technical code'. According to Feenberg, the technical code is the rule in which technologies are realized in a social context with biases reflecting the unequal distribution of social power. Feenberg postulated instrumentalization theory as a means of uniting the insights of substantivist understandings of technology in which technology reduces and enframes natural processes and elements into raw materials for extraction, and the constructivism of contemporary historians and sociologists who argued that technology is nothing more than extension of natural human processes. Feenberg unites the pessimistic distrust of technology's potential from the philosophy of technology, courtesy of Heidegger and Marcuse, and Habermas's accommodating and forgiving interpretation of technological usage in his theory. Instrumentalization theory states that technology must be analyzed at two levels, the level of our original functional relation to reality (Primary) and the level of implementation and design (Secondary). Both the primary and secondary levels of instrumentalization theory contain contingent elements and ontological operations which help distinguish between both analytical spaces. Most importantly, instrumentalization theory makes explicit the dual nature of technological processes in its' deconstruction of the manner in which technology simultaneously constrains material social processes, while also providing access to expedited and efficient digital applications leveraged by new technologies.

Andrew Feenberg's instrumentalization theory is a meta-theory, as it combines Heideggerian, Marcusian and Habermasian social critiques on the philosophy of technology, in addition to insights from case studies related to Science and Technology Studies. At its core, however, instrumentalization theory provides a way to understand how socially dominant values are embedded in technological artifacts following their creation and use. Feenberg refers to the relationship of the social and technical requirements within an arrangement of technical objects as technical code. "A technical code is the realization of an interest or ideology in a technically coherent solution to a problem. Although some technical codes are formulated explicitly by technologists themselves, I am seeking a more general analytic tool that can be applied even in the absence of such formulations" (Feenberg, 52). The technical code, according to Feenberg, determines the manner in which technology enframes and discloses new technical relationships and understandings of reality to the human subject. In other words, the technical code constitutes a new way of interacting and engaging with the world, whether implicitly or explicitly, through technology.

It is the technical code which dictates the behavior and conceptualization of technological processes. It is a product of technological essence and the interaction between human subjects and technical object. It is the final product of the instrumentalization process as a whole. The instrumentalization process occurs on two levels, which Feenberg refers to as primary and secondary. Both levels of instrumentalization contains multiple processes which work to support the analysis of particular technological arrangements. Primary instrumentalization consists of the processes of decontextualization and reduction, described by Heidegger through his revealing and standing reserve respectively and Marcuse's reductionism. Feenberg aptly refers to this decontextualization-reduction effect described in primary instrumentalization as 'de-worlding', while the mediation of ethical and aesthetic designs and the recontextualization of technical subjects and objects are generalized as 'disclosing'. Secondary instrumentalization observes the mediation of ethics and recontextualization of technical objects, as inspired by Marcuse' dialectics. This critical theory on the philosophy of technology reads as having potential analytical applications when applied to present day technologies, namely social media.

Primary Instrumentalization

In primary instrumentalization technological functions are **decontextualized** and **reduced** from everyday life. Later the user is positioned to relate to them. The decontextualization and reduction processes of the primary level all occur under a **distancing effect**, where the function and the subject are reduced for maximum manipulation and control by those who design the technology. When examined through the lens of primary instrumentalization theory, humans are continuously decontextualized through computer usage. This process is more easily grasped when viewed and understood as a version of Heidegger's revealing. "The computer simplifies a full-blown person into a 'user' in order to incorporate him or her into the network. Users are decontextualized in the sense that they are stripped of body and community in front of the terminal and positioned as detached terminal subjects" (Feenberg, 59). It is in the repositioning of people into users in a technical space that in turn reveals new opportunity or affordances for new technical actions. This process can be applied to multiple other technological arrangements as well, to follow the earlier example, such as digital social networks.

The technical decontextualization of the human being into a technical object immediately positions them for what Feenberg describes as ‘distanced control’ or ‘operational autonomy’. “Operational autonomy of management and administration positions them in a technical relation to the world, safe from the consequences of their own actions… it enables them to reproduce the conditions of their own supremacy at each iteration of the technologies they command” (Feenberg, 53). When the process of reduction occurs, a technical object is reduced to a manageable conceptual term in order for it to be easily controlled within the conventions of science and technology in the next level of instrumentalization. It is important, however, at this point, to distinguish Feenberg’s decontextualization-reduction process from the Heideggerian concept of enframing which inspired it. Whereas Heidegger viewed the process of ‘de-worlding’ as something that could permanently alter the contextual valuing of a human being in a world where technology has influence over nature, Feenberg views this process to be impermanent within instrumentalization theory. “Primary instrumentalization involves decontextualization, which shatters pre-existing natural arrangements. Of course, no decontextualization can be absolute. The process is always conditioned by secondary instrumentalizations which offer a partial recontextualization of the object in terms of various technical and social requirements” (Feenberg, 57). Thus, Feenberg’s theory creates an opportunity for an escape from the trappings of Heideggerian determinism and the inevitable destiny of human conversion into standing reserve. Instrumentalization theory is flexible and dynamic, it enables us to conceptualize a way out of our current technological paradigm. A paradigm where the technical code that rules the behavior and unspoken laws of technology and men can be reevaluated and rewritten. Whereas Heideggerian notions of technology removes human will from any possible solution to the issue of ethical technological usage, Feenberg reinvigorates the power of human will and control by highlighting the continued and ever-present significance of human engineers on the process of ethical decision-making in technological design and implementation. While technology certainly contains transformative metaphysical and metacognitive properties, it still exists as a mechanism of the human will. As such, it is incumbent on us as humans who are subservient to multiple technical codes to reexamine the manner, we use technology to relate to one another, our reality, and ourselves.

In his Critical Theory of Technology, Feenberg cites the German American philosopher Albert Borgmann in order to establish how 'de-worlding' occurs in digital environments. Borgmann states "that computer networks de-world the person, reducing human beings to a flow of data the 'user' can easily control" (Feenberg, 59). This Heideggerian-esque reduction which occurs within Feenberg's primary level is indicative of the hierarchical relationship between the technical object (e.g. the user/consumer) and the technical mediator. When examined via primary instrumentalization, the interactions and content generation of users with accounts on applications such as Facebook are reduced to meta-data for machine learning algorithms to incorporate into their data set. At this stage in our socio-technical analysis, the manner in which our data is represented and respected would fall to engineers and network administrators, which still holds true generally, however machine learning algorithms have come to represent a 'buffer' technical mediator of sorts in social media. The issue of how machine learning algorithms are implemented to collect, organize and classify the information they incorporate into their data sets, not to mention the manner in which that data informs the behavior and bias of said algorithms, are of great ethical relevance in a primary instrumentalization analysis.

Machine learning in social media infrastructure parallels instrumentalized 'de-worlding' when examining the manner in which Facebook user's meta-data is mined and utilized to inform algorithmic functions and processes. When Facebook's DeepFace Convolutional Neural Network algorithm engages in facial recognition for an image or video, an individual's very face is reduced to a series of data points to be incorporated into another algorithmic data set. When explained through instrumentalization theory, Feenberg explains that within this process the "two levels are analytically distinguished. No matter how abstract the affordances identified at the primary level, they carry social content from the secondary level in the elementary contingencies of a particular approach to the materials" (Feenberg, 50). This description of the process of instrumentalization demonstrates how interconnected the dual primary and secondary instrumentalization levels are, as well as the how they represent the recursive connectivity of ethical design implementation and society.

Secondary Instrumentalization

In secondary instrumentalization the focus lies on the social, political and cultural forces which influence design choices. This analytical level addresses primary instrumentalization by **systematizing**, technically incorporating, the reduced functions, whereby decontextualized technical objects are combined with each other and re-embedded in the natural environment. Technical objects can then be **mediated** by Actors (designers) for aesthetic and ethical considerations ("A Critical Theory of Technology" 50). Feenberg utilizes the example of the cutting of a tree to create lumber to further highlight the manner in which primary and secondary instrumentalization inform each other's socio-technical ethical constraints. "Cutting down a tree to make lumber and building a house with it are not the primary and secondary instrumentalizations respectively. Cutting down a tree 'decontextualizes' it but in line with various technical, legal and aesthetic considerations determining what trees can become lumber of what size and shape and are salable as such" (Feenberg, 50). Let's expand on this, prior to a tree being cut down or **decontextualized** the loggers must address numerous social and technical requirements imposed on them by their corporate employers and political leaders. Laws such as the U.S. Lacey Act imposes or **mediates** strict limitations on what types of trees can be logged. This mediation has been in turn informed by subsequent decontextualization processes. Illegal logging (primary instrumentalization) informs the prohibition of logging certain species of trees (secondary instrumentalization), which in turn informs the new decontextualization practices of the loggers who are now performing an adjusted primary instrumentalization.

Regardless of how abstract any of the affordances which are decontextualized and reduced at the primary level are, they still carry social content from the secondary level through their basic technical provisions for unforeseen circumstances. Within secondary instrumentalization, the higher-level processes of mediation and recontextualization are implemented in order to utilize the technical affordances identified in the primary level in what Feenberg defines as 'disclosing'. Borrowing once again from Heidegger's revealing, Feenberg's disclosing operates in much of the same way. It "is a complementary process of realization which qualifies the original functionalization by orienting it toward a new world involving those same objects and subjects" (Feenberg, 50). It is during the process of 'disclosing' that Feenberg's Marcusian concern with the democratizing power of technology comes into play. Marcuse's dialectic critical theory on technology can be utilized to understand both Feenberg's attitude towards the democratizing force of technology and its capability to generate technology inscribed with the dominant values of those who control the functional processes of the technical arrangement. While primary instrumentalization certainly does reduce people into technical objects, it is through the constraints inherent within the technology and the various modern cultures that utilize them that those in positions of operational autonomy are informed on the possibilities for ethical and aesthetic mediation prior to recontextualization. It is useful now to return to the real-world example of the computer network, as a means of examining disclosing. As the process of decontextualization occurs, simultaneously "a highly simplified world is disclosed to the user which is open to the initiatives of rational consumers. They are called to exercise choice in the world" (Feenberg, 59).

It is evident when viewed through the lens of secondary instrumentalization that the primary generator of the technical code which governs the behavior of any machine learning model employed on social media are the corporate actors and technical operators who oversee Facebook's overall data infrastructure. In 'How Facebook Scales Machine Learning', Jamal Robinson demonstrates the manner in which Facebook systematizes and mediates the social content that has been reduced by machine learning on the primary level. Robinson, an enterprise technologist in practice, breaks down the manner in which machine learning algorithms are scaled to leverage specific processes, such as the Facebook Newsfeed and Facial Recognition. His deconstruction is taken from a presentation made by Yangqing Jia, Facebook's Director of Artificial Intelligence Infrastructure. Facebook relies on three categorical pillars when creating and deploying machine learning models: Frameworks, Platforms, and Infrastructure. Frameworks are needed to create, migrate and train machine learning models. Platforms are utilized for model deployment and management. Finally, the infrastructure represents the hardware that is needed to compute machine learning workloads and store data for future algorithmic processes. Here Robinson helps us establish that Facebook's approach to mediation and systematization is directly informed by the specific software and hardware limitations that come with maintaining and storing a world's worth of meta-data.

Unlike Feenberg's conception of mediation, which implies the importance of the role of a specific engineer in how a technological product embodies secondary instrumentalization, in Facebook it is the interplay between the frameworks, platforms and infrastructure that determines how this secondary process occurs. As such, if we wish to achieve a more conscious awareness of the role of mediation in Facebook's use of machine learning we must pay attention to how the interplay between frameworks, platforms and infrastructure inform the 'technical code', or imperative, of machine learning models. The secondary instrumentalization of Facebook's use of machine learning is further complicated by its use of an orchestration engine. An orchestration engine is the technical process of running multiple applications or services to automate a process. Facebook's orchestration engine is FBLearner and is Facebook Inc's primary method of managing its' multiple machine learning models which can be further broken down into 3 components: the FeatureStore, FB Learner Flow, FB Learner Predictor. In Robinson's description, the FeatureStore is used for data manipulation and storing machine learning features, where features represent an individual measurable property or characteristic from observed data (Robinson 2019). This classificatory distinction between data and features in the machine learning scaling infrastructure is key in trying to gain an appreciation for the manner in which M.L. works with and utilizes user-information, a technical process that contains a significant overlap with Feenberg's notion of 'technical code.' FB LearnerFlow manages the specific workflow processes required during the training of machine learning models, as well as determining what specific hardware requests will be needed in order to deploy specific models on Facebook. FB LearnerFlow is what enables the easy reuse of algorithms for different products and can scale them to run thousands of processes simultaneously. The FB LearnerPredictor component of Facebook's orchestration engine is used to serve the models that other applications use to make inferences against it. In other words, it is what enables Facebook's algorithms to predict what a user may or may not do, through inference of the collected data and features. When viewed from a macro-perspective, Feenberg's notion of systematization and mediation has become automated through Facebook's orchestration engine. The primary role of human engineers in the maintenance of Facebook's instrumental systematization and mediation processes appears in the construction of Facebook's material Framework, use of Platforms and Workflow Infrastructure. By automating the mediation of algorithm deployment and efficacy machine learning algorithms have usurped humans as technical mediators and subsequently perform the secondary instrumentalization that builds Facebook's technical code. These kinds of artificially mediated technical codes have proven to be harmful and pervasive throughout not only social media, but important public service programs as well. This understanding of the infrastructure of machine learning algorithms on social media can serve as a discursive space to discuss the ways in which instrumentalization theory describes the ethical and technical elements that are essential in ML implementation.

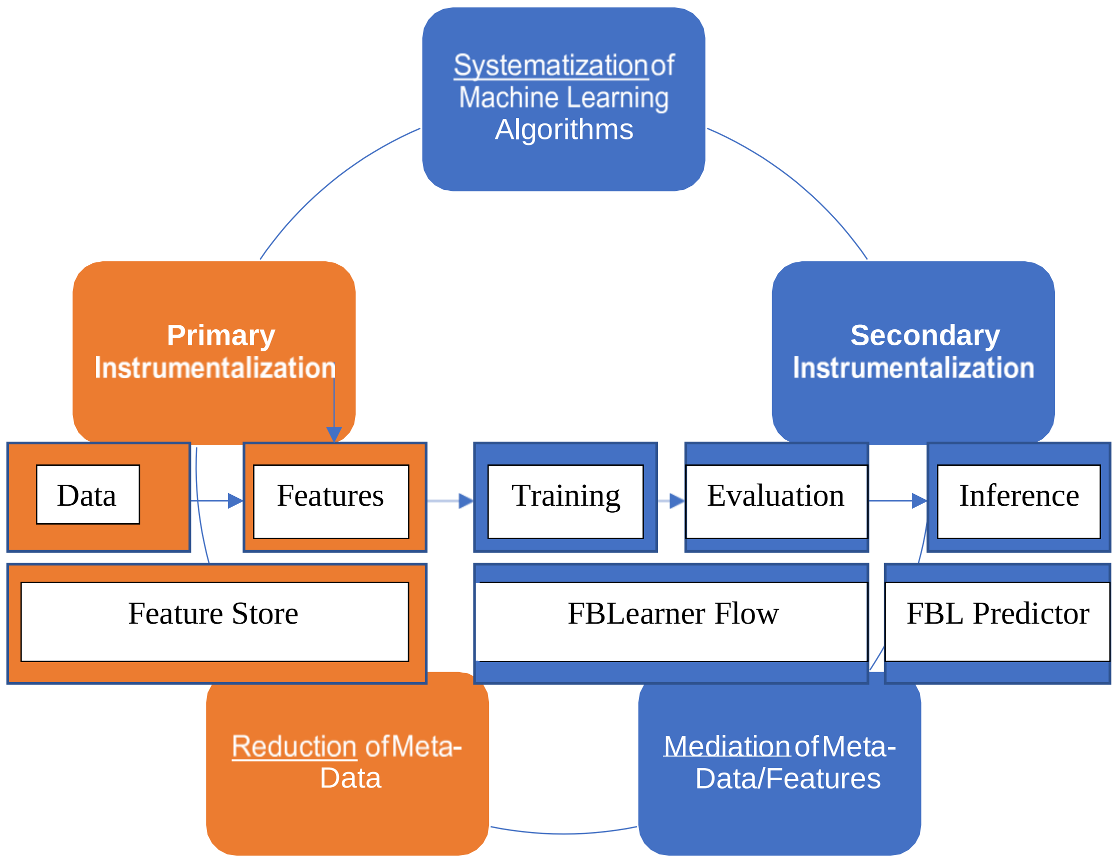

Figure 1 Instrumentalization Process in Facebook's Machine Learning Orchestration Engine FBLearner

Section 2: Machine Learning: Ubiquity and Socio-Technical Opacity

Machine learning relies on an automated process that extracts patterns from data and then models that pattern to generate predictive decisions from data present in any available information set. Supervised machine learning techniques automatically learn a model of the relationship between a set of descriptive features and a target feature based on a set of historical examples, or instances. This 'model' can then be used to make further predictions for new instances (Kelleher et. al. 3). Facebook uses machine learning algorithms for classification, ranking, and content understanding devices. According to Hazelwood, et al. in 'Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective', the applications leveraged by machine learning include but are not limited to the Facebook Newsfeed, Serving Advertisements, Search Functions, Classifying Objects, Facial Recognition, and Language Translation. The machine learning algorithms used by Facebook to enable these applications to function properly include Multi-Layer Perceptrons (MLP), Support Vector Machines (SVM), Gradient Boosted Decision Trees (GBDT), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) (Hazelwood et al. 3).Facebook's machine-learning artificial intelligence 'ecosystem', as Hazelwood describes it, can be categorized into 3 primary sections: frameworks, platforms, and infrastructure (Hazelwood et al. 1). Hazelwood, et al., expand on Robinson's earlier explanation of the Facebook machine learning infrastructure in their explanation of the Facebook 'ecosystem'. Frameworks are needed to create, migrate and train machine learning models, while platforms are used for model deployment and management. Infrastructure, as Robinson demonstrated earlier, is needed to compute workloads and store data.

Facebook was selected from among all the various social media platforms to apply Feenberg's instrumentalization due to its ubiquity in modern culture, in addition to its historic significance in popularizing the brand of interactive computer-mediated technologies. What would appear to be the inherent technical essence, to borrow from Heidegger, of social media is the facilitation of the creation and sharing of textual, visual and audio-visual information, ideas, career interests and multiple other forms of expression via generated communities and networks. However, this is only true within the context of the general Facebook user. When viewed from a macro-perspective, a second essence or technical code reveals itself in the form of meta-data acquisition, used for storage and retrieval for machine learning inference. The existence of a dual technical code in Facebook's use of machine learning, parallels Feenberg's dual-level primary and secondary instrumentalization processes. As such, it makes sense to analyze Facebook's page ranking algorithm and how it is utilized to value and prioritize content for members of its computer-mediated communities through instrumentalization theory because of its global presence and high-level of influence on the political and social perspectives of its users. Facebook has also found itself at the center of multiple allegations of the misuse of user-information. In 2016 Facebook delivered the data of approximately 87 million Facebook users to Cambridge Analytica, a British consulting firm. As J. Isaak and Mina J. Hannah declare in their article, 'User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection', the delivery of "personally identifiable information (PII) of more than 87 million unsuspecting Facebook users to Cambridge Analytica has fueled growing interest in the debate over technology's societal impact and risks to citizens' privacy and well-being. It is clear that national government institutions demonstrably lack the ability to anticipate technology's future impact on the rights and duties of its citizens, much less its' impact on the structure of society, ideological divides, and political schisms among its citizens and the expansion of identity politics promoted by isolated social and news gathering echo chambers" (Isaak and Hannah, 56). Using instrumentalization theory to examine the technical code of Facebook's use of machine learning algorithms is useful, in order to highlight the need for a greater social awareness of digital infrastructures and the power they hold over us when we interact with them. Unlike pre-Web 2.0 technologies which Feenberg originally described in his theorizations of instrumentalization in the early '90s, Facebook is in and of itself a perpetual and continuing exemplification of the instrumentalization framework.

This form of analysis is also pertinent and relevant from another point of social exigency. As Safiya Umoja Noble discusses in *Algorithms of Oppression*, "The near ubiquitous use of algorithmically driven software, both visible and invisible to everyday people, demands a closer inspection of what values are prioritized in such automated decision-making systems" (Noble, 16). In her description of how the Google search algorithm framework generated sexualized content as the top results for African American Women, Noble demonstrates the manner in which machine learning algorithms operating within Google's search engine mediated and compounded the generation of these racist results. This distanced control of social perceptions vis a vis search results demonstrates the pervasive influence of machine learning across digital spaces. "What each of these searches represents are Google's algorithmic conceptualizations of a variety of people and ideas. Whether looking for autosuggestions or answers to various questions or looking for notions of what is beautiful or what a professor may look like… Google's dominant narratives reflect the kinds of hegemonic frameworks and notions that are often resisted by women and people of color" (Noble, 40). Noble and Feenberg are aligned in their understanding that technological arrangements of all kinds must be challenged to mitigate the creation of persecutory technological outcomes. Feenberg's stance embodies the philosophical and conscientious approach that individuals must engage in to liberate themselves. Whereas Noble's call for increased regulations of internet activity, citing the Cyber Civil Rights Initiative's ruling on the distribution of non-consensual pornography, represents a much more grounded solution. Feenberg states that it is through a democratic consideration for the technical processes of any technical arrangement that issues of oppression in those arrangements can be addressed. In regard to how we should address issues of technological marginalization through policy, I agree with Noble's approach wholeheartedly. However, any policy that is created to address the issue of technological persecution and marginalization must also be supported by flexible theories which enable us to begin to conceptualize solutions for the unique technical conditions present in every digital space. When viewing Noble's analysis of Google's use of machine learning, instrumentalization theory suggests that machine learning algorithms contained the recursive processes of de-worlding and disclosing of web pages for user interaction, albeit in a manner which concretizes harmful patriarchal offline biases.

Virginia Eubank's *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, examines the manner in which automated decision-making technology disproportionately impacts the lives of poor working-class households in the United States. Eubanks begins her analysis by detailing her personal experience being red flagged by an automated algorithm following her husband's intensive hospitalization. According to Eubanks testimony, she was only able to gauge what the possible cause to her family's red flagging was due to her educational background and the economic resources she and her partner had access to.

Eubanks uses her personal experience to highlight that while instances of such problematic automated red-flagging occasionally break through into the middle-class, the consequences of this reliance on algorithms and other automated technologies impact the life possibilities of poor working-class Americans. Eubanks supports this claim by detailing case studies exemplifying this new form of social marginalization. A prime case comes in the form of an examination of the manipulative political practices of Main Republican Governor Paul LePage and his use of automated eligibility to track the purchase behaviors and patterns of TANF (Temporary Assistance for Needy Families) recipients by tracking their EBT expenditures. LePage managed to use the suspicious monetary activity of 0.03% of Maine TANF recipients to disparage the remaining 99.97%, and pass legislation which required the submission of all cash receipts to expedite State audits. "Since the Great Recession, my concern about the impacts of high-tech tools on poor and working-class communities has increased. The skyrocketing economic insecurity of the last decade has been accompanied by an equally rapid rise of data-based technologies in public services: predictive algorithms, risk models, and automated eligibility systems. Massive investments in data-driven administration of public programs are rationalized by a call for efficiency. Technologies of poverty management are not neutral" (Eubanks, 9) Since then, the American government has massively overhauled its use of automated technology as it relates to the security, health insurance coverage, access to social services and beyond of low wage working class families and minority groups.

In 'Dissecting Racial Bias in an Algorithm Used to Measure the Health of Populations' Obermeyer et. al. found evidence of racial bias in an algorithm that was being used to measure the health of populations, as well as their eligibility for insurance in the United Kingdom. The algorithm would make it so that Black patients assigned the same level of risk would be classified as sicker than White patients. Obermeyer et. al concluded that algorithmic bias reduced the number of Black patients identified for extra care by more than half. The reason for this bias was due to the algorithm's use of low health costs as a proxy for health needs, in short it conflated health costs with health needs. As a result, less money was spent on Black patients who had the same level of need, and the algorithm falsely concluded that Black patients were healthier than equally sick White patients. (Obermeyer et. al., 421).

Each of these case studies highlights the global reach of machine learning algorithms, as well as the devastating real-world impacts of their usage. The general theme of this historical turn towards machine learning automation, is that companies and government agencies will always opt for the most expedient technical solution for longstanding social problems. In the words ascribed to George Santayana, "those who do not learn history are often doomed to repeat it." In the case of machine learning, perhaps not so much a lesson unlearned as a lesson not heeded. Deferment to expedient technical solutions versus ethical ones, reveals yet another need for a reorientation of our relationship to automation and exponential innovation.

The Functional Ethic of Machine Learning Algorithms

Arlindo Oliveira in *the Digital Mind* provides a simplified description of Machine Learning as it relates to artificial intelligence which is helpful in contextualizing ML models as inherently neutral technology in regard to their function. According to Oliveira, the quintessential problem of machine learning consists, in its most simplified form, of inferring general rules or behaviors from a number of specific concrete experiences. This general process is referred to as inductive learning. Inductive learning is performed by learning a general rule from a set of labeled instances. Oliveira provides multiple practical examples and analogies to describe different ML models, such as GBDT and K-Nearest Neighbor, methodologically referred to as similarity-based learning. Oliveira cites David Hume, and his critique of induction-based learning in order to critique the inference methods of machine learning models that are currently being deployed in various industries, including social media. Hume points out that induction from past experience cannot provide guaranteed results, as our ability to learn from experience exists only because there is some regularity in the data that we are able to explore. Oliveira refers to this inductive bias as the generator of the variety of different algorithmic machine learning models we see today. Oliveira goes on to demonstrate that all learning algorithms have the same inherent outcomes when their respective performances are averaged across all possible problems in a given domain. Commonly referred to as the *No Free Lunch Theorem,* it describes the fact that all learning algorithms are equally good if a preference for a specific learning bias can't be established. When incorporating the *No Free Lunch Theorem* into a rhetorical analysis of the functional ethics of machine learning algorithms one comes to understand that no one model is more ethical or unethical in its implementation than another. As such, the inherent functional ethic of machine learning algorithms when employed in public production environments is one of general neutrality. This perceived neutral functional ethic in turn lends itself to obfuscating secondary ad hoc and corporate technical ethics which motivate their purposes, and also highlights the fact that any and all algorithms are susceptible to a dominating technical code which holds the potential to perpetuate obfuscation indefinitely. As we established in the previous sections, the technical code that comes to dominate the neutral function of machine learning comes in the form of the digital infrastructure which models are deployed into.

Safiya Umoja Noble demonstrates how machine learning directs social media in problematic ways when she describes the manner in which search results, managed and produced by machine-learning algorithms have the tendency of generating disparaging and racially insensitive results. When examining Noble's critique of Google within the context of the *No Free Lunch Theorem* we come to realize that while machine learning was certainly the vehicle by which racist and sexually explicit results were generated, one cannot ascribe the originating power of the technical code to it. While Noble describes this form of technological social marginalization within the context of identity and social inclusion, it hits at the core of what Feenberg has set out to describe and correct in his creation of the instrumentalization (primary and secondary) analytical heuristics. The lack of consideration for the social impact of the technological design and implementation of automated technologies, highlights the need for further consideration of the social and dialectic variables that are at play when technical design decisions are being made (secondary instrumentalization). Noble's analysis highlights what can go wrong when decisions at the secondary level of technological instrumentalization impact the reality of its users at the primary level (Noble 64). Similarly, in 'Racist in the Machine', cyber security specialist Megan Garcia highlights the need for companies and governments to "pay attention to the unconscious and institutional biases that seep into their algorithms" (Garcia 115). Garcia notes that distorted data can skew results in web searches, home loan decisions, or photo recognition software. This analysis of machine learning challenges Oliveira's claims of total technological neutrality, as well as our conceptions of what it means for digital technology to work 'well'. Machine learning works in the sense that it gets any singular job done, but an algorithm has no sense of social liberty or cultural history. It does not care if a minority family can't make EBT withdrawals following a red flagging. An algorithm has no country, nor does it hold national interests or even domestic ones. It does not care if a politician disseminates lies and falsehoods so as to cling onto power.

In addition to various accounts of algorithmic design misuse levied against machine learning, DeepFake facial modification algorithms being a prime example, the controversy surrounding machine learning's use also extends to targeted advertisements and the ethics of corporate data mining. When informed by Noble's social analysis and the No Free Lunch Theorem, Feenberg's instrumentalization theory suggests that machine learning algorithms focus the recursive and simultaneous processes of de-worlding and disclosing within the web pages they are employed and meta data inscriptions from user-generated content. This can occur in a manner which can concretize harmful patriarchal offline biases when socio-technical design concerns are not addressed. In other words, it is the algorithmic bias of the machine learning model which informs the technical and ethical constraints of the production environment on the secondary level of instrumentalization. When the algorithmic bias of a particular machine learning algorithm is informed by poor data, the implications related to its usage when it is deployed into a production environment are directly tied to the specific engineers and designers who designed the digital infrastructure in which it was deployed. As explained in the previous section, the methodology for data acquisition and implementation stems from the infrastructure which guides and maintains the process of data storage, use and inference. When viewed via instrumentalization theory, questions pertaining to the manner in which executive decisions related to machine learning algorithms ability to mine user data reveal the manner in which technical problems are relegated to corporate control and ethical mediation. As such giving more attention to the ethical mediation, or lack thereof, of the impact of the algorithmic bias of machine learning algorithms on the data of Facebook users, as well as the countless other public services leveraged by machine learning, must become a central concern of American Science and Technology policy going forward.

The Socio-Technical Opacity of Machine Learning in Social Media and the Obfuscation of Algorithmic Bias

The last few sections have examined the manner in which algorithmic processes and the infrastructure that enable their deployment to constitute a socio-technical opacity, but the question obviously arises, so what? Surely, human beings have successfully engaged with black-boxed technology in the past without severe social repercussions. This is true to an extent, but it would be a grave mistake to compare present digital technologies to other technological innovations of the 20th century, like the telephone or radio. With those technologies, the technical code is more or less easily discerned in the immediacy of the technical product, in the case of the phone and radio it is the successful audio transmission of a voice from one location to another. However, the opacity present in machine learning masks its' dual technical codes, one informed by the users it interacts with and assists, and the other by its creators and the infrastructure it exists in. Opacity creates the illusion of technological neutrality and fogs human understanding of the influence technology has on the development of how they interact with the world. In a world that is becoming increasingly constituted by automated technologies, to lack an appropriate understanding of their functions is akin to lacking basic awareness of your own human rights. The persistence of socio-technical opacity in relation to automated technologies, machine learning algorithms in particular, erodes and effaces the democratic potential and promise of new digital technologies.

While Noble and Garcia establish an exigency for the creation of a more clear-cut discourse on machine learning models in social media, Dr. Jenna Burrell introduces us to the concept of opacity in machine learning implementation. Machine learning algorithms "are opaque in the sense that if one is a recipient of the output of the algorithm (the classification decision), rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs. Additionally, the inputs themselves may be entirely unknown or known only partially. What are the reasons for this state of not knowing? Is it because the algorithm is proprietary? Because it is complex or highly technical? Or are there perhaps, other reasons?" (Burell, 1). Burrell distinguishes between three different forms of socio-technical opacity present in machine learning algorithm models across all technological industries and potential implementations. The first form of opacity relates to opacity as intentional corporate or state secrecy, particularly when they engage in the use of algorithms for ranking, recommending, trending, and filtering, which is often used to secure attention from the public. The second form of opacity that Burrell deconstructs is opacity as technical illiteracy, which stems from the fact that understanding code and the design of algorithms is a specialized skill. This knowledge continues to remain out of reach of the majority of the population. Code used in programming languages, such as Python, have a particular syntax and language that must be learned. These languages follow logic-based rules that require exactness in spelling, grammar, and sequencing for an algorithm to function properly. According to Burrell "good code does double-duty. It must be interpretable by humans (the original programmer or someone adding to or maintaining the code) as well as by the computational device. Writing for the computational device demands a special exactness, formality, and completeness that communication via human languages does not" (Burrell 4). The third and final form of opacity reflects opacity as the way algorithms operate at the scale of their application. Burrell argues that there are particular "challenges of scale and complexity that are distinctive to machine learning algorithms" (Burrell 5).

As Burrell notes in her article, opacity is tied not only to the total number of lines or pages of code but to other material conditions, such as the number of team members on the engineering team and the number of interlinkages between modules or sub-routines. (Burrell 4). Burrell observes that this poses further challenges to both scholars and industry professionals who wish to understand what exactly occurs when machine learning algorithms are deployed in various online services. According to Burrell, “though a machine learning algorithm can be implemented simply in such a way that its logic is almost fully comprehensible, in practice, such an instance is unlikely to be particularly useful. M.L. models that prove useful (specifically, in terms of the ‘accuracy’ of classification) possess a degree of unavoidable complexity” (Burrell 5). Burrell points out that it is the *heterogenous* characteristics of the data (features) which adds complexity to the code. The interplay between large data sets and the code in the mechanism of the algorithm is ultimately what yields this analytical complexity (which compounds opacity). Burrell goes on to provide examples to better understand this complexity and provides suggestions for overcoming the barriers to opacity by means of visualizing neural networks, in addition to inspecting the opacity of spam filtering to examine classificatory discrimination. Evaluating the manner in which the socio-technical opacity of machine learning algorithms in social media constrains social processes and technological understanding requires an understanding of the material infrastructure that designers and engineers at companies like Facebook Inc., are themselves constrained by when producing machine learning algorithms.

Section 3: An Instrumentalized Analysis of Machine Learning & the Need for a Heightened Socio-Technical Awareness in Science and Technology Policy

We would like to believe that social media is a perfectly neutral networking tool to maintain communication with our friends, loved ones and the world at large, however, scholars from the Frankfurt School and theorists inspired by them, like Feenberg, would disagree as to the existence of a truly neutral technological arrangement. "Neutrality generally refers to the indifference of a specific means to the range of possible ends it can serve … There is no such thing as technology as such. Today we employ this specific technology with limitations that are due not only to the state of our knowledge but also to the power structures that bias this knowledge and its applications. This really existing contemporary technology favors specific ends and obstructs others" (Feenberg, 182). According to Lars Backstrom, the Engineering Manager for NewsFeed Ranking at Facebook, there are "as many as 100,000 individual weights that produce NewsFeed. The three original EdgeRank elements- Affinity, Weight, Time Decay- are still factors in NewsFeed ranking, but other things are equally important" (McGee, 2014). Outside of the standard Affinity, Weight and Time Decay weight factors, the new algorithm also considers a user's relationship settings, post types, spam reporting, network exploration, device considerations, and story bumping.

Facebook Inc.'s Technical Code

Facebook has experienced multiple iterations of its page ranking algorithm but for the purposes of this analysis shall be limited to EdgeRank and its current machine learning based algorithm. The earlier iteration, EdgeRank, collected, organized, and ranked undiscovered content based on three elements: affinity (the proximity of content to the technical user), weight (the action a user took when interacting with content) and time decay (which values newer content over pre-existing content). "Every item that shows up on the NewsFeed is considered an object. If you have an object in the NewsFeed (say a status update), whenever another user interacts with that object, they are creating an edge, which includes actions like tags and comments" (Kincaid, 2010). When each edge factor was multiplied by the other, the resulting value represented the contents' relevancy score represented algorithmically as $\sum \mu w d$. Affinity, weight, and decay were all presented to the user via their content interface. It is in this way that the EdgeRank algorithm incorporates Feenberg's primary and secondary instrumentalizations simultaneously, much like the tree being cut for wood the algorithmically generated content on the NewsFeed has already been mediated for worth. When a user creates an edge factor through a de-worlding interaction (primary) the content is at the same time undergoing a new disclosing when the EdgeRank algorithm recalculates the content's relevancy score (secondary).

This speaks to the presence of Facebook's public facing technical code, where machine learning algorithms are utilized to enhance the user's experience on social media, through improvements in user to user and user to content interactions. However, as we've already established, Facebook's use of machine learning models underscores a second technical code. A technical code that is informed by Facebook's artificial intelligence infrastructure. Burrell's analysis and explanation of opacity in machine learning model deployment hits at the essence of this secondary technical code. A technical code that is supported by an infrastructure designed to harvest terabytes of personally identifiable information and meta-data for process enhancement and vendor access to user data. However, providing ease of access to our personal information for outside-third parties is merely an extremely powerful side-effect of machine learning's true technical code. When examined through instrumentalization theory, the functional purpose of machine learning is obfuscated by the technical code that situates the Facebook user as a confused and ignorant resource. Facebook's second code does not assist socio-technical opacity, rather it is quite literally the essential embodiment of opacity. Facebook relies on an incredibly opaque understanding of the function of machine learning models inputs and outputs to secure their control of all socialization processes on the network. As Burrell notes, the primary motivation behind this technical code's valuing of opacity stems from a combination of technical opacity related to the high calculus of machine learning, in addition to the opacity generated by scaling ML models to Facebook's massive infrastructure. This is not to say that Facebook does not have a responsibility to attempt to clear this opacity, rather that it highlights an intense vulnerability that vendors of ill-intent can manipulate for their own political and monetary gain.

What Ethical Concerns Surrounding Machine Learning Implementation Should American Science and Technology Policy Consider?

Because of its' capacity to highlight both the problematic and positive ethics of technology, I believe Feenberg's theory of instrumentalization can serve as the theoretical foundation for future considerations for Science and Technology policy related to machine learning and other ubiquitous socialization technologies with socio-reductive characteristics. More than 3 years after H.R. 5051, the OPEN Government Data Act, attempted to create a means to publish all non-federally restricted data in an open source format in order to produce a standardized use of big data for both federal and public use, we have now entered an age where private companies are able to dictate the rules of information management and metadata commodification through machine learning. Since the 2016 American Presidential election, which saw one of the most blatant violations of consumer privacy in internet history when the personally identifiable information of 87 million Facebook users was made available to Cambridge Analytica, not to mention the use of Facebook's vulnerable infrastructure by Russian hackers to disseminate political and social falsehoods, questions about the integrity of America's digital society remain unanswered. While Cambridge Analytica was dissolved in 2018 for allegations of bribery and sexual "honey-potting" on the behalf of their clients, looking at Facebook today, it is not clear if attempts made by Mark Zuckerberg or his Director of Artificial Intelligence have solved the issue of information privacy. After all, Facebook does not sell personal information but rather makes our personal information available to outside vendors. If Facebook's automated digital infrastructure provides machine learning with its opaque technical code, then the rules of ethical machine learning use falls to the vendors who are on the receiving end of Facebook's ML driven meta-data acquisitions. This reliance on the supposed ethics of third-party vendors is deeply troubling.

In his examination of Brad Parscale, the digital director for President Donald Trump's 2020 re-election campaign, McKay Coppins of the Atlantic sees no evidence of an ethical approach to targeted use of machine learning. Parscale Media, of which Parscale is the founder, was first noticed by the Trump Organization in 2011. Parscale continued working for the Trump Organization before he started work on the 2016 presidential campaign. Parscale "seemed to have no reservations about the kind of campaign Trump wanted to run. The race-baiting, the immigrant-bashing, the truth-bending- none of it seemed to bother Parscale. While some Republicans wrung their hands over Trump's inflammatory messages, Parscale came up with ideas to more effectively disseminate them." (Coppins). During the general election Parscale's team used Google and Facebook to spread inexpensive ads fomenting notions of the threat of radical Islamic terrorists and Hillary Clinton's criminality. Parscale plans to continue implementing this approach to Trump's 2020 digital ad campaign, and Facebook's technical code of opacity will assist him and his team in doing just that. McKay refers to this opaque shadowy interaction with machine learning on the behalf of political advertisers and vendors as a disinformation architecture. The disinformation architecture of the Trump campaign, and those like it, benefits directly from public misunderstanding of the manner in which their personal data is used for targeted advertisement.

This problem also extends internationally. McKay cites Peter Pomerantsev's book *This is Not Propaganda* and the case of 'P', a Filipino political consultant who utilized targeted advertising to successfully test a disturbing social experiment. As a student in college, 'P' learned of the 'Little Albert' experiment, where a child was psychologically conditioned to fear furry animals after exposing him to a lab rat following a loud noise. 'P' took inspiration from 'Little Albert' and created multiple Facebook groups under the pretense of providing a space for Filipinos to talk about daily events in their communities. However, after the group gathered roughly 100,000 members 'P' "began posting local crime stories and instructed his employees to leave comments falsely tying the grisly headlines to drug cartels. The pages lit up with frightened chatter. Rumors swirled; conspiracy theories metastasized. To many, all crimes became drug crimes" (Coppins). Little did the members of these Facebook groups know that 'P' had designed these groups to boost the popularity of then-aspiring presidential candidate, Rodrigo Duterte whose rise to international power was launched by a stern promise to eliminate drug cartels. Since 2016 Duterte's Presidency has been characterized by hundreds of extrajudicial killings, many of the victims having been drug addicts with no connection to organized crime. This is one of the more extreme examples of how a lack of conscientious awareness of the pervasive influence that an automated digital space has on human behavior effaces not only our social conceptions of truth but our ability to form consensus. These cases of third parties misusing algorithmic targeting degrades our ability to participate effectively in democracy, by extension technical opacity aids in this erosion. From all appearances, and via our analysis of Facebook's general machine learning infrastructure using instrumentalization theory it is clear that it is still just as vulnerable to manipulation by unethical third-party mediators as it was in 2016. This lack of forward momentum in corporate and institutional accountability in regard to machine learning model implementation on social media presents the need for two primary points of public policy to address this concern. Namely, the development of a national regulatory body for machine learning model implementation on social media and other digital service platforms as well as the creation of education programs designed around the understanding and use of coding.

Regulation requires the hard work of examining the technical code of newly emerging machine learning infrastructures and evaluating them through national standardized training and inference processes with diverse sample data sets representing features from meta-data collected from minority groups, different political organizations, and verified and unverified news. This is an adaptation of an earlier attempt by federal agencies to do just this work in other industries. In 2019 the U.S. Food & Drug Administration proposed a regulatory framework for modifications to artificial intelligence software as a medical device [SaMD]. The FDA proposal outlines several guidelines for standardizing the regulation of machine learning algorithms prior to their public implementation. "The FDA's proposed TPLC [Total Product Life Cycle] approach is based on the following general principles that balance the benefits and risks, and provide access to safe and effective AI/ML-based SaMD: 1. Establish clear expectations on quality systems and good ML practices; 2. Conduct premarket review for those SaMD that require premarket submission to demonstrate reasonable assurance of safety and effectiveness and establish clear expectations for manufacturers of AI/ML-based SaMD to continually manage patient risks throughout the lifecycle; 3. Expect manufacturers to monitor the AI/ML device and incorporate a risk management approach and other approaches outlined in 'Deciding When to Submit a 510(k) for a Software Change to an Existing Device' Guidance18 in development, validation, and execution of the algorithm changes (SaMD Pre-Specifications and Algorithm Change Protocol); 4. Enable increased transparency to users and FDA using postmarket real-world performance reporting for maintaining continued assurance of safety and effectiveness" (Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning [AI/ML]-Based Software as a Medical Device [SaMD]). This white paper proposal is a promising step in a potential right direction. It represents a genuine willingness from the Federal government to engage with the prospect of seriously tackling the issue of algorithmic regulation in important service sectors.

The FDA's approach to machine learning regulation provides scholars and policy makers interested in expanding such regulation into a broader American Science and Technology policy related to the use of automated targeting technologies a codified approach to a non-invasive outcome-based regulatory process. The Food & Drug Administration's regulatory policy also emphasizes reconfiguring algorithms that are currently deployed in public services, by adjusting a model's approach to data management and retraining it to provide greater transparency for users, in this case health service providers and patients. Here the FDA's regulatory proposal provides some promise of a possible solution to the perpetuation of disinformation infrastructures compounded by socio-technical opacity. Regulation should and must strive for transparency and work to limit the creation of opaque notions of machine learning model implementation and scaling.

A national regulatory body for machine learning algorithm models designed for social media would have to work to ensure that machine learning processes that are responsible for opaque understandings of the influence of third-party advertisers and organizations, are more transparent for users. Taking more inspiration from the FDA's approach to the regulation of machine learning in software for medical devices, possible best practices for a general regulatory body should include (1) relevance of available data for user to user and user to content interaction, (2) data acquired in a consistent, relevant manner, and generalizable manner that aligns with the algorithms intended use and modification plans, (3) an appropriate separation between training, tuning, and test data sets and (4) an increased emphasis on **transparency** of the output of the algorithm aimed at social media users.

However, the creation of a national regulatory body for the regulation of machine learning algorithm model usage in social media applications is not enough to achieve the desired goal of cultivating a heightened social consciousness about the influence of the technical codes of third-party disinformation infrastructures. Machine learning regulation must be supported by educational initiatives designed around providing children and young adults a clear understanding of the transformative social power of automated technologies and algorithms. Fortunately, there are newly emerging examples of potential curricular approaches on cultivating a general awareness of machine learning and other automated technologies. Blakely H. Payne, a doctoral candidate at the Massachusetts Institute of Technology, has developed a curriculum to teach children about integral concepts such as algorithmic bias and deep learning. "She tested the week-and-a-half long program with about 225 fifth-through-eigth grade students at David E. Williams Middle School in Coraopolis, PA outside Pittsburgh. […] Payne developed the course of study with input from computer science teachers and researchers at the Harvard Graduate School of Education. Her *unplugged* curriculum mainly uses pen, paper and craft supplies so that teachers can adapt it for their classrooms, regardless of budget or technological know-how. Each 45-minute lesson typically includes a short lecture and demonstration, followed by a group activity and open-ended discussion" (Ma). The learning outcomes of Payne's experimental course encourage students to: "(1) understand the basic mechanics of artificial intelligence systems, (2) understand that all technical systems are socio-technical systems and that these systems are not neutral sources information, but instead serve political agendas, (3) recognize that there are many stakeholders in a given socio-technical system and that the system can affect these stakeholders differentially, (4) apply both technical understanding of AI and knowledge of stakeholders in order to determine a just goal for a socio-technical system, (5) consider the impact of technology on the world" (Payne). The learning outcomes for Payne's 'Ethics for Artificial Intelligence Curriculum for Middle Schooler's' address the ontological concerns of machine learning algorithms revealed through Feenberg's instrumentalization theory. In particular, the development of an awareness of social media's reliance on a technical code based in socio-technical opacity. While certain aspects of socio-technical opacity will always exist, such as the opacity related to scaling algorithms for global service platforms, an educational effort is essential in order to address the social impacts of automated technology from the ground-up.

These policy considerations are motivated by rising instances of machine learning misuse via harmful algorithmic bias not only in the social media industry, but health, banking and national security. This dual-policy approach, based on establishing a firm regulatory system for machine learning in addition to a new educational curriculum to inform the public on the dangers of opacity, is the most practical and responsible means of reclaiming the democratic potential of digital spaces and automated technologies. If we are to follow the advice of Andrew Feenberg, and truly aim for the true democratization of technology, then we must engage in the hard and continuous work of aligning policy with the values we espouse in the classroom. Social media's digital infrastructure should not become a dark reflection of humanity's baser instincts and motivations. When viewed through the lens of Feenberg's instrumentalization theory we see spaces where the obfuscating power of machine learning's technical code can be challenged and possibly rectified. It is essential that social scientists, scholars in Science and Technology Studies, educators and politicians address this force that threatens to obfuscate our understanding of current events, the plight of minority populations, and the inscription of social prejudice. As Feenberg writes, “a social account of the essence of technology enlarges democratic concerns to encompass the technical dimension of our lives. It offers an alternative to both the ongoing celebration of technology triumphant and the gloomy Heideggerian prediction of technocultural disaster” (Feenberg, 17). If we truly wish to preserve the integrity of the democratic experiment, then we must also be concerned with the continued and perpetual democratization of technology through progressive policy and education initiatives.

Works Cited

Adorno, Theodor W., and Max Horkheimer. *Dialectic of Enlightenment*. London: Verso, 2016.

Burrell, Jenna. "How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms." *SSRN Electronic Journal*, 2015. https://doi.org/10.2139/ssrn.2660674.

Coppins, McKay. "The Billion-Dollar Disinformation Campaign to Reelect the President." *The Atlantic*, Atlantic Media Company, 7 Feb. 2020, www.theatlantic.com/magazine/archive/2020/03/the-2020-disinformation-war/605530/.

Feenberg, Andrew. *Critical Theory of Technology*. New York: Oxford University Press, 1991.

Feenberg, Andrew. *Alternative Modernity: the Technical Turn in Philosophy and Social Theory*. Berkeley: University of California Press, 2008.

Feenberg, Andrew. "Marcuse or Habermas: Two Critiques of Technology." *Inquiry*39. Accessed July 1, 2019. https://www.sfu.ca/~andrewf/books/Marcuse_or_Habermas_Two_Critiques_of_Technology.pdf.

Garcia, Megan. "Racist In The Machine: The Disturbing Implications Of Algorithmic Bias." *Human Rights Documents Online*, n.d. https://doi.org/10.1163/2210-7975_hrd-1216-2016004.

Habermas, Jürgen. *Toward a Rational Society*. Cambridge: Polity Press.

Hazelwood, Kim, Sarah Bird, David Brooks, Soumith Chintala, Utku Diril, Dmytro Dzhulgakov, Mohamed Fawzy, et al. “Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective.” *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, February 24, 2018. https://doi.org/10.1109/hpca.2018.00059.

Isaak, Jim, and Mina J. Hanna. “User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection.” *Computer*, vol. 51, no. 8, 2018, pp. 56–59., doi:10.1109/mc.2018.3191268.

Johnson, Carolyn Y. “Racial Bias in a Medical Algorithm Favors White Patients over Sicker Black Patients.” *The Washington Post*, WP Company, 25 Oct. 2019, https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/.

Kelleher, John D., Brian MacNamee, and Aoife DArcy. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, MA: MIT Press, 2015.

Kilmer, and Derek. “Related Bills - H.R.5051 - 114th Congress (2015-2016): OPEN Government Data Act.” *Congress.gov*, 26 Apr. 2016, https://www.congress.gov/bill/114th-congress/house-bill/5051/related-bills.

Kincaid, Jason, and Jason Kincaid. “EdgeRank: The Secret Sauce That Makes Facebook's News Feed Tick.” *TechCrunch*, TechCrunch, 23 Apr. 2010, techcrunch.com/2010/04/22/facebook-edgerank/.

Ma, Michelle. “How to Teach Kids About AI.” *The Wall Street Journal*, Dow Jones & Company, 13 May 2019, www.wsj.com/articles/how-to-teach-kids-about-ai-11557759541.

Marcano, Andrew E. "Critical Theory and Social Media: Alternatives and the New Sensibility." Master's thesis, University of Pennsylvania, 2018. August 5, 2018. Accessed July 15, 2019. https://repository.upenn.edu/ppe_honors/35/.

Marcuse, H. *One Dimensional Man*. London: Abacus, 1974.

Marcuse, Herbert, Kellner, Douglas (ed.). Some Social Implications of Modern Technology. *Technology, War and Fascism*. London and New York: Routledge Press, 1998.

McGee, Matt. "EdgeRank Is Dead: Facebook's News Feed Algorithm Now Has Close To 100K Weight Factors." *Marketing Land*, 16 Aug. 2013, marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908.

Noble, Safiya Umoja. *Algorithms of Oppression How Search Engines Reinforce Racism*. New York: New York University Press, 2018.

Obermeyer, Ziad, et al. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science*, vol. 366, no. 6464, 2019, pp. 447–453., doi:10.1126/science.aax2342.

Oliveira, Arlindo L. *The Digital Mind: How Science Is Redefining Humanity*. Cambridge, MA: MIT Press, 2018.

Payne, Blakeley H. An Ethics of Artificial Intelligence Curriculum for Middle School Students. MIT Media Lab, Massachusetts Institute of Technology, Cambridge, August 2019. *MIT MEDIA LAB*, https://docs.google.com/document/d/1e9wx9oBg7CR0s5O7YnYHVmX7H7pnITfoDxNdrSGkp60/edit#heading=h.1et5vs39qkyh.

Pilkington, Ed. "Digital Dystopia: How Algorithms Punish the Poor." *The Guardian*, Guardian News and Media, 14 Oct. 2019, https://www.theguardian.com/technology/2019/oct/14/automating-poverty-algorithms-punish-poor.

"Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning [AI/ML]-Based Software as a Medical Device [SaMD]." *Discussion Paper and Request for Feedback.* U.S. Food & Drug Administration, 4 April 2019, https://www.regulations.gov/document?D=FDA-2019-N-1185-0001

Robinson, Jamal. "How Facebook Scales Machine Learning ." Medium. Medium, February 3, 2019. https://medium.com/@jamal.robinson/how-facebook-scales-artificial-intelligence-machine-learning-693706ae296f.

White, Jeffrey, and Rick Searle. *Rethinking Machine Ethics in the Age of Ubiquitous Technology*. Information Science Reference, An Imprint of IGI Global, 2015.