ORIGINAL ARTICLE

Examining embedded apparatuses of AI in Facebook and TikTok

Justin Grandinetti[1]

Abstract

In popular discussions, the nuances of AI are often abridged as "the algorithm", as the specific arrangements of machine learning (ML), deep learning (DL) and automated decision-making on social media platforms are typically shrouded in proprietary secrecy punctuated by press releases and transparency initiatives. What is clear, however, is that AI embedded on social media functions to recommend content, personalize ads, aggregate news stories, and moderate problematic material. It is also increasingly apparent that individuals are concerned with the uses, implications, and fairness of algorithmic systems. Perhaps in response to concerns about "the algorithm" by individuals and governments, social media platforms utilize transparency initiatives and official statements, in part, to deflect official regulation. In the following paper, I draw from transparency initiatives and statements from representatives of Facebook and TikTok as case studies of how AI is embedded in these platforms, with attention to the promotion of AI content moderation as a solution to the circulation of problematic material and misinformation. This examination considers the complexity of embedded AI as a material-discursive apparatus, predicated on discursive techniques—what is seeable, sayable, knowable in a given time period—as well as the material arrangements—algorithms, datasets, users, platforms, infrastructures, moderators, etc. As such, the use of AI as part of the immensely popular platforms Facebook and TikTok demonstrates that AI does not exist in isolation, instead functioning as human–machine ensemble reliant on strategies of acceptance via discursive techniques and the changing material arrangements of everyday embeddedness.

Keywords AI · Algorithms · Transparency · Platforms · Material-discursive

1 Introduction

While forms of artificial intelligence (AI) are increasingly embedded in a growing number of everyday practices, the use of learning AI as part of social media platforms demonstrates how these processes are made colloquially accessible, yet simultaneously opaque and controversial. Most generally, learning AI identifies relations, detects and uncovers patterns, classifies and transcribes, and decides and recommends (Murphy 2012; Goodfellow et al. 2017). In popular discussions, the nuances of AI are often abridged as "the algorithm" (Agrawal 2016), as the specific arrangements of machine learning (ML), deep learning (DL) and automated decision-making (ADM) on social media platforms are typically shrouded in proprietary secrecy punctuated by vague press releases and transparency initiatives. That is, while members of the general public are likely unable to explain the intricacies of the machine learning models underpinning processes of suggestion, recommendation, and moderation found on popular social media platforms, there is a rising popular acknowledgment that "the algorithm" matters.

For example, speculation as to how algorithms prioritize certain content on platforms is an essential conversation for those looking to monetize a social media presence. Celebrities and influencers tag posts with "for the algorithm" to try to make their content trend on platforms like Instagram and TikTok. Despite the opacity in popular discussions of AI, what *is* clear to researchers is that on social media platforms algorithms function in various ways to recommend content (Kumar 2019), personalize ads (Boerman et al. 2018), aggregate news stories (Thurman and Schifferes 2012), and moderate problematic material (Gillespie 2020). Prima facie, these possibilities offer potential convenience for users, along with greater efficiency and revenue for platforms. Nevertheless, both scholarly and popular assessments warn

✉ Justin Grandinetti
  jgrandin@uncc.edu

1 Communication Studies, The University of North Carolina at Charlotte, Charlotte, USA

of the algorithmically assisted proliferation of disinformation along with the formation of "echo chambers" (where individuals only interact with information that conforms to, not challenges, their beliefs) and their digital counterpart "filter bubbles" (how personalized filters like search engines facilitate a unique universe of information for each of us) (Pariser 2011; Benton 2016). These terms came to public prominence in explanations of the role social media played in the 2016 US Presidential elections, and remain unfortunately relevant in the midst of circulating conspiracy theories and disinformation regarding the COVID-19 pandemic.

Perhaps related to these issues of algorithmically assisted disinformation circulation, recent studies show that individuals are concerned with the risks and fairness of greater reliance on AI automated decision-making (Lee 2018; Araujo et al. 2020). It should also come as no surprise that scholarship has emphasized the complexity of how algorithms are understood in both colloquial and popular discussions (Dourish 2016) as well as "folk theories" characterizing algorithms as confining, practical, reductive, intangible, and exploitative (Ytre-Arne and Moe 2020). Film and television, and in particular the genres of science fiction and horror, have long reflected societal anxieties surrounding the adoption of new technologies, as well as fears of disruption and disconnection once technologies become embedded and essential (Gunning 1991). To this end, contemporary shows like Netflix's *Black Mirror* create moments of psychological uneasiness through AI reanimations of the deceased based on their social media history, whereas others like HBO's *Westworld* features an antagonistic AI system that manipulates all human behavior and includes (fairly ridiculous) dialogue such as "The truth is that a human is just a brief algorithm—10,247 lines." Such popular depictions and discussions about algorithms and AI evoke a wide array of responses, ranging from indifference and ambivalence, to ambition and possibility, to fear and reverence. While I wish to avoid delving into the difficult questions of how and why individuals have come to be concerned with various arrangements of AI in this particular paper, what matters here is that some amalgamation of big data, algorithms, and learning AI elicits substantial trepidation within the public imagination.

Social media platform officials and spokespeople seem to understand the importance of selling the public and regulators on the benefits of AI. Specifically, both Facebook and TikTok have unveiled recent algorithmic accountability initiatives in response to controversies regarding issues of data collection, privacy, security, content moderation, and recommendation (TikTok 2020a; Alexander 2020; Sonderby 2020; Robertson 2020; Vincent 2020; Wang 2020). It is certainly possible to dismiss these tried-and-true tactics by big tech as hollow attempt to placate the user concerns and respond to threats of government regulation through the promise of transparency (Crain 2016). And, for the moment, it does appear that both Facebook and TikTok were able to forestall oversight by the US government (Kelly 2020; Allyn 2020). The end goals of platform providers notwithstanding, these initiatives along with official statements by platform representatives offer insight into the embeddedness of AI in everyday practices, how AI is continually defined and redefined, and how AI is made banal, concealed, and infrastructural.

In the following paper, I turn to a mix of algorithmic transparency initiatives and statements by representatives of Facebook and TikTok as case studies of how AI is embedded in these platforms, with attention to the promotion of AI content moderation as a solution to the circulation of problematic material and dis- and misinformation. Here, I add to conversations on AI transparency that include the variety of forms transparency takes (Walmsley 2020), how the ethical issues of AI are portrayed in media coverage and public discourse (Ouchchy et al. 2020), the relationship between transparency and how the public perceives AI decision-making as legitimate or not (de Fine Licht and de Fine Licht 2020), as well as the promise of AI for content moderation (Gillespie 2020). Additionally, this analysis of Facebook and TikTok's AI and algorithmic promotion strategies serves as a critical supplement to scholarly accounts that treat algorithmic systems as diffuse sociotechnical systems (Seaver 2017) that "live in dynamic relation to other material and discursive elements of software systems and the setting that produce them" (Dourish 2016: 1). Following the lead of these critical accounts, I position AI not merely as a discrete computational procedure, but instead as a material-discursive apparatus, in which there is an ongoing and intra-active relationship between the "material (re)configurings or discursive practices that produce material phenomena in their differential becoming" (Barad 2007: 170). Certainly, there is a material world of matter forms, but representationalist scholarship often subordinates matter to that of language. Instead, Barad emphasizes a cyclical relationship between matter and its representations—between the material and discursive—in shaping one another. AI is predicated on discursive techniques—what is seeable, sayable, knowable in a given time period—as well as the material arrangements—algorithms, datasets, users, platforms, infrastructures, servers, etc. To wit, Barad's attention to the material and discursive serves as a model for considering how boundaries between things are produced as well as how individuals conduct themselves in relation to other humans, nonhumans, systems, infrastructures, platforms, institutions, etc. Assessment of the ongoing becoming of artificial intelligence, both in terms of how AI is composed and how AI is understood, requires attention to the complex strategies of integration and acceptance of AI. Stated more simply, the use of AI as part of the immensely popular platforms of Facebook and TikTok demonstrates that AI does not exist in isolation, instead of functioning as human–machine ensemble (Møhl 2020) reliant on discursive

techniques that promote AI as technological panacea, as well as new material arrangements of everyday embeddedness, often in the form of AI content moderation.

Assessing what exactly *are* the material-discursive apparatuses of Facebook and TikTok's AI is a challenging task. Here, Barad's defractive analysis functions as a strategy for mapping and describing phenomena that resists a singular, authoritative reading (Taylor and Ingraham 2020). That is to say that while there are existing definitions of exactly what constitutes artificial intelligence, there are also contradictory ways in which AI is discussed by computer scientists, platform representatives, tech companies, government officials, members of the public, etc. as well as complex forms of how AI is embedded and experienced in everyday life. This paper specifically focuses on social media algorithmic transparency initiatives and attendant AI controversies of Facebook and TikTok to highlight both the changing material arrangements of AI compositions within social media platforms, as well as how AI is discursively positioned to the public. Doing so requires examination of an archive that includes popular press coverage of these platforms, technology news, statements from platform representatives, transparency initiatives, and official statements and posts from Facebook and TikTok.

2 AI definitions, terminology, and history

Though the cultural fascination with "the algorithm" functions as shorthand for a variety of learning AI models, it is critical to clarify what exactly AI entails—especially because "AI" and "algorithm" are used in inconsistent ways by tech companies and in public dialogue. Take for instance two of Google's most popular features "Search" and "Translate." These features are not all that technologically dissimilar, and both involve AI techniques. However, few consider Google searches AI, while Google's translation feature is often characterized as artificial intelligence due to the perceived intellectual skill of translating language (Gunkel 2020). Other examples such as IBM's chess playing Deep Blue show how something formerly considered intelligent AI can be "demoted" to the status of "just another computer application" over time (Gunkel 2020). In contrast to this AI demotion, machine learning processes that have been critical to industry optimization and problem-solving for decades have more recently been "promoted" to the status of AI by tech companies (Jordan 2019). Defining AI, accordingly, requires attention to evolving and contested definitions of what constitutes intelligence, as well as acknowledgement that AI characterizations are often a branding and promotion tool for tech companies.

Most of us have become accustomed to the bold promises of technological hype, designed to excite consumers, court investors, and placate governments; but these strategies also have a dark side, in that hype cycles hide technological flaws and limitations from critical consideration (Milne 2020). Hype and over-excitement about technologies can be limiting and blinding. Perhaps most importantly, these discursive characterizations are intertwined with technological trajectories. Technologies like AI do not suddenly appear; these computational processes have a long history that includes the emergent and dominant discursive and rhetorical positionings by tech companies (Natale and Ballatore 2020). More specific to AI, Natale and Ballatore (2020) contend that the myths of AI are driven by the recurrence of analogies and discursive shifts, imaginings of the future in terms of how the limitations and shortcomings of AI will be overcome, and the controversies of AI. What matters most is that the myths of a thinking machine capable of transcending human intelligence (often to utopian or apocalyptic ends) continue to shape the social presence of contemporary AI forms (Natale and Ballatore 2020). The case studies following this section encounter some of the same type of discursive mythology as Natale and Ballmore's article, including emphasis on how Facebook and TikTok's AI is wrapped up in platform controversies and overblown promises of improved AI moderation of content. Moreover, the accounts from Facebook and TikTok inevitably encounter such terminological slippage due to the fuzziness in which transparency initiatives and company representatives discuss the uses of algorithms, big data, and AI within their platforms. Whether it comes from tech company hype, opaque popular discussions, and even scholarly literature, AI vocabulary is often used in such opaque and contradictory fashion, but I wish to provide some clarification and context on the subject before delving into each case study. This involves attention to some intertwined strands of AI scholarship: the changing historical trajectory and evolutions of AI techniques, imaginings of both AI and intelligence, and the contemporary everyday forms of and colloquial discussions about AI.

Regarding AI's historical trajectory, Gunkel (2020) notes that there have been multiple "boom and bust cycles" in terms of the dominant paradigm of AI over time, beginning with symbolic or good old-fashioned AI (GOFAI) in the 1950s. GOFAI differs greatly from more contemporary forms of "connectionist" machine learning—particularly due to a division of how intelligence is conceptualized (Franklin 2014; Mühlhoff 2019). That is, GOFAI characterizes intelligence and cognition as achieved by manipulating human-readable symbols according to underlying rules; by contrast, machine learning models use algorithms and data sets to identify structural relations among datasets, to detect and uncover patterns, and to classify, transcribe, and make decisions (Murphy 2012; Goodfellow et al. 2017). The contemporary dominance of the learning AI paradigm can be attributed, in large part, to advances in hardware architectures as

well as a landscape of data collection, storage, and processing obtained through more and more instances of human participation (Mühlhoff 2019). For example, the practices of viewing, sharing, liking, tagging, clicking, retweeting, etc. on social media platforms generates enormous amounts of data. Such activities have been characterized by Bruns (2007) as produsage—the produser is a productive user who engages in the technosocial networked environments. Produsage accounts for the fact that participants on platforms are not producers in a traditional, industrial sense in creating a physical product; these acts of participation accumulate, and this data is in turn stored and used as training/verification data for machine and deep learning models (Mühlhoff 2019). Taken together, the shift to the learning AI paradigm over time has two key implications worth highlighting. First, learning AI is often embedded, invisible, and reliant on data collected from user interactions on platforms. And second, while contemporary forms of AI are capable of increasingly impressive feats, the goal of creating a machine capable of high-level reasoning and thought similar to the human mind has *not* been achieved. The learning AI embedded in social media, for instance, might be capable of flagging a problematic post, but lacks the ability to offer reasoned justification and analysis for why such a post might be an issue. More on this subject later.

Despite their often muddied and interchangeable use by tech companies and in popular discussions, essential to compositions of learning AI are the elements of big data and algorithms. The algorithm has origins in the texts of Ninth Century Muslim mathematician Muhammed ibn Musa al-Khwarizmi in referring to a mathematical or logical term for a set of instructions (Introna and Wood 2004: 18). There are natural overlaps between big data and AI, in that AI needs large amounts of data to function. The key difference between big data algorithms and AI rests in the degree of automation: "big data aims at enhancing decision-making, AI goes further aiming at automating decision-making" (Strauß 2018: 4). To this end, machine learning uses large datasets *and* layers of algorithms to learn to perform tasks through pattern recognition and probability. For instance, machine learning processes can be used to achieve automated decision-making, in which algorithmic correlations allow for conclusions to be reached for simple and complex problems, often with minimal supervision (Allen and Masters 2020). Machine learning requires human supervision and training via pre-structured datasets, but the subset of deep learning utilizes layers of algorithms compose artificial neural networks modelled after the neurons of the human brain that are capable of unsupervised learning from unstructured datasets (Jan et al. 2019; Arkoudas and Bringsjord 2014; Goodfellow et al. 2017). These distinctions aside, the learning AI paradigm overall "is less about replacing human cognitive labor by an intelligent machine but about

embedding and harvesting human cognition in computing networks through new forms of labor and machinized power relations" (Mühlhoff 2019: 3). Stated differently, the expansion of AI is largely driven by sociotechnical shifts—in particular, the paradigm of big data collection—that captures human participation in an ever-increasing variety of contexts.

This context on artificial intelligence is relevant to Facebook and TikTok for a number of reasons—perhaps the most important of which comes in the form of pushing back on narratives of technological neutrality and objectivity. It is hardly a novel contention to claim that technologies are political in reflecting, maintaining, and transforming relations of power (Latour 1990; Slack and Wise 2005). As noted by Feenberg (2006), technologies are no mere tools under the control of humans; instead, technologies and society are intertwined and influence one another. That is, instead of falling into technological determinism (where technology functions as driving causal agent for a cultural change) or cultural determinism (in which social values or megamachines such as capitalism drive what is created, used, and adopted), theorists instead have turned to examinations of how humans and non-humans are arranged in constellations that function toward certain logics and goals, enable relations of power, facilitate flows of capital and data, and contribute to processes of subjectivation. Or, perhaps put even more simply, culture and technology are inexorably intertwined—and these technosocial interrelations often reinforce historical inequities. For example, despite his status as a celebrated twentieth century architect, Robert Moses designed intentionally low overpasses on the Southern State Parkway to keep buses of low-income and minority demographics from travelling to Long Island (Winner 1980). Unfortunately, similarly problematic racial biases are found in algorithmic and AI compositions as well (O'Neil 2016; Noble 2018). Recent attention to the deep-learning powered autoregressive language model GPT-3 shows how bias can be made inherent to AI, as GPT-3 trained on large swaths of internet data shows ample evidence of troubling racial, gender, and religious biases (Brown 2020). Returning for a moment to the fact that algorithms and training datasets (often captured via everyday practices and interactions on platforms) are critical to learning AI necessitates reminder of how AI is fundamentally a sociotechnical composition. To this end, algorithms may be better understood not merely as set of instructions, but instead as performing entities, or "actualities that select, evaluate, transform, and produce data" (Parisi 2013). All this is to emphasize that AI algorithmic design is not objective, but reflective of a platform's goals; AI is not neutral, but instead beholden to the inherent biases (or perhaps less nefariously, the lack of critical awareness) of those selecting AI training data, as well as the biases of those whose data was used for the training.

Taken together, changing historical trajectories, cultural imaginings, colloquial and expert discourses, and computational and infrastructural techniques of AI are all relevant in understanding exactly what AI is and how it is embedded in everyday practices. Sure, there are attempts to nearly define artificial intelligence, but the myriad sociotechnical constellations of AI, as well as what is seen, said, and known about AI across popular, corporate, and academic discourses resists one singular reading. Such a subject, accordingly, is well-suited for Barad's defractive analysis. More specifically, the AI embedded in social media platforms is, in many ways, both discursively and materially different from the knowledge techniques, imaginings, and technical compositions of historical forms of AI like GOFAI. Here, the complexities in both composition and uses of learning AI make clear that artificial intelligence is not simply a unified machine nor evolution of computational techniques, but an ever-changing sociotechnical apparatus (Dourish 2016; Seaver 2017; Møhl 2020). Examining the AI of TikTok and Facebook as sociotechnical apparatus requires a material-discursive investigation that takes into account the meaning, knowledge practices, and social/cultural power of artificial intelligence as it is embedded within various platforms, infrastructures, and practices.

3 Case study: Facebook transparency reports and initiatives

Facebook is a platform that needs little introduction. Launched in 2004 initially as an Ivy League University-only social network, Facebook self-reports 2.8 billion monthly active users worldwide, barring a few exceptions in countries that block the platform (Facebook Investor Relations 2020). Due to its massive user base along with continual efforts to add new features and buyout (or squash) rivals, Facebook joins Amazon, Apple, Google, and Microsoft in comprising the Big Five tech giants in the US. The sheer magnitude and everyday embeddedness of Facebook blurs the line between platform and infrastructure (Helmond et al. 2019), with scholars questioning the considerable ethical and legal implications of what would happen if Facebook went down permanently (Öhman and Aggrawal 2019). Of course, Facebook's fame is perhaps only surpassed by its infamy. The platform's data breaches, laissez-faire attitude to data privacy, misuse of data, ad discrimination, anti-competitive practices, and role in the spread of disinformation and the rise of violent political extremism are issues well beyond the scope of this paper. In light of this ever-growing list of controversies, it can seem like Facebook's co-founder, chairman, CEO, and controlling shareholder Mark Zuckerberg has become a fixture of Capitol Hill in testifying before Congress about Facebook's practices, though these

hearings have led to little in terms of actual regulatory oversight. And in spite of negative press and scholarly critique, users continue to engage with Facebook and seem relatively undeterred by data collection practices and privacy issues (Lupton and Southerton 2021).

As previously noted, social media platforms like Facebook employ a variety of embedded learning AI that identifies relations, detects and uncovers patterns, classifies information, recommends content, moderates and flags problematic material, auto-generates pages, and delivers targeted advertisements. These embedded uses of learning AI are often flashpoints for some of the most controversial aspects of Facebook. The list of recent examples is too many to elucidate in this paper, but includes the thriving presence of white supremacist content on Facebook, in which Facebook's interface not only failed to redirect users away from white supremacist pages, but also directed users to far-right content (Tech Transparency Project 2020). Facebook's template is driven by forms of categorization that made data entered into the platform algorithm-ready (Bucher 2018). For example, Facebook's signup interpolates users into a system where they are encouraged to share contact and basic info, preferred pronouns, gender identity, relationship status, interests, photos, favorite movies, TV, books, and music, and in some countries, that they are an organ donor. Disturbingly related is that analysis of hate groups affiliated with white supremacy found that roughly 64% of these pages were auto-generated by Facebook's interface— that is, these pages were not made by users, but instead created when Facebook's AI scraped information from profiles and suggested a pre-made page for content based on this info. (Tech Transparency Project 2020). Research into Facebook's advertisement algorithms also encountered evidence of bias across racial and gendered lines despite selecting what Facebook considered neutral targeting parameters (Ali et al. 2019). These racially charged controversies are joined by the fact that Facebook struggles to contain the spread of mis- and disinformation across the platform (Breland and Levy 2021). Taken together, Facebook is clearly well-aware of its increasingly negative perception and actively works to counter these negative narratives.

Facebook has provided biannual transparency reports since 2013, which enumerate government requests for data, efforts to safeguard user data, context restrictions, instances of deliberate internet disruptions by governments around the world as pertaining to Facebook, and intellectual property reports (Sonderby 2020). While Facebook does provide a list of instances where content violating community standards was removed, there are sparse details as to how this content moderation occurred (either by AI or human moderation). As explained by Bucher (2018), calls for greater transparency do not always achieve the desired results, and too much information can also make the visible invisible in blinding

us from seeing clearly what Facebook does provide: a slick web interface that allows users to search through information on government requests for data, policies, enforcement, moderation, etc. On these transparency pages, there is both abundance of and vagueness of information. That is, one can easily become lost in the Facebook's countless rabbit holes of hyperlinks, policies, changes to policies over time, commitments, interactive charts, etc. These Kafkaesque guidelines are so numerous that Facebook itself even lost a piece of its own moderation policy internal guidance for three years (Facebook Oversight Board 2021). Further much of Facebook's public-facing transparency information the is platitudinal and reliant on quantified instances instead of truly transparent examples.

The story of Facebook's use of AI is, additionally, anything but completely transparent. That is, both official statements by Facebook and its representatives, as well as popular press and scholarly accounts are often vague about the use of algorithms and/or AI on the platform. One place where Facebook *is* forthcoming is in emphasizing the company's numerous investments in AI. Facebook has released open-AI tools, and has previously touted their Rosetta AI system as key to assisting in moderating problematic content (Cambridge Consultants 2019). Of course, moderation is one of the biggest challenges for social media platforms, and Facebook's well-publicized gaffes and shortcomings are, perhaps, why the company looks to the veneer of technological solutions as way to placate growing moderation concerns. A recent breakdown of how moderation occurs on the platform elucidates how posts believed to have violated the company's rules are flagged either by users or machine learning filters, then human moderators sort through flagged posts for removal (Vincent 2020). This process functioned chronologically in the past—posts were dealt with in the order that they were reported. Facebook has since moved to using machine learning algorithms to prioritize what posts are seen first by moderators based on the criteria of virality, severity, and likelihood content is breaking the rules (Vincent 2020). Although Facebook provides few details on exactly how these criteria are weighted, visualizations of how AI is integrated into the moderation process make clear that the Facebook sees moderation as a human-AI synthesis, characterized as more "proactive" than "reactive" (Vincent 2020). Facebook is quick to promote the success of this enhanced AI integration, as a 2020 Community Standards Enforcement Report notes that 95% of more than 22 million posts including hate speech were initially flagged by AI (Bidar 2020). Here, Barad's attention to the material and discursive offers a strategy for examining how AI is embedded. Over time, the integration of AI moderation changes as a material arrangement; AI goes from a process that operates to flag posts independently of moderators to one that is, instead, integrated with moderators. Stated differently,

the old moderation system was a two-fold process: user-flagged posts were filtered and ranked by human moderators while machine learning classifiers also sent posts to human moderators for review. Under the new system of AI integration, both user-reported content and ML flagged content is filtered, ranked, and de-duplicated by machine learning AI *before* human review. The changing material arrangement of AI integration to moderation, nevertheless, exists because of a shift in the discursive imaginings of AI. Instead of functioning independently to flagged posts, AI instead now works "side-by-side" in a manner of speaking with human moderators in allowing a more efficient proactive prioritization of flagged content. Platform users themselves might be unaware of how this process occurs, but they do experience algorithms—often through what they see or do not see on a platform. That is, "by using algorithms they also do things to [the algorithms]—modulate and reconfigure them in both discursive and material ways" (Bucher 2018: 95). Put another way, the felt and imagined everyday encounters with algorithms and learning AI matters, and this manifests in legitimate and illegitimate accusations of platform moderation bias.

Despite these transparency efforts by Facebook, the reality of the platform's successful AI integration is far murkier—particularly when it comes questions of what and why certain content is flagged and removed. In spite of the fact that "censorship" on social media generally and Facebook specifically has become a rallying cry for those whose content is removed, recent analysis has found no evidence of platform bias against rightwing groups (Barrett and Sims 2021). But Facebook does censor plenty of political content. Recently leaked Facebook internal guidelines maintains a list of "recognized crimes", designed to cater to more repressive governments that pressure Facebook to censor content in certain countries pertaining to sexuality, peaceful protests against governments, and "controversial" historical events (Hern 2021). In 2020, the company had to update its so-called race blind hate speech algorithm, which treated posts critical of men, white people, and America created by historically marginalized groups and minorities the same as hate speech, Holocaust denial, and racist stereotypes (Jibilian 2020). More precisely, these instances demonstrate a lack of historical context and critical reasoning by machine learning models, which were trained to treat a host of content dealing with race, gender, and geopolitics as evenly problematic. Accordingly, the AI-assisted moderation strategies of Facebook demonstrate the complexity of embedding AI into everyday practices. Algorithmic systems are situated in larger socioeconomic contexts, as well as larger systems that are often implicit in injustice and institutionalized oppression (Dencik et al. 2019; Johnson 2014). Further, algorithmic systems all-too-often function toward biased ends across racial and historically marginalized lines (Noble

2018; Ochigame 2020). When taken as a totality, the integration of AI moderation into Facebook demonstrates that algorithmic systems are far from neutral, but instead reflect power relations, histories of inequity, geopolitical pressures, and implicit biases.

It is also possible that Facebook's greater recent emphasis on AI as a solution to the spread of misinformation is related to other bad press about moderation conditions on the platform. A 2019 exposé shed light on the horrid circumstances of Facebook's content moderation, as moderators must navigate the stress and mental health impacts of viewing images of graphic sexual content, child pornography, murders, suicides, conspiratorial content, and extreme violence, all for low wages with little in terms of wellness support (Newton 2019). Moderators occupy what is tragically a critical job in stopping the spread of horrendous content on social media, but it is also unlikely a position any person is mentally suited for. Despite promises of improved conditions for moderators, however, recent accounts detail that little has actually changed (Dickey 2020). The harsh realities of moderation are, in and of itself, a topic worthy of both further scholarly analysis and widespread societal meditation on the costs versus the benefits of social media (see Gray and Suri 2019 for a more detailed breakdown of the hidden and taken-for-granted labor of moderation). Such a topic is also relevant to the integration of artificial intelligence into everyday practices, as the increased promotion of AI allows platforms like Facebook to minimize the role of human moderators in the eyes of the public and regulators. This discursive positioning, by extension, provides multiple benefits to Facebook in appearing more humane to moderators while at the same time espousing more so-called neutral and objective moderation of content.

Despite the rhetoric and possibilities of what AI is capable of, there is an important human element to these compositions that often goes overlooked. Not only is learning AI often reliant on both datasets derived from interactivity on platforms, but machine learning AI also requires human training and intervention. Such is the ghost labor (Wakefield 2021) or ghost work of AI (Gray and Suri 2019). The recent explosion of AI uses has been billed by tech companies as a revolution of automation, but the reality is that the human labor powering apps, websites, and AI is made intentionally hidden and opaque (Gray and Suri 2019). Put perhaps most directly, "AI is simply not as smart as most people hope or fear" (Gray and Suri 2019). AI is reliant on micro-work for AI preparation and training, AI "impersonation" in which humans are more in the loop of tasks than tech companies have the public believe, and AI verification, where humans are tasked assessing the work of AI or making decisions based on AI recommendations (Tubaro et al. 2020). The importance of human moderator intervention is certainly relevant to Facebook—even when accounting for the recent

proactive approach to moderation prioritization, where machine learning helps to filter, rank, and de-duplicate problematic content. More simply, the removal of content is still subject to human review (and by extension continual exposure to rule-breaking content). Facebook's account of how machine learning filters analyze posts—what the company calls "whole post integrity embeddings" (WPIE)—demonstrates the current challenges of AI-only moderation (Vincent 2020). While WPIE machine learning algorithms are able to judge various elements of a given post, AI is unable to work out what images, captions, the relationship to the poster reveals (Vincent 2020). The limits of AI's "intelligence" are laid bare. Although machine learning is able to flag problematic aspects of a post, it is unable to engage in the kind of critical thought and reasoning required to discern if, for instance, a post labelled "special treats" and a picture of Rice Krispies squares are merely baked goods or code for THC edibles (Vincent 2020). Moderators are, consequently, an essential if all-too-often invisible aspect of AI's integration into social media platforms. Like any kind of infrastructure, attention is rarely given to the sociotechnical composition of AI moderation until there are hiccups, mistakes, and outright breakdowns. Importantly, a Barad-inspired defractive analysis offers possibility for mapping this complex sociotechnical arrangement. That is, the human element of AI content moderation may be left out of Facebook's discursive accounts of AI, but this component is a critical part of how AI successfully (or unsuccessfully) functions on platforms.

Transparency initiatives and statements by Facebook representatives demonstrates the importance of a material-discursive account of how AI is embedded in everyday practices. Facebook is one of the biggest tech companies in the world, and the discursive hype machine of AI characterizations generated by the platform have power to shape perceptions of what AI can do. Attention to discursive practices demonstrate how AI offers a vague panacea against the growing threats of misinformation and extremist content on the platform, in turn functioning as a response to official government regulation and oversight. Zuckerberg himself has emphasized the power of AI to flag and remove 99% of content related to both ISIS and al-Queda on Facebook, but these numbers are self-reported by the platform (Harwell 2018). Following Barad's emphasis on the cyclical relationship between the material and discursive, it is also critical to note that material limitations of AI remain an issue for Facebook and other platforms. Zuckerberg stated in 2018 that he believes that AI will be capable of comprehending linguistic nuances and flag potential risks within five to ten years, an estimate challenged by experts (Harwell 2018). At the heart of the material-discursive compositions of Facebook's embedded AI, then, is a tension—what AI can do versus what Facebook wants the public to think AI can do.

Accounts of how algorithms are part of larger ecologies emphasize how algorithmic processes can be used to spread propaganda (Treré 2018), but one must wonder how much Facebook's algorithmic transparency initiatives and statements border on platform propaganda. Or, put another way, there is a chiasmus between the technocratic veneer of how Facebook positions AI as solution to the maligned social media platform's failings versus the very real limitations of Facebook's use of AI for the purposes of selection, advertisement, recommendation, and moderation.

4 Case study: TikTok accountability and transparency initiatives

TikTok is a short-form mobile streaming app featuring user-uploaded audio-visual content of sixty seconds or less. The app is available in over 150 countries around the world, and is the international version of the Chinese-owned app Douyin, launched in 2016 by parent company ByteDance. Douyin itself is a rebranded version of the app Musical.ly, which debuted in Shanghai in 2014 (Tidy and Smith Galer 2020; Kaye et al. 2020). As such, the story of TikTok and Douyin is one of parallel platformization, in that both apps are "deployed in vastly different contexts and have thus far managed to survive as emerging platforms in two opposing but comparable oligopolistic platform ecosystems" (Kaye et al. 2020). Compared to more established social media platforms like Facebook, TikTok has existed for a relatively short amount of time; nevertheless, the platform's rise is nothing short of meteoric. TikTok was the most downloaded app in 2020 (Bellan 2020) reaching over two billion downloads (Qu 2020) second only to Facebook in average-user minutes per month (Statista 2021). A leaked pitch deck from TikTok further details that the platform's 732 million global active monthly users engage with the app for an average of 89 min per day and open the app an average of 19 times per day (Ingham 2021). TikTok's user base is disproportionally young, as roughly 60% of US TikTok users are between 16 and 24 years old (Rocque 2019). While younger audiences have been scrolling through near-endless video content that often takes the form of dances, pranks, unboxings, performances, viral trends, and music tie-ins for some time now, many older adults only became aware of the platform following former US President Donald Trump's executive action that ordered a ban of the app in the United States if ownership in the platform was not sold by ByteDance (Carvajal and Kelly 2020 2020). The Trump administration's rationale cited that TikTok "automatically captures vast swaths of information from its users" that "threatens to allow the Chinese Communist Party access to Americans' personal and proprietary information" (Carvajal and Kelly 2020). Of course, even before President Trump lost reelection to his successor President Joe Biden, the deadline for ByteDance to sell TikTok passed without an actual ban taking place.

The TikTok ban in the United States mirrors similar discussions by governments around the world, as Pakistan, Japan, Indonesia, and Australia all debated similar action, whereas the Indian government actually enforced a ban of TikTok starting in mid 2020 (Iyengar 2020). Relatedly, in response to recent protests coordinated against President Vladimir Putin on social media, the Russian government fined TikTok for failing to remove so-called inappropriate political content (Leo and Lunden 2021). The general complexities of blocking web traffic as proxy battle for international politics (Merrill and Weber 2020) as well as the specific geopolitical tensions underlying TikTok bans (Gray 2021) have been taken up in detail in other accounts. But what matters to my own examination of TikTok's transparency initiative is twofold. First, it must be acknowledged that these proposed and actual bans are largely rooted in xenophobia and overblown security concerns. This is not to downplay the well-documented human rights abuses of the Chinese Communist Party, but geopolitics, economic competition, and leveraging a nebulous fear of China are likely more accurate drivers of TikTok bans than true security concerns. To this end, forensic accounts of TikTok show some concerns about data collection, but acknowledge that the apps' safeguards and policies "seem to cover enough ground for a user" (Neyaz et al. 2020: 56). Similarly, Jia and Ruan (2020) found that both TikTok and Douyin comply with privacy and data protection measures like the EU's General Data Protection Regulation in being attentive to users and policies from different geographical regions. Of course, the European Consumer Organization recently organized and filed complaints against TikTok with the European Commission, claiming that TikTok's terms and conditions are unfair to content creators, that terms of data collection are unclear to children and teenagers, and that the platform fails to protect underage users from both harmful content and hidden forms of marketing by influencers (Ikeda 2021). That being said, TikTok does not seem to collect any more data, nor any more meaningful data than platforms like Facebook do (not that such a statement should be read as a ringing endorsement of either TikTok or Facebook). Such realities have led to calls of hypocrisy—centralized data collection by big tech companies and the potentiality that these companies could turn this data over to government agencies deserves our attention regardless of platform country of origin (Panday 2020). Second, the complexities of proposed or actual TikTok bans notwithstanding, it is clear that ByteDance takes the geopolitics and bad press surrounding TikTok seriously enough to pursue transparency initiatives aimed at deflecting regulation through obfuscation and confusion.

Barad's defractive analysis allows attention to how TikTok integrates AI requires attention to discursive

positionings by the platform, along with the changing computational techniques of AI integration in the app. Similar to Facebook, TikTok began providing its own transparency reports released biyearly beginning in December 2019. These reports—although brief—detail the number legal requests the platform has received for data, as well as information on changing content moderation guidelines. Perhaps most notably, in early 2020 TikTok revealed plans to open "Transparency and Accountability Centers" in both Los Angeles and Washington DC. Though the openings of these locations were put on hold due to the Covid-19 pandemic, TikTok did eventually allow some researchers to virtually view the company's data storage, moderation practices, and inspect the platform's algorithms (TikTok 2020b; Vincent 2020). Comments from former TikTok CEO Kevin Mayer underscores the nature of this initiative:

> We believe our entire industry should be held to an exceptionally high standard. That is why we believe all companies should disclose their algorithms, moderation policies, and data flows to regulators. We will not wait for regulation to come, but instead TikTok has taken the first step by launching a Transparency and Accountability Center for moderation and data practices. Experts can observe our moderation policies in real-time, as well as examine the actual code that drives our algorithms. This puts us a step ahead of the industry, and we encourage others to follow suit (Newton 2020b).

As noble as Mayer's comments may seem, the idea that tech companies should disclose their algorithms had already been pushed by US lawmakers as potential solution to growing concern about TikTok's Chinese-based ownership. Accordingly, TikTok's transparency center initiative is likely an attempt to avoid government regulation, and, at the time, impending domestic ban.

TikTok's algorithm has long been central to conversations about the platform's success—something ByteDance has not been shy to promote. For example, in 2018 ByteDance claimed to be "one of the first companies to launch mobile-first products powered by machine learning technology" in order to "combine the power of artificial intelligence with the growth of mobile internet to revolutionize the way people consume and receive information" (Byford 2018). Representatives from ByteDance similarly noted that "artificial intelligence powers all of Bytedance's content platforms" in analyzing, text, images, and videos to serve users with targeted content (Byford 2018). And such targeting and personalization is critical to keeping users engaged. When a new user opens TikTok, the platform's algorithm finds eight popular but diverse videos to show them (Newton 2020a). From this point onward, TikTok gathers data "to map a user's preferences in relation to similar users and group them

into "clusters." Simultaneously, it also groups videos into 'clusters' based on similar themes, like 'basketball' or 'bunnies'" (Newton 2020a). TikTok's (2020a, b) own breakdown of the platform's recommendation algorithm further explains that recommendations are based on factors that include user interactions (videos that are liked or shared, followed accounts, comments posted), video information (captions, sounds, hashtags), and device and account settings (language preference, device type, country settings). Although there has been speculation that TikTok's AI is capable of recognizing images in uploaded videos for the purposes of categorization, recommendation, and moderation, it does appear that, similar to YouTube and Facebook, TikTok's machine learning models primarily relies on metadata like description, tags, time, and location of video upload (Knight 2019). Taken together, how TikTok employs learning AI toward the goals of recommendation, selection, and personalization is relatively unsurprising—TikTok is doing what other forms of social media do, the platform simply does it more successfully. Materially, algorithmic integration is very much in line with other platforms. Of course, whether found in official statements by TikTok representatives, as part of transparency initiatives, or in breakdowns of TikTok, the same conceptual fuzziness around artificial intelligence again rears its head—AI and algorithms are used interchangeably to describe TikTok's core functionality. The discursive promotional strategies of TikTok represent a hype cycle that obfuscates as much as it clarifies.

One additional point of emphasis in TikTok's transparency initiatives is concern with aforementioned filter bubbles. The delivery of personalized infospheres that conform and not challenge a user's viewpoints goes hand-in-hand with rising populism and political radicalization. Moreover, it is not only adults, but in particular younger individuals that are more likely to get their news information from social media platforms like TikTok (Swart 2021). On social media generally and TikTok specifically, scholars have called for greater moderation of problematic mental health representations and content (Gerrard 2020) and the spread of extremist content and hateful speech (Weimann and Masri 2020). These concerns are joined by examinations of algorithmic bias when it comes to categories of race, class, and gender (Bishop 2021), as well as how TikTok contributes to images of postfeminist neoliberal ideals of girlhood through promotion of a narrow range of prominent influencers (Kennedy 2020). It is, perhaps, in response to these issues that TikTok's transparency initiative specifically emphasizes the goal of bursting filter bubbles, promising that the platform's recommendation AI interrupts repetitive patterns and duplicate content, diversifies recommendations, and safeguards the viewing experience by moderating problematic content (TikTok 2020a, b). These goals of honing the algorithmic recommendation system for the purposes of popping filter

bubbles and combatting the spread of misinformation has been similarly echoed by TikTok's Product Lead Richard Huang (Newton 2020a). Whether or not TikTok succeeds where other social media platforms have consistently failed remains to be seen; yet, TikTok's recent discourse on AI and algorithms promotes these computational mechanisms as solution and not a problem.

Finally, in similar fashion to Facebook, the TikTok's discursive framings of AI within the platform often obfuscate the amount of human labor required for content moderation. As previously noted, promises of AI moderation often render invisible the massive amounts of content moderation work done by human labor—often exposing these workers some of the most damaging and horrendous content on the web (Tubaro et al. 2020; Wakefield 2021). What is also rendered invisible is how platforms use moderation practices to limit or outright exclude content from "undesirable users." In leaked documents from 2019, TikTok moderators were told to remove content of users who are/or look to have an: "abnormal body shape", "chubby", "obese or too thin", "ugly facial looks", "facial deformities", "senior people with too many wrinkles" or who are shooting video in an environment that is "shabby", "dilapidated", "slums", or "rural fields" (Biddle et al. 2020). In addition to removing fairly standard content in terms of criminal behavior, pornography, nudity, etc., TikTok's guidelines included wide-reaching bans on content that includes "defamation toward civil servants, political, or religious leaders", as well as the content that "endangers national security" or constitutes an "uglification or distortion of local or other countries' history" specifically citing "Tiananmen Square incidents" as example (Biddle et al. 2020). Such accusations of problematic moderation on TikTok are not new, as the platform has been previously accused of intentionally suppressing content from Black, disabled, and LGBTQ creators (Burke 2019). More recent forensic analysis of TikTok's censorship by researchers was inconclusive, but found that Douyin does clearly restrict some political terms in searches (Lin 2021). Such practices are far cry from the utopian moderation safeguards promoted by TikTok itself. Put another way, the discursive positionings of TikTok by platform representatives and transparency initiatives do not neatly align with the material constellations of content moderation.

Like the case study of Facebook, TikTok's transparency initiatives and official statements demonstrate the importance of a Barad-inspired material-discursive account of how AI is embedded in the platform, and, by extension, into everyday life. The discursive positionings of TikTok's algorithm by platform representatives are dedicated to transparency, bursting filter bubbles, and efficiently moderating problematic content—all desirable goals. The messy reality of human and AI moderation, however, challenges discursive positionings of a transparent, objective, and desirable

content moderation on TikTok. Here, the material compositions of how AI and moderator labor is arranged on TikTok further demonstrates the complexity of how AI is embedded in platforms and to what ends it functions. A material-discursive account of TikTok's algorithmic initiatives and statements sheds light on the complexity of embedding AI in everyday practices via an immensely popular platform. TikTok's AI integration reliant on the knowledge techniques and discursive positionings of a utopian AI potentiality—but what also matters are the human–machine compositions and ghost work that makes the platform run. Like the example of Facebook, the material and discursive are both cyclical and contradictory, together shaping how AI is characterized and operates on these platforms.

4.1 Embedded, integrated, and increasingly essential

Many of us encounter embedded forms of AI every day, often without even realizing it. In a platform era, AI matters to so much of what we view (or do not view), what gets promoted and what gets lost, what counts and legitimate and what gets automatically flagged or removed. And because these integrated forms of AI are so invisible and taken-for-granted, their power often goes unnoticed and unaccounted for—at least as long as they work well. Most social media users are fortunate to avoid graphic and illicit content in their personalized feeds, but recent societal issues with the spread of dis-and misinformation, along with new attention to algorithmically-generated filter bubbles have led to a general awareness that "algorithms matter." The strategies by social media platforms like Facebook and TikTok in promoting how well their AI functions, then, has implications for the evolving definition and societal understanding of what AI is. Defining AI is something that requires attention to historical context and trajectory, neuroscience, philosophy, computational techniques, corporate branding, colloquial discourse, fictional representations, and the changing definition of "intelligence." It's why Deep Blue can go from AI chess mastermind in one decade to mere computational process in another. This is, in part, because the goals of AI have shifted from replicating human intelligence to augmenting it. It is also due to changing technological techniques and infrastructures capable of collecting data via the platform model, storing previously unpredicted amounts of data on servers, and processing data in the form of machine and deep learning algorithms and neural networks.

In taking steps to examine how exactly AI becomes embedded and integrated into everyday life, this paper has made a case for conceptualizing AI as a material-discursive apparatus in a continual process of becoming. Here, attention to the discursive considers "that which constrains and enables what can be said" (Barad 2007: 170) emphasizing

"the role of discourses as loci where knowledge is tied with cultural and social power" in conditioning material bodies, events, and institutions (Huhtamo and Parikka 2011: 8). Put another way, technical media arise out of distributed discursive networks (bodies of knowledge and techniques) but then crystalize, cut their roots, become mobile, intermix, and recombine (Monea and Packer 2016: 3147). Said more plainly, the uses and embeddedness of AI in everyday interactions requires attention to both the changing compositions of which AI is a part, as well as general acceptance of data aggregation and faith in algorithmic processes.

I wish to conclude with a few takeaways from this paper's examination of algorithmic transparency initiatives from Facebook and TikTok as a way to examine AI embeddedness. First, it must be acknowledged that Facebook, TikTok, and social media more generally are far from the only places algorithms and AI are embedded into everyday practices; however, the immense popularity of social media and ensuing societal discussion situates these platforms at the forefront of public understanding of AI. Yet, far from providing a simple definition or example of AI, these transparency initiatives, statements from platform representatives, and the ensuing public/scholarly discussion contribute to a conceptual fuzziness around AI—artificial intelligence, algorithms, machine learning, and big data are often used in interchangeable and contradictory ways. Additionally, discussion of embedding AI in social media platforms are clearly responsive to larger cultural and scholarly conversations on filter bubbles (Pariser 2011; Benton 2016), content moderation (Gillespie 2020), the spread of misinformation (Swart 2021). Yet again, I wish to avoid simple arguments of causality (for instance, we simply cannot know for sure if popular attention to these issues *caused* TikTok and Facebook to respond with transparency initiatives). Instead, what matters more is how these transparency initiatives promise both a "look behind the curtain" of social media platform's use of AI, while also functioning to further blur some of the terminology of AI, algorithms, etc. These initiatives and dialogue from representatives of Facebook and TikTok are, after all, part of how AI is discussed, accepted (or not), and integrated in everyday sociotechnical practices.

These case studies also add to scholarly literature regarding AI transparency initiatives (Ouchchy et al. 2020; Walmsley 2020; de Fine Licht and de Fine Licht 2020). The goal of such initiatives is often to respond to and ultimately deflect government regulation and oversight through promises of internal review and greater transparency—and these tactics are clearly employed by both Facebook and TikTok. Finally, it is worth mentioning how the conjunction of promoting the possibilities of AI and the promises of AI's potential in moderating content in these transparency initiatives function to minimize the invisible ghost labor of those who work to create and train AI, as well as moderators tasked with the Sisyphean task of removing problematic content from these platforms.

Attention to how AI is embedded into society requires continued assessment of the discursive positionings and changing material compositions of contemporary AI in the various forms it may take. Inspired by this task, the case studies in this paper provide such material-discursive attention to AI vis-à-vis social media platforms like Facebook and TikTok. The two are massively popular platforms that users are likely to engage with multiple times in a given day, and are, most importantly, sites in which users engage with a variety of embedded AI processes. Artificial intelligence—particularly the kind found on these social media platforms—requires our participation and our data. Consequently, the challenge for platform providers like Facebook and TikTok in embedding AI is to promote algorithmic possibilities while also minimizing the downsides and limitations of these sociotechnical processes. The task of scholarly accounts is to assess how strategies promoting the benefits of AI contribute to the ongoing material-discursive arrangements of what AI is, what AI is becoming, and how AI is integrated into everyday practices.

Declarations

References

Agrawal A (2016) What do social media algorithms mean for you? Forbes. https://www.forbes.com/sites/ajagrawal/2016/04/20/what-do-social-media-algorithms-mean-for-you/?sh=1de5a4e6a515

Alexander J (2020) TikTok reveals some of the secrets, and blind spots, of its recommendation algorithm. The Verge. https://www.theverge.com/2020/6/18/21296044/tiktok-for-you-page-algorithm-sides-engagement-data-creators-trends-sounds

Ali M, Sapieznski P, Bogen M, Korolova A, Mislove A, Rieke A (2019) Discrimination through optimization: How Facebook's

ad delivery can lead to skewed outcomes. In: Proceedings of the ACM on human–computer interaction 2019

Allen M, Masters D (2020) Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making. ERA-Forum 20(4):585–598. https://doi.org/10.1007/s12027-019-00582-w

Allyn B (2020) Trump's TikTok sell-by date extended by 15 days. NPR. https://www.npr.org/2020/11/13/933916944/trump-ordered-tiktok-to-be-sold-off-but-then-ignored-the-deadline

Araujo H, Helberger N, Kruikemeier S, de Vreese C (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 35(3):611–623. https://doi.org/10.1007/s00146-019-00931-w

Arkoudas K, Bringsjord S (2014) Philosophical foundations. In: Frankish K, Ramsey W (eds) The Cambridge handbook of artificial intelligence. Cambridge University Press, Cambridge, pp 34–63

Barad K (2007) Meeting the Universe Halfway: quantum physics and the entanglement of matter and meaning. Duke University Press, Durham

Barrett PM, Sims JG (2021) False accusation: the unfounded claim that social media companies censor conservatives. NYU Stern Center for Business and Human Rights. https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/60187b5f45762e708708c8e9/1612217185240/NYU+False+Accusation_2.pdf

Bellan R (2020) The top social media apps of 2020, according to Apptopia. Forbes. https://www.forbes.com/sites/rebeccabellan/2020/12/03/the-top-social-media-apps-of-2020/?sh=218911a01d4e

Benton J (2016) The forces that drove this election's media failure are likely to get worse. Neiman Lab. https://www.niemanlab.org/2016/11/the-forces-that-drove-this-elections-media-failure-are-likely-to-get-worse/

Bidar M (2020) Facebook touts use of artificial intelligence to help detect harmful content and misinformation. CBS News. https://www.cbsnews.com/news/facebook-artificial-intelligence-harmful-content-misinformation/

Biddle S, Ribeiro PV, Dias T (2020) Invisible censorship: TikTok told moderators to suppress posts by "Ugly" people and the poor to attract new users. The Intercept. https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/

Bishop S (2021) Influencer management tools: algorithmic cultures, brand safety, and bias. Soc Media Soc. https://doi.org/10.1177/20563051211003066

Boerman S, Kruikemeier S, Zuiderveen Borgesius F (2018) Exploring motivations for online privacy protection behavior: insights from panel data. Commun Res. https://doi.org/10.1177/0093650218800915

Breland A, Levy P (2021) Report: Facebook Still Struggles to Contain Misinformation. Mother Jones. https://www.motherjones.com/politics/2021/03/facebook-misinformation-avaaz/

Brown et al (2020) Language models are few-shot learners. John Hopkins University, Open AI. https://arxiv.org/pdf/2005.14165.pdf

Bruns A (2007) Produsage: a working definition. Produsage.org. http://produsage.org/node/9

Bucher T (2018) If…Then: algorithmic power and politics. Oxford University Press, New York

Burke C (2019) Tiktok admits to hiding content made by fat, LGBTQ, and disabled users. The Daily Dot. https://www.dailydot.com/irl/tiktok-fat-lgbtq-disabled-creators/

Byford S (2018) How China's Bytedance became the world's most valuable startup. The Verge. https://www.theverge.com/2018/11/30/18107732/bytedance-valuation-tiktok-china-startup

Carvajal N, Kelly C (2020) Trump issues orders banning TikTok and WeChat from operating in 45 days if they are not sold by Chinese parent companies. CNN. https://www.cnn.com/2020/08/06/politics/trump-executive-order-tiktok/index.html

Cambridge Consultants (2019) Use of AI in online content moderation. 2019 Report produced on behalf of Ofcom. https://www.cambridgeconsultants.com/sites/default/files/uploaded-pdfs/Use%20of%20AI%20in%20online%20content%20moderation.pdf

Crain M (2016) The limits of transparency: data brokers and commodification. New Media Soc 20(1):88–104. https://doi.org/10.1177/1461444816657096

de Fine LK, de Fine LJ (2020) Artificial intelligence, transparency, and public decision-making: why explanations are key when trying to produce perceived legitimacy. AI Soc 35:917–926. https://doi.org/10.1007/s00146-020-00960-w

Dencik L, Hintz A, Redden J, Treré E (2019) Exploring data justice: conceptions, applications, and directions. Inf Commun Soc 22(7):873–881. https://doi.org/10.1080/1369118X.2019.1606268

Dickey MR (2020) Facebook content moderators demand safer working conditions. Tech Crunch. https://techcrunch.com/2020/11/18/facebook-content-moderators-demand-safer-working-conditions/

Dourish P (2016) Algorithms and their others: algorithmic culture in context. Big Data Soc. https://doi.org/10.1177/2053951716665128

Facebook Investor Relations (2020) Facebook reports fourth quarter and full year 2020 results. Facebook. https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx

Facebook Oversight Board (2021) Case decision 2021-006-IG-UA. Facebook Oversight Board. https://www.oversightboard.com/decision/IG-I9DP23IB

Feenberg A (2006) What is philosophy of technology? Defin Technol Lit. https://doi.org/10.1057/9781403983053_2

Franklin S (2014) History, motivations, and core themes. In: Frankish K, Ramsey W (eds) The Cambridge handbook of artificial intelligence. Cambridge University Press, Cambridge, pp 15–33

Gerrard Y (2020) The COVID-19 mental health content moderation conundrum. Soc Media Soc 6(3):1–5. https://doi.org/10.1177/2056305120948186

Gillespie T (2020) Content moderation, AI, and the question of scale. Big Data Soc 7(2):1–5. https://doi.org/10.1177/2053951720943234

Goodfellow I, Bengio Y, Courville A (2017) Deep learning. MIT Press, Cambridge

Gray JE (2021) The geopolitics of 'platforms': the TikTok challenge. Internet Policy Rev. https://doi.org/10.14763/2021.2.1557

Gray ML, Suri S (2019) Ghost work: how to stop silicon valley from building a new global underclass. Houghton Mifflin Harcourt

Gunkel D (2020) An introduction to communication and artificial intelligence. Polity Press, Cambridge

Gunning T (1991) Heard over the phone: the Lonely Villa and the de Lorde tradition of the terrors of technology. Screen 32(2):184–196. https://doi.org/10.1093/screen/32.2.184

Harwell E (2018) AI will solve Facebook's most vexing problems, Mark Zuckerberg says. Just don't ask when or how. The Washington Post. https://librarylink.uncc.edu/login?url=https://www-proquest-com.librarylink.uncc.edu/blogs-podcasts-websites/ai-will-solve-facebook-s-most-vexing-problems/docview/2024339331/se-2?accountid=14605

Helmond A, Nieborg DB, van der Vlist FN (2019) Facebook's evolution: development of a platform-as-infrastructure. Internet Hist 3(2):123–146. https://doi.org/10.1080/24701475.2019.1593667

Hern A (2021) Facebook leak underscores strategy to operate in repressive regimes. The Guardian. https://www.theguardian.com/technology/2021/mar/23/facebook-leak-underscore-strategy-operate-repressive-regimes

Huhtamo E, Parikka J (2011) Introduction: an archaeology of media archaeology. In: Huhtamo E, Parikka J (eds) Media archaeology: approaches, applications, and implications. University of California Press, Berkeley, pp 1–21

Ikeda S (2021) TikTok faces barrage of new complains from EU consumer protection groups. CPO Magazine. https://www.cpomagazine.com/data-privacy/tiktok-faces-barrage-of-new-complaints-from-eu-consumer-protection-groups/

Ingham T (2021) TikTok users watch the app for an average of 89 minutes per day—and there's 732M of them worldwide. Music Business Worldwide. https://www.musicbusinessworldwide.com/tiktok-users-watch-the-app-for-an-average-of-89-minutes-per-day-and-theres-732m-of-them-worldwide/

Introna L, Wood D (2004) Picturing algorithmic surveillance: the politics of facial recognition systems. Surveill Soc 2(2/3):177–198

Iyengar R (2020) This is what it's like when a country actually bans TikTok. CNN. https://www.cnn.com/2020/08/13/tech/tiktok-ban-trump-india/index.html

Jan B, Farman H, Khan M, Imran M, Ul Islam I, Ahmad A, Ali S, Jeon G (2019) Deep learning in big data analytics: a comparative study. Comput Electr Eng 75:275–287. https://doi.org/10.1016/j.compeleceng.2017.12.009

Jia L, Ruan L (2020) Going global: comparing Chinese mobile applications' data and user privacy governance at home and abroad. Internet Policy Rev. https://doi.org/10.14763/2020.3.1502

Jibilian I (2020) Facebook is changing its 'race-blind' hate speech algorithm to prioritize flagging content it deems 'the worst of the worst,' after backlash over removing Black users' posts. MSN. https://www.msn.com/en-us/news/us/facebook-is-changing-its-race-blind-hate-speech-algorithm-to-prioritize-flagging-content-it-deems-the-worst-of-the-worst-after-backlash-over-removing-black-users-posts/ar-BB1bBOod

Johnson JA (2014) From open data to information justice. Ethics Inf Technol 16:263–274. https://doi.org/10.1007/s10676-014-9351-8

Jordan M (2019) Artificial intelligence—the revolution hasn't happened yet. MIT Press. https://hdsr.mitpress.mit.edu/pub/wot7mkc1/release/9

Kaye DBV, Chen X, Zeng J (2020) The co-evolution of two Chinese mobile short video apps: parallel platformization of Douyin and TikTok. Mob Media Commun. https://doi.org/10.1177/2050157920952120

Kelly M (2020) Republicans learned to live with Big Tech in latest CEO hearing. The Verge. https://www.theverge.com/2020/11/17/21572268/senate-judiciary-committee-dorsey-zuckerberg-hearing-graham-sasse

Kennedy M (2020) 'If the rise of the TikTok dance and e-girl aesthetic has taught us anything, it's that teenage girls rule the internet right now': TikTok celebrity, girls and the Coronavirus crisis. Eur J Cult Stud 23(6):1069–1076. https://doi.org/10.1177/1367549420945341

Knight W (2019) This technique can make it easier for AI to understand videos. Wired. https://www.wired.com/story/technique-easier-ai-understand-videos/

Kumar S (2019) The algorithmic dance: YouTube's Adpocalypse and the gatekeeping of cultural content on digital platforms. Internet Policy Rev 8(2):1–21. https://doi.org/10.14763/2019.2.1417

Latour B (1990) Technology is society made durable. Sociol Rev 38(1):103–131. https://doi.org/10.1111/j.1467-954X.1990.tb03350.x

Lee MK (2018) Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. Big Data Soc 5(1):1–16. https://doi.org/10.1177/2053951718756684

Leo H, Lunden I (2021) TikTok emerges as political battleground in Navalny-stirred Russia. Tech Crunch. https://techcrunch.com/2021/02/11/tiktok-emerges-as-a-political-battleground-in-navalny-stirred-russia/

Lin P (2021) TikTok vs Douyin: a security and privacy analysis. Citizen Lab. https://citizenlab.ca/2021/03/tiktok-vs-douyin-security-privacy-analysis/

Lupton D, Southerton C (2021) Beyond wicked facebook: a vital materialist perspective. In: McKenzie J, Patulny R (eds) Emotional landscapes, dystopia and future imaginaries. Bristol University Press, Bristol

Merrill N, Weber S (2020) Web site blocking as a proxy of policy alignment. First Monday. https://doi.org/10.5210/fm.v26i1.11415

Milne G (2020) Smoke and mirrors: how hype obscures the future and how to see past it. Little, Brown Book Group, London

Møhl P (2020) Seeing threats, sensing flesh: human–machine ensembles at work. AI Soc. https://doi.org/10.1007/s00146-020-01064-1

Monea A, Packer J (2016) Media genealogy and the politics of archaeology. Int J Commun 10:3141–3159

Mühlhoff R (2019) Human-aided artificial intelligence: or, how to run large computations in human brains? Toward a media sociology of machine learning. New Media Soc 22(10):1868–1884. https://doi.org/10.1177/1461444819885334

Murphy K (2012) Machine learning: a probabilistic perspective. MIT Press, Cambridge

Natale S, Ballatore A (2020) Imagining the thinking machine: technological myths and the rise of artificial intelligence. Convergence 26(1):3–18. https://doi.org/10.1177/1354856517715164

Newton C (2019) The trauma floor: the secret lives of Facebook moderators in America. The Verge. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

Newton C (2020a) Three takeaways from a visit to TikTok's new transparency center. The Verge. https://www.theverge.com/interface/2020/9/11/21430822/tiktok-transparency-visit-tour-algorithms-for-you-page

Newton C (2020b) TikTok has a bold new plan to win over regulators. The Verge. https://www.theverge.com/interface/2020/7/31/21348172/tiktok-algorithms-transparency-accountability-review-lawmakers-michael-beckerman-interview

Neyaz A, Kumar A, Krishnan S, Placker J, Liu Q (2020) Security, privacy and steganographic analysis of FaceApp and TikTok. Int J Comput Sci Secur 14(2):38–59

Noble S (2018) Algorithms of oppression: how search engines reinforce racism. New York University Press, New York

Ochigame (2020) Informatics of the oppressed. Logic 11. https://logicmag.io/care/informatics-of-the-oppressed/

Öhman C, Aggrawal N (2019) What if Facebook goes down? Ethical and legal considerations for the demise of Big Tech. Internet Policy Rev 9(3):1–17. https://doi.org/10.14763/2020.3.1488

O'Neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Penguin Books, New York

Ouchchy C, Coin A, Dubljević V (2020) AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. AI Soc 35:927–936. https://doi.org/10.1007/s00146-020-00965-5

Panday J (2020) The hypocrisy of a US, TikTok ban. Internet Governance. https://www.internetgovernance.org/2020/07/28/the-hypocrisy-of-a-u-s-tiktok-ban/

Pariser E (2011) The Filter bubble: what the internet is hiding from you. Penguin Press, New York

Parisi L (2013) Contagious architecture: computation, aesthetics, and space. MIT Press, Cambridge

Qu T (2020) TikTok and China version Douyin surpass 2 billion download milestone, underlining continued appeal. South China Morning Post. https://www.scmp.com/tech/apps-social/article/3082285/tiktok-and-china-version-douyin-surpass-2-billion-download

Robertson A (2020) Facebook says AI has fueled a hate speech crackdown. The Verge. https://www.theverge.com/2020/11/19/21575139/facebook-moderation-ai-hate-speech

Rocque SR (2019) The history of TikTok. Teen Vogue. https://www.teenvogue.com/story/tiktok-what-is-it

Seaver N (2017) Algorithms as culture: some tactics for the ethnography of algorithmic systems. Big Data Soc. https://doi.org/10.1177/2053951717738104

Slack JD, Wise JM (2005) Culture and technology: a primer, 2nd edn. Peter Lang, New York

Sonderby C (2020). Our continuing commitment to transparency. Facebook. https://about.fb.com/news/2020/11/biannual-transparency-report/

Statista (2021) Most popular social media apps in the U.S. 2019, by engagement. Statista. https://www.statista.com/statistics/579358/most-popular-us-social-networking-apps-ranked-byengagement/

Strauß S (2018) From big data to deep learning: a leap towards strong AI or 'Intelligentia obscura'? Big Data Cognit Comput 2(3):1–19. https://doi.org/10.3390/bdcc2030016

Swart J (2021) Experiencing algorithms: how young people understand, feel about, and engage with algorithmic news selection on social media. Soc Media Soc 2021:1–11. https://doi.org/10.1177/20563051211008828

Taylor N, Ingraham C (2020) Introduction: clickable media in a plastic world. In: Taylor N, Ingraham C (eds) LEGOfied: building blocks as media. Bloomsbury Academic, New York, pp 1–22

Tech Transparency Project (2020) White supremacist groups are thriving on Facebook. Tech Transparency Project. https://www.techtransparencyproject.org/sites/default/files/Facebook-White-Supremacy-Report.pdf

Thurman N, Schifferes S (2012) The future of personalization at news websites: lessons from a longitudinal study. Journal Stud 13:775–790. https://doi.org/10.1080/1461670X.2012.664341

Tidy J, Smith Galer S (2020) TikTok: the story of a social media giant. BBC News. https://www.bbc.com/news/technology-53640724

TikTok (2020a) How TikTok recommends videos #ForYou. TikTok. https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

TikTok (2020b) TikTok to launch Transparency Center for moderation and data practices. TikTok. https://newsroom.tiktok.com/en-us/tiktok-to-launch-transparency-center-for-moderation-and-data-practices

Treré E (2018) From digital activism to algorithmic resistance. In: Meikle G (ed) The Routledge companion to media and activism. Routledge, London and New York, pp 367–375

Tubaro P, Casilli A, Coville M (2020) The trainer, the verifier, the imitator: three ways in which human platform workers support artificial intelligence. Big Data Soc 7:1–12. https://doi.org/10.1177/2053951720919776

Vincent J (2020) TikTok is opening up its algorithm and challenging competitors to do the same. The Verge. https://www.theverge.com/2020/7/29/21346390/tiktok-algorithm-moderation-policy-transparency-china

Wakefield J (2021) AI: Ghose workers demand to be seen and heard. BBC News. https://www.bbc.com/news/technology-56414491

Walmsley J (2020) Artificial intelligence and the value of transparency. AI Soc. https://doi.org/10.1007/s00146-020-01066-z

Wang C (2020) Why TikTok made its user so obsessive? The AI algorithm that got you hooked. Towards Data Science. https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked-7895bb1ab423

Weimann G, Masri N (2020) Research note: spreading hate on TikTok. Stud Confl Terror. https://doi.org/10.1080/1057610X.2020.1780027

Winner L (1980) Do artifacts have politics? Daedalus 1980:121–136

Ytre-Arne B, Moe H (2020) Folk theories of algorithms: understanding digital irritation. Media Cult Soc. https://doi.org/10.1177/0163443720972314