

WHY INFORMED CONSENT FORMS ALONE CANNOT PROTECT US FROM FALSE INFORMATION: DEFENDING HUMAN AGENCY AMIDST ALGORITHMS IN THE BUSINESS MODEL OF SOCIAL MEDIA

Nathanael V. Navarro | University of the Philippines - Diliman;
Global Compassion Coalition

ABSTRACT:

The advancement of communication technology has allowed us to have quick access to relevant information as much as we are also frequently bombarded with false information. Many theorists argue that this was the result of the optimization of greed and profit, overtaking how we create trust between and among the agents of social media. In particular, this study will show that the frequency of encountering disinformation and misinformation has so much to do with the algorithmic operations behind social media, as these are the main instruments that social media companies use to highly personalize content through their stimulating recommendations of options that users are to take for themselves. Crucial to this issue is the fact that algorithms used by social media are merely serving as cold conduits of data based on “highly engaging content” to optimize the watch time of users, which, in turn, becomes a profiling tool to place targeted advertisements. If we treat this whole scheme as internet research for the operation and development of social media platforms, I argue that this should be subject to ethical considerations. Coming from the purview of research ethics, I will explain why informed consent forms with various terms and conditions of digital platforms are not enough to effectively guard the trust that we value in human connection. This paper will ultimately elucidate, using Onora O’Neill and Neil Manson’s theories, why trust in human interaction is so complex, which will entail rethinking how we appreciate the value of informed consent. Rather than siding with the misconception that informed consent forms must include all information that social media users, as research participants, might encounter in the use of social media to merely create a cover for legal liabilities and a ready-made response for any negative feedback and critical remarks from various participants and stakeholders of the research, I argue that we must pay attention to the inherent complexity of human communication. Trust can only be secured in any type of human communication when we acknowledge that it is open to ambiguity, inferentially rich, and always presumes shared background conditions of epistemic, practical, and ethical commitments. As a propositional act, consenting is always limited in its knowledge but is never abusive. Despite the mistrustful energies and the uncertainties that the operations of algorithms entail, this essay attempts to give hope for us to agentively redirect our lives towards trust-building in social media.

Keywords: false information, social media, research ethics, informed consent forms, trust-building

INTRODUCTION

Coming from the definition of Annette Markham and Elizabeth Buchanan, the 2022 National Ethical Guidelines for Research Involving Human Participants (NEGRIHP)¹ defines Internet Research as a study of “*how people use and access the internet,*” and/or one that “*utilizes or engages in data processing, analysis, or storage of datasets, databanks, and repositories available via the internet,*” and/or one that “*studies large scale production, use, and regulation of the internet by governments, industries, corporations, and military forces.*”² Needless to say, Internet Research, just like any other type of research, is also subject to ethical considerations. Can one also take such a definition to mean that even those who manage social media sites can be considered researchers, essentially because of their activities in relation to social media users? What if those behind the business of social media point out that their intention is not scholarly but merely for the sake of profit-driven interests? What is intriguing, however, in the 2022 edition of NEGRIHP is the foreword of the late bioethicist Dr. Leonardo de Castro. As de Castro once quipped: “*Driven by powerful consumer-oriented market forces, technological advances pose double-edged challenges. The benefits beckon with disarming irresistibility while risks lurk in the background, hardly palpable to the ordinary person.*”³ My hypothesis is that if the internet – social media, in particular – is instrumentalized by businesses to profit from their operation through the use of research elements in the developmental and managerial aspects of their existence, then they should also be subject to ethical examination to prevent the unjustifiable harms inflicted on the well-being of their users. I hold the view that relational imbalance and knowledge barriers, which hinder the objective evaluation of risks, benefits, and burdens, are also coercive and create vulnerability in research partnerships, especially those with lucrative outcomes.⁴

¹ The referencing of the 2022 NEGRIHP in this research does not mean that the essay will not also critically examine other research practices and socio-political phenomena beyond the Philippines. The author finds the reference to the 2022 NEGRIHP as highly relevant, as it is deemed to be an inclusive and equitable way to manage and develop research endeavors compared to many research guidelines.

² Philippine Health Research Ethics Board (Ad Hoc Committee for Updating the National Ethical Guidelines), *2022 National Ethical Guidelines for Research Involving Human Participants* (Bicutan, Taguig City: Philippine Council for Health Research and Development, 2022), 122. I applied this ethical guideline even when discussing issues beyond the Philippines, because I believe that the elements of it that evoke and promote equity, inclusiveness, and justice are more prominently displayed compared to other guidelines in research ethics.

³ Philippine Health Research Ethics Board, *2022 NEGRIHP*, vii.

⁴ Leonardo De Castro, “Exploitation in the Use of Human Subjects for Medical Experimentation: A Re-Examination of Basic Issues,” *Bioethics* 9 (July 1995): 262. DOI: 10.1111/j.1467-8519.1995.tb00360.x.

Even if businesspeople use the 2022 NEGRIHP to define what research ethics means, I strongly believe that such an ethical guideline cannot stand as a gobbledygook to deter us from ethical mindfulness in the business operations behind the administration of social media. In fact, in the Philippines, as a general ethical guideline in all research involving human participants, the participation of a consentor to a consentee (i.e., one requesting for acceptance of participation after the disclosure of necessary information about the research) must have in mind a justification based on its “social value” as founded on real contributions of expanding understanding about crucial issue/s and the promotion of well-being. As the 2022 NEGRIHP spelled out the ethical dimension behind the procedural means of securing informed consent: “*By informing the potential participants of the purposes of the research project [through] repetition and explanation, answering their questions as they arise, ensuring that they understand each procedure, and obtaining agreement from them, researchers elicit their informed consent, and in doing so, manifest respect for their dignity and autonomy.*”⁵ This only means that any research enterprise, including those of the internet and social media administrators, must not be employed with only an eye for profit at the expense of its humane purpose. An efficient communication and meaningful understanding, ipso facto, is a must in getting consent, as this is not merely a ritualistic act of the bureaucracy to escape legal liabilities after the failure to secure informed consent from human participants. This communication will not only give an element of intelligibility for the comprehension of human participants, but it will also enable the researchers to know if there are harmful consequences in the conduct of research beyond what they expected.

The problem, however, lies in the algorithmic decision-making prevalent now in the management of social media. It is a problem that opens the complexity of accountability and agency, as the data-trained set of rules executes a decision with almost nothing and sometimes without direct human facilitation. This becomes possible when a machine-learning-derived algorithm becomes highly reliant on the user-provided data to create unique sets of predictions of how to resolve a problem or create an arrangement in a certain transaction and/or usage of a social media platform. This means that there is a limitation to what a consentor may acquire in an informed consent form, that is, even before the use of a social media platform. Rather than offering a very granulated and nuanced explanation of its business operations, the potential social media user – as a potential consentor – may only get information containing the platform’s system functionality: “*The logic, significance, envisaged consequences, and general functionality of an automated decision-making system, e.g., the system’s requirements specifications, decision trees,*

⁵ Philippine Health Research Ethics Board, 2022 NEGRIHP, 16.

pre-defined models, criteria, and classification structures.”⁶ Meaning to say, it is impossible to provide a specific ex-ante explanation when a consentee tries to get informed consent from a potential user of a social media platform. But before labeling this limitation as a sinister ingenuity for masking vital information from the users, it is important to note that a large number of social media users make the use of algorithmic decision-making necessary despite the foreseen and admitted lapses in its operation. After all, even in human intervention, lapses are still inescapable. Another point to also consider is that such a decision system, according to many, is premised on the intention “*to minimize humans’ predetermined or biased decision criteria and, instead, let the machine – with its unprecedented computational capacity for big data – find insights and make decisions for humans. Thus, in varied commercial contexts – especially those in which highly advanced machine learning technologies are used – an [ex-ante] explanation can often hardly be specific or complete about which personal data the machine will ultimately use, the ways they will be used, or what kinds of inferences or insights will ultimately be gleaned from computational processes.*”⁷ However, this intention misses the fact that societal biases can be reinforced by the algorithm. Consider what Latanya Sweeney found out in her study: Google ads for arrest records were displayed more often when searching for names usually associated with Black people than white people because as the algorithm detects how often many people click arrest records of Black people than white people, the algorithm modifies what is to be shown more frequently because of the “popularity” it spotted.⁸

What makes our problematization sensible is that we have to first understand what is wrong with this systemic manipulation; otherwise, we are simply like a dog with a bone, missing the meat of our difficulty. According to Shoshana Zuboff, this algorithmic operation, as grounded on an advertising model, is problematic because it has the tendency to treat the data from users as raw materials that leads to further “commodification of subjectivity.” Massive automation is highly revered for more profit at the expense of more humane transactions. This logic of accumulation in this period of “surveillance capitalism” gives rise to “*methods of production of ‘big data’ from small data and the ways in which ‘big data’ are valued reflect the formal indifference that characterizes the firm’s relationship to its population of ‘users.’ Populations are the sources from which*

⁶ Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,” *International Data Privacy Law* 7 (2017): 78. DOI: <https://doi.org/10.1093/idpl/ix005>.

⁷ Tae Wan Kim and Bryan R. Routledge, “Why a Right to Explanation of Algorithmic Decision-Making Should Exist,” *Business Ethics Quarterly* 32 (January 2022): 86. DOI: 10.1017/beq.2021.3.

⁸ Latanya Sweeney, “Discrimination in Online Ad Delivery,” *Communications of the Association of Computing Machinery* 56 (2013): 44-54. DOI: <http://dx.doi.org/10.2139/ssrn.2208240>.

*data extraction proceeds and the ultimate targets of the utilities such data produce.”*⁹ As such, rather than simply defining big data with too much technical detachment – as consisting of massive, rapidly expanding datasets in numerous forms that are extremely intricate and too large for conventional data systems to manage and analyze effectively – one must also worry about the power that it has to change individual’s behavior, how we interact, and how we perceive reality. Following the critical observation of Zuboff, one may argue that there is a need to explore the duplicitous nature of this profiteering behind datafication and digitization of our actions, as we use our technologies to restore the democratic platform of our power to politically question even the seemingly subtle forms of control. No matter how opaque this awareness is, the serious concern is how we make ourselves understand the very condition that makes power asymmetry possible: that is, the trajectory of this information economy in which “*we need to know how it is collected, what it is used for, how it is stored, who it is shared with, how it is sold, and what happens if the security of this data is breached.*”¹⁰ The problem, after all, is not technological advancement *per se*, but the particular “business model” that espouses an incentive “*to find customers ready to pay to modify someone else’s behavior.*”¹¹ The problem is in the hidden power that allows the algorithm to create a correlation of data from an individual’s activity between and among other users, which makes possible the “theorizing” of the most probable arrangement of prolonged and effective engagement – that is, of the most attention-captivating – more predictable for advertising managers. Jaron Lanier pictured this business model as making the algorithm reduce emotions, attention, and political issues in the game of signal optimization for the sake of monetary incentives.¹²

Despite the aforementioned issues, the complacency of most social media users rests on another important political factor: the libertarian perception that strong regulation is tantamount to an asphyxiating restriction of individual rights to express and communicate.¹³ At the other end of the spectrum, due to political ends of ideological correctness, for instance, the Chinese state uses computational propaganda that harnesses

⁹ Shoshana Zuboff, “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization,” *Journal of Information Technology* 30 (2015): 79. DOI: <https://doi.org/10.1057/jit.2015.5>.

¹⁰ Jane Andrew et al., “Data Breaches in the Age of Surveillance Capitalism: Do Disclosures Have a New Role to Play?,” *Critical Perspectives on Accounting* 90 (2023): 6. DOI: <https://doi.org/10.1016/j.cpa.2021.102396>.

¹¹ Jaron Lanier, *Ten Arguments for Deleting Your Social Media Accounts Right Now* (New York: Henry Holt and Company, 2018), 26.

¹² Lanier, *Ten Arguments for Deleting Your Social Media Accounts Right Now*, 32. Also see Onora O’Neill, *A Philosopher Looks at Digital Communication* (Cambridge: Cambridge University Press, 2022), 101–104.

¹³ Onora O’Neill, “Some Limits of Informed Consent,” *Journal of Medical Ethics* 29 (2003): 5. DOI: [10.1136/jme.29.1.4](https://doi.org/10.1136/jme.29.1.4).

targeted censorship with a lack of transparency and justifiability. From DNS poisoning and other IP blocking of foreign websites to the close monitoring of domestic internet, the Great Firewall and Golden Shield Project either redirect to irrelevant sites and/or block the internet users from websites that they are accessing or manually work to block domestic websites and sensitive keywords and/or hashtags that may contain information that the Chinese government does not want. However, one must not take away the algorithm from this issue. Through an algorithmic scheme, these sinister ways of censorship – automatic and manual – eventually combine: As a recently sensitive topic develops, a fresh automatic filtration will come out, which is trailed by periodic manual changes to ensure that the filter is effectively censoring only necessary content and not those that are merely detected as related to unwanted information.¹⁴ While this censorship may sound damaging to our freedom, one must also keep an eye on the libertarian ideology behind social media administration that brings leniency and deception. Unsound conspiracy theories also spread like wildfire due to the internet, as in the case of *Info Wars* (a conspiracy website) using the YouTube videos of Bill Gates presenting in a TED Talk in 2015, which prefigured how the October 2019 pandemic simulation in the Brazilian Amazon forest would take place as a preparation to a so disastrous an epidemic that possibly can cause human extinction.¹⁵ This intentional distortion of facts made many people think that Gates had anything to do with the engineering and discharge of the COVID-19 pandemic, as he was already manufacturing a vaccine against the viral infection.¹⁶ This is because the advertising model that makes us access various websites for free can also be abused for profit itself, as in the case of the Red Cross, an organization committed to health and risk reduction, inadvertently placed ads beside

¹⁴ Gillian Bolsover, “China: An Alternative Model of a Widespread Practice,” in *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, eds. Samuel C. Woolley and Philip N. Howard (New York: Oxford University Press, 2019), 226.

¹⁵ Raul Rodriguez-Ferrandiz, “The Pandemic and Its Apostles: Conspiracy Theories in Pandemic Mode,” in *Digital Totalitarianism: Algorithms and Society*, ed. Michael Filimowicz (New York: Routledge, 2022), 75. DOI: 10.4324/9781003173304-4.data

¹⁶ In a published article, the biologist Kristian Andersen clarified that it is impossible for COVID-19 to be an artificially engineered product because it needs to be seen that it is a combination of known viruses. COVID-19, especially its spike protein, exhibits unique and unparalleled characteristics that cannot be easily replicated, yet it shares common features with viruses found hosted in pangolins and bats. Therefore, the most plausible theory suggests that the virus was transmitted from animals to humans, a phenomenon known as zoonosis. See Kristian G. Andersen, et. al., “The Proximal Origin of SARS-CoV-2,” *Nature-Medicine* 26 (2020), 450-452. DOI: <https://doi.org/10.1038/s41591-020-0820-9>.

COVID-19 misinformation.¹⁷ In fact, Global Disinformation Index, a nonprofit organization, reported in September 2019 that Google was placing ads with 70% of many of the dubious sites (also generally the most curiosity-riveting and attention-captivating for internet users) notorious for fact-checking organizations, creating a ballpark figure of a total of two-hundred-thirty-five million dollars of ad exchange revenue.¹⁸ If in China, the algorithm is used for very undemocratic purposes with explicit ideological ends, one of the protective rationales used by Facebook in rejecting the Biden campaign to remove a video advertisement is their so-called defense of “free press” and “political speech” – even if it means making people falsely believe that Biden provided a sort of “quid pro quo” amounting to a billion-dollar foreign aid to Ukraine on the condition that it will stop investigating a company connected to Biden’s son.¹⁹ Indeed, corruption of power, information asymmetry, and manipulation get in the way of genuinely using social media for truthful democratization.

Is the foregoing discussion to be interpreted as allowing glitches and erroneous decisions to be made by algorithmic operations? To allay this worry necessitates reasonable constraints on the power of the individualistic sense of autonomy, which haphazardly labels any form of control as plainly “paternalistic.” It comes as no surprise that the demand for structural reform of data privacy is pushed by a recognition that “*complex data-intensive information flow [has] long replaced the linear data collection practices with clearly articulated responsible actors.*”²⁰ In particular, there is a limitation to the transparency that can be given in what we currently call “informed consent” since the complexity of algorithmic operation becomes more and more intricate and ever-changing in the processing of personal data.²¹ This includes, *inter alia*, the future processing of personal data, its sharing and disclosure, other implicitly defined benefits to be generated from such, and other possible risks and harms. Some may well argue that the solution is to force dominant companies to open their algorithms. But even though we know that the concentration of a great amount of feedback data on Amazon’s and Google’s dominance

¹⁷ Issie Lapowsky, “In a Letter to Pichai, Top Philanthropists Slam Google for Placing Charity Ads on Disinfo Sites,” Protocol, August 13, 2020, <https://www.protocol.com/google-ads-charities-disinformation-sites>.

¹⁸ Global Disinformation Index. “The Quarter Billion Dollar Question: How is Disinformation Gaming Ad Tech?” Global Disinformation Index, September 22, 2019, https://www.disinformationindex.org/files/gdi_ad-tech_report_screen_aw16.pdf.

¹⁹ Cecilia Kang, “Facebook’s Hands-Off Approach to Political Speech Gets Impeachment Test,” *New York Times*, October 8, 2019, <https://www.nytimes.com/2019/10/08/technology/facebook-trump-biden-ad.html>.

²⁰ Alexandra Giannopoulou, “Algorithmic Systems: The Consent is in the Detail?,” *Internet Policy Review* 9 (March 2020): 7.

²¹ Cf. Philippine Health Research Ethics Board, 2022 *NEGRIHP*, 18-19.

is due largely to the fact of the innovative feedback effect that these companies get from their wide user base, the suggestion is still couched to the seemingly illusory premise that such disclosure can catch the ever-changing adaptive system, as the bolt-like learning of the system from new data only gives a glimpse into the past.²² Moreover, it is not a far-off possibility that even with the smaller competitors benefitting from more political data-sharing mandates, if at all that is possible, the companies may still coalesce with one another in embracing a manipulative kind of advertising model as they are merely all for profit, by hook or by crook – that is, they are just equally competing to answer “Who is the more attention-gravitating of them all?” and nothing more.

Bigger than these worries, perhaps, is the idea that informed consent is required in the face of “radical ignorance,” that is, when undergoing experimentation in which certainties of necessary information are impossible to obtain. But, as Etye Steinberg argues, the validity of consent – in contrast to informed consent – can be secured even if we are radically ignorant about the algorithmic operations of social media – by protecting the autonomy of the social media users and not merely the indiscriminate “interests” of those who manage the algorithmic operations behind social media platforms.²³ This means that we need to create normative and institutional settings that place consenting individuals – even under imperfect information – to place ethical restrictions on the power of tech administrators seeking consent to avoid exploiting the vulnerability of individuals. As Steinberg analogized, this ethical intervention of valid consent under imperfect information can also be gleaned when it comes to experimental treatment of a terminal patient concerning “*state regulation, hospital ethics committees, professional norms and codes of conduct for medical practitioners*” that are secured, and ensuring that institutions live up to these ethical standards.²⁴ This will also ease the patient or technology user from being overwhelmed with Terms and Use agreements, which, in the first place, cannot exhaustively put everything into detail. Steinberg’s claim, however, lies in two fundamental assumptions: “*First, we do not know what of the information that is made available to the AI will end up being important for its decision. Second, we don’t know what information the AI will gain access to in the future.*”²⁵ These assumptions seem to proceed with an idea of privacy self-management through consent intermediaries by which a social media user can keep track and visualize the data flows from consent metadata and other data sources down to initial data collectors and online ad platforms

²² Viktor Mayer-Schönberger and Thomas Ramge, *Reinventing Capitalism in the Age of Big Data* (New York: Basic Books, 2018), 167-171.

²³ Etye Steinberg, “AI, Radical Ignorance, and the Institutional Approach to Consent,” *Philosophy and Technology* 37 (2024): 20. DOI: <https://doi.org/10.1007/s13347-024-00787-z>.

²⁴ Steinberg, “AI, Radical Ignorance, and the Institutional Approach to Consent,” 8.

²⁵ Steinberg, “AI, Radical Ignorance, and the Institutional Approach to Consent,” 6.

(and reevaluate this flow and block when intended), but without worrying that this will only create a simplistic and individualistic cost-benefit analysis to the problem without effectively changing the problematic aspects of the business model.²⁶ This can, hypothetically speaking, make intermediaries so powerful that it limits the choices of the individuals who are overly reliant on the decisions, options, and interpretations presented to them. It is, therefore, important to investigate if it is technologically impossible for some principles and ethical considerations to serve as “humanizing” limits to how the algorithmic operations will work rather than simply wheedling with the idea that we are not in control, of whatever sort, to what is significant to our decision-making – one that holds together the internet with accountability and ethical purposefulness in the limits of information and data management.

In the next sections of the study, I will argue why trust is so precious that we cannot let informed consent forms define our genuine agreement with social media companies to communicate and interact with others, not merely for efficiency but also for truthfulness. While we can treat algorithmic operations as internet research subject to ethical standards, an informed consent form is just a surface of the underlying elements and problems that merit our discursive attention. The succeeding discussions will attempt to show how our agency should not only be protected by informed consent but also how our agential character can protect us from deception, manipulation, and knowledge-lopsidedness.

WHAT IS TRUST?

With all the deception and manipulation we have gotten ourselves into, we often fall into the follies of embracing an illusory understanding of trust. Onora O’Neill reminds us that the nature of trust lies with all the uncertainties that we may feel about it, rather than fantasizing about a life without disappointments and of complete reliability. This is the very reason why we value a relationship filled with trust, as cheating and betrayal are not unavoidable in this world of human freedom. It is also for this reason that O’Neill criticizes the obsession of contemporary society to dovetail “trustworthy” expectations with the advancement of technology, whereby we make ourselves hamstrung by the illusion of “complete transparency.” As Onora O’Neill delivered a rather enigmatic and demanding ethical intervention on how we define trust:

To place trust reasonably[,] we need to discover not only *which* claims or undertakings we are invited to trust, but *what* we might reasonably think

²⁶ Tuukka Lehtiniemi and Yki Kortesniemi, “Can the Obstacles to Privacy Self-Management be Overcome? Exploring the Consent Intermediary Approach,” *Big Data & Society* 4 (2017): 7-9. DOI: 10.1177/2053951717721935.

about them.²⁷

Reasonably placed trust requires not only information about the proposals or undertakings that others put forward, but also information about those who put them forward. Gullible people who put their trust eagerly in blind dates, or pyramid selling schemes, or snake oil merchants[,] and other unlikely purveyors of sure-fire magic do so on the basis of patently inadequate evidence about those who make the proposals they accept; they get our pity or derision, but hardly our sympathy. We reserve that sympathy for people who cannot judge those who deceived them. If we are to place trust with assurance[,] we need to know *what* we are asked to believe or accept, and *who* is soliciting our trust.²⁸

This elucidation from O'Neill puts us in the difference between misinformation and disinformation: the former deals with believing and spreading false information without the intention to deceive other people, while the latter radiation of falsehood is a deliberate intention to destroy, harm, and pervert others' capacity to think and act.²⁹ At the heart of this message is the belief that the lack of transparency is not the archenemy of trust, but rather the intention to deceive. Unsorted information circulated for the sake of transparency may only lead to misinformation. What we need is an active inquiry beyond the concurrence of various information and claims. Trust arises from the motivation to "*trace specific bits of information and specific undertakings to particular sources on whose veracity and reliability we can run some checks.*"³⁰ This limited aim of trust and the active inquiry it entails is contrary to the pressure of enveloping ourselves with an unfailing mode of guarantees and auditing. Various professional, organizational, and societal aims are not reducible to quantifiable and strenuous targets of the bureaucratic system. Obsessing with accountability requirements might distort the essence of various social processes and organizational practices.³¹ The frequency of examinations to memorize concepts might shrink the time for more meticulous learning and research to foster critical thinking. Doctors who are disproportionately required to do meticulous record-keeping might find less time to converse with their patients and study their medical conditions. Requiring burdensome procedures to record complaints might discourage people from disclosing corrupt practices in government. All of these make us

²⁷ Onora O'Neill, *A Question of Trust: The BBC Reith Lectures 2002* (Cambridge: Cambridge University Press, 2002), 64.

²⁸ O'Neill, *A Question of Trust*, 64-65.

²⁹ O'Neill, *A Question of Trust*, 70.

³⁰ O'Neill, *A Question of Trust*, 76.

³¹ O'Neill, *A Question of Trust*, 49.

misdirected from the true purpose of our actions, giving us the misleading idea that audit is an end in and of itself. Instead of blindly accepting imposed indicators as the measurement of success and trust, we can scrutinize the integrity of our undertakings “by asking to whom the new audit culture makes professionals and institutions accountable, and for what it makes them accountable.”³²

Trust, according to O’Neill, is not based on informed consent; the latter is merely an expression of the presupposition of the former. The value of informed consent, then, is not that it is the foundation of trust, but it is merely an attempt to express that an action is standing against deception and coercion, which, if there is a failure to be curbed and eliminated, become the conditions that undermine trust.³³ More than transparency, informed consent, to express trust effectively, needs information that is limited yet checkable; one that does not need the inundation of non-assessable information whose accuracy cannot be provided. In the first place, we throw ourselves into a snafu when we reduce everything to be trusted to what is elaborated in the informed consent form, for it is a great mistake to articulate every causal explanation in a propositional attitude, as in the case of consenting. Because of the natural opacity that consent possesses, it falls short of describing – and being transparent about – all the possible logical implications and equivalences that a proposition may entail. O’Neill cited what happened in Alder Hey Children’s Hospital as an example of this reality: while some of the parents consented to remove the human tissue of their children, some suddenly demurred that they had not agreed to the removal of the patient’s organs simply because they thought that their consent did not directly articulate the removal of organs but only of human tissue – failing to consider that fact that it is impossible to remove a tissue without removing an organ, as the former constitutes the latter.³⁴ Not only is it unethical to overwhelm the patient with irrelevant information, but it is also logically impossible to be fixated on being completely specific in consent forms without arriving at unnecessary complexity.

If we understand the nature of propositional attitudes as being of opaque character, then we can understand why it is futile to trust with full certainty, much less to expect a complete description of every possible concern in an informed consent form. O’Neill even argued that we must not think that the ethical value of informed consent forms simply zero in on elaborate statements, securing of signatures, and formal documentation of consent to avoid possible lawsuits, objections, or disgruntlement – even if averting these forms of opposition by the consenter might mean simply accepting any form

³² O’Neill, *A Question of Trust*, 52.

³³ O’Neill, *A Question of Trust*, 86-87.

³⁴ O’Neill, “Some Limits of Informed Consent,” 6.

of damage and abuses inflicted by the consentee.³⁵ Not all actions and procedures under a medical intervention, for instance, can be explicitly stated in an informed consent form to create a long-lasting record without any need for interpretation and inferential knowledge. For instance, there is no need for an additional statement in an informed consent form to record an implied agreement that can be seen when a patient lifts one's arm and stretches it out for a medical doctor to take a blood sample through an injection. According to Neil C. Manson and Onora O'Neill, we must locate the value of trust within the agency of people engaged in real communication; one that makes information not merely for slashing, chopping, and/or severing "chunks," "packets," "signals" of messages from one another like "quasi-physical stuff" that can be "acquired," "stored," "hidden," and/or "accessed" according to how we perceived these to "function." This is what Manson and O'Neill think of as the "container/conduit model" of communication, widespread nowadays. Trust is needed because most of the elements in human communication are out of the control of a single agent due to the culture, beliefs, and underlying assumptions, among others, that each person variably holds. This is also the reason why misunderstanding arises even in the face of elaborate, penned arrangements and conditions, just as it becomes mysterious that even if we are strangers to one another, without explicit and overly sophisticated written agreement about actions and transactions, we oftentimes understand one another. As Manson and O'Neill explained the prevailing saturation of the epistemic and practical significance of what they call the "agency" model of communication:

First, we risk downplaying the fact that we know something only if we have sets of rationally evaluable commitments and if we also have the rational competencies that are required to grasp how propositions stand as evidence for, or in favor of the truth of, other propositions. [...] Second, we risk downplaying the fact that the *significance* that knowledge has for an agent (e.g., the difference it makes to how she acts) depends upon that agent's cognitive and practical commitments.³⁶

The underlying point that we want to stress here is that *information acquires ethical and normative significance not because it is about certain special aspects of the world, but because it can or may be used in certain inferences and actions*. These inferences and actions always depend on background

³⁵ Onora O'Neill, "Informed Consent and Public Health," *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 359 (July 2004): 1134. DOI: 10.1098/rstb.2004.1486.

³⁶ Neil C. Manson and Onora O'Neill, *Rethinking Informed Consent in Bioethics* (Cambridge: Cambridge University Press, 2007), 107.

knowledge, and on competencies, commitments[,] and interests *other* than the mere possession of specific information.³⁷

For Manson and O'Neill, the practical and cognitive commitments are discounted whenever we understand human communication as reducible to the "transference" of semantic content, thereby limiting the agentive vision of human action. The genuine act of informing is downplayed when we simplistically point to the seemingly "context-free," "transportable," and discrete "materials" that we call "information," which means brushing into the rug the normative framework and practical considerations between the speaker and the audience – ultimately involving a certain amount of uncertainty. For the most part, the communicative transaction entails the intention of the people involved, their capabilities, their beliefs, and expectations about one another.³⁸ To illustrate this, suppose you went home from work and your wife suddenly asked you, "*What do you want for dinner?*," to which you would reply, "*It's up to you. Just make sure that it's my comfort food.*" The phrase "*my comfort food*" could be interpreted in many ways, especially if you are a stranger to the speaker. But because you and your wife have an intimate connection, you do not have to exhaust yourself further by enumerating your "*comfort foods*." If you separate "*want*" from the rest of the context, you fail to realize that your wife asked you for the "*dinner*" you want to eat later (rather than next week or next year). Your wife's question is also purposive in the sense that she wants to know what you want to eat, for she might ask this question with the purpose of testing if you already want to talk to her after you had a misunderstanding last week, or because she wants to know if you miss how she cooks for comfort food. The sentence "*It's up to you.*" could be interpreted by your wife as going to the grocery store before cooking your dinner, or she will interpret that with a sense of immediacy so that you can be comforted as she knows that you are tired from work, and you want to feel better. This propositional nature of communication helps us understand better how the evaluative bearing of action and the reasons for action are formed. It is also for this reason that we understand the inferential fertility and contextual relations of our proposals, as we do not have to engage in the meticulous task of knowledge discovery nor proceed to an intricate chain of reasoning, but we only need to draw upon the stock of knowledge that we have. Thus, it takes some trust (and a considerable amount of uncertainty), so to speak, to read between the lines. Take another example to illustrate this: If you want Neo and Trinity (suppose they are your friends) to join you in a private meeting regarding the plans of your organization next year, and all of you do not want other members of your organization to know those plans, you must altogether have a practical commitment to make sure that only among yourselves, in that intended circle of friends, will be

³⁷ Manson and O'Neill, *Rethinking Informed Consent*, 107-108.

³⁸ Manson and O'Neill, *Rethinking Informed Consent*, 41.

able to know where you all will meet at the same time (e.g., using private messaging application, placing a small paper containing private information inside the personal bag of your invited friends, etc.). This invitation from you must be understood by Neo and Trinity, say, by using a secret code that other people cannot understand (but not with an obscure one that Neo and Trinity cannot understand), or by telling them the details about the secret passage to this secret meeting place. While defining trust in human communication seems nebulous when taken from a more complex and realistic perspective, it creates three requirements, according to Manson and O'Neill, in which the speech acts should play against the background of common agreement of multifarious sets of commitments, social expectations, and/or collective ways of understanding in which the commitments of others can be adjusted: first, sharing of a language between speaker and audience; second, shared background knowledge about the world, reality, and other social conventions to bring about the idea of feasibility; and third, the speakers must attempt to grasp the limited capacities and competencies of one's audience.³⁹

WHAT IS WRONG WITH THE ALGORITHMIC SYSTEM IN RELATION TO THE VALUE OF INFORMED CONSENT

In this part of the research, I will give a philosophical commentary as well as accurate descriptions of the cavernous work and implications of algorithmic operations on YouTube, X (née Twitter), Facebook, and the internet. Emphasizing the role of algorithms in the proliferation of misinformation and disinformation, I will link this phenomenon, or better yet, political manipulation and negligence, in relation to the value of trust in human communication and the limits of informed consent. The main points of this section of the study draw a philosophical elucidation of what constitutes the commitment towards trust-building in human interaction from the three requirements that Manson and O'Neill discussed.

It is a common understanding that the players must understand the rules of the game, and they cannot do so without having a shared language to communicate with one another. Otherwise, there is no common understanding upon which all problems and other issues can be deliberated and resolved fairly. This same truth about human communication in which trust is secured is also enshrined in the 2022 NEGRIHP, for it instructs that informed consent forms must be written in English and in a local language that is “*appropriate to the level of understanding of research participants*.”⁴⁰ Transparency, as one of the important elements of research ethics, needs the clarity of language to build bridges between research participants and the researchers. Trust is intended to be guaranteed in any research enterprise, even when disruptions and questionable aspects

³⁹ Manson and O'Neill, *Rethinking Informed Consent*, 56-57.

⁴⁰ Philippine Health Research Ethics Board, *2022 NEGRIHP*, 47.

of research arise. For this reason, the 2022 NEGRIPH requires researchers to state their affiliation and affinities, as well as their conflicts of interest that might affect the objectivity and integrity of the research and the relationship between the participants and the researchers.⁴¹ Such a digital bridge is necessary especially now that the internet and the algorithmic decision-making process may amplify disinformation and misinformation because of the business model adopted, as in the case of X that has a translated *Terms of Service* in Filipino, which explicitly states that their users will encounter advertisements of various sorts in exchange for their access to their services.⁴² As we have introduced, this is a crucial issue since the claim that algorithmic operations in social media administration can be beneficial for users in deciding on recommendations offered to them by the digital medium that they use must also square to the fact that the user's activities – both the deviation and conforming to the range of recommendations – are used to train the algorithm as a feedback loop for its predictive scheme. According to the Reuters Digital News Report in 2024, 59% of their participants around the world expressed concern about the distinction between real and fake news on the internet.⁴³ Such digital uneasiness is not far from recent surveys in the Philippines. In the survey conducted by the Social Weather Station in December 2021, 51% of adult Filipinos find it hard to recognize misinformation on social media.⁴⁴ Pulse Asia survey conducted in September 2022 showed that 68% of Filipino adults encountered political fake news on social media.⁴⁵ Indeed, it is necessary not only to translate in Filipino the *Terms of Service* and other consensual elements in the use of social media platforms but also to meaningfully and intelligibly articulate the conditions that stabilize the existence of the business of social media. Without this understanding, Filipinos remain unaware of how the business model that they consented to in the use of social media makes them susceptible to receiving misinformation and disinformation. It is as if we are expecting to know that we are shortchanged for something that we do not know how to count, and who should be held accountable.

⁴¹ Philippine Health Research Ethics Board, 2022 NEGRIPH, 28.

⁴² X. “Mga Tuntunin ng Serbisyo,” last modified November 15, 2024, <https://x.com/tl/tos>.

⁴³ Adrian Parungao. “Concern over Fake News Increases – Reuters Study,” last modified June 19, 2024, <https://globalnation.inquirer.net/240039/fwd-reuters-report-on-world-trust-concern-on-real-and-fake-in-news-and-news-avoidance>.

⁴⁴ Richa Noriega. “51% of Adult Filipinos Find Difficulty in Discerning Fake News,” last modified February 17, 2022, <https://www.gmanetwork.com/news/topstories/nation/822135/sws-51-of-adult-filipinos-find-difficulty-in-discerning-fake-news/story/>.

⁴⁵ Gabriel Pabico Lalu. “‘Fake News’ a Problem in PH? 9 in 10 Filipinos Agree, Says Pulse Asia,” last modified October 11, 2022, <https://newsinfo.inquirer.net/1678248/fake-news-a-problem-in-ph-9-in-10-filipinos-agree-says-pulse-asia>.

Digital echo chambers and filter bubbles emerge because the algorithmic operation itself cannot ascertain the truth about the autonomy of the user deciding in relation to the algorithmic recommendations, or whether the users were merely swayed by the attractive recommendations.⁴⁶ It follows that the algorithm tends to shape the decision-making of the users according to the patterns that it creates by processing the commonality of those observed behaviors of users in the past. Its engineering is built on the assumption that such an operation will create efficiency and optimization of engagements. Informed consent about the use of the netizen's data is crucial in this issue since the algorithmic operations' raw materials come from the large-scale collection of submitted data in their profiling and predictive scheme.⁴⁷ Such a pattern-seeking scheme also affects the advertisements to be shown to the users derived from the categories placed in the algorithm, consumption patterns, and other observed online behavior.

Not only is the algorithm modified through the responses fed by the users, but the users can likewise be modified through their use of social media. A telling example of this would be the rise of disinformation and misinformation, which is all the more exacerbated by the atomistic management of data in which the burden of controlling the collection of data is placed alone in the consent of a single person rather than creating a systemic review of informed consent forms and community-wide engagements in the investigation of the wide implications of social media use vis-à-vis the use of personal data.⁴⁸ Indeed, there is a double-edged sword that must be dealt with. But the issue becomes murky when we realize that the problem is not directly caused by social media sites in the sense that they are invariably financing fake news for a propaganda, but it can be magnified when one looks at how social media companies profit from disinformation through political advertisement, which makes many users prone to absorb disinformation. More than being a search engine, Google's revenue mostly comes from its being an advertising company, as it has already overtaken traditional media to become the dominant ad medium in the world as it increased "*to nearly one hundred billion dollars, and it accounted for eighty percent of the company's combined revenue.*"⁴⁹ One way that Google does this is by allowing itself to become an ad platform where organiza-

⁴⁶ Swathi Meenakshi Sadagopan. "Feedback Loops and Echo Chambers: How Algorithms Amplify Viewpoints," last modified February 4, 2019, <https://theconversation.com/feedback-loops-and-echo-chambers-how-algorithms-amplify-viewpoints-107935>.

⁴⁷ Sriya Sidhar. "Collective Consent for Algorithmic Accountability," last modified December 2, 2020, <https://www.internetjustsociety.org/collective-consent-for-algorithmic-accountabilitybaaa8d68>.

⁴⁸ Jeni Tennison. "Community Consent," last modified January 17, 2020, <https://www.jenitennison.com/2020/01/17/community-consent.html>.

⁴⁹ Noah Giansiracusa, *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More* (New York: Apress, 2021), 152.

tions can advertise whatever they want to be included in the leading Google searches, as they “purchased” particular keywords. The other way that Google is profiting from its advertising power is by becoming a middleman to facilitate how various clients intend to display their ads on third-party properties. This second way of Google’s advertising allows popular websites to generate income by allowing these websites to become another channel where ads can be displayed. Those ad-hosting websites are included in the Google Display Network (GDN), where several ads are added every day, too many to be manually managed, and are automatically powered through algorithmic transactions.

According to the 2017 report of the Campaign for Accountability (CfA), the Terms of Service that Google used did not restrict the Google Display Network from spreading disinformation.⁵⁰ Negligence can also be seen in Google’s dashboard, which only allows advertisers to choose political content without dwelling on the difference between false information and quality journalism. Furthermore, according to the CfA report, the anonymization option offered to ad-hosting websites only created a funnel for attention-captivating disinformation to be financially incentivized, as the advertisers tend to choose the web traffic in anonymous ones compared with safe, non-anonymous ones. Thus, this proclivity towards financial conflict of interest getting in the way of pursuing truthfulness must be explicitly stated by Google through its intelligible and meaningful explanation of its business model to consenters of informed consent forms of its *Terms of Service*.⁵¹ Such an urgency does not only come from legitimate advertisers suffering from their damaged brand reputation as their ads are posted on hyper-partisan and misleadingly sensational websites, but also because the netizens absorb misinformation, which contradicts the principle of justice as a crucial element of research ethics. As the 2022 NEGRIHP succinctly reminds us of this ethical dimension: “*It is unjust for one group in society to bear the costs of research while another group reaps its benefits. Research should not worsen existing health and social inequities.*”⁵² If one takes a look at the April 2024 version of the *Terms of Service* of YouTube for their Partner Program in monetizing various channels, there was a Filipino-translated statement that explicitly directs YouTube channels to avoid merely intending to gravitate viewers to their channels without genuine entertainment or anything to learn from it.⁵³ This is, indeed, a great stipulation to address misinformation and disinformation. In this same vein, in their Business Help Center (as

⁵⁰ Campaign for Accountability. “Report: Google Makes Millions from Fake News,” last modified October 30, 2017, <https://campaignforaccountability.org/work/how-google-makes-millions-off-of-fake-news/>.

⁵¹ Corinne Tan, “The Curious Case of Regulating False News on Google,” *Computer Law and Security Review* 46 (2022): 11. DOI: <https://doi.org/10.1016/j.clsr.2022.105738>.

⁵² Philippine Health Research Ethics Board, 2022 NEGRIHP, 26.

⁵³ YouTube Tulong. “Mga Patakarán sa Pag-Monetize ng Channel sa YouTube,” last modified August 12, 2024, <https://support.google.com/youtubelanswer/1311392?sjid=2322536943219315613-AP#zippy=>.

translated in Filipino language), Facebook also explicitly requires its publishers, creators, and third-party providers to follow Partner Monetization Policies (PMP) to get the tools for monetization; one of which is the prohibition of the use of some formats such as those that cradle false information, which was labeled by “third-party checker,” and what the platform deemed as “debatable social issues” and other contents that induce “physical and psychological distress.”⁵⁴ The question, therefore, is: Are the third-party checkers working efficiently and effectively to prevent the spread of misinformation and disinformation? Asking that question also pivots us to give adequate attention to the neglected dimension of social media interaction: fostering critical thinking among netizens through reasonable discussions, which might be simply labeled by Facebook itself as “debatable” and “distressing.” Following O’Neill’s theory on trust-building, we must transition from merely investigating the language and how the *Terms and Conditions* are articulated by social media companies in their informed consent forms to move towards the practical affairs, that is, the feasibility and effectiveness of fact-checking in social media through the company’s efforts, social media users, and third-party fact-checkers. Without this practical side of our study, we end up writing words of high ideals without actively and constantly making efforts to make falsehood largely blotted out of our lives. For instance, in 2020, when YouTube decided to rely on a machine learning approach to combat videos spreading false information, its former Chief Product Officer said that they were suddenly alarmed by the huge number of videos being taken down, as these are not as precise as what humans could do to fact-check YouTube content.⁵⁵ In 2022, over 80 fact-checking institutions around the world reproved YouTube’s ineffectiveness in fact-checking videos. In particular, these organizations lambasted the lack of specific labels to categorize content as already “debunked,” the ineffective prevention of recommendation algorithms to promote repetitive offenders, and the uneven fact-checking services to arrive at a language-specific approach to arrest false information, rather than merely focusing on the English language.⁵⁶ These are indications that show the significance of understanding the context of deception and manipulation instead of merely taking down posts with keywords that are usually associated with disinformative pro-

⁵⁴Meta Business Help Center. “Mga Patakaran sa Pag-monetize ng Content,” https://tl-ph.facebook.com/business/help/1348682518563619?locale=tl_PH.

⁵⁵ YouTube Official Blog. “Protecting our Extended Workforce and the Community,” last modified March 16, 2020, <https://blog.youtube/news-and-events/protecting-our-extended-workforce-and/>. Cf. James Vincent. “YouTube Brings Back More Human Moderators After AI Systems Over-Censor,” last modified September 21, 2020, <https://www.theverge.com/2020/9/21/21448916/youtube-automated-moderation-ai-machine-learning-increased-errors-takedowns>.

⁵⁶ Loreben Toquero. “Global Fact-Checkers Demand from YouTube Effective Action Against Disinformation,” last modified January 13, 2022, <https://www.rappler.com/technology/global-fact-checkers-demand-youtube-effective-action-disinformation/>.

paganda. These are serious implications of disproportionate reliance on automatic yet inflexible control without fostering regulations anchored on situation differences and conceptual judgments.⁵⁷ It also creates an agentive standard towards content regulation that only relies on individual fact-checking in spades, which does not acknowledge individual limitations of the competencies of social media users. It is, for this reason, that we need to understand consent not as a static result but as a continuous process that is revocable anytime that the one who consented no longer wants to participate.⁵⁸ While not all data given is retractable and its dissemination is perhaps impossible to undo, it is the responsibility of researchers, in this case, the social media administrators, to inform participants about the risks and their ineffectiveness to immediately put an end to the spread of false information.

In November 2016, Mark Zuckerberg posted on his Facebook account that his company was taking things seriously. The catch, however, is that restricting free speech can inordinately limit our freedom to express ourselves, which Zuckerberg himself ostensibly considered in his Facebook post. Admitting that striking a balance between these two considerations is “*both technically and philosophically*” complex, Zuckerberg shared that his company’s move is to somehow give the autonomy for social media users to discern the falsity of an information, as Facebook and its human resources do not want to be “*arbiters of truth.*”⁵⁹ In other words, Facebook still believes that social media users have the individual responsibility to guard themselves against untruthful information. If one looks back at O’Neill’s theory, one may recall that one of the requirements in trust-building is knowing the limited capacity and competency of one’s audience. It must be noted that Facebook did not actually abandon helping its users detect false information, but their approach would be “less epistemological” and used instead a technical system that predicts whether or not a particular content is likely to be flagged as false. But this content moderation through the use of machine learning must be doubted, as a report of the Guardian in February 2019 found that even newly created Facebook accounts – without friends or records of activities – would show autocomplete recommendations as “vaccine truth movement,” “vaccine resistance movement,” “vaccine

⁵⁷ Leonardo D. De Castro and Isidro Manuel C. Valero, “Ethics Education Needs More than the Four Principles: Bioethics Discourse in a Community of Inquiry,” in *Global Education in Bioethics* (Switzerland: Springer International Publishing, 2018), 69-80.

⁵⁸ Minerva C. Rivas Velarde et al., “Consent as a Compositional Act – A Framework that Provides Clarity for the Retention and Use of Data,” *Philosophy, Ethics, and Humanities in Medicine* 19 (2024): 8. DOI: 10.1186/s13010-024-00152-0.

⁵⁹Mark Zuckerberg. Facebook status update, last modified November 18, 2016, https://www.facebook.com/zuck/posts/10103269806149061?__mref=message_bubble.

re-education,” among others, when one entered “vaccine” in Facebook’s search bar.⁶⁰ In the said report, it was also shown that anti-vaccination Facebook pages and groups have overridden other search results. The problem with Facebook, even in the midst of the COVID-19 pandemic, is that it was hesitant to ban anti-vax content and merely discouraged posts that might induce “imminent physical harm.”⁶¹ It is as if Facebook were only waiting for the moment when the achievement of herd immunity against the virus would be extremely impossible. What depletes, time and again, the capacity of users to critically examine sources is the situation in which the alternative sources – that is, the truthful ones – are submerged under the deluge of fallacious ones. This is happening because social media companies failed to address the root causes of the problems. For instance, even if Twitter started banning false content on its platform, it was found in a certain study that it failed to address the many bot activities found in non-official Twitter applications that spread deceptive information on Twitter’s platform itself.⁶² That has been the case in Twitter despite the fact that it could easily be contained since the “*tweeting/retweeting network was more connected and homogenous for fake news than it was for traditional news; the fake news network was also more tightly intertwined with the network of right-wing news tweets than it was for center and left-wing news.*”⁶³

Insider information obtained by *Bloomberg News* in 2019 revealed that business leadership on YouTube is unwilling and incapable of changing its algorithm for the sake of intensifying its “engagement” on its platform, even if this means tolerating misleading videos, discriminatory content, and unethical messages.⁶⁴ Giansiracusa argued that this is mainly because YouTube adopted a deep learning algorithm, and was not given any effective regulation for it, as it is used for accumulating profits. Therefore, despite the explicit articulation that a ban may be imposed by YouTube on channels that repeatedly violate its *Terms of Service* and *Community Guidelines*,⁶⁵ recommending videos on the

⁶⁰ Julia Wong. “An Update on Our Work to Keep People Informed and Limit Misinformation,” last modified February 01, 2019, <https://www.theguardian.com/media/2019/feb/01/facebook-youtube-anti-vaccination-misinformation-social-media>.

⁶¹ Guy Rosen. “An Update on Our Work to Keep People Informed and Limit Misinformation,” last modified April 16, 2020, <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>.

⁶² Alexandre Bovet and Hernan Makse, “Influence of Fake News in Twitter During the 2016 US Presidential Election,” *Nature Communications* 10 (2019): 1-14. DOI: 10.1038/s41467-018-07761-2.

⁶³ Giansiracusa, *How Algorithms Create and Prevent Fake News*, 191.

⁶⁴ Mark Bergen. “YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant,” last modified April 2, 2019, <https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant>.

⁶⁵ Cf. YouTube Tulong. “Mga Pagwawakas ng Channel o Account,” <https://support.google.com/youtube/answer/2802168?sjid=7851885953287863292-AP>.

homepage and the “up next” lists based on personalized content from an almost infinite stream or newsfeed, and with the intention not merely to maximize the views of each YouTube video but for the maximization of watch time, the platform becomes a home to ad-hosting videos that is also a host of attractive yet distorting contents. Its cause, as Giansiracusa described, is mainly due to letting the algorithm work by “*whittling down the vast ocean of videos on YouTube to a small pool of a few hundred videos the user might like based on the user’s watched video history, keyword search history, and demographics*” followed by the ranking process of this small pool of videos starting “*from highly recommended to least highly recommended*” videos in the platform and “*including details on the user’s previous interactions with the channel the video is from – such as how many videos the user has watched from this channel and when was the last time the user watched a video from this channel.*”⁶⁶

No less than the then-Chief Product Officer of YouTube, Neal Mohan, revealed in 2018 that the “recommended videos” account for 70% of the total amount that users have spent on their watch time.⁶⁷ This fact may sound unalarming until juxtaposed with a study in 2020 showing that 60% of those who believe in conspiracy theories about the COVID-19 pandemic were produced by radiation from 5G towers depended on YouTube as their source of information.⁶⁸ The influence of algorithms could also be observed in Brazilian elections when the far-right political party of Jair Bolsonaro attributed their large engagement to videos uploaded on YouTube to support their extremist propaganda and conspiracy theories.⁶⁹ According to Giansiracusa, however, YouTube did not completely throw away other countervailing videos to these conspiracy stories; what was happening was that the YouTube algorithm was simply pushing further the psychological drivers of emotions – fear, curiosity, hate, and anger – as it remains a cold neutral instrument to merely optimize watch time.⁷⁰ These provocations and triggers become the magnetizing pool of engagement and watch time that clouds the reasonable assessment of users since they are assuming that a “hidden truth” is lurking somewhere, as several videos have been seemingly corroborating one another’s claims.

⁶⁶ Giansiracusa, *How Algorithms Create and Prevent Fake News*, 73.

⁶⁷ Joan Solsman. “YouTube’s AI is the Puppet Master Over Most of What You Watch,” last modified January 10, 2018, <https://www.cnet.com/tech/services-and-software/youtube-ces-2018-neal-mohan/>.

⁶⁸ Rory Cellan-Jones, “Coronavirus: Social Media Users More Likely to Believe Conspiracies,” last modified June 18, 2020, <https://www.bbc.com/news/technology-53083341>.

⁶⁹ Max Fisher and Amanda Taub, “How YouTube Radicalized Brazil,” last modified August 11, 2019, <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html>.

⁷⁰ Giansiracusa, *How Algorithms Create and Prevent Fake News*, 79.

CAN INFORMED CONSENT ALONE PRESERVE THE TRUST WE NEED?

For Onora O'Neill, the simple answer to this question is no. As we have seen, informed consent becomes a simplistic barometer to measure how we have attained trust when, in fact, we have only shown "trustworthiness." Our obsession with "trustworthiness," according to O'Neill, comes from our fascination with and resignation to the bureaucratic forms of thinking that foster the overreliance on determinate commands and recording of evidence. From a philosophical lens, O'Neill figured out that this phenomenon largely arose from the dubiety for so-called "non-experts" to audit the experts who can shape the determinate structure of how we should collectively conduct our political and other organizational frames. More sharply, O'Neill criticizes the idea of having a "life algorithm," which can allegedly escape underdetermined action without realizing that actions, which "*must always be complemented and implemented by the exercise of judgment; and practical judgment, including ethical judgment, is not a matter of arbitrary choice. [...] We do not generally dither about which of many possible non-deceptive acts to do, because much of the indeterminacy is resolved by the fact that we are always pursuing a range of aims, and simultaneously meeting numerous constraints.*"⁷¹ This is not quite surprising, especially in the exercise of ethical research practice that generally promotes the delicate striking of the balance between possible risks and harm to human participants and the potential societal benefits in the knowledge formation of the research endeavor to be carried out. One may, therefore, criticize the predominant view of social media companies to once and for all bar research access from relational metadata without distinguishing the difference between legitimate scholarly endeavors and those motivated by narrow interests.⁷² Social biases, for instance, are impossible to investigate without network data that uses personal data, for it is impossible that every processing of data should have explicit informed consent in the case of research participants ranging from thousands to millions. For instance, two professors from Northeastern University conducted research concluding that supervised learning in algorithmic prediction – even if initially commanded by non-biased metrics – can reinforce historical biases by chasing patterns that it receives based on recorded users' behavior, as in the case of posting advertisements for job opportunities appearing predominantly in a certain class or biologi-

⁷¹ Onora O'Neill, *Autonomy and Trust in Bioethics* (Cambridge: Cambridge University Press, 2004), 124.

⁷² Anja Bechmann and Jiyoung Ydun Kim, "Big Data," in *Handbook of Research Ethics and Scientific Integrity*, ed. Ron Iphofen (Switzerland: Springer Nature, 2020), 432-436. DOI: https://doi.org/10.1007/978-3-319-76040-7_18-1. Also see Manson and O'Neill, *Rethinking Informed Consent*, 116-127.

cal sex.⁷³ This, indeed, poses a more nuanced and agentic approach to corporate secrecy by transporting data from a rigid understanding of privacy to scholarly security of data confidentiality. Is it not ironic that social media sites cannot curb false information for the sake of “freedom of expression,” while suppressing the development of research dedicated to studying conditions that make disinformation possible? From that realistic perspective, informed consent forms cannot be “fully” informing, but should strive to be “relevantly” informing.⁷⁴ That is, it should enable society not only to conform to standardized procedures to record evidence, nor simply to reduce informed consent to making research participants sign lengthy forms or tick specific boxes to reinforce the idea of so-called “precise” acceptance of terms and conditions. Informed consent forms are merely instruments to ensure a more significant element in social research: “*As relational beings, researchers and participants live in a community, the good of which they ought to foster and promote. Researchers are to always relate to their research participants as kapwa-tao, a fellow human being with whom they have a common dignity, rights, and duties and with whom they are to practice reciprocity and solidarity with each other.*”⁷⁵ Looking at this philosophical understanding gives us the idea that autonomy in research ethics is not merely about individualistic claimants of unsubstantiated demands of so-called “rights” without any obligatory underpinning. O’Neill even argued that we need to recover trust from the threats of unbridled corporatization, marginalization, and deception, which solidifies the understanding of principled autonomy that has a relational bearing – we must give, refuse, assess, and exchange reasons that should be accessible to others. As we have seen, trust necessitates a shared channel of communication, especially in language. As O’Neill gave a Kantian interpretation of this dilemma:

We may approach the topic in an intuitive way by asking why *any* specific ways of thinking and acting should be thought of as having a general authority that would lead us to speak of them as *reasoned*, while others do not. [...] We do not offer others reasons for believing if we communicate with them in ways that we know they will find unintelligible; we do not offer others for acting if we present them with proposals for action that we know they will find unadoptable. To make either thinking or proposals for action followable by others we must impose structure on them: we must make our underlying principles of communication and proposals for ac-

⁷³ Muhammad Ali et al., “Discrimination Through Optimization: How Facebook’s Ad Delivery Can Lead to Biased Outcomes,” *Proceedings of the ACM on Human-Computer Interaction* 3 (November 2019): 1-30. DOI: <https://doi.org/10.1145/3359301>.

⁷⁴ O’Neill, *Autonomy and Trust in Bioethics*, 156.

⁷⁵ Philippine Health Research Ethics Board, 2022 *NEGRIHP*, 86.

tion that are law-like or principles; if we do not, we do not offer reasons for thinking or for acting.⁷⁶

Tying the loose ends, we see why informed consent does not come from an antecededly imposed yardstick of reason as wielded by a dictator, nor does trust in informed consent spring from unruly and capricious interests of individuals simply because they are unrestricted in doing everything that they want. Molding public goods and safety merely according to individualistic inclination is absurd. Such has been the case in the refusal to be vaccinated has been rejected without any medical justification, but merely out of blanket mistrust of possible risks.⁷⁷ As a result, free-riders who enjoyed a high level of immunity before might suffer from communicable diseases when the proportion of the vaccinated population has decreased significantly. An alternative to this extremely individualistic outlook – also an effective move to restore trust – is to limit the sources of misplaced mistrust, which is perpetrated by those who abuse the “freedom of speech” and “freedom of the press” by spreading salacious content, factoids, evidence-lacking report, half-truth dissemination of information, and other sensationalizing strategies simply to get the most of out the popularity and trend. We need to distinguish reasonable regulation from mere censorship.

While we have realized that trust comes from the realistic recognition that human communication and transactions are normative and not completely explicit, we also must ensure that robust conventions will give a sufficient degree of protection from disinformation. Our practical and epistemic commitments to trust involve demanding our government to require media outlets and other distributors of public information to be dedicated to “*practices such as setting out evidence, indicating its sources, and explaining its limitations, as well as structures that provide (for example) for forms of quality control, external assessment or peer review and for sanctions on those who invent or falsify evidence.*”⁷⁸ Ultimately, it creates a helpful demarcation between materials that exercise the value of the “freedom of self-expression” and those that are purported to be reporting the truth but are actually distorting our perception of facts. In this sense, trust is not merely a dispensable by-product of self-expression, but is essentially the truth about making human communication under the ethical purview of directing information to be assessable by and accessible to its intended audience. Facebook indeed closed down the Twinmark Media Enterprises in 2019 after the latter earned a revenue of approximately 400 million pesos from spreading clickbait content of fallacious claims to discount the critics of

⁷⁶ O'Neill, *Autonomy and Trust in Bioethics*, 91.

⁷⁷ O'Neill, “Some Limits of Informed Consent,” 4.

⁷⁸ O'Neill, *Autonomy and Trust in Bioethics*, 182.

the Rodrigo Duterte administration,⁷⁹ but it galvanized another way of profiting from hateful, highly intriguing, and false content not directly related to politics by acquiring various Facebook groups and pages where the write-ups in their websites can be shared.⁸⁰ This is because Twinmark, according to its former employees, is mainly profit-driven, and not merely holding on to guard a political ideology or a leader without anything in return. Twinmark's activities are, indeed, far different from the political motives of those who peddle the Pizzagate conspiracy theories and QAnon's conspiratorial views that aim to heighten the "heroic" personality of Donald Trump, who allegedly was facing a connivance of global elites and "deep state" agents trying to protect a worldwide child sex trafficking syndicate. This prompted the FBI to release a document in May 2019 saying that conspiracy theories can be considered as new terrorist threats.⁸¹ However, after considering YouTube as a so-called "private forum," the US Court of Appeals for the Ninth Circuit ruled that the social media company has the discretion to take down content without abridging the freedom of speech; this was after the PragerU, a conservative channel, accused YouTube of illegally silencing conservative voices on its platform. However, it is obvious that there is a need to philosophically re-examine the grey areas between public and private forums because a social media company like YouTube can be so powerful and enormous in influencing the lives of many citizens of a society.⁸² More explicitly, society has to ask itself whether an activity is prohibiting falsehood or simply denying "legal yet controversial" ideas and information from a source that has been beleaguered with negative receptivity. Perhaps Facebook could superimpose labels on content that they find controversial to some audiences and direct them to further assessable information in order to avoid the situation in which people are merely swayed by public outrage and mob energies on digital platforms. A telling example that hinders achieving such a measure would be the combination of two phenomena: Meta allowing *Free Basics* Program to enter a partnership with telecommunication companies in the Philippines by making people access Facebook without any fee but only allowing a certain portion

⁷⁹ Jonathan Corpus Ong and Jason Vincent A. Cabañes, *Politics and Profit in the Fake News Factory: Four Work Models of Political Trolling* (Riga, Latvia: NATO Strategic Communications Centre of Excellence, November 2019), accessed December 01, 2024, https://stratcomcoe.org/cuploads/pfiles/5december_report_philippines.pdf.

⁸⁰ Camille Elemia, "Cat-And-Mouse Game: Twinmark Fake Network Still Thrives on Facebook," last modified December 23, 2020, <https://www.rappler.com/newsbreak/investigative/remnants-twinmark-media-enterprises-fake-network-still-thrives-facebook-after-takedown/>.

⁸¹ Jana Winter, "Exclusive: FBI Document Warns Conspiracy Theories are a New Domestic Terrorism Threat," last modified August 2, 2019, <https://www.yahoo.com/news/fbi-documents-conspiracy-theories-terrorism-160000507.html>.

⁸² Jon Brodtkin, "First Amendment Doesn't Apply on YouTube; Judges Reject PragerU Lawsuit," last modified February 27, 2020, <https://www.bbc.com/news/technology-51658341>.

of information (not accessing external links to see the full article or source of the information shared), a thing abused by those who peddle fallacious information by allowing catchy headlines, captions, and pictures for their content to gather engagement, which the algorithm favorably detects to override the fact-checking of legitimate news sources that are not attention-grabbing for netizens.⁸³

O'Neill also pointed out that we need to recover trust from the proactive fights against deception and manipulation, rather than simply relying on an informed consent form. This can be illustrated in the political and legal climate of the United States. When Newsmax, One America News, Gateway Pundit, and many Trump-connected lawyers were sued by Dominion Voting Systems for falsely defaming their services by spreading rumors that they were deleting votes for Trump during the 2020 election, those peddlers of falsehoods then was protected by Section 230 of the 1996 Communications Decency Act (CDA).⁸⁴ What happened was that there were two subsections of the law that contradict one another: Section 230(c)(1) prevents the platforms from suffering with liabilities for illegal content displayed on their websites that was primarily user-generated or posted by other third parties, while Section(c)(2) gives freedom to platforms to regulate contents posted in their websites but without any requirement to do so even in harmful contents. The positioning of these two subsections was deemed appropriate as it was assumed that the companies would do whatever it takes to police their platform in order to protect their branding. But when the advertising models came in, engagement became the pinnacle parameter of profit. While it may be well argued that changing this law could stifle competition as it would require colossal financial resources for new social media businesses to thrive, it could be recommended that the legal system adopt an equitable way to require all businesses to undertake “reasonable steps” – proportionate to its influence, financial capacity, and reach – to control harmful contents rather than using the law to be an alibi.⁸⁵ This will demolish the commonly ballyhooed difficulty of separating the intentional toleration of disinformation and the unintentional spread of false information, which, more often than not, is the result of the lack of political will on the part of the companies to put a stop to the tides of information flooding our social media accounts. Furthermore, a great leap to this dilemma was when the Delaware State Court ruled that Fox News abused its freedom of speech by merely relying on the First Amendment at the expense of the reputational harm incurred on the part of the

⁸³ Kelly Grounds and Madelyn Koff. “Disinformation, Disruption, and the Shifting Media Ecosystem in the 2022 Philippines Election,” last modified May 5, 2022, <https://www.asiapacific.ca/publication/election-watch-philippines-dispatch-4-social-media-use>.

⁸⁴ Giansiracusa, *How Algorithms Create and Prevent Fake News*, 199.

⁸⁵ Michael D. Smith and Mashall W. Van Alstyne, “It’s Time to Update Section 230,” last modified August 12, 2021, <https://hbr.org/2021/08/its-time-to-update-section-230>.

Dominion, especially because media outlets, by nature, are expected to support truthful democracy by constantly striving for accuracy rather than unjustified “claims” and baseless “opinions.”⁸⁶ In the end, we need political and legal structures empowering us to fight against the abuses of companies that tolerate the spread of false information, rather than laws that can be circumvented to protect the self-interest and pernicious agenda of those who profit from making us immersed in attractive yet untruthful content. We need to safeguard the political stability of democracy grounded in truth rather than tolerating mob rule under the pretext of “free communication.”⁸⁷

Tom Beauchamp observed that we have made historical progress from merely treating informed consent forms as a guard against possible culpability towards the enabling and protecting of meaningful choices of patients and research participants. In this sense, we treat people as ends-in-themselves and not merely instrumentalized for the means to our manipulative, harmful, and/or deceptive ends. Rather than becoming subsidiary to prevailing rules and practices, Beauchamp observed an ethical awareness that wants autonomous choice and moral adequacy of research benchmarks to be the reference of societal evaluation. In the case of algorithms, we may expect merely a “broad consent” to its operation as its performance cannot completely be described in information sheets and informed consent forms, but this should never be accepted as a sign not to protect sensitive personal information, to violate our privacy and confidentiality, and to be harmed by such through oppressive power relations.⁸⁸ To analogize, we need not expect what is impossible by trying to attain a so-called higher standard of competency in the face of higher risks, say, in “experimental therapy,” simply because a “standard therapy” can easily give us cognitive exacting clarity about the consequences of the treatment.⁸⁹ We may not expect all the processing of artificial intelligence’s deep learning to be laid down in our hands at any moment, but we see that our choice is reasonable when we approve an informed consent form because it explicitly states that it will not use our personal data simply to manipulate us like puppets and be under the gaze of data-extracting advertising model for nothing but to sell our attention, dignity, and autonomy. Instead of making a claim that we have no right to be guaranteed ethical treatment from social media companies simply because we cannot have an a priori grasp of our “fate,” so to

⁸⁶ Laurence H. Tribe, et al., “Fox News’ Recent Setback Against Dominion is a Major Victory for the First Amendment,” last modified April 11, 2023, <https://www.justsecurity.org/85918/fox-news-recent-setback-against-dominion-is-a-major-victory-for-the-first-amendment/>.

⁸⁷ O’Neill, *A Philosopher Looks at Digital Communication*, 55.

⁸⁸ Tom L. Beauchamp, “Informed Consent: Its History, Meaning, and Present Challenges,” *Cambridge Quarterly of Healthcare Ethics* 20 (2011): 519-520. DOI: 10.1017/S0963180111000259.

⁸⁹ David Rudge, “Do Unknown Risks Preclude Informed Consent?,” *Essays in Philosophy* 4 (2003): 114-115. DOI: 10.5840/eip2003422.

speak, in the reality of algorithm, as research participants in wide-scale internet research, we need to assert our autonomy by addressing the concern how we become vulnerable to the power structures caused by hiding negligence and abuses in the spread of false information vis-à-vis the advertising business model and its ideological premises.⁹⁰ Thus, we need to define our vulnerability in a contextual way and in the face of the realities we encounter as we participate in research rather than making it a backdoor solution to escape difficulties, however necessary they are to be confronted and resolved.

CONCLUSION:

It is not merely for the sake of efficiency that we should embrace how algorithms affect our lives. It is not merely in the name of profit that we should bow down to big tech companies and treat informed consent forms as the defining measure of our protection from misinformation and disinformation. While informed consent forms cannot explicitly articulate all the algorithmic operations possible, as they now mainly use deep learning – escaping any rigid definition of its predictive performance – this does not amount to letting ourselves be emptied of our autonomy through relentless data extraction for whatever the power that be may decide to make use of it. In this essay, I have shown that a written agreement must not be treated as the final route of human transactions since we decide matters based on normative and inferential backgrounds that we commonly share. Thus, the predictive scheme used in social media administration must also be shaped by the ethical commitments that we share, defining regulations to ensure that truthfulness remains as a reference to how we understand the reality we live in. These ethical commitments must be stated in informed consent forms for social media users to understand the complexity behind the algorithm and how it is dependent on a larger business model of advertisement. A language that is meaningful and intelligible is necessary for social media users from diverse cultural backgrounds to understand the scope of their control in the business development and operation of social media. These are some crucial ways in which social media users, who consented to participate in internet research, are to be treated justly as “research participants,” or better still, as “research partners,” and not merely as disposable appendages to the machines of big tech companies.

More than highlighting the significance of informed consent forms in the restoration of trust in the wearisome and mistrustful world we now live in, this essay has shown that

⁹⁰ Wil C. van den Hoonard, “‘Vulnerability’ as a Concept in its Own Prison,” in *Handbook of Research Ethics and Scientific Integrity*, ed. Ron Iphofen (Switzerland: Springer Nature, 2020), 584-585. DOI: <https://doi.org/10.1007/978-3-030-16759-2>.

there are broader and deeper sources of the problems that we should uncover, which the limitations of the overreliance on informed consent forms cannot simply take away. We also have to admit that algorithms are important in the realistic operation of business for it to be efficient, despite the numerous transactions and users it has to face. Hence, these interconnected problems also demand expanding our capacity to tolerate the development of our critical thinking skills as we use social media as a digital space where we can freely access and assess information. We must refuse how social media companies, in some instances, normalize the promotion of falsehoods as if it is what “freedom of expression” is for. Therefore, an informed consent form is not the panacea to combat the lack of media literacy and the distorted understanding of human freedom. To be sure, we need legal structures to be reformed, but to do so would require the courage to fight for how we value trust, so precious to be left to what others have to say, without any discursive connection in this world we call the internet.

BIBLIOGRAPHY:

- 2022 National Ethical Guidelines for Research Involving Human Participants. Bicutan, Taguig: Philippine Council for Health Research and Development, 2022.
- Ali, Muhammad, et al., “Discrimination Through Optimization: How Facebook Ad Delivery Can Lead to Biased Outcomes.” *Proceedings of the Human-Computer Interaction* 3 (November 2019): 1-30. DOI: <https://doi.org/10.1145/3359301>.
- Andersen, Kristian G., et al. “The Proximal Origin of SARS-CoV-2.” *Nature-Medicine* 26 (2020): 450-452. DOI: <https://doi.org/10.1038/s41591-020-0820-9>.
- Andrew, Jane, et al. “Data Breaches in the Age of Surveillance Capitalism: Do Disclosures Have a New Role to Play?.” *Critical Perspectives on Accounting* 30 (2015):1-16. DOI: <https://doi.org/10.1016/j.cpa.2021.102396>.
- Bechmann, Anja, and Jiyoung Ydun Kim. “Big Data.” In *Handbook of Research Ethics and Scientific Integrity*, edited by Ron Iphofen, 427-444. Switzerland: Springer Nature, 2020.
- Bergen, Mark. “YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant.” Bloomberg, April 2, 2019. <https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant>.
- Beauchamp, Tom L. “Informed Consent: Its History, Meaning, and Present Challenges.” *Cambridge Quarterly of Healthcare Ethics* 20 (2011):
- Bolsover, Gillian, “China: An Alternative Model of a Widespread Practice.” In *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, edited by Samuel C. Woolley and Philip N. Howard, 212-238. New York: Oxford University Press, 2019.

- Bovet, Alexandre, and Hernan Makse. "Influence of Fake News in Twitter During the 2016 US Presidential Election." *Nature Communications* 10 (2019): 1-14. DOI: 10.1038/s41467-018-07761-2.
- Brodkin, Jon. "First Amendment Doesn't Apply on YouTube; Judges Reject PragerU Lawsuit." *BBC*, February 27, 2020. <https://www.bbc.com/news/technology-51658341>.
- Campaign for Accountability. "Report: Google Makes Millions from Fake News." Accessed October 31, 2024. <https://campaignforaccountability.org/work/how-google-makes-millions-off-of-fake-news/>.
- Cellan-Jones, Rory. "Coronavirus: Social Media Users More Likely to Believe Conspiracies." *BBC*, June 18, 2020. <https://www.bbc.com/news/technology-53083341>.
- De Castro, Leonardo D. "Exploitation in the Use of Human Subjects for Medical Experimentation: A Re-Examination of Basic Issues." *Bioethics* 9 (1995): 259-286.
- De Castro, Leonardo D., and Isidro Manuel C. Valero, "Ethics Education Needs More than the Four Principles: Bioethics Discourse in a Community of Inquiry." In *Global Education in Bioethics*, edited by Henk ten Have, 69-80. Switzerland: Springer International Publishing, 2018.
- Elemia, Camille. "Cat-And-Mouth Game: Twinmark Fake Network Still Thrives on Facebook." *Rappler*, December 23, 2020. <https://www.rappler.com/newsbreak/investigative/remnants-twinmark-media-enterprises-fake-network-still-thrives-facebook-after-takedown/>.
- Floridi, Luciano, et al. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7 (2017): 76-99. DOI: <https://doi.org/10.1093/idpl/ix005>.
- Fosher, Max, and Amanda Taub. "How YouTube Radicalized Brazil." *New York Times*, August 11, 2019. <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html>.
- Giannopoulou, Alexandra. "Algorithmic Systems: The Consent is in the Detail." *Internet Policy Review* 9 (March 2020): 1-19. DOI: 10.14763/2020.1.1452.
- Giansiracusa, Noah. *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More*. New York: Apress, 2021.
- Global Disinformation Index. "The Quarter Billion Dollar Question: How is Disinformation Gaming Ad Tech?." *Global Disinformation Index*, September 22, 2019. <https://www.disinformationindex.org/research/2019-9-1-the-quarter-billion-dollar-question-how-is-disinformation-gaming-ad-tech/>.
- Ground, Kelly, and Madelyn Koff. "Disinformation, Disruption, and the Shifting Media Ecosystem in the 2022 Philippines Election." *Asia Pacific Foundation of Canada*, May 5, 2022. <https://www.asiapacific.ca/publication/election-watch-philippines-dispatch-4-social-media-use>.

- Kang, Cecilia. "Facebook's Hands-Off Approach to Political Speech Gets Impeachment Test." *New York Times*, October 8, 2019. <https://www.nytimes.com/2019/10/08/technology/facebook-trump-biden-ad.html>.
- Kim, Tae Wan, and Bryan R. Routledge. "Why a Right to Explanation of Algorithmic Decision-Making Should Exist." *Business Ethics Quarterly* 32 (January 2022): 75-102. DOI: 10.1017/beq.2021.3.
- Lalu, Gabriel Pabico. "'Fake News' a Problem in PH? 9 out of 10 Filipinos agree, Says Pulse Asia." *Inquirer.net*, October 11, 2022. <https://newsinfo.inquirer.net/1678248/fake-news-a-problem-in-ph-9-in-10-filipinos-agree-says-pulse-asia>.
- Lanier, Jaron. *Ten Arguments for Deleting Your Social Media Accounts Right Now*. New York: Henry Holt and Company, 2018.
- Lapowski, Issie. "In a Letter to Pichai, Top Philanthropists Slam Google for Placing Charity Ads on Disinfo Sites." *Protocol*, August 13, 2020. <https://www.protocol.com/google-ads-charities-disinformation-sites>.
- Lehtiniemi, Tuukka, and Yki Korstesniemi. "Can the Obstacles to Privacy Self-Management be Overcome? Exploring the Consent Intermediary Approach." *Big Data & Society* 4 (2017): 1-11. DOI: <https://doi.org/10.1177/20539517177219>.
- Manson, Neil C., and Onora O'Neill. *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press, 2007.
- Mayer-Schonberger, Viktor, and Thomas Range. *Reinventing Capitalism in the Age of Big Data*. New York: Basic Books, 2018.
- Meta Business Help Center. "Mga Patakarán sa Pag-Monetize ng Content." Accessed October 3, 2024, <https://support.google.com/youtubelanswer/1311392?s-jid=2322536943219315613-AP#zippy=>.
- Noriega, Richa. "51% of Adult Filipinos Find Difficulty in Discerning Fake News." *GMA News*, February 17, 2022. <https://www.gmanetwork.com/news/topstories/nation/822135/sws-51-of-adult-filipinos-find-difficulty-in-discerning-fake-news/story/>.
- O'Neill, Onora. *A Philosopher Looks at Digital Communication*. Cambridge: Cambridge University Press, 2022.
- _____. *A Question of Trust: The BBC Reith Lectures 2022*. Cambridge: Cambridge University Press, 2002.
- _____. *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press, 2004.
- _____. "Informed Consent and Public Health." *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 359 (July 2004): 1133-1136. DOI: 10.1098/rstb.2004.1486.
- _____. "Some Limits of Informed Consent." *Journal of Medical Ethics* 29 (2003):

- 4-7. DOI: 10.1136/jme.29.1.4.
- Ong, Jonathan Corpus, and Jason Vincent A. Cabanes. *Politics and Profit in the Fake News Factory: Four Work Models of Political Trolling*. Latvia: NATO Communications Centre of Excellence, November 2019. Accessed November 15, 2024. https://stratcomcoe.org/publications/download/5december_report_philippines.pdf.
- Parungao, Adrian. "Concern Over Fake News Increases – Reuters Study." *Inquirer.net*, June 19, 2024. <https://globalnation.inquirer.net/240039/fwd-reuters-report-on-world-trust-concern-on-real-and-fake-in-news-and-news-avoidance>.
- Rivas Velarde, Minerva C., et al., "Consent as a Compositional Act – A Framework that Provides Clarity for the Retention and Use of Data." *Philosophy, Ethics, and Humanities in Medicine* 19 (2024): 8. DOI: 10.1186/s13010-024-00152-0.
- Rodríguez-Ferrandiz, Raul, "The Pandemic and Its Apostles: Conspiracy Theories in Pandemic Mode." In *Digital Totalitarianism: Algorithms and Society*, edited by Michael Filimowicz, 62-83. New York: Routledge, 2022.
- Rosen, Guy. "An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19." Accessed October 3, 2024. <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>.
- Rudge, David. "Do Unknown Risks Preclude Informed Consent?" *Essays in Philosophy* 4 (2003): 110-118. DOI: 10.5840/eip2003422.
- Sadagopan, Swathi Meenakshi. "Feedback Loops and Echo Chambers: How Algorithms Amplify Viewpoints." *The Conversation*, February 4, 2019. <https://theconversation.com/feedback-loops-and-echo-chambers-how-algorithms-amplify-viewpoints-107935>.
- Sidhar, Sriya. "Collective Consent for Algorithmic Accountability." *Internet Just Society*, December 2, 2020. <https://www.internetjustsociety.org/collective-consent-for-algorithmic-accountabilitybaaa8d68#:~:text=The%20need%20to%20rethink%20the%20individualistic%20framework%20of%20'informed'%20consent.&text=Over%20the%20course%20of%20our,by%20a%20variety%20of%20entities>.
- Smith, Michael D., and Marshall W. Van Alstyne. "It's Time to Update Section 230." *Harvard Business Review*, August 12, 2021. <https://hbr.org/2021/08/its-time-to-update-section-230>.
- Steinberg, Etye. "AI, Radical Ignorance, and the Institutional Approach to Consent." *Philosophy and Technology* 37 (2024): 1-26. DOI: <https://doi.org/10.1007/s13347-024-00787-z>.
- Sweeney, Latanya. "Discrimination in Online Ad Delivery." *Communications of the Association of Computing Machinery* 56 (2013): 1-36. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2208240.

- Solsman, Joan. "YouTube's AI is the Puppet Master Over Most of What You Watch." CNET, January 10, 2018. <https://www.cnet.com/tech/services-and-software/youtube-cs-2018-neal-mohan/>.
- Tan, Corinne. "The Curious Case of Regulating False News on Google." *Computer Law and Security Review* 46 (2022): 1-14. DOI: <https://doi.org/10.1016/j.clsr.2022.105738>.
- Tennison, Jeni. "Community Consent." Jeni Tennison, January 17, 2020. <https://www.jenitennison.com/2020/01/17/community-consent.html>.
- Toquero, Loreben. "Global fact-Checkers Demand from YouTube Effective Action Against Disinformation." *Rappler*, January 13, 2022. <https://www.rappler.com/technology/global-fact-checkers-demand-youtube-effective-action-disinformation/>.
- Tribe, Lawrence H., et al. "Fox News' Recent Setback Against Dominion is a Major Victory for the First Amendment." *Just Security*, April 11, 2023. <https://www.justsecurity.org/85918/fox-news-recent-setback-against-dominion-is-a-major-victory-for-the-first-amendment/>.
- Van den Hoonaard, Wil C. "'Vulnerability' as a Concept in its Own prison." In *Handbook of Research Ethics and Scientific Integrity*, edited by Ron Iphofen, 577-588. Switzerland: Springer Nature, 2020.
- Vincent, James. "YouTube Brings Back More Human Moderators After AI Systems Over-Censor." *The Verge*, September 21, 2020. <https://www.theverge.com/2020/9/21/21448916/youtube-automated-moderation-ai-machine-learning-increased-errors-takedowns>.
- Winter, Jana. "Exclusive: FBI Document Warns Conspiracy Theories are a New Domestic Terrorism Threat." Yahoo News, August 2, 2019. https://www.yahoo.com/news/fbi-documents-conspiracy-theories-terrorism-160000507.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFcNajm-wdCbfsVknva0sp8ADrZK3Lt61bckeQh8bSPAMe05u4CDe85Sl9ue07GOB-Wh10U-d1i-ULAqjeAmFLvU_kcXNdeujKtZmXIFHjFaI7vub9bQmZ3ZgerLasdqiVPR3lx-UP70-HTfTf2--TrpXXjXrkZ4WLPqBEiHwRIBk7E.
- Wong, Julia. "An Update on Our Work to Keep People Informed and Limit Misinformation." *The Guardian*, February 1, 2019. <https://www.theguardian.com/media/2019/feb/01/facebook-youtube-anti-vaccination-misinformation-social-media>.
- X. "Mga Tuntunin ng Serbisyo." Accessed December 2, 2024, <https://x.com/tl/tos>.
- YouTube Official Blog. "Protecting our Extended Workforce and the Community." Accessed October 3, 2024, <https://blog.youtube/news-and-events/protecting-our-extended-workforce-and/>.
- YouTube Tulong. "Mga Pagwawakas ng Channel o Account." Accessed October 2, 2024, <https://support.google.com/youtube/answer/2802168?s-jid=7851885953287863292-AP>.

- YouTube Tulong. "Mga Patakarán sa Pag-Monetize ng Channel sa YouTube." Accessed October 2, 2024, <https://support.google.com/youtube/answer/1311392?sjid=2322536943219315613-AP#zippy=>.
- Zuboff, Shoshana. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30 (2015): 75-89. DOI: <https://doi.org/10.1057/jit.2015.5>.
- Zuckerberg, Mark. "Facebook Status Update." Accessed last November 3, 2024, https://www.facebook.com/zuck/posts/10103269806149061?__mref=message_bubble.