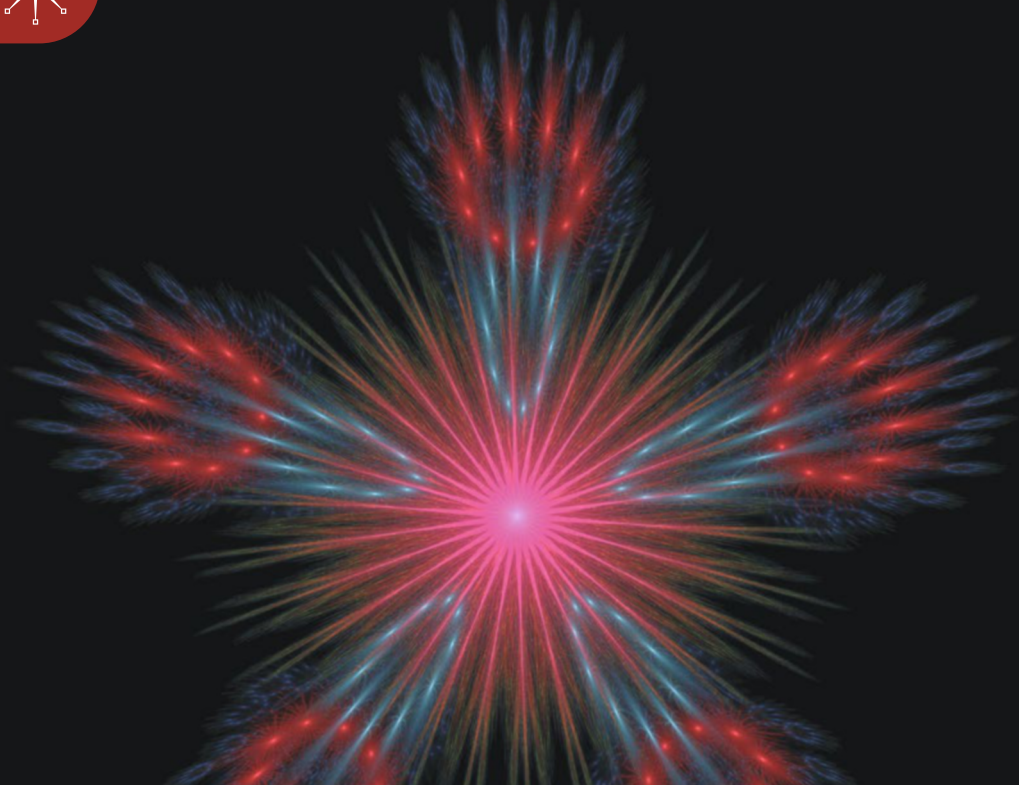




RADICAL THEOLOGIES AND PHILOSOPHIES




Dialectic of Digital Enlightenment

Reclaiming Radical Philosophy for Our Times

Edited by Bhabani Shankar Nayak

palgrave
macmillan

Editor

Bhabani Shankar Nayak 
Guildhall School of Business and Law
London Metropolitan University
London, UK

ISSN 2634-663X ISSN 2634-6648 (electronic)
Radical Theologies and Philosophies
ISBN 978-3-031-95468-9 ISBN 978-3-031-95469-6 (eBook)
<https://doi.org/10.1007/978-3-031-95469-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: © Tim Bird via Getty Images

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.



Participatory Democracy in the Age of AI: A Habermasian Approach

Asaf Bar-Tura

INTRODUCTION

Critical theorists of recent decades have not only delineated the political nature of technology, but also its malleability, and the possibility for a more democratic technological society. More recently, developments in Artificial Intelligence (AI) technologies not only shape how we communicate, but also how we think, what we expect, and how we relate to others. Therefore, critical theory of technology must examine the possibilities for society to shape technology while dialectically addressing the ways in which technology is shaping society.

My thesis here includes three main arguments that build one upon the other: *First*, a contemporary critical theory of technology needs a comprehensive political-theoretical framework through which to critique technology and its role in a democratic society. *Second*, Jurgen Habermas's theory of law and democracy offers a necessary framework through which to ground a critical theory of technology. *Third*, equipped with the

A. Bar-Tura (✉)
New York, NY, USA

© The Author(s), under exclusive license to Springer Nature
Switzerland AG 2025

B. S. Nayak (ed.), *Dialectic of Digital Enlightenment*, Radical
Theologies and Philosophies,
https://doi.org/10.1007/978-3-031-95469-6_13

Habermasian discourse-based democratic framework, I offer some considerations for public reason in the age of AI.

PART ONE: THE NORMATIVE AND PRACTICAL DEFICITS IN CRITICAL THEORY OF TECHNOLOGY

Habermas and Technology

My first argument is that current discussions in critical theory of technology lack a comprehensive political-theoretical framework through which to critique technology such as AI and its role in a democratic society.

In the philosophical discussions about technology over the last century, we can broadly distinguish between *essentialist* and *constructivist* approaches. This distinction classifies theories according to answers they provide to certain questions about technology. Such questions include: Is the development of technology and the trajectory of this development under human control? Is this trajectory predetermined by the very nature of technology? Does technology have a nature or essence? If so, what is it? Do technologies inherently dictate values, or are they valueless means to value-laden ends? *Essentialist* approaches attribute to technology an essence that cannot be altered (for better or worse),¹ while *constructivist* approaches emphasize the social aspects of technology and the ways in which it can be reformed toward socially determined ends (Kaplan, 2009; Feenberg, 1999).² Indeed, one of the main efforts of the constructivist approach as a critical theory of technology is to restore the possibility of agency within the technological realm, a possibility that has been placed in serious doubt by many twentieth-century thinkers.

Habermas's foray into questions concerning technology was in his essay "Technology and Science as 'Ideology'," in which he responds to Herbert

¹ Habermas's early work can be construed as endorsing an essentialist approach to technology, and in particular the neutrality approach. According to this approach, a tool is taken to be neutral and can be used for good or bad purposes. Technology is considered to embody a universal rationality that is independent of social forces. In other words, there is no such thing as morally good or bad technology, only good or bad users. The neutrality approach is sometimes referred to as instrumentalism since it views technologies as mere instruments for human activities; as value-neutral means to value-laden human ends.

² Constructivist theories cannot avoid attributing some essence to technology (otherwise it seems unclear what it would mean to talk about technologies at all). However, constructivists emphasize the social processes that operate beyond this (relatively thinly conceived) essence.

Marcuse's call for a *new* science which will produce *new forms* of technology. Put succinctly, Habermas argues that liberation cannot be achieved by transforming technology because technology cannot be altered. For Habermas, technology is essentially the unburdening of needs that are rooted in human nature through purposive-rational action and substituting other means for human labor. On a fundamental level, he understood technology as related to the interests of humans in general, and not to the interests of specific groups or classes. Therefore, to the extent that human nature itself is not fundamentally altered, technology cannot be altered as well (Habermas, 1992, pp. 120–122; Feenberg, 1999, p. 7). Furthermore, Habermas is critical of Marcuse's assertion that the instincts can serve as a ground for critical theory, claiming that such a proposition relied too heavily on speculations about human nature that could not be verified. As an alternative, Habermas argued that a critical foundation could be found in the very structure of everyday language, which became a fundamental tenet of Habermas's thought (Ingram, 1990, p. 107).³

According to Habermas, advanced-capitalist societies are divided between a lifeworld, which is governed by norms of communicative interaction, and a system governed by "steering imperatives" of money and power. This distinction is meant to capture the communicative practices of everyday life on the one hand, while on the other hand recognizing the systemic forces that operate in society and which, if not controlled, come to colonize, or dominate, the lifeworld (Habermas, 1975, pp. 1–8; Kellner, 2000, p. 272).⁴ In this Habermasian framework, in modern societies technology properly relates to the level of systems (work and administration) and not the lifeworld. Habermas accepts the autonomy of technical (instrumental) rationality in a limited role of facilitating systems of labor and technical administration, while emphasizing the role of communicative reason in the lifeworld. Therefore, Habermas can be taken to maintain that technology is neutral in its proper sphere, while outside that sphere it causes various social pathologies in modern societies (Feenberg, 1999,

³Habermas has suggested the turn to language as early as 1968 in "Knowledge and Human Interests."

⁴Habermas views Marx's most important contribution to social theory to be the "account of social evolution as a separation (abstraction or uncoupling) of a self-regulating legal and economic system from a meaningful lifeworld" (Ingram, 2010, p. 310). One can easily see this contribution in Habermas's distinction between system and lifeworld. Moreover, Habermas considers this process of differentiation as an achievement of modern societies, as compared with traditional ones.

p. 152). Habermas critiques Adorno insofar as he thinks Adorno has only attended to instrumental rationality when considering the dialectic of enlightenment, and has not considered the emancipatory potential in communicative rationality (Krakauer, 1998, p. 12).

Critical Theory of Technology and Its Deficits

Contrary to earlier approaches to technology (including Habermas's), which tended to emphasize a necessary technological rationality, in recent decades philosophers have begun to construct more empirical and historical views of technology, and to understand it in its actual uses in social contexts (Ihde, 2001; Achterhuis, 2001). This (broadly) *constructivist* approach argues that society simultaneously shapes technology as technology shapes society (Kaplan, 2009).

Andrew Feenberg frames his constructivist critical theory of technology through the debate between Habermas and Marcuse regarding technology. One of Feenberg's most significant attractions to Habermas is the latter's attempt to rehabilitate the prospect of a rational, democratic, emancipated society—a prospect placed in serious doubt by Adorno and Horkheimer—by pointing to communicative rationality and the process of reaching intersubjective understanding. Feenberg, then, is inspired by Habermas's insistence on carrying forward the project of an emancipated democratic society that nonetheless does not eschew a rational ideal.

However, Feenberg's thinking quickly diverges from Habermas's. While Habermas segregates technology (and instrumental rationality) to the realm of the system, and argues for an emancipatory (communicative) rationality in the lifeworld, Feenberg aims to bring the emancipatory dimension of rationality into technology itself. His main critique of Habermas has to do with Habermas's notion of "differentiation," according to which well-functioning (non-pathological) modern societies can maintain a healthy differentiation between the system (which would include technology and its inherent instrumental rationality) and the lifeworld (characterized by uninhibited communication). Feenberg asserts that Habermas underestimates the extent to which the problems technology is meant to solve, and the technological solutions offered, are shaped by social interests (which would be communicatively considered in the lifeworld), and are not simply a result of neutral instrumental action (Feenberg, 2011, p. 869). If Habermas rejected the possibility of an alternative technology, one not guided by the principles of instrumental action,

Feenberg embraces this possibility but presses for a more empirically based approach that would provide guidelines for reform.⁵ Feenberg emphasizes that the issue of how particular design choices are made over other choices is an inherently political question.⁶ What Feenberg can teach us about AI is that any technological development does not follow only technological imperatives. Social choices intervene in the selection of the problem definition as well as its solution (Feenberg, 1999, p. 79).⁷

A critical theory of technology, then, understands technology as a site of social struggle, and adheres to the following principles: (1) Technical design is not determined by a general criterion such as efficiency, but by a social process; (2) this social process is not about fulfilling “natural” human needs, but concerns the cultural definition of needs; (3) competing definitions reflect competing visions of modern society realized in different technical choices. Feenberg is optimistic that this politics of technology can be carried out at the “micro” level; that actors within technically

⁵ Feenberg’s theory has evolved from an explicitly socialist theory in his early book *Critical Theory of Technology*, to a more reform-oriented approach that is not necessarily tied to a socialist (or even anti-capitalist) vision.

⁶ Feenberg qualifies this Marcusean view about the socially determined nature of technology by distinguishing between primary and secondary levels of instrumentalization of objects (the primary level is essential to technology while the secondary level is socially shaped).

⁷ An illustration of this point can be found in the use of Geographic Information Systems (GIS) in environmental surveying and planning. GIS can be defined here as a means of integrating spatial and non-spatial information into a single computer system for analysis and graphic display (think of Google Maps as a popular example). It has been argued that the use of GIS for policy-making is less likely to favor special interests in cases where what is surveyed is a physical environment, since there is little room for value judgment. However, as one study asserts, even here concerns of justice emerge: “During the former apartheid era in South Africa’s Soil and Irrigation Research Institute a maximum 12 percent slope angle for plow land was set. This was based on the requirements of mechanized cultivation and GIS land suitability analyses was carried out accordingly. This slope angle reflected the Institute’s viewpoint and constituency as hand hoeing and animal plowing, as practiced by the majority of black farmers, allows cultivation on much steeper slopes” (Cinderby, 1999, p. 306). Though the decision to set the maximum slope angle suitable for plowing at twelve percent seems innocent, in fact the chosen slope angle reflected the practices of (mostly) white farmers using mechanized farming techniques, while the practices of traditional (mostly black) farmers go unrecognized by the GIS technology. This example shows that the design of the technical system to survey slope angles is in itself value-laden. It expresses a favoring of “efficient” mechanized cultivation limited to certain slope angles, and thus chooses slope angles as the determinant data to be found. A different cultural approach, identifying the problems differently, would have resulted in an altogether different technology; or, at the very least, it would have made this particular technology irrelevant to the case at hand.

mediated systems are able to identify a “margin of manoeuvre” within such systems, and alter them (Feenberg, 1999, pp. 83–105; Feenberg, 1991, pp. 66–89).

While Feenberg points to the possibility of “democratizing technology” through the appropriation and redefinition of technologies by users on the micro level, I now turn to two deficits of this analysis—normative and practical—that Habermas can help us overcome. To understand the *normative deficit*, we turn to Feenberg’s analysis of actual technologies that were altered by their users, such as the Minitel in France (Feenberg, 1999, p. 126). Critics have responded to these analyses that while certain user-driven initiatives may be desirable, “such changes of technology in response to consumers’ initiatives or preferences follow the logic of market rationalization, not democratization” (Doppelt, 2006, p. 89; Doppelt, 2001). This critique by Doppelt demarcates an important distinction between consumer power and democratic agency. In other words, Feenberg’s empirical examples do not provide a normative standard for what we mean by designing technologies *for democracy* (Doppelt, 2006, pp. 87–92). Without an underlying theory of democracy, we have no normative resources from which to draw when arguing for the empowerment of hitherto excluded interests. We would have no standard with which to judge whether particular designs are “democratic” or ought to be “democratized.” This perhaps could be Habermas’s most important contribution to thinking about AI.

The second challenge for Feenberg’s analysis is a pragmatic deficit, namely, the concern that some technologies may be too deeply entrenched in social systems to be reformed. A useful framework for thinking this through is Thomas Hughes’s analysis of four stages in the development of a system: invention and development; technology transfer; system growth; and substantial momentum (Stump, 2006, p. 11; Hughes, 1983). According to this analysis, in the last stage (substantial momentum) the technology in question shapes society in such a deep way that it is unclear what kind of resistance to that technology is possible beyond the point where substantial momentum is attained.⁸ In other words, some

⁸Using Hughes’s framework, Tyler Veak compares the development of electrical grids to the Internet network: “[Hughes] compares the development of electrical systems in Chicago, London, and Berlin and shows how each [particular] context transfigured the shape of the electrical system. [...] Nevertheless, Hughes claims that by the 1930s, all three systems were homogenized by the market demands of utilitarian efficiency. [...] As in the case of the Internet, electricity was hailed as a liberatory technology—emancipating the common person

technological systems cannot be easily transformed, certainly not by individual users. Reasons for this may be technical, or economic, but they may also be cultural. That is, some technologies may be so deeply integrated into our lifeworld that life without them can hardly be imagined, and alternative designs that do not change their social function would not change their primary social effect.⁹

In order to address these concerns, we need a theory of democracy upon which we can base a critical theory of technology. Furthermore, Habermas's discourse theory of democracy is well suited to provide this foundation. It allows us to articulate the ground for a critique of communicative practices, and to point to a set of standards to which we hold others accountable. We stand to gain an ethical and political (democratic) logic for the framework within which negotiations of technical designs occur.

PART TWO: HABERMAS'S DISCOURSE THEORY AS A FRAMEWORK FOR A CRITICAL THEORY OF TECHNOLOGY

The Foundations of Habermas's Discourse Ethics

To understand the normative framework for democracy that we find in Habermas, we start with his theory of modernity. Max Weber analyzed modernity as entailing a process of rationalization. In this process, the world becomes "disenchanted" of its perceived inherent values. Influenced by Weber's analysis of rationalization, Adorno and Horkheimer argued for an inherent dialectic in the process of enlightenment. Since the result of this process was that nothing has inherent value, all becomes a potential means to particular ends. The role of reason becomes not determining what is good, right, and of value; rather, reason takes on merely an instrumental role in determining the best way to achieve the desired ends. Thus, so the argument goes, the enlightenment process has produced

from the drudgery of everyday life. But in the end, we find ourselves more deeply embedded in a system over which we have no control and no way out—that is, short of dropping out completely. Like London, we are all forced to capitulate to the standard (e.g., Microsoft) of the present (Internet) system" (Veak, 2000, p. 232).

⁹Albert Borgmann expresses similar concerns (Tijmes, 2001, p. 19).

instrumental reason as the only form of reason available, which, in a capitalist society, facilitates domination (Horkheimer, 1992, pp. 38–42).¹⁰

Habermas criticizes his mentors' analysis of the dialectic of enlightenment, arguing that their analysis neglects another form of reason, namely, communicative reason. It is the underlying (rational) structure of communicative action that holds the emancipatory potential lamented by the first generation of critical theorists.¹¹ For Habermas, the primary role of the critical theorist is to expose the impediments to free communication through which participants' preferences can be worked out intersubjectively.

There is also a clear Kantian foundation to Habermas's discourse-ethical theory. Habermas, like Rawls and others, takes on the task of rehabilitating a moral theory based on reasons by "analyzing the conditions for making impartial judgments of practical questions, judgments based solely on reasons" (Habermas, 1990, p. 43). The task is to show how reason can lead to a judgment with an "ought" character that has a justified claim to universal validity. For Habermas, to say I *ought* to do X is to say I have *good reasons* for doing X. Instead of talking about normative claims as claims to *truth*, Habermas suggests it is better to talk about them as claims to *validity* (Habermas, 1990, pp. 55–56). The result is that the moral point of view becomes a universally justified intersubjective *procedure* (this procedural aspect of Habermas's moral theory will show up in his theory of law and democracy as well).

For Habermas, claims to validity are inherently claims we are willing to defend against criticism. This is because a claim to validity cannot be arbitrary, and hence, we have reasons for this claim that we are willing to put forth (Habermas, 1990, p. 56; Habermas, 1998, p. 35).¹² In a way

¹⁰For Habermas's presentation of Horkheimer and Adorno's critique of the dialectic of enlightenment, see: Habermas (1984, Chapter Four, Section Two).

¹¹As Ingram points out, Habermas traced the error of the "dialectic of enlightenment" insofar as it ignores the intersubjective nature of communicative action back to its origins in Kant's transcendental philosophy of "subject-centered" reason. According to Habermas, Kant is not the sole perpetrator of this neglect. Habermas goes back to Descartes in this assertion, and continues all the way up to Sartre and Heidegger, and praises Kierkegaard, for example, for insisting on intersubjective philosophy (Ingram, 2010, pp. 4–21).

¹²Habermas asserts that actors may make three different claims to validity in their speech acts when oriented to reaching agreement. First, claims to truth are made when referring to something in the objective world. For example, empirical propositions ("It is raining outside") lay a claim to truth and may admit of falsity. We justify, or redeem, these sorts of claims by means such as empirical observation. A second kind of claim is claims to truthfulness (or

reminiscent of Kant’s categorical imperative, Habermas introduces the Discourse Principle as a principle with which we test the validity of normative claims:

(D) Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse*. (Habermas, 1990, p. 66)¹³

Discourse Ethics and the Democratization of Technology

Habermas argues that reason resides inherently in political communications. Thus, the reflective character of reason—that is, that upon reflection we recognize the rationality inherent in common deliberation—can stand as the source of legitimation for deliberative politics. No doubt invoking Kant’s notion of the “public use of reason,” Habermas explains that the public use of uninhibited communication has two dimensions, cognitive and motivational. The cognitive dimension includes the free processing of information and reasons (and is presupposed in communicative interaction). The motivational dimension, which bolsters both social integration and legitimacy, involves the actors’ inclinations to accept reasons given on free and rational grounds. What is more, intersubjectively shared convictions that result from deliberation form the very medium of social integration (Habermas, 1998, p. 35).

Habermas argues that in order to engage in argumentation at all, speakers must strive to and counterfactually presuppose an “ideal speech situation (ISS)” (Habermas, 1990, pp. 134–135). This ideal speech situation entails that all participants understand the argumentation process to be a cooperative search for the truth and are motivated to agree or disagree solely on the basis of “the unforced force of the better argument.” As we

sincerity), which refer to something in one’s own subjective world. These may include, for example, statements about one’s beliefs or emotions (“I believe it is raining outside”; “I am disappointed that it is raining outside”). The third kind of claims to validity is claims to rightness, which are made when referring to something in the shared social world. This would include normative claims (“Everyone ought to have equal rights”). For Habermas, the motivation of discussants to accept all three of these claims lies in the ability of the speaker to redeem them discursively by offering reasons, or, in the case of claims to truthfulness, by demonstrating behavior consistent with one’s claims (Habermas, 1990, pp. 58–59).

¹³As we will see, the Discourse Principle can help guide policy-making related to AI regulations.

saw in the Discourse Principle, Habermas asserts that ideally this speech situation would include, or at least take into account, all those potentially affected by the issue at hand. Thus, the discursive forum constituting such a debate would be the “communication community of those affected” (Habermas, 1998, p. 228). This ethic of discourse means that decisions about AI and its future cannot be limited to scientists, technologists, and experts.¹⁴

The rationality that grounds communicative action—the capacity to understand the speech of the other, to adhere to the “force of the better argument,” and finally to reach consensus—provides a solid foundation for developing discursive norms for public debate, and for the critique of various forms of societal domination, oppression, and manipulation that distort free processes of communication.¹⁵ The ISS model is a tool for critique and an aspirational model for democratic social communication, and it involves a set of counterfactual “pragmatic presuppositions” of rational consensus that serve as a regulative ideal (Kellner, 2000, pp. 270–271; Rehg, 2009, p. 27).¹⁶

¹⁴For a strong argument against leaving the future of AI to technologists, see Marcus (2024). Furthermore, Rehg, Latour, and others have highlighted that science and technology are often portrayed as fields in which the truth is not affected by social dynamics and politics. However, in Rehg’s words, “to transform a controversial claim into fact requires social resources beyond the laboratory. Specifically, the researchers must get other people, groups, and institutions interested in the claim” (Rehg, 2009, pp. 75–79). What these thinkers point to are the social resources underlying expert roles and argumentation processes in the public sphere.

¹⁵Regarding the orientation to consensus, Habermas stresses the difference between (1) the conditions necessary for the discursive generation of a rationally motivated consensus, and (2) the conditions necessary for negotiating a fair compromise. Discourse ethics recognizes the need for creating (1), and not settling for (2).

¹⁶It is precisely because “the demanding communicative presuppositions of rational discourses can only be approximately fulfilled” that Habermas emphasizes the need for democratic *procedures* (Habermas, 1998, pp. 230–234). The focus on the structure of procedures also allows to differentiate between varying types of discourses, a distinction that would bear on the question of the imperative of inclusion in the communication community. For example, following Dworkin, Habermas distinguishes between discourses of justification and discourses of application. Discourses of justification discuss whether a norm is just in general, while discourses of application discuss whether this norm ought to be applied in a specific case. For Habermas, discourses of justification and discourses of application would weigh differently the principle of including all those affected (*ibid.*, pp. 217–229). He explains that in the case of ethical-political questions (as opposed to moral ones) “those affected” can include only those who have a shared ethical-political tradition (*ibid.*, p. 108). However, when discussing moral (not ethical-political) questions, there is room to admit the participation of nonmembers as well (*ibid.*, p. 183).

Habermas's conceptualization of the ISS can provide normative guidance for the democratization of technology in two ways. First, Habermas's framework conceives of actors engaging each other with an aim of reaching consensus, which will serve as a foundation for some form of collaborative action. This ideal, which may possibly be approximated in actual speech situations, can serve as a guide for the institutionalization of discourse or the critique of systematically (and technologically) distorted communication. Second, the concept of ISS can serve as a critical tool in examining consensus, or, in the context of the democratization of technology, examining a technical code, that was actually established. In other words, we can ask whether the established technical code is genuine, consensual and democratic, or perhaps based on domination. Thus, even if the ISS is never actualized, it still serves as a critical standard against which every actually realized consensus can be tested (McCarthy, 1975, pp. xvii–xviii).

The Dialectic of Technology and Social Discourse

When considering the development of a discourse ethical framework that could account for the processes of negotiating technical codes, a dialectical relationship between technology and social discourse emerges. Namely, such an analysis must take into account that technology is not only the *object* of discourse; it is also a *means* by which this discourse takes place, and shapes the very nature of this discourse. To a large extent, technology constitutes the ways in which we engage in communicative action. As Habermas explains, the lifeworld “not only forms the *context* for the process of reaching understanding but also furnishes *resources* for it. The shared lifeworld offers a storehouse of unquestioned cultural givens from which those participating in communication draw agreed-upon patterns of interpretation for use in their interpretive efforts” (Habermas, 1990, p. 135). That is, all deliberations about technological designs draw upon a lifeworld increasingly infused with communication technologies. As Simon Cooper highlights, what we mean by democratic participation must be clearly articulated, since the very meaning of political participation is being transformed by technology itself. As we shift with technology to looser, more abstract modes of being-in-the-world, and being-with-others, “the

settings that have always grounded social life and any sense of a *cooperative ethic* are destabilized” (Cooper, 2006, p. 35; Bar-Tura, 2011).

The realization of Habermas’s Discourse Principle requires an open and vibrant public sphere. Consequently, if we are to examine the dialectic of discourse and technology, we must investigate the impact of technology on the public sphere and public discourse carried out within it. The next step, then, is to see how Habermas’s discourse framework as an ethical theory informs a political theory of democracy that normatively assesses the circulation of power and flow of communications in the public sphere.

Habermas’s Procedural Paradigm of Politics and the Public Sphere

Habermas argues that the same structure we find in language (played out in communicative action), we also find in broader social relations. In both cases, this structure entails a tension between facticity (real discourses) and validity (ideal discourses). We begin to see the tension between facticity and validity when we see that, on the level of language, these discursive norms appeal to a counterfactually ideal discourse. That is to say, the redeeming of these claims to validity always involves a counterfactual idealization of the circumstances of the communicative exchange, since the implicit assumptions are often at odds with the real context in which the exchange takes place.¹⁷

In the transfer of the tension between ideal norms of discourse and the empirical unfolding of real discourses from the ethical realm to the political one, the need to coordinate action involves the need for social integration. The question then arises: where do we look for a legitimate medium that will serve as a sort of bridge between facts and norms, thus maintaining a valid social order? Habermas argues that in modern societies, the medium that operates in the space of tension between social facts and social norms—thus providing a coordinating medium and hence a means of social integration—is law (Habermas, 1998, p. 27; Rasmussen, 1994, p. 24).¹⁸

¹⁷What is important to see here is that the counterfactual idealizations on which communicative action rests are unavoidable. One cannot speak without implicitly making these claims. For example, even the liar makes an implicit claim to sincerity; otherwise, she wouldn’t be lying but merely mistaken (Habermas, 1998, pp. 3–5).

¹⁸We can see here how Habermas’s theory of modernity, which informed his ethical framework, returns to inform his political framework as well. On this he comments that “the

Habermas explains that the legal form serves as a mechanism of mutual understanding regarding the norms that guide strategic interactions. An important dimension of this form of mutual understanding is that the actors themselves reach this understanding. Custom, tradition, or coercive authority does not impose it. Thus, though binding and hence stabilizing, the legal form can make a claim to rationality, and hence legitimacy (Rasmussen, 1994, p. 25). Law is ultimately legitimated by appeal to its origin in democratic will-formation based in communicative action. At the same time, law creates a framework in which actors can legitimately act strategically and are “unburdened” of the need to interact communicatively (Hedrick, 2010, p. 109).¹⁹

In order to arrive at a legal code that is legitimate, Habermas argues that we must posit certain sets of rights that would guarantee the legitimacy of the legal process. These include freedoms that could be categorized as those guaranteeing one’s private autonomy (as an individual) and those guaranteeing one’s public autonomy (as a member of the community).²⁰ Importantly, while traditional liberal political theory has

constitution of the legal form became necessary to offset deficits arising with the collapse of traditional ethical life” (Habermas, 1998, p. 113). Law as a newly emerging coordinating medium provides a dual function hitherto sustained by tradition: it provides directives for action (a set of authoritatively binding norms), while at the same time providing individuals with the cultural knowledge needed to anticipate the way others will behave as well as what behavior is expected by others and by the state.

¹⁹ Habermas explains that actors may not only *act* strategically, but may also take a strategic attitude toward law itself: “Legitimate law is compatible only with a mode of legal coercion that does not destroy the rational motives for obeying the law: it must remain possible for everyone to obey legal norms on the basis of insight. In spite of its coercive character, therefore, law must not *compel* its addressees to adopt such motives but must offer them the option, in each case, of foregoing the exercise of their communicative freedom and not taking a position on the legitimacy claim of law, that is, the option of giving up the performative attitude to law in a particular case in favor of the objectivating attitude of an actor who freely decides on the basis of utility calculations” (Habermas, 1998, p. 121).

²⁰ Habermas initially identifies three categories of rights that refer to private autonomy: the first and most basic is the classic liberal notion of “the right to the greatest possible measure of equal individual liberties.” He then asserts that the following two categories of rights are necessary in order to ensure the first. Hence, the second set of rights includes basic rights of membership (“being a member in a voluntary association of consociates under law”). The third set of rights would elaborate the protection afforded to individuals under the law in the form of actionable rights. These categories of rights, which pertain to an individual’s private autonomy, are not sufficient since legitimate law is premised on the participation of individuals in generating it. Hence, the basic system of rights must include categories that guarantee public autonomy. These include two additional categories. The first category pertains to

prioritized private autonomy over public autonomy, and has seen political rights as derived from, and aimed at guaranteeing, individual rights, Habermas takes a different approach, according to which public and private autonomy are co-original, and are given equal weight (Habermas, 1998, pp. 127–128). This is because all rights, when instituted as positive law, are legitimated by their origin in democratic legislative procedures; and these procedures, in turn, require public autonomy (Hedrick, 2010, p. 113; Rasmussen, 1994, pp. 29–30). Habermas’s thinking here reminds us that to think about rights in the context of AI must go beyond rights to privacy, and include rights to publicity—the rights to effective participation in the public sphere.²¹

The public sphere forms opinions, but defers decision making to the institutionalized political process. The quality of discursive public opinion formation depends on the ability of the public sphere to serve as a place to clarify issues, pose questions, and assert arguments, a function we find lacking in oppressive regimes (Habermas, 1998, pp. 360–362). Thus, for Habermas what constitutes a public sphere as public is the nature of the *discourse* it enables, understood through the procedural mechanisms by which this discourse is enacted. Furthermore, he recognizes that a fundamental worry from a democratic perspective is the possibility of administrative and social power operating without accountability to the public in the periphery. Hence the important role of communicative processes in informal publics in the periphery. According to Habermas, the public sphere should operate as a “warning system” that signals to formalized discursive forums (such as parliaments) about social problems. It is, if you will, the canary in the coal mine of democracy. As such, any analysis of communicative processes regarding technological designs must account for how new technologies are changing the flow of communications in the political public sphere, and what problems arise as a result of these changes.

rights of participation in the political process; the rights “to equal opportunities to participate in processes of opinion- and will-formation.” The final category pertains to rights that guarantee the material conditions necessary for a more or less equal opportunity to exercise the other rights. (Habermas, 1998, pp. 122–127).

²¹ Habermas’s procedural approach posits those basic categories of rights that would enable *real* participants to carry out a discourse with respect to the concrete rights that they self-legislate. Therefore, the determination of particular rights, the concrete implementation of a system of rights, is the task of concrete political communities.

PART THREE: PUBLIC REASON IN THE AGE OF AI

Rethinking Media Power

To think about AI through Habermas's ethical and political theory is to examine the dialectic of discourse and technology, namely, how we should design AI for democracy, while at the same time considering AI's role in facilitating public discourse.

Key to Habermas's procedural paradigm of law and democracy is the concept of *media power*, the mechanism through which topics of social concern can be raised in civil society and reach a large audience for public debate. In the age of AI, we see that media power is becoming ever more concentrated in large corporations, and barriers to compete with these corporations are becoming higher. To analyze these barriers, we can think of AI in terms of three layers, which are also true for the internet as a whole: (1) the hardware layer, (2) the code layer, and (3) the content layer (Hindman, 2009, pp. 39–40). An additional layer, which continues to rise in prominence in competition between large AI developers, is the search function.²² The algorithms designed by companies for their AI products and associated search engines are key to understanding user patterns and the content presented to users. Furthermore, when discussing the consolidation of media power, we must also consider the immense infrastructure necessary for AI, which also creates a significant barrier to meaningful market entry for anyone other than large corporations. This is often referred to as the “infrastructure divide.”²³

The media power accumulated by corporations that make the most popular AI products, and the influence of these products on the flow of communication and public discourse, means that these products' design merits public discussion based on democratic values, informed by Habermas's framework. His discourse-based political theory highlights the question: How does a topic gain recognition as controversial, as worthy of public debate, to begin with?²⁴

²² Google has embedded its AI product Gemini in Google Search, and Open AI (which is controlled by Microsoft) has also integrated a search function in its main AI product, ChatGPT.

²³ A study conducted as far back as 2009 already showed that Google spent as much on physical equipment as a typical telephone company (Hindman, 2009, p. 85).

²⁴ Habermas's procedural approach highlights the questions of what comes up for public discussion, what is a matter of public concern and that the struggle to make something public is a struggle for justice (Benhabib, 1992, pp. 79–80).

Furthermore, in considering the ways in which media power is dominated by a few profit-driven corporations, we can draw upon Habermas's warning that securing equitable participation in the discursive process requires attention to encroachment on public and private autonomy by the economic system just as much, if not more, than encroachment by government and the administrative system (Habermas, 1998, pp. 263–264). This means being as concerned with AI's impact on the public sphere as we are with its impact on our private sphere.

Designing for Trust

As we have seen, Habermas articulates ideal speech situations in which the “force of the better argument” prevails in public discourse. Rational persuasion can be understood to include three necessary components: (1) An attempt to provide reasons, and suggestion that the interlocutor ought to accept those reasons as a result of considered judgment; (2) an appeal to reasons that have some minimal degree of plausibility; and (3) an appeal to reasons that present a genuine *attempt* to be relevant to the issue at hand (Blair, 2012, p. 75).²⁵ That said, it is clear that not all forms of persuasion are indeed rational in this strict sense. For example, we often use images in persuasion. We can think of security camera footage clearly showing that a certain person committed a bank robbery being presented to persuade that someone else is innocent. In such a case, the image itself plays a central role in the argument being made (Aspeitia, 2012, p. 359). The ability of AI technology to generate artificial and yet highly realistic images (still and video) poses serious concerns for the ability to trust such argumentation in the public sphere. If one cannot even trust that their interlocutor is in fact human, and that the image they are seeing is authentic, then shared communicative action breaks down (Bar-Tura, 2011).

Furthermore, as Sanford Goldberg has shown, we inherently rely on others (semantically and epistemically) when we speak and make

²⁵ Blair notes here that “[t]his understanding of what makes persuasion rational implies that the classifications of cases of attempted or actual persuasion as rational, irrational or nonrational are in principle contestable. What counts as ‘minimally credible,’ ‘some measure of pertinence’ or ‘engaging the intellect’ will in some cases be controversial, for these are properties with vague borderlines and some cases will fall within those penumbras. But this feature is not a flaw: precision about such concepts as rationality (cf., practicality, efficiency) is a false ideal” (Blair, 2012, p. 75).

arguments. One obvious aspect of epistemic reliance on others when constructing arguments is the use of testimonial knowledge. In such cases, we take the testimony of another person as supporting evidence for our position. Our argument then relies on the knowledge provided by the testifier, and a key component in subsequent evaluation of our argument becomes the reliability attributed to the testifier and her testimony. That reliance on testimony necessitates the practice of *semantic* (in addition to epistemic) deference, which means relying on expert knowledge in order to recover the *meaning* of concepts in the testimony.²⁶ The speaker is implicitly relying on the expert for any further explication of the concept the speaker is deploying when providing testimony. We rely on a wide community of testifiers and experts not just for their knowledge, but for the meaning of our speech. Hence, semantic and epistemic deference as a practical function in public reason must be considered when designing AI, since users may defer to AI generated content and consider the AI algorithm as “expert.”

Thus, in order to design AI for democracy, it needs to be designed for trust. This might include regulation requiring any image or text generated by AI to be labeled as such, and done so in a way that cannot be altered. We need to know if a particular user on a social media platform is human or bot. We need to know whether a video is of real events or artificially generated. Just like we have laws about labeling food in certain ways, because we recognize that it’s a matter of public health, so too laws about labeling AI are a matter of the republic’s health.

AI and Our Political Imagination

Sherry Turkle has pointed to important consequences of software design for our political imagination. She remarks that since the 1980s users of computer software have become less and less interested in understanding how the software works and put more emphasis on functioning effectively

²⁶ Goldberg defines “expertise” as “the state of having specialized background knowledge (or at least justified belief) in a given domain, where the knowledge in question is organized in a manner that allows for easy access and use in appropriate circumstances” (Goldberg, 2009, p. 582). Furthermore, Goldberg clarifies that testimonial knowledge need not be acquired from experts (I can rely on my friend to know who won the Yankees game, but that does not make him an expert). Hence, *epistemic* reliance on others need not involve experts at all. The reliance on experts becomes more important for this discussion when considering *semantic* deference.

within the software design. In a sense, users put more emphasis on striving to play the game well than on questioning the rules of the game. Fostering this type of attitude has political impacts. In Turkle's words, it can compromise our "sense that understanding is accessible and action is possible" (Turkle, 2003, pp. 20–24). The invigoration of the political imagination is especially important in a society in which established norms and conventions are deeply entrenched. The more AI gains "substantial momentum," the more important it is to concern ourselves with the values and norms its designs embody.

A critical theory of AI must ask whether and how these technologies and their design expand or contract the political imagination, and whether they provide avenues for challenging the status quo. In the words of Evgeny Morozov, a critical approach "will trace how these technologies are produced, what voices and ideologies are silenced in their production and dissemination, and how the marketing literature surrounding these technologies taps into the zeitgeist to make them look inevitable" (Morozov, 2013, p. 356).

Heated debates about issues of privacy and intellectual property with respect to AI are prevalent. There are major concerns that data about citizens are being collected and used in a variety of ways, often without the explicit knowledge of the citizens. Though these concerns for privacy are worthwhile, they have overshadowed the process in which the digital public sphere has lost its public nature. This process is in accord with the liberal and neo-liberal ethos, which takes as its primary concern the right of the individual to maintain her own private sphere of autonomy. However, the public autonomy of citizens to shape the digital public sphere and meaningfully participate in it is being stymied in various ways.

When brought to its extreme, we see that lack of privacy and lack of publicity are two sides of the same coin. A society in which there is no privacy at all is clearly totalitarian. But so is a society in which no participatory public sphere exists. When we overly emphasize citizens' private autonomy—their right to pursue their own private interests with little regard for a meaningful concept of the public good—we risk losing citizens' ability (and motivation) to actively shape their social conditions.

Simons and others have offered new perspectives on the role of technology regulation and governance, if we are to have a flourishing democracy in the age of AI (Simons, 2023). We must debate the role that AI should play in the structures of our public sphere. Should we require public audits of AI algorithms when a company holds a certain share of the

market? Should we require OpenAI and Google to tell us how ChatGPT and Gemini generate their responses to our queries? Should we require social media companies to tell us how they determine what is “trending” and what “goes viral”? Should we require Facebook to tell us how it determines what we see in our “feed”? If such transparency would divulge trade secrets, should we at least require that an audit of these technologies be conducted by a public agency such as the FCC? These are important questions for public deliberations as we consider the ways in which digital technologies structure the public sphere.²⁷

CONCLUSION

As with any technology, the way in which AI is designed and developed embodies a certain view of the world as it should be, not just a description of the world as it is. And, like the technologies themselves, arguments in public discourse are also designed by drawing on social resources, and the ways in which arguments are presented involve values. By uncovering and exposing these dynamics, we broaden our understanding of what public reason is, what we can expect of it, and how we can utilize it when thinking through AI and democracy.

First, we must change the discourse about AI. We must encourage a discourse that questions the basic rules currently governing our technologies. AI is not a force of nature, and could be designed to fit our communal and democratic aspirations. The Habermasian political-philosophical framework is well positioned to inform discussions about the role of technology in society, since it focuses on communication and deliberation, which lie at the foundation of the design process.

In our thinking we must politicize technologies and the public sphere they help shape. There are strong currents in public discourse that aim to de-politicize the public sphere in the sense that they discourage the conversation about the technology’s design and its political effects. To democratize technology ultimately means uncovering the political implications of particular designs. Users may demand certain designs, but this does not mean that they are driven by considerations of democracy. To develop, design, and regulate AI through a Habermasian lens, is to do so in light of the regulative ideals of public reason, communicative action, and participatory democracy.

²⁷For an analysis of the ways in which social biases are embedded in algorithms, see Noble (2018).

REFERENCES

- Achterhuis, H. (2001). Introduction: American philosophers of technology. In H. Achterhuis (Ed.), *American philosophy of technology: The empirical turn*. Indiana University Press.
- Aspeitia, A. A. B. (2012). Words and images in argumentation. *Argumentation*, 26, 355–368.
- Bar-Tura, A. (2011). Between virtual reality and the real: Cyber subjectivity and ideology critique. *Humanities and Technology Review*, 30, 25–56.
- Benhabib, S. (1992). Models of public space: Hannah Arendt, the liberal tradition, and Jürgen Habermas. In C. Calhoun (Ed.), *Habermas and the public sphere*. MIT Press.
- Blair, J. A. (2012). Argumentation as rational persuasion. *Argumentation*, 26, 71–81.
- Cinderby, S. (1999). Geographic information systems (GIS) for participation: The future of environmental GIS? *International Journal of Environment and Pollution*, 11(3), 304–315.
- Cooper, S. (2006). The posthuman challenge to Andrew Feenberg. In T. J. Veak (Ed.), *Democratizing technology: Andrew Feenberg's critical theory of technology*. SUNY Press.
- Doppelt, G. (2001). What sort of ethics does technology require? *The Journal of Ethics*, 5(2), 155–175.
- Doppelt, G. (2006). Democracy and technology. In T. J. Veak (Ed.), *Democratizing technology: Andrew Feenberg's critical theory of technology*. SUNY Press.
- Feenberg, A. (1991). *Critical theory of technology*. Oxford University Press.
- Feenberg, A. (1999). *Questioning technology*. Routledge.
- Feenberg, A. (2011). Modernity, technology and the forms of rationality. *Philosophy Compass*, 6(12), 865–873.
- Goldberg, S. (2009). Experts, semantic and epistemic. *Nous*, 43(4), 581–598.
- Habermas, J. (1975). *Legitimation crisis*. Beacon Press.
- Habermas, J. (1984). *The theory of communicative action, Vol. I: Reason and the rationalization of society*. Beacon Press.
- Habermas, J. (1990). *Moral consciousness and communicative action*. MIT Press.
- Habermas, J. (1992). Technology and science as 'ideology'. In D. Ingram & J. Simon-Ingram (Eds.), *Critical theory: The essential readings* (pp. 117–150). Paragon House.
- Habermas, J. (1998). *Between facts and norms: Contributions to a discourse theory of law and democracy*. MIT Press.
- Hedrick, T. (2010). *Rawls and Habermas: Reason, pluralism, and the claims of political philosophy*. Stanford University Press.
- Hindman, M. (2009). *The myth of digital democracy*. Princeton University Press.

- Horkheimer, M. (1992). Means and ends. In D. Ingram & J. Simon-Ingram (Eds.), *Critical theory: The essential readings*. Paragon House.
- Hughes, T. (1983). *Networks of power: Electrification in Western society 1880–1930*. Johns Hopkins University Press.
- Ilde, D. (2001). Forward. In H. Achterhuis (Ed.), *American philosophy of technology: The empirical turn*. Indiana University Press.
- Ingram, D. B. (1990). *Critical theory and philosophy*. Paragon House.
- Ingram, D. B. (2010). *Habermas: Introduction and analysis*. Cornell University Press.
- Kaplan, D. M. (2009). Introduction. In D. M. Kaplan (Ed.), *Readings in the philosophy of technology*. Rowman and Littlefield.
- Kellner, D. (2000). Habermas, the public sphere and democracy: A critical intervention. In L. E. Hahn (Ed.), *Perspectives on Habermas*. Open Court.
- Krakauer, E. L. (1998). *The disposition of the subject: Reading Adorno's dialectic of technology*. Northwestern University Press.
- Marcus, G. (2024). *Taming Silicon Valley: How we can ensure that AI works for us*. MIT Press.
- McCarthy, T. (1975). Translator's introduction. In J. Habermas (Ed.), *Legitimation crisis*. Beacon Press.
- Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Rasmussen, D. M. (1994). How is valid law possible? A review of Faktizität und Geltung by Jürgen Habermas. *Philosophy and Social Criticism*, 20(4), 21–44.
- Rehg, W. (2009). *Cogent science in context: The science wars, argumentation theory, and Habermas*. MIT Press.
- Simons, J. (2023). *Algorithms for the people: Democracy in the age of AI*. Princeton University Press.
- Stump, D. J. (2006). Rethinking modernity as the construction of technological systems. In T. J. Veak (Ed.), *Democratizing technology: Andrew Feenberg's critical theory of technology*. SUNY Press.
- Tijmes, P. (2001). Albert Borgmann: Technology and the character of everyday life. In H. Achterhuis (Ed.), *American Philosophy of Technology: The Empirical Turn*. Indiana University Press.
- Turkle, S. (2003). From powerful ideas to PowerPoint. *Convergence*, 9(2), 19–25.
- Veak, T. (2000). Whose technology? Whose modernity? Questioning Feenberg's Questioning Technology. *Science, Technology, & Human Values*, 25(2), 226–237.