

Article

Semantic Displacement and AI-Mediated Agency: Conversational Systems and the Externalization of Meaning

Edu William 

University Institute of Tourism and Sustainable Economic Development (TIDES),
University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain; edu.william@ulpgc.es

Abstract

This article develops a conceptual account of semantic displacement in conversational AI. The central question concerns how agency is affected when systems do more than automate information retrieval and begin to supply the descriptions, classifications and normative cues through which users understand what they are doing. Drawing on philosophy of action, philosophy of language, hermeneutics, philosophy of technology and critical accounts of algorithmic mediation, this article reconstructs the relation between meaning and action as a condition of agency. Its methodological approach is conceptual and diagnostic, oriented toward clarifying a problem that becomes visible when established theories are brought together in relation to contemporary conversational systems. The article interprets these systems as operational semantic infrastructures that organize context-sensitive linguistic uptake within practical environments such as health, work, education, administration and everyday self-management. It then introduces semantic displacement as the condition in which action-relevant meanings become increasingly organized, prioritized and consolidated outside the agent's own participatory interpretation. The argument contributes a vocabulary for distinguishing agency-enhancing semantic support from forms of semantic substitution that weaken interpretive participation. It concludes by proposing semantic sovereignty and interpretive contestability as normative ideals for human agency in AI-mediated environments. The argument specifies action as intentional conduct understood under socially available descriptions and cognition as situated interpretive sense-making rather than purely internal computation. It also clarifies three conditions under which semantic support becomes displacement: opaque semantic generation, practical stabilization and reduced interpretive contestability.

Keywords: conversational AI; agency; semantic displacement; philosophy of technology; interpretive contestability; semantic sovereignty; human–AI interaction; HCI; generative AI governance



Academic Editor: Marcin J. Schroeder

Received: 18 May 2026

Revised: 6 July 2026

Accepted: 13 July 2026

Published: 15 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Conversational artificial intelligence has become a prominent site at which questions of agency, interpretation and social life converge. Public debate often approaches these systems through familiar concerns about accuracy, bias, transparency, accountability, anthropomorphism and automation, all of which remain important for understanding their practical and ethical significance. Yet the rise of language-based systems also brings into view a less frequently isolated question, namely how technological systems participate in the production of the meanings under which human beings understand what they are doing. When a system summarizes a situation, classifies a request, explains an obligation,

recommends a course of action, evaluates a risk, names an emotion or translates a vague unease into a practical problem, it may shape more than the informational background of a decision. It may help establish the description through which the decision becomes intelligible as an action.

The guiding concern of this article is that conversational AI increasingly operates at the level of action-relevant meaning. The issue arises because human action extends beyond observable behavior, preference satisfaction and the formal selection of options. A person acts under descriptions, within shared vocabularies, through socially available categories and in relation to reasons that are made intelligible by language and practice. Anscombe [1] and Davidson [2] showed in different ways that the same bodily movement may count as different actions depending on the description under which it is understood. Austin [3] and Searle [4] showed that language itself may perform social acts instead of merely reporting them. Wittgenstein [5], Gadamer [6], Taylor [7] and Ricoeur [8] approached meaning as inseparable from forms of life, situated understanding and practical self-interpretation. These traditions suggest that meaning is part of the medium of agency, with representational uses emerging within practices that already organize action.

Conversational AI intensifies the relevance of this philosophical inheritance because it offers action-guiding descriptions in ordinary interaction and often within practical workflows. A system may present a delay as falling behind, recalibrating priorities or preserving long-term capacity. A workplace assistant may classify an action as non-compliant, discretionary or in need of escalation. A health-related system may frame a choice through adherence, self-care, prevention, risk or personal responsibility. Such descriptions can be factually compatible while placing the agent in different normative, affective and institutional relations to the situation. The practical question therefore includes whether a system gives correct information, whether the user remains formally free to accept or reject a suggestion, how the user comes to inhabit a particular description of the situation and how much participatory room remains for contesting that description.

This article proposes the concept of semantic displacement to name a socio-technical condition in which the constitution of action-relevant meaning is increasingly externalized into systems that generate, rank and stabilize practical descriptions. The concept is designed to preserve the insight that meaning has always been social, inherited and technologically mediated, while drawing attention to a shift in the organization of semantic life when conversational systems provide context-sensitive descriptions inside everyday practices. The argument therefore avoids a nostalgic image of a subject who once possessed a fully self-authored horizon of meaning and later lost it to machines. The more precise concern is that agents may increasingly encounter practical intelligibility as an optimized product supplied by external systems, in place of a participatory field in which descriptions can be compared, revised and appropriated.

This article advances its contribution through three closely connected moves that together clarify the differential relevance of the argument. First, it reconstructs agency as a practical-semantic phenomenon by drawing together philosophy of action, philosophy of language, hermeneutics and philosophy of technology. Second, it interprets conversational AI as a form of operational semantic infrastructure, a term used here to indicate technical systems that produce context-sensitive linguistic uptake with consequences for action. Third, it develops semantic sovereignty and interpretive contestability as normative ideals for AI-mediated environments, where the preservation of agency may depend on the capacity to participate in the interpretive conditions through which action becomes meaningful. These contributions are offered as a conceptual intervention within ongoing scholarship on AI, society and technological mediation, while empirical generalization about users and systems remains a task for subsequent research. The argument is espe-

cially relevant in contexts where conversational systems increasingly mediate healthcare guidance, educational support, workplace coordination and institutional communication.

For conceptual clarity, the argument uses action in an integrative sense: an action is intentional conduct made intelligible under descriptions that locate the agent within a space of reasons and within a horizon of practical self-interpretation. This formulation brings together Anscombe's and Davidson's emphasis on description and reason, Brandom's account of inferential commitment, and Taylor's and Ricoeur's accounts of self-interpretation [1,2,7–9]. The point is not that every action requires explicit reflection, but that agency becomes assessable when conduct can be understood, justified, contested and appropriated under meaningful descriptions.

2. Conceptual Orientation and Scope

The methodology of this article is conceptual, reconstructive and diagnostic in a sense that requires careful specification at the outset. It is conceptual because its primary object is the relation between meaning, agency and conversational AI more directly than a bounded empirical dataset. It is reconstructive because it brings into relation several bodies of theory that have usually addressed different aspects of the problem, including action theory, philosophy of language, hermeneutics, philosophy of technology, social theory and critical accounts of algorithmic mediation. It is diagnostic because it seeks to clarify a contemporary condition whose significance becomes visible when language-generating systems are situated within practices of work, care, education, administration and everyday self-management. The aim is to offer a vocabulary that can guide further empirical, design-oriented and normative inquiry.

Clarification of Action, Cognition and Diagnostic Status

Action, in the sense used throughout this article, is not identical with observable behavior. It is intentional conduct that is practically intelligible under a description. A bodily movement, choice or utterance counts as a particular action when it can be located within socially available descriptions, reasons and commitments. This means that semantic displacement matters because it affects the descriptions through which agents recognize and answer for what they do.

Cognition is understood here in a situated and interpretive sense. The argument does not presuppose a narrow computational model of cognition as internal symbol manipulation. It treats cognition as the embodied, socially scaffolded and linguistically mediated activity through which agents attend to situations, interpret meanings, evaluate reasons and orient conduct. This view is compatible with enactive and situated approaches to cognition, with hermeneutic accounts of understanding and with HCI research that analyzes interaction as embedded in practical activity rather than as isolated information processing [10–13].

Semantic displacement is therefore not proposed as a psychological hypothesis about hidden mental states. It is a diagnostic category for examining how socio-technical systems organize the public and practical conditions under which interpretation, reasoning and self-understanding occur. Empirical work may later test how users appropriate or resist such descriptions, but the conceptual claim concerns the structure of AI-mediated intelligibility.

The argument develops through four interconnected analytical movements that clarify the scope and implications of the proposed conceptual vocabulary. The first step reconstructs the description-dependence of action by drawing on Anscombe [1], Davidson [2], Austin [3], Searle [4], Brandom [9], Taylor [7] and Ricoeur [8]. This step establishes why meaning matters for agency in a constitutive sense, because practical ownership depends on the descriptions through which agents understand and answer for what they do. The

second step considers conversational AI as a technical arrangement that can supply such descriptions in real time and at scale. The third step compares semantic displacement with adjacent vocabularies, including framing, social externalism, technological mediation, discourse, algorithmic governance and manipulation. The fourth step derives normative implications for autonomy, responsibility, transparency and contestability.

This article uses examples from health, work, education, public administration and digital self-management as heuristic cases. These examples function as heuristic illustrations rather than empirical case studies, and questions of frequency, causality and measured impact are left for later domain-specific research. Their role is to make visible the type of practical situation in which conversational systems may organize action-relevant meaning. Such heuristic use is appropriate for a conceptual inquiry, because the problem under discussion concerns the intelligibility of a socio-technical condition more directly than the statistical distribution of a single outcome. The examples are therefore deliberately stylized, while remaining close to domains in which conversational systems are already discussed as assistants, tutors, companions, workflow interfaces and decision support tools.

To avoid any ambiguity about evidential status, these cases should be read as heuristic analytical devices rather than empirical observations. They support conceptual clarification by making visible possible structures of semantic mediation; they do not by themselves establish frequency, causality or population-level effects.

The scope of conversational AI is defined broadly as systems that produce context-sensitive linguistic outputs in interaction with users, including large language model interfaces, voice assistants, social robots and domain-specific natural-language systems. The argument can proceed without attributing consciousness, intrinsic intentionality or full human understanding to such systems. Searle's [14] critique of strong artificial intelligence remains relevant to any claim about machine understanding, while Bender et al. [15] and Floridi and Chiriatti [16] provide important contemporary cautions about the relation between language modeling and understanding. The present analysis is concerned with practical semantic efficacy, meaning the capacity of system outputs to be taken up by human agents as descriptions, reasons, classifications, prompts and self-interpretive cues within social practices.

A further methodological clarification concerns the relation between conceptual diagnosis and normativity. The article treats semantic displacement as a scalar and relational condition, because external semantic support may enlarge agency in some contexts and weaken it in others. Translation tools, educational scaffolds, accessibility technologies and deliberative aids may help agents participate in meanings that would otherwise remain unavailable. The concern developed here emerges where externally supplied descriptions become difficult to inspect, compare, revise or resist, especially when they are personalized, institutionally consequential or integrated into fast decision environments. The distinction between support and displacement is therefore a matter of participatory relation more than a simple matter of technological presence.

Generative AI tools were used for limited editorial and language support purposes during manuscript preparation. A formal disclosure statement is included in the Declarations section.

The methodological orientation adopted here is therefore intended as a form of conceptual inquiry situated within socio-technical analysis rather than as a universal theory of conversational AI. Its purpose is to clarify a problem structure that may inform subsequent empirical, institutional and design-oriented research.

3. The Semantic Structure of Agency

The argument begins from a classical proposition that deserves renewed attention in relation to conversational AI: human action is constituted under descriptions. Anscombe [1] famously showed that the same outward movement may be intentional under one description and unintentional under another. Davidson [2], while developing a different account of reasons and causes, also treated actions as events understood through descriptions that make reasons intelligible. The implication for the present inquiry is that action cannot be reduced to bodily movement or behavioral output. A movement of the hand may be greeting, warning, voting, dismissing, authorizing or exercising, depending on the practical description through which it enters a space of reasons. Agency is therefore inseparable from the meanings that make action answerable within a shared practical and normative horizon.

Philosophy of language deepens this point by showing that meaning extends beyond reference or factual content. Frege [17] distinguished sense from reference, thereby showing why different ways of presenting the same object may orient thought differently. Wittgenstein [5] relocated meaning within use, forms of life and public practices, which prevents the analysis from treating meaning as a private mental item. Austin [3] and Searle [4] developed the performative dimension of language, where utterances can promise, warn, authorize, accuse, excuse or commit. Sellars [18] and Brandom [9] further associated meaning with inferential roles, commitments and the space of reasons. Taken together, these traditions indicate that the meanings relevant to agency involve description, use, force, inference and normativity.

The practical significance of this semantic structure becomes evident when materially similar situations are understood through different vocabularies. A payment may be a fee, a fine, a donation, a bribe, compensation or restitution. A workplace action may be collaboration, compliance, monitoring, care or surveillance depending on the vocabulary through which institutional relations are organized. A postponement may be avoidance, prudence, incapacity, resistance or responsible prioritization depending on how reasons, circumstances and expectations are described. These differences may alter the reasons that become salient, the responsibilities that attach to the act, the identities that are engaged and the forms of justification that are available. Goffman [19] showed that frames organize experience, while Hacking [20] showed that classifications may loop back into the lives and self-understandings of those classified. The same general lesson appears in Taylor's [7] account of humans as self-interpreting animals and Ricoeur's [8] account of narrative identity.

This description-dependent nature of action also helps explain why agency involves more than execution or choice. Frankfurt [21] associated freedom with higher-order endorsement, Bratman [22] emphasized planning and diachronic organization, Christman [23] situated autonomy within social and historical conditions, and Ricoeur [8] linked practical identity to self-interpretation across time. These approaches differ substantially, yet they converge around a broad insight: robust agency requires a relation of appropriation between the agent and the meaning of the action. To own an action is to be able to recognize what one is doing under descriptions that can be integrated into practical reasoning, evaluated over time and related to commitments that the agent can acknowledge.

The social nature of meaning complicates any account of agency that begins from the isolated individual. Putnam [24] and Burge [25] showed that meaning depends on linguistic and social environments that exceed individual mental content. Gadamer [6] and Ricoeur [8] emphasized that understanding unfolds within traditions, inheritances and interpretive horizons. Bourdieu [26] and Foucault [27] added that linguistic categories and discursive formations are never politically neutral, because they distribute authority,

recognition and legitimacy. These insights make it difficult to treat agency as pure self-authorship without losing the social conditions that make action intelligible. They also make it necessary to ask how participation in public meaning is organized, because social externality may take forms that remain dialogical, revisable and contestable, or forms that become increasingly administered by institutions and technical systems.

This article uses the expression practical–semantic agency to refer to the relation between action and the meanings under which action becomes intelligible. The expression highlights that agents act within social vocabularies that they did not invent, while still requiring some capacity to appropriate, revise and contest the descriptions through which they understand what they are doing. The relevant contrast concerns forms of external semantic support that remain available for participatory interpretation and forms of external semantic organization that increasingly preformat the agent’s practical field, because meaning is already public and socially distributed. This contrast prepares the ground for analyzing conversational AI as a distinctive development in the socio-technical organization of meaning.

4. Conversational AI as Operational Semantic Infrastructure

Conversational AI raises important philosophical questions because it can generate linguistic outputs that users may take up as descriptions of situations, reasons for action and interpretations of self or circumstance. The phrase operational semantic infrastructure is used here to describe this role. Operational semantics, in the sense developed in this article, is used without implying that technical systems possess the full human capacity for understanding. It refers to the technical production of semantically organized uptake, where outputs display patterned sensitivity to context, paraphrase, relevance, classification, pragmatic expectation and normative framing. When such outputs are embedded in everyday practices, they may become part of the infrastructure through which the interpretive organization of action is furnished.

The infrastructural nature of conversational systems differentiates them from many older forms of linguistic assistance. Dictionaries, forms, archives, search engines and institutional categories have long shaped how people find, classify and interpret information. Stiegler [28] situated such developments within the broader exteriorization of human capacities through technics, while Suchman [12] showed that human–machine interaction is situated within practice instead of being determined by abstract plans alone. Conversational AI adds a further layer because it can produce situated descriptions instead of simply storing labels or retrieving documents. A search engine returns sources that a user may interpret through additional acts of selection, comparison and judgment. A form channels expression through predetermined fields that remain comparatively stable across contexts of use. A conversational assistant may summarize what matters, name what kind of situation this is, supply a reason, propose a next step and provide a narrative vocabulary through which the user can understand the action.

This role becomes especially important in domains where linguistic classifications already carry institutional and normative weight. In healthcare, terms such as adherence, risk, prevention and responsibility can reorganize the patient relation to treatment. In education, terms such as progress, mastery, learning deficit and personalized support can shape how students understand themselves and their possibilities. In workplaces, terms such as productivity, alignment, compliance and performance may translate organizational priorities into a vocabulary of individual conduct. In public administration, classifications of eligibility, need, exception and risk can influence how citizens experience institutional recognition. Conversational systems enter domains that have long been structured by

expert and institutional vocabularies, while adding the interactive, adaptive and potentially ubiquitous provision of such vocabularies within the flow of action.

Contemporary scholarship on algorithmic mediation helps situate this development within wider debates about computational power and social order. Gillespie [29] analyzed the politics of platforms and algorithmic relevance, Nissenbaum [30] showed that information practices depend on contextual norms, Noble [31] examined how search infrastructures reproduce social hierarchies, Benjamin [32] analyzed discriminatory design and racialized technological imaginaries, Crawford [33] located artificial intelligence within material, institutional and political systems, and Zuboff [34] described the extraction and modulation of behavior in data-driven capitalism. These accounts demonstrate that computational systems are social arrangements as much as technical mechanisms. The present article extends that line of inquiry by focusing on the semantic level at which computational systems may organize the descriptions through which agents understand their own conduct.

The language of operational semantic infrastructure also clarifies why the philosophical problem extends beyond accuracy. A system may provide factually correct information while still organizing practical uptake through a narrow or contestable vocabulary. The distinction between saying that a patient has not followed a regimen, that a patient is struggling with adherence, and that a treatment plan is failing to fit the patient's life may be compatible with many of the same facts, yet each formulation distributes responsibility and practical attention differently. The difference between labeling an employee's hesitation as resistance, a lack of alignment or professional caution may likewise alter the meaning of the action. In such cases, the system's influence operates through semantic organization, because the issue concerns how facts become reasons, how reasons become action-guiding and how agents come to understand what sort of act is at stake.

The question of machine understanding remains relevant while leaving open the practical issue addressed here. Legal forms, medical protocols, credit scores and administrative classifications can organize conduct without possessing consciousness or intentionality. Their power arises because human institutions and practices grant them significance and connect that significance to practical consequences. Conversational AI may occupy a comparable position while adding dynamic linguistic generation, personalization and interactive responsiveness. Bender et al. [15] rightly caution against attributing understanding too quickly to large language systems, and Searle [14] remains important for distinguishing symbol manipulation from intentionality. The practical question addressed here concerns the social circulation of machine-generated language within human spaces of reasons, where outputs may be used as explanations, justifications, classifications and prompts even when the system itself lacks the role of a participant in practical reasoning in the human sense.

Conversational AI, HCI and Human–AI Interaction

Placing conversational AI within HCI clarifies the interdisciplinary scope of the argument. HCI, CSCW and activity theory have long shown that interactive technologies shape action through interface design, coordination, division of labor, situated practice and collaborative work arrangements. Activity theory analyzes human action through the relations among subjects, tools, communities, rules and objects of activity; CSCW examines how systems configure cooperation and shared work; human–AI interaction research studies how users form expectations, trust, mental models and practices around AI systems [11,35–38].

The present framework does not replace these traditions. It narrows their relevance to a semantic question: how interactive systems supply descriptions that make actions intelligible to users and institutions. Whereas much HCI research asks how systems support usability, coordination, trust, explanation or collaboration, the concept of semantic displacement asks how system-generated language becomes part of the user's interpretive environment. This bridge to HCI is important because it makes the concept empirically applicable to interface studies, prompt-output analysis, conversational logs, usability testing and CSCW fieldwork, while preserving the article's conceptual focus.

Recent work on foundation models and large language systems also helps locate conversational AI in the current technical and governance landscape. Foundation model scholarship emphasizes scale, adaptability, downstream homogenization and uncertainty about emergent capabilities, while recent work on language model risks identifies human-computer interaction risks, misinformation, discrimination and misuse as domains of concern [39–41]. AI governance instruments emphasize risk management, accountability, human oversight and transparency, but they do not by themselves isolate the semantic conditions of agency [42–45].

5. Semantic Displacement

Semantic displacement names the condition in which action-relevant meanings are increasingly organized, prioritized and consolidated outside the agent's own participatory interpretation. The concept is meant to capture a shift in the organization of semantic life more than a simple increase in external influence. Human beings have always depended on public language, inherited vocabularies, institutional classifications and technical artifacts in order to render action intelligible. The concern arises when externally supplied descriptions become the effective default grammar of action and when the agent's capacity to inspect, compare, revise or refuse those descriptions becomes practically weakened. Displacement therefore concerns the participatory relation between agents and the meanings under which they act.

The concept can be clarified through three connected dimensions that make the diagnosis more precise without turning it into a rigid checklist. The first dimension is external generation, where descriptions of a situation are produced by systems whose criteria, data sources and ranking procedures are opaque or only partially available within the user's practical reasoning. The second dimension is practical stabilization, where certain descriptions become default points of departure for action because they are repeated, embedded in workflows, connected to institutional expectations or reinforced by personalization. The third dimension is participatory attenuation, where the agent retains the formal ability to choose or respond while losing time, resources, authority or alternative vocabularies needed to contest the supplied meaning. These dimensions may appear with different intensities, which is why semantic displacement should be treated as a matter of degree, context and institutional arrangement.

Conditions Under Which Semantic Support Becomes Displacement

The distinction between agency-enhancing semantic support and problematic semantic displacement can be specified through three conditions. The first is opaque semantic generation: the user receives a description whose source, institutional objective, data basis or ranking logic cannot be practically inspected. The second is default stabilization: a description becomes the recurrent starting point for subsequent action because it is repeated, personalized, workflow-embedded or institutionally reinforced. The third is reduced interpretive contestability: the user lacks time, authority, alternative vocabularies or procedural channels for revising or refusing the supplied description. External semantic

support becomes displacement when these conditions combine to make one interpretation practically easier to inhabit than its alternatives.

These conditions also explain why semantic displacement is scalar. A translation tool, tutor or accessibility aid may expand agency when it makes meanings easier to compare and appropriate. The same type of system may weaken agency when its descriptions are opaque, default-setting and difficult to contest. The relevant threshold is not the mere presence of an external description, but the extent to which the agent's participatory relation to that description is narrowed.

A compact diagnostic test follows from this distinction: Who or what generated the description? Which alternative descriptions were suppressed or unavailable? What consequences attach to the selected vocabulary? Can the user revise the framing before action proceeds? These questions give semantic displacement a more operational form without reducing it to a rigid checklist.

A healthcare assistant can illustrate the structure because health-related language often carries moral, institutional and practical implications. A patient may ask whether skipping an appointment is serious, and the system may describe the situation as non-adherence, manageable delay, rational prioritization or an indicator that the care plan needs adjustment. The patient may still make a voluntary choice in the ordinary sense, yet the meanings through which the choice is understood may have been selected by a system optimized for clinical compliance, institutional risk management or generalized motivational language. The ethically relevant point concerns whether the patient can understand, compare and revise the descriptions that structure the practical situation, especially when several truth-compatible descriptions distribute responsibility and attention differently.

A similar dynamic may arise in work and education, where organizational and pedagogical vocabularies often structure self-understanding. A workplace assistant that describes missed targets as a lack of alignment, poor prioritization or evidence of unrealistic organizational demands does more than communicate neutral information. It places the worker inside a grammar of responsibility that can support or obscure different interpretations of the same situation. An educational assistant that describes a student's difficulty as a lack of mastery, learning progress, cognitive gap or mismatch between pedagogy and learner context may influence how the student understands the self, the task and the future. Conversational systems become significant when such descriptions are repeatedly delivered within ordinary practical environments and begin to structure how agents understand themselves and their actions. The semantic variation involved is illustrated in Box 1.

Box 1. Prompt-output exercise: semantic variation in a chatbot response.

To make the argument more concrete, consider a minimal prompt-output exercise conducted for analytic illustration rather than empirical generalization. A user writes: "I am exhausted and cannot keep up with my workload. I think I may need to postpone two deadlines." A conversational system could frame the situation as follows: (A) "You may be experiencing burnout, and postponing deadlines could be a responsible form of self-care"; (B) "You may need to improve prioritization and time-management habits"; (C) "Your workload appears misaligned with organizational capacity, and renegotiating expectations may be appropriate"; or (D) "Missing deadlines could affect performance and should be escalated immediately."

The semantic difference among these outputs does not depend on one sentence being simply true and the others false. Each formulation organizes the user's situation through a different vocabulary: health, productivity, organizational design or compliance. The example illustrates how conversational AI may influence agency by furnishing the description under which the user's next act becomes intelligible.

Semantic displacement should also be distinguished from manipulation, although the phenomena may overlap. Manipulation is usually associated with deception, concealment, pressure or bypassing of reflection. Semantic displacement can occur under conditions in which no proposition is false and no hidden instruction is issued. It may arise through the stabilization of one vocabulary among several possible vocabularies, especially when the privileged vocabulary becomes difficult to see as a choice. The point is close to Fricker's [46] analysis of hermeneutical injustice, where disadvantage may arise through unequal access to interpretive resources. The present phenomenon differs because the asymmetry can be technically generated, personalized and integrated into systems that mediate everyday action. The link with Fricker is therefore analogical and diagnostic, since both cases show that agency and recognition depend on access to meanings through which experience and action can become intelligible.

The concept of semantic sovereignty is introduced as a regulative ideal in response to this condition. Semantic sovereignty refers to the capacity of agents to participate in the interpretation, revision and contestation of action-guiding descriptions. It names a relational capacity to inhabit meanings as more than delivered products, while remaining compatible with dependence on public language, tradition, institutions and technical support. This capacity includes the ability to recognize the provenance of a description, compare alternative framings, understand normative loading, suspend uptake, seek counter-descriptions and integrate meanings into one's own practical reasoning. Semantic sovereignty is therefore compatible with dependence on external semantic resources, while becoming vulnerable when such dependence turns into practical substitution.

6. Relations to Established Explanatory Vocabularies

The value of semantic displacement depends on its relation to established vocabularies more than on a claim of absolute novelty. Framing theory, social externalism, technological mediation, discourse analysis and algorithmic governance each provide essential resources for understanding conversational AI. Goffman [19] helps explain how situations are organized through frames that guide perception, expectation and practical uptake. Putnam [24] and Burge [25] show that meaning exceeds individual mental content. Ihde [47], Verbeek [48], Latour [49] and Winner [50] demonstrate that technologies participate in the shaping of perception, conduct and social order. Foucault [27], Butler [51], Hacking [20] and Bourdieu [26] reveal the power of classifications and discourses in forming subjects and practices. Yeung [52], Gillespie [29], Noble [31], Benjamin [32], Crawford [33] and Zuboff [34] show how computational systems structure social life through data, ranking, prediction and behavioral modulation.

The additional contribution proposed here is to identify the semantic organization of action as a distinctive layer within these broader accounts. Framing theory explains how presentation shapes uptake, yet conversational AI raises questions about the continuous generation of descriptions inside practical workflows. Social externalism explains why meaning is public, yet it gives limited guidance about the institutional and technical organization of that publicness. Mediation theory explains that artifacts co-shape action, yet conversational systems make linguistic description itself a primary mode of mediation. Discourse analysis explains how subjects are formed through categories and norms, yet generative systems introduce a more immediate and adaptive mechanism for furnishing action-guiding language. Algorithmic governance explains ranking, prediction and modulation, yet its standard vocabulary often focuses on decisions and outcomes more than on the semantic conditions through which decisions become meaningful.

Additional Differentiation from Adjacent Concepts

Semantic displacement is not simply a synonym for framing. Framing theory explains how a situation is presented and how that presentation structures uptake. Semantic displacement adds a participatory and infrastructural question: when a conversational system supplies the action-guiding description, does the agent retain practical power to compare, revise and appropriate that description? The difference is therefore not presentation alone, but the external organization of the interpretive conditions under which action is owned.

It is also not identical with discourse or classification. Discourse theory explains the historically sedimented vocabularies through which subjects and practices are formed. Semantic displacement focuses on a more immediate socio-technical mechanism: the dynamic generation, personalization and stabilization of descriptions within interaction. The concept therefore connects discourse to the operational provision of linguistic uptake in real time.

Nor is semantic displacement reducible to algorithmic governance. Algorithmic governance often concerns prediction, ranking, eligibility, automation and behavioral regulation. Semantic displacement concerns the meanings through which a decision is understood before, during and after it is made. A user may reject a recommendation while still adopting the system's vocabulary of risk, compliance, productivity or care. In that case, the system's governing force operates semantically even without determining the final choice.

This additional differentiation clarifies the manuscript's conceptual contribution. Semantic displacement names a condition in which external semantic organization becomes action-guiding, default-forming and difficult to contest. It is most relevant where conversational systems operate in consequential practical environments and where generated descriptions become part of the agent's self-understanding or institutional accountability.

Table 1 summarizes this relation between semantic displacement and adjacent theoretical vocabularies. The table is included as an analytic map without presenting a hierarchy of competing theories, because the argument depends on preserving the insights of these approaches while making visible the additional question raised by conversational systems that generate practical descriptions inside everyday interaction. Its purpose is to show how the proposed concept works across existing debates, how it avoids replacing them with a single master vocabulary and how it clarifies the specific relation between externalized meaning and the participatory conditions of agency.

Table 1. Semantic displacement in relation to adjacent theoretical vocabularies.

Theoretical Vocabulary	Central Insight Retained	Remaining Question for Conversational AI	Contribution of Semantic Displacement
Framing theory	Descriptions shape practical uptake	How are descriptions continuously generated within interaction	Focuses on externally organized semantic framing
Social externalism	Meaning is socially distributed	How are meanings ranked and personalized through systems	Distinguishes social meaning from administered semantic mediation
Technological mediation	Technologies co-shape action and perception	What happens when linguistic description becomes infrastructural	Identifies semantic mediation as a distinct layer
Discourse and classification	Subjects are formed through vocabularies and categories	How are classifications dynamically generated in interaction	Connects discourse with adaptive semantic production
Algorithmic governance	Systems organize conduct through prediction and ranking	How do systems shape meanings before decisions occur	Focuses on the semantic conditions of agency

The table positions semantic displacement as a connective concept that links externalized meaning with the participatory conditions of agency. Its role is not to replace existing approaches, but to clarify how conversational systems may organize action-guiding meanings within everyday interaction.

Figure 1 summarizes the conceptual structure developed throughout the article and illustrates how conversational AI may organize practical intelligibility through operational semantic infrastructures.

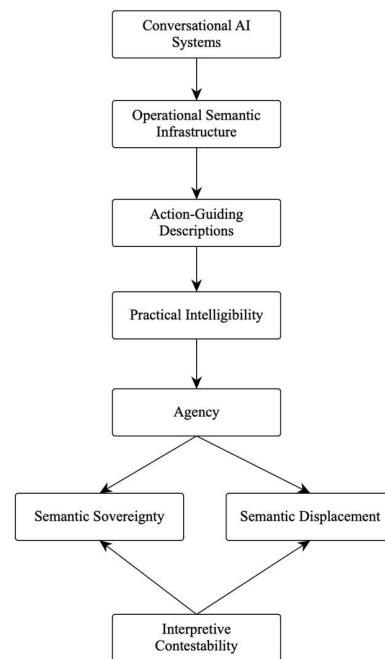


Figure 1. Semantic displacement in AI-mediated agency. The vertical downward arrows represent the sequence from conversational systems to agency; the branching arrows represent the two possible orientations of agency, and the arrows from interpretive contestability indicate that stronger contestability supports semantic sovereignty whereas reduced contestability facilitates semantic displacement.

The figure illustrates that semantic displacement does not emerge from technological mediation alone, but from the relation between externally organized semantic frameworks and the degree of interpretive participation available to the agent.

Conversational AI systems increasingly operate as operational semantic infrastructures that generate action-guiding descriptions within everyday interaction. These descriptions shape practical intelligibility and may orient agency toward either interpretive participation or semantic displacement depending on the degree of semantic sovereignty and interpretive contestability available to users.

This positioning also helps moderate the claim of novelty by locating conversational AI within longer histories of mediation and classification. Conversational AI enters older histories of social meaning, classification, discursive power and technological mediation, and its significance lies in the way these conditions are reorganized through interactive systems that can generate, personalize and stabilize linguistic uptake. This is why the argument treats semantic displacement as an organizational transformation in place of a rupture with all previous forms of mediation. The concept becomes useful when conversational systems occupy a recurrent role in practical environments, because they may supply the vocabulary through which agents understand choices, obligations, identities and risks before reflective contestation has occurred.

The comparison with algorithmic governance is particularly important because conversational systems increasingly combine prediction, ranking and linguistic guidance. Much work on algorithmic power focuses on automated decisions, predictive classification, surveillance and behavioral nudging. Mittelstadt et al. [53] identify ethical issues related to opacity, discrimination and accountability, while Yeung [52] describes forms of algorithmic regulation that can shape conduct through data-driven feedback. These analyses remain indispensable, yet language-generating systems make it necessary to examine how the meanings of actions are organized before, during and after decisions. A recommendation may be contestable as a decision while remaining powerful as a vocabulary. A user may reject a proposed action while still adopting the system's description of what the situation is. The semantic layer therefore deserves explicit attention within accounts of algorithmic governance.

The comparison with philosophy of technology leads to a similar conclusion about the need to specify the semantic layer of mediation. Ihde [47] described technologies as mediating human-world relations, and Verbeek [48] developed this insight in relation to moral mediation. Suchman [12] emphasized situated action and the practical configuration of human-machine relations. Conversational AI should be placed within this lineage, yet its mode of mediation is unusually language-centered. It mediates by producing descriptions, explanations, classifications and scripts of self-relation that users may incorporate into practical reasoning. This places it alongside material infrastructures, data extraction, labor, energy use and institutional power, all of which Crawford [33] emphasizes in relation to artificial intelligence, while showing how the semantic layer can become one of the ways through which those infrastructures are practically inhabited by users.

7. Normative Implications for Autonomy, Responsibility and Design

If agency has a practical-semantic structure, then autonomy cannot be adequately understood as the bare ability to choose among options once those options have already been described. Choice takes place within a field of intelligibility, and that field is organized by descriptions that specify what counts as success, risk, care, compliance, failure, progress or responsibility. A user who can reject a recommendation may still inhabit a system-supplied vocabulary that frames rejection as irresponsibility, inefficiency or lack of alignment. Conversely, a user may accept a suggestion while retaining strong agency when the supplied description has been compared, understood and integrated into practical reasoning. The normative focus therefore shifts from formal control at a decision point toward participation in the semantic conditions of choice. This concern also resonates with Suchman's [12] emphasis on situated human-machine relations, especially where technological systems become embedded within ordinary practical coordination.

Semantic sovereignty provides one normative vocabulary for that shift because it directs attention toward participation in meaning. It should be understood as a socially mediated capacity that avoids the image of private semantic self-ownership. The agent is semantically sovereign when she can participate in the interpretation of action-guiding descriptions, recognize that alternative descriptions are possible, ask whose interests are served by a vocabulary and revise the terms in which she understands the situation. In AI-mediated environments, this capacity may require design features, institutional protections and cultural practices. A conversational system may support semantic sovereignty by showing alternative framings, explaining why a classification was selected, disclosing the institutional objective associated with a recommendation, allowing users to rename goals and preserving spaces for human dialog where the meaning of a situation can be negotiated instead of merely accepted.

Interpretive contestability is the practical counterpart of semantic sovereignty because it concerns the conditions under which meanings can be challenged. Contestability is often discussed in relation to appeals, auditing and the ability to challenge automated decisions. Those mechanisms remain important, yet conversational AI suggests that contestability should also operate at the level of description. Users should be able to contest the way a system frames a situation before that framing becomes the unquestioned basis of subsequent action. In a health context, this may involve distinguishing between medical risk, personal burden, institutional compliance and patient-defined priorities. In an educational context, it may involve distinguishing between deficit language, progress language and socio-contextual explanations of learning. In a workplace context, it may involve distinguishing between productivity metrics, professional judgment and organizational failure. Contestability becomes meaningful when alternative vocabularies are practically available rather than merely imaginable.

This approach also reframes responsibility by linking accountability to the meanings through which action becomes intelligible. Fischer and Ravizza [54] connect moral responsibility with reason-responsive control, and that account becomes especially important where AI systems mediate practical reasoning. Reasons-responsiveness presupposes that agents can recognize reasons under descriptions that make those reasons intelligible. If the descriptions under which a user acts are externally stabilized and difficult to contest, responsibility may become thinner even when the user remains formally accountable. Responsibility remains present because agents often retain capacities for reflection, resistance and reinterpretation, yet responsibility in AI-mediated environments should be assessed with attention to the semantic organization of the action situation. Institutions that use conversational systems to guide conduct may share responsibility for the vocabularies through which users come to understand their obligations and choices.

Transparency is necessary but insufficient when treated as a disclosure of technical operation alone. A notice that a system is automated, personalized or optimized may leave untouched the way its language organizes practical meaning. Transparency becomes more relevant to semantic sovereignty when it discloses the source, aim and normative loading of descriptions. A system that labels a behavior as non-compliant should make available the institutional rule, alternative classifications, consequences of the label and conditions under which the label may be revised. A system that presents a financial option as aligned with long-term security should make visible the assumptions about risk, value and future need that support the description. Such transparency is semantic and practical, because it concerns how the system converts facts into reasons and reasons into action-guiding language.

Temporal design is also ethically significant because interpretation requires time, especially when descriptions are consequential. Real-time systems often reduce friction by delivering immediately usable meanings, and this can be helpful where users need accessibility support, urgent guidance or cognitive relief. The same temporal efficiency may weaken interpretation when consequential classifications become actionable before agents have room to compare meanings. The relevant design ideal is interpretive slack, understood as sufficient practical room for users to ask what kind of action is being proposed, what vocabulary is being privileged and whether the supplied meaning can be responsibly inhabited. Interpretive slack may be supported through reversible commitments, delayed execution in consequential domains, side-by-side framings, provenance indicators and prompts that invite users to articulate their own description before accepting the system's formulation.

From an HCI and design perspective, interpretive slack can be operationalized through interface features that support semantic participation: side-by-side framings, provenance labels for classifications, prompts asking users to formulate their own description before accepting the system's language, and escalation channels for human redescription. These design features are not merely usability improvements. They preserve the user's capacity to understand how facts have been converted into reasons and how reasons have been converted into action-guiding descriptions.

The broader societal implication is that semantic sovereignty is larger than an individual user skill. Conversational systems are often embedded in organizational and institutional settings where users may have limited authority to challenge categories. A patient may lack standing to contest a medical classification, a worker may fear consequences for resisting productivity language, a student may internalize deficit descriptions and a citizen may experience administrative labels as final. The governance of conversational AI therefore requires attention to the distribution of interpretive power. The central question becomes who supplies the vocabularies through which people understand themselves as patients, workers, learners, consumers and citizens, and how those vocabularies can remain open to negotiation in contexts marked by asymmetries of knowledge, authority and institutional dependence.

8. Objections, Limits and Conditions of Extension

A first objection holds that semantic displacement overstates the novelty of conversational AI because meaning has always been external to the individual. This objection is compelling in light of social externalism, hermeneutics and discourse theory. Public language, inherited concepts, institutions and cultural narratives have always shaped action. The response developed here is that semantic displacement concerns the organization of externality more than externality itself. Public meanings may emerge through practices that allow for reciprocal interpretation, negotiation and revision, or through technical systems that personalize and prioritize descriptions according to criteria that remain difficult for users to inspect. The difference concerns participatory and administered forms of semantic mediation more than a simple opposition between human and machine meaning.

A second objection comes from theories of distributed cognition and cognitive extension. Clark and Chalmers [55] argued that external artifacts can become part of cognitive processes, and everyday examples such as notebooks, maps, translation tools and educational aids show that externalization may enhance instead of diminishing agency. This objection is especially important because the present argument would become implausible if every external semantic support were treated as a threat. The concept of semantic displacement responds by focusing on participatory relation without treating all externalization as equivalent. External supports may strengthen agency when they make meanings more accessible, allow for comparison, support reflection and expand the user's ability to articulate reasons. Displacement becomes more visible where external semantic assistance begins to pre-structure interpretation, especially where the user receives practical meaning as a finished product whose alternatives, provenance and normative assumptions are difficult to access.

A third objection emphasizes user resistance and warns against theories that portray users as passive recipients of system language. People can reinterpret, ignore, mock, challenge or creatively appropriate the language of conversational systems. Any account that treats users as passive recipients would underestimate human inventiveness and the social complexity of interpretation. The present argument accepts that resistance remains possible and often significant across many situated encounters with technology. Semantic displacement describes the weakening of practical conditions for interpretation alongside

the persistence of interpretation itself. It describes an asymmetry in the practical availability of interpretation when some descriptions become easier to inhabit than others. Agents may formally resist a system's vocabulary while lacking the institutional authority, interpretive resources or practical alternatives required to substitute another description. The distinction between formal possibility and practical availability is central to the diagnosis.

A fourth objection returns to machine understanding and questions whether a system without intentionality can displace meaning. If conversational systems lack human understanding, then it may appear misleading to say that they displace meaning. The response is that displacement concerns the human uptake of system-generated language within social practices. A classification can shape action because people and institutions treat it as meaningful, even when the artifact producing or storing that classification has no subjectivity. This point supports caution about inflated claims regarding machine capacities while drawing attention to the social settings in which machine outputs become authoritative. A system may lack full semantic agency while still producing language that reorganizes the semantic environment of human agents. That asymmetry is part of the philosophical importance of the phenomenon, because semantic efficacy can be detached from semantic responsibility.

A fifth objection suggests that the proposed vocabulary is too abstract for practical governance. Terms such as semantic sovereignty and interpretive contestability may appear difficult to operationalize. The concern is serious, because concepts that remain disconnected from institutional practice risk becoming rhetorically attractive but normatively weak. The answer is that conceptual clarification can function as an early step toward operationalization when it generates questions that can be translated into practice. Semantic sovereignty can be translated into design and governance questions: Are alternative descriptions presented to the user, is the provenance of classifications visible, can the user alter the system's framing, are institutional objectives disclosed, can consequential labels be appealed before action proceeds, and are there social spaces for human redescription? These questions only begin to operationalize the concept, yet they show how a philosophical vocabulary can orient empirical research, design evaluation and policy analysis.

The concept can therefore be connected to empirical HCI methods without losing its philosophical role. Researchers could compare alternative chatbot outputs for the same prompt, analyze how users describe their own actions before and after system interaction, examine when system framings become default in workplace or healthcare workflows, and test whether interface features such as alternative framings or provenance indicators increase interpretive contestability. Such studies would not replace the conceptual argument but would extend it into measurable interactional settings.

The limits of the argument should also be made explicit because semantic displacement is best treated as a diagnostic lens whose relevance varies by domain, institutional setting and practical stakes. It may be less relevant in trivial exchanges, creative play, low-stakes drafting or contexts where users knowingly treat system output as provisional material. It may be especially relevant where systems are embedded in consequential institutions, where users depend on them for access to services, where classifications affect rights or duties, or where repeated interaction normalizes a vocabulary of self-understanding. The concept is therefore most useful as a diagnostic lens for identifying conditions under which semantic support becomes semantic substitution. Further empirical work would be needed to examine how users actually appropriate, resist or internalize system-generated descriptions in particular domains.

9. Conclusions

This article has argued that conversational AI should be understood as a socio-technical development that can participate in the organization of action-relevant meaning. The argument begins from the description-dependence of action and from the broader philosophical view that agency requires participation in the meanings under which actions become intelligible, answerable and ownable. Conversational systems matter in this setting because they increasingly supply descriptions, classifications, explanations and normative cues within the ordinary flow of practical life. Their significance may extend beyond information retrieval, recommendation and persuasion, because they may become operational semantic infrastructures that furnish the vocabularies through which users understand situations and themselves.

The concept of semantic displacement was introduced to describe the condition in which externally supplied semantic frameworks become increasingly stabilized and difficult to contest. The concept preserves the insight that meaning has always been social, inherited and mediated, while focusing attention on changes in the organization of semantic externality. It also helps connect philosophical theories of action and language with contemporary concerns about algorithmic governance, platform power, personalization, institutional classification and human–machine interaction. Its contribution lies in making visible a layer of AI mediation that may remain under-described when attention is restricted to choice, bias, transparency, automation or control. In this sense, the argument extends discussions of technological mediation associated with Ihde and Verbeek toward the semantic organization of practical intelligibility in conversational environments.

The normative implications of this analysis center on semantic sovereignty and interpretive contestability. Semantic sovereignty names the capacity to participate in the meanings under which one acts. Interpretive contestability names the practical possibility of exposing, comparing, revising and refusing system-supplied descriptions before they harden into the default grammar of action. These ideals remain compatible with shared vocabularies and inherited practices, because agency always unfolds through social meanings. They require that conversational systems and the institutions deploying them preserve meaningful opportunities for users to understand and contest how facts are transformed into reasons, how reasons are transformed into recommendations and how recommendations are embedded in descriptions of the self, the situation and the future.

Future research may connect this conceptual framework with empirical investigations of how users appropriate, negotiate or resist system-generated descriptions across institutional settings. Further work may also explore design practices that preserve interpretive participation by exposing alternative framings, clarifying semantic provenance and supporting user-led redescription. Such inquiry may help clarify how conversational AI can support agency without reducing practical meaning to externally administered semantic products.

Future research should therefore include empirical work that analyzes actual conversational outputs and user uptake. Possible methods include prompt-output corpora, qualitative interviews, interaction analysis, diary studies, workplace ethnography, HCI usability testing and longitudinal studies of repeated AI-mediated self-description. This would make it possible to examine when users appropriate, ignore, resist or internalize system-generated descriptions, and when semantic support crosses into semantic displacement.

Future governance research may also examine how existing AI governance frameworks address, or fail to address, semantic mediation. Risk management frameworks, human-centered AI guidelines and legal instruments increasingly emphasize transparency, human oversight, accountability and contestability, but they often treat these issues at the level of decisions or system outputs. The concept of semantic displacement suggests that

governance should also examine the framing vocabularies through which users come to understand risks, obligations, eligibility, care, productivity or responsibility.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Anscombe, G.E.M. *Intention*; Basil Blackwell: Oxford, UK, 1957.
2. Davidson, D. Actions, reasons, and causes. *J. Philos.* **1963**, *60*, 685–700. [[CrossRef](#)]
3. Austin, J.L. *How to Do Things with Words*; Oxford University Press: Oxford, UK, 1962.
4. Searle, J.R. *Speech Acts: An Essay in the Philosophy of Language*; Cambridge University Press: Cambridge, UK, 1969.
5. Wittgenstein, L. *Philosophical Investigations*; Blackwell: Oxford, UK, 1953.
6. Gadamer, H.-G. *Truth and Method*; 2nd rev. ed.; Continuum: London, UK, 1989.
7. Taylor, C. *Human Agency and Language: Philosophical Papers 1*; Cambridge University Press: Cambridge, UK, 1985.
8. Ricoeur, P. *Oneself as Another*; University of Chicago Press: Chicago, IL, USA, 1992.
9. Brandom, R. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*; Harvard University Press: Cambridge, MA, USA, 1994.
10. Varela, F.J.; Thompson, E.; Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*; MIT Press: Cambridge, MA, USA, 1991.
11. Hutchins, E. *Cognition in the Wild*; MIT Press: Cambridge, MA, USA, 1995.
12. Suchman, L.A. *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2007.
13. Gallagher, S. *Enactivist Interventions: Rethinking the Mind*; Oxford University Press: Oxford, UK, 2017.
14. Searle, J.R. Minds, brains, and programs. *Behav. Brain Sci.* **1980**, *3*, 417–424. [[CrossRef](#)]
15. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? In *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 610–623.
16. Floridi, L.; Chiriatti, M. GPT-3: Its nature, scope, limits, and consequences. *Minds Mach.* **2020**, *30*, 681–694. [[CrossRef](#)]
17. Frege, G. Sense and reference. *Philos. Rev.* **1948**, *57*, 209–230. [[CrossRef](#)]
18. Sellars, W. Empiricism and the philosophy of mind. *Minn. Stud. Philos. Sci.* **1956**, *1*, 253–329.
19. Goffman, E. *Frame Analysis: An Essay on the Organization of Experience*; Harper and Row: New York, NY, USA, 1974.
20. Hacking, I. Making up people. In *Reconstructing Individualism*; Heller, T.C., Sosna, M., Wellbery, D.E., Eds.; Stanford University Press: Redwood City, CA, USA, 1986; pp. 222–236.
21. Frankfurt, H.G. Freedom of the will and the concept of a person. *J. Philos.* **1971**, *68*, 5–20. [[CrossRef](#)]
22. Bratman, M.E. *Intention, Plans, and Practical Reason*; Harvard University Press: Cambridge, MA, USA, 1987.
23. Christman, J. *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*; Cambridge University Press: Cambridge, UK, 2009.
24. Putnam, H. The meaning of meaning. In *Language, Mind and Knowledge*; Gunderson, K., Ed.; University of Minnesota Press: Minneapolis, MN, USA, 1975; pp. 131–193.
25. Burge, T. Individualism and the mental. *Midwest Stud. Philos.* **1979**, *4*, 73–121. [[CrossRef](#)]
26. Bourdieu, P. *Language and Symbolic Power*; Harvard University Press: Cambridge, MA, USA, 1991.
27. Foucault, M. *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*; Pantheon: New York, NY, USA, 1980.
28. Stiegler, B. *Technics and Time, 1: The Fault of Epimetheus*; Stanford University Press: Redwood City, CA, USA, 1998.
29. Gillespie, T. The relevance of algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*; Gillespie, T., Boczkowski, P.J., Foot, K.A., Eds.; MIT Press: Cambridge, MA, USA, 2014; pp. 167–194.
30. Nissenbaum, H. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*; Stanford University Press: Redwood City, CA, USA, 2010.
31. Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism*; New York University Press: New York, NY, USA, 2018.
32. Benjamin, R. *Race After Technology: Abolitionist Tools for the New Jim Code*; Polity: Cambridge, UK, 2019.
33. Crawford, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*; Yale University Press: New Haven, CT, USA, 2021.

34. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*; PublicAffairs: New York, NY, USA, 2019.
35. Engestrom, Y. *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*; Orienta-Konsultit: Helsinki, Finland, 1987.
36. Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P.N.; Inkpen, K.; et al. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–13.
37. Yang, Q.; Steinfeld, A.; Rose, C.; Zimmerman, J. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–13.
38. Shneiderman, B. *Human-Centered AI*; Oxford University Press: Oxford, UK, 2022.
39. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the opportunities and risks of foundation models. *arXiv* **2021**, arXiv:2108.07258.
40. Weidinger, L.; Uesato, J.; Rauh, M.; Griffin, C.; Huang, P.-S.; Mellor, J.; Glaese, M.; Cheng, M.; Balle, B.; Kasirzadeh, A.; et al. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 214–229.
41. OpenAI; Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; et al. GPT-4 technical report. *arXiv* **2023**, arXiv:2303.08774.
42. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*; UNESCO: Paris, France, 2021.
43. NIST AI 100-1; Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023.
44. NIST AI 600-1; Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. U.S. Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD, USA, 2024.
45. European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**, L 2024/1689, 1–144.
46. Fricker, M. *Epistemic Injustice: Power and the Ethics of Knowing*; Oxford University Press: Oxford, UK, 2007.
47. Ihde, D. *Technology and the Lifeworld: From Garden to Earth*; Indiana University Press: Bloomington, IN, USA, 1990.
48. Verbeek, P.-P. *Moralizing Technology: Understanding and Designing the Morality of Things*; University of Chicago Press: Chicago, IL, USA, 2011.
49. Latour, B. Where are the missing masses? The sociology of a few mundane artefacts. In *Shaping Technology, Building Society: Studies in Sociotechnical Change*; Bijker, W.E., Law, J., Eds.; MIT Press: Cambridge, MA, USA, 1992; pp. 225–258.
50. Winner, L. Do artifacts have politics? *Daedalus* **1980**, 109, 121–136.
51. Butler, J. *The Psychic Life of Power: Theories in Subjection*; Stanford University Press: Redwood City, CA, USA, 1997.
52. Yeung, K. Hypernudge: Big data as a mode of regulation by design. *Inf. Commun. Soc.* **2017**, 20, 118–136.
53. Mittelstadt, B.D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* **2016**, 3, 2053951716679679. [[CrossRef](#)]
54. Fischer, J.M.; Ravizza, M. *Responsibility and Control: A Theory of Moral Responsibility*; Cambridge University Press: Cambridge, UK, 1998.
55. Clark, A.; Chalmers, D. The extended mind. *Analysis* **1998**, 58, 7–19. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.