

Critical Theory of Symbolic Artificial Intelligence

Stefan Schlobach¹

¹Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Abstract

Due to the unprecedented AI-hype, there is increasing interest in AI, also from a sociological perspective. Critical Theory studies societal phenomena, such as AI, in relation to ideologies and power relations. While existing studies mostly focus on Deep Learning-based AI, the power dimension of Symbolic-AI (S-AI) is less in focus. This paper suggests some research questions to stimulate a critical discussion.

We will specifically look at three aspects: how Knowledge is represented and reasoned with, what Knowledge is constructed in applications, and which processes are used for this. With those dimensions in mind, we will discuss Symbolic-AI and ideology, the political economy of (symbolic)AI, and the impact on individuals of such systems. This paper is not a systematic analysis, nor does it claim methodological soundness, but is intended to increase awareness within the S-AI community of our societal responsibilities.

Keywords

Symbolic AI, Critical Theory, Ideology, Power Relation, left-Hegelian Epistemology

1. Why we need a Critical Theory of Symbolic AI

AI is a highly transformative technology that is ubiquitous in our personal lives and society at large. The hype around Deep Learning and Generative AI has led to several interdisciplinary studies of AI as a societal phenomenon, e.g. from an ethical, psychological, and sociological perspective. Recently, it has also been discussed using Critical Theory, which studies societal systems in terms of ideologies and power relations and their impact on individuals (see, e.g. [1]). Some of the arguments and results directly transfer to symbolic AI, others do not. The questions at the heart of KCAP research, such as how to reason with knowledge and how to create (engineer) symbolic formal representations of knowledge¹ suggest the need for a separate critical analysis. Working in the tradition of Tarsky-style semantics [2], it is tempting to claim that S-AI-based technologies are, by construction, domain independent and thus necessarily ideology-free. We might also believe that building an ontology, as an exercise of building domain-specific shared conceptualizations and formalizations [3, 4], maybe even community-driven, will be value-free/ideology-free. I want to challenge this view with this short paper, focusing on Critical Theory for Knowledge Representation, semantic technologies, logics and ontologies.

The main conjecture of this paper is: Symbolic AI conserving existing power relations, acting as a multiplier and accelerator, and thus strengthens capitalism in its current form - with its excesses of totalitarianism and fascism - while claiming an aura of objectivity and futurism. This paper wants to raise awareness that **Science, also of (Symbolic) AI is political, whether we want it or not.**

1.1. Basic Terminology

Critical Theory was started in the interwar period by Horkheimer, Adorno and Marcuse [5] (among others), with the goal of analyzing how modernization, profitability, and automation impact individuals in society. Their *Frankfurter Schule* targeted technology among others and its impact on (economic) politics and power. Since, more multidimensional and disciplinary Critical Theory includes perspectives from cultural and gender studies to post-modernism and post-colonialism.

K-CAP 2025 Companion Volume

✉ k.s.schlobach@vu.nl (S. Schlobach)

🆔 0000-0001-7116-9338 (S. Schlobach)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹I will use the (capitalized) term Knowledge to refer to explicit symbolic knowledge.

This paper will introduce some questions to stimulate discussions. Before, two core concepts need introduction: an **AI assemblage** [1] or apparatus is a *"heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral, and philanthropic propositions – in short, the said as much as the unsaid"* [6]. In other words, we refer to AI not as a technology only, but as the entire ecosystem around it. An **ideology** means that *"certain ideas, connected to power relations in society, become constructed as given truths and objectives, even though many other truths and objectives might be possible"* ([1], p. 63), or as consciousness arising from social, political, and economic relationships, that is often implicit, unquestioned, and seemingly unquestionable, thus hiding the fact that specific truths and objectives are constructed by social relations, rather than epistemologically valid. Assemblages are created by and create explicit views of the world. Identifying hidden assumptions is essential to ensure that our work contributes to diversity and democratization, and not to strengthen oppression.

2. A Critical Theory of Symbolic AI: some Questions

AI has a sufficiently prominent status in our lives that special attention, also from a sociopolitical as well as critical perspective is warranted. Even when AI is considered mostly from an algorithmic, data and methodological perspective, it directly impacts human lives through banking, shopping, entertainment, work, and other social contacts. In this paper, though, I claim that AI as an assemblage is strongly related to societal and political power-relations, even Symbolic AI.

Many of the following questions have been studied with respect to AI in general, specifically for Deep Learning-based approaches with their reliance on (large) datasets (see [1, 7, 8, 9] and others). However, I want to argue that similar questions should be asked about Symbolic and Knowledge-based AI.

2.1. What are the underlying ideologies that created or shaped Symbolic AI?

AI development is closely related to earlier, existing ideologies, such as solutionism, technocracy, innovationism, Californian Ideology and others [10]. Those ideologies, often summarized as techn-utopia, suggest that technology is the key to reaching a future utopian society (often long-term), at the price of human-oriented and more short-term solutions, using almost unlimited resources to push AI technology while exploiting (and shaping) an extreme form of liberal capitalism. Various works have studied the relation of the current AI ideology, which is mostly driven and exploited by few monopolistic companies, with highly ideological super-rich and politically active leaders (e.g. [11], [7] or [1]).

Most symbolic AI formalisms are highly influenced by these ideologies. Most of our research depended on company or military funding and with a predominantly North/Western perspective. Domains are mostly economically selected, resources dominated by powerful companies or organizations, and methodologies 'biased' towards privileged groups, e.g., highly educated professionals. In a recent paper at EKAW 2026 [12], we use Semantic Web technology, a Knowledge Graph and an Ontology, to *unravel implicit power relations in symbolic AI research* with some interesting preliminary results.

But the influence of ideologies is more subtle: the commercialization is ingrained in the world, even the academic world, competitions for scarce resources focus on (financially) hyped topics, often determined by economic, political and ideological interests. This relates to the original work of the earlier Critical Theorists. Horkheimer argues, e.g. in [13] following a notion introduced by Max Weber in [14], that applying rationality in a purely instrumental way ultimately strengthens oppression and existing power relations. I claim that such instrumental rationality (Zweckrationalität) occurs in two different ways in the context of AI: AI systems operationalization instrumental rationality and AI systems are used in purely instrumental ways.

2.2. What are the ideologies created or strengthened by Symbolic AI?

Knowledge is highly dependent on who makes it, imposing world views that inherently represent existing power; instead of polyvocality [15], where in a system different perspectives and "voices" are

reflected in style and content. AI based systems often critically depend on explicit representations of conceptualizations, often as Ontologies or Knowledge Graphs. Every specific modeling choice, both for the conceptualization, the representation, and even the representation formalisms, heavily impacts IT systems and, in this way, our daily lives.

Knowledge (in the broad sense), with its fundamental role in human society, is a result of a social discourse, and thus ideological. Right from the start, symbolic Knowledge played a central role in AI, through formalizing this knowledge to enabling technology and, more widely, to provide ordering principles and abstraction for conceptualizations. Knowledge is rightly considered an immense value and asset for humanity, both in its conceptual form and in its explicitly formalized form. However, it is worth pointing out that this immense power of knowledge comes with a huge potential for misuse, often unconsciously, and highly impacted by the ideology underlying its creation.

FOL-based formalisms, such as Relational Databases and Description Logic KBs, e.g., force conceptualizations into a Boolean paradigm, which often does not represent under-privileged parts of society with more complex identities. Selection of application domains is influenced by economic and political power, and crowd-based Knowledge, such as DBpedia or Wikidata, under-represents minorities and contains human bias. And even Knowledge Engineering methods focus on reusable Knowledge patterns, thus “favoring” applications of known types and thus disadvantaging more atypical solutions, such as non-profit collaborative work [16].

Following the arguments of the first generation of the Frankfurter Schule (e.g. [17], [18], [13]), in a recent positions paper [19], we analyzed the prevalent methodologies and epistemology of the disciplines of Knowledge Representation and Engineering. We argue that our disciplines’ attempt to categorize reality systematically leads to a distortion of the social world, in a manner contoured by socio-structural power relations. This happens in two ways, first by reducing most AI tasks to tasks depending on categorization and secondly, by conceptualizing the world in static ontologies and datasets, which distort and entrench an asocial and undemocratic reality, as well as result in systematized biases the systems based on such conceptualizations. Such anti-positivist insights, and the related critique of instrumental rationality, are also developed in the work of Andrew Feenberg [20], who shows how positivist epistemologies serve to distort reality.

This way, AI-based support (including symbolic AI-based) helps to conserve, and often even reinforce existing power relations, acting as a multiplier and accelerator, and thus preserving capitalism in its current form, while claiming an aura of objectivity and futurism.

2.3. Does Symbolic AI negatively impact the political economy?

AI, also symbolic AI, stimulates the platform economy[21] (such as Uber or Amazon), and the recent reliance on automation for Knowledge Construction not only contributes to high energy consumption, but to precarious working conditions by outsourcing digital labor².

Other tangible examples happen when human behavior has to be adapted to suit Computer Systems. Take, e.g., job applicants, who have to optimize their CV according to quantitative, measurable qualities to pass AI filters³. The previous example points to another danger of the technology, the reduction of diversity and human features towards quantifiable and instrumental qualities.

Finally, a frequent manifestation related to the ideology of solutionism is that (symbolic) AI contributes uncritically to the suggestion to solve societal problems by increasing investments into technology, rather than by addressing the underlying sociological problems.

The most impact-full contribution of symbolic AI to the political economy is probably through its epistemology: its focus on abstraction and categorization of assumed identities, which cement current ways of thinking, which are often undemocratic and steeped in inequalities and social injustice.

In [19], we argue for a novel methodology to symbolic AI based on left-Hegelian dialectical thinking. The petrification of the world into classes, tables, triples, graphs and formulas does not reflect the complex nature of a contradictory reality in flux. While our KR and KE methods remain descriptive,

²<https://time.com/6247678/openai-chatgpt-kenya-workers/>

³<https://www.eurofirms.com/es/en/how-to-create-a-resume-that-will-pass-the-ai-ats-filter/> (consulted 21/7/2026).

assuming (and reinforcing) existing realities, our conceptualizations should be ontological and two-dimensional, i.e. modeling not just what *is*, but what *ought to be* [17].

2.4. Does the application of AI across societies reduce humans to subjects of the assemblage and contribute to their alienation?

Given the prominent role of AI in the societal discourse, there is a direct effect of conceptualization on individuals. Most directly, this is felt by people and organizations outside the main-stream. The predominantly binary view of society and technology of S-AI formalisms has, e.g., a significant impact on gender fluid people, but also organizations and companies that differ from the norm. This not only might violate their identity, it could suggest exclusion from societal consensus and often has direct economic consequences. Even the choice of formalism impacts power-relations. In a society heavily influenced by logic-based rhetoric and discourse, the rational positivist paradigm impacts its subjects.

Traditional themes of early Critical Theory, based on Marx (e.g. [22]) and Lukács [23], are alienation, reification, and commodification. Arguably social networks, the World-Wide Web, and novel AI technology contribute to those phenomena. *Alienation* happens when people feel disconnected from their products, work, communities and previous relations. Through the AI supported atomization of processes, workers become unconscious elements of automation and get distanced from the product of their work and the work itself. As creative and social beings, Human therefore also get distanced from their own identity as humans, and get reduced to objects of a production system. *Reification* turns human social relations into objects and their characteristics. For social media and AI, humans are valuable data provider, while their needs become dominated by technological needs of big tech companies. *Commodification* means that humans and their actions and relations are merely seen as economic commodities. Marcuse [17] offers a similar critique with respect to dehumanization (one-dimensionality) in advanced industrialized societies. In future work we will focus on the role of symbolic AI in this respect.

2.5. Symbolic AI enforces "instrumental rationality". Does this paves the way to totalitarianism?

Based on Weber's work [14], the first generation of the Frankfurter Schule, such as Horkheimer and Adorno in "Dialectic der Aufklärung" [18] or Marcuse in "One-dimensional Man" [17] critiqued the effect of *instrumental rationality*, as well as technocracy and industrialization, on society. Automation and formalization of reasoning through mathematics and technology push societal truths into the realm of metaphysics, thus removing normative, societal and ethical rationality from the scientific discourse. Mühlhoff [11] and McQuillan [7], among others, analyzes the close connection between the AI-centered tech industry with anti-democratic and totalitarian ideologies.

Horkheimer and Adorno were often criticized for their dystopian and bleak view towards the future. Their work has been continued and modernized over the past decades and inspired many philosophers of Science and Technology to pursue a more positive type of critical theory for technology, e.g. [24, 20], studying conditions for which technology could be used critically, and as vehicles to improve democracy, equality and social values. Other work (e.g. [1]) have given great overviews over those new critical theories of AI, related to ethical AI, bias studies, feminist critique or postcolonial theory, to name but a few. Given the current societal, economic and political developments in the global context, though, a revisit of the original skepticism of the almost 100 year old original literature might be warranted.

3. Conclusion

This paper is an attempt to stimulate discussions on the sociological, political, and philosophical context of the symbolic aspect of the AI assemblage. The intention is not to judge, but to provide tools to decompose S-AI and the ideology it creates and operates in. Only if we understand the power relations

that our own research helps creating and the negative societal effects it might have, will we be able to use our expertise to overcome injustice and help true democratization.

Social science has an impressive history of untangling complex societal problems and a critical analysis of the socio-political context of AI is crucial. It is time that we, as AI researchers, realize that **Artificial Intelligence, including symbolic AI, is political**, and that we have to act upon this insight.

References

- [1] S. Lindgren, *Critical Theory of AI*, Polity Press, 2023.
- [2] H. Field, Tarski's theory of truth, *The Journal of Philosophy* 69 (1972) 347–375.
- [3] T. R. Gruber, A translation approach to portable ontology specifications, *Knowledge acquisition* 5 (1993) 199–220.
- [4] N. Guarino, D. Oberle, S. Staab, What is an ontology?, *Handbook on ontologies* (2009) 1–17.
- [5] T. Lijster, *Frankfurter Schule*, number 82 in *Elementaire Deeltjes*, Athenaeum - Polak Van Gennep, 2023.
- [6] V. Larroche, *The dispositif: A concept for information and communication sciences*, John Wiley & Sons, 2019.
- [7] D. McQuillan, *Resisting AI: An anti-fascist approach to artificial intelligence*, Policy Press, 2022.
- [8] M. Coeckelbergh, *The political philosophy of ai*, *Critique of Thought* (2023).
- [9] E. M. Bender, A. Hanna, *The AI con: How to fight big tech's hype and create the future we want*, Random House, 2025.
- [10] T. Gebru, P. Torres, *The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence*, *First Monday* 29 (2024). doi:10.5210/fm.v29i4.13636.
- [11] R. Muehlhoff, *Künstliche Intelligenz und der neue Faschismus*, Reclam, 2025.
- [12] A. Wróbel, E. Meregalli, S. Schlobach, How to unravel implicit power relations in symbolic ai research, in: *Proceedings of EKAW 2026*, Springer, 2026.
- [13] M. Horkheimer, *Eclipse of Reason*, OUP, 1947.
- [14] M. Weber, *Economy and Society: An Outline of Interpretive Sociology*, University of California Press, Berkeley, CA, 1978. Two volumes.
- [15] M. van Erp, V. de Boer, A polyvocal and contextualised semantic web, in: *The Semantic Web*, Springer International Publishing, Cham, 2021, pp. 506–512.
- [16] A. Volda, E. Harmon, B. Al-Ani, Homebrew databases: complexities of everyday information management in nonprofit organizations, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 2011, p. 915–924. doi:10.1145/1978942.1979078.
- [17] H. Marcuse, *One Dimensional Man; Studies in the Ideology of Advanced Industrial Society*, 1964.
- [18] T. Adorno, M. Horkheimer, *Dialektik der Aufklärung*, 1946.
- [19] N. Harris, S. Schlobach, P. Hitzler, *Dialectics of knowledge*, in: *Proceedings of EKAW 2026*, Springer, 2026.
- [20] A. Feenberg, *Critical Theory of Technology*, 1991.
- [21] N. van Doorn, Platform capitalism's social contract., *Internet Policy Review*, 11(1), 2022 (????). doi:<https://doi.org/10.14763/2022.1.1625>.
- [22] K. Marx, *Economic and Philosophic Manuscripts of 1844*, Dover Publication, 2007. Written in 1844.
- [23] G. Lukács, *History and Class Consciousness: Studies in Marxist Dialectics*, MIT Press, 1971.
- [24] G. Delanty, N. Harris, *Critical theory and the question of technology: The Frankfurt School revisited.*, *Thesis eleven* 1 (2021) 88–108.

Declaration on Generative AI

The authors has not employed any Generative AI tools.