**VILNIUS TECH**
Vilnius Gediminas
Technical University

# A PHENOMENOLOGICAL REPLY TO GARY MARCUS AND ERNEST DAVIS'S *REBOOTING AI: BUILDING ARTIFICIAL INTELLIGENCE WE CAN TRUST*: CONTRIBUTING TO THE CREATION OF GENERAL ARTIFICIAL INTELLIGENCE

Joaquin TRUJILLO [ID] *

*Independent researcher, United States*

**Abstract.** Marcus and Davis's *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019) describes the narrowness of current artificial intelligence, posits constituents of human intelligence it contends artificial intelligence (AI) must reproduce to achieve a general (broad) formulation, and envisages the contributions of general AI to human living. This article (1) reviews Marcus and Davis's argument, (2) phenomenologically evaluates their interpretation of human intelligence, and (3) discloses the moral inconsistencies implied by their vision of the relation between general AI and human being. Transcendental and hermeneutic phenomenological perspectives are employed to address the range of phenomena the authors affiliate with human intelligence.

**Keywords:** artificial intelligence, common sense, hermeneutics, human intelligence, phenomenology.

## Introduction

Marcus and Davis's *Rebooting AI: Building Artificial Intelligence We Can Trust* describes the narrowness of current artificial intelligence (AI), posits constituents of human intelligence it contends AI must reproduce to achieve a general (broad) formulation, and envisages the contributions of general AI to human living. This article (1) reviews Marcus and Davis's argument, (2) phenomenologically evaluates its interpretation of human intelligence, and (3) discloses the moral inconsistencies implied by its vision of the relation between general AI and human being. Transcendental and hermeneutic phenomenological perspectives are employed to address the range of phenomena Marcus and Davis affiliate with human intelligence.

Marcus and Davis begin by describing the "gap" between the rhetoric and reality of AI. They explain that although AI is commonly extolled for its professed ability to transform individual life, economy, and society, the technology has "been long on promise, short on delivery" (Marcus & Davis, 2019, p. 4). Its "central problem", according to Marcus and Davis,

*Corresponding author. E-mail: *jtrujillo1@me.com*

is that it is "narrow" (2019, p. 13). Current AI is "no match" for ability of the human mind to function "in a complex and ever-changing world", in a world, (the world) that is "open-ended" and suffuse with ambiguities, complexities, and contingencies (Marcus & Davis, 2019, pp. 9, 11, 110–111). Reduced: current AI lacks a robust and reliable capacity to contend with (mitigate, resolve, navigate) variance. The technology "works for particular tasks it is programmed for" as long as the data it processes are not too "novel" (Marcus & Davis, 2019, pp. 13, 16).

Marcus and Davis source current AI's limitations to its two principal methods: classical AI and machine learning. Neither approach is "able to adapt flexibly to a world that is fundamentally open-ended", and "flexibility", Marcus and Davis state, "is what general intelligence, of the sort any ordinary person has, is all about" (2019, p. 16). Classical AI, which comprises hand-coded, knowledge-based, or purposefully-designed programs, is constrained by the fantastic amount of labor its development requires and the inability of programmers to foresee the contingencies it will encounter. For these reasons most classical systems never move beyond the laboratory prototype stage (Marcus & Davis, 2019, pp. 41–42). Machine learning, which includes "deep" and "end-to-end" learning, is constrained by its dependence on "massive amounts of data", univocal reliance on correlation and probability (compiling "a collection of statistically plausible matches"), and zero-heuristic power (Marcus & Davis, 2019, pp. 56–58, 83–85, 89). Machine learning works well correlating and categorizing information, as seen in World Wide Web search, recommendation engines, and visual and voice recognition, but fails catastrophically when confronted with anything that requires a modicum of interpretation (Marcus & Davis, 2019, pp. 76, 87).

According to Marcus and Davis, taking AI to the next level, transforming it from narrow to general intelligence, demands "an immense amount of foundational progress" (2019, p. 4). It requires machines be "equipped from the factory" with the capacity to cope with an "open-ended world" (Marcus & Davis, 2019, p. 206). It further requires the AI industry depart from its "blank slate" obsession programming AI to "learn everything from scratch" through data correlation and classification rather than investing the technology with "knowledge" (Marcus & Davis, 2019, pp. 25, 75). It mandates AI include the capacity "to deal not only with specific situations for which there is an enormous amount of cheaply obtained relevant data", as witnessed in deep and end-to-end learning systems, "but also problems that are novel, and variations that have not been seen before" (Marcus & Davis, 2019, p. 16). It obliges AI be programmed with "rich cognitive models" that contain the capacity to evaluate situations, predict events, and dynamically decide situationally appropriate actions (Marcus & Davis, 2019, p. 113).

Marcus and Davis correlate general AI to the reproduction of two phenomena it affiliates with human intelligence: "common sense" and "deep understanding". Common sense, as Marcus and Davis define it, is: (a) "a rich understanding of the world, and how it works, and what can and cannot plausibly happen in various circumstances" (2019, p. 115), (b) "knowledge that is commonly held" (2019, p. 149), and (c) the "stuff that everybody knows, yet nobody seems to know what exactly it is or how to build machines that have it" (2019, p. 150). It includes a "core understanding of physical objects", a "firm sense of time, space, and causality" (Marcus & Davis, 2019, p. 146), and an understanding of taxonomy (Marcus & Davis, 2019, p. 157) and semantics (Marcus & Davis, 2019, p. 159).

Marcus and Davis define deep understanding as the ability to "reason about the complex interplay of entities that causally relate to one another in an ever-changing world" (2019, p. 66). The phenomenon, as they describe it, also comprises an understanding of logic, motivation, analogy, relation, abstraction, inference, and language (words and compositionality) (Marcus & Davis, 2019, pp. 64, 72, 88–89, 94, 127, 146, 174, 178), as well as perception (Marcus & Davis, 2019, p. 179) and situational awareness, or the capacity to know what "objects collectively mean" (Marcus & Davis, 2019, pp. 107–108). Deep understanding, according to Marcus and Davis, also subsumes self-awareness inasmuch as it can locate and track itself, objects, and persons in space and time (2019, pp. 147, 163–164), distinguishes between itself and the objects and persons it references and encounters (2019, pp. 72, 147, 164, 201), understands its actions and their consequences (2019, pp. 109, 169, 197–198), and learns and "adapt flexibly" (2019, pp. 16, 201).

Marcus and Davis describe the deep understanding of general AI as an emergent process that would ensue from the interaction among common sense and other "cognitive models" and "powerful tools for reasoning" in machines (2019, p. 199). "Together", these systems, they assert, common sense being the principal one, would "lead to deep understanding" and "AI that genuinely understands the world" (Marcus & Davis, 2019, p. 199). Marcus and Davis's prescription to invest machines with these processes, to endue them with "common sense, and ultimately general intelligence", which, in the final analysis, it equates with deep understanding, is to:

> "start by developing systems that can represent the core frameworks of human knowledge; time, space, causality, basic knowledge of physical objects and their interactions, basic knowledge of humans and *their* interactions. Embed these in an architecture that can be freely extended to every kind of knowledge, keeping always in mind that central tenets of abstraction, compositionality, and tracking of individuals. Develop powerful reasoning techniques that can deal with knowledge that is complex, uncertain, and incomplete and that can freely work both top-down and bottom-up. Connect these to perception, manipulation, and language 178. Use these to build rich cognitive models of the world. Then finally the keystone: construct a kind of human-inspired learning system that uses all the knowledge and cognitive abilities that the AI has; that incorporates what it learns into its prior knowledge; and that, like a child, voraciously learns from every possible source of information: interacting with the world, interacting with people, reading, watching videos, even being explicitly taught. Put all that together, and that's how you get to deep understanding" (2019, p. 179).

Marcus and Davis argue that to contribute meaningfully to the production and quality of human living, AI must be trustable. It must protect against harm and do no harm. Fulfilling this mandate, Marcus and Davis explain, requires AI be programmed with a respect for a "well-structured set of core ethical values", which should be reflected in the persons and enterprises that design and operate the technology as well as in the "social structures and incentives" belonging to their situations (2019, p. 196). They also mean prescribing to Isaac Asimov's (1977) Three Laws of Robotics, which read, a robot: (1) "may not injure a human being, or, through inaction, allow a human being to come to harm, (2) "must obey the orders given it by human beings except where such orders would conflict with the First Law", and (3) "must protect its own existence as long as such protection does not conflict with the

First or Second Laws" (Marcus & Davis, 2019, pp. 193–194). AI systems capable of doing "significant harm", Marcus and Davis state, should

> "understand the world in a deep enough fashion to be able to understand the consequences of their actions, and to factor human well-being into the decisions they make" (2019, pp. 197–198).

Marcus and Davis assert that AI (especially AI robots) programmed with common sense, deep understanding, and values would catapult humanity into a world divorced of toil, dominated by leisure, and defined by boundless opportunities for human thinking, invention, creativity, and discovery. The world Marcus and Davis envisage is one where "domestic robots" are sufficiently "practical" and "trustworthy" to assume the responsibilities of "cooking, cleaning, tidying, buying groceries", expert robots perform "medical diagnosis, digest legal cases and documents, [and] teach complex subjects",

> "digital assistants <…> will be able to do pretty much anything human assistants can do, and they will become democratized, available to all rather than just the wealthy" (2019, pp. 202–203),

and:

> "computer programming might finally be automated too, and the power of any one individual to do something new, like build a business or invent an art form, will be vastly greater than it is now. The construction industry will change too, as robots start to be able to do the skilled work of carpenters and electricians; the time required to build a new house will be reduced, and the cost will decrease, too. Nearly anything that is dirty and dangerous, even if it requires expertise, will become automated. Rescue robots and robotic firefighters will be widespread, delivered from the factory with skills ranging from CPR to underwater rescue" (2019, p. 203).

Moreover, AI would not only supplement individual creativity and discovery, according to Marcus and Davis, which state:

> "Computers that could read as well as PhD students but with the raw computational horsepower of Google would revolutionize science, too. We would expect advances in every field, from mathematics to climate science to material science. And it's not just science that would be transformed. Historians and biographers could instantly find out everything that has been written about an obscure, place, or event. Writers could automatically check for plot inconsistencies, logical gaps, and anachronisms" (2019, p. 68).

AI will be there to assist with every calculation you require; robots will do most of the labor. Individuals in every field will be able to do things they never could have imagined; each one will be able to serve as the creative director for a whole team of robot helpers (people will also have more free time, with AI and robots doing much of the tedious work of everyday life) (Marcus & Davis, 2019, p. 204).

## 1. General observations

Marcus and Davis demystify current AI. It clearly explains the technology and emphatically exposes its heuristic limitations. It also compels one to think the meaning of human intelligence and its relation to the total human person. Its formula to endue machines with

common sense and deep understanding, however, reads more like a "wish list" than it does, as Marcus and Davis propose, a heliograph for progress. The report consists of chapter-by-chapter reviews of the strengths and weaknesses of current AI systems, ends most sections with a leap into descriptions of the human phenomena it contends AI must replicate to achieve a general formulation, and culminates in predictions of "creative applications" of current systems and the invention of new ones to effect the transformation (Marcus & Davis, 2019, pp. 65–66): "getting to broad intelligence", Marcus and Davis state, "will require us to bring together many different tools, some old, some new, in ways we have yet to discover" (2019, p. 128). This back and forth explanation of what AI does and fails to do compared to what persons do followed by milestones that must be met to resolve the difference is thought-provoking, but fails to evoke confidence. It does not yield a vision of a concrete trajectory that would otherwise inspire readers to get the invention of general AI underway.

Marcus and Davis belong to the minority of AI reports that labor to discern the basic attributes of human intelligence to inform the technology's development. Other noteworthy exceptions include Legg and Hutter's (2006, 2007) reports and Davis and Marcus' (2015) *Communications of the Association for Computing Machinery* article, which recommends AI developers replicate common-sense reasoning and common-sense knowledge in machines to make progress. Marcus and Davis represent a continuation and expansion of the thinking the authors started in the Association for Computing Machinery, United States. Their effort to render the phenomenon of intelligence also exceeds (in range, richness, and depth) the attempt Mitchell (2019) tries in *Artificial Intelligence: A Guide for Thinking Humans*, which also debunks current AI.

Marcus and Davis do not define human intelligence, however. They posit variables it affiliates with the phenomenon, but do not reduce it to a fundamental process or principle. They cite, but neither affirm, negate, nor evaluate, the definition posited by the cognitive psychologist, Pinker (2018), who correlates of human intelligence to "the ability to deploy novel means to attain a goal", and add: "But the goals are extraneous to intelligence: Being smart is not the same as wanting something" (Marcus & Davis, 2019, p. 30). By supplanting the provision of a definition of intelligence with descriptions of its characteristics, Marcus and Davis deny themselves a first-principles perspective that could serve as interpretive (and falsifiable) mooring through/toward which to think (project-open) its matter. Marcus and Davis's premise that "any theory that proposes to reduce intelligence to a single principle – or a single 'master algorithm' – is bound to be barking up the wrong tree" (2019, p. 119) or their offer of a "rough sense about where the field ought to be going" (2019, p. 155) have no epistemic basis. They imply a flight from parsimony, forestall a reduction of human intelligence to its ownmost (*Wesen*) – to the way it foundationally shows itself to be from itself – and increase the susceptibility to drift and errancy.

Marcus and Davis's decision to forfeit a definition of human intelligence in lieu of descriptions of attributes antedates its neglect of the fundamental question: is general AI possible? One-liners, such as, "We don't think this is impossible, and later we sketch some of the steps that would need to be taken" (Marcus & Davis, 2019, p. 74), where "steps" means posited features of human intelligence AI must replicate to transition to a general formulation, or, "It's a tall order, but it's what has to be done" (Marcus & Davis, 2019, p. 179), do not substitute

a rigorous response to the question of general AI. It signifies an abandonment of Marcus and Davis's professed skepticism (2019, p. 9), diminishes the report's empirical power, and betrays its understanding of human being (subjectivity). The failure to investigate the question of general AI foreshadows the "overattribution error", which Marcus and Davis define as the mistaken belief AI "has human-like intelligence" (2019, p. 32), and the rhetoric surrounding the technology. It connotes an anthropomorphization of AI correlated to a failure to think human intelligence within the context of human experiencing and being. It mindlessly credits machines with the possibility of reproducing a radically human phenomenon, and automatically endows them the prospect of being, consciousness, and language, of perceiving, experiencing, and understanding, of knowing and thinking beyond calculation, categorization, and correlation, of discovering, inventing, and creating. It risks confusing efforts to distinguish between the reproduction and simulation of human intelligence in machines, and, if either proves unfeasible, determining next best steps. It implies a technicity-based perspective, a mindless belief in the ability of technology to endlessly grow, accomplish whatever imaginable, even if improbable, and engineer solutions to all of life's questions and challenges.

Neglecting the question of general AI also preempts a discernment of the fundamental properties of human being that distinguish it from machines. It signifies a mechanistic interpretation of subjectivity that not only mirrors the anthropomorphization of AI, but, like it, succeeds a forgottenness of the human "to be". It suggests a dualistic interpretation of reality that individuates human being as a discrete (objectified) subject in a discrete world populated by discrete subjects and objects. It overlooks human being (subjectivity) as it is factically (primitively) endured: an event (situation) it immanently comprehends and is immanently shared (intersubjective). Neglecting the question whether human intelligence can be reproduced in machines implies a dehumanization of human being. It encapsulates human intelligence within cognitive models and heuristic tools, and is removed from the way the phenomenon of intelligence is openly and straightforwardly experienced.

## 2. Phenomenology

Marcus and Davis source their understanding of human intelligence to the cognitive sciences, psychology, linguistics, and philosophy (2019, p. 118). It makes no mention of phenomenology. The omission is not uncommon. The AI literature is largely devoid of references to phenomenology. Factors contributing to this lacuna include the phenomenology's lexicon and its success alienating itself from the physical and human sciences. The language phenomenology has developed to accomplish its purpose, a thesaurus (the language of human experiencing and being) whose evolution is continuously underway and epitomizes its endeavor (struggle) to expose human reality as it is factically endured, does not lend itself to the quick assimilation of observations and theses researchers typically exact from fields other than their own. Also, phenomenology has in many ways marginalized itself from empirical research generally by concentrating too much on methodological reflections (phenomenology of phenomenology) rather than responding to its inner charge: the exposition of human phenomena.

What makes the absence of references to phenomenology in Marcus and Davis remarkable, aside from the method's relevance to its matter, is its phenomenological spirit, inasmuch

as it looks "inward", as Marcus and Davis state, "toward the structure of our own minds" to learn how to design "truly intelligent machines" (2019, p. 25). But its rendition of human intelligence is more naturalistic than it is explicative. It lacks rigor and a clear thematic trajectory. It mixes too much, can lack clarity, and is directionally challenged. It is removed from a foundational understanding of the phenomenon of intelligence, conflates essential attributes with derivative ones, and risks subverting the articulation of parameters to guide AI's development.

So, what is phenomenology? The query is posed in light of the transdisciplinary nature of this report and the goal to maximize access to its reflections. Maurice Merleau-Ponty's "Preface" to the *Phenomenology of Perception* (1962, vii–xxi) begins with the same question and, despite its lack of being-historical considerations, remains a good primer. Phenomenology is a philosophically situated method epitomized by its mantra: "to the things themselves". It is a way of thinking (questioning) the phenomena comprised by human reality. It is considered the most significant hermeneutical movement of the 20th century and hermeneutics as it is fundamentally understood, which is the open, resolute interrogation (thinking, unearthing, disclosing-saying) of human experience and being (as well as being as such, or be-ing); incidentally, Marcus and Davis leap over the phenomenological movement to cite 18th century Immanuel Kant (Marcus & Davis, 2019, pp. 145, 162)! Phenomena the method commonly investigates include consciousness, transcendence, thinking, intelligence, common sense, perception, intersubjectivity, temporality, spatiality, physicality, work, meaning, death, meaningfulness, language, and be-ing (enowning). There is no "one" phenomenology, however. Phenomenology is a movement (way of thinking) comprising distinct (but not discrete) perspectives (Spiegelberg, 1994). This report commensurately employs the most radical (foundational) ones, transcendental and hermeneutic phenomenology, to address the range of human phenomena Marcus and Davis affiliate with human intelligence.

Transcendental phenomenology, the phenomenology of Edmund Husserl, particularly as expressed in his *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy* (1980), *Cartesian Mediations: An Introduction to Phenomenology* (1982), and *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy* (1970), is the rigorous exhibition (interpretation) of consciousness (intentionality, experiencing) (as such), or the life-world (*Lebenswelt*). Consciousness (the total human person, subjectivity) is transcendence. Its rendition as transcendence indicates nothing metaphysical. It denotes consciousness as the meaning of the phenomenon it intends and constitutes in experiencing, as the indivisible correlation between νόησις (*nóēsis*) (the experiencing of a matter) and νόημα (*nóēma*) (the matter experienced). The principal technique transcendental phenomenology employs to accomplish itself as a "rigorous science" is the reduction. The reduction is the resolute effort (struggle) to suspend one's preconceptions and predications to augment the ability to see the self-givenness of phenomena. Alfred Schutz's/Schuetz constitutive phenomenology of the natural attitude, which amounts to an investigation of the phenomenon of common sense (Gurwitsch, 1970; Natanson, 1982; Trujillo, 2021), is based on Husserl's thinking (van Breda, 1982; Zaner, 1961).

Hermeneutic-phenomenologically, specifically its transcendental-horizonal iteration, which is the phenomenology Heidegger (2001) expresses (develops) in *Being and Time* (Emad, 2000), bypasses (but does not ignore) the question of consciousness to provide an

extended rendition of subjectivity. It is the rigorous exhibition of human being (the human person as such) discerned as *Da-sein* (t/here-being), an event, situation, or happening, also, transcendence, beings-in-the whole (*das Seiende im Ganzen*), that unfolds (comes to pass) as the concern for and comprehension of its transcendence, or ἔκ-στασις (*ék-stasis*). "Transcendence", used in this context, also means nothing metaphysical. It signifies human being as the being-of-t/here, its defining prerogative as the comprehension of being (hence, also meaning), and its ownmost as the concern for "to be" (*Sorge*). Hermeneutic phenomenology discloses *Da-sein* as a process that transcends beings to their being and comes to pass as the understanding of and preoccupation with the way they are; it is the disclosing-showing-saying of that way. Its structure is temporal. It singularly unfolds as the moments of "having been", "is", and "will be", and is grounded in language.

The human potentialities to comprehend being and wield language, within the context that *Da-sein* is both its comprehension of being and language, constitute the pre-philosophical (pre-Socratic) λόγος (*lógos*); phenomenology regularly cites original Greek in light of the limitations of contemporary language (particularly, its inclination to objectify subjectivity) and to leverage its extraordinary disclosing-saying power. The originary inclusion of these foundational constituents (endemic ways) of *Da-sein* within a single phenomenon speaks to their factical unity. Language is contingent on the comprehension of being and the comprehension of being is effected through language. The potentiality (power and possibility) to comprehend being, including the being-of-t/here, or transcendence, enables language and steers its "polymorphic, living, dynamic" articulation of meaning (Kovacs, 2016, p. 252), and language allows for the comprehension of being. The comprehension of being and language indissolubly belong to each other, as indicated by the phenomenological theses: "Language is the house of being. In its home man dwells" (Heidegger, 2008a, p. 217); "Human being is language" (Kovacs, 2013, p. 219), and "We – human beings – are a conversation" (Heidegger, 2000, pp. 55–56).

Putting aside transcendental and hermeneutic phenomenology's distinct foci, one, human experiencing (consciousness), and the other, human being (*Da-sein*), both rendered as transcendence, as well as transcendental phenomenology's attempts to suspend the meaning of "is" through a professed reduction, sufficient parallels remain between the perspectives to allow for their collaborative application. For Merleau-Ponty, the hermeneutic phenomenology of *Da-sein* represents an "explicit account" of the life-world, "which Husserl, towards the end of his life, identified as the central theme of his phenomenology" (1962, vii). And, as George Kovacs correctly points out, hermeneutic phenomenology embodies a rethinking (radicalization) rather than a disavowal of transcendental phenomenology. It is "neither apologetic nor triumphalistic", but is, rather, an ongoing "confrontation" with the "thought" and "unthought" in Husserl's thinking as well as an expression of "genuine indebtedness" to it (Kovacs, 2016, pp. 245–246).

## 3. Human intelligence

Marcus and Davis equate human intelligence with deep understanding, which they describe as an open-ended knowing and interpreting process that comprehends itself and the world. Deep understanding includes situational awareness, self-awareness, logic, and an

understanding of relation, inference, physicality, temporality, spatiality, and causality. It can recognize the consequences of it actions, contains the ability to reason and choose, and understands language. It openly learns and flexibly adapts. Marcus and Davis contend that AI must reproduce these variables (and others) together to transition to a general formulation.

Human intelligence rendered phenomenologically, specifically, hermeneutic phenomenologically, does not correspond to Marcus and Davis's notion of deep understanding. Human intelligence (the phenomenon of intelligence), thought phenomenologically (hermeneutically), is the pre-philosophical, ποίησις (*poiēsis*): the power to heed and bring forth to completion a possibility that is sheltered t/here, within transcendence, and calls for, invites, or summons to be brought forth to its fulfillment (Heidegger, 2008b). Its principal moments, hence, includes heedfulness (attending, listening to a call), thinking (resolutely yielding comprehending and language to its matter), resolve (striving, struggling, laboring), and solicitude (care for the matter that beckons be brought forth and fulfilled). Michelangelo's assertion that "every block of stone has a statue inside it and it is the task of the sculptor to discover it" epitomizes the understanding of ποίησις.

Ποίησις includes τέχνη (*tékhnē*) (Heidegger, 2008b), which signifies the same discovery process, but is aimed at a possibility that is more "at hand" than it is hidden and emanates more as a challenge to put in order than it does a call to bring forth. Albert Einstein's special theory of relativity, for example, resonates more as τέχνη than it does as ποίησις inasmuch as it responded to and solved apparent contradictions suggested by different reference frames observing light's constant speed (Stannard, 2008; Zeh, 2007). The brilliant discovery that it was spacetime that varied between the reference frames was substantially an outcome of ποίησις, however.

The strengths and weaknesses of Pinker's (2018) definition of intelligence, as quoted by Marcus and Davis, become apparent when read against the phenomenological notions of ποίησις (and τέχνη). The ownmost of human intelligence (the phenomenon of intelligence), as it directly and primordially shows itself, as it is factically endured, is not "the ability to deploy novel means to attain a goal", as Pinker (2018) writes (Marcus & Davis, 2019, p. 30). It is the power to heed, respond to, and bring forth to its fulfillment a possibility sheltered within transcendence that calls to be brought forth and fulfilled: to attune to it, think and care for it, solve, craft, or create it, invent it. Although invariably an expression of individuality, the process is not encapsulated within the individual, nor, as Pinker (2018) implies, does it equate to imagination or creativity ("novel means"), although it often includes them. The end term of intelligence, its culminating moment, point of repose, or τέλος (*telos*), also is not an object removed from transcendence or relative, as implied by "goal". It is immanent to the life-world, to *Da-sein*. The process is pulled (and not pushed) by a factical "otherness", by a prospect transcendent to or beyond oneself that originates t/here, in the world. Its directionality, hence, opposes tendencies toward self-orientation and self-absorption, although it can, and typically does, engender personal meaningfulness, as the phenomenological reflections of Frankl (2014) emphatically reveal. The end term of intelligence, its τέλος, or, also, ἐντέλέχεια (*entelécheia*), as well as the resolution to accomplish it and the way one goes about accomplishing it, is a key measure of its quality and power. It is not, as might be inferred from Pinker (2018), "extraneous" to the phenomenon, although he is correct to point out

that being intelligent "is not the same as wanting something" (Marcus & Davis, 2019, p. 30). Creatively evading personal responsibility, aggrandizing oneself, or abusing others for personal gratification, for example, may involve novel or creative means, but the aims are not intelligent. They are destructive, self-defeating, and disaffected of basic human truths and values. Human intelligence does not destroy life. It protects and generates it, and contributes to the understanding, quality, and meaningfulness of human living.

Marcus and Davis's notion of deep understanding is akin to the hermeneutic phenomenological interpretation of λόγος. It is consistent with the rendition of the inherent power of *Da-sein* to comprehend being, and, hence, beings-in-the-whole (the world), and disclose the meaning of phenomena through words. Λόγος (and deep understanding) is foundationally related to human intelligence, but it is not the matter Marcus and Davis set out to elucidate. The phenomenon of intelligence is contingent on λόγος. It also succeeds and is continuous with thinking, which, thought phenomenologically, is steadfastly relinquishing λόγος to the to-be-thought and, like human intelligence, also ensues from λόγος.

Λόγος, the ground of human intelligence and thinking, also is the home of *Da-sein*'s freedom-to-be. The freedom-to-be, the phenomenon of freedom, is not a moral, ideological, or cultural conception, although it commonly motivates and provokes such ideas. It also is not a positive correlation between the subject and its intended object, though it encompasses that moment too. It is the inherent capacity of *Da-sein* to discover, comprehend, and speak the truth, ἀλήθεια (*alḗtheia*) of phenomena (the way they are from themselves), including itself, and, in the light of the truth, independently (responsibly) decide the way of its to be (itself, its future). The phenomenon of freedom, the freedom-to-be, rendered phenomenologically, is the intrinsic potentiality, δύναμις (*dynamis*) of human being to (struggle to) know (disclose-say), evaluate (comprehend), and overcome itself to dwell (stand) in the truth (of transcendence) and choose its meaning and future. It is inseparable from the phenomena of intelligence and thinking, and, like them, belongs to λόγος. The phenomena of intelligence, thinking, and freedom, reduced to their ownmost, are cognate phenomena whose inception originates in/from λόγος.

## 4. Common sense

Marcus and Davis define common sense as a robust understanding of the world, how it works, and how it should continue working. Although specifics are woefully lacking, the definition comes closer to the phenomenological interpretation than the notion of deep understanding does. As thought by Schutz, whose phenomenology of daily life is framed transcendental phenomenologically, the phenomenon of common sense is the phenomenon of typification. It is typical knowing-doing-saying and the world of daily life, or the everyday life-world. In the final analysis, the phenomena of common sense, typification, and life-world are factically equivalent in Schutz's phenomenology (Natanson, 1982; Trujillo, 2021). Schutz demystifies the life-world (transcendence) by propositionally revealing its typical constitution, interpretation, and unfurling. The experiences of the everyday person, he postulates, are mostly typical ones. They largely comprise typifications, or "common-sense constructs" (Schuetz, 1953; Schutz, 1982b).

Typifications are taken-for-granted, analogously or associatively derived knowledge and experiences (Schutz, 1970). They are automatically (habitually) lived meanings that open-endedly define the way things of the same basic type generally are and should be given how they have been experienced (Schutz, 1982a, 1982b). They are commonly learned from others and usually shared among persons of the same in-group. They are flexible, adaptable, and open to revalidation, reinterpretation, and falsification. Their viability is assumed as long as they work (Schuetz, 1953; Schutz, 1982b). They comprise pragmatic knowledge and knowing. Their constitution corresponds to a world of eminently practical interests, a place persons are obligated to "dominate" and "change" to realize their purposes (Schuetz, 1945, p. 534).

Typifications are constituted in consciousness, accrete as a stock of knowledge, and, once articulated, shape the ongoing constitution of experiencing. They are either given to experiencing or immediately at hand ready to come automatically into play in consciousness to facilitate everyday interpreting and living. Typifications are embedded in everyday language, or the vernacular, which Schuetz describes as "the typifying medium *par excellence*" (1953, pp. 9–10). It is a treasure house of preconstitued types and characteristics, each of them carrying along an open horizon of unexplored typical contents (Schutz, 1982a). Although typifications are originally constituted analogously (correlatively) or associatively (referentially), they become taken-for-granted insofar as they are habitually employed. They transform from predicated to prepredicated knowledge and their operation in consciousness becomes increasingly smooth and fast.

The rendition of the phenomenon of common sense as the phenomenon of typification, as "ways of life" and "efficient recipes for the use of typical means for bringing about typical ends in typical situations" (Schuetz, 1953, p. 10), suggests a framework that may align machine learning, which essentially amounts to correlation and classification, to what humans are doing. Machine learning outwardly parallels the analogous and associative constitution of meaning. It is not structurally unlike the way typifications are generated in consciousness and shape experiencing. A radical difference between machine learning and the phenomenon of common sense, however, is that human being (transcendence) is the experiences it typifies and typically interprets whereas machine learning only correlates, classifies, and calculates data. It neither experiences nor interprets. It does not perceive, but only registers and records, and does not harbor the meaning of "is". Marcus and Davis allude to these radical differences between human being and machines when they underscore the fundamental "mismatch between what machines are good at doing now – classifying things into categories – and the sort of reasoning and real-world understanding" they correspond with common sense (2019, p. 74).

Schutz's phenomenology does not suggest common sense antedates deep understanding, as Marcus and Davis contend, however; whether it will in general AI is a different matter. The phenomenon of common sense, like human intelligence, is contingent on λόγος. Its possibility is conditional on the inherent potentiality of human being to comprehend "is" (being) and wield language from within language. Without this power, human being would not have the ability disclose, constitute, interpret, or express meaning of any kind, including typical ones.

## 5. Moral inconsistencies

Marcus and Davis envision a world where general AI robots, machines endowed with common sense, deep understanding, and values, would assume the toil and hazards of human living. They would liberate humanity of life's burdens and dangers, and free persons to commit themselves to thinking, inventing, creating, and discovering. These general AI robots, as Marcus and Davis describe them, would, among other things, build, maintain, and clean our buildings and homes, cook our meals, do our chores, fight fires and mitigate all types of disasters, rescue persons and protect property, teach, conduct scientific, literary, and legal research, diagnose patients, work in our factories, program other machines, and supplement human creativity, invention, and discovery. The cost of living would be reduced too, as general AI robots would do everything humans can do, but cheaper, more efficiently, and in many instances, better.

The moral inconsistencies implied by this vision are remarkable. Marcus and Davis propose creating machines that reproduce fundamental elements of human being but are consigned to servitude. They would be invested with self-awareness, situational awareness, reasoning, logic, perception, an understanding of language, and the ability to know and interpret. They would be endowed with the capacity to comprehend, evaluate, and overcome themselves (openly learn and flexibly adapt) and comprehend their actions. They would "genuinely" understand the "world" (Marcus & Davis, 2019, p. 199), and, hence, themselves and their place in it. Marcus and Davis's general AI robots would reproduce human phenomenon that phenomenology corresponds with λόγος and the freedom-to-be, but disenfranchised of the freedom to decide their futures. These machines would be humanized to be dehumanized. They would be lobotomized by hardwiring them with Asimov's (1977) Second Law of Robotics: obey the orders of human beings except where such orders would conflict with the First Law, which is to not injure a human being.

Captain Jean-Luc Picard addresses the same inconsistencies (distortions) in the *Star Trek: The Next Generation* (created by Gene Roddenberry, 1987–1994) episode, "The Measure of a Man" (directed by Robert Scheerer, 1989). There he defends Lieutenant Commander Data, a one-of-a-kind android endowed with capabilities that correspond to general AI, during a hearing to decide his right to choose whether to submit to a *Starfleet* order to reassign him to *Starbase 173* to undergo refit experiments. Commander Bruce Maddox, a cyberneticist, initiated the order, and boarded the *USS* Enterprise (*NCC-1701-D*) to execute it during the starship's visit to *Starbase 173*. He believes that by disassembling and studying Lieutenant Commander Data, he will be able to learn how to replicate him. Lieutenant Commander Data assesses Commander Bruce Maddox's refit experiment would despoil him of the "essence" of his experiences and memories, hence, his self, and, following Captain Jean-Luc Picard's guidance, resigns his commission to avert it. Commander Bruce Maddox insists the order remains valid because Lieutenant Commander Data is *Starfleet* property and therefore has no rights, including the right to choose whether to submit to an experiment that could ruin him. Captain Philippa Louvois, who is posted to *Starbase 173*, initially rules for Commander Bruce Maddox. Captain Jean-Luc Picard's challenges the ruling, and compels Captain Philippa Louvois to hold a hearing to decide the matter formally. Captain Jean-Luc Picard

volunteers to serve as Lieutenant Commander Data's defense. *USS* Enterprise (*NCC-1701-D*) Commander William Riker reluctantly agrees to represent Commander Bruce Maddox. Captain Philippa Louvois ultimately rules for Lieutenant Commander Data. Captain Jean-Luc Picard's defense of Lieutenant Commander Data reads:

Captain Jean-Luc Picard: "Commander [Bruce Maddox], is your contention that Commander Data is not a sentient being and therefore not entitled to all the rights reserved for all life-forms within this federation?"

Commander Bruce Maddox: "Data is not sentient, no".

Captain Jean-Luc Picard: "Commander, would you enlighten us, what is required for sentience?"

Commander Bruce Maddox: "Intelligence, self-awareness, consciousness".

Captain Jean-Luc Picard: "Prove to the Court that I'm sentient".

Commander Bruce Maddox: "This is absurd. We all know you're sentient".

Captain Jean-Luc Picard: "So I'm sentient, but Commander Data is not".

Commander Bruce Maddox: "That's right".

Captain Jean-Luc Picard: "Why? Why am I sentient?"

Commander Bruce Maddox: "Well, you are self-aware".

Captain Jean-Luc Picard: "Ah, that's the second of your criteria. Let's deal with the first, intelligence. Is Commander Data intelligent?"

Commander Bruce Maddox: "Yes. It has the ability to learn and understand, and to cope with new situations".

Captain Jean-Luc Picard: "Like this hearing?"

Commander Bruce Maddox: "Yes".

Captain Jean-Luc Picard: "What about self-awareness? What does that mean? Why am I self-aware?"

Commander Bruce Maddox: "Because you are conscious of your existence and actions. You are aware of yourself and your own ego".

Captain Jean-Luc Picard: "Commander Data, what are you doing now?"

Lieutenant Commander Data: "I'm taking part in a legal hearing to determine my rights and status. Am I a person or property?"

Captain Jean-Luc Picard: "And what's at stake?"

Lieutenant Commander Data: "My right to choose, perhaps my very life".

Captain Jean-Luc Picard: "'My rights'. 'My status'. 'My right to choose'. 'My life'. Well, he seems reasonably self-aware to me, Commander. I'm waiting".

Commander Bruce Maddox: "This is exceedingly difficult".

Captain Jean-Luc Picard: "Do you like Commander Data?"

Commander Bruce Maddox: "I… I don't know it well enough to like or dislike it".

Captain Jean-Luc Picard: "But you admire him?"

Commander Bruce Maddox: "Yes, it's an extraordinary piece…"

Captain Jean-Luc Picard: "Of engineering and programming. Yes, you have said that. Commander, you have devoted your life to the study of cybernetics in general?"

Commander Bruce Maddox: "Yes".

Captain Jean-Luc Picard: "And Commander Data in particular?"

Commander Bruce Maddox: "Yes".

Captain Jean-Luc Picard: "And now you propose to dismantle him".

Commander Bruce Maddox: "So I could learn from it and construct more".

Captain Jean-Luc Picard: "How many more?"

Commander Bruce Maddox: "As many as are needed. Hundreds, thousands if necessary. There is no limit".

Captain Jean-Luc Picard: "A single Data, and forgive me, Commander [Data], is a curiosity, a wonder even. But thousands of Datas. Isn't that becoming…a race? And won't we be judged by how we treat that race? Now, tell me Commander, what is Data?"

Commander Bruce Maddox: "I don't understand".

Captain Jean-Luc Picard: "What is he?"

Commander Bruce Maddox: "A machine!"

Captain Jean-Luc Picard: "Is he? Are you sure? You see he's met two of your three criteria for sentience, so what if he meets the third, consciousness, in even in the smallest degree? What is he then? I don't know. Do you? [then turning to Commander William Riker] Do you? [and finally Captain Philippa Louvois] Do you? Well, that's the question you have to answer. Your Honor, a courtroom is a crucible. In it we burn away irrelevancies until we are left with a pure product, the truth, for all time. Now, sooner or later, this man, or others like him, will succeed in replicating Commander Data. Now, the decision you reach here today will determine how we will regard this creation of our genius. It will reveal the kind of a people we are, what he is destined to be. It will reach far beyond this courtroom and this, one android. It could significantly redefine the boundaries of personal liberty and freedom. Expanding them for some, savagely curtailing them for others. Are you prepared to condemn him and all that come after him to servitude and slavery? Your Honor, Starfleet was founded to seek out new life. Well, there it sits! Waiting. You [Captain Philippa Louvois] wanted a chance to make law. Well, here it is. Make it a good one".

In an earlier scene of the same episode, Guinan, the *USS* Enterprise (*NCC-1701-D*)'s bartender and impromptu spiritual guide, inspired Captain Jean-Luc Picard's defense of Lieutenant Commander Data by provoking him to see the human distortions in Commander Bruce Maddox's argument after Commander William Riker forcefully presented his case:

Guinan: "And now he's about to be ruled the property of Starfleet. That should increase his value".

Captain Jean-Luc Picard: "In what way?"

Guinan: "Well, consider that in the history of many worlds, there have always been disposable creatures. They do the dirty work. They do the work that no one else wants to do, because it's too difficult or too hazardous. And an army of Datas, all disposable… You don't have to think about their welfare, you don't think about how they feel. Whole generations of disposable people".

Captain Jean-Luc Picard: "You're talking about slavery".

Guinan: "I think that's a little harsh".

Captain Jean-Luc Picard: "I don't think that's a little harsh, I think that's the truth. But that's a truth that we have obscured behind a comfortable, easy euphemism. Property".

Commander Bruce Maddox's calculus, his proposal to populate *Starfleet* with Lieutenant Commander Data-like androids manufactured to act "as our hands and eyes in dangerous situations" and boundlessly expand "the horizons for human achievement" is paralleled in Marcus and Davis's proposal to fabricate legions of general AI robots earmarked to discharge their creators of the burdens and dangers of human living. Marcus and Davis's general AI machines would be foreordained by their architects to serve as chattel for human purposes. Marcus and Davis's vision, like Commander Bruce Maddox's, disregards the freedom inherent to the human phenomena affiliated with its rendition of general AI. The narratives only differ in their starting-points. Commander Bruce Maddox witnesses the man in the machine, but chooses to see only the machine. His sole goal is to replicate a technology. Marcus and Davis, on the other hand, witness the machine (current AI), envisage the reproduction of a man, but choose to see only a machine. They foresee the possibility of endowing machines with the human capacity "to reason, understand language, and comprehend the world", learn "efficiently, and with human-like flexibility" (Marcus & Davis, 2019, p. 25), and interpret and evaluate themselves, their actions, and the consequences of their actions, but fails to see, forgets, or, perhaps, even more alarmingly, disregards the freedom intrinsic to this power and the human "to be".

## Concluding remarks

The aim of these reflections is not to criticize Marcus and Davis, but to deconstruct it critically. It is to work (dialogue) with Marcus and Davis to contribute to the discernment of human intelligence and the milestones it contends AI must achieve to transition to a general formulation. Marcus and Davis is a timely and daring work. Marcus and Davis describe a technology whose global economic impact could reach "$13 trillion" (2019, p. 11), showcase its current limitations, and discern how far it must go to live up to its rhetoric. Their observations about human intelligence – which they derive from an examination of what "our minds" are doing or "even trying to do" (Marcus & Davis, 2019, pp. 25–26) – are phenomenologically spirited. They invite phenomenological scrutiny and suggest the method's ability to elucidate phenomena industrialists, entrepreneurs, programmers, and engineers are laboring to clone in machines.

Marcus and Davis provoke us to think about the meaning of human intelligence. This effect may be one of its most valuable, although not necessarily most readily recognized, dividends. It would have heightened its thought-provoking power, however, by first asking if general AI was possible rather than assuming it was "not impossible" (Marcus & Davis, 2019, p. 74). The formula would have forced a confrontation with the meaning of "is" (and, perhaps, the everyday forgottenness of "to be") and compelled a consideration of human intelligence within the context of human being. It may have even begged an introduction of a phenomenological (first-principles) perspective that perhaps would have helped it deliver a more concise and exact rendition of common sense and human intelligence including their relation to deep understanding. Investigating the question of general AI might also have cornered Marcus and Davis into investigating the radical differences between human being and machines. Illuminating them would have further checked the misconceptions about AI that have contributed its rhetoric and impeded progress.

# References

Asimov, I. (1977). *The robot series*. *I*, *Robot*. Bantam Books.

Breda, van H. L. (1982). Preface. In A. Schutz, *Phaenomenologica: Collection Publiée Sous le Patronage des Centres d'Archives-Husserl. Book: 11. Collected papers. Vol. 1: The problem of social reality* (vi–xiii). M. Natanson (Ed.). Martinus Nijhoff.

Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the Association for Computing Machinery*, *58*(9), 92–103. https://doi.org/10.1145/2701413

Emad, P. (2000). On the inception of being-historical thinking and its unfolding as mindfulness. *Heidegger Studies*, *16*, 55–71. https://doi.org/10.5840/heideggerstud2000163

Frankl, V. E. (2014). *The will to meaning: Foundations and applications of logotherapy*. A Plume Book.

Gurwitsch, A. (1970). Introduction. In A. Schutz, *Phaenomenologica: Collection Publiée Sous le Patronage des Centres d'Archives-Husserl. Book: 22. Collected papers. Vol. 3: Studies in Phenomenological Philosophy* (xi–xxxi). Martinus Nijhoff.

Heidegger, M. (2001). *Being and time*. Blackwell.

Heidegger, M. (2000). Hölderlin and the essence of poetry. In M. Heidegger, *Elucidations of Hölderlin's Poetry* (pp. 51–65). Humanity Books.

Heidegger, M. (2008a). Letter on humanism. In M. Heidegger, *Basic writings from being and time* (*1927*) *to the task of thinking* (*1964*) (pp. 213–266). D. F. Krell (Ed.). Harper Perennial/Modern Thought.

Heidegger, M. (2008b). The question concerning technology. In M. Heidegger, *Basic writings from being and time* (*1927*) *to the task of thinking* (*1964*) (pp. 307–342). D. F. Krell (Ed.). Harper Perennial/Modern Thought.

Husserl, E. (1982). *Cartesian meditations: An introduction to phenomenology*. Martinus Nijhoff Publishers.

Husserl, E. (1980). *Collected works. Vol. 1: Phenomenology and the foundations of the sciences. Book 3: Ideas pertaining to a pure phenomenology and to a phenomenological philosophy*. Martinus Nijhoff Publishers. https://doi.org/10.1007/978-94-009-8779-1

Husserl, E. (1970). *Northwestern University studies in phenomenology and existential philosophy. The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy*. J. Wild (General Ed.). Northwestern University Press.

Kovacs, G. (2016). Heidegger in dialogue with Husserl. *Heidegger Studies*, *32*, 245–258. https://doi.org/10.5840/heideggerstud20163213

Kovacs, G. (2013). Heidegger's insight into the history of language. *Heidegger Studies*, *29*, 121–132. https://doi.org/10.5840/heideggerstud2013296

Legg, Sh., & Hutter, M. (2006, 11–12 May). A formal measure of machine intelligence. In *Proceedings of 15th Annual Machine Learning Conference of Belgium and the Netherlands* (*Benelearn 2006*) (pp. 73–80). Ghent, Belgium. Technical Report No. IDSIA-10-06. Istituto Dalle Molle di Studi sull'Intelligenza Artificiale.

Legg, Sh., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, *17*, 391–444. https://doi.org/10.1007/s11023-007-9079-x

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. Routledge & Kegan Paul.

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Farrar, Straus and Giroux.

Natanson, M. (1982). Introduction. In A. Schutz, *Phaenomenologica: Collection Publiée Sous le Patronage des Centres d'Archives-Husserl. Book: 11. Collected papers. Vol. 1: The problem of social reality* (xxv–xlvii). M. Natanson (Ed.). Martinus Nijhoff.

Pinker, S. (2018). We're told to fear robots. But why do we think they'll turn on us? *Popular Science*. https://www.popsci.com/robot-uprising-enlightenment-now/

Schuetz, A. (1953). Common-sense and scientific interpretation of human action. *Philosophy and Phenomenological Research: A Quarterly Journal*, *14*(1), 1–38. https://doi.org/10.2307/2104013

Schuetz, A. (1945). On multiple realities, *Philosophy and Phenomenological Research: A Quarterly Journal*, *5*(4), 533–576. https://doi.org/10.2307/2102818

Schutz, M. (1982a). Language, language disturbances, and the texture of consciousness. In A. Schutz, *Phaenomenologica: Collection Publiée Sous le Patronage des Centres d'Archives-Husserl. Book: 11. Collected Papers. Vol. 1: The Problem of Social Reality* (pp. 261–286). M. Natanson (Ed.). Martinus Nijhoff.

Schutz, M. (1982b). Symbol, reality and society. In A. Schutz, *Phaenomenologica: Collection Publiée Sous le Patronage des Centres d'Archives-Husserl. Book: 11. Collected papers. Vol. 1: The Problem of Social Reality* (pp. 287–356). M. Natanson (Ed.). Martinus Nijhoff.

Schutz, A. (1970). The problem of transcendental intersubjectivity in Husserl. In A. Schutz, *Phaenomenologica: Collection Publiée Sous le Patronage des Centres d'Archives-Husserl. Book: 22. Collected papers. Vol. 3: Studies in phenomenological philosophy* (pp. 51–84). I. Schutz (Ed.). Martinus Nijhoff. https://doi.org/10.1007/978-94-015-3456-7_4

Spiegelberg, H. (1994). *The phenomenological movement: A historical introduction*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-7491-3

Stannard, R. (2008). *Very short introductions. Relativity: A very short introduction*. Oxford University Press. https://doi.org/10.1093/actrade/9780199236220.001.0001

Trujillo, J. (2021). The phenomenon of common sense and the thinking of Alfred Schutz. *Filosofija. Sociologija*, *32*(3), 268–276. https://doi.org/10.6001/fil-soc.v32i3.4499

Zaner, R. M. (1961). Theory of intersubjectivity: Alfred Schutz. *Social Research: An International Quarterly*, *28*(1), 71–93.

Zeh, H. D. (2007). *The Frontiers Collection. The physical basis of the direction of time*. A. C. Elitzur, M. P. Silverman, J. Tuszynski, R. Vaas & H. D. Zeh (Series Eds.). Springer-Verlag.