

The Rhetoric of Non-Knowledge in Popularizations of Artificial Intelligence: Harari’s *Homo Deus* and Tegmark’s *Life 3.0*

Jürgen Meyer

In this article I investigate different levels of non-knowledge in recent popularizations of science by examining two bestselling monographs on Artificial Intelligence (hereafter referred to as AI): historian Yuval Noah Harari’s *Homo Deus: A Brief History of Tomorrow* (2016) and physicist Max Tegmark’s *Life 3.0: Being Human in the Age of Artificial Intelligence* (2017). They belong to a wider field of futurological literature, which includes such works as Ray Kurzweil’s *The Singularity is Near: When Humans Transcend Biology* (2005), Nick Bostrom’s *Superintelligence: Paths, Dangers, Strategies* (2014), Amy Webbs and Andrew Hessel’s *The Genesis Machine* (2023), or Karen Hao’s *Empire of AI: Dreams and Nightmares in Sam Altman’s OpenAI* (2025). Their common feature is the discussion of the current knowledge, as well as the imponderability, of AI as a technological extension of the human body, mind, and society. They also evaluate the possibilities and obstacles in the future development of a super-intelligent AI, or an artificial general intelligence (AGI) that exceeds the human mind, the latter defined by Bostrom as “any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest” (Bostrom 26). Therefore, all these texts have many parallels in content and subject; however, they also display clear contrasts in their respective professional attitudes towards science, technology, and their representation to a more general public. The following discussion distinguishes three different foci: first, the treatment of the different degrees of non-knowledge emerging from discursive structures in the texts: on the personal, intersubjective and systemic levels; second, the discursive and narrative strategies, inquiring how the authors deal with their own historical position as writers who discuss not only the past and the present, but also a yet unknown future; and third, the construction of the implied recipient of these studies, i.e., the way the two authors communicate with their readers, by guiding them through the argument and giving them space for developing their own attitudes towards AI.

Since the last focus frames the first two, it is necessary to begin by elucidating two contrasting approaches to analysing science popularization, drawing on insights from Social Science Studies in general, and Science Communication in particular. Here, it is the branch of Public Awareness of Science (henceforth PAS), which has fashioned the communication channels and media connecting the scientific disciplines and the public in a very different way than what may be referred to as a divisive “deficit model” (Cortassa 447) of “genuine” vs “popularized knowledge”, critically reviewed and outlined by Hilgartner (524).

Science Communication Approaches: “Deficit Model” vs. “PAS”

According to the deficit model in science communication, there are two major channels by which scientific and technological expert knowledge enters the public sphere: either in material and often commodified shape, i.e., as marketable products (such as increasingly smart home technologies), or as a public medium, in textual, audio-visual, analogue or digital format (e.g., as science popularization, or science blog or vlog). In the first case, before many an innovative hi-tech prototype is given

permission to be circulated as a commodity in a more or less extensive segment of the public economy, it will be closely monitored for its safety and security standards; i.e., there will be a detailed and often lengthy risk assessment, according to national and international norms and standards (fig. 1, "Public Material Sphere", in the bottom right quadrant).

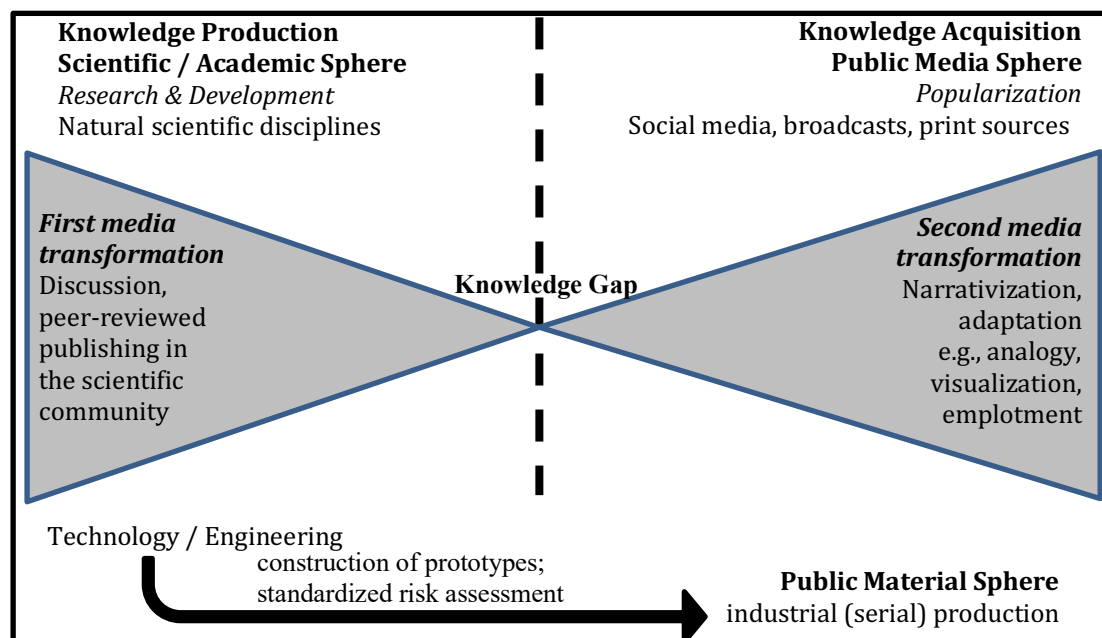


Figure 1: The deficit model of science communication (Source: JM)

The deficit model divides the epistemological field into the segments of “science” and “the public”. As the locations of “knowledge production” and “knowledge acquisition” (Whitley 3-10), they are separated by a deep knowledge gap. The usually competitive character of the first transformative process from laboratory action to academic publication turns, in the second step, into an often cumulative and adaptive representation of the different phases of generating scientific or technological knowledge. Here, the audience finds a discussion of research questions as starting points for the journey to results considered as reliable, valid and credible. However, only a fraction of scientific research and development will be featured in popularizations: usually the most blatant failures and the most revolutionary “success stories” are narrativized. These narratives, and with them the new insights, are represented as a happy, not least heroic, ending after a series of failures, obstacles and intellectual, methodological or technological dead ends. The transformative discourse requires converting the competitive character of actual research into a narrative emplotment with protagonists and antagonists, engaged in controversies between, and cooperation with, natural persons as well as institutions (Gregory and Miller 108-14). Commonly used popularizing strategies include representational techniques such as point of view-technique, chronology, linguistic register, and suspense, as well as rhetorical-stylistic means, for instance imagery including analogy or metaphor and multimodal strategies combining text and visualization (Hesse, Kompa, Nappo, and Fröschl). Many of these strategies occur also in academic texts as well as in science popularizations, but will be much more frequently used in the latter.

Ultimately, the deficit model may be traced back to the idea of the “two cultures” with its basic assumption of a deep communicative rift between the sciences and the humanities (Snow 2). If Snow limited his criticism to the attitudinal rivalry

and mutual incomprehension between these two intellectual spheres, the "deficit model" generalizes it, expanding the scope to an adverse relationship between the sciences and a general public caused by the deep knowledge gap and a lack, or deficiency, of cognition and understanding on the side of the public. Increasing reservations and uneasiness with this oversimplified dichotomous image in the last two decades of the 20th century gave rise to an emerging field of Social Science Studies and Science Technology Studies. Instead, the focus shifted to the complex interplay of communicative relations within the sciences (Woolgar and Latour) and to those between the sciences and different parts of society, past and present (Shapin). Many insights from this newly established area share the conviction that "the scientific document is, in effect, a social artifact shaped by and functioning in the social milieu – the milieu of science and the milieu of the surrounding world" (Locke 19). This also means that the boundaries between the sciences and society are "fluid and 'slippery'" (Einsiedel 5), and that the agents participating in this communicative interaction are much more complex than the dichotomy suggests.

Particular attention to the knowledge transfer from the sciences to various audiences is paid by studies on PAS. Within this framework, the transformations in fig. 1 above are re-conceptualized in terms of a continuity as a multilateral, dialogic information exchange between the scientific community and the public media sphere (fig. 2).

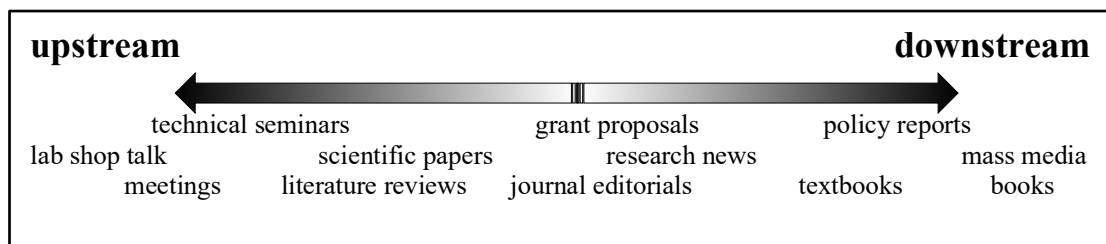


Figure 2: "Contexts in which Scientific Knowledge is Communicated" (Source: Hilgartner, 528, adapted)

Although the model displays a hierarchical relation in the spatial metaphors "upstream" for the intradisciplinary, and "downstream" for the public communication, it closes the former gap between the two poles with a wide range of actors and media formats (Hilgartner 524-25). These correspond to a variety of audiences within and without the sciences, each informed by different degrees of knowledge, education, contexts and interests. At this point, innovative knowledge has left its secured space within the boundaries of science and engineering discourses, and has disseminated into the fields of economy, law, politics, ethics, and education. Thus, scientific and non-scientific players become equally involved as participants in the knowledge production process, which therefore cannot be located solely either in the laboratory or in the lecture theatre: in fact, is also found in such places as funding committees, political, legal or ethical commissions, and other adjacent areas with a more or less immediate influence on the scientific enterprise. To overcome the inherent "epistemic asymmetry" between scientists and the public, Cortassa replaces the pejorative public "deficit" by its re-evaluation as a positive "premise" for any endeavour in successful science communication (451-54). This change of perception leads to new demands regarding popular science writing, including the call for less formal and technical language than the textual and discourse-specific conventions known from peer-reviewed publishing.

The above sketch of the two opposing conceptualizations of science communication in light of the binary "deficit model" and of the continuous PAS approach will be a significant key to comprehending the way the two texts analysed here encourage or discourage their recipients to engage in critical, knowledge-based interaction with the scientific and technological topics they present. Before doing so, it seems necessary to outline the conceptual focus on knowledge and non-knowledge.

Non-Knowledge in Science Writing

Usually, widely accepted, i.e. in the Kuhnian sense paradigmatic epistemic positions constitute the foundation for a reliable exploration and explanation of valid knowledge formations in different disciplines such as historiography, (auto)biography, and in particular that of scientific writing. Such scientific knowledge is often achieved over years, decades or even centuries of theorizing and experimenting in and out of a laboratory, and of publishing and discussing in intradisciplinary as well as interdisciplinary exchanges, frequently involving the scientific community worldwide. More generally, Paul refers to the multiplicity of "schematic, oppositional distinctions ["pure/abstract/theoretical versus applied/empirical, particular/local versus universal/global, scholarly (expertise) versus popular/ vernacular (understanding)"]", which "codify how to talk about knowledge formations and in any given moment may even mark knowledge as relevant (or irrelevant), true (or false), usable (or unusable), sophisticated (or lacking complexity), and so on in specific situations and for various purposes." (Paul 172)

However, non-knowledge in such non-fiction seems to be a paradox. The representation of the absence of epistemic knowledge, or blind spots in scientific endeavour, is itself both varied and complex. A few manifestations of non-knowledge have a historic dimension, insofar as a new episteme may replace an older one, causing either a macro-scale "paradigm change" or an evolutionary development of the episteme, filling the "normal science" gaps with new insights (Kuhn 52). Knowledge may even actively be withheld from a wider public, contributing to the degree of (mis-) information circulating within a society (Auguscik and Broders 80-82). The emergent field of "agnotology" brims with controversial terms such as "non-knowledge" and "ignorance" (Proctor 1-36). Proctor refines the concept of "ignorance as native state (or resource), ignorance as lost realm (or selective choice), and ignorance as deliberately engineered and strategic ploy (active construct)" (2-3). In contrast, Japp distinguishes a range of degrees of (non-) knowledge: First, scientific knowledge ("decision-makers' or originators' perspective of normalization rhetoric"); second, specific non-knowledge ("frequently linked with dissent among experts"); third, unspecific non-knowledge ("frequently linked with (aggressive) claims [in the public]") (Japp 229-30, incl. notes 10-12). The underlying terminological and conceptual uneasiness leads others to question the term "non-knowledge" as such, arguing that this was a contradiction in terms, because non-knowledge assumes the existence of knowledge which it negates at the same time: "anyone referring to ignorance cannot avoid making claims to know something about who is ignorant of what" (Smithson 210). In view of this controversy, it may be useful to separate two different levels of ignorance, one at which there is a meta-cognitive knowledge-gap, leading to the opposition of an awareness about one's individual, sometimes deliberately chosen (non-) knowledge and an unconscious level of ignorance. The other level refers to the content in which the knowledge of (factual) themes and (procedural) methods are under review (see fig. 3).

Following the distinction of personal and interpersonal negative knowledge formations (the "lacuna" or "desideratum", Bouissac 463), the next section shows that it is possible to discern different discursive strategies in the representation of individual (= cognitive), disciplinary (= historical) and systemic (= principal) non-knowledge formations in popularizations (Meyer 230-31).

META-LEVEL		
Awareness		Ignorance
knowing that one does not know denying to know (= choosing to keep ignorant)		not knowing that one does not know (business) secrets, taboo (= being kept ignorant)

SUBJECT LEVEL		
individual	disciplinary	systemic
what / how do I know? what do I not know?	is the number π finite? what is life / consciousness?	what will we know in the future? do we live in a multiverse?

Figure 3: Levels of (Non-)Knowledge (Source: JM)

Non-Knowledge in *Homo Deus*: Turning Science into Dogma

Harari's *Homo Deus* is divided into four parts, in their structure resembling a fugue, in which the first part serves as a prelude for the following three movements. They elaborate the earlier exposition, with the representation of the arguable human victory over plague, famine, and death and the ongoing struggle for achieving spiritual, rather than scientific goals, in particular "immortality, bliss and divinity" (Harari, *Homo Deus*, 75). Harari constructs an interlocutor who attributes to science the theological status of becoming a quasi-religious system with its foundation in axiomatic sets of belief. Arguing that "the rise of humanism also contains the seeds of its downfall" (75), it turns out that it is predominantly Harari's own mode of discussing isolated fragments of the respective scientific discourses that he eventually aligns, by often unconvincing analogy, with vaguely epistemic categories such as "myth" and religious "dogma". In the final pages of *Homo Deus*, he arrives at the following set of conclusions:

1. Science is converging on an all-compassing dogma, which says that organisms are algorithms and life is data-processing.
2. Intelligence is decoupling from consciousness.
3. Non-conscious but highly intelligent algorithms may soon know us better than we know ourselves. (Harari 462)

The first statement is a leitmotif throughout his narrative, although Harari rarely explains to which particularly "dogmatic" writings or authorities he refers, except dropping an occasional book title such as Kurzweil's *The Singularity is Near*. Regarding the latter publication, even though Kurzweil explains the scientific concept of "singularity" and explains its different usage in mathematics, astrophysics and computer technologies (Kurzweil, *The Singularity is Near*, 22-23), Harari recognizes in the title an echo of the scriptural phrase "the kingdom of heaven is near" and labels it a "book of prophecies" (Harari, *Homo Deus*, 444). In this quasi-religious realm of "dataism", live organisms can be, and will be, reduced to mere algorithms in the

course of scientific progress (372, 382, 402, 428, 445 and 452). The explicit assumption of a transformation from science and technology to a substitute religion will, Harari says, lead to the extinction of the human race by ever more efficient and powerful algorithms.

In opinions such as these, Harari's interlocutor eschews, and in fact precludes, any in-depth discussion of AI-research and development activities that are specifically focused on machine-learning and the operation of many-layered neural networks created for processes such as decision making or image recognition (in contrast to this avoidance of such a discussion, consider Tegmark's technical elaborations in *Life 3.0*, Tegmark 61-66). Technically, the processes in such a neural network may be divided into nine different transformational stages (including "input embeddings", i.e., the transformation of verbal input into numerical representation, and "multi-head attention", i.e., the identification of "different types of relationships between tokens", IBM website on Transformer Models; see also Kurzweil, *The Singularity is Nearer*, 19-24). Strikingly, it has remained impossible for researchers to comprehend how exactly such an advanced AI system works, and which specific internal operations within the neural network give rise to a particular result: "Since the changes throughout those millions of connections were so complex and minute, researchers aren't able to exactly determine what is happening. They just get an output that works" (Gershgorn n.p.). It is this lack of comprehensibility that has become one of the biggest concerns in the current debate about the chances and challenges, potentials and risks, manageability and unpredictability of AI, although this lack of understanding of such a system is still a far cry from anything that might be considered as evidence of a machine consciousness, let alone sentience.

By avoiding to lead this discussion on a disciplinary level (and, for instance, considering it as a problem that will eventually be solved, given better analytic tools observing the machine operations), and by covering up this lack of objective discussion with his interlocutor's hyperbolic, biased rhetoric, Harari predicts that "the individual is becoming a tiny chip inside a giant system that nobody really understands" (Harari, *Homo Deus*, 449). This culmination is an updated, computerized version of the 18th-century mechanistic metaphor of the "cog in the machine", generalizing that "[n]o one is [...] capable of connecting all the dots and seeing the full picture ["in artificial intelligence, nanotechnology, big data or genetics"]" (59).

The third of the above statements gives the reader a teleological perspective with a strong anti-chilastic twist. Based on the conviction that "[t]hrough the details are obscure, we can nevertheless be sure about the general direction of history" (53), the author's mode of presenting the future is linear – the apocalypse being inevitable, though occurring, possibly unexpectedly, at an unpredictable moment:

humans might be reduced from engineers to chips, then to data, and eventually we might dissolve within the torrent of data like a clump of earth within a gushing river. Dataism thereby threatens to do to *Homo sapiens* what *Homo sapiens* has done to all other animals. (Harari, *Homo Deus*, 460; emphasis in original)

Harari applies two strategies for his attempt at downplaying his own disciplinary limitations regarding AI: an analogical style, and a dogmatic mindset. The book borrows from a variety of scientific notions and concepts enmeshed in an interdiscursive and, more often than not, incommensurable conglomeration of

historiography, politics, synthetic biology, neuroscience, AI, social sciences, economics, and religion studies. The multiplicity of these discourses evokes the construction of many analogies, which bring together ideas from several fields but which, here, often lead to a logical dead end. The stylistic device simulates an academic standing reminiscent of the Renaissance *uomo universale*. With these strategies, however, Harari violates one of the "unwritten rules of popular science writing" mentioned by Gregory and Miller: by entering more than two fields of research which he discusses as a historian rather than as a scientist, he fails to "stick strictly to a specific area of expertise" (82). Moreover, the interlocutor's figurative language, when focusing on the historicist dimension of non-knowledge, is enriched with imagination to envision almost exclusively biased, usually negative narrations representing a dystopian AI-governed world with humans as slaves, or no humans at all (or any natural life, for that matter). The style in *Homo Deus* abounds in a figurative language of scriptural tonality (as the "torrent of data" in the "gushing river"). Given this interdiscursive mixture, the text remains faithful to its title by injecting a hyperbolic religious discourse into the representation of scientific topics. In fact, the phrase *Homo Deus* seems to parody the Linnaean taxonomy established in biology: a biological species (*homo*) and a religious concept (*deus*). The subtitle, too, creates further tension in the concept of historiography as a means of predicting the future, rather than as a reconstruction of the past. Thus, the title, as epitome of the whole rhetorical strategy in the text, creates a contradiction in terms that evokes Samuel Johnson's famous 18th-century characterization of metaphysical wit in his "Life of Cowley", "yoking the most heterogeneous ideas by violence together". Although in later literary criticism this phrase was construed affirmatively as a high-quality standard of seventeenth-century poetry, Johnson used the phrase to criticize the obscurity of the Metaphysical poets' imagery, pointing at the "occult resemblances in things apparently unlike". If anything, he concluded the reader's "improvement [is] dearly bought" (Johnson 16).

Harari's frequent use of such conceited, sometimes outlandish analogies prevents his readers from arriving at an in-depth understanding of the complex scientific or technological topics discussed in the book. In the following example Harari's interlocutor constructs a capitalist correlative to the historiographical trope of "the king's two bodies", when he implicitly alludes to Ernst Kantorowicz's classical study which distinguishes a king's biological body from the spiritual institution of medieval kingship and its theological sanctification. This dichotomy only serves to short-circuit the ritualized tribute paid to ancient rulers by their subjects with the social status of celebrities in Horkheimer and Adorno's bleak vision of a modernist "culture industry":

Just like [an ancient Egyptian] pharaoh, Elvis [Presley] too had a biological body, complete with biological needs, desires and emotions. Elvis ate and drank and slept. Yet Elvis was much more than a biological body. Like pharaoh, Elvis was a story, a myth, a brand – and the brand was far more important than the biological body. (Harari, *Homo Deus*, 185)

This historical analogy culminates in an oversimplifying generalization: "[i]f the Sumerian gods remind us of present-day company brands, so the living-god pharaoh can be compared to modern brands such as Elvis Presley, Madonna, or Justin Bieber" (185). Thus, Kantorowicz's specific disciplinary trope "the king's two bodies" itself is distorted by this forced (in Johnson's terminology: "conceited") relativism, and the

medieval historicist concept is extended both to categorically different (ancient, polytheistic) metaphysical belief systems and to the present stereotype of the profane capitalist commodification of people who are revered as 'stars'.

If Harari uses such far-fetched analogies, it does not come as a surprise if he expects his reader to follow suit. There are many occasions which instruct the reader to "think about" (187), "suppose" (237), "assume" (425), or "consider" (137, 461), serving mainly to emulate the speculative and, as shown above, often misleading analogies that characterize the style in *Homo Deus*. Presenting his expectation of a data apocalypse as a future inevitability, Harari concludes with only a lukewarm retraction: "the scenarios outlined in this book should be understood as possibilities rather than as prophecies" (461). Elsewhere he relativizes his own predictions as a mere spice "pepper[ing] this book", serving to other ends than to "discuss our present-day dilemmas" (75). However, by reserving the term "prediction" for his own views, and ascribing only the "idle prophecies" (445) to the party of supposed dataists, Harari attempts to persuade his reader to accept his interlocutor as the only correct, righteous authority. It turns out that he constructs a science-alienated, dogmatic interlocutor who addresses an AI-ignorant reader. He assigns to himself the role of a mediator who envisions an inevitable and ineluctable future: "once technology enables us to re-engineer human minds, *Homo sapiens* will disappear, human history will come to an end and a completely new kind of process will begin, which *people like you and me* cannot comprehend" (53, my emphasis). Although this interlocutor can anticipate the future, he does not know let alone understand the details: His attitude shows his insistence on an unsurmountable "epistemic asymmetry" (Cortassa 457) between the profession (here: AI) and the general public, including himself. Viewed in this light, AI becomes an occult, esoteric discourse transcending all human intellectual capacity.

Despite its logical non-sequiturs, historic anachronisms and manipulative rather than persuasive rhetoric, *Homo Deus* was a huge international success documented by 40 translations listed in the English *Wikipedia* article, which also mentions that it was longlisted for the Wellcome Book Prize, an award for books dealing with medicine. *Time* ranked it as No. 4 among the top ten of non-fiction books published in 2017 (Howorth); similarly, *Der Spiegel* listed it as one of the top five best sold books in the same year (Anon.). However, if lay recipients, many critics included, accept this rhetoric, the reason may be that the interlocutor successfully manages to picture them as intellectual victims in a world that is too complex for them to comprehend, thus creating a strong dichotomy of "us" (the common-sensical lay people) vs "them" (culminating less in the human scientists or engineers than in the personification of evil, super-intelligent algorithms controlling everything else with or without consciousness). This, however, is no popularizing mode of representing complex scientific or technological contents, but a polemic use of anti-scientific rhetoric that turns readers away from the main subject. In consequence the all-too simplistic but multi-discursive, highly eclectic rhetoric in *Homo Deus* encourage populist thinking, despite Harari's apparent awareness of its dangers (understood as a combination of anti-liberal, anti-pluralist, and anti-progressive attitudes) and his own anti-populist stance abundantly professed in *Nexus: A Brief History of Information Networks from the Stone Age to AI*. To do Harari justice, in his "Epilogue" to this later book he concedes his limited knowledge on AI at the time of writing *Homo Deus*:

though I have no background in the technical aspects of computer science, I suddenly found myself, post-publication, with the reputation of an AI expert. This opened the doors to the offices of scientists, entrepreneurs and world leaders interested in AI and afforded me a fascinating, privileged look into the complex dynamics of the AI revolution. (Harari, *Nexus*, 395)

Although this new "privileged look" has allowed him a more focused representation of AI, his concept of a "dataist dogma", even though it is no longer referred to explicitly in the text, continues to pervade his argument in *Nexus*: Again, Harari constructs many more discursive homologies arguably connecting the Abrahamic religions ("Judaism, Christianity and Islam", *Nexus* 73-74) and "Silicon Valley", his metonymy for the global AI industry. Referring to the process of "AI canonization" (*Nexus* 399), he constructs another theology-driven analogy by pointing to the arguable parallels between Church fathers, whose compiling and editing of different text sources shaped the Christian Bible, and modern software authors, who select and collect algorithms from programming codes: "The present-day equivalents to Bishop Athanasius are the engineers who write the initial codes for AI, and who choose the dataset on which the baby-AI is trained" (400), on the historical canonization of the Bible and the role of Bishop Athanasius, 74-77 and 85-88). Thus, both books feature their respective interlocutors' prophetic, teleological inter-discourse, reduced in *Nexus* to the arguable convergences of religion and AI: in fact, they tell stories of AI that are shaped by a highly dogmatic science-sceptic mindset.

Non-Knowledge as Thinking Outside the Box in *Life 3.0*

Tegmark, like Harari a relative stranger to AI technology but with a scientific background as a physicist, follows a different strategy to gain credibility among his readers. Beyond his disciplinary expertise, he presents himself as peer to a circle of high-profile scientists and powerful tech-entrepreneurs – including such figures as Elon Musk (Tegmark 30-32 and 321-27), Google co-founder Larry Page (30-31), philosopher David Chalmers (284), and neurologist Christoph Koch (295-96), who are all introduced by highly subjective and personal character vignettes. All these scientists are involved in the administrative structures of the Future of Life Institute (FLI), a research and development organisation founded by Tegmark in 2014 and initially sponsored by Musk, forming an elite network that represents, in motivation and policy, the "Beneficial AI Movement" outlined in a section of *Life 3.0* (33-37).

Like Harari, Tegmark navigates within the discursive strands of different academic and scientific disciplines. In various chapters, he addresses the use of AI in the army, biology, ethics, engineering, medicine, and philosophy, and in discrete discussions weighs the potentials and limitations of AI technologies, which may bear some relevance for the different readers' evaluation of AI. In the final chapter he discusses the subject of "Consciousness" (281-315) as a multidisciplinary field involving highly specialized research areas such as neurosurgery, brain anatomy, physiology, medical imaging technologies, and philosophy, all of which are disciplines represented by scientists in Tegmark's FLI network, and whom he quotes at length as his main sources. In contrast to Harari's bookish armchair representation of AI in *Homo Deus*, however, Tegmark presents himself as an active participant in AI research, with open doors to (and for) other actors', i.e., his network colleagues', special research areas. He creates the impression that, despite his own disciplinary limitations, he is able to tackle the problem of defining and locating consciousness due to his scientific education as well as a creative, synthetic mindset. On the one

hand, this results in a positive, though not uncritical bias towards the collective research activities he discusses in *Life 3.0*. On the other hand, by highlighting the questions *not* answered by these specialists' research success, he self-confidently offers a non-conventional way of answering these questions.

The premise for the elaborate discussion at the end of *Life 3.0*, raising the key question whether a future super-intelligent AI may or may not be capable of developing consciousness, echoes Harari's assumption of a "great decoupling" of intelligence and consciousness (in statement no. 2, quoted above). Tegmark, however, gives this discussion a different turn: Complicating Harari's dichotomy of a naturally ethical biological intelligence and an essentially ethic-free artificial intelligence, Tegmark distinguishes three functional levels of intelligence outlined in a "Terminological Cheat Sheet", or glossary of terms as he uses them throughout his book: First, "narrow intelligence" corresponds to the "ability to accomplish a narrow set of goals"; second, "general intelligence" is the "ability to accomplish virtually any goal, including learning"; and finally, "universal intelligence" aspires to the "ability to acquire general intelligence given access to data and resources" (all quotes, 39). A key point in understanding intelligence is its independence of its material carrier: "intelligence doesn't require flesh, blood or carbon atoms" (67), but it may take any shape within a physical or biological environment, as long as it obeys "the laws of physics" (55). In accordance with Bostrom's hint that superintelligence "could be – indeed, it is most likely that they will be – extremely alien" (Bostrom 53), Tegmark prepares his reader for an emergence of super-intelligence in yet unknown non-biological forms. It is this fact which humans may have to pay most attention to, implying an AI's "ability to accomplish any cognitive task at least as well as humans", and which may lead to an artificial "super-intelligence", denoting anything that signifies "general intelligence far beyond human level" (Tegmark 39). However, on the way to this third level of intelligence, there are several uncertainties in the development of technological devices: "Nobody knows for sure what the next blockbuster computational substrate will be, but we do know that we're nowhere near the limits imposed by the laws of physics" (69).

These early reflections in his book on intelligence lead to the extensive discussion about the possibility to design AI tools that have consciousness, or even sentience, i.e., the ability to feel and experience like a human. Outlining the boundaries of accepted empirical and theoretical knowledge, the final chapter stages non-knowledge about consciousness as one of the "thorniest philosophical topics of all" and as "controversial" (281), indeed as "a hopeless waste of time" (287) for serious scientists. First of all, there is "no undisputed correct definition of the word [consciousness]" (283); second, the impossibility to locate it as an organic or physiological state of affairs raises the question, "which parts of your brain *are* responsible for consciousness" (297, emphasis in original); third, the moment when a mere perception becomes conscious is also contested ground (297-98), and finally there are questions about different degrees of consciousness (308). By offering a definition steeped in information technology and particle physics, Tegmark surpasses the disciplinary limits of earlier, futile psychological and neurological attempts to achieve and understanding consciousness. While organic, biological consciousness may be described by physical processes traced back to the interaction of neuronal activity, he introduces the radically different idea of "*physical correlates of consciousness*" (299) as a possible means to understanding synthetic, artificial machine consciousness. Moreover, he defines four abstract principles which are a

minimum of “*necessary* conditions”, but leaves open the question whether they are “*sufficient* to guarantee consciousness” (304):

Principle	Definition
Information principle	A conscious system has substantial information-storage capacity.
Dynamics principle	A conscious system has substantial information-processing capacity.
Independence principle	A conscious system has substantial independence from the rest of the world.
Integration principle	A conscious system cannot consist of nearly independent parts.

Table 1: Tegmark’s “Necessary Conditions for Consciousness” (304)

From these axioms Tegmark infers that “[i]f consciousness is the way information feels when being processed in certain ways, then it must be substrate-independent; it’s only the structure of the information processing that matters, not the structure of the matter doing the information processing” (304). This view enables him to argue that consciousness does not have to be developed from an organic, biological system (such as the human brain), but that it is also possible to develop consciousness in a non-organic form. In sum, the interlocutor transcends the boundaries of Tegmark’s own discipline, theoretical physics, and at the same time steps into the field of those sciences that seek evidence for parallels between biological “consciousness” or “sentience” and their equivalents in non-organic materials. His proverbial thinking outside the box allows him to offer a solution that is not easily accessible, let alone acceptable in any of the conventional disciplinary approaches (psychology, anatomy, philosophy). In contrast to Harari’s multi-discursive but manipulative interlocutor with a habit to stimulate the reader’s imagination, Tegmark clearly points at his subjective viewpoint, frequently using the phrase “I think” (304-05) where he cannot rely on scientific facts but presents his own interpretation of the discussion about consciousness. Science-optimist that he is, Tegmark points to the historicity of the problem, implying that it is a question of time rather than a matter of principle that the mysteries about intelligence and consciousness will eventually be resolved: “the answer depends on time!” (289).

In his lengthy discussion of consciousness, Tegmark invokes a historic, temporary state of individual and disciplinary non-knowledge, which may at some point in the future be overcome and integrated into the (then governing) episteme. As one possible candidate, he mentions the quantum computer, qualifying his statement with another, but more fundamental reservation: to render such a future computer marketable and accessible for a wider public, science still needs to improve, and prove with more reliable evidence than at present that “quantum physics works as we think it does” (70). The hypothesis of substrate-independent manifestations of intelligence, including sub-atomic environments, finally moves into another field of disciplinary non-knowledge in physics: the – to date – incomplete merger of quantum mechanics and relativity theory in a Grand Unified Theory which is key to the development of highly efficient, super-intelligent quantum computer technologies. Effectively, Tegmark’s approach to a material interpretation of intelligence as substrate-independent entity place his ideas in the ancient Atomistic, pre-Socratic tradition of Democritus and Leucippus.

Non-Knowledge and the Role of the Reader

In *Life 3.0* Tegmark, like Harari, practises a highly speculative storytelling technique that informs long passages of the argument, such as the different future scenarios in which AI technologies cooperate or compete with, or even destroy, humanity. However, although both writers share common ground in their motivation to warn their readers of uncontrollable super-intelligent AI systems, as well in the assumption that "[t]he race toward AGI is on, and we have no idea how it will unfold" (Tegmark 161), the preceding case studies have shown that Tegmark's attitude is less pessimistic and alarmist than Harari's firm belief that the human species is set on a one-way course towards an AI-induced apocalypse.

In accordance with Tegmark's less feverish attitude towards AI, *Life 3.0* begins with a prefatory "Tale of the Omega Team" (3-21). It is the utopian parable of a super-intelligent AI system with the telling name "Prometheus" (literal meaning: "one with foresight, visionary", *Britannica*), implemented successively and stealthily controlled by a human syndicate referred to as the Omega team, which leads the world to a supposedly better future by gradually abolishing any of the known political and economic (capitalist or other) structures worldwide. In contrast to Harari's apocalyptic teleology, this utopian tale carries strong chiliastic implications: The Omega plan to "tak[e] over the world" (15), including its economy, media and politics, serves their only goal to "assume the role of a world government" (21), bringing peace, wealth, health, democracy and justice to the world. The bad guys, "a few dictators and others", the narrator reports, "were all toppled in carefully orchestrated coups or mass uprisings" (21). Further into the book, describing an unprecedented (expected but hypothetical) "intelligence explosion" in digital technologies, Tegmark boldly outlines not only "the Next 10,000 Years" (161-201), but even the "Next Billion Years and Beyond" (202-47). In chapter 8 of the main text, this parable and its vision is divided into a sequence of alternative stories: "Prometheus", in keeping with the mythological figure's status as "supreme trickster" (*Britannica*), is sketched in more ambivalent roles and shapes, for instance as a "Libertarian Utopia", the "Benevolent Dictator", or the "Egalitarian Utopia", but also negative ones leading to the AI as "Zookeeper" of humans, as a totalitarian autonomous surveillance system as in Orwell's *Nineteen Eighty-Four*, or as the key to complete "Self-Destruction" of all earthly life-forms. Thus, *Life 3.0* shows how the power of the imagination may be deployed as a legitimate heuristic device that helps envisioning both the chances and challenges of future AI systems – the fictional mode retracts any claim on veracity in what is being said in the tale. If there is no definitive answer to a question, as to how generative AI will develop in the future, the narrative multiplication of imaginary, possible worlds enable a wider range of interpretations, both positive and negative, rather than a fact-based description or a moralistic assessment of AI systems alone.

The preceding comparison has shown that *Homo Deus* adheres to the "deficit model" and the idea of a "literacy monologue" dominated by the author's rather conceited interlocutor, while *Life 3.0* addresses a more participatory reader, who is invited to take influence on the discussion about a further development of AI, and who is expected to be able to assess the potential chances and risks in this process, thus offering to the reader the possibility of an "epistemic dialogue" (all quotes Cortassa 450). Given their different degree of scientific expertise, both authors fashion their interlocutors with a very different set of rhetorical strategies and imaginative mindsets by which they represent various gaps of knowledge, both on the levels of scientific content and of representational discourse.

In fact, both books feature different kinds of non-knowledge formations representating an AI-enhanced world projected into the distant future. Harari, in *Homo Deus*, hides his individual knowledge gaps from the reader with an indirect, image-ridden multidiscursive, often vague, but usually authoritative discourse. Whilst his storytelling does little to fill the knowledge gap between reader and the technological-scientific discourse, Tegmark chooses the opposite strategy for *Life 3.0*. Although his highly prestigious position as a scientist with a whole network of colleagues behind him suggests an even wider knowledge gap between his storyteller and the audience, he repeatedly invites his readers to join (Tegmark 45) the ethical and technological discussion as it was initiated on two conferences, organized by the Future of Life Institute. The first was held in Puerto Rico in January 2015 (35-37 and 320-25), the second followed two years later in Asilomar, CA, culminating in a manifest titled "The Asilomar AI Principles" (329-331). This document outlines the limits and rules for future AI research and development as they seemed feasible at the time of his writing.

In line with the PAS approach in science communication, Tegmark does not address his readership as a uniform mass, but shows his awareness of a multifaceted, diverse and (perhaps all too) science-friendly reader community. He inscribes several reader groups with different AI-backgrounds into his text: at the end of the first chapter he distinguishes expert readers and novices in the field of AI (47), and even considers non-linear reading techniques practised by those "who like[-] skipping around" (46), who may still profit from end-of-chapter summaries with key points (in boxes labelled "the bottom line"). Finally, in the last of his twelve AI-scenarios (titled "What Do *You* Want?", 200-01), he concludes that "it's a mistake to passively ask 'what will happen', as if it were somehow predestined! [...] we should instead ask: 'What *should* happen? What future do we want?'" (159; orig. emphasis). Clearly, he appeals to an (inter-) active reader to take the initiative and participate in the technological, political, legal and educational negotiations that help to establish a consensual agreement over AI and its place in human society: "we humans need to continue and deepen this conversation about our future goals, so that we know in which direction we steer" (201).

Coda

Since the publication of the two books discussed in this article, much innovative research has yielded products which, at the time of their publication, seemed to lie ahead in a distant future. Particularly generative AI has been implemented in many areas of society and, most prominently, become accessible in everyday life to a global public in shape of advanced Large Language Models (LLMs), changing jobs and working conditions in many writing professions (such as journalism, but also academic publishing) for better or worse. In addition, analytic AI has increasingly been used in many controversial security and control systems, but also in such areas as medical diagnosis and in sustainable agriculture, to name but a few. AI has become a significant economic factor and, as predicted by earlier futurologists, it has created new job profiles, while making others obsolete. Moreover, AI has become an ecological factor, demanding vast amounts of energy to run, for instance, LLMs hosted in big computer farms in which the systems are not only trained by experts, but also fed with data by users. Not least, AI has claimed a discursive omnipresence in everyday media, which thus contribute to increase a public awareness about it, although – it may be safely assumed – only a small minority of recipients and users is sufficiently informed about the technological and scientific intricacies of AI.

On a more abstract level, and integrating as well as elaborating many of the Asilomar AI Principles mentioned above, the European AI Act, formally passed in July 2024 (European Union, *OJEU* 2024/1689), is – to date – the first substantial supranational AI-governance instrument of its kind worldwide, classifying four different types of AI according to its respective risk factor for life and society. In this legal framework, the highest risk level itself reflects some further disciplinary non-knowledge relating to existing and future AI apps alike: Not only does it aim at the containment of the developing, using and circulating autonomous AI systems, generative or analytic, that may have an existential impact on humans, either as regulating tools of governance and supervision (such as social scoring systems) or as military weaponry (e.g., autonomously operating drones). It also defines a set of requirements that is meant to devise an efficient, ethically value-based risk-management, including “the quality and relevance of data sets used, technical documentation and record-keeping, transparency and the provision of information to deployers, human oversight, and robustness, accuracy and cybersecurity. Those requirements are necessary to effectively mitigate the risks for health, safety and fundamental rights” (*OJEU* 2024/1689, (66)). Whether this Act is sufficient to guarantee these rights, and whether the underlying humanist ethics will be accepted beyond the range of the European Union is a question that only time will tell, and is itself a matter of disciplinary non-knowledge.

Works Cited

- Adorno, Theodor W., and Max Horkheimer. *Dialectic of Enlightenment*, Verso Books, 2016.
- Anon., "Die meistverkauften Sachbücher des Jahres" ["Non-Fiction Bestsellers of the Year"]. *Der Spiegel*, 31.12.2017.
- "Artificial Intelligence Act." European Union. *Official Journal of the European Union (OJEU)*, Regulation 2024/1689. 12 July 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>. Accessed 30 March 2026.
- Auguscik, Anna, and Simone Broders. "Limits of Knowledge – Knowledge of Limits: The Productiveness of Ignorance, Non-Knowledge, and Agnotology." *Anglistik: International Journal of English Studies*, vol. 33, no. 2, 2022, pp. 77-88.
- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*, Oxford UP, 2014.
- Bouissac, Paul. "The Construction of Ignorance and the Evolution of Knowledge". *University of Toronto Quarterly: A Canadian Journal of the Humanities (UTQ)*, vol. 61, no. 4, 1992, pp. 460-72. <https://doi.org/10.3138/utq.61.4.460>
- "Prometheus (Greek God)." *Encyclopedia Britannica*. February 6, 2026. <https://www.britannica.com/topic/Prometheus-Greek-god>. Accessed 29 December 2025.
- Cortassa, Carina. "In Science Communication, Why Does the Idea of a Public Deficit Always Return? The Eternal Recurrence of the Public Deficit." *Public Understanding of Science*, 2016, Vol. 25, no. 4, pp. 447-459. <https://doi.org/10.1177/0963662516629745>
- Dilley, Roy, and Thomas G. Kirsch (eds.). *Regimes of Ignorance: Anthropological Perspectives on the Production and Reproduction of Non-Knowledge*, Berghahn, 2015.
- Dwivedi, Yogesh K., et al. "Opinion Paper: 'So what if ChatGPT wrote it?' Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy". *International Journal of Information Management*, vol. 71, Aug. 2023, pp. 1-63. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Fröschl, Martina. "The Door of Science Visualization". *Doors to Hidden Worlds: The Power of Visualization in Science, Media and Art*, edited by Alfred Vendl and Martina Fröschl, de Gruyter, 2023, pp. 39-64.
- Gershgorn, Dave. "AI is now so complex its creators can't trust why it makes its decisions". *Quartz*, 7 Dec. 2017. <https://qz.com/1146753/ai-is-now-so-complex-its-creators-cant-trust-why-it-makes-decisions>. Last update 20 July 2022.
- Gregory, Jane, and Steve Miller. *Science in Public: Communication, Culture and Credibility*, Basic Books, 1998.
- Hao, Karen. *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI*, Penguin P, 2025.
- Harari, Yuval Noah. *Nexus: A Brief History of Information Networks from the Stone Age to AI*, Fern Press, 2024.
- . *Homo Deus: A Brief History of Tomorrow*, Vintage, 2016.
- Hesse, Mary B. "Language, Metaphor, and a New Epistemology". *The Construction of Reality*, edited by Michael R. Arbib and Mary B. Hesse, Cambridge UP, 1986, pp. 147-170.

- Hilgartner, Stephen. "The Dominant View of Popularization: Conceptual Problems, Political Uses". *Social Science Studies*, vol. 20, no. 3, 1990, pp. 519-539. <https://doi.org/10.1177/030631290020003006>.
- "Homo Deus: A Brief History of Tomorrow." *Wikipedia*. https://en.wikipedia.org/wiki/Homo_Deus:_A_Brief_History_of_Tomorrow. Accessed 30 March 2026.
- Howorth, Claire. "The Top 10 Non-Fiction Books in 2017". *Time*, 21 November 2017.
- IBM. "What are Transformer Models?" *IBM Think: Tech News, Education and Events*, no date. <https://www.ibm.com/think/topics/transformer-model>
- Japp, Klaus. "Distinguishing Non-Knowledge". *Canadian Journal of Sociology / Cahiers canadiens de sociologie*, vol. 25, no. 2 (Spring), 2000, pp. 225-238.
- Johnson, Samuel. "Life of Cowley". *Samuel Johnson: The Lives of the Poets. A Selection*, edited by Roger Lonsdale, Oxford UP, 2009, pp. 5-53.
- Kantorowicz, Ernst. *The King's Two Bodies: A Study in Political Theology*. With a new introduction by Conrad Leyser, Princeton UP, 2016.
- Kompa, Nikola. "Insight by Metaphor: The Epistemic Role of Metaphor in Science". *Physics and Literature: Concepts – Transfer – Aestheticization*, edited by Aura Heydenreich and Klaus Mecke, de Gruyter, 2021, pp. 23-45.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*, U of Chicago P, 1996.
- Kurzweil, Ray. *The Singularity is Nearer: When We Merge with AI*, Vintage, 2024.
- . *The Singularity is Near: When Humans Transcend Biology*, Penguin Books, 2005.
- Locke, David. *Science as Writing*, Yale UP, 1992.
- Meyer, Jürgen. "... all those pretty pebbles on the shoreline of knowledge': Fashioning the limits in Science Popularizations". *Anglistik: International Journal of English Studies*, vol. 34, no. 3, 2023, pp. 217-233.
- Nappo, Francesco. "Revolutionary Analogies". *Rethinking Thomas Kuhn's Legacy*, edited by Yafeng Shan, Springer, 2024, pp. 229-252.
- Paul, Heike. "Knowledge". *Critical Terms in Future Studies*, edited by Heike Paul, Palgrave Macmillan, 2019, pp. 171-177. https://doi.org/10.1007/978-3-030-28987-4_27
- Proctor, Robert N. "Agnotology. A Missing Term to Describe the Cultural Production of Ignorance (and Its Study)". *Agnotology: The Making and Unmaking of Ignorance*, edited by Robert N. Proctor and Londa Schiebinger, Stanford UP, 2008, pp. 1-36.
- Sundvall, Scott. "Artificial Intelligence". *Critical Terms in Future Studies*, edited by Heike Paul, Palgrave Macmillan, 2019, pp. 29-34. https://doi.org/10.1007/978-3-030-28987-4_6
- Smithson, Michael J. "Social Theories of Ignorance". *Agnotology: The Making and Unmaking of Ignorance*, edited by Robert N. Proctor and Londa Schiebinger, Stanford UP, 2008, pp. 209-229.
- Shapin, Steven. "How to be Antiscientific". *Perspectives on Science*, vol. 3, 1995, pp. 255-275; repr. in *The One Culture? A Conversation about Science*, edited by Jay A. Labinger and Harry Collins, U of Chicago P, 2001, pp. 99-115.
- Snow, Charles Percy. *The Two Cultures*. Introduction by Stefan Collini. Cambridge UP, 1999.
- Tegmark, Max. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Vintage Books, 2017.

- Tegmark, Max et al. "Pause Giant AI Experiments: An Open Letter". *Future of Life Institute*. 22 March 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Webbs, Amy, and Andrew Hessel. *The Genesis Machine: Our Quest to Rewrite Life in the Age of Synthetic Biology*. Public Affairs/Hachette Group Books, 2023.
- Whitley, Robert. "Knowledge Producers and Knowledge Acquirers". *Expository Science: Forms and Functions of Popularization*, edited by Terry Shinn and Robert Whitley, Reidel, 1985, pp. 3-28.
- Woolgar, Steven, and Bruno Latour. *Laboratory Life: The Construction of Scientific Facts*. Introduction by Jonas Salk. With a New Postscript by the authors. Princeton UP, 1987.