Minds and Machines (2024) 34:5
https://doi.org/10.1007/s11023-024-09670-4

Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences

Luciano Floridi[1,2] · **Anna C Nobre**[3]

Published online: 25 April 2024

Abstract
The article discusses the process of "conceptual borrowing", according to which, when a new discipline emerges, it develops its technical vocabulary also by appropriating terms from other neighbouring disciplines. The phenomenon is likened to Carl Schmitt's observation that modern political concepts have theological roots. The authors argue that, through extensive conceptual borrowing, AI has ended up describing computers anthropomorphically, as computational brains with psychological properties, while brain and cognitive sciences have ended up describing brains and minds computationally and informationally, as biological computers. The crosswiring between the technical languages of these disciplines is not merely metaphorical but can lead to confusion, and damaging assumptions and consequences. The article ends on an optimistic note about the self-adjusting nature of technical meanings in language and the ability to leave misleading conceptual baggage behind when confronted with advancement in understanding and factual knowledge.

Keywords Artificial intelligence · Carl schmitt · Cognitive science · Conceptual borrowing · Neuroscience

Artificial intelligence (AI) can be confusing in many ways. The dizzying developments in software and hardware are beyond most of us. But perhaps the deepest source of confusion arises from AI's technical vocabulary. Imbued with terms from

✉ Luciano Floridi
luciano.floridi@yale.edu

[1] Digital Ethics Center, Yale University, 85 Trumbull St., New Haven, CT 06511, USA

[2] Department of Legal Studies, University of Bologna, Via Zamboni, 27/29, Bologna 40121, Italy

[3] Wu Tsai Institute, Department of Psychology, Yale University, 100 College St., New Haven, CT 06510, USA

brain and cognitive sciences (BCS, this includes Cognitive Science and Neuroscience), AI acquires unwarranted biological and cognitive properties that taint its understanding in society. In turn, the scientific disciplines concerned with understanding how the brain supports cognition and behaviour have increasingly borrowed from informational and computational sciences that paved the way for AI, flattening the most complex and perplexing biological entity into mere calculating machines.

AI scientists speak of "machine learning", for example. The term was coined (or perhaps popularised, the debate seems open) by Arthur Samuel in 1959 to refer to "the development and study of statistical algorithms that can learn from data and generalize to new data, and thus perform tasks without explicit instructions".[1] But this "learning" does not mean what brain and cognitive scientists mean by the same term when referring to how humans or animals acquire new behaviours or mental contents, or modify existing ones, as a result of experiences in the environment. Similarly, AI scientists use "hallucinations" to describe errors or deviations in the output of a model from grounded, accurate representations of the input data. These are a far cry from the disturbing perceptual experiences lacking external stimuli (those are *our* hallucinations). As we shall see presently (Table 1), the list continues.

The crosswiring between neuroscientific and computational terms in AI and BCS is problematic beyond just taking some metaphorical liberty. To get to the bottom of the confusion, we need to take a step back and start from an influential idea by Carl Schmitt.

In his classic *Political Theology: Four Chapters on the Concept of Sovereignty* (1922, see now (Schmitt, 2005), Schmitt famously remarks that

> All significant concepts of the modern theory of the state are secularised theological concepts not only because of their historical development—in which they were transferred from theology to the theory of the state, whereby, for example, the omnipotent God became the omnipotent lawgiver—but also because of their systematic structure, the recognition of which is necessary for a sociological consideration of these concepts. (Chap. 3)

For example, political concepts such as "sovereignty", "state of exception" (where normal laws are suspended), "sovereign will", "omnipotence of the law", and "legitimacy" (through historical precedence) can be traced back to theological concepts.[2] Schmitt argues that the secularisation process involved translating theological concepts into political ones. This process of *conceptual borrowing* did not eliminate the structure or influence of theological concepts, but instead recontextualized them into a secular framework. This is not just a historical observation but also a severe critique. Conceptual borrowing diminishes the scrutiny of political concepts because of their well-assimilated theological roots. Modern political concepts have not fully

[1] https://en.wikipedia.org/wiki/Machine_learning. We replaced "unseen" with "new", where "new" means at least new to the machine.

[2] Respectively: divine authority, which in theology has the ultimate power to decide above and beyond the law; the theological concept of "miracle" as an extraordinary event that transcends the natural order as defined by God; God's will; God's omnipotence; and how religious authority is often justified by ancient scriptures and traditions.

emancipated themselves from their theological origins, and the power dynamics and decision-making processes in politics still reflect the structures established in religious thought.

Schmitt's observation was insightful, and the phenomenon of conceptual borrowing can be generalised to other disciplines. When new sciences emerge, they lack the technical vocabulary to describe and communicate their unique phenomena, problems, hypotheses, observations, formulations, theories, etc. There is a pressing need to be precise, clear, consistent, and economical; to agree on definitions, promote standardisation… Yet, unavoidably, the scientific developments outpace the maturing of linguistic conceptualisations. The asymmetry generates a technical vocabulary gap, often filled by inventing new terms – sometimes using Greek or Latin translations and other times adopting and adapting technical terms from other established disciplines.

Science is full of conceptual borrowing. Indeed, a history of science written from a conceptual-borrowing perspective would be fascinating and revealing. It could investigate rhetorical issues (e.g., in the appropriation of scientific language by policy-making), uncover power struggles of "semantic solidifications" (who "owns" which terms and hence controls related concepts, such as "emergence"[3]), and link conceptual borrowing dynamics with critical insights from the social construction of technology theory and conceptual blending in cognitive linguistics. Scientific conceptual borrowing is widespread, happening whenever a new discipline emerges. But, as Schmitt rightly stresses, it is not neutral. Every technical term is part of a network of conceptual structures to which it remains linked, providing contextual constraints and exerting semantic influences and powers. When grafting terms from one discipline to another, these terms, therefore, carry additional baggage and implications. Depending on the alignment and relationship between the disciplines, the baggage can add value, confuse, or misguide.

In some cases, scientific conceptual borrowing can be straightforward and natural. Take the example of how biochemistry inherited its vocabulary from its parent fields – biology and chemistry. In other cases, borrowed terms can take surprising turns in their meaning, such as when the nascent chemistry field drew on the more established alchemy practices. Consider the term "alcohol". It comes from the Arabic "al-kuḥl" (الكحل), which refers to a fine metallic powder, often made from antimony, used as an eyeliner, and obtained through *sublimation*, a term in alchemy referring to the process of transforming a solid directly into a vapour, which then recondenses to form a purified solid. Alchemists ended up associating the term "al-kuḥl" simply with refining or extracting the essence of a substance. Eventually, the meaning narrowed to indicate the "spirit" or "essence" commonly extracted from fermented grain or fruit, what we now understand as ethanol or ethyl alcohol. Today, alcohol is any organic compound with one or more hydroxyl (-OH) groups bound to a saturated carbon atom, with ethanol (drinking alcohol) being the most well-known among them.

We caution that, in the case of conceptual borrowing between AI and BCS, the extra baggage carried by grafted terms has insidious negative consequences. As a newborn discipline studying and engineering successful forms of agency, AI devel-

[3] We are very grateful to Jessica Morley for calling our attention to this point and providing the relevant example, and Claudio Novelli.

oped very quickly compared to other disciplines and needed to borrow its vocabulary from related fields. Cybernetics was available at the time, though, intriguingly, it failed to gain traction as an academic field (Gagliano and Gehl 2008). Cybernetics provided AI with many technical expressions such as "adaptive system", "autonomous agent", "control theory", "cybernetic organism (cyborg)", "feedback loop", "signal processing", and "system dynamics". Indeed, given the scope of AI and its inclusion of some robotics, it may be the rightful heir of cybernetics' technical vocabulary. Other disciplines included logic, computer science, and information theory. We shall come back to them presently. But, most importantly, AI found it helpful to borrow from sciences linked to human and animal agency and behaviour, and their biological footings, most notably cognitive/psychological sciences, and neuroscience.

The phenomenon of AI's conceptual borrowing from BCS has been growing since the work of Alan Turing (Turing, 1950), who influentially drew parallels to human intelligence and behaviour to conceptualise how machines might eventually mimic some aspects of biological cognition. But, perhaps the most problematic borrowing came with the generation of the label of the field itself: "*Artificial Intelligence*". John McCarthy was responsible for the brilliant, if misleading, idea. It was a marketing move, and, as he recounted, things could have gone differently:[4]

> Excuse me, I invented the term 'Artificial Intelligence'. I invented it because we had to do something when we were trying to get money for a summer study in 1956, and I had a previous bad experience. The previous bad experience [concerns, McCarthy corrects himself and says] occurred in 1952, when Claude Shannon and I decided to collect a batch of studies, which we hoped would contribute to launching this field. And Shannon thought that 'Artificial Intelligence' was too flashy a term and might attract unfavorable notice. And so, we agreed to call it 'Automata Studies'. And I was terribly disappointed when the papers we received were about automata, and very few of them had anything to do with the goal that at least I was interested in. So, I decided not to fly any false flags anymore but to say that this is a study aimed at the long-term goal of achieving human-level intelligence. Since that time, many people have quarrelled with the term but have ended up using it. Newell and Simon and the group at Carnegie Mellon University tried to use 'Complex Information Processing', which is certainly a very neutral term, but the trouble was that it didn't identify their field, because everyone would say 'well, my information is complex, I don't see what's special about you'. *The Lighthill Debate* (1973) [Punctuation added for readability purposes].

[4] "In 1973, professor Sir James Lighthill was asked by Parliament to evaluate the state of AI research in the United Kingdom. His report, now called the Lighthill report, criticized the utter failure of AI to achieve its 'grandiose objectives'. He concluded that nothing being done in AI couldn't be done in other sciences. He specifically mentioned the problem of 'combinatorial explosion' or 'intractability', which implied that many of AI's most successful algorithms would grind to a halt on real world problems and were only suitable for solving 'toy' versions. The report was contested in a debate broadcast in the BBC 'Controversy' series in 1973. The debate 'The general purpose robot is a mirage' from the Royal Institute was Lighthill versus the team of Michie, McCarthy and Gregory. The report led to the near-complete dismantling of AI research in England." https://youtu.be/pyU9pm1hmYs?si=Ygt8EhSXgQJpBk6D.

The psychologically permeated terms that followed since artificial "intelligence" have continued to generate problems. Back to our first example. The "learning" in "machine learning" carries the positive value of the original concept and exerts *influence* over the interpretation of the qualities of the computational systems. It also links the concept to other original, equally anthropomorphic concepts such as "unlearning" (Bourtoule et al., 2021). Above all, once you speak of "machine learning", it becomes natural to wonder whether machines can learn – not just metaphorically – but in the biological and psychological sense. One assumes or seeks similarities between machine and human learning, running the risk of under-scrutiny. Indeed, a booming cottage industry is currently exploring how the properties and algorithms of human and machine learning relate, for example, by comparing language abilities in children and large language models. One wonders about the extent to which the endeavour is misguided and derails scientists from exploring the most relevant biological and psychological vs. informational and computational processes within BCS and AI in turn.

Biological and psychological terms in AI are abundant. Table 1 offers some examples other than "machine learning" and "hallucinations.

Today, AI is replete with terms that have technical meanings only vaguely related, if at all, to the precise sense in which they occur in their original scientific context. Consider, for example, "attention", an extremely popular term recently introduced in machine learning (Vaswani et al. 2017) (Table 2). In BCS, the technical term refers broadly to the processes of prioritising neural or psychological signals that are

Table 1 Examples of borrowed terminology in AI

Adaptation - How AI systems modify to accomplish tasks over time better.
BDI (belief-desire-intention) - Architecture designed for logical programming languages for artificial agents' models (belief), goals (desire) and choices (intention).
Computer vision – The field of artificial intelligence enabling computers to acquire and process visual data.
Embodiment - Property of AI that uses physical interaction to ground representations/control.
Emergence - Property of complex, decentralized AI systems whereby "intelligent" behaviour arises from component interactions.
Feature extraction - Techniques for deriving high-level descriptors from raw input data.
Memory - How AI systems store data, enabling retrieval of past computational states/outputs.
Neuron - The basic processing units of artificial neural networks.
Neuroplasticity - The ability of artificial neural networks to change their structures and connections.
Perceptron - A basic neural network unit that performs threshold logic.
Sensorimotor coordination – "Reflexive" AI behaviours connecting perception to action in real-time.
Sensory processing - How early neural layers in AI systems analyse input data.
Stimulus - External inputs to artificial neural networks that activate neurons.
Synapse - The connections between artificial neurons that strengthen or weaken based on signals passed across them.

Table 2 Descriptions of "Attention" in AI and in Cognitive Science in Wikipedia

Artificial intelligence	Cognitive science
Attention is a mechanism, within neural networks, particularly transformer-based models, that "calculates 'soft' weights for each word, more precisely for its embedding, in the context window." *Wikipedia*	*Attention* is the concentration of awareness on some phenomenon to the exclusion of other stimuli. It is a process of selectively concentrating on a discrete aspect of information, whether considered subjective or objective.' *Wikipedia*

Table 3 Examples of BCS' technical vocabulary borrowed from information theory and computer science

Architecture - Overall design and organization principles of neural systems.
Capacity - The maximum amount of information that can be coded by a neural population.
Channel - The conduit transmitting information between brain regions (e.g., axonal pathways).
Circuit - Specific interconnected pathways underlying functions like vision or movement.
Coding - Representing information via distinct patterns of neural activity.
Control processes - mechanisms by which cognitive processes regulate information processing.
Decoding - Figuring out what information is encoded in observed neural activity patterns.
Encoding - The process by which sensory inputs are transformed into neural representations.
Filtering - Network-level mechanisms that regulate information flow and streams.
Information - The meaningful content carried by spike trains and neural population responses.
Information processing - cognitive processes involved in perception, learning, memory, and decision-making as analogous to the processing of information in a computer.
Modulation - Neural tuning properties that imbue activity with diverse signals.
Multiplexing- Encoding multiple streams of data into a single communication channel.
Parallel processing The ability of the brain to analyse or solve problems using many concurrent pathways.
Sampling - Methods to estimate neural population characteristics from limited measurements.
Signal-to-noise ratio - Measure of neural fidelity that depends on reliability versus variability.
Synchronization - Temporal coordination of activity within and between brain regions.
Transmission - Propagation of signals between neurons and brain areas.

relevant to guide adaptive behaviour within the current context (Nobre & Kastner, 2014) and is often preceded by further qualifiers (e.g., selective, spatial, object-base, feature-base, cross-modal, or temporal attention). The meaning in machine learning differs dramatically, as attested even by the Wikipedia entries (Table 2). It is a case of polysemy,[5] if not of homonymy:[6] the scientific differences between the two concepts are profoundly significant, the similarities superficial and negligible. The superficial similarities in the definitions are also insignificant, yet the psychological and biological baggage exerts alluring semantic power that pushes hard toward more anthropomorphism. The ability of AI systems to pay attention, learn, and hallucinate… further fuels AI projects, research programs, and business strategies. Unfortunately, but unsurprisingly, this leads to recurrent "AI winters" (Floridi, 2020)

The term "Artificial Intelligence" – and the extensive conceptual borrowing to establish the field of studies to which it refers as an academic discipline – are problematic, not only for all the reasons highlighted by Schmitt and for the confusion that they keep generating, but also because of the semantic crosswiring with the emergence and co-development of BCS, engaged in their own conceptual borrowing.

As they rapidly advanced, BCS borrowed the technical and quantifiable constructs from information theory and computer sciences, framing the brain and mind as *computational* and *information processing systems*. For example, in the influential book launching Cognitive Psychology as a distinctive new field, (Neisser, 1967) states that the "task of a psychologist trying to understand human cognition is analogous to that of a man trying to discover how a computer has been programmed. In particular, if the program seems to store and reuse information, he would like to know by what "routines" or "procedures" this is done." (p. 6). Table 3 provides some telling examples of terms borrowed by BCS. In many ways, the enterprise has been highly successful, providing a scientific and empirical hold for investigating the properties and biological basis of the most elusive of entities – the subjective human mind. However, sometimes it may go too far. For example, it is not uncommon for computational neuroscientists to use ingenious analytical and imaging methods to identify brain areas, tracking the values of variables in computational operations attributed to brain circuits (e.g., in reinforcement learning or Bayesian models), as if the brain were really running these computational functions mathematically.

The overall result is an impoverished reductionist view in which the subjective qualities of the mind are more sidestepped than understood. For example, patterns of brain activity required for, or that correlate with, psychological phenomena are taken as sufficient explanations. The vivid, experiential contents of our minds are flattened into sustained activations or functional states in neuronal populations, and the moment of willed choices are reduced to firing rates or activation levels reaching a decision boundary.

Today, the two lines of conceptual borrowing have led AI to speak anthropomorphically about machines and algorithms that are not intelligent, and brain and

[5] Polysemy occurs when a single word has multiple meanings, remotely related, that can be disambuigated by context, like "table" (furniture or organised data).

[6] Homonymy occurs when two distinct words share the same spelling or pronunciation, but have unrelated meanings, like "bank" as a sloping ground alongside a river, and "bank" as a business.

cognitive sciences to reduce intelligent biological agents to mere informational and computational systems. The short circuit between the two vocabularies was inevitable. The situation generates confusion in those who do not know better and believe that AI is intelligent, in those who know better but have faith that AI will create some super-intelligent systems, and in those who may or may not know better but do not care and exploit the confusion for their purposes and interests, often financial. Some of the support for a sci-fi kind of AI is not just the outcome of an anthropomorphic interpretation of computational systems but also of a very impoverished understanding of minds.

What can be done to tackle this conceptual mess? Probably nothing in terms of linguistic reform. Languages, including technical ones, are like immense social currents: nobody can swim against them successfully, and they cannot be contained or directed by *fiat*. AI and BCS will keep using their terms, no matter how misleading they may be, how many resources they will make one waste, and how much damage they may cause in the wrong hands or contexts. AI will continue to describe a computer as an artificial brain with mental attributes – attending, learning, memorising, reasoning, and understanding information; brain and cognitive sciences will continue to flatten the brain and mind into a biological computer – encoding, storing, retrieving, processing, and decoding signals through input-output mechanisms.

However, linguistic history itself offers reasons to be optimistic. Better understanding and more facts shape the meaning of words and improve how they are used. Even the strongest current must bend when it encounters new obstacles. For example, we still use expressions like "sunrise" ("the sun rises") and "sunset" ("the sun sets") even if nobody (well, probably almost nobody) believes that the sun goes anywhere with respect to our planet. The geocentric model has long been abandoned. Language has kept the expressions but upgraded the meanings.

Let us close this article with an analogy that offers reasons to be optimistic. In the late 18th century, the Scottish inventor James Watt was instrumental in developing and improving the steam engine during the Industrial Revolution. To enlist new customers, he needed to show how the engine outperformed horse labour. So, he measured the work done by draft horses in coal mines. He observed that a mining horse could turn a mill wheel once every minute, lifting approximately 33,000 pounds by one foot in one minute, and thus defined the standard unit of one horsepower as moving 550 foot-pounds per second. The conceptual borrowing worked, and "horsepower" was universally adopted to quantify steam engine power relative to animal labour. Today, horsepower remains the standard unit to measure an engines' mechanical power output. Of course, nobody is looking for hooves and manes inside an engine. So, there is hope. One day, if we are lucky, people will treat AI more like HP and stop looking for the cognitive or psychological properties inside informational and computational systems.

Acknowledgements Many thanks to Emmie Hine, Claudio Novelli, and Jessica Morley, for their very helpful feedback on previous versions of this article.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

References

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., Lie, D., & Nicolas Papernot. (2021). and. Machine unlearning. In *2021 IEEE Symposium on Security and Privacy (SP)*, 141 – 59. IEEE.

Floridi, L. (2020). AI and its New Winter: From myths to realities. *Philosophy & Technology*, *33*, 1–3.

Gagliano, R. (2008). and John Gehl. 'Whatever happened to cybernetics?', *Ubiquity*, 2008: Article 3.

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts: New York).

Nobre, A. C., & Kastner, S. (2014). *The Oxford handbook of attention*. Oxford University Press: Oxford; New York).

Schmitt, C. (2005). *Political theology: Four chapters on the concept of sovereignty*. University of Chicago Press: Chicago).

Turing, A. M. (1950). 'Computing machinery and intelligence', *Mind*, 59: 433 – 60.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Aidan, N., & Gomez Łukasz Kaiser, and Illia Polosukhin. 2017. 'Attention is all you need'. *Advances in Neural Information Processing Systems*, 30.