TAFHIM: IKIM Journal of Islam and the Contemporary World 17 No. 1 (June 2024): 23–55

Philosophical Underpinnings of Artificial Intelligence and the Concept of Human Soul in Islam: Some Issues at the Interface

Juris Arrozy*
jurisarrozy21@gmail.com

Wendi Zarman**
wendizar@gmail.com

DOI: http://doi.org/10.56389/tafhim.vol17no1.2

Abstract
Can machines think like humans do? Is the mind essentially a physical entity like a machine? A particular view on the nature of the human mind and intelligence, i.e. the computational theory of mind, appears either to be the basis of, or to

* Researcher at Institut Pemikiran Islam dan Pembangunan Insan (PIMPIN) Bandung, Indonesia. He obtained his bachelor and master's degrees in Electrical Engineering from Institut Teknologi Bandung, Indonesia in 2017 and 2018, respectively. In 2019, he started working as a PhD researcher at the Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands. Aside from his formal research in the field of power electronics, his other research interest includes the relationship between Islam and modern science & technology. He is the main author of the paper.

** Director of Institut Pemikiran Islam dan Pembangunan Insan (PIMPIN) Bandung and lecturer at Universitas Komputer (UNIKOM), Indonesia. He obtained his bachelor and master's degrees in Physics from Institut Teknologi Bandung, Indonesia in 1999 and 2002, respectively. In 2012, he completed his doctoral degree in Islamic Education from Universitas Ibnu Khaldun, Bogor, Indonesia. He is the supervisor of the first author.

have taken inspiration from, the field of Artificial Intelligence. The theory aims to explain intelligence by only resorting to physical explanation. This might run counter to the worldview of Islam since Islam acknowledges the existence of a spiritual substance, i.e., the *nafs* (soul), in addition to the body and affirms its role in explaining human intelligence. Therefore, this article discusses some issues stemming from the interface between the philosophical underpinnings of Artificial Intelligence and Islam. Two strands of the computational theory of mind, i.e. strong symbol system hypothesis and connectionism, are elaborated. The Islamic conception of the human soul is adopted from the works of Syed Muhammad Naquib al-Attas. It is shown that whereas the computational theory of mind regards intelligence and knowledge as purely physical, the Islamic conception of the human soul argues that the soul and the knowledge imprinted upon it are non-physical. The disagreement is further by analysing examples from the current AI limitations: "adversarial examples" in visual abstraction, syntax-semantics distinction, and abduction as a "leap" in reasoning.

Keywords
Artificial Intelligence, computational theory of mind, strong symbol system hypothesis, connectionism, *nafs*, *rūḥ*, *ʿaql*, *qalb*.

Introduction

Can machines think like humans do? Is the mind essentially a physical entity like a machine? From mythology to natural philosophy, the imagery of thinking machines has been one of the most recurring symbolism for life and intelligence.[1] For

1. Adrienne Mayor, *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology* (Princeton: Princeton University Press, 2018), 1–3; and Minsoo Kang, *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Cambridge: Harvard University Press, 2011), 116–132.

instance, the ancient Greek mythical figure Talos is described as an "animated metal machine in the form of a man, *able to carry out complex human-like actions*…"[2] Other mythologies such as Prometheus' first human and Pandora also exhibit the theme of "made, not born" intelligence.[3]

In *Leviathan*, Thomas Hobbes (1588–1679) resorts to the imagery of machines when describing human life. He writes:

> For seeing life is but a motion of limbs, the beginning whereof is in some principal part within; why may we not say, that all *automata* (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the *heart*, but a *spring*; and the *nerves*, but so many *strings*; and the *joints*, but so many *wheels*, giving motion to the whole body, such as was intended by the artificer?[4]

He also views reasoning as computation[5]—a claim that earned him title "grandfather of AI."[6] The imagery of machines is more explicit in Julien Offray de La Mettrie's (1709–1751) *Man a Machine*, where he argues that thought very much depends on "specific organisation of the brain and of the whole body,"[7]

---

2. Ibid., 7. Emphasis in italics are mine.
3. Ibid., 1.
4. Thomas Hobbes, *Leviathan*, ed. John C.A. Gaskin (Oxford: Oxford University Press, 1998), 7.
5. William Molesworth, ed., *The Collected Works of Thomas Hobbes*, 12 vols. (London, Routledge, 1992), 1:3. Hobbes writes "By ratiocination, I mean *computation*. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another. Ratiocination, therefore, is the same with addition and substraction." Hobbes does not seem to limit computation (combination of addition and substraction) only to numbers. In *Leviathan*, he also writes "These operations [addition and substraction] are not incident to numbers only, but to all manner of things that can be added together, and taken one out of another." See Hobbes, *Leviathan*, 27.
6. John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge: The MIT Press, 1985), 23.
7. Julien Offrey de La Mettrie, *Machine Man and Other Writings*, trans. and ed. Ann Thomson (Cambridge: Cambridge University Press, 1998), 26. See also the editor's introduction on page xiv.

effectively aligning with the imagery of machines[8] and dismissing the soul as "merely a vain term of which we have no idea."[9] He is depicted as "[a] radical materialist taking the final logical step of jettisoning the notion of an immaterial, transcendent soul in man and turning him into an organic automaton and *nothing* more."[10]

As Karl Popper (1902–1994) later notes, "de La Mettrie's doctrine that man is a machine has today perhaps more defenders than ever before among physicists, biologists, and philosophers; especially in the form of the thesis that *man* is a *computer*."[11] It has been observed that there is a tendency among scientists and philosophers to model the operation of the brain on the most fashionable technology of the day.[12] Therefore, the field of Artificial Intelligence (AI) is arguably the best place to set the context today.[13]

Since its inception, AI is progressing and even in some cases dominating intellectual domains that were previously regarded as human-exclusive. Chess and Go are good examples of intellectually demanding games where current AI can outperform humans[14], despite both being previously described as too complex

8. Ibid. The full text is written as follows: "But since all the soul's faculties depend so much on the specific organisation of the brain and of the whole body that they are clearly nothing but that very organisation, *the machine is perfectly explained!*" Emphasis in italics are mine.
9. Ibid.
10. Kang, *Sublime Dreams*, 130.
11. Karl R. Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford: Clarendon Press, 1979), 224.
12. Jack Copeland, *Artificial Intelligence: A Philosophical Introduction* (Oxford: Blackwell Publishers), 182.
13. When describing the thesis "man is a computer," Popper made a reference to Alan Turing's paper on computing machinery and intelligence, which is arguably one of the most influential forerunner in the field of AI. See Alan Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 434–460.
14. In 1997, IBM Deep Blue defeated the world chess champion Garry Kasparov by 3.5–2.5. In 2016, DeepMind AlphaGo defeated the world Go champion Lee Sedol by 4–1.

to be mastered by a machine and requiring human expertise.[15] Machine translation (such as Google Translate) can translate sentences across languages with arguably decent performance. ChatGPT released by OpenAI in 2022 can seemingly understand human language and provide reasonable responses. AI-powered self-driving cars like Waymo and Tesla are already available on the market. All these achievements seem to give the impression that machines can produce intelligent behaviour. It is even claimed by some that human-level AI is possible within the foreseeable future.[16] Others speculate even further on the possibility of superintelligence.[17]

Appearing as a basis of or having taken inspiration from the field of AI, there are some attempts to explain the phenomena of (human) intelligence[18] as a byproduct of computational systems. This is the stance of the computational theory of mind—a position that views the mind as a computational system.[19] In this regard, human intelligence is just one implementation

15. For such typical comments about chess and Go, see for example: Hubert L. Dreyfus and Stuart E. Dreyfus, *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (New York: Free Press, 1986), 49; Alan Levinovitz, "The Mystery of Go, the Ancient Game That Computers Still Can't Win," last modified May 13, 2014, https://www.wired.com/2014/05/the-world-of-computer-go/.
16. For surveys of expert opinions regarding this matter, see for example Seth D. Baum et al., "How Long until Human-level AI? Results from An Expert Assessment," *Technological Forecasting and Social Change* 78 (2011): 185–195; Martin Ford, *Architects of Intelligence: The Truth About AI From the People Building It* (Birmingham: Packt Publishing, 2018), 528–529.
17. Bostrom tentatively defines superintelligence as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest." See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014), 26. A similar concept is also proposed earlier by Irving John Good defines ultraintelligent machine as "a machine that can far surpass all the intellectual activities of any man however clever." See Good, "Speculations Concerning the First Ultraintelligent Machine," in *Advances in Computers*, ed. Franz L. Alt and Morris Rubinoff (New York: Academic Press, 1965), vol. 6, 31–88.
18. In this paper, the term "intelligence" is broadly understood by its colloquial usage as the ability to perform cognitive actions such as visual perception, language, mental abstraction, logic, understanding, etc.
19. For good overviews of the computational theory of mind, see Michael Rescorla, "The Computational Theory of Mind," September 21, 2020, https://plato.stanford.edu/entries/computational-mind/.

of the broader "laws" governing any intelligent agents such as humans, animals, computers, etc. Consider these two statements in the Dartmouth: "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence *can in principle be so precisely described that a machine can be made to simulate it*;"[20] and by Allen Newell (1927–1992) and Herbert Simon (1916–2020): "A physical symbol system *has the necessary and sufficient means* for general intelligent action."[21] According to the computational theory of mind, intelligence can emerge out of a purely physical system and does not require any non-physical explanation.[22] This might run in contrast with the Islamic worldview since Islam acknowledges the existence of a spiritual substance, i.e., the *nafs* (soul),[23] in addition to the existence of the body and affirms its role in explaining human intelligence.[24] Moreover, the *rūḥ* (spirit) is something about which humans are given little knowledge.[25]

However, to the author's best knowledge, an extensive treatment of the issues from the Islamic standpoint is still scarce. For example, articles written by Amana Raquib et al. are focused on the ethical concerns of AI, therefore making little reference

20. See John McCarthy et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," *AI Magazine* 27, no. 4 (2006): 13. Emphasis in italics are mine. This proposal is widely considered to be the starting point of AI as a field. John McCarthy also first coined the term 'Artificial Intelligence' here.
21. Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the Association for Computing Machinery* 19, no. 3 (1976): 116. Emphasis in italics are mine.
22. According to John Haugeland, resorting to the immaterial souls in explaining intelligence would rule out AI from the start. It must be assumed first that human intelligence is (or at least could be) realised in matter—such as the brain. See Haugeland, *Artificial Intelligence: The Very Idea*, 256.
23. Also known as *rūḥ* (spirit), *'aql* (intellect), or *qalb* (heart) depending on the state of such a spiritual substance. See Syed Muhammad Naquib Al-Attas, *Prolegomena to the Metaphysics of Islām: An Exposition of the Fundamental Elements of the Worldview of Islām* (Kuala Lumpur: ISTAC, 1995), 148.
24. Ibid. The term *'aql* is used when the human soul is involved in intellection and apprehension.
25. *Sūrat al-Isrā'* (17):85.

to the implications of AI on the nature of intelligence and the soul.[26] Articles by Mahmoud Dhaouadi,[27] Hamza Tzortzis,[28] and Osman Bakar[29] address some philosophical implications of AI from the Islamic perspective. However, the current technological achievements of AI are not addressed and the difference between the computational theory of mind and the Islamic conception of the human soul in viewing the nature of intelligence, knowledge, and other related issues such as perception, language, and inference is not elaborated further.

Thus, this article discusses some issues stemming from the interface between the philosophical underpinnings of Artificial Intelligence and Islam. Two strands of the computational theory of mind, i.e., the strong symbol system hypothesis and connectionism, are elaborated. The Islamic conception of the human soul is adopted from the works of Syed Muhammad Naquib al-Attas. It is shown that whereas the computational theory of mind regards intelligence and knowledge as purely physical, the Islamic conception of the human soul argues that the soul and knowledge imprinted upon it are non-physical. The disagreement is also exemplified by analysing examples from the current AI limitations: (1) "adversarial examples" in visual abstraction, (2) syntax-semantics distinction, and (3) abduction as a "leap" in reasoning.

---

26. Amana Raquib, Bilal Channa, Talat Zubair, and Junaid Qadir, "Islamic Virtue-based Ethics for Artificial Intelligence," *Discover Artificial Intelligence* 2, no. 11 (2022); Talat Zubair, Amana Raquib, and Junaid Qadir, "Combating Fake News, Misinformation, and Machine Learning Generated Fakes: Insight's from the Islamic Ethical Tradition," *ICR Journal* 10, no. 2 (2019): 189–212.
27. Mahmoud Dhaouadi, "An Exploration into the Nature of the Making of Human and Artificial Intelligence and the Qur'ānic Perspective," *American Journal of Islam and Society* 9, no. 4 (1992): 465–481.
28. Hamza Andreas Tzortzis, "Does Artificial Intelligence Undermine Religion?" last modified June 30, 2020, https://sapienceinstitute.org/does-artificial-intelligence-undermine-religion/.
29. Osman Bakar, "The Clash of Artificial and Natural Intelligences: Will It Impoverish Wisdom?," in S Abdallah Schleifer (ed.), *The Muslim 500: The World's 500 Most Influential Muslims 2023* (Amman: The Royal Islamic Strategic Studies Centre, 2023), 218–222.

AI and the Computational Theory of Mind (Computational Theory of Mind)

This section focuses only on selected works that are relevant to the computational theory of mind as the philosophical underpinnings of AI. Two branches of computational theory of mind, namely the strong symbol system hypothesis (SSSH) and connectionism, are emphasised. This is due to the former's relevance to the dominant AI paradigm in its early days, i.e., symbolic AI,[30] and the latter's relevance to the current dominant AI paradigm, i.e., artificial neural networks (ANN).[31]

In 1936, Alan Turing (1912–1954) proposed an imaginary machine later called the "Turing machine."[32] The Turing machine is an imaginary computer capable of performing symbol manipulations and storing memory. Interestingly, by "computer," he means a *human* computer performing the procedures or algorithm—already drawing the parallel of human and machine.[33] Later, his 1950 paper[34] deals with the possibility of mechanising all of human intelligence[35] where he also devised a test to answer

30. Keith Frankish and William M. Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence* (Cambridge: Cambridge University Press, 2014), 43. Alternatively, SSSH is also known as physical symbol system hypothesis or classical computational theory of mind.
31. Matt Carter, *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence* (Edinburgh: Edinburgh University Press, 2007), 187, 199–200.
32. Alan Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, *2nd Series* 42 (1936): 544–546.
33. Alan Turing is claimed to grow up with the notion that the human body is a machine. It is even argued that the casual association of machine and human is characteristic of Turing's work. See Charles Petzold, *The Annotated Turing: A Guided Tour through Alan Turing's Historic Paper on Computability and the Turing Machine* (Indianapolis: Wiley Publishing, 2008), 57, 61, and 68. For an in-depth account on the biography of Alan Turing, see Andrew Hodges, *Alan Turing: The Enigma* (New York: Simon & Schuster, 1983).
34. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433–460.
35. Nils J. Nilsson, *The Quest of Artificial Intelligence* (New York: Cambridge University Press 2010), 37.

the question "can a machine think?" called the "imitation game."[36] Turing truly believes in the idea of thinking machine, claiming that "I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking *without expecting to be contradicted*."[37]

The parallel drawn between humans and computers in terms of symbol manipulation finds shelter under the umbrella of the strong symbol system hypothesis (SSSH). Briefly defined, SSSH is a view that only universal symbol systems are capable of thought.[38] The hypothetical Turing machine is shown to be capable of symbol manipulation. Conversely, it is also possible to fully formalise the process of human thought with formal logic. Thus, it is no surprise that symbol systems and symbol manipulation become the core of SSSH and symbolic AI research. This view is already implied in the Dartmouth proposal that states "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."[39] It is described further in John McCarthy's (1927–2011) paper titled *Programs with Common Sense*,[40] where he attempts to utilise first-order logic to represent

36. Also known as the Turing test. One (easier to explain) version of the test is described as follows: imagine one person interrogating two person, all in separate rooms. However, one of the interrogated person is actually a machine/computer. If the machine is able to fool the interrogator into thinking that it is a person, the machine is said to pass the Turing test and therefore sufficiently able to exhibit some kind of intelligence/ intelligent traits.
37. Turing, "Computing Machinery and Intelligence," 442. Emphasis in italics are mine.
38. Jack Copeland, *Artificial Intelligence: A Philosophical Introduction*, 82 and 180. It is differentiated from the softer version of the hypothesis called symbol system hypothesis (SSH) that views it is *possible* to construct a universal symbol system capable of thinking. SSH only considers the possibility of universal symbol system being able to think whereas SSSH views that only universal symbol system is capable of thinking.
39. McCarthy et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," 13.
40. McCarthy, "Programs with Common Sense," *Symposium on Mechanization of Thought Process* (1958), 1–15.

information in a computer. Allen Newell and Herbert Simon share the same spirit with SSSH in their paper *Computer Science as Empirical Inquiry*, stating: "A physical symbol system *has the necessary and sufficient means* for general intelligent action."[41]

If SSSH is conceived, then what really matters in intelligence is not its building blocks such as biological neurons, silicon transistors, electromagnetic relays, etc., but rather what is called the symbol token. Consequently, it is perceived that intelligence is multiply realisable/substrate independent.[42] This idea is apparent in Herbert Simon and Allen Newell's paper *Information Processing in Computer and Man*. By seeing thought as information processing à la the symbol manipulations, Simon and Newell conceive that:

> since the thinking human being is also an information processor, it should be possible to study his process and their organisation independently of the details of the biological mechanisms—the "hardware"—that implement them.[43]

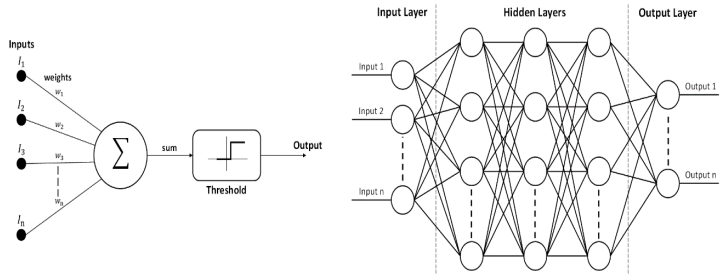
They also propose three propositions in the paper, one of them is that information-processing theories of human thinking can be formulated in computer programming languages and can be tested by simulating the predicted behaviour with computers.[44]

Connectionism departs from a different starting point from SSSH. Instead of relying on symbol manipulations, connectionists take inspiration from the human brain and attempt to recreate it artificially, i.e., artificial neural networks (ANN). This is initiated by Warren McCulloch (1898–1969)

41. Newell and Simon, "Computer Science as Empirical Inquiry: Symbols and Search," 116. Emphasis in italics are mine.
42. Copeland, *Artificial Intelligence: A Philosophical Introduction*, 81; and Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (New York: Penguin Books, 2018), 58.
43. Simon and Newell, "Information Processing in Computer and Man," *American Scientist* 2, no. 3 (1964): 281.
44. Ibid., 282.

and Walter Pitts (1929–1969) in 1943, where they propose a simple mathematical model of a neuron unit as a weighted sum of multiple inputs which produces a binary output (0/1) depending on whether the threshold is fulfilled or not (Figure 1 left).[45] They also showed that networks of such units could construct any Boolean operation (and, or, not) and thus could construct any possible computation.[46] In 1958, Frank Rosenblatt (1928–1971) uses the model to explain how a biological system can store and process information from the physical world, with an emphasis on visual perception.[47]

Figure 1 (left) a simple mathematical model of an artificial neuron; (right) multilayer ANN architecture.



Unlike SSSH and symbolic AI, in connectionism knowledge is encoded in the weights and threshold of a neuron unit and not in the relation between symbols.[48] Thus, a multilayer ANN (Figure 1 right) can hold more knowledge since there

45. Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5 (1943), 115–133.
46. Frankish and Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence*, 16.
47. Frank Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organisation in the Brain," *Psychological Review* 65, no. 6 (1958): 386–408.
48. Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (London: Pelican, 2019), 17.

are more interactions and interconnections between neuron units. It turned out to have contributed to the breakthrough of deep learning in many AI tasks, the most notable being image recognition.[49] A review article about deep learning in 2015 documented the success of deep learning in tackling some of the most complicated challenges such as image recognition and natural language understanding.[50]

Therefore, the two main strands of the computational theory of mind—SSSH and connectionism—embark from a pure physicalist view of intelligence. SSSH accentuates the primacy of a physical symbol system capable of symbol manipulations as the core of intelligence. Meanwhile, connectionism, having taken inspiration from the human brain, views intelligence as weighted interactions and interconnections between neuron units. These are deemed sufficient in explaining (human) intelligence according to their proponents.

The Concept of the Human Soul: An Islamic Perspective

The existence of the human soul is well-corroborated in the Islamic tradition, as indicated by both the scriptural evidence and Islamic scholars' exposition. For example, it is well noted in the Qur'ān that the Divine *spirit* (*rūḥ*) is breathed into human at the moment of creation.[51] In another well-known verse, it is said that humans are given limited knowledge about the spirit.[52]

The soul's preoccupation with the activity of thinking is also indicated by numerous verses. It is mentioned that "Your Lord knoweth best what is in your *hearts* (*nufūs*, plural of *nafs*)…"[53]

49. Ibid., 72–73; and Alex Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems* 25 (2012): 1097–1105.
50. Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep Learning," *Nature* 521 (2015): 436–444.
51. *Sūrat al-Ḥijr* (15):29; *Sūrat al-Sajdah* (32):9; and *Sūrat Ṣād* (38):72.
52. *Sūrat al-Isrā'* (17):85.
53. *Sūrat al-Isrā'* (17):25.

Another term for the heart, i.e., *qalb*, is also associated with the activity of thinking, as shown in the verse "…so their *hearts* (*qulūb*, plural of *qalb*) [and minds] may thus learn wisdom…"[54] The term intellect (*ʿaql*) also appears as a verb *(ʿaqala-yaʿqilu*) in various verses, emphasising those who (do not) understand (*yaʿqilūn*) the sign of God.[55]

The conceptualisation of the human soul in Islam is also explained by numerous scholars such as Abū Zayd al-Balkhī (850–934), al-Farābī (870–950), Ibn Sīnā (980–1037), al-Ghazālī (1058–1111), and Abū al-Barakāt al-Baghdādī (1080–1165)—just to mention a few. Already in the 9th century, al-Balkhī states that man is composed of body and soul.[56] The soul is also busy with thinking, memories, and reflection.[57] Al-Farābī explains how the sensitive, imaginative, and rational faculties of the soul contribute to the perception of forms (*ṣūrah*; pl: *ṣuwar*) and intelligible (*maʿqūlāt*) in order to attain knowledge.[58] Ibn Sīnā gives an extensive treatment on the aspects of the human soul such as its faculties, perceptual processes, the levels of intellect, etc.[59]—which is later followed by al-Ghazālī.[60] Al-Baghdādī discusses other topics related to the soul such as its proof of existence, its immateriality, and its relation to the body.[61]

---

54. *Sūrat al-Ḥajj* (22):46.
55. For example, see *Sūrat al-ʿAnkabūt* (29):43; *Sūrat al-Baqarah* (2):171; and *Sūrat al-Rūm* (30):24.
56. Malik Badri, *Abū Zayd al-Balkhī's Sustenance of the Soul: The Cognitive Behavior Therapy of a Ninth Century Physician* (London: The International Institute of Islamic Thought, 2013), 28.
57. Ibid., 60.
58. Osman Bakar, *Classification of Knowledge in Islam* (Cambridge: The Islamic Texts Society, 1998), 48–64.
59. See Fazlur Rahman, *Avicenna's Psychology* (Westport: Hyperion Press, 1952).
60. According to al-Attas, al-Ghazālī's exposition of the human soul in *Maʿārij al-Quds* is largely derived from the *Kitāb al-Najāt* and *Kitāb al-Shifāʾ* by Ibn Sīnā, although he also added important modifications of his own. See al-Attas, *Prolegomena*, 167.
61. See Wan Suhaimi Wan Abdullah, *Abū al-Barakāt al-Baghdādī on the Human Soul: An Exposition of Some Major Problems of Psychology* (Kuala Lumpur: Pertubuhan Pendidikan Futuwwah, 2021).

In the contemporary setting, Syed Muhammad Naquib al-Attas is one of the scholars that has extensively restated and recontextualised the Islamic conception of the human soul, especially in The Nature of Man and the Psychology of the Human Soul and On Justice and the Nature of Man.[62] Al-Attas' exposition is chosen as the framework of the study for the following reason: (1) his elucidation of the Islamic conception of the human soul is in alignment with the scholars mentioned above, thus making him a good contemporary representative of the Islamic tradition;[63] (2) his recontextualisation makes his works suited to address modern discourse such as AI; and (3) his unique definition of knowledge as "the arrival of meaning in the soul" and "the arrival of the soul at meaning"[64] serves as an important framework in analysing the nature of human and artificial intelligence and its relation to knowledge.

According to al-Attas, man has a dual nature. He is at once body and soul, physical being and spirit.[65] Man is then defined as a "rational animal" (*ḥayawān nāṭiq*). The term "rational" (*nāṭiq*) points to an innate faculty of knowing that apprehends the meaning of the universals and that formulates meaning, which involves judgment, discrimination, clarification, and distinction.[66] The articulation of symbolic forms into meaningful patterns is the outward, visible and audible expression of the inner, unseen reality called the intellect (*'aql*).[67]

62. Al-Attas, *The Nature of Man and the Psychology of Human Soul: A Brief Outline and Framework for an Islamic Psychology and Epistemology* (Kuala Lumpur: ISTAC, 1990). This monograph is included in the fourth chapter of his *Prolegomena*; and Idem, *On Justice and the Nature of Man: A Commentary on Sūrah al-Nisā' (4): 58 and Sūrah al-Mu'minūn (23): 12–14* (Kuala Lumpur: IBFIM, 2015).
63. In fact, al-Attas' exposition in the *Prolegomena* is largely derived from the works of al-Ghazālī and Ibn Sīnā.
64. Al-Attas, *Prolegomena*, 133.
65. Ibid., 143.
66. Ibid., 121–123. See also Idem, *On Justice*, 32; Muhammad Zainiy Uthman, *al-Attas' Psychology* (Kajang Selangor: Akademi Jawi Malaysia, 2022), 34.
67. Ibid., 122.

Al-Attas also affirms that all knowledge ultimately comes from God.[68] However, the soul is not merely a passive recipient but instead an active one.[69] Therefore, al-Attas defines knowledge as both "the arrival of meaning in the soul" and "the arrival of the soul at meaning."[70] The seat of knowledge in man is a spiritual substance which is variously referred to in the Holy Qur'ān sometimes as his heart (*al-qalb*), or his soul or self (*al-nafs*), or his spirit (*al-rūḥ*), or his intellect (*al-ʿaql*).[71]

Al-Attas further illustrates how the human intellect plays part in the perception and acquisition of knowledge. The human soul has perceptive power, which is actualised by the external senses and internal senses.[72] The external senses are responsible for the reception of particulars whereas the internal senses are instrumental in the perception and intellection of universals.[73] The human intellect through internal senses participates in the process of abstraction from sensible particular entities, of the intelligible universals, the entirety of which is described by al-Attas as "an epistemological process towards the arrival at meaning."[74]

Al-Attas also notes that the intellect is not in possession of intelligible realities, but rather intelligible forms (which is a reflection of the former). It is the Active Intelligence (*al-ʿaql al-faʿāl*), which ultimately refers to God, that turns the potential intelligibles into actual intelligibles and allows the intellect to

68. Ibid., 133.
69. Ibid., 14.
70. Idem, *Prolegomena*, 133.
71. Ibid., 143–44.
72. Ibid., 149.
73. Ibid. Al-Attas also explains the five internal senses and their role in the processes of perception: the common sense (*al-hiss al-mushtarak*) is the perceiver of forms, the representative faculty (*al-khayāliyyah*) is the conserver of forms, the estimative faculty (*al-wahmiyyah*) is the perceiver of meanings or intelligible forms, and the retentive (*al-ḥafiẓah*) and recollective (*al-dhākirah*) faculties are the conserver of meanings. The imaginative faculty (*al-mutakhayyilah*) is a special one since it perceives and acts upon intelligible forms.
74. Ibid., 156.

perceive them, just as light coming from the sun illuminates an object and makes them visible to the eye.[75] Note that this is in alignment with al-Attas's definition of knowledge as "the arrival of meaning in the soul" and "the arrival of the soul at meaning" since God is the source of origin of the intelligible realities but the human soul must participate in deliberate effort in perceiving them to attain knowledge.

To summarise, the nature, recipient, and source of knowledge are all non-physical entities. The meanings and intelligible forms imprinted upon the soul do not have physical qualities since they are already abstracted from their accidental attachments such as quantity, quality, space, and position.[76] The internal senses also are not in need of physical intermediaries for their acts of perception, although al-Attas notes that their various functions are localised in certain regions of the brain.[77] This non-physicalist view of human intelligence and the nature of knowledge shall be contrasted with the physicalist view stemming from SSSH and connectionism in the next section.

Issues at the Interface

One obvious issue emerging from the interface between the two is the nature of intelligence and its relation to knowledge acquisition. The computational theory of mind regards knowledge as purely physical, either in the form of a physical symbol-manipulating system (SSSH) or interaction and interconnection between neuron units (connectionism). Consequently, intelligence is reduced to computation *à la* Turing machine. It is self-contained—making no reference to the external world, only to its pre-defined internal rules. This is why it is argued that intelligence is substrate independent, because a computational system can be realised in many forms, i.e., biological system, electronics system, etc.

75. Ibid., 161–162 and 164–165.
76. Ibid., 150.
77. Ibid., 154.

In contrast, the Islamic conception of the human soul sees knowledge as non-physical. It is the reflection of the external non-physical "objects" (intelligible realities) that are imprinted on the non-physical soul. Physical systems such as the human brain are just intermediaries and not the locus of knowledge.

Since the computational theory of mind regards knowledge as purely physical, it logically follows that the source of knowledge only resides with particulars, because the physical world consists of particulars and not universals. This is another disagreement with the Islamic conception of the human soul, since the latter regards knowledge (i.e., meanings and intelligible forms) imprinted upon the soul are already abstracted from material or physical attachments. The process indeed begins with the perception of particulars. However, it does not stop there. The intellect also performs abstraction on them.[78] Furthermore, a material or physical entity can neither receive nor contain intelligibles, since the physical entity is divisible and it is not possible for intelligibles to be divisible too (were it to reside in a physical entity).[79]

Since according to the computational theory of mind the mind is a computational system, and since a computational system acts exactly according to pre-defined internal rules, the processes of intelligence can be said to be reducible to formal rules. Deduction and induction are two of the most popular formal rules, and they are also used in AI. It can even be claimed that the foundation of SSSH and symbolic AI is deduction whereas the foundation of connectionism and ANN is induction.

78. This is also the reason why the human intellect or soul is able to make a universal concept out of particular objects, such as the concept of chair out of particular chairs in the physical world. Furthermore, the human intellect is also able to construct "concepts of concepts," such as the concept of infinity. The former example belongs to the category of primary intelligibles (*al-maʿqūlāt al-ūlā*) whereas the latter belongs to the category of secondary intelligibles (*al-maʿqūlāt al-thāniyah*). The internal senses act upon the secondary intelligibles, which is pure abstractions of the matter. See Ibid., 156.
79. Ibid., 163.

While the Islamic conception of the human soul affirms that the intellect can perform intellection in a systematic manner (i.e. following the rules of induction and deduction), it does not mean that the act of intellection itself is reducible to the formal rules governing it. The intellect through its imaginative faculty can appraise intelligible forms in orderly and non-orderly orders.[80] It also emphasises the content of knowledge (intelligible forms) and not only the formal rules.

These are some of the general disagreements stemming from the interface between the computational theory of mind and the Islamic conception of the human soul. The disagreements will be elaborated further by the case studies from the current AI limitations: (1) "adversarial examples" in visual abstraction, (2) syntax-semantics distinction, and (3) abduction as a "leap" in reasoning.

Visual Perception and Abstraction

At a first glance, deep learning seems to be able to perform abstraction such as recognising and categorising images. It uses a technique called deep convolutional neural networks (DCNN) combined with a large data set to perform feature extraction from the training samples. By extracting hierarchical patterns such as edges, corners, gradations, shapes, and others from the picture's pixels, DCNN can perform image recognition and categorisation at high accuracy.[81]

The success of computer vision is largely due to the addition of more hidden layers which enables the hierarchical representation of images. This might indicate that the problem of visual perception and abstraction is computational, as suggested by Hans Moravec when he estimates that the human's retina

80. Ibid., 153.

81. For a concise explanation on how DCNN is used in visual perception-related tasks such as object recognition and categorisation, see Mitchell, *Artificial Intelligence*, 72–103.

resolution is about 500x500 pixels and able to process 10–50 frames per second, translating to 1 billion computer calculations per second.[82]

However, it is debatable whether DCNN computational way of visual perception by pixels processing and features extraction is the same as humans' visual perception. One of the main sources of doubt is the so-called "adversarial examples," where images created by slightly modifying an easily-classifiable exemplar in a way that was imperceptible to humans but could cause dramatic misclassification by computers.[83] For instance, one research discovered that by alternating very small and specific changes to its pixels, an image previously classified correctly with high confidence by AlexNet (an image recognition system using DCNN) as "school bus" is misclassified as an "ostrich" instead.[84] Another research creates a computer program that could create spectacle frames with specific patterns that fool a face-recognition system into confidently misclassifying the wearer as another person (one of them bizarrely misclassified as female actress Milla Jovovich!).[85] Questions are then raised: why computers are fooled and not humans? How do humans differ from computers in terms of performing visual perception and abstraction?

An alternative explanation is possible if visual perception is not viewed strictly as a computational problem. The perception of particulars in humans is done by the external senses, i.e., the

82. Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence* (Cambridge: Harvard University Press, 1988), 58–59.
83. Cameron Buckner, "Deep Learning: A Philosophical Introduction," *Philosophy Compass* 14 (2019): 13.
84. Christian Szegedy et al., "Intriguing Properties of Neural Networks," *arXiv preprint arXiv:1312.6199* (2013): 6; and Mitchell, *Artificial Intelligence*, 129.
85. Mahmood Sharif et al., "Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition," *CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (New York: ACM, 2016), 1528–1540; and Mitchell, *Artificial Intelligence*, 134.

eyes.[86] However, abstraction does not happen there, nor does the perception processes stop there. The information collected by the external senses is brought to the internal senses which "perceive internally the sensual images and their meanings, combine or separate them, conceive notion of them, preserve the conceptions thus conceived, and perform intellection on them."[87] It is the imaginative faculty with the mediacy of estimative faculty and common sense that perceives and "combines and separates them [forms] in an act of classification [and] adds to them and takes away from them so that the soul may perceive their meanings and connect them with the forms or images."[88] It is further noted that when the meanings are imprinted upon the soul, the intellect has already abstracted them from accidental attachments such as quantity, quality, space, and position.[89]

Therefore, it is not feature extraction of particulars that defines the abstraction process in humans. The soul receives the particulars from the external senses, but it is also capable of performing "isolation of single universals from particulars by way of abstraction of their meanings from matter."[90] This might explain why humans are not affected by the adversarial examples mentioned earlier and computers are: because computers perceive only particulars (pixels) and therefore are gullible to pixels manipulation. Humans also have their own "adversarial examples" in other forms such as optical illusions. However, this is in favour of the view that human perception is not purely computational but instead context-dependent.[91] The gullibility of humans to optical illusions is possible to have been caused by the estimative faculty that presides over judgment "by an instinctive interpretation of the image perceived" and not in an analytical way.[92]

86. Al-Attas, *Prolegomena*, 149.
87. Ibid., 149–150.
88. Ibid., 153.
89. Ibid., 150.
90. Ibid., 166.
91. Mitchell, *Artificial Intelligence*, 136.
92. Al-Attas, *Prolegomena*, 152.

Language and Understanding

Language is one of the key aspects of intelligence. Thus, it is no wonder that the infamous Turing test is essentially a language test. For years, AI researchers have been trying to create computer programs that can understand human language. There have been numerous successes such as virtual assistants (Siri, Alexa), automated translation (Google Translate), and the recently published ChatGPT by OpenAI.

How does a computer understand human language? Earlier natural language processing (NLP) programs used symbolic rule-based approaches inspired by grammar and linguistic rules.[93] In line with the SSSH, first-order logic is thought to be sufficient for representing language and knowledge.[94] This is apparent in John McCarthy's program "advice taker," where the program employs first-order logic to "draw immediate conclusions from a list of premises….either [in] declarative or imperative sentences."[95] However, this approach is largely forgotten now and overshadowed by the connectionist approach.

A connectionist approach is to "train" the ANN by feeding it massive language data sets. A variant of neural networks called the recurrent neural networks (RNN) is used here to process sequences of words at variable lengths.[96] However, neuron units can only take numbers as their input. Thus, each word from the data sets is assigned a number.[97] After that, each word is placed into a "semantic space" where the more related words tend to be placed close to each other.[98] Once all the words in the vocabulary are properly placed in the semantic space, the

93. Mitchell, *Artificial Intelligence*, 226.
94. Frankish and Ramsey, *The Cambridge Handbook of Artificial Intelligence.*
95. McCarthy, "Programs with Common Sense," 1.
96. Mitchell, *Artificial Intelligence*, 232.
97. Ibid., 238.
98. For example: the word "mother" with "female," "father," "grandmother," and so on.

meaning of a word can be represented by its location in space.[99] Algorithms such as *Word2vec* are used for this.[100] With this method, a network deduces that the closest words to "France" are "Spain," "Belgium," and "Netherlands" without being told concepts such as "country" or "European country." It can also correctly answer analogy problems such as "man is to woman as *king* is to __ (*queen*)" by subtracting the word vectors for *man* and *woman* and adding the result to the word vector *king*.[101]

Despite the promising results (especially by the connectionist approach), it still leaves questions about whether the computer actually *understands* the language. It has been argued that computation (be it in formal logic in SSSH or statistical analysis in connectionism) is sufficient for language understanding. For example, John Haugeland says that in AI;

> if you take care of the syntax, the semantics will take care of itself"[102] Ray Kurzweil also states that "the mathematical techniques that have evolved in the field of artificial intelligence…are mathematically very similar to the methods that biology evolved in the form of the neocortex. If *understanding language* and other phenomena *through statistical* analysis does not count as true understanding, then humans have no understanding either.[103]

However, there has not been a shortage of criticism, especially under the argument of Chinese room thought experiment[104] and the Winograd Schema Challenge[105]

99. Mitchell, *Artificial Intelligence*, 241.
100. For a concise explanation of *Word2vec*, see ibid., 242–247.
101. Ibid., 247–248. The closest word in the semantic space to the result turned out to be the word "queen."
102. Haugeland, *Artificial Intelligence: The Very Idea*, 100.
103. Ray Kurzweil, *How to Create A Mind: The Secret of Human Thought Revealed* (New York: Penguin Books, 2012), 7. Emphasis mine.
104. John R. Searle, "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3 (1980): 417–457.
105. Hector J. Levesque et al., "The Winograd Schema Challenge," *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning* (2012), 552–561.

commonsense reasoning test.[106] The Chinese room thought experiment is summarised as follows:

> Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a database) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.[107]

The Winograd Schema Challenge, on the other hand, is a test of commonsense reasoning in a machine. It involves questions that are for humans but tricky for machines, such as: "The trophy doesn't fit in the brown suitcase because it is too small. What is too small?" Surprisingly, it is difficult for machines to pass the Winograd Schema Challenge.[108] It challenges the view that a large language model based on statistical analysis is sufficient for understanding.[109] The Chinese room thought

106. The word "commonsense" here is understood generally as the ability to reason correctly in practical sense and not the "common sense" in the sense of *al-hiss al-mushtarak*. To avoid confusion, the word "commonsense" (without space) will be used to refer to the former and the word "common sense" will be used to refer to the latter.

107. Robert A. Wilson and Frank C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences* (Massachusets: The MIT Press, 1999), 115.

108. For a concise overview of the Winograd Schema Challenge and its relation to the problem of understanding in machine, see Melanie Mitchell, "What Does It Mean for AI to Understand?," *Quanta Magazine,* 16th December *2021*, https://www.quantamagazine.org/what-does-it-mean-for-ai-to-understand-20211216/.

109. Levesque et al., "The Winograd Schema Challenge," 554. See also Hector J. Levesque, "On Our Best Behaviour," *Artificial Intelligence* 212, no. 1 (2014): 27–35.

experiment presents a point that mere symbol manipulation (syntax) is not sufficient for real-world understanding (semantics), whereas the underwhelming results from WSC confirm the issue raised by the CRA. Ultimately, it presents a question about the nature of language: is it based solely on pre-defined internal rules (ex: grammar, statistical correlation between words, etc.)? Or is it an outward expression of intelligible forms imprinted upon the human soul?

It seems that the Islamic conception of the human soul is leaning towards the latter. Al-Attas places a strong emphasis on language as one of the defining features of man. Man is *hayawān nāṭiq*, a "language animal."[110] The root of language is inherent in the cognitive faculty of the soul, i.e., the rational soul.[111] The soul is responsible for the formulation of meaning through judgment, discrimination, distinction, and clarification.[112] Therefore, language is a reflection of meaning, because meaning is an "intelligible form…which a word, an expression, or a symbol is applied to denote it."[113]

Therefore, language is an outward expression of the intelligible form imprinted upon the soul, whose source is the Active Intelligence.[114] It is not an emergent property from a physical symbol system (syntax) completely detached from meanings and semantics as SSSH advocates. It is also not an emergent property from the statistical correlation of symbols as the connectionist would argue. The soul is indispensable for the apprehension of meaning.

110. Al-Attas, *Prolegomena*, 122.
111. Idem, *On Justice*, 31.
112. Ibid., 32 and *Prolegomena*, 122.
113. Ibid., 123.
114. Ibid., 161.

Knowledge and Inference

Earlier, it was stated that the intellect or soul can appraise intelligible forms in an orderly and non-orderly manner. The predisposition of the intellect to the activity of reasoning cannot be reducible to formal rules (i.e., induction and deduction) although it certainly can follow such rules. This raises questions: is it limited only to deduction and induction alone to exhibit human-level intelligence? If human reasoning is not reducible to induction and deduction, is there, then, any other type of reasoning that humans have but AI cannot replicate?

Deduction has two problems. The first one is the syntax-semantics dilemma addressed previously.[115] The other limitation is that it never adds knowledge because the conclusion necessarily follows if the premises are true. Causation also cannot be inferred directly from the premises.[116] Induction tries to fix the first problem as it learns from data, but it still suffers from the classic "correlation does not imply causation" dilemma, as the causation cannot be directly inferred from the induction itself.[117] This is also apparent in deep learning, as computer scientist Yoshua Bengio (1964–present) says, "deep [neural networks] tend to learn surface statistical regularities in the dataset rather than higher-level abstract concepts."[118]

---

115. Formally known as the "symbol grounding problem": how can the semantic interpretation of a formal symbol system be made intrinsic to the system. See Stevan Harnad, "The Symbol Grounding Problem," *Physica D* 42 (1990): 335–346.

116 For the limitation of deduction in relation to inference, knowledge, and AI, see Erik J. Larson, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do* (Cambridge: Harvard University Press, 2021), 106–115.

117. For an in-depth discussion about correlation and causation, see Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (New York: Basic Books, 2018).

118. Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (New York: Vintage Books, 2019), 62. This also explains the adversarial examples problem, as the DCNN only looks for common features instead of inferring high-level abstract concepts of the picture in the dataset.

If deduction and induction are not sufficient, then is there any mode of reasoning that humans can do but AI cannot? Computer scientist Erik J. Larson (1971–present) draws from philosophers Charles Sanders Peirce (1839–1914) that "thinking is not a calculation but a leap, a guess."[119] Humans do not necessarily think in the inductive or deductive mode in their strict sense, but also in abductive inference (abduction). However, the problem with abduction is that it is conjectural by nature.[120] Larson gives an example of abduction in its natural and formal form:

> If it is raining, the streets are wet // A → B
> The streets are wet // B
> Therefore, it's raining // ∴ A

Not only it is a fallacy of affirming the consequent error (and thus cannot be formalised in a computer), but it also requires knowledge about the causal relation between rain and wet streets. Abduction views an observed fact as a sign that points to a feature in the world.[121] It may also explain why the current AI lacks commonsense, because it "doesn't fit into logical frameworks like deduction or induction"[122] Commonsense and the understanding of the world are prerequisites of abduction, and both cannot be fully formalised (if not at all).

From the Islamic perspective, knowledge involves abstraction of sensibles into intelligibles.[123] Therefore, it requires an understanding of the world that comes from the abstraction

119. Larson, *The Myth of Artificial Intelligence*, 94.

120. Ibid., 172.

121. Ibid., 163. The conclusion that understanding of the world is vital to inference and production of knowledge is also shared by philosopher of science Alan Chalmers (b.1939) in terms of scientific method. See Alan F. Chalmers, *What Is This Thing Called Science?*, 3rd ed. (Indianapolis: Hackett Publishing Company, 1999), chapters 1–4 (especially chapter 4).

122. Ibid., 161.

123. Al-Attas, *Prolegomena*, 156.

of the observed objects. In addition, the imaginative faculty may appraise forms and meanings in an orderly or non-orderly fashion.[124] This might explain the "leap" part in abduction since the soul is not bounded by mechanistic induction and deduction—although the soul can use induction and deduction as well.[125] Furthermore, the soul in the possessive intellect stage possesses "first principles established by premises upon which rest self-evident truths…obtained not by means of deduction nor by verification"[126] This also might explain why commonsense is possessed by humans but not AI, since commonsense is a property of the soul.

Conclusion

This article addresses some issues stemming from the interface between the philosophical underpinnings of AI and the concept of the human soul in Islam. The computational theory of mind and its two strands, SSSH and connectionism, are taken as the philosophical underpinnings of AI whereas al-Attas' exposition is adopted in representing the Islamic conception of the human soul.

It is shown that the computational theory of mind regards intelligence and knowledge as purely physical. Intelligence is reduced to a formal system in SSSH or weighted interaction and interconnection between neuron units in connectionism. Thus, the computational theory of mind implicitly assumes that intelligence and knowledge reside in particulars such as physical systems. This is opposed to the Islamic conception of the human soul which regards the source, nature, and recipient of knowledge as non-physical in nature. The human intellect is the recipient of intelligible forms or universals from the Active

124. Ibid., 153.

125. Ibid., 154 and 166.

126. Ibid., 159. For example, the apprehension of the truths in the statement that the whole of something is greater than the parts.

Intelligence, thus intelligence and knowledge are not reducible to physical entities.

The disagreements between the two positions are further exemplified by analysing some cases from the current AI limitations. Three issues are discussed: (1) visual perception and abstraction, (2) language and understanding, and (3) inferential knowledge. The examples show the current AI limitations in understanding, as shown in the cases of "adversarial examples" and visual abstraction, syntax-semantics distinction (through the Chinese room thought experiment and Winograd Schema Challenge), and abduction as a "leap" in reasoning. The limitations come from the physicalist approach of the computational theory of mind that disregards the notion of non-physical entities such as the soul and intelligible forms/universals, whereas the two are vital in explaining human intelligence.

For future research, the issues of consciousness and the spiritual aspect of man can also be elaborated. From the Islamic perspective, consciousness is imaginal and intellectual in nature.[127] It is not an emergent property of any physical system. Influenced by the notion of AI, there is also a tendency to limit the human mind only in terms of rationality.[128] This is in stark contrast with the Islamic conception of the human soul that places the perfection of the heart as the primary aspect of the perfection of man.[129]

127. Al-Attas, *Prolegomena*, 167–168.

128. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Harlow: Pearson Education Limited, 2022), 19 and 58.

129. Muhammad Zainiy, *al-Attas' Psychology*, 55.

References

Abdullah, Wan Suhaimi Wan. *Abū al-Barakāt al-Baghdādī on the Human Soul: An Exposition of Some Major Problems of Psychology*. Kuala Lumpur: Pertubuhan Pendidikan Futuwwah, 2021.

Al-Attas, Syed Muhammad Naquib. *On Justice and the Nature of Man: A Commentary on Sūrat al-Nisā' (4):58 and Sūrat al-Mu'minūn (23):12–14*. Kuala Lumpur: IBFIM, 2015.

_____. *Prolegomena to the Metaphysics of Islām: An Exposition of the Fundamental Elements of the Worldview of Islām*. Kuala Lumpur: ISTAC, 1995.

_____. *The Nature of Man and the Psychology of Human Soul: A Brief Outline and Framework for an Islamic Psychology and Epistemology*. Kuala Lumpur: ISTAC, 1990.

Badri, Malik. *Abū Zayd al-Balkhī's Sustenance of the Soul: The Cognitive Behavior Therapy of a Ninth Century Physician*. London: The International Institute of Islamic Thought, 2013.

Bakar, Osman. *Classification of Knowledge in Islam*. Cambridge: The Islamic Texts Society, 1998.

_____. "The Clash of Artificial and Natural Intelligences: Will It Impoverish Wisdom?," in Abdallah Schleifer (ed.), *The Muslim 500: The World's 500 Most Influential Muslims 2023*. Amman: The Royal Islamic Strategik Studies Centre, 2023.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Buckner, Cameron. "Deep Learning: A Philosophical Introduction." *Philosophy Compass* 14 (2019): 1–19.

Carter, Matt. *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh: Edinburgh University Press, 2007.

Chalmers, Alan. *What Is This Thing Called Science?* 3rd. Indianapolis: Hackett Publishing Company, 1999.

Copeland, Jack. *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell Publishers, 1993.

De La Mettrie, Julien Offray. *Machine Man and Other Writings*. Edited by Ann Thomson. Translated by Ann Thomson. Cambridge: Cambridge University Press, 1996.

Dhaouadi, Mahmoud. "An Exploration into the Nature of the Making of Human and Artificial Intelligence and the Qur'ānic Perspective," *American Journal of Islam and Society* 9, no. 4 (1992): 465–481.

Dreyfus, Hubert L. Dreyfus and Stuart E. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press, 1986.

Ford, Martin. *Architects of Intelligence: The Truth About AI From the People Building I*t. Birmingham: Packt Publishing, 2018.

Frankish, Keith, and William M. Ramsey. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, 2014.

Good, Irving John. Speculations Concerning the First Ultraintelligent Machine. Vol. 6, in *Advances in Computers*, edited by Franz L. Alt and Morris Rubinoff, 31–88. New York: Academic Press, 1965.

Harnad, Stevan. "The Symbol Grounding Problem." *Physica D* 42 (1990): 335–346.

Haugeland, John. *Artificial Intelligence: The Very Idea.* Cambride: The MIT Press, 1985.

Hobbes, Thomas. *Thomas Hobbes Leviathan*. Edited by John C. A. Gaskin. Oxford: Oxford University Press, 1998.

Hodges, Andrew. *Alan Turing: The Enigma*. New York: Simon & Schuster, 1983.

Kang, Minsoo. *Sublime Dream of Living Machines: The Automaton in the European Imagination*. Cambridge: Harvard University Press, 2011.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* (2012): 1097–1105.

Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Penguin Books, 2012.

Larson, Erik J. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge: Harvard University Press, 2021.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning." *Nature* 521 (2015): 436–444.

Levesque, Hector J. "On Our Best Behaviour." *Artificial Intelligence* 212, no. 1 (2014): 27–35.

Levesque, Hector J., Ernest Davis, and Leora Morgenstern. "The Winograd Schema Challenge." *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning*, 2012. 552–561.

Levinovitz, Alan. *The Mystery of Go, the Ancient Game That Computers Still Can't Win*. May 12, 2014. Accessed March 30, 2022. https://www.wired.com/2014/05/the-world-of-computer-go/.

Marcus, Gary, and Ernest Davis. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Vintage Books, 2019.

Mayor, Adrienne. *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton: Princeton University Press, 2018.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, Claude E. Shannon. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." *AI Magazine* 27, no. 4 (2006): 12–14.

McCarthy, John. "Programs with Common Sense." *Symposium on Mechanization of Thought Processes*, 1959. Teddington. 1–15.

McCulloch, Warren S., and Walter Pitts.. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics* 5 (1943): 115–133.

Mitchell, Melanie.. *Artificial Intelligence: A Guide for Thinking Humans*. London: Pelican, 2019.

_____. *What Does It Mean for AI to Understand?* December 16, 2021. https://www.quantamagazine.org/what-does-it-mean-for-ai-to-understand-20211216/#.

Molesworth, Sir William, ed. *The Collected Works of Thomas Hobbes*. 12 vols. London: Routledge, 1992.

Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Massachusetts: Harvard University Press, 1988.

Newell, Allen, and Herbert A. Simon. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the Association for Computing Machinery* 19, no. 3 (1976): 113–126.

Nilsson, Nils J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. New York: Cambridge University Press, 2010.

Pearl, Judea, and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books, 2019.

Petzold, Charles. *The Annotated Turing: A Guided Tour through Alan Turing's Historic Paper on Computability and the Turing Machine*. Indianapolis: Wiley Publishing, 2008.

Popper, Karl R. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press, 1979.

Raquib, Amana, Bilal Channa, Talat Zubair, and Junaid Qadir. "Islamic Virtue-based Ethics for Artificial Intelligence," *Discover Artificial Intelligence* 2, no. 11: (2022).

Rescorla, Michael. *The Computational Theory of Mind*. Edited by Edward N. Zalta. September 21, 2020. Accessed March 31, 2022. https://plato.stanford.edu/entries/computational-mind/.

Rosenblatt, Frank. "The Perceptron: A Probabilistic Model for Information Storage and Organisation in the Brain." *Psychological Review* 65, no. 6 (1958): 368–408.

Rahman, Fazlur. *Avicenna's Psychology*. Westport: Hyperion Press, 1952.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th edition. Harlow: Pearson Education Limited, 2022.

Searle, John R. "Minds, Brains, and Programs." *The Behavioral and Brain Sciences* 3 (1980): 417–457.

Seth D. Baum, Ben Geoertzel, and Ted G. Goertzel. "How Long Until Human-level AI? Results From an Expert Assessment." *Technological Forecasting & Social Change* 78 (2011): 185–195.

Sharif, Mahmood, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. "Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition." *CCS'16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. New York: ACM, 2016, 1528–1540.

Simon, Herbert A., and Allen Newell. "Information Processing in Computer and Man." *American Scientist* 52, no. 3 (1964): 281–300.

Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. "Intriguing Properties of Neural Networks." *arXiv preprint* arXiv:1312.6199 (2013): 1–10.

Tegmark, Max. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Penguin Books, 2018.

Turing, Alan. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–460.

Turing, Alan. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42 (1936): 230–265.

Tzortzis, Hamza Andreas. *Does Artificial Intelligence Undermine Religion?* June 30, 2020. Accessed March 31, 2022. https://sapienceinstitute.org/does-artificial-intelligence-undermine-religion/.

Uthman, Muhammad Zainiy. *Al-Attas' Psychology*. Kajang Selangor: Akademi Jawi Malaysia, 2022.

Wilson, Robert A., and Frank C. Keil. *The MIT Encyclopedia of the Cognitive Sciences*. Massachusets: The MIT Press, 1999.

Zubair, Talat, Amana Raquib, and Junaid Qadir. "Combating Fake News, Misinformation, and Machine Learning Generated Fakes: Insights from the Islamic Ethical Tradition." *ICR Journal* 10, no. 2 (2019): 189–212.