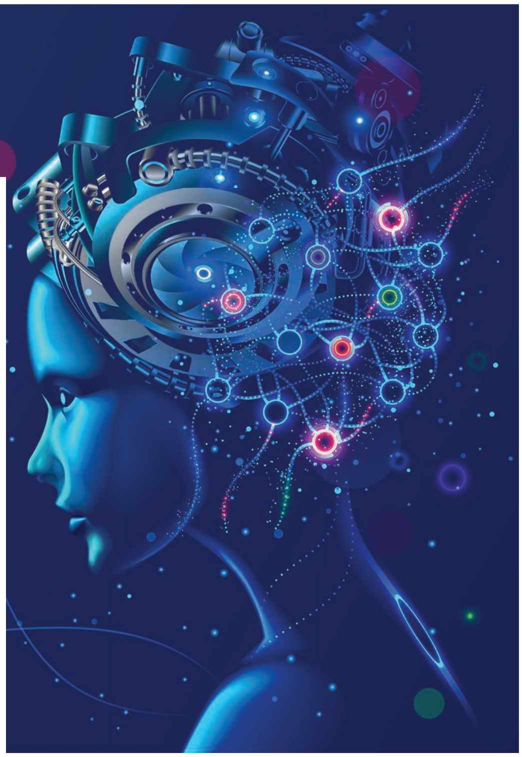

# Artificial General Intelligence: Pressure Cooker or Crucible?

**Moshe Sipper,** Ben-Gurion University of the Negev

*Turning toward nature, I contend that biological artificial general intelligence design is based on five pillars. I propose a crucible approach that would "melt" these elements by using (the equivalent of) computational "heat."*

Artificial intelligence (AI), a broad field that involves the ongoing pursuit to render machines capable of performing intelligent tasks, has taken the academic and industrial worlds by storm in a breathtakingly short time span. The current state of the art is nothing short of astounding, with advances in recent years that would have been considered futuristic a mere decade ago. Yet, despite all this progress, most scientists would agree that we are still in the "narrow AI" era. While a face recognition system might perform superbly at recognizing those allowed to enter a secure facility, it would be far from trivial to render the system capable of sifting through X-ray images. Few, if any, would aver that a machine performing some complex task is intelligent, conscious, and self-aware.

Artificial general intelligence (AGI), or strong AI, is defined as "an emerging field aiming at the building of 'thinking machines'; that is, general-purpose systems with intelligence comparable to that of the human mind (and perhaps ultimately well beyond human general intelligence)."[1] There are other definitions of AGI, which emphasize qualities other than analytical prowess, for example, creativity and emotionality; I consider herein the analytical, which is, arguably, the focus of most research in the area. AGI is, in fact, a recent term, intended to encompass the original idea of strong AI since mainstream AI research has turned toward domain-dependent and problem-specific methods and solutions.

Examining the history of AI reveals a division into epochs, each characterized by a research zeitgeist embracing a central methodology that was seen as the path toward what we would now refer to as AGI. We have witnessed epochs of symbolic reasoning, expert systems, intelligent agents, and probabilistic reasoning, culminating in the

*Digital Object Identifier 10.1109/MC.2022.3169399*
*Date of current version: 4 July 2022*

 0018-9162/22©2022IEEE

**EDITOR HSIAO-YING LIN**
IEEE Member; hsiaoying.lin@gmail.com

current era of machine- and deep learning.[15] Every epoch seems to have manifested the belief that the then-central paradigm would ultimately lead to true, self-aware, thinking machines.

AI pioneers in the 1950s and 1960s firmly believed that AGI would arrive on the scene within a generation. Simon[18] predicted in 1965 that "machines will be capable, within twenty years, of doing any work a man can do." In 1970, Minsky wrote that "within a generation … the problem of creating artificial intelligence will substantially be solved."[2] The AI character HAL 9000 in the iconic 1968 movie *2001: A Space Odyssey* best portrays what was believed at the time to be attainable by the beginning of the millennium. Some 20-odd years after the film's "deadline," estimates for the arrival of AGI differ widely, ranging from a decade to never.

An example of a classical approach to AGI is the Cyc project, which began in the 1980s, with the aim of assembling an encyclopedic ontology and knowledge base that spanned basic concepts and rules about how the world functions. "Cyc leverages symbolic reasoning rather than machine learning (ML). Symbolic reasoners were dubbed Good Old Fashioned Artificial Intelligence (GOFAI) by John Haugeland."[5] Interestingly, Cyc is still alive and kicking, having survived the vicissitudes of a field in continual flux.

Contemporary approaches to AGI focus, unsurprisingly, on the currently reigning paradigm of deep learning. Clune[3] recently proposed an AI-generating algorithm based on three pillars: 1) metalearning architectures, 2) metalearning the learning algorithms themselves, and 3) generating effective learning environments. This approach shares my thinking on the possibilities of computational learning and the importance of environments. Another recent approach is that of brain-inspired AI,[9] which uses principles of brain science (for example, cognition, inference, memory, and intelligence) to build AI algorithms. Herein, the idea is to more closely mimic naturally occurring systems and structures. I consider the natural inspiration as paramount and will circle back to it in the following.

Recent research into data-driven and knowledge-aware AI has noted the increasing difficulty in explaining AI models.[10] In many domains, explanations may not only be necessary but indeed legally required (for instance, in medicine), and the subfield of explainable AI aims to provide said explanations. This line of research mostly deals with deep learning-based techniques, wherein data-driven methods afford explanations stemming from task-related data, and knowledge-aware methods use extraneous knowledge to furnish explanations.[10] Indeed, the ability to explain one's reasoning is a hallmark of human intelligence.

I think of these current, heavily learning-based concepts, as "pressure cooker" approaches: apply enough "pressure"—that is, more (perhaps much more) of the current technology of choice—and AGI will ultimately emerge. The pressure cooker did not deliver the promised AGI in previous eras of AI. As to whether it will deliver this time around is a question still up in the air. I wish to propose a different path that may be worth exploring in parallel.

In 1952, Miller and Urey set up an experimental investigation into the molecular origins of life by conducting a chemical experiment that simulated the conditions thought at the time to be present on the early Earth and by testing the chemical origin of life under those conditions.[11] They introduced molecules thought to exist in early Earth's primitive atmosphere into a closed chamber and simulated lightning discharges by supplying the system with electrical current. After a few days, they observed that the flask contained organic compounds, some of which were amino acids that serve as essential building blocks of protein (interestingly, decades later, scientists examining sealed vials preserved from the original experiments found well more than 20 different amino acids, far more than originally reported).

The Miller–Urey experiment's importance, to my mind, is not so much in its results (which have come under some debate) but in the fundamental question the work posed—and in the rigorous experimental method the authors proposed by way of answering said question. They wished to explore the origins of life and set out to define the environment they believed to be appropriate for such an experimental study. In the same vein, I propose to pursue AGI from its origins and aim to define the necessary desiderata by turning toward nature, hitherto the only "designer" of AGI entities. Granted, huge strides in biology notwithstanding, our understanding of nature remains limited. Yet, I still think that our current grasp of biological AGI can guide us in the building of AGI machines, particularly by focusing on the following five elements[21]: 1) an underlying physical universe that supports 2) evolution, 3) learning, 4) a complex environment, and 5) replication. It now behooves us to reify these elements, and I wish to offer one such possible reification (others can likely be proposed).

For the first element—a "universe" wherein all computation takes place—I propose cellular automata (CA). CA are dynamical systems in which space and time are discrete.[4] CA consist of an array of cells (usually of dimensionality 1 or 2), each of which can be in one of a finite number of possible states, updated through discrete time steps according to a local, identical interaction rule. CA exhibit three notable features, namely, massive parallelism, locality of cellular interactions, and simplicity of basic components (cells). The (local) operation of a cell is dictated by a so-called rule table, which defines the cell's next state given its current state and the states of its neighbors. With today's advances in hardware, CA can be made to perform extremely fast for 2D and possibly even 3D grids. Figure 1(a) provides an example of CA performing an image correction task.[19] Interestingly, in the 1960s, Fredkin[6] speculated that our universe might, at its deepest level, operate like CA.

For the evolutionary element, the natural candidate is the field of evolutionary algorithms (EAs), wherein core concepts from evolutionary biology—inheritance, random variation, and selection—are harnessed in algorithms that are applied to complex computational problems.[21] EAs, whose origins can be traced to the 1950s and 1960s, have come into their own during the past two decades. EA techniques have been shown to solve numerous difficult problems from widely diverse domains. As argued by Kannappan et al.,[8] who reviewed research on evolving human-competitive machine intelligence, "Surpassing humans in

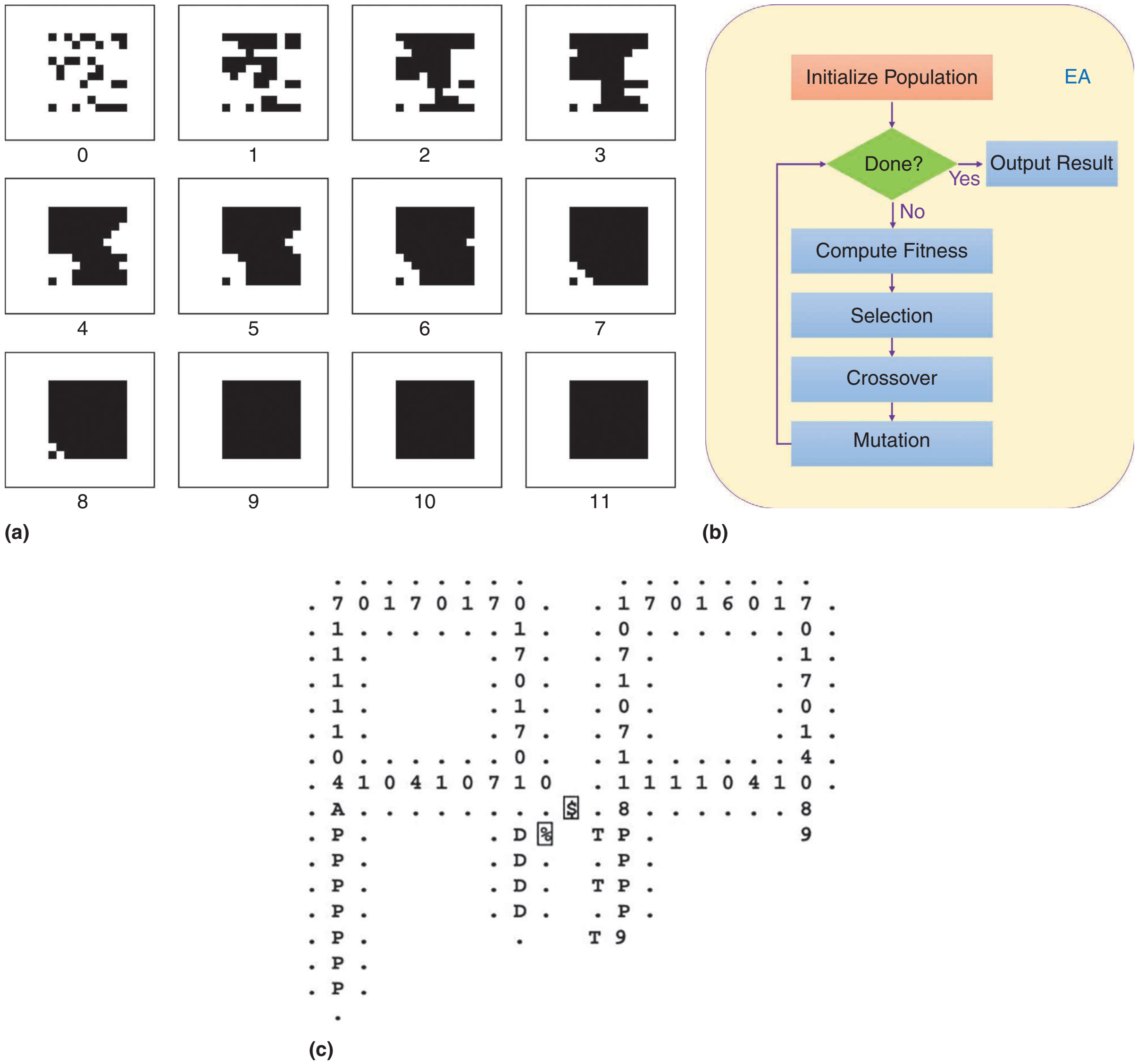


**FIGURE 1.** (a) 2D CA computing the boundary of a rectangle through discrete time steps (shown beneath each panel), with all cells working in parallel.[19] The output is attained by having the rectangle filled out at the final time step. (b) A canonical evolutionary algorithm (EA). (c) A programmable self-replicating loop in 2D CA. A numerical value denotes a cellular state. "Sheath" states are denoted by dots; P denotes a state belonging to the set of program states; D denotes a state belonging to the set of data states; and A is a state that indicates the position of the program. The loop is shown in the midst of self-replication, after which the program will run on the data.[13]

the ability to solve complex problems is a *grand challenge*, with potentially far-reaching, transformative implications." A canonical EA is shown in Figure 1(b): essentially, an initial population of candidate solutions is generated (usually at random), after which the algorithm cycles through a fitness–selection–variation loop until an acceptable solution is found. Note that Figure 1(b) does not delineate a single algorithm but rather a meta-algorithm, representing the basic prototype of a broad family of algorithms, referred to as *evolutionary*.

The learning element would straightforwardly consist of machine- and deep learning algorithms, which today constitute the focal paradigm of AI. As for a complex environment, several candidates might come to mind. I propose games, which have been suggested as AGI testbeds[16] because they require many human-level skills for achieving excellence: quick reactions, visual understanding, motor coordination, path planning, decision making, tradeoff evaluation, predicting future states, physics comprehension, handling incomplete information, and understanding narrative. Games possess essential qualities that arguably render them complex enough for the AGI challenge, exhibiting key properties of complex environments by being changeable, diverse, unpredictable, surprising,[14] nondeterministic, fully or partially observable, and rewarding (regarding the latter, Silver et al.[17] recently argued that "reward is enough"; that is, "intelligence … can be understood as subserving the maximisation of reward").

Finally, regarding the replication element, we may look into the decades of research into artificial self-replication. In the late 1940s, eminent mathematician and physicist John von Neumann became interested in the question of whether a machine can self-replicate, that is, produce copies of itself. The study of artificial self-replicating structures and machines has been taking place since then.[12,20] Much of this work is motivated by the desire to understand the fundamental information-processing principles and algorithms involved in self-replication, independent of their physical realization. An understanding of these principles could prove useful in a number of ways. It may advance our knowledge of biological mechanisms of replication by clarifying the conditions that any self-replicating system must satisfy and by providing alternative explanations for empirically observed phenomena. The fabrication of artificial self-replicating machines can also have diverse applications ranging from nanotechnology to space exploration. Much of this research has been studied within the context of CA [Figure 1(c)].[20]

By computational analogy to the Miller–Urey setup—and, in contrast, to a pressure cooker—I propose a crucible that would "melt" the preceding five elements by using "heat." This approach, which I term *artificial general intelligence crucible* (*argil*), represents a different path to AGI. Figure 2 depicts

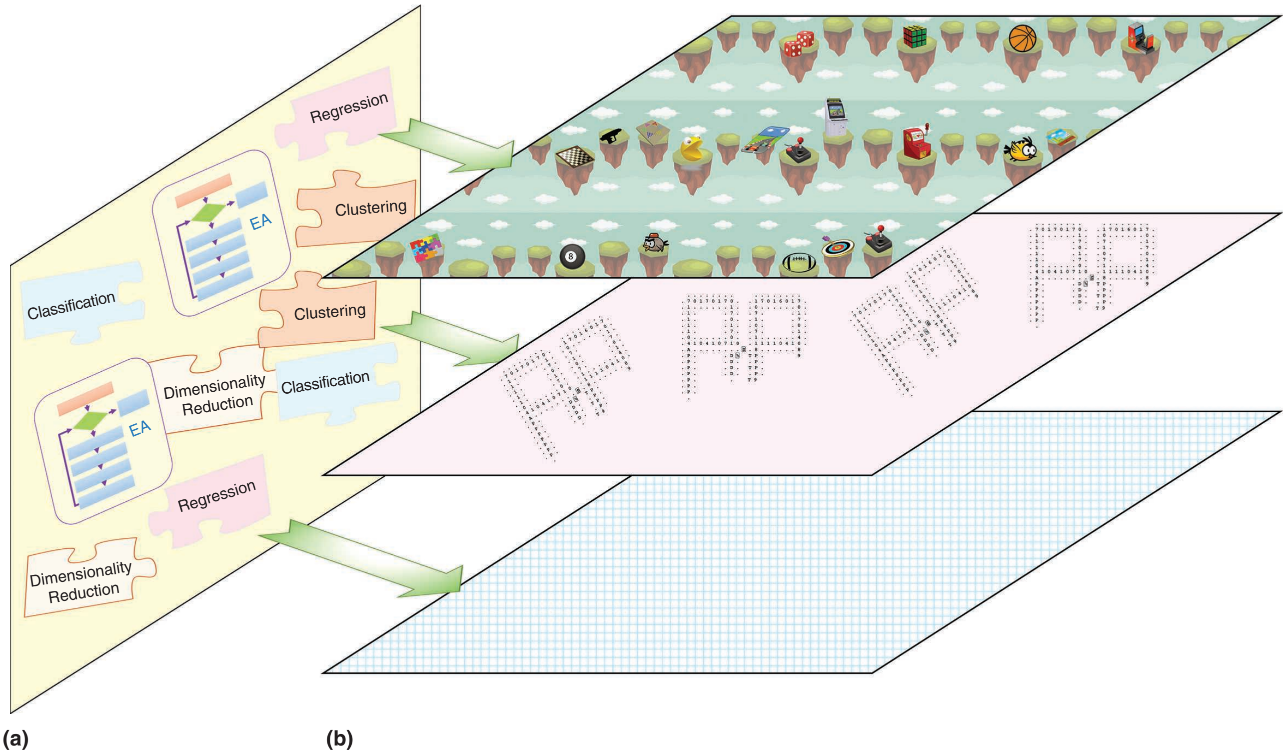


**FIGURE 2.** The argil approach. (a) The EA and ML mechanisms pertinent to all vertical layers. (b) The physical layer of CA "universe" (bottom), CA-based structures (center), and games environment (top).

the idea for the overall argil setup. The following two main challenges present themselves:

1. *Melting*: For AGI to emerge, the five elements must come to work as a cohesive one.
2. *Attaining heat*: To wit, there must be the computational wherewithal—theory, algorithms, and hardware—necessary to unite the elements.

Perhaps the argil approach will serve as clay from which to mold AGI.

Games can be divided into many groups and subgroups, a partial list of which includes board games, card games, dice games, video games, mobile games, and puzzles (groups are not necessarily mutually exclusive). We might first aim for an AGI prototype that performs well on several games that belong to a single group. Two cornerstones of AGI are general-purpose abilities and humanlike intelligence. General-purpose abilities will be had by complexifying the environment through the addition of different kinds of games. To adjudicate humanlike intelligence, we might consider a Turing Test-like[22] scenario adapted to game environments, for example, as proposed by Hingston.[7] An AGI able to evolve, grow, learn, and perform on a par with humans on multiple games belonging to multiple categories would surely find many valuable uses outside the domain of games.

It is told of Albert Einstein that he once gave an exam to his graduating class—the exact same exam he had given the previous year. His teaching assistant, thinking this was the result of the professor's absentmindedness, alerted Einstein: "Excuse me, sir …" he began timidly.

"Yes?" said Einstein.

"Sir, it's about the test you just handed out …"

Einstein waited patiently.

"I'm not sure you realize it, but this is the exact same test you gave out last year."

Einstein paused for a moment and then replied with equanimity, "Yes, it is the same test. But the answers have changed."

Perhaps we can come up with different answers to decades-old questions. ■

## ACKNOWLEDGMENTS

I am grateful to Jenny Racah and Yair Kohn for our many insightful discussions. I thank Hsiao-Ying Lin for several helpful suggestions.

## REFERENCES

1. AGI Society. http://www.agi-society.org/ (Accessed: Jul. 26, 2021).
2. N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, 1st ed. London, U.K.: Oxford Univ. Press, 2014.
3. J. Clune, "AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence," 2020, *arXiv:1905.10985*.
4. E. F. Codd, *Cellular Automata*. New York, NY, USA: Academic, 1968.
5. Cycorp. https://cyc.com/faq/ (Accessed: Feb. 16, 2022).
6. E. Fredkin, "Digital mechanics," *Physica D*, vol. 45, nos. 1–3, pp. 254–270, 1990, doi: 10.1016/0167-2789(90)90186-S.
7. P. Hingston, "A Turing test for computer game bots," *IEEE Trans. Comput. Intell. AI Games*, vol. 1, no. 3, pp. 169–186, 2009, doi: 10.1109/TCIAIG.2009.2032534.
8. K. Kannappan *et al.*, "Analyzing a decade of human-competitive ('HUMIE') winners: What can we learn?" in *Genetic Programming Theory and Practice XII*, R. Riolo, W. P. Worzel, and M. Kotanchek, Eds. Cham, Switzerland: Springer Int. Publishing, 2015, pp. 149–166.
9. N. K. Kasabov, *Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence* (Springer Series on Bio- and Neurosystems). vol. 7. New York, NY, USA: Springer-Verlag, 2019.
10. X.-H. Li *et al.*, "A survey of data-driven and knowledge-aware eXplainable AI," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 29–49, 2022, doi: 10.1109/TKDE.2020.2983930.
11. S. L. Miller, "A production of amino acids under possible primitive earth conditions," *Science*, vol. 117, no. 3046, pp. 528–529, 1953, doi: 10.1126/science.117.3046.528.
12. B. Naveh, M. Sipper, D. Lancet, and B. Shenhav, "Lipidia: An artificial chemistry of self-replicating assemblies of lipid-like molecules," in *Proc. 9th Int. Conf. Simul. Synthesis Artif. Life*, MIT Press, 2004, vol. 9, p. 466.
13. J.-Y. Perrier, M. Sipper, and J. Zahnd, "Toward a viable, self-reproducing universal computer," *Physica D, Nonlinear Phenomena*, vol. 97, no. 4, pp. 335–352, 1996, doi: 10.1016/0167-2789(96)00091-7.
14. E. M. Ronald, M. Sipper, and M. S. Capcarrère, "Design, observation, surprise! A test of emergence," *Artif. Life*, vol. 5, no. 3, pp. 225–239, 1999, doi: 10.1162/106454699568755.
15. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Hoboken, NJ, USA: Pearson Education, 2020.
16. T. Schaul, J. Togelius, and J. Schmidhuber, "Measuring intelligence through games," 2011, *arXiv:1109.1314*.
17. D. Silver, S. Singh, D. Precup, and R. S. Sutton, "Reward is enough," *Artif. Intell.*, vol. 299, p. 103,535, Oct. 2021, doi: 10.1016/j.artint.2021.103535.
18. H. A. Simon, *The Shape of Automation for Men and Management*, vol. 13. New York, NY, USA: Harper & Row 1965.
19. M. Sipper, *Evolution of Parallel Cellular Machines*. Heidelberg, Germany: Springer-Verlag, 1997.
20. M. Sipper, "Fifty years of research on self-replication: An overview," *Artif. Life*, vol. 4, no. 3, pp. 237–257, 1998, doi: 10.1162/106454698568576.
21. M. Sipper, *Machine Nature. The Coming Age of Bio-Inspired Computing*. New York, NY, USA: McGraw-Hill, 2002.
22. A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. LIX, no. 236, pp. 433–460, 1950, doi: 10.1093/mind/LIX.236.433.

**MOSHE SIPPER** is a professor of computer science at Ben-Gurion University of the Negev, Beer Sheva, 84105, Israel. Contact him at sipper@bgu.ac.il.