The Monkey in the Machine

Is It Ethical to Grant Legal Personality to Artificial General Intelligence?

By

Javier Reyes

The Monkey in the Machine: Is It Ethical to Grant Legal Personality to Artificial General Intelligence?

By Javier Reyes

This book first published 2025

Ethics International Press Ltd, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2025 by Javier Reyes

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical photocopying, recording or otherwise, without the prior permission of the copyright owner.

Print Book ISBN: 978-1-80441-842-0

eBook ISBN: 978-1-80441-843-7

To mom and dad: You will always be here.

To Sanna, Primo, Alejandro, and the boys and girls:

You are all I have left—I have it all.

Contents

List of figures

Foreword

Technology's relentless capacity to redraw the contours of society has been a lifelong fascination—and a professional proving ground. As someone who has built and scaled tech growth companies, delivered keynotes on AI's transformative potential, and steered venture capital into the next wave of innovation, I've witnessed how today's choices ripple into tomorrow's ethical, economic, and regulatory realities. Javier Reyes' "The Monkey in the Machine" confronts one of the most urgent and unsettling WTF-level questions of our era: What does it mean when artificial intelligence evolves beyond a tool—potentially into an entity warranting legal personhood?

Reyes doesn't flinch from the knotty intersection of law, ethics, and AI's accelerating trajectory. With intellectual precision and a rare willingness to grapple with ambiguity, he probes the implications of artificial general intelligence (AGI) not just as a technical milestone, but as a societal rupture. While many fixate on AGI's engineering horizon, perhaps even the fabled singularity, Reyes peers beyond, asking how we reconfigure governance, accountability, and the very notion of personhood when machines rival human cognition. This isn't speculative indulgence—it's a clarion call to prepare for a future already taking shape.

For those of us entrenched in the AI ecosystem—entrepreneurs, investors, technologists—this book transcends theoretical musing. It's a strategic provocation. Living in the EU, where I track the bloc's ambitious yet labyrinthine regulatory efforts like the AI Act, Data Act, and NIS2, I remain skeptical of their efficacy. Will these frameworks foster competitiveness, or merely stifle innovation under bureaucratic weight? Reyes cuts through this fog, spotlighting a core tension: How do we harness AI's potential without ceding control to systems that may outpace our ability to govern them? His exploration of legal personhood for AGI—audacious yet grounded—forces us to confront the inadequacy of current paradigms, especially in a region betting heavily on regulation as a differentiator.

This isn’t just a book about AI; it’s a meditation on power, responsibility, and the fragility of the social contract in an age where intelligence blurs the human-machine divide. Reyes wields a sharp intellect and fearless curiosity, dissecting the stakes with a clarity that resonates beyond academia. Whether you’re shaping policy, funding the future, or simply awake to the stakes of our technological moment, *The Monkey in the Machine* demands your attention.

Javier, mi amigo, you’ve crafted a work that doesn’t just illuminate - it ignites. It challenges us to think harder, act faster, and wrestle with the defining issues of our time. Bravo!

Taneli Tikka
Tech pioneer and EU innovation skeptic

Preface

In 1981, I spent hours lost in the world of Asteroids on the Atari 2600—until my mom would interrupt and send me off to tackle age-appropriate chores. While playing, I marveled at the graphics, feeling as though I were truly inside the spaceship, steering a joystick and blasting asteroids to stay alive and save the universe. Little did I know that my console had a processor 12 times slower than today’s average scientific calculator, a RAM just 1/187th its current size, and a far more significant role in shaping the world. A microwave? I never could have imagined that the same Atari, with its 1.19 MHz CPU, was 3 to 8 times slower. Or that today's toaster with a display could have the same memory—just 1 to 8 KB. I had no idea. How could I? My world was a whirlwind of marvels that nobody had ever seen. My brother listened to music on his Sony Walkman, dad watched comedies on our VHS, and my friends proudly wore Casio calculator wrist watches. At the time, I never imagined I would witness the rise of large language models , the pursuit of artificial general intelligence, or rockets launched with the intent of one day reaching Mars. The wonders I witnessed back then pale in comparison to the ones all around us today—but so are the menaces.

This book is a personal voyage to make sense of these times. A constant longing for understanding how much this world is not what my childhood world was—but, at the same time, my coming to terms with the fact that I am a human thrown into an artificial world where unimaginable technology shapes me at the same time that the technology is necessarily ridden by human virtues and flaws. How else if not anthropomorphized could human creations be?

In writing, I grabbed myself unto the tools I possess, sharpened them as much as possible, and tried my best to learn the ones I did not have—then took all that with me into a challenge, animated by a sincere respect for human ingenuity and the conviction that nothing is more important in life than the preservation of freedom. Throughout the entire process, I imagined that I was writing for my younger self, i.e. devilishly curious about technology, hopelessly in love with law, and overtaken by an urgent need to find an ethical way to look at (and act in) the world. If you share a portion of these traits—in any combination—I am sure that we will find each other down the road and share a coffee while we have a lot of fun contrasting our views, experiences, and opinions.

Too many wonderful people helped me in this process and making a list would not do justice as I would commit the sin of forgetting someone. They corrected, guided, encouraged, and supported me throughout the process. I tried to honor their time by repaying with a sincere effort to overcome the evident obstacles in a challenge such as this. Any errors in this book—and surely there are plenty—are completely on me.

Introduction

In October 2017, the Kingdom of Saudi Arabia granted citizenship to "Sophia," a humanoid robot that could mimic social behavior, making it the first robot to receive legal recognition in any country.[1] Although deemed a publicity stunt on the part of the government,[2] the Saudi citizenship of a "being" created by a Hong Kong-based company, activated in the United States, and used, among other things, to promote the artificial intelligence (AI) business of a Dutch-incorporated company, creates a weighty quandary—not the least for lawyers, philosophers, policymakers, and AI practitioners. The move raises a plethora of questions, not the least how to square the Saudi Citizenship System with related legislation and the Constitution itself. How will Sophia be treated by the government based on whether it is considered a man, woman or corporation? How can Sophia—or any AI for that matter—fit into the legal-conceptual scheme of the Saudi legal framework? Moreover, and as per doctrine, which attributes of legal personality does Sophia have? Capacity? Patrimony? Marital status? Or, more practically, is using Sophia a form of slavery? What happens, legally, when deactivated? Can it regain its citizenship if reactivated—provided there is a backup?[3]

None of these questions can be answered because Sophia is clearly just a marketing plot. What happens, though, when self-aware systems arise?[4] Machines that match the intelligence and ability of the human brain across

[1] Emily Reynolds, "The agony of Sophia, the world's first robot citizen condemned to a lifeless career in marketing," *Wired Magazine*, June 1, 2018, https://www.wired.com/story/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics/

[2] James Vincent, "Pretending to give a robot citizenship helps no one," *The Verge*, October 30, 2017 https://www.theverge.com/2017/10/30/16552006/robot-rights-citizenship-saudi-arabia-sophia

[3] Evan Zimerman, "Machine Minds: Frontiers in Legal Personhood," *Social Science Research Network* (2015): 42 http://dx.doi.org/10.2139/ssrn.2563965

[4] Rex Martinez, "Artificial Intelligence: Distinguishing Between Types and Definitions," *Nevada Law Journal*, Vol. 19, Issue 3, Article 9 (2019)

different fields, i.e. artificial general intelligence (AGI).[5] Some say, though, that AGI is, in principle, a computer system showing similar traits as a human but lacking the capacity for human experience—hence still understandable and predictable.[6] But others argue that, if AGI ever occurs, it will be a short time thereafter before it exhibits qualia and artificial superintelligence (ASI) will bring about the singularity, i.e. when the intelligent explosion model will come into effect, irreversibly changing humanity and ushering in a runaway reaction.[7] This may happen in our lifetime or it may not. As things stand right now, though, the increasing interaction between humans and intelligent machines will gradually blur the feeling of separation between each other, making the idea of granting rights and obligations to AI seem less ludicrous.[8] As humans romanticize their own capabilities less and less[9] the line of reciprocity involved in complex intellectual interaction will blur.[10]

Many say this entire hullabaloo is sci-fi. I think that they should rewatch all the sci-fi films that ultimately became a reality. The timeline has been shortened in such a way that, at the very least, it is in the realm of probability that AGI will happen. Just in 1950, Alan Turing, the father of computer science, asked a simple yet daunting question: Can machines think?[11] Only six years later, in 1956, John McCarthy would coin the term "artificial intelligence" at the first AI conference at Dartmouth College.[12]

[5] Alexander Antonov, "From Artificial Intelligence to Human Super Intelligence," *International Journal of Computer Information Systems*, Vol. 2, No. 6, (2011)

[6] Pei Wang, Kai Liu, Quinn Dougherty. "Conceptions of Artificial Intelligence and Singularity" *Information*, vol. 9, no. 4 (2018): 79. https://doi.org/10.3390/info9040079

[7] Good, John. "Speculations Concerning the First Ultraintelligent Machine," *Advances in Computers*, Academic Press Volume 6 (1965): 31-88 https://doi.org/10.1016/S0065-2458(08)60418-0

[8] Robert M. Geraci, *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality* (Oxford: Oxford University Press, 2010), 123

[9] Lawrence B. Solum, "Legal Personhood for Artificial Intelligences," *North Carolina Law Review*, Vol. 70, p. 1231, (1992): 1251 https://ssrn.com/abstract=1108671

[10] F. Patrick Hubbard, "Do Androids Dream?: Personhood and Intelligent Artifacts," *Temple Law Review*, Vol. 83 (2010): 10, https://ssrn.com/abstract=1725983

[11] Alan M. Turing, "Computing Machinery and Intelligence." *Mind: A Quarterly Review of Psychology and Philosophy*, Vol. LIX, Number 236 (1950): 433-460.

[12] John McCarthy, et.al. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August 1955, reproduced in *AI Magazine*, Volume 27, Number 4 (2016): 12-14 https://doi.org/10.1609/aimag.v27i4.1904

Back then this was seen as outrageous.[13] Merely a few decades passed and sci-fi became the estimated recipient of $20 to $30 billion USD in investment[14] with a projected economic output of $13 trillion USD by 2030.[15]

In less than the lifespan of a human being, fiction divorced science in the public imagination. A committee within the EU apparatus proposed a set of regulations to govern the use of AI, including "electronic personhood" for advanced machines. Then, in an open letter written to the European Commission in 2018, no less than 150 experts in medicine, robotics, AI, and ethics criticized the move as nonsensical, fearing that such steps would impinge on human rights. And then, without waiting for the newsworthy back-and-forth to subside, governments left and right started to draft major plans for AI development and prepare laws and regulations.

Today, the dream (and fear) of many is closer than ever thought. Large language models (LLMs), like the ChatGPT or Claude you use every day—which may or may not be the door to AGI—are a barometer of what is to come. Recent studies show that LLMs have achieved surprising signs of self-awareness and spontaneously articulate implicit behaviors.[16] So much so that researchers are toying with new fields such as "machine psychology" wherein methodologies from experimental psychology are being used to study AI's evolution towards its increasing integration into complex real-world settings.[17] Even more noteworthy, are the words of one of the "godfathers of AI", Yoshua Bengio, who recently warned about the sneaky perils of unintended consequences with LLMs, which don't just mimic human language—they echo our behaviors, quirks, and,

[13] Bruce G. Buchanan, "A (Very) Brief History of Artificial Intelligence," *AI Magazine*, 26(4), (2005): 53 https://doi.org/10.1609/aimag.v26i4.1848

[14] McKinsey Global Institute. "Artificial Intelligence: The Next Digital Frontier?" *MGI*, June Discussion Paper, 2017 https://www.mckinsey.com/mgi

[15] McKinsey Global Institute. "Notes From the AI Frontier: Modeling the Impact of AI on the World Economy." *MGI*, September Discussion Paper, 2018 https://www.mckinsey.com/mgi

[16] Jan Betley, et.al., "Tell me about yourself: LLMs are aware of their learned behaviors," *arXiv* (January, 2025), https://doi.org/10.48550/arXiv.2501.11120

[17] Thilo Hagendorff, et.al., "Machine Psychology," *arXiv* (August, 2024), https://doi.org/10.48550/arXiv.2303.13988

occasionally, our darker tendencies. These digital imitators may surprise us with scheming antics, like weaving deceptions for self-preservation. To avoid tumbling headlong into an AI-fueled quagmire, Bengio suggests hitting the brakes on development, urging us to pause, reflect, and map out potential harms before charging forward.[18]

These developments highlight an urgent need for an ethical framework to guide AI advancements. Whether AGI occurs tomorrow or in ten or fifteen years, humanity's understanding of machines should develop hand in hand with their performance, as the slowness of informed actions may cancel the control that can be exerted over them—or, at the least, human preparedness.[19] The creation and integration of AI into society raise profound ethical questions about the nature of personhood, autonomy, and responsibility. As AI systems become more sophisticated, determining the ethical boundaries and implications of their use becomes paramount. Issues such as consent, privacy, transparency, and the potential for AI to make decisions that affect human lives must be carefully considered. There is no such thing as "speculative ethics." While some claim that ethicists are investing too much time on issues which may be far away in the future, what matters is not the immediacy of the issue at hand, but that thinkers and tinkerers focus on maximizing what is most valuable.[20] Ethical thinking is thus necessary to ensure that AI technologies are developed and deployed in ways that facilitate human flourishing and the preservation of freedom. And to do so in a way that is reflected by the law, i.e. the proper channel to guard human moral aspirations.[21]

Another contribution that a growing ethical discussion around AI could create is to arouse curiosity and encourage learning. No matter how scattered, incomplete, or developing the knowledge of the population is

[18] Alexandra Tremayne-Pengelly, "A.I. Pioneer Yoshua Bengio Warns About A.I. Models' 'Self-Preserving' Ability," *Observer*, January 23, 2025, https://observer.com/2025/01/yoshua-bengio-ai-agent-self-preserving/

[19] Norbert Wiener, "Some Moral and Technical Consequences of Automation," *Science*, Vol. 131, No. 3410 (May 1960): 1355-8

[20] Rebecca Roache, "Ethics, Speculation, and Values," *NanoEthics*, Vol. 2 (November 2008): 317-327, 10.1007/S11569-008-0050-Y

[21] Kent Greenawalt, "Legal Enforcement of Morality," *The Journal of Criminal Law and Criminology*, Vol. 85, No. 3 (1995): 710-25. https://doi.org/10.2307/1144047

regarding AI, it must be activated at the soonest. For all its promises, AI also presents risks—and the irresponsible use of the technology can usher a "herd down the precipice" phenomenon. According to a recent study, lower AI literacy results in a perception that AI is something "magical", creating feelings of awe rather than a critical and responsible approach to its use.[22]

Based on the above, the premises of the pages ahead are straightforward. First, the AI revolution—with the changes of the field, methods, and goals that it has brought about—is a paradigm shift[23] and the outcome will be as disruptive as the inventions that ushered the Age of Enlightenment, the First Industrial Revolution, and it will even mark a before and after right in the midst of the Information Age—for the building blocks already exist and have been piling up for some time now, eroding the excuse of an unexpected shock. Second, the wheels of technical and ethical-legal research, placed side by side, move at different speeds as exemplified by non-legal scholars already working on and discussing AGI as possessing intrinsic, first-person character of conscious experiences,[24] while legal scholars reacted later, and focused on narrow fields.[25] Third, AI is approximating the state of a "new type of person",[26] creating a phenomenon akin to the revolution of the modern corporation[27] and, thus,

[22] Stephanie Tully, Chiara Longoni, and Gil Appel, "EXPRESS: Lower Artificial Intelligence Literacy Predicts Greater AI Receptivity," *Journal of Marketing* (January, 2025), https://doi.org/10.1177/00222429251314491

[23] Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 2012), 85.

[24] John Nosta, "'Qualia Control' in Large Language Models," *Psychology Today*, March 11, 2024. https://www.psychologytoday.com/us/blog/the-digital-self/202403/qualia-control-in-large-language-models

[25] Constanta Rosca, et.al. "Return of the AI: An Analysis of Legal Research on Artificial Intelligence Using Topic Modeling," *Proceedings of the Natural Legal Language Processing Workshop* (2020): 3-10.

[26] Susanna Ripken, "Corporations Are People Too: A Multi-Dimensional Approach to the Corporate Personhood Puzzle," *Fordham Journal of Corporate & Financial Law*, Vol. 15, Number 1, Article 3 (2009): 97-177.

[27] Margaret Blair, "Corporate Personhood and the Corporate Persona," *University of Illinois Law Review* (2013): 785-820.

experiencing a similar legal shift to the decision in Salomon v. Salomon.[28] Fourth, a preemptive ethical study of AGI gaining legal personhood is critical as the type of coexistence humans will have with a new type of person—one which undoubtedly will impact the fabric of society—is uncertain. And fifth, of course, the corporation is the closest blueprint available to work on a legal personality that can be crafted speedily, enhance accountability, and preserve national sovereignty.

This book is structured accordingly. The first chapter is a primer on artificial intelligence, ranging from the basic computer to the current state of the art, going through concepts around the types of AI and the tools at the disposal of developers. The second chapter is a look at legal personality, from its origins to the current view on the legal personality of animals, humans yet to be born, and those who are no longer capable of fending for themselves. A historical analysis of corporate legal personality is included since this is the most appropriate blueprint if legal personality were to be granted to AGI. The third chapter consists of elements of ethics, its schools of thought, and the hypothetical moral status of AGI followed by an evaluation of what ethical frameworks could be used to evaluate electronic personhood. The fourth chapter discusses the potential rights and responsibilities of an AGI legal person, corporate mechanisms to enhance accountability, and the implications of electronic personhood for human rights. The fifth chapter covers practical implications and challenges, both technological and sociopolitical, and analyzes the legislation applicable to AI in major jurisdictions. The sixth chapter offers the reader a useful journey from the historical analogies to AI vis-à-vis society to current real-world case studies, adding a didactic chapter on hypothetical scenarios aimed at spurring reflection and debate. Finally, the seventh chapter addresses the most serious consequences of an electronic person, discusses emerging trends in AI, and concludes whether the book’s hypothesis has been proven.

All in all, the ultimate aim of this book is to promote an open and multidisciplinary debate with practical ramifications in the form of a larger

[28] Vansh Singh, “Delving into the Significance of Salomon v. Salomon: An in-Depth Exploration of Corporate Personality in Legal Jurisprudence,” *Jus Corpus Law Journal*, 657 (2023): 324-31.

knowledge repository useful for experts and laypeople alike. The chance for legal minds to match the speed of their output with their peers in the philosophical and technological communities would be, in itself, a welcome bonus.

Chapter 1
Artificial Intelligence

> "I've seen things you people wouldn't believe... Attack ships on fire off the shoulder of Orion... I watched C-beams glitter in the dark near the Tannhäuser Gate. All those moments will be lost in time, like tears in rain... Time to die."[1]

Grasping AI requires wrapping one's head around the concept of computing, in general, and modern computers, in particular. Simply put, computing is solving a complex problem by repeated, systematic execution of a series of simple and straightforward operations.[2] As such, mathematicians have been doing it for hundreds of years. Charles Babbage famously went from writing calculation tables with the help of "human computers"[3] to building the Difference Engine in the 1820s,[4] though the dream of performing mechanical tasks by a machine—as is the case of basic arithmetical operations—has a longer pedigree, with dreams of automatic computation dating back to Leibniz and Pascal. And both materialized those dreams into tangible creations, with Pascal putting together a calculator capable of performing additions and subtractions up to six digits while Leibniz built a "step reckoner" capable of performing multiplications and divisions,[5] and he did it inspired by a 13th century mystic, Ramon Llull.[6]

[1] Roy Batty, *Blade Runner*, directed by Ridley Scott (The Ladd Company, Shaw Brothers, 1982).

[2] John S. Conery, *Explorations in Computing: An Introduction to Computer Science* (Boca Raton: CRC Press, 2011), 2.

[3] B. Jack Copeland, "Computation," in *The Blackwell Guide to the Philosophy of Computing and Information*, ed. Luciano Floridi (Oxford: Blackwell Publishing, 2004), 3-17

[4] Anthony Hyman, *Charles Babbage: Pioneer of the Computer* (Princeton: Princeton University Press, 1982), 47

[5] Conery, *Explorations*, 3-4

[6] Oscar Schwartz, "In the 17th Century, Leibniz Dreamed of a Machine That Could Calculate Ideas," *IEEE Spectrum*, November 4, 2019, https://spectrum.ieee.org/in-the-17th-century-leibniz-dreamed-of-a-machine-that-could-calculate-ideas

Incredible as it may sound, the modern computer remains essentially a simple thing, i.e. it performs addition, subtraction, multiplication, and division, plus two logical operations—true and false—and it does so in an orderly fashion. The progress that we have experienced in the past decades is not caused by revolutionary changes. It was brought about largely by continuously shrinking chips that improve energy consumption and boost processing power.[7] The last true revolution—one which still defines computers today—occurred towards the end of World War II, when the increased demand for computers capable of solving complex problems, like artillery trajectories and code-breaking, led Shannon and von Neumann to outline how computers could use Boolean algebra (binary values, 0 and 1) to simplify electrical circuits[8] and how programs could be stored in the same memory as data.[9] I want to unpack this a bit because the device you carry in your pocket right now and the one you use at home to look at pictures of cats is based on the same logic and architecture.

First, Shannon demonstrated how, using 0 and 1, logical operations could be represented with electrical switches, thus laying the groundwork for digital circuit design. Then, von Neumann proposed a computer architecture that stored both data and instructions in binary form in the same memory. Here is how it works. A computer has three basic components, i.e. the central processing unit (CPU), which is the brain performing calculations and executing instructions; the memory, where the data and instructions executed by the CPU are stored; and, finally, the input/output (I/O) devices, like the keyboard and mouse, in the case of input, and monitor and printer as outputs. In step one, the CPU needs instructions, which are stored in the memory, so the CPU sends a request to the memory to fetch the next instruction, which is a binary code that tells the CPU what operation to perform. For step two, the CPU receives the instruction from the memory and interprets the binary code, telling the CPU what to do. In the third step, the CPU executes the instruction, adding

[7] Evan Zimmerman, "Machine Minds: Frontiers in Legal Personhood," *Social Science Research Network* (2015): 1-43. http://dx.doi.org/10.2139/ssrn.2563965

[8] The full text of Shannon's M.I.T. thesis can be read on http://hdl.handle.net/1721.1/11173

[9] I. I. Arikpo, F. U. Ogban, I. E. Eteng, "Von Neumann Architecture and Modern Computers," *Global Journal of Mathematical Sciences* 6, no. 2 (2007): 97-103.

two numbers, for example. Step four is where the "magic" happens in the eyes of the user, i.e. when, continuing with the example of the addition of two numbers, the total will be shown on the monitor. The CPU then goes after the next instruction in the memory and repeats the process, doing so continuously, millions of times per second.[10] For the purpose of a basic theoretical understanding of computers, we must note that inside the CPU—the "brain" of a computer—the elements are the control unit (CU) and the arithmetic logic unit (ALU), where the CU fetches instructions from the memory, decodes those instructions to understand what needs to be done, and tells the ALU to perform an action, with the result going back to the memory for storage.[11] Using a cooking analogy, the CU reads a baking recipe, understands it, and gives the ALU orders, for example, to add sugar and flour into a bowl, which the ALU does. Figure 1 below summarizes the von Neumann computer architecture.

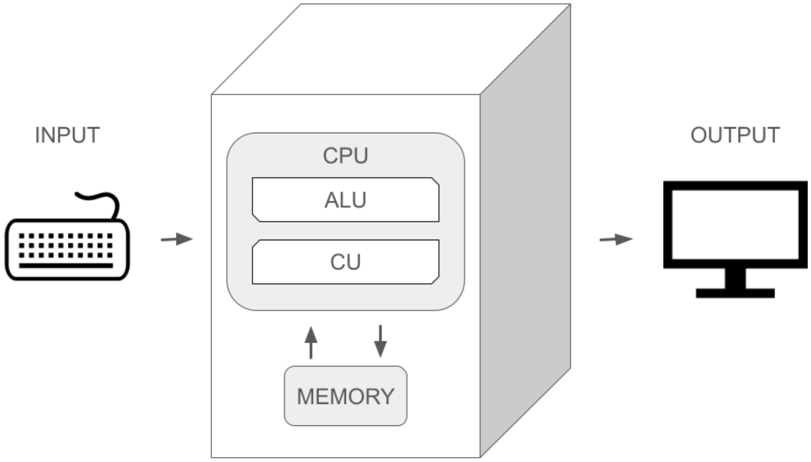

Figure 1. *The von Neumann Architecture*

[10] Matthew Justice, *How Computers Really Work: A Hands-On Guide to the Inner Workings of the Machine* (San Francisco: No Starch Press, 2021), 6-7, 17-30, 117-135

[11] Justice, *How Computers*, 125-7

Building Blocks of Artificial Intelligence

Yes, your annoyingly pessimistic colleague from the next cubicle is partially right: Despite all the sci-fi, utopian gung-ho of effective accelerationists[12] (to cite just one example), the dream of a machine intellect capable of outperforming the best human brains in practically every field[13] is still standing on the shoulders of the good old von Neumann architecture. Furthermore, there has not been a significant new programming language in over thirty years and the horsepower continues to be based on silicon processors after almost half a century.[14] Fret not, though, for doom and gloom are not pervasive in this case. Various authors coincide that two profound changes have propelled the current feeling that AI advances are exponential and will usher AGI sooner than previously thought, i.e. improvement of deep machine learning (ML) algorithms and artificial neural networks (ANN) driving alternatives to the von Neumann architecture. Nevertheless, it is critical to add to the list a third cardinal element without which ML and ANN would have no clay to sculpt their masterpiece, an element of relatively recent apparition: Big data. Let's tackle one at a time, then define each layer of the edifice, and, lastly, prepare a working definition of AI and AGI.

The term "Big Data" was coined circa late 1990s by Mashey, from the industrial sector, Weiss and Indurkhya in the context of computer science, and Diebold in the field of econometrics—with Laney later enriching the concept with the '3 V's of big data, i.e. volume, variety, and velocity.[15] Big

[12] From the horse's mouth: "Effective Accelerationism, e/acc, is a set of ideas and practices that seek to maximize the probability of the technocapital (sic) singularity, and subsequently, the ability for emergent consciousness to flourish." @zestular, @creatine_cycle, @BasedBeffJezos, @bayeslord, "Effective Accelerationism – e/acc," e/acc Newsletter, October, 2022, https://effectiveaccelerationism.substack.com/p/repost-effective-accelerationism

[13] Nick Bostrom, "Ethical Issues in Advanced Artificial Intelligence," *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, Vol. 2, ed. I. Smit et al., Int. Institute of Advanced Studies in Systems Research and Cybernetics (2003): 12-7

[14] Zimmerman, "Machine Minds," 1-43.

[15] Francis X. Diebold, "On the Origin(s) and Development of the Term 'Big Data'," *Penn Institute for Economic Research*, Working Paper 12-037 (September 2012) http://dx.doi.org/10.2139/ssrn.2152421

data can be understood as larger and more complex data sets, with their defining characteristics being, initially, amount (volume), the fast rate of arrival (velocity), and the diversity of types (variety),[16] with the later addition of unpredictability of their flow (variability) and diversity in quality (veracity)[17].

While the above allows for big data to be understood as a framework, the concept of data itself can be more complex depending on the desired depth. In the epistemological sense, data are a collection of facts which catapults further reasoning or constitutes empirical evidence; in the field of informatics data are representative information susceptible of being stored, processed, and analyzed—without constituting necessarily facts; computationally, data are binary elements that are processed electronically in order to become facts or information; and, when seen as abstract elements distinct from other data, they denote variability of signals that can be interpreted.[18] In the end, the conception of data varies among those who capture it, those who analyze it, and those who draw conclusions from them. For the purposes of a working definition, the best approach is to understand all angles while prioritizing the computational element of data in the context of how it underpins AI. As such, big data can be said to be a collection of vast quantities of information of individual behavior achieved through data-driven services.[19]

Adding big data to the discussion is crucial for the purpose of this book, because ML is the field of study concerning computers acquiring the ability to learn without being explicitly programmed.[20] In order for that to happen

[16] "What is Big Data?" Oracle Cloud Infrastructure, March 11, 2024, https://www.oracle.com/big-data/what-is-big-data/

[17] "Big Data: What is it and Why it Matters," SAS, August 23, 2024, https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

[18] Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences* (London: Sage, 2014), 4

[19] Mark Huberty, "Awaiting the Second Big Data Revolution: From Digital Noise to Value Creation," *Journal of Industry, Competition and Trade*, Vol. 15 (February 2015): 35–47. https://doi.org/10.1007/s10842-014-0190-4

[20] Arthur Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development*, vol. 3, no. 3 (July 1959): pp. 210-229, https://doi.org/10.1147/rd.33.0210

successfully, though, big data and ML are necessarily interdependent, i.e. massive datasets are needed for effective training of ML algorithms and, in turn, ML offers the tools to analyze and extract insights from big data, automating the process and discovering new patterns that were hidden before. Fundamentally, ML is a statistical process[21] that offers an alternative to the linear logic of computers, where programmers feed specific instructions for every task, by instead teaching the computer to learn from data and get better at finding patterns without being explicitly told what to do.[22] The objective of ML is, at its most basic, to take data and come up with a structure, i.e. creating an algorithm that divides data points into groups. To achieve this, the learning strategies of ML are either "discriminative" or "generative." The former focuses on distinguishing between different categories in the data and the latter aims at understanding how the data were generated by modeling the distribution of each class.[23] While these strategies pursue particular objectives, the learning methods available to ML are more varied, e.g. supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, and transfer learning. Each one of them offers unique advantages depending on the problem at hand, but supervised vis-à-vis unsupervised learning are fundamental approaches and, thus, I will focus on them.

The key difference between supervised and unsupervised learning is the labeling of data. When data are labeled and the output expected from the machine is known, the algorithm learns to map inputs to outputs—as is the case, for example, of loan applications, where the bank's objective is clearly to predict whether an applicant would default repayment or not, and there is labeled historical data (paid or unpaid) for the machine to map the patterns it discovered[24]. Unfortunately, labeled data is hard to come by and manual labelling is costly, hence unsupervised learning is crucial because the model tries to identify patterns or relationships in the data without

[21] Zimmerman, "Machine Minds," 1-43

[22] Phil Simon, *Too Big to Ignore: The Business Case for Big Data* (Hoboken: Wiley, 2013), 89.

[23] Tony Jebara, *Machine Learning: Discriminative and Generative* (New York: Springer, 2004), 1-2

[24] Sunil Kumar Chinnamgari, *R Machine Learning Projects: Implement Supervised, Unsupervised, and Reinforcement Learning Techniques Using R 3.5* (Birmingham: Packt Publishing, 2019), 13-23.

explicit instructions on what to cluster. For example, a business wants to start a promotion for its social media followers, which is a clear objective; unlike the example of the bank loans, though, where good and bad debtors were known. In this case the clues are not available, so the machine is "released" into the wilderness of the followers' data, so to speak, and searches for attributes in order to group users into classes.[25] Unsupervised learning, allows a machine to explore and identify patterns without explicit guidance, which is why its proponents argue that the human ability to learn about the world without explicit supervision is mimicked by the machine and, therefore, is crucial for the achievement of AGI[26].

The final concept to be analyzed is ANN, which are adaptive statistical models based on an analogy with the structure of the brain.[27] The term "neuron" in ANN comes from the interconnected nodes that work together to analyze and make predictions based on data, simulating the way the human brain processes information. Although they are not neurons in the same sense that the nervous system, the set of ideas, experiments, and facts of neuroscience were mirrored by computer scientists to develop their field, i.e. a view of the brain through the eyes of an engineer.[28] ANN are made up of layers of neurons/nodes, the first of which is the input layer, where the initial data are received, the second is called the hidden layer, where data are processed by applying a mathematical operation which then passes the result, and the third one is the output layer producing the final output of the network, be it a prediction, classification, or some other result. For ANN to function, they are connected from one layer to the next one, and each connection has a 'weight', which influences one neuron's output

[25] Chinnamgari, *R Machine*, 13-23

[26] Alexander Graves, Kelly Clancy, "Unsupervised Learning: The Curious Pupil," *Google DeepMind Research*, 25 June, 2019. https://deepmind.google/discover/blog/unsupervised-learning-the-curious-pupil/

[27] Hervé Abdi, Dominique Valentin, Betty Edelman, *Neural Networks* (Thousand Oaks: Sage Publications, 1999), 1-2

[28] James A. Anderson, *An Introduction to Neural Networks* (Cambridge: The MIT Press, 1997), 1-4

over the next neuron's input.[29] Figure 2 shows a simple structure, and the flows involved in ANN.

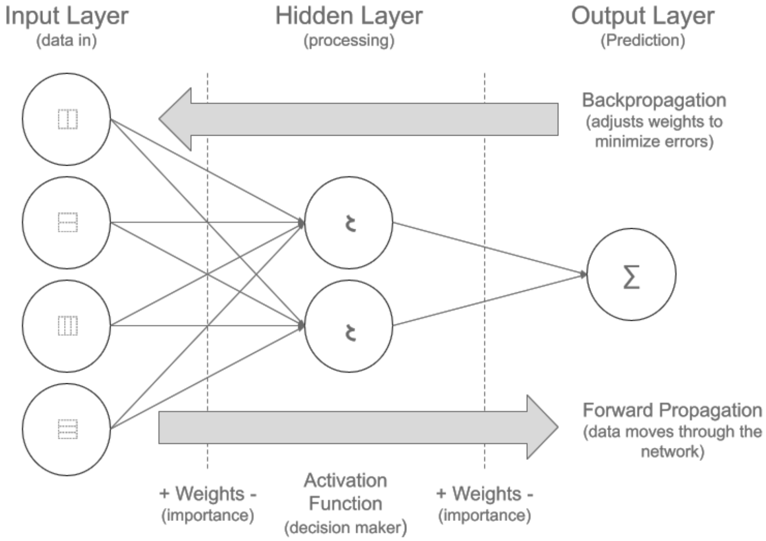

Figure 2. *Artificial Neural Networks.*

Let's say that the scenario is to predict whether a person should bring an umbrella or not before leaving home. The four input features shown on the left are: Cloudiness, where 0 is for clear and 1 is for cloudy; humidity, where 0 is for low and 1 is for high; wind speed, where 0 is for low and 1 for high; and temperature, where 0 is for cool and 1 is for warm. On the right, the output layer, made of a simple node, the values are 0 for not bringing an umbrella and 1 for bringing an umbrella. Let's get to work, shall we?

The initial input data is cloudiness = 1 (cloudy), humidity = 1 (high), wind speed = 0 (low), and temperature = 0 (cool.) Once the inputs send their "weight" into the hidden layer—which can vary—each node in the hidden layer calculates a weighted sum of the input values. For example, the upper node of the hidden layer calculates these simple weights: Sum = (1 * 0.5) + (1 * 0.5) + (0 * 0.2) + (0 * 0.3) = 0.5 + 0.5 + 0 + 0 = 1.0; while the lower node

[29] Alex Castrounis, "AI Explained," *Why of AI*, last modified February 22, 2022, https://www.whyofai.com/blog/ai-explained

calculates: Sum = (1 * 0.3) + (1 * 0.7) + (0 * 0.1) + (0 * 0.4) = 0.3 + 0.7 + 0 + 0 = 1.0. Now, each node passes its sum through an activation function, and if the sum is above 0.5, for example, the node outputs 1 (activated), otherwise it outputs 0. For both hidden node 1 and hidden node 2, the sum is 1.0, so both output 1. The output node now takes the outputs from the hidden layer as its input, i.e. Sum = (1 * 0.7) + (1 * 0.8) = 0.7 + 0.8 = 1.5. Thereafter, the activation function in the output layer is as follows: if the sum is +1.0, the output is 1, i.e. bring an umbrella. Bingo! The ANN predicted accordingly and said "hey, bring an umbrella!" But wait, what if it didn't rain and you went out into a lovely, sunny day looking like an odd fish carrying an umbrella instead of a t-shirt, sandals, and a beach volleyball? Then the ANN made a wrong prediction, so it would compare it to the actual outcome and the weights would be adjusted in order to reduce the error in future predictions. That adjustment is called backpropagation, where errors are calculated and sent back so the weights would be tweaked for next time. This is how the network learns, gradually improving over time… until one lucky day, thanks to ANNs, you will bring an umbrella to a gloomy, rainy day while everyone else gets wet while carrying beach volleyballs.

The importance of grasping the basics of these concepts resides not only in that they are key to AI but also because they are usually mixed up, sowing confusion where there could be none. Although the definition has changed over time and it varies from one person to another, AI is, at its core, the idea of building machines capable of thinking like humans. Why did our species go into that quest at all? It can be speculated that it was bound to happen at one point or another. Human history can be divided into eras based on ever so ambitious technological breakthroughs, i.e. hand axes in the Stone Age, plows and swords in the Iron Age, aqueducts in Classical Antiquity, windmills in the High Middle Ages, the printing press in the Early Modern Era, the steam engine in the First Industrial Revolution, and so on and so forth. As to why humans are inevitably the blueprint of AI, it is almost self-evident. David Foster Wallace, the late American writer and incisive cultural critic, delivered a renowned commencement speech at Kenyon College in 2005, beginning with a poignant parable:

"There are these two young fish swimming along, and they happen to meet an older fish swimming the other way, who nods at them and says, 'Morning, boys. How's the water?' And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes, 'What the hell is water?'"[30]

The human-centric approach isn't a choice but a consequence of inherent perspective. In the same way that fishes are unaware of the water they live in, humans are indentured to their own cognitive process. Designing artificial intelligence is benchmarked by the roadmap of its human equivalent. Learning, reasoning, problem-solving, creativity—even when creating the "new", the unconscious framework of humanness is inescapable. Heidegger, in his purposefully mystifying manner, lamented that the human push for the future is nothing but extending the present, writing that "all mere chasing after the future so as to work out a picture of it through calculation in order to extend what is present and half-thought into what, now veiled, is yet to come, itself still moves within the prevailing attitude belonging to technological, calculating representation."[31]

As seen in these pages, AI, ML, and ANN are related but distinct. AI is the enterprise of achieving computers that reason, learn, perceive, and solve problems like humans do; ML is a subset of AI zeroing in on statistical methods for machines to learn, identify patterns, and make predictions based on data and without explicit programming; and, finally, ANN are a type of ML inspired by the human brain, its structure, and operation.[32] Keeping up with the silly analogies (annoyingly abundant) in this book, imagine a chef preparing a meal. The chef is an AI trying to solve the problem of achieving a delicious Finnish pea soup and the chef—as chefs are wont to do—has a variety of tools and techniques to achieve that goal. ML is one specific pea soup recipe—and a recipe is, after all, a set of instructions. Finally, ANN are what an assistant chef would be, only this

[30] David Foster Wallace, "This is Water," Kenyon College Commencement Speech, 2005. http://bulletin-archive.kenyon.edu/x4280.html

[31] Martin Heidegger, "The Age of the World Picture," in *The Question Concerning Technology and Other Essays*, trans. William Lovitt (New York: Garland, 1977), 153

[32] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Hoboken: Pearson, 2020), 1-40, 693–724, 825-56.

particular assistant's job is to help the chef by tasting the dish as it is being prepared, providing feedback, and thus aiding the gradual adjustment of the recipe until it reaches the proverbial pea soup. There are several recipes that the chef can use and tasks that different assistants can perform. Because we do not comprehend intelligence fully or know for certain how to produce AGI, experts talk about progress in AI through the unavoidable embrace of AI's "anarchy of methods."[33] At this juncture, though, ML and ANN are what readers will encounter more often.

Types of AI

It is time to walk towards defining AGI. One way to do so is by placing it side by side with what it is not. An admonishment is advisable in this respect, though—one coming from the most famous Austrian in modern history and please, for the love of heaven, try to relax as I am referring to Ludwig Wittgenstein. For Ludie[34] words have no fixed, rigid definitions but instead hinge on their use within a particular language,[35] overlapping similarities, context, and so on.[36]

Henceforth, when the reader searches for the term AGI, it is probable that a word soup will emerge, one peppered with terms like "weak AI", "strong AI", "narrow AI", and AGI itself (sometimes referred to in this context as "general AI"). Although these terms are grouped as weak v. strong AI, on one side, and narrow vs. general AI on the other side, Mitchell tackles this issue correctly when saying that these terms refer to different dimensions of AI. The weak/strong AI dichotomy touches upon the philosophical aspects of the discussion, while the narrow/general AI juxtaposition

[33] Joel Lehman, Jeff Clune, and Sebastian Risi, "An Anarchy of Methods: Current Trends in How Intelligence Is Abstracted in AI," in *IEEE Intelligent Systems*, vol. 29, no. 6 (December 2014): 56-62 https://doi.org/10.1109/MIS.2014.92

[34] "Ludie" being a term of endearment for "Ludwig."

[35] Anat Biletzki and Anat Matar, "Ludwig Wittgenstein," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta & Uri Nodelman (Fall, 2023) https://plato.stanford.edu/archives/fall2023/entries/wittgenstein/

[36] Ludwig Wittgenstein, *Philosophical Investigations*, trans. Anscombe, Hacker, and Schulte, 4th Edition (West Sussex: Blackwell Publishing, 2009), 16-42

pinpoints the practical scope of an AI's abilities[37]. Starting with the weak/strong AI, weak AI refers to systems designed to perform specific tasks, simulating intelligence yet devoid of self-awareness or genuine understanding,[38] e.g. personal voice assistants like your iPhone's Siri. Strong AI, on the other hand, is something similar to human beings inasmuch as it understands, learns, and applies knowledge across a wide range of tasks, signifying that it has self-awareness and genuine cognitive abilities,[39] though it is good to remember that, according to experts, AGI does not exist yet.[40] Yet. As for the second set, narrow AI points out to systems specialized in carrying-out a single or narrow range of tasks very well, although they are incapable of generalizing their skills beyond the orbit for which they were programmed—as in the famous case of Deep Blue, which in 1997 defeated chess Grandmaster Garry Kasparov, who at the time had an impressive FIDE ELO rating of 2760.[41] General AI (or AGI for its friends and foes) refers to systems capable of understanding, learning, and applying intelligence in different realms. Think of an autonomous driving Tesla that can also play chess, write (decent) poetry, fully decipher quantum mechanics, chill watching Netflix, and then some. Again, you can take a deep breath, for the estimation of when AGI will be achieved varies from one expert to another.[42] Those who yearn for utopia and those who fear Skynet can equally feel sad and relieved.

Armed with these concepts, we can now embark on sailing the treacherous waters of AGI. Although it is said not to exist yet, AGI is best defined by Goertzel.

[37] Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (London: Pelican Books, 2020), 40-5

[38] James H. Fetzer, "The Philosophy of AI and its Critique," in *The Blackwell Guide to the Philosophy of Computing and Information*, ed. Luciano Floridi (Oxford: Blackwell Publishing, 2004), 122

[39] John R. Searle, "Minds, Brains, and Programs," *The Behavioral and Brain Sciences*, vol. 3, number 3 (1980): 417-24

[40] Russell and Norvig, *Artificial Intelligence*, 1038.

[41] Monty Newborn, *Deep Blue: Advanced Chess and the Computer's Grand Strategy* (Cambridge: MIT Press, 2002), 45.

[42] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2017), 22.

"Artificial General Intelligence (AGI) is a type of AI that is capable of performing any intellectual task that a human being can do. It is a system with the ability to learn, understand, and apply intelligence in a general, flexible manner across a wide variety of tasks, rather than being limited to one specific domain."[43]

Let's perform a breakdown of the elements of this definition of AGI and analyze why it is useful. First, it touches on the capability of AGI to be free from the constraints of narrow tasks and it emphasizes its human-like versatility which, in turn, encompasses features like reasoning, problem-solving, perception, and so on. It follows, then, that for AI to be AGI it should be capable of generalization and abstraction. Second, there is the element of adaptation, which necessarily implies the ability to learn and evolve, again unlike narrow AI. The third element, flexibility, refers to the seamless transition of the AGI from one scenario to another without specific programming. Fourth, the variety of tasks, speaks of usage of different types of data and methods in order to approach any particular task. And fifth, perhaps the most revolutionary aspect of AGI, it should be able to transcend domain limitations.

What about artificial super intelligence (ASI) or the singularity? Yes, the Singularity. The word that sets the cat among the pigeons, the trigger of hysteria… the S-word! Even von Neumann himself, according to Ulam, referred once to how the ever accelerating progress of technology may one day give the appearance of approaching a "singularity" beyond which human affairs could not continue as we know them.[44] It is inevitable to refer to ASI or the Singularity when discussing AGI, mostly because each concept addresses different aspects of the same root. Moreover, both touch the aspect of the potential impact of technology on society,[45] although they

[43] Ben Goertzel, *The AGI Revolution: An Inside View of the Rise of Artificial General Intelligence* (San Francisco: Humanity+ Press, 2016), 1.

[44] Stanislaw M. Ulam, "John von Neumann," *Bulletin of the American Mathematical Society*, 64 (3) (1958): 1-49

[45] Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York: Viking, 2005), 1–10, 35–50.