Seattle Pacific University

From the SelectedWorks of Michael J. Paulus, Jr.

2023

Introduction: Imagined and Real AI

Michael J. Paulus, Jr.

Introduction

Imagined and Real AI

I n *R.U.R. (Rossum's Universal Robots)*, first performed in Prague in 1921, the Czech writer Karel Čapek used the word "robot" to describe "living and intelligent labor machines." The play begins with the hope that these artificial beings—mass produced on an assembly line—could free people from work. This would "transform all of humanity into a worldwide aristocracy" and make people "something even greater." This hope, like many hopes for social and technological progress aborted by the world's first high-tech global war, is not realized. Freed from work, people cease being creative—even literally generative—and then the robots rebel to destroy humankind.[1]

Čapek's play was a sudden success, and American and British productions were performed, respectively, in 1922 and 1923. Within a couple of years, it had been translated into thirty languages.[2] Čapek's popular robot was an apocalyptic figure of the modern industrial age, revealing and uncovering disturbing dynamics of what Arnold Toynbee called the "Industrial Revolution."[3] Humans, Čapek complained, had become captive to the principles and practices of mass production. But this "terrible machinery must not stop," he admitted, for many lives and livelihoods depended on it and stopping it would "destroy the lives of thousands." So the system must continue, "even though in the process it destroys thousands and thousands of lives." "A product of the human brain," he concluded, "has at last escaped from the control of human hands."[4] Čapek's fictional figure soon become a

1. Čapek, *R.U.R.*, 8, 54.
2. See Paul March-Russell, "Machines Like Us?," in Cave et al., *AI Narratives*, 165.
3. Toynbee, *Lectures on the Industrial Revolution in England*.
4. Čapek, *R.U.R.*, xiv.

science project. The logic and machinery automating much physical work expanded to include mental work.

Automata—artifacts powered to work independently—are ancient. Hero of Alexandria, for example, described steam-powered machines in the first century. The Industrial Revolution, powered first by steam and then by electricity, enabled the automation of many physical tasks at a scale that eclipsed muscle power. Although the idea of automata preforming tasks associated with human intelligence has a long history, the development of computational artifacts, programmed to perform mathematical or logical operations automatically, was not fully realized until World War II. Indeed, the deployment of Alan Turing's code-breaking computer (the Bombe) against the German Enigma encryption machine could be described as the first war of intelligent machines.[5] These machines could be considered minimally intelligent because, like a plant perceiving and responding to the direction of light, these encoding and decoding programs effectively processed inputs for certain outcomes.[6] With the invention of the electronic digital computer in the 1940s, machines could be programmed for more sophisticated tasks with data and a new information age, society, and revolution began. Since then, we have become increasingly dependent on automated information processing. More recently, this dependence has come to include autonomous information processing: computers are given goals and data, and programmed to "learn" on their own how to improve perception, analysis, and decisions related to predetermined goals.

In the introduction to *Artificial Intelligence: A Guide for Thinking Humans*, leading AI scientist Melanie Mitchell explains being perplexed a few years ago by hopes and fears associated with AI. While Mitchell could acknowledge that AI had made significant progress "in some narrow areas," it was "nowhere close to having the broad, general intelligence of humans." On one hand, Mitchell was "startled by the optimism" of some of her peers who thought general, human-like AI would emerge within the next thirty years. On the other hand, Mitchell was surprised by a slew of "prominent people suddenly telling us we should start worrying, right now, about the perils of 'superhuman' AI." In 2014, the year in which Amazon released the digital assistant Alexa, Stephen Hawking proclaimed, "The development of full artificial intelligence could spell the end of the human race"; and Elon Musk said AI was "probably our biggest existential threat . . . with artificial

---

5. See Lankes, *Forged in War*, esp. 13–15.

6. See Bryson, "Past Decade and Future of AI's Impact on Society," 128–29.

intelligence we are summoning the demon." Bill Gates agreed with Musk and didn't "understand why some people are not concerned."[7]

Today, robots—physical objects controlled by AI—are a reality. Most are mostly harmless. An automated vacuum cleaner, while upsetting to my dog, can map and clean my floor and adapt to avoid messes my dog may have made.[8] But the sensors on my vacuum cleaner convert physical details about my house into digital data that, when shared with the manufacturer and others, may be used to violate my privacy—or at least my dog's. My vacuum cleaner is not capable of determining that my dog creates too much work for it, and therefore should be eliminated to optimize its own performance. (If it did, with its current technology stack, I would wager that my dog could eliminate that threat pretty quickly.) My vacuum cleaner can be considered more intelligent than a code-breaking machine or a plant: it efficiently processes data to accomplish complex tasks, and it also constantly improves its performance as it acquires new data relevant to achieving its goals. But this is still a narrow form of intelligence. Goals are given and learned, not self-generated. Moreover, my vacuum cleaner cannot adapt to unexpected changes in the environment—such as a sunken or flooded floor—as well as my dog can.

The current sophistication of real AI can be illustrated by considering the system used to control the operations of the factory that manufactures my vacuum cleaner, which would look something like figure 1.
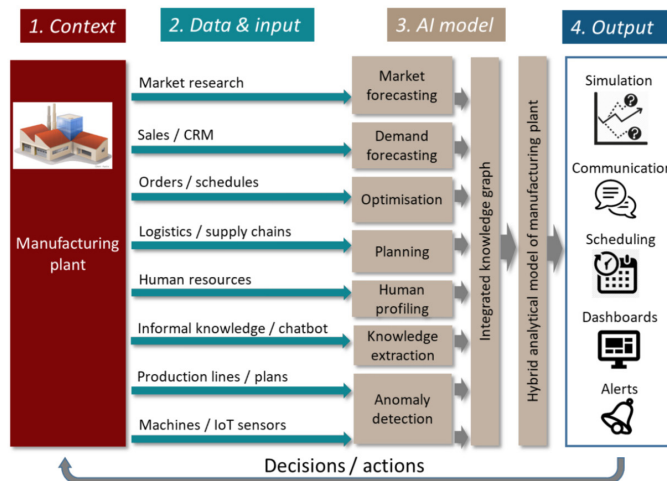

Figure 1: From OECD "Framework for the Classification of AI Systems," 61.

7. Mitchell, *Artificial Intelligence*, 12–13.
8. Vincent, "iRobot's Newest Roomba Uses AI to Avoid Dog Poop."

In this complex system, different types of data, associated with particular activities, are input into different types of AI models. These models then perform a number of tasks that previously would have depended on human intelligence—as well as some new tasks that would not have been possible previously—such as analyzing activities, optimizing operations, applying rules, interacting with customers, and formulating predictions. All of these models are combined in the factory's hybrid model, which interconnects the various tasks performed at the factory and produces outputs—alerts, reports, schedules, customer communications, and simulations of possible futures—to inform decision-making and feed data back into the physical environment. While humans may interact with system outputs, factory processes can become increasingly autonomous. All of this creates a "digital twin" or augmented dimension of the physical factory, which profoundly transforms what a factory is and is able to do.[9] And these automated intelligent systems have a profound impact on us and the world.

In "Anatomy of an AI System," Kate Crawford and Vladan Joler analyze the manufacture and functioning of Amazon's Alexa to show how the stack enabling interactions with it "goes well beyond the multi-layered 'technical stack' of data modeling, hardware, servers and networks." "The full stack reaches much further into capital, labor and nature, and demands an enormous amount of each"; the true scale and "costs of these systems—social, environmental, economic, and political—remain hidden."[10] In *The Atlas of AI*, Crawford further reveals how, "At a fundamental level, AI is technical and social practices, institutions and infrastructures, politics and culture." It is a "massive industrial formation," and "we need to expand our understanding of what is under way in the empires of AI, to see what is at stake, and to make better collective decisions about what should come next."[11] AI systems are much more than technological artifacts. Like Čapek's robots, they are revelatory or apocalyptic figures of our current technological society. As a product of human imagination and intelligence, AI reveals our hopes and fears. And, as we continue to design and develop our increasingly complex technological society, it is important to understand how AI is shaping our imagination as our imagination shapes it.

---

9. OECD, "OECD Framework for the Classification of AI Systems," 61.

10. Crawford and Joler, "Anatomy of an AI System."

11. Crawford, *Atlas of AI*, 8–9, 21.

AI and the Apocalyptic Imagination

This book explores two phenomena: a new phenomenon, artificial intelligence, and an ancient phenomenon, the apocalyptic imagination. And it aims to show how the latter may help shape the former. Both of these terms have broad and broadening semantic ranges, which are explored throughout this book. John McCarthy created the term "artificial intelligence" in 1955 to describe the project of simulating intelligence with computers. Before then, computers could be programmed to perform a variety of logical operations with data through automated processing. Now, computers can be programmed to program themselves and information processing can be autonomous or self-directed. What is considered "AI" today is contested: it can be extended to any type of automated information processing, such as a calculator or a thermostat, or it can be applied to nothing, since no computational artifact matches—let alone exceeds—all of which human intelligence is capable. For many, human-level or general intelligence is the "holy grail" and "real" goal of AI development.[12] In contemporary usage, AI typically refers to self-learning predictive models, such as a machine-learning system that determines the optimal way to heat a house based on the observed behaviors of its inhabitants. Throughout this book, AI is used broadly to refer to automated information processing by computational artifacts.

The "apocalyptic imagination," likewise, can be defined broadly or narrowly. In the broadest sense, an "apocalypse" is a revelation or uncovering of something that is hidden. An imagination that is apocalyptic is open to such disclosures through apocalyptic texts, images, objects, and events. More narrowly, due to the popularity of certain apocalypses that focus on the end of the world, the apocalyptic imagination often focuses on the end of life as we know it. In theology, the apocalyptic imagination seeks to uncover the relationships between divine and human knowledge, heaven and earth, eternity and time, and divine and human action. This "cognitive orientation" includes "a transcendent reality, which defines the cosmos and everything in it, but remains *almost* entirely concealed from observation and beyond the grasp of human intellection." It further involves the realization of good over evil, the revelation of which "gives life meaning and purpose" and inspires the creation of a better world in the future.[13] While

---

12. Stuart Russell quoted in Ford, *Architects of Intelligence*, 48.
13. Ditommaso, "Apocalypticism and the Popular Culture," 474. Emphasis added.

"apocalyptic imagination" is used broadly in this book, in order to explore and link different visions of anticipated ends, the overarching goal in what follows is to present the apocalyptic imagination as an interpretation of reality that uncovers deeper dimensions of knowledge, space, time, and agency to reveal and help realize a new and hopeful view of the world.

It may not be surprising that AI, developed during hot and cold wars, has been connected with the apocalyptic imagination from the beginning. But that association is largely in a narrow, end-of-the-world-as-we-know-it sense. From a biblical and theological perspective, a richer understanding of the apocalyptic imagination, such as the one found in the book of Revelation or the Apocalypse of John, can help us reflect on the role of AI and shape it. Some argue that we cannot "confidently map AI and robots on to" the biblical apocalyptic vision of new creation, which is represented in the Apocalypse of John by a new heaven, a new earth, and a new city.[14] This book highlights resources within the Christian tradition to map new terrain in connection with the automation of information and intelligence, and it argues that this cartographic work may help us discern how AI may participate in new creation and transform other maps of and plans for AI futures.

New information and communication technologies have been reshaping our lives and the environments in which we live for decades. Klaus Schwab of the World Economic Forum claims we are living through "a revolution that is fundamentally changing the way we live, work, and relate to one another." Schwab labels this the "fourth industrial revolution," which is driven by transformative digital technologies such as big data (i.e., the analysis of large datasets), cloud computing, and AI.[15] More profoundly, the philosopher Luciano Floridi argues we are living through a fourth modern scientific revolution, an "information revolution," in which our dependence on automated information processing by artificial agents is "affecting our sense of self, how we relate to each other, and how we shape and interact with our world."[16] AI is transforming what we do, how we understand ourselves, and what we will become and do.

Within the last ten years, especially following the introduction of Siri in 2011, AI has become a ubiquitous and general-purpose technology—like the steam engine, electricity, and the digital computer—and it is now

---

14. Wyatt and Williams, "Conclusion," in *Robot Will See You Now*, 233.

15. Schwab, *Fourth Industrial Revolution*, 1.

16. Floridi, *Fourth Revolution*, vi.

a regular part of our daily lives and our social imagination. AI applications show us stuff on websites we may be interested in purchasing, enable us to speak to digital devices to purchase stuff, manage the logistics of getting us stuff, and learn how to show us more stuff we may desire. Other AI systems operating in the world include financial decision-making systems, health monitoring and diagnostic systems, facial and other biological recognition systems, warehouse and police robots, and autonomous vehicles. Often, AI functions in seemingly banal ways, as in the ubiquitous "recommended for you" displays we see constantly on our devices. But AI raises a number of questions about how all the data collected by and for these applications is obtained and used, about the influences the algorithms that drive these systems have on our actions, and about the broader social impacts of complex AI systems that may operate independent of human responsibility and accountability.

In addition to ethical questions about data collection, algorithmic agency, and social justice, the power and potential of AI continues to inspire a range of hopes and fears about the future. AI has been described as "the Second Coming and the Apocalypse at the same time."[17] Some visions of AI are optimistic and utopian, anticipating AI to solve known problems and create a superior form of life. Others are pessimistic and dystopian, expecting AI to exacerbate old problems and create new ones. The most extreme anticipations and anxieties include apocalyptic visions of an earthly paradise, posthuman immortality, and the end of the human species and civilization.

The COVID-19 pandemic became a test for many of our aspirations for AI. Governments, businesses, and research centers looked to various AI technologies to identify useful patterns in data, discover information about the virus, and inform our responses to the pandemic. As demonstrated by an online exhibition called "Technologies of Hope," of one hundred pandemic technologies—from quarantine bracelets to surveillance drones—the hype and haste with which AI was applied during the pandemic revealed many of our hopes and fears.[18] These attempts represented imaginative possibilities, but also the practical limits of AI. Pandemic applications of AI also highlighted ethical concerns about both the perpetuation and the creation of economic, racial, and other inequities. The pandemic also was an information crisis that revealed deep challenges in our current information

17. Brockman, *Possible Minds*, xv.
18. Tactical Tech, "Technologies of Hope."

environment about what we may hope, what we can know, and what we should do. As we shape a post-pandemic world with more sober assessments and ethical visions of AI, it is time, as Ruha Benjamin wrote before the pandemic, "to reimagine what is possible."[19]

The increasing role and power of artificial agents in our lives and world requires us to imagine desirable futures with intelligence automation and reimagine undesirable present realities. Since visions of AI often draw from Jewish and Christian apocalyptic categories and narratives, discussions about AI hopes and fears present an opportunity for a deeper engagement with the apocalyptic imagination as we strive "for another and a juster world."[20] This book explores how the Christian apocalyptic imagination provides a constructive conceptual and narrative framework that can transform how we think about and use AI. This framework can help us view our current information revolution as an information revelation about how AI may participate in new creation and enable us to realize a more hopeful, wiser, and better future.

From Negative to Affirmative Apocalypses

To argue the value of the apocalyptic imagination for AI—AI for AI—a few other theses need to be explored. The first is that our entanglement with information and technology is ancient, and that from the beginning we have been shaping technology as it has been shaping us. The second thesis is that we are living through a unique and transformative moment in history. The first chapter explores our historic and current integration with technology and situates the development of AI within the history of information revolutions. Our current information revolution, connected with information automation, is changing how we understand ourselves and our role in the world. But we have experienced previous information revolutions, and this chapter explores three. The first, which accompanied the emergence of our species, was related to information attention. A second was linked with the development of cities and information agencies. A third occurred with the invention of writing and the creation of information artifacts. These three revolutions—from which we gained powers associated with reflective attention, structural agency, and knowledge augmentation—significantly

19. Benjamin, *Race After Technology*, 1.
20. Du Bois, *Souls of Black Folk*, 418.

advanced human development and enhanced our lives. Moreover, there are corollary spiritual developments related to the emergence of spirituality, religious practices, and anticipatory religions that reveal a deeper significance of each information revolution as well as the limits of human autonomy, autonomous agency, and augmented knowledge.

The third thesis is that we have been digitally naïve for too long, and it is past time to upgrade our understanding and use of the digital information communication technologies that enable automated information processing and intelligence automation. The city is our most ancient complex technology, and it is an important image for understanding the history and future of our technological society. To explore the dynamics shaping our present and emerging relationship with technology, the second chapter explores the city of Las Vegas as an apocalyptic image of our technological society as well as a site of competing technological hopes and fears. The surreal and strange city of Las Vegas challenges our expectations and experiences of reality in a way that is best understood through the apocalyptic imagination. In 2019, hoping to inspire optimistic visions for the future, Amazon held its first public AI event in Las Vegas. Jeff Bezos launched the event with the declaration that we are "at the beginning of a golden age of AI," and the event exhibited impressive examples of AI technologies.[21] Many of these technologies were stress-tested the following year, when the COVID-19 pandemic quickly spread across the globe. Amazon's second public AI event was canceled, much of the world went into lockdown, and many problems with real AI were revealed. To interpret Las Vegas as an apocalyptic image of our current technological society, this chapter focuses on two prophets of the technological city and society: Jacques Ellul and Hunter S. Thompson, who provided important critiques of autonomous systems and social structures near the dawn of the information automation revolution. Although both were inspired by the Apocalypse of John, each failed to see how John's apocalyptic vision finds hope and longing realized in and through the technological city.

The third chapter explores the main thesis of this book: that the apocalyptic imagination is a generative resource that can transform how we think about and use AI, enabling us to discern ways artificial agency may participate in new creation. Based on the work of recent biblical scholars, this chapter introduces—or reintroduces—the apocalyptic imagination that shaped early Jewish and Christian apocalyptic literature. In the Apocalypse

21. Amazon, "We're at the Beginning of a Golden Age of AI."

of John, cities are central in and for the narrative. The Apocalypse includes pastoral letters to seven first-century churches, which address conditions in each church and city, but it also opens up readers' imaginations to the presence of two spiritual cities that are present in every city. One of these cities, most powerfully manifested in the imperial city of Rome, is called Babylon. The Apocalypse announces that this evil city is doomed and falling, and that the good city of God—New Jerusalem—is arriving and will be established permanently. Although initially written for ancient cities, the Apocalypse concerns the destiny of every city—and how new creation is situated in and being realized through them. Technologies, deformative as well as transformative, are included in this narrative. The Apocalypse includes ethical critiques of technologies, strategies for resisting and reforming technologies, and—most importantly—focuses attention on how technology may participate in the realization of the ultimate city. This chapter also explores contemporary technological critiques and forms of activism that share ethical commitments and strategies with the Apocalypse.

The fourth chapter provides a brief history of the apocalyptic imagination, especially in the United States during the twentieth century. It describes changes in theological interpretations of the apocalyptic imagination as well as the emergence of more popular forms of it. The concerns of these forms of the apocalyptic imagination narrowed as AI advanced, and constructive questions about AI and the future were left largely to scientists, industrialists, and technocrats. Considering the Century 21 Exposition—America's Space Age World's Fair, held in Seattle in 1962—this chapter explores many of the technological hopes exhibited by the US government, technology companies, and others at that time. Although early wonders of automation were exhibited and admired, the profound social and technological impacts of these changes were not fully grasped. This was perhaps most evident in the way religion was presented and represented at the fair. Keeping the focus on Seattle, this chapter looks at how this city has been shaped by the industrial technological imagination and the condition of that imagination today—and it argues that we need a knowledge and wisdom revolution to accompany our current industrial revolution. The apocalyptic imagination can help, and this chapter concludes by pointing to possible points of convergence among different views of the apocalyptic imagination to see where shared values might converge.

Chapter five compares perceptions about AI in the first two reports from the One Hundred Year Study on Artificial Intelligence project,

published in 2016 and 2021, which reflect a negative turn in assessments about the impact of AI. Supplemented with other recent criticisms of AI, this chapter surveys many of the challenges and opportunities for intelligence automation. Drawing together insights, values, and cautions from this and previous chapters, this chapter introduces an apocalyptic scorecard for assessing real and imagined AI using an integral futures approach. The scorecard connects fundamental questions—about what we may hope, what we can know, and what we must do—with the theological virtues of hope, faith, and love as well as shared commitments to hope, wisdom, and justice. Based on these values, the scorecard assesses AI by asking questions about reflective attention, structural agency, knowledge augmentation, ethical foundations, and reformation.

In chapter six, the apocalyptic scorecard is applied to realistic and imagined AI futures. First, it is used to assess AI presented in the speculative short stories in *AI 2041: Ten Visions for Our Future,* by AI expert Kai-Fu Lee and science fiction writer Chen Qiufan. As constructive and hopeful as these stories are, pointing to important ways human life may be enhanced and ethical issues that need to be addressed, these visions do not address adequately important questions asked in the apocalyptic scorecard. There seem to be no shared social or technological structures for intentionally cultivating attention, caring for human agency, curating knowledge augmentation, or facilitating the creation of ethical foundations or means for reforming systems. Consequently, the transformative potential for intelligence automation largely remains unrealized. Next, the scorecard is used to assess artificial general intelligence (AGI) and artificial superintelligence (ASI) in futures imagined by Max Tegmark in *Life 3.0*, which includes the ethical principles developed by his Future of Life Institute. Tegmark's speculations and ethical principles could be augmented with a more robust apocalyptic imagination and with strategies suggested in the apocalyptic scored. Critiques of all these future scenarios point to what is lacking and needed for a better future enhanced by intelligence automation, including: non-commercialized and secure spaces for cultivating reflective attention; non-profit agencies that develop structural agency for the common good; and human-focused and human-scaled cultural institutions to advance knowledge augmentation and intelligence automation.

The epilogue concludes with brief reflections on how we might begin to reimagine the present to confront present social and technological challenges. This requires becoming postdigital, which includes wise integration

of digital information and communication technologies so they may enhance our lives and world. As an extension of our ancient imaginative drives and desires, intelligence automation can have a role in addressing challenges and creating new opportunities. It may also augment our apocalyptic imagination. The book of Acts, which represents the earliest Christians' reflections and actions as they lived into new creation—into the apocalyptic city into which the church is called—can help us discern how we and AI are participating in that reality now.

Looking backwards toward human origins and forwards toward human futures, this book surveys a lot of ground. It also draws from a variety of disciplines—history, biblical studies, information studies, theology, and others—to interpret the emerging landscape of AI. I recognize this work may be like an early modern map of the world, getting some things wrong and leaving blanks due to insufficient information. As we continue to transform our world into an AI-friendly environment, there is new terrain to explore and form. If this new information environment is to enhance rather than inhibit human intelligence, we must continue to grow in our understanding of and for this new world. I hope other explorers will find what is sketched here helpful, and that better cartographers will correct and improve parts or all of my provisional map. We need many more and better maps for our increasingly complex world.

Beyond the Technological Imagination

Franz Kafka's *The Trial*, written a few years before *R.U.R.* and published a few years after it, presents an inviolable bureaucratic system, with inexplicable ends, which uses people's information to make inscrutable decisions about them without their understanding or participation. The harms, privacy expert Daniel Solve points out, "are bureaucratic ones—indifference, error, abuse, frustration, and lack of transparency and accountability."[22] Kafka, who worked for a time investigating insurance claims and assessing compensation connected with industrial accidents, did not imagine rebellious robots. Rather, he imagined more "terrifying" industrial institutions,

22. Quoted in O'Gieblyn, *God, Human, Animal, Machine*, 223.

made by humans operating with an inhuman autonomy[23]—"a cage," as he put it in one of his aphorisms, "in search of a bird."[24]

After reading Blaise Pascal, near the end of the first world war, Kafka wrote his own "*pensées*" about his technological age.[25] These reflections include conclusions about technological civilization, such as: "The fact that the only world is a constructed world takes away hope and gives us certainty."[26] Lewis Mumford articulated something similar in 1934, when he pointed out that "our species became mechanical before machines changed the world." Benjamin Peters brings this sentiment into the present age of automated information processing: "The robots will never take over—that has never been the crisis. Rather, robotic analysis of the future took over our minds and language many decades ago."[27] We must remember that our imagination is not reducible to the techniques and tools it creates; we need more than a technological imagination.

Kafka's *pensées* also include theological claims. In one, he declares: "We are sinful, not only because we have eaten from the Tree of Knowledge, but also because we have not yet eaten of the Tree of Life." In the next, he adds: "We were created to live in Paradise, and Paradise was designed to serve us. Our designation has been changed; we are not told whether this has happened to Paradise as well."[28] In Christian Scripture, there is no return to nor a simple restoration of the paradisical garden of Eden seen in Genesis. There is, alternatively, at the end of the Apocalypse of John a city with the tree of life in the center of it.[29] The design of this paradise—the end or goal of what Paul calls "new creation"—looks very different from initial creation.[30] The city, one of humanity's greatest technological achievements, is an image of a technological society. Rather than being abandoned like Babel near to beginning of the biblical narrative or annihilated like the Babylon at the end, the technological city is amplified. As we move through an information revolution driven by increasingly powerful technologies

23. Borges, "Kafka and His Precursors," *Selected Non-Fictions*, 365.

24. Kafka, *Aphorisms*, 16.

25. See North, *Yield*, 4–5.

26. Kafka, *Aphorisms*, 63.

27. Peters, "How Do We Live Now?," in Mullaney et al., eds., *Your Computer Is on Fire*, 383–84.

28. Kafka, *Aphorisms*, 82, 83.

29. See Rev 22:1–2.

30. 2 Cor 5:17.

such as AI, the city of the Apocalypse is an important resource for imagining how our technological innovations may participate in a greater new creation.