EPILOGUE

AWE, DREAD, HOPE

DEEP LEARNING AND ANCIENT STORIES

Ancient myths articulated timeless hopes and fears
about artificial life, human limits, and immortality.
What could we—and Artificial Intelligence—learn from the classical tales?

THE MIX OF exuberance and anxiety aroused by a blurring of the lines between nature and machines might seem a uniquely modern response to the juggernaut of scientific progress in the age of technology. But the hope—and trepidation—surrounding the idea of artificial life surfaced thousands of years ago in the ancient Greek world. Imaginative myths expressed and struggled with the awe, dread, and hope summoned by the creation of animated statues, attempts to surpass human limits, and the pursuit of immortality. This is a discussion one might say that the ancient Greeks began.[1]

The question of what it meant to be human obsessed the ancient Greeks. Time and again, their stories explore the promises and perils of staving off age and death, enhancing mortals' capabilities, replicating nature. The complex network of myths about Prometheus, Jason and the Argonauts, Medea, Daedalus, Hephaestus, Talos, and Pandora—all raised basic questions about the boundaries between biological and manufactured beings.

The most enduring, best-loved Greek myths—along with many other long-forgotten ancient tales—spin thrilling adventures well worth knowing for their own sake. But when we recognize the old stories as inquiries into *biotechne* (*bios*, life; *techne,* craft), these "science fictions" of antiquity take on new significance. Deeply imbued with metaphysical insight and

forebodings about divine and human manipulation of natural life, the mythical stories seem startlingly of our moment.

The fantasies of imitating and augmenting life inspired haunting dramatic performances on the stage and indelible illustrations in classical vase paintings, sculpture, and other artworks. Meanwhile, in about 400 BC the philosopher-engineer Archytas caused a sensation with the first mechanical bird in flight. By the Hellenistic era, Heron of Alexandria and other brilliant engineers were devising a multitude of automated machines driven by hydraulics and pneumatics. The Greeks recognized that automata and other artifices in natural forms—whether imagined or actual—could be either harmless or dangerous, and they could be used for work, sex, spectacle, or religion, or to inflict pain or death. Clearly, *biotechne*, both real and imaginary, fascinated the ancients.

Taken together, the myths, legends, and lore of past cultures about automata, robots, replicants, animated statues, extended human powers, self-moving machines, and other artificial beings, and the authentic technological wonders that followed, constitute a virtual library and museum of ancient wisdom and experiments in thinking, a priceless resource for understanding the fundamental challenges of biotechnology and synthetic life on the brink today. A goal of this book has been to suggest that on deeper levels the ancient myths about artificial life can provide a context for the exponential developments in artificial life and Artificial Intelligence—and the looming practical and moral implications. I hope that rereading those ancient stories might enrich today's discussions of robotics, driverless cars, biotechnology, AI, machine learning, and other innovations.

We saw how the god Hephaestus made a fleet of "driverless" tripods that responded to commands to deliver food and wine. Even more remarkable was the covey of life-size golden female robots he devised to do his bidding. According to Homer, these divine servants were in every way "like real young women, with sense and reason, strength, even voices, and they were endowed with all the learning of immortals." More than twenty-five hundred years later, Artificial Intelligence developers still aspire to achieve what the ancient Greeks imagined that their god of technological invention was capable of creating.

Hephaestus's marvels were envisioned by an ancient society not usually considered technologically advanced. Feats of *biotechne* were

dreamed up by a culture that existed millennia before the advent of robots that win complex games, hold conversations, analyze massive mega-data, and infer human desires. But the big questions are as ancient as myth: Whose desires will AI robots reflect? From whom will they learn?

In 2016, an experiment in AI machine learning became a cautionary tale, when Microsoft invented the teenage fem-chatbot Tay. Intricately programmed to mimic neural networks in the human brain, Tay was supposed to learn from her human "friends" on the social network Twitter. She was expected to articulate conversational gambits without filters or behavioral supervision. Within hours of Tay's going live on Twitter, malicious followers conspired to cause the bot to morph into a tweeting troll spewing racist and sexist vitriol. Within days, Tay was terminated by her makers. Her easily corrupted learning system dampened optimism about self-educating AI and smart robots, but only momentarily. Tay's replacement, Zo (2107) was supposedly programmed to avoid chatting about religion and politics, but she too went rogue on Twitter.[2]

In Greek myth, the capstone of Hephaestus's divine laboratory was the female android commissioned by Zeus. To punish humans for accepting the technology of fire stolen by Prometheus, Zeus commanded Hephaestus to fabricate Pandora (chapter 8). Each of the gods endowed the artificial maiden with a human trait: beauty, charm, knowledge of the arts, and a deceitful nature. As the vengeful god's AI agent, Pandora executed her mission to unseal a jar of disasters to plague humankind forever. She was presented as a wife to Epimetheus, a man known for his impulsive optimism. As we saw, Prometheus warned humankind that Pandora's jar should never be opened. Are Stephen Hawking, Elon Musk, Bill Gates, and other prescient thinkers the Promethean Titans of our era? They have warned scientists to halt or at least slow the reckless pursuit of AI, because they foresee that once it is set in motion, humans will be unable to control it. "Deep learning" algorithms allow AI computers to extract patterns from vast data, extrapolate to novel situations, and decide on actions with no human guidance. Inevitably AI entities will ask—and answer—questions of their own devising. Computers have already developed altruism and deceit on their own. Will AI become curious to discover hidden knowledge and make decisions by its own logic? Will those decisions be ethical in our human sense? Or will AI's ethics be something "beyond human?"

Released from Pandora's jar—much like the computer viruses let loose by a sinister hacker who seeks to make the world more chaotic—misfortune and evil flew out to prey upon humans for as long as the world exists. In simplistic fairy-tale versions of the myth, the last thing to flutter out of Pandora's box was *hope*. But in darker versions, the last thing in the jar was "anticipation of misfortune." And Zeus had programmed Pandora to slam down the lid, trapping foreknowledge inside. Deprived of the ability to anticipate the future, humankind was left with what we call "hope." As was true of Epimetheus, foresight is not our strong point.

Yet foresight is crucial as human ingenuity, curiosity, and audacity continue to push the frontiers of biological life and death and the melding of human and machine. Our world is, of course, unprecedented in the scale of techno-possibilities. But that unsettling oscillation between techno-nightmares and grand futuristic dreams—that is timeless. The ancient Greeks understood that the quintessential attribute of humankind is always to be tempted to reach "beyond human," and to neglect to envision consequences. We mirror Epimetheus, who accepted the gift of Pandora and only later realized his error.

In 2016, Ray Crowder, an engineer at Raytheon, created three miniature learning robots. He gave the robots classical names: Zeus, Athena, and Hercules. With neural systems modeled on those of cockroaches and octopuses, the little solar-powered robots were endowed with three gifts: the ability to move, a craving for darkness, and the capacity to recharge in sunlight. The robots quickly learned to travel and soon understood they must venture into excruciating light in order to recharge or die. This seemingly simple learning conflict of these creatures that were *made, not born*, parallels human "cognitive economy," in which emotions help the brain allocate resources and strategize. Other AI experiments are teaching computers how human strangers convey goodwill to one another, and how mortals react to negative and positive emotions.[3]

Since Hawking warned that "AI could spell the end of the human race," some scientists are proposing that we could teach human values and ethics to robots by having them read stories. "Fables, novels, and other literature," even a database of Hollywood movie plots, could serve as a kind of "human user manual" for AI computers. One such system is named Scheherazade, in homage to the heroine of *The One Thousand and One Nights*. Scheherazade was the legendary Persian philosopher-storyteller

who had memorized myriad tales from lost civilizations. She saved her own life by reciting these enchanting stories to her murderous captor, the king. The first stories uploaded into the Scheherazade AI were simple narratives that show computers examples of how to behave like good rather than psychotic humans. With the goal of empathetic interactions with human beings and appropriate responses to their emotions, more complex narratives would be added to the computer's repertoire. The idea is that stories would be valuable when AI entities achieve the human mental tool of "transfer learning," symbolic reasoning by analogy, to make appropriate decisions without human guidance.[4]

Computers may be modeled on human brains, but human minds do not work just like computers. We are learning, for example, that our cognitive function, self-reflection, and rational thinking depend on emotions. Stories appeal to emotions, *pathos*, the root of *empathy*, sharing feelings. Stories continue to be alive as long as they summon strong, complicated emotions, as long as they still resonate with real dilemmas, and as long as they are good to think with. We have seen how Greeks and other ancient societies told themselves stories to try to understand humankind's yearning to exceed biological limits and to imagine the consequences of those desires. The insights and wisdom in such myths might deepen our discourse about AI.

Biotechne stories, perpetuated over millennia, are a testament to the persistence of thinking and talking about what it is to be human and what it means to simulate life. We are hardwired to hear, tell, and remember stories. As George Zarkadakis reminds us, stories "are the most powerful means available to our species for sharing values and knowledge across time and space."[5] This raises an intriguing possibility.

Might myths about artificial life in all its forms, like the examples gathered in this book, play a role in teaching AI to better understand humankind's conflicted yearnings? Perhaps some day AI entities will be able to absorb mortals' most profound wishes and fears as expressed in mythic musings about artificial life. Perhaps AI beings might somehow grasp the tangled expectations and fears we have of AI creations. Through learning that humans foresaw their existence and contemplated some of the quandaries the machines and their makers might encounter, AI entities might be better able to comprehend—even "empathize" with—the quandaries that they pose for us.

The rise of a Robot–Artificial Intelligence "culture" no longer seems far-fetched. AI's human inventors and mentors are already building the Robot-AI culture's *logos* (logic), *ethos* (moral values), and *pathos* (emotions). As humans are enhanced by technology and become more like machines, robots are becoming infused with something like humanity. We are approaching what some call the new dawn of Robo-Humanity.[6] When that day comes, what myths and stories will we tell ourselves? The answers will shape how and what our AI creations will learn too.