1 Introduction

December 1992. Berlin. Aglaopheme becomes one with the electric guitar. Her six artificial brains just got reset, her floating-point synapses sprayed with fresh Gaussian noise, leaving her clueless face to her body-environment. An impish childlike spirit, ready to enact her own world. To explore and exploit the most intimate connections between outputs and inputs. To become and make become.

The air is charged with static. The robotic siren hesitantly plucks the E string as she moves the slide in a downward fashion, producing a shrieking, distorted sound. Thus begins a long and monotonic solo, as Aglaopheme obsessively plays that E tone, over and over and over.

Nicolas Baginsky is standing in front of her, mesmerized. The robot never behaved this way before. There must be a bug somewhere in the source code, a glitch in one of the rules governing her behavior. Maybe one of the pieces of hardware, like the pickup or the effects pedal, is not functioning properly. Yet Baginsky listens, enthralled by the strangeness of Aglaopheme's performance as his own brain tries to make sense of the robot's neurotic behavior.

Twenty minutes pass that feel like twenty hours. The vibrations of the E string feed through the guitar pickup and then to a set of distortion devices. The filtered sound waves are converted into a frequency spectrum that is fed back into the robot's six artificial neural networks. Each net emits a digital signal, competing for attention, and for the first time since her synapses have been reset, Aglaopheme's A net screams louder than the E net, and the guitar-bot nonchalantly pinches the corresponding string, as if emerging from her torpor.

During the rest of Berlin's Second Electronic Art Syndrome festival—a three-day, twenty-four-hour event near Alexanderplatz—the robot began producing more complex sonic improvisations, self-organizing around the physical qualities of the sound spectrum, developing a style of her own. For the next two decades, Aglaopheme (see figure 1.1) would practice, rehearse, and perform, constantly adapting to the different sound environments to which she was exposed. Over these years, other robotic music players joined with her to form a jazz improvisation band called The Three Sirens. All music was played live, without any score, the result of the robots' self-adapting interactions that came by jamming together in rehearsals and shows.

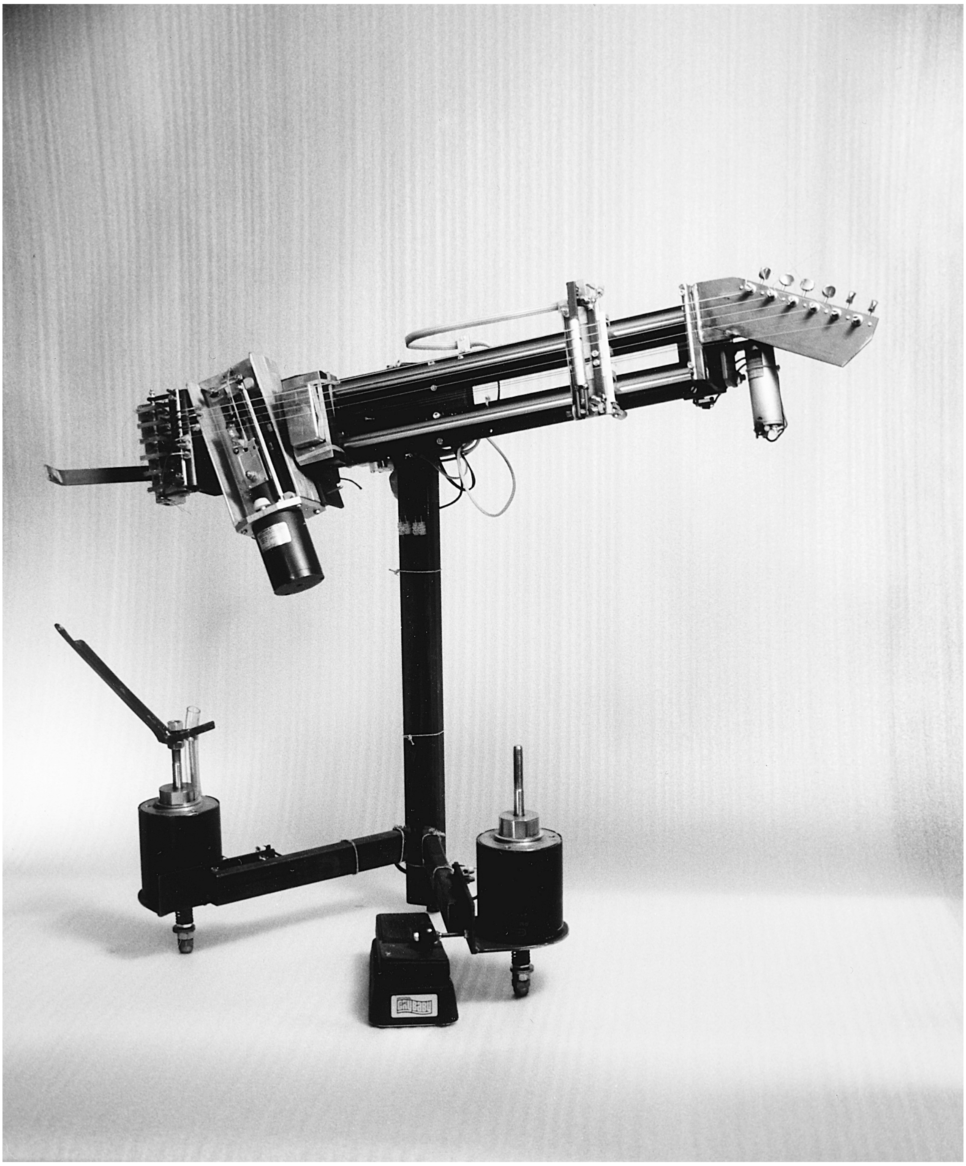

Figure 1.1
Nicolas Baginsky, *Aglaopheme*, 1992. Guitar robot. Courtesy of Nicolas Baginsky.

The Three Sirens constitutes an early example of a growing field of practice within contemporary digital art that engages with a form of *artificial intelligence* (AI) known as *machine learning*. Machine learning proposes to endow machines with intelligence not by programming them directly with logic rules but rather by allowing them to program themselves by learning from their experience. Almost left for dead by the end of the 1990s, this approach has gained impressive impetus since the mid-2000s thanks to critical breakthroughs in fundamental research that triggered an unprecedented interest by the commercial sector and more recently in artistic circles. This has given rise to a loosely defined artistic movement, intimately related to previous computational artistic practices such as cybernetics art, artificial life art, and evolutionary art that I call machine learning art.[1]

Machine learning is a branch of artificial intelligence that allows computers to learn from experience rather than by being "explicitly programmed" (Samuel, 1959). In the case of *The Three Sirens*, the robots learn from the sounds they record live from their environment—including the sound they themselves produce. This data is processed by a training process, an algorithm that gradually adjusts a mathematical function mapping inputs (sound waves) onto outputs (plucking the indicated string).

This work shares many similarities with speech recognition software found on modern mobile devices such as Alexa or Siri. Most state-of-the-art speech recognition algorithms used today are also based on machine learning. Like Baginsky's robots, these systems are subjected to streams of audio data from which they extract regularities. In the case of speech recognition applications, the sound information is associated with specific phonemes that have been previously annotated by humans. The algorithm then tries to predict the right phonemes given new streams of audio inputs, readjusting itself using the tagged data.

These two examples share a common technology, but what truly distinguishes them is how and why they exploit it. Whereas a speech recognition system can be evaluated using a specific metric that measure its ability to perform well in its given task (i.e., translating sounds into text), Baginsky is not really interested in understanding how his robots come to make their decisions, let alone in measuring the accuracy of their musical performances—he simply decides whether he likes or dislikes what he hears. His interest lies in the possibilities that the technology offers and in the fact that these robotic performers are developing a unique, specific style, not by direct, hand-crafted programming but rather by being exposed to the world and finding their own way through it. Perhaps more profoundly, Baginsky is interested in what these adaptive entities, and the process of making them, can teach him about music.

Recent breakthroughs in machine learning have sparked a "4th industrial revolution" (Schwab, 2016) in which adaptive computational systems are rapidly overtaking intellectual tasks in a diversity of fields such as medicine, transportation, and finance. As AI researcher Max Welling has stated, "where steam engines replaced physical labor during the industrial revolution, smart algorithms will soon replace mental labor in what some have dubbed the second machine age" (Welling, 2016). The spearhead of this revolution, *deep learning* (LeCun, Bengio, & Hinton, 2015; Goodfellow, Bengio, & Courville, 2016), involves using several interconnected layers of artificial neurons to represent and interpret patterns present in huge quantities of data. These highly "disruptive technologies" (Bower & Christensen, 1995) significantly alter the way business and society operate, in areas as diverse as self-driving cars, automated medical diagnosis, smart financial trading systems, autonomous weaponry, data mining, and surveillance. Indeed, economy, labor, justice, and the environment, are only a few of the many disciplines that are being transformed to their core by AI.

Deep learning is the last offshoot of a technological lineage that originated in the late 1940s with the science of *cybernetics*, and further expanded in the 1980s with *connectionism*, an approach in cognitive science and AI that rests on using simplified mathematical models of neural networks found in the human brain. Its emergence since the beginning of the millennium is inseparable from the increased access to raw computing power—in particular, due to the development of graphical processing units (GPUs) incidentally pushed by the game and cinema industries—and the exponential growth of data, thanks to the explosive expansion of the internet as a platform for mass social media. AI has become an industrial science. It also has become more accessible: with the development of the internet and mobile devices, AI now lives with us at all time, within reach of our fingertips.

With deep neural architectures computing billions of software neurons and trillions of synaptic connections on GPU clusters owned by the largest IT companies, attuned to our everyday actions in the most unobtrusive, steady, and inexorable fashion, the digital world we were used to, with its recognizable and explainable decision-making procedures based on hand-coded heuristics, is already gone. We are moving into a new era, in which pervasive, seemingly organic algorithms feeding on statistics are replacing rule-based systems, adaptively coupling to humanity in all-encompassing, distributed processes of control and optimization. To understand this new age, we need to extricate ourselves from an outdated vision of computational systems as formal, rule-based, logical constructs and start seeing them for the biologically inspired, statistically driven, agent-based, networked entities they have become.

Our accelerated move into a world populated by these adaptive, autonomous, and uncanny forms of computation constitutes the realization of a particular form of cybernetic society that is highly contingent on the interests of big business[2] and government powers. Artist Memo Akten highlights how in the same way that World War II gave rise to digital computers and the Cold War gave us the internet, today "the mass surveillance related to the War on Terror and Internet business models are giving us Artificial Intelligence and Deep Learning" (Akten, 2016b). In critical need of being reappropriated, deconstructed, torn apart, and democratized, artificial intelligence has thus become a critical space of engagement in the twenty-first century.

Myths and Misconceptions

Despite the increased use of machine learning in many facets of contemporary industrial and commercial culture, until recently one area in which it had not made a meaningful impact was the field of artistic practice. This is certainly not true anymore: in the last decade, there has been an explosive interest in artificial intelligence and machine learning from the art world, with exhibitions such as *Uncanny Valley: Being Human in the Age of AI* (Young Museum, San Francisco, 2020–2021), *AI: More Than Human* (Barbican Centre, London, 2019), *Deep Feeling: AI and Emotions* (Petach Tikva Museum, Tel Aviv, 2019), *D3US EX M4CH1NA* (LABoral, Gijón, Spain, 2019), *Entangled Realities: Living with Artificial Intelligence* (House of Electronic Arts, Basel, 2019) *I Am Here To Learn: On Machinic Interpretations of the World* (Frankfurter Kunstverein, 2018), and *Machines Are Not Alone: A Machinic Trilogy* (Chronus Art Center, Shanghai, 2018). The 2019 edition of the Prix Ars Electronica included a new category called Artificial Intelligence & Life Art, a testimony to the established relevance of AI-oriented approaches to digital arts.

This frenzied enthusiasm is accompanied by a number of myths and misconceptions that complicate the analysis of the situation. Here are some of them.

Myth 1: Artificial intelligence, machine learning, and deep learning are one and the same. This confusion arises because the term artificial intelligence is employed in at least three different ways. The first, which is the one used in this book, refers to a broad field of research that spans many competing approaches. One of these approaches is machine learning, which focuses on designing computer algorithms that can learn on their own. Deep learning is a specific approach within machine learning that uses a particular type of learning system known as artificial neural networks. The second definition of AI reserves the term to state-of-the-art systems while previous approaches are considered devoid of intelligence. According to this definition, in our day and age, only deep learning and other advanced forms of machine learning should be called AI, thus AI is often used as a synonym for these cutting-edge techniques. Finally, the third meaning of the term, more frequently used in everyday parlance, concerns artificial agents that may or may not rely on machine learning, as in the expression "an AI created this masterpiece."

Myth 2: Machine learning art is new. Machine learning can be traced back to the early days of cybernetics in the 1940s. The expression *machine learning* first appeared in the 1950s around the same time as *artificial intelligence*. Artists have been using adaptive or learning computational systems since then, through various artistic movements such as systems art, algorithmic art, robotic art, and evolutionary art. Yet, the presence of such approaches in artistic works is often hard to trace because they are frequently used more as metaphors than as actual techniques. For example, the definitions of concepts such as *learning*, *adaptation*, and even *artificial intelligence* used by artists often differ greatly from the corresponding scientific definitions.

Myth 3: Machine learning can create art without artists. The idea of machines that can altogether replace artists is far from new. Consider, for example, Jean Tinguely's *Métamatics* series of drawing machines from the late 1950s, or Harold Cohen's painting program AARON, which he developed from 1973 to his death in 2016. Although some machine learning systems produce fascinating results, in fact, as we show in this book, machine learning art still requires a lot of labor. While some of the tasks usually associated with computer programming are borrowed, other cumbersome and often expensive tasks arise, such as the building of huge data sets, the fine tuning of training algorithms, and a lot of preprocessing and postprocessing. More importantly, even when some of the choices are left to a machine, art always involves a number of decisions that can be made only by the author of the work.

Myth 4: Machine learning will soon give rise to superhuman intelligence and creativity. This is a common myth about machine learning and also more generally about technologies. The appearance of every new technology has triggered a dread of human obsolescence. For example, in the early twentieth century, the Futurists claimed that mechanical technologies would soon overtake humanity (Versari, Doak, Evans, Bellow, & Curtin, 2016). With regard to current-day machine learning, although opinions diverge on the matter, the scientific community seems to largely agree that current systems are very limited. Although some systems currently in place are impressive, they are still limited to very narrow tasks and require a lot of examples to be trained. They have no common sense and are unable to apply knowledge outside of the problem on which they have been trained. A defining trait of creativity is the ability to "think outside the box," to use one's intuition to come up with ideas that rattle the status quo. Machines are still far from being able to do that—although they may be in a distant future.

Understanding Machine Learning Art

As a field of research, new media arts have not often been a sustained topic of study for art historians, leaving a void that is only starting to be addressed. For the most part, it is new media artists themselves who have started building some of the theoretical tools for understanding their discipline through analyzing their own practices. By comparison, machine learning has been an essential part of the AI ecosystem since the 1950s. Although its role has often been peripheral, its presence has been exponentially growing since the deep learning revolution of the mid-2000s, largely due to machine learning's unprecedented success in tackling major AI-related problems. Machine learning is thus a critical concept whose increasing presence in our world has vast sociotechnical repercussions.

As these technologies are becoming increasingly popular and readily available, there currently exist almost no conceptual guidelines or theoretical frameworks for how to make these works and think about them. There has been some groundwork concerning generative and artificial life practices and concepts such as self-regulation, evolution, and emergence (Kac, 1997; Tenhaaf, 2000; Whitelaw, 2004), but there has been a smaller amount of rigorous work on machine learning and adaptive computation by artists.

The renaissance of machine learning that we have been experiencing since the mid-2000s is occurring in the context of what Simon Penny calls the *crystallization* of new media (Penny, 2017) pushed by strong market forces that undervalue the kind of independent and experimental artistic practices that existed in the 1980s and 1990s. It has become crucial to start building aesthetic theories of machine learning systems in order to allow for a better understanding of artworks that use them, comprehend the processes entailed in working with them artistically, and reposition the role of the artist within this new landscape.

This book seeks to lay the first blocks of such a conceptual framework to comprehend machine learning within the field of new media art. It attempts to bring some clarity to the early stages of the emerging industrial revolution through historical, practical, and theoretical examinations of machine learning in the arts. Through this, it aims to offer conceptual tools, accounts of practice, and historical perspectives to contemporary new media artists, musicians, composers, writers, curators, and theorists, in order to help them grasp what machine learning systems are and how they are related to experimental new media art practice, and to suggest ways that artists can engage with them and learn how to use them. This book does not set out to teach specific techniques but rather intends to translate basic definitions and challenges to nonscientific audiences while it connects them to core issues in new media art.

In order to do so, we put machine learning systems on the operating table and carefully dissect them, examining their different dimensions and components through the lens of art practice. While pulling them apart, we discuss the aesthetic and artistic affordances and significance of each of these elements, while showing ways that artists engage with these elements. The goal of this process is to reveal the inner workings of machine learning and how it can and does operate within art practice. This process is reflected in the organization of the core chapters, each of which focuses on one of the three core components of learning systems: training processes, models, and data.

The book situates machine learning within new media art by studying its relationships with key concepts in art such as indeterminacy, materiality, representation, and authorship. Examples of artworks and creative technologies from a wide variety of domains and formats are presented and discussed, with a focus on works produced by independent artists who are critically engaging with machine learning technologies rather than relying on off-the-shelf systems. As an artist who works with machine learning, I also, whenever it is relevant, bring examples from my own research and practice. This approach serves to illustrate the significance of deep learning technologies for the evolution of new media art in the twenty-first century and beyond as well as the contribution of artists to the field of machine learning.

Why Machines Should Learn

In the Western world, intelligence is often conflated with rational thinking, mathematics, and logic. Back in the 1950s, most artificial intelligence experts thought that the Holy Grail of AI was to perform mathematically oriented tasks such as proving theorems or playing strategic games such as chess and checkers. However, it turned out that such problems are relatively easy for computers to solve because they exist in definite domains based on logical rules and symbols. For example, at any given point in a chess game, there is a finite set of permissible moves, as certain kinds of actions are forbidden, such as moving a piece halfway between two squares.

By contrast, most real-world problems requiring intelligence are very different from playing board games. For example, although specialized work such as translation, financial trading, teaching, research, and medical diagnosis and treatment must follow sets of rules and guidelines, those tasks require a great deal of intuition and experience. Moreover, many tasks that may not seem to require much intelligence because we do them without thinking—such as moving, talking, recognizing objects, or driving a vehicle—are really hard for computers to accomplish.

Take the example of walking. How do we walk? At first glance, one might believe that walking could be expressed with a simple algorithm:

Step 1: Put one foot in front of the other.

Step 2: Repeat.

However, this procedure does not take into account all the dynamics involved in bipedal motion. It may represent the broad picture of most situations, but it does not address possible conditions that would require additional effort, such as moving across irregular surfaces or climbing; nor does it address more challenging situations such as being put off balance or carrying a weight. In fact, bipedal walking in humans is an extremely complex sensorimotor activity involving the coordinated control of muscles in many parts of the body.

The truth is that we do not really know *how* we walk.

But we surely know one thing about walking: we were not born walkers. Gradually, through trial and error and with supervision from our guardians, we *learned* how to walk, one baby step after the other.

How can we program computers to do the things that we do and to know the things that we know when we do not even know *how* we do or know these things? Machine learning suggests that we let computers learn from their experience, just like we did for walking on two feet. Machine learning is hence directly related to the biologically rooted concept of *adaptation*, which refers to a "process whereby a structure is progressively modified to give better performance in its environment" (Holland, 1992, 7). Most machine learning systems learn iteratively, observing flows of data, incrementally refining their understanding of the problem they are trying to solve.

Supervised, Unsupervised, and Reinforcement Learning

Machine learning algorithms are often divided into three subcategories, corresponding to three different types of *tasks*. In *supervised learning* (by far the most commonly used approach), the system learns from labeled data, that is, data for which the appropriate output has been assigned (usually by a human being). For example, imagine a database of pictures in which each has been tagged as an image of either a dog or a cat. The goal of the system is to learn from this information to become good at differentiating cats from dogs in pictures. In other words, given a not previously encountered image of either a cat or a dog, the machine learning algorithm must guess accurately which animal that it represents—in other words, its category or class.

Unsupervised learning is used to make inferences from data sets that do not have such labels. Different outcomes may be desired, such as extracting a more compact representation of the data (e.g., dimensionality reduction or representation learning) or separating the data into different groups. Using the previous example of images of dogs and cats, imagine that we give a database of images to the learning system, and this time the images are unlabeled. We ask the system to classify the images into two unspecified categories. Depending on the database and the configuration of the system, it could decide to differentiate between dogs and cats, but because the classification has been left to the machine learning system, it could instead choose to separate the images into dark and bright, or colored and gray, depending on the data set, the type of machine learning system, and other system characteristics.

Finally, *reinforcement learning* concerns situations in which an artificial agent[3] is evolving into an environment and needs to learn how to behave optimally within it. Good decisions are reinforced by giving the system positive rewards, and bad actions are given negative rewards (i.e., punishments). Common uses of reinforcement learning include robot control, financial trading, delivery management, and adaptive agents for game AI.

For example, imagine a trading software agent that tries to maximize its gains on the stock market. The program chooses either to buy or to sell some shares, on the basis of its observation of the market, which can include the prices of other shares and other information sources such as date and time, financial news, and so forth. On the basis of these decisions, the system receives a reward proportional to the money it has won (positive reward) or lost (negative reward). Over time, the system should learn how to make more profitable decisions. Another common example of reinforcement learning concerns a robot that moves in an environment and tries to collect items while also returning to its recharging station before its batteries are depleted. In this case, the robot needs to autonomously find a balance between exploring the space and managing its power.

Although these techniques might seem highly abstract and mathematically distanced from how we normally understand the processes of learning and growing in biological systems, supervised, unsupervised, and reinforcement learning each corresponds to a form of learning found in real life. Hence, supervised learning is about learning with a guide such as a teacher or a reference document (consider, for example, a children's book that shows pictures of animals along with their names). Unsupervised learning is about acquiring knowledge about the world through basic observations, such as how children can learn the fundamental laws of physics by playing with blocks (or later in life, by playing with fire). Reinforcement learning covers situations in which agents are rewarded (or punished) for their actions within the world, such as when a dog is fed a treat after bringing back a stick or when a young child trips on one of their blocks.

These categories do not exist in isolation. Quite permeable, they often share models and algorithms, as the research carried out in one domain can often be applied to another. One famous example of this contributed to the resurrection of interest in neural computation and machine learning in the mid-2000s, when scientists discovered a method to train multiple layers of neurons in supervised learning and reinforcement learning systems by using unsupervised learning to facilitate training the lower-level layers of the neural architecture (Hinton, Osindero, & Teh, 2006).

Components of a Machine Learning System

Machine learning systems can be further qualified by three constituents that interoperate: a *training process*, a *model*, and *data* (see figure 1.2). These items represent interdependent

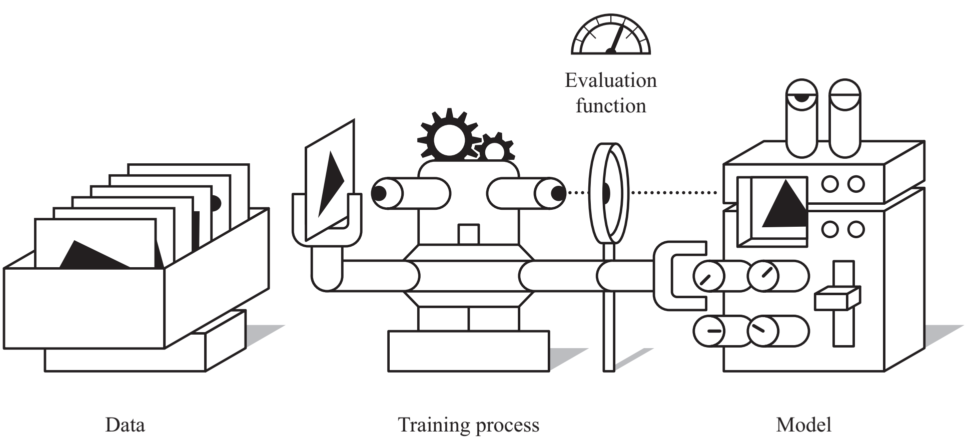

Figure 1.2
Components of a machine learning system. The *training process* trains a *model* over a set of *data* using an *evaluation function* to measure the performance of the model. Drawing by Jean-François Renaud.

dimensions of a learning system that influence its outcomes—in particular, when applied to art, its aesthetic potentialities.

Machine learning systems are trained on sets of examples that represent the empirical knowledge to which they have access. The *data* made available to the algorithm is one of the fundamental elements that influence the system's behavior and performance: the system cannot acquire knowledge beyond the data that it is fed, unless it comes a priori, encoded in the data itself or in the system. An example usually consists of a group of numerical values, each representing a dimension of the learning space. For example, a data set consisting of 10×10 grayscale images would typically be represented as a series of points, each with one hundred different values (10×10).

The knowledge the system has about the world is contained in a structure called the *model*. In the same way a scale model of a sailboat both represents the original ship while scaling it down to a more portable format by removing the irrelevant details, machine learning models can often be understood as more compact versions of the training data. To work well, a good model must be sufficiently complex to represent the important characteristics of the source data. However, it should not be too precise, as it then risks becoming to specific and unable to generalize to new examples outside of the training set.[4]

There are many different kinds of such models, each with its own strengths and weaknesses. For example, artificial neural networks contain artificial neurons connected by synaptic *weights* (i.e., numerical values that represent the strength of the connection between two neurons). Such models can represent a wide range of mathematical functions and are usually considered good at recognizing patterns—hence their popularity in computer vision and speech recognition, among other applications. Another kind of model is a binary genetic code that represents a computer program, such as used in genetic programming (GP): such a model could potentially implement any algorithm, and could therefore be characterized as a general problem solver.[5]

The model and the data set are, in essence, inert structures. A third component, the *training process*, binds them together by using the data to adjust the model. To guide its decisions, this procedure uses an *evaluation function* (also called cost, fitness, or reward function, depending on the context in which it is used) that measures the performance of the model over the data points (Alpaydin, 2004, pp. 35–36).

A machine learning system can thus be summarized as follows. Given a certain kind of task (supervised, unsupervised, or reinforcement learning), a learning algorithm adjusts a model to improve its performance (measured using an evaluation criterion) over a data set. While this is roughly true across all fields of applications of machine learning, there exist many variations within the techniques that are suitable for each of these components.

Machine learning therefore provides a generic framework for problem-solving that challenges not only traditional AI approaches but computer science in general. Rather than trying to solve a problem by designing a computer program that directly addresses it, machine learning suggests putting together different components (data, model, training process, and evaluation function) and letting the system find the solution autonomously. Although this approach did not work very well for many decades (due mainly to insufficient data and computing resources), since the mid-2000s it has moved from the background to the forefront of AI and has become a catalyst of profound social transformations.

From Cybernetics to Deep Learning

The triumph of machine learning in the twenty-first century builds on more than five decades of research in computer science. Deep learning, an approach to machine learning that is inspired by the hierarchical and self-organizing nature of the brain, has enabled giant leaps in achieving (and sometimes overcoming) human-level performance on challenging problems such as computer vision and speech recognition, propelling the field of artificial intelligence into a new industrial era that promises to disrupt the very fabric of society. Yet, deep learning is merely the latest milestone in artificial intelligence, and more broadly, in humanity's long, relentless journey of inquiry into the workings of living systems and human intelligence.

One can trace the first concepts that led to the emergence of machine learning in the 1980s and then to its re-emergence in the 2000s, to the interdisciplinary science of *cybernetics*. Born during the post-war period in the United States and England, cybernetics sought to understand the workings of the brain and, in this endeavor, to comprehend fundamental mechanisms governing both organic and computational systems. Cyberneticians designed adaptive and autonomous machines (Walter, 1950; Ashby, 1954) and laid the foundations of new theories of control and communication (Wiener, 1961; Shannon, 1948).

Some cyberneticians tried to address the workings of human cognition by looking at the brain's most basic units: neurons. Walter Pitts and Warren S. McCulloch showed that simple interconnected neurons could be used to model logical gates (McCulloch & Pitts, 1943). Grey Walter and Ross Ashby created artificial devices that tried to simulate neurological mechanisms through feedback loops between interconnected units. At the end of the 1950s, following the work of neuroscientist Donald Hebb, psychologist Frank Rosenblatt proposed a neural-inspired system that could recognize patterns: the perceptron (Rosenblatt, 1957). Around the same time, Oliver Selfridge proposed a structure made up of sets of layered units of neurons for image recognition, which he called the pandemonium (Selfridge, 1959). Plagued by important practical and theoretical issues, the perceptron and the pandemonium nonetheless embodied core ideas forming the basis of contemporary deep learning systems: that the key to artificial intelligence lies in the design of autonomous systems made of stacked layers of self-oganizing units (i.e., neurons), in which each layer learns an increasingly higher-level degree of representation of what the system observes.

Although cybernetics as a scientific field has more or less vanished, its significance in the development of computer technologies cannot be ignored, particularly in regard to the development of artificial intelligence and machine learning. What is less known is its influence on the society and culture of the 1960s, particularly in the field of contemporary art, in which cybernetics was closely related to movements such as conceptual art, performance art, and kinetic art. These revolutionary approaches attempted to move beyond the materiality of the art object, toward artworks conceived as computer programs and artificial systems (Burnham, 1968).

This paradigm shift from objects to systems also resonates with an important theory of mind that emerged in the postwar era, which would profoundly impact the conception of the human subject in the Western world in subsequent decades. This current of thought, which we refer to as *computationalism* (also known as *cognitivism*), is a particular form of representationalism, a theory of mind that rests on the notion that we do not experience the world directly but rather through a representation of it. Computationalism posits that human cognitive capabilities are equivalent to computation—in other words, that intelligence is realized by applying operations over sets of symbols that represent the world. Computationalist theory claims that the brain is simply the *hardware* substrate that runs the *software* of such mental capacities. Consequently, it holds that a computer program able to reproduce human cognitive performances should be deemed intelligent even if it runs on a silicon-based machine (Turing, 1950).[6] For computationalists, cognition is purely functional, hence immaterial. It is not defined by, or limited to, human brains, let alone subjective experience. Therefore, they argue, it is theoretically possible to design cognitive processes on computers.

Computationalism was concomitant with the appearance and development of artificial intelligence in the 1950s, in parallel to cybernetics. Artificial intelligence as a field set as its core goal the study of how computers could simulate human intelligence. In those years, two notable approaches were already present, and they would be collaborating and competing throughout the history of AI. The first approach, *symbolic AI*,[7] attempts to endow computers with intelligence by directly programming them to be smart. The second approach, *machine learning*, argues that rather than trying to explicitly implement intelligence in machines, we should instead seek to teach them how to learn by themselves. The same way that rule-based AI is tied to computationalism, machine learning seems intimately related to connectionism, a theory of mind that posits that cognition happens through multiple parallel interactions between interconnected units such as neurons.[8]

In the first stage of AI history, symbolic AI rapidly gained impetus as computer programs were shown to perform incredibly well on tasks deemed difficult for humans, such as playing strategy games, whereas at the same time, connectionist learning systems such as the perceptron were shown to have severe theoretic limitations (Minsky & Papert, 1969). Scientists such as Marvin Minsky and Seymour Papert argued in favor of rule-based computing and heuristics, using the powerful calculation features of computers to solve problems with brute force. By the end of the 1970s, however, symbolic AI research plateaued and public funding came to a halt. AI entered a period of disfavor, later called the first *AI Winter*. Although research did not stop completely, interest in cybernetics and systems decreased in the artistic world as the Western world entered the 1980s, in favor of explorations of other computer capabilities such as graphics and sound production.

In the mid-1980s, interest in AI surged once again with the revival of neural network research, tied mainly to new discoveries in training more complex forms of neural nets directly inspired by perceptrons. In parallel, some rule-based approaches to AI regained popularity through the development of expert systems, which aim to transfer the know-how of human experts into a set of logic rules.

In addition to these two concurrent approaches to artificial intelligence, a crucial new field directly related to cybernetics appeared at the end of the 1980s: artificial life (ALife). Inspired by cybernetics, complexity theory, chaos theory, and artificial intelligence, ALife seeks to study living systems in their mode of operation, specifically by simulating living processes with the computer. ALife rests on a bottom-up approach that makes use of the raw power of the computer to simulate complex interactions between numerous units and then observes the result. ALife researchers explore how applying simple algorithmic instructions to low-level units can generate complex patterns at higher levels that sometimes look like living processes and organisms.

At the end of the 1980s, both expert systems and neural networks had shown severe limitations in real-life settings, and interest in AI faded into a second AI winter. In response to these challenges, combining the successes of both ALife and machine learning, and downplaying the importance of representation in AI systems, Rodney Brooks suggested an alternative to AI called New AI. In opposition to both connectionism and symbolic AI, Brooks argued that intelligence did not need a representation of the world and that cognition could not be detached from happening in a situated/embodied manner: in other words, that the body was using the world as its own model. Therefore, he argued that the only way that AI could make real progress, was to design robots that could interact in their world and to gradually build them and teach them how to act within that environment.

In the 1990s, both ALife and New AI became important sources of inspiration for new media artists. One notable example is the influence of ALife and complexity theory on video games; a number of important simulation games involving complex phenomena were given wide exposure through popular titles such as *SimCity* and *Civilization*. A particularly apropos example is Will Wright's 1990 game *SimEarth: The Living Planet*, which allows the player to supervise the development of a planet through indirect means such as varying its volcanic activity, erosion, rainfall, and albedo. The main scenario follows the different eras of the history of an earth-like planet, from the formation of the crust to the appearance of the first oceans, and then to the emergence of life and civilizations.

In contemporary art, ALife art became an art form in itself, for example, through the work of Nell Tenhaaf, Susie Ramsay, and Rafael Lozano-Hemmer, who in 1999 created the Art and Artificial Life International Awards (VIDA) with the support of Fundación Telefónica. The prize, which ran for sixteen years, sought to support artistic inquiries in the field of artificial life. Robotics artists Louis-Philippe Demers, Ken Rinaldo, and Bill Vorn, who were all directly inspired by Rodney Brooks's New AI, are among the winners of this award. Also influenced by Brooks, artist and media theorist Simon Penny came up with the term *aesthetics of behavior* to describe the kind of work made by the creation of an artificial agent that interacts with the real world (Penny, 2000, 2017).

From the mid-1990s to the mid-2000s, the development of the internet produced massive amounts of data, while at the same time computing power increased, in particular through the development of graphical processing units (GPUs), developed in response to the growing entertainment industry (video games and special effects), which specialize in matrix multiplication—exactly the kind of expansive mathematical operations required to compute extremely large artificial neural networks. These circumstances, together with support for fundamental research in Canada through the Canadian Institute for Advanced Research (CIFAR), created the context for a renaissance of neural network-based machine learning—sometimes referred to as the Canadian AI conspiracy. Although connectionist AI had been widely abandoned in the early 2000s, on the algorithmic side critical breakthroughs in the mid-2000s enabled the effective training of neural networks with many hierarchical layers, making it possible for such systems to reproduce and even exceed human performance on tasks deemed extremely difficult (such as computer vision, speech recognition, and text translation) in a truly autonomous fashion, by analyzing raw information without the need to rely on a priori knowledge or heuristics designed by humans.

After more than fifty years of research, artificial intelligence technology was finally ripe for the taking. Giant corporations such as Google and Facebook began aggressively hiring top researchers of the field, often literally buying their labs, granting them not only outstanding salaries and funds but perhaps more importantly, access to huge sets of data on which to carry out their research, in the hope of leading the development of their current and future products.

Within the scope of only a few years, corporate investment in artificial intelligence has skyrocketed and seems to grow exponentially, even prompting fear of an economic bubble. But there are strong signs that this is more than just hype. AI-driven technologies that looked like science fiction a few years ago are now on the market, including self-driving cars, speech-to-speech translation, and personal assistants such as Alexa and Siri. These, however, might be just the tip of the iceberg. We are experiencing a technological revolution at least as important as (and probably much more important than) the one following the appearance of the internet; the current revolution is already having a profound impact on society, similarly to previous large-scale industrial transformations. In much the same way that the industrial revolution of the eighteenth and nineteenth centuries brought societies into the first machine age, in which machines assisted humans in carrying out physical tasks, artificial intelligence is the driving force of a second machine age, in which smart algorithms are replacing cognitive tasks (Brynjolfsson & McAfee, 2014).

Machine learning represents an immense potential for humanity that goes far beyond marketable applications such as self-driving cars and personalized advertisement. Moreover, these technologies present important socio-political and ethical issues that threaten democracy itself, such as the proliferation of fake news through AI-supported social media bubbles. Eminent deep learning experts Yoshua Bengio and Geoffrey Hinton have emphasized that the technology, which was developed largely through publicly funded fundamental research over decades, should not merely profit the private sector but should expand to public services such as health care and education, as well as other areas.

Art is one alternative territory of exploration for machine learning's potential. New media scholars Joline Blais and Jon Ippolito suggest that digital art acts as antibodies against technological invasions of the cultural and social body. "Science," they claim, "has always offered us a future, and sometimes even a promise to repair the dangers it has unleashed on us in previous generations. But in an age when technology seems increasingly to have a mind of its own, art offers an important check on technology's relentless proliferation." (Blais & Ippolito, 2006, p. 9)[9]

A Shift in Paradigm

The coming of age of machine learning has triggered a mix of fear and excitement in the media as well as in academia, which has turned contemporary discourse about AI technologies into a highly polarized debate. One camp warns against the dire impacts of AI on the labor market, such as robots and algorithms rapidly replacing humans in fields such as transportation, logistics, and office support;[10] and the emergence of a much dreaded technological *singularity* following which AI will supplant humans as the superior intelligent species, with possibly dire consequences that could lead to the extinction of the human race (Kurzweil, 2006). On the other side of the debate, techno-optimist choirs chant the libertarian utopia of a postwork, postdemocratic world in which humanity's problems will be smoothly resolved by benevolent artificial learning systems; more moderate voices point to the concrete benefits of machine learning in health care and education and believe that the advantages outweigh the drawbacks.

In the 1950s and 1960s, the cyberneticians dreamed of a society regulated by smart, self-regulating, adaptive systems similar to those found in the human brain. This loosely organized group of interdisciplinary researchers suggested a complete change in paradigm about the way we consider technology and how it operates in the world. They criticized technologies of the past for their lack of adaptivity and autonomy driven by a human-centric worldview that sought to control nature (Pickering, 2010), and suggested an alternative vision of technological development that broke with these outdated principles.

What if technologies were designed to adapt themselves to natural processes and entities, rather than the other way around? Can we envision technologies that are not meant to control nature but rather to take part in an ecosystem, trying to survive while allowing other processes to flow? Can we give artificial agencies the right to make mistakes? Can we allow them to be gracefully weak, imprecise, and hesitant, just as we are? In the field of AI, what would happen if we moved beyond the ideal of optimization and control, toward the most open-ended paradigm of adaptation as a living process?

Although its story is deeply rooted in cybernetics (Goodfellow, Bengio, & Courville, 2016), current-day machine learning has not embraced the cyberneticians's utopian dream of self-regulating technologies, relying instead on a relatively traditional engineering culture that attempts to efficiently solve concrete, measurable *problems* such as recognizing patterns or predicting future quantifiable events; in other words, attempting to gain control over nature.

The potential repercussions of the industrial development of machine learning, for good and for bad, are immense. On the bright side, consider, for example, how automated translation technologies facilitate access to information beyond linguistic frontiers; how self-driving intelligent cars can potentially reduce traffic and accidents; and how image-based pattern recognition can improve the quality of medical diagnosis and help reduce suffering. Yet, as many observers have pointed out, the increasing presence of machine learning since the mid-2000s through rapid industrial deployment by major multinationals is problematic in many ways. As a source of important debates, many believe these technologies might in fact lead to increased inequalities and power imbalances, and fragilize democracies. As examples, think about the jobs lost due to automation in the transportation industry, the deep ethical implications of autonomous weaponry, the AI-aided fragmentation of society through the reinforcement of media bubbles and the dissemination of fake news, and the potentially nefarious implications of using learning technologies for crime prediction and human profiling.

These discussions are critically needed and require the attention and participation of all sectors of society. As machine learning is likely to become one of the most important industrial technologies of the twenty-first century, how can artists engage in the material and intellectual debates that it brings forward? How can they work creatively and independently with a technology that has been aggressively privatized and is increasingly reliant on an industrial complex based on social media and advertising? Consider for example how Google's DeepDream project,[11] despite its attempt to make it open to the public as a creative tool, is inseparable from Google's access to massive data, computing power, and scientific expertise.

With their capacity to work both critically and creatively with material and experiential questions, artists have a unique standpoint for reflecting on the complex issues that surround machine learning. Art can suggest alternative ways of engaging with machine learning systems and imagining our relationship with them now and in the future. But how can artists work with technologies that seem so contingent on access to big databases, big computers, and big expertise? How can they approach algorithms that are largely meant for problem-solving and optimizing—both of which that have little to do with the arts? In other words, how can they relate to a field that has everything to do with engineering, science, and business and seems utterly disconnected from contemporary forms of artistic expression?

As a way to approach these questions, consider the existence of a rich historical tradition in new media art of creators working with adaptive and self-organizing technologies such as machine learning. Since the ascent of modern computers after World War II, artists and other creative practitioners have been exploring self-organizing systems, artificial intelligence, and adaptive computation as base materials for the creation of aesthetic experiences. As early as the 1950s, artists were creating adaptive robots and generative works using cybernetic systems. Important movements such as Jack Burnham's systems aesthetics, Roy Ascott's cybernetics art, and robotic art and artificial life art marked the development of new media art from the postwar era onward. This tradition goes hand in hand with contemporary discourses surrounding the nature of life and cognition, such as autonomy, chaos, emergence, and the generation of novelty—what artificial life researcher Takashi Ikegami calls *living technology* (Ikegami, 2013).

When we compare machine learning art to documented approaches in new media art that make sure of computational systems such as artifical life art (Langton, 1995; Tenhaaf, 2008; Penny, 2009) and situated robotic art (Brooks, 1999; Penny, 2013), one of the most important differences to keep in mind is that these approaches are bottom-up in nature, relying on the iterative building of emergence and self-organization by human-based trial and error. The artist engaged in these practices uses computation to simulate artificial life forms, looks at the result, tentatively changes a few things, and tries again until satisfied. In other words, they act as an adaptive device themselves, making choices among indeterminate processes.

Machine learning suggests a different way to deal with self-organization, in which one assembles different ingredients (data, model, and training process) but lets the emergent system find its own way to achieve its goals, hence handing more control to the machine. This results in a different relationship with the machine that is closer to experimental science or to a form of a collaboration between the artist and the machine. It allows finer control over outcomes than with a purely emergent procedure. It also gives more options because the artist can still directly control the goal of the system in real time (as in ALife simulations) or intervene more indirectly by tweaking data, model, and/or evaluation function.

We are moving into a world with computer technologies that are increasingly adaptive, whereas similar kinds of systems were previously found only in natural phenomena. These pervasive systems are newly meaningful for artists and cultural theorists because they suggest new approaches to working with self-organizing systems and open up novel ways to understand what it means to be alive and human. Machine learning also challenges the notion that artistic creation is a purely human-centric practice, as the creative agency becomes diffused between humans and machines that couple with one another. Finally, their rapid development in the age of big data and massive concentrations of wealth and power makes them a critical engagement space for the arts.

There are important challenges for making art with machine learning. First, machine learning usually requires enormous amounts of data, which are difficult to generate or even access, because most of the largest databases are privately owned. Second, computer power is still relatively expensive, although costs have been steadily dropping. Third and perhaps more importantly, artists often lack the technical skills to work with these technologies in a meaningful way.

Finally, in most cases, machine learning is an optimization process for problem-solving that attempts to maximize or minimize an evaluation function over time. For example, in supervised learning classification applications such as detecting cats and dogs in images, we usually try to minimize the number and range of errors made by the system. But art is specifically *not* about optimization because there is no objective evaluation function to minimize. There is no such thing as the best painting, just as there is no such thing as the best joke. Preferences are subjective and not mutually exclusive—for example, it is not unusual for someone to have many favorite movies.

Fortunately, these difficulties are not insurmountable. Many of these issues can in fact be circumvented, as artists have needs and goals that differ from those of scientists and engineers. It could be appropriate to use smaller databases and less computer-intensive learning systems for a large variety of artistic applications. Furthermore, most of the technology is developed under a very open culture. Even when carried out within the walls of tech giants, research is for the most part made publicly available, and many big IT companies play an active role in making tools available to the public under open source licenses. Computational power is likely to increase fast in coming years, and the new generation of creative open hardware has multiple cores and GPUs.

On the positive side, machine learning technologies are becoming increasingly easy to use. For example, neural-based machine learning used to be a sort of dark craft because it required extensive tuning and massaging of the data. One of the advantages of deep learning is that algorithms are now able to work with raw data, which is a huge gain for users and makes working with such systems much easier. There are reams of free software tools and online tutorials to learn about these techniques, and the democratization of these technologies through education is likely to become central in the development of societies in the coming decades.

Chapter Breakdown

This book aims to provide conceptual tools, accounts of practice, and historical perspectives to understand and address machine learning technologies from an artistic standpoint. The text is organized in three parts generally following the different components that characterize machine learning systems: training process and evaluation functions (part 1); models and machines (part 2); and data (part 3). By dissecting the scientific description of learning algorithms and connecting their properties with artistic questions, I aim to establish a comprehensive framework that artists, musicians, composers, writers, curators, and media theorists can use to approach machine learning in works of art and within larger cultural questions. Throughout the book, I thus bring together an overview of the scientific theories, concepts, and definitions attached to the various components of learning machines, expressed in an accessible way; an examination, supported by examples, of the opportunities for artistic exploration and exploitation of machine learning, either through the application of off-the-shelf techniques in their intended use or by practices of hacking and hijacking; and the main limitations, challenges, and constraints of these components of machine learning algorithms, in the context of artistic creation.

This book attempts to tackle head-on aesthetic and practical issues within the intricate landscape of machine learning art, maneuvering between questions of art and science, human and machine, and bodies and processes. Following posthumanist scholar Rosi Braidotti's concept of *zigzagging* (Braidotti, 2013, p. 164), it embraces a nonlinear way of hobbling through this murky territory, using the materiality of machine learning systems themselves as a guide. As it builds on my own research-creation work as a transdisciplinary artist-researcher, I will also, at times, share examples and perspectives from my own practice and experience.

The first part of the book delves into questions surrounding the training process. Chapter 2 positions the learning loop as an optimization process, therefore seemingly antithetical to the arts, which are specifically nonpurposeful and nonoptimizable. Art practices differ from those of scientists and engineers in being more process driven than goal driven. I thus argue that research in the field of computational creativity and creative AI that attempts to reproduce human-level creativity in a given artistic domain can be misleading, because such research often misunderstands the fundamental principles and values of contemporary art. In chapter 3, we look specifically at alternative approaches employed by artists to hijack the training process by playing with evaluation functions. Recalling the origins of machine learning in cybernetics, where an agent adapts to its environment, in chapter 4, I propose a framework to understand the aesthetic properties of adaptive behaviors.

The second part examines what constitutes the true outputs of machine learning systems: models. In chapter 5, we consider how these self-organized *black boxes* act in ways that often defy human understanding and why these qualities provide a fertile ground for new types of practice and art forms. We then examine how different species of machine learning systems afford different kinds of artistic practices and aesthetic qualities: chapter 6 deals with parametric models and genetic algorithms, chapter 7 with shallow connectionist learning, and chapter 8 with deep learning.

The third and last part of the book focuses on the role of data in machine learning art. Chapter 9 shows how artists use data as a raw material to shape machine learning systems and how it impacts the creative process. Chapter 10 makes the argument that machine learning allows novel forms of algorithmic remixes through the collection of data and the reuse of pretrained models. Finally, chapter 11 examines the correspondence between observation and generation in machine learning systems and how biases operate in these contexts. I conclude the book in chapter 12 by zooming out of the materiality of machine learning art, addressing broader issues related to the relationships artists establish with machine learning systems, the impact of machine learning on the art world and curatorial practices, and the sociopolitical implications of machine learning art in the twenty-first century.

By organizing the chapters around different perspectives over machine learning and its connections with new media art, I hope that this book can provide an understanding of the fundamental design of machine learning algorithmic structures to the layperson, while at the same time framing such technologies within larger historical and conceptual spaces, and hence becomes a reference from which to draw knowledge and inspiration long after it has been read.