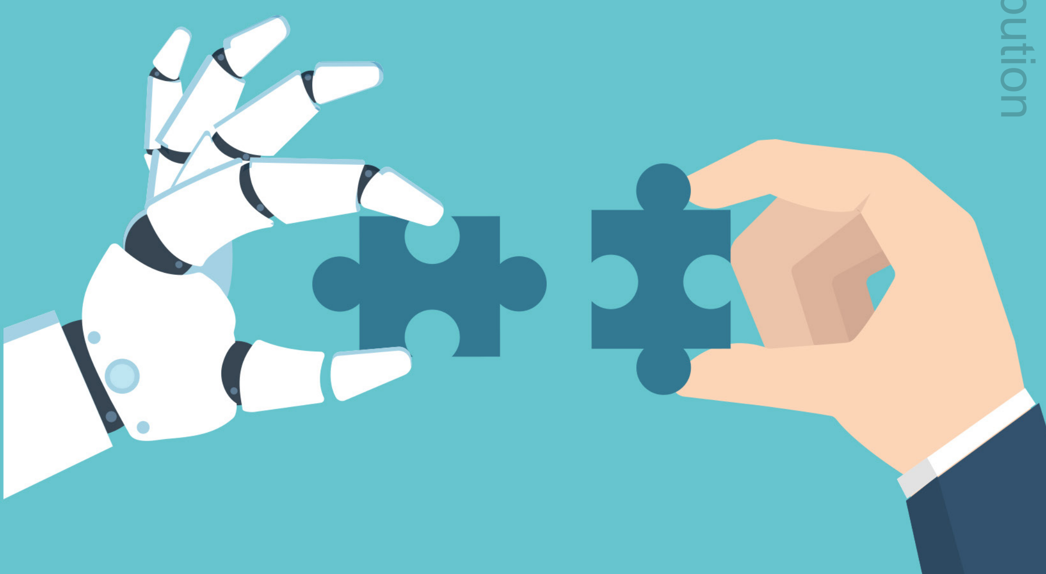
MARIYA YAO, MARLENE JIA, ADELYN ZHOU

Applied Artificial Intelligence

A HANDBOOK FOR BUSINESS LEADERS



"The authors of *Applied AI* have incredible depth of expertise and experience in AI, and they make the complex topic accessible to everyone. It's rare to find technology experts as engaging and thought-provoking."

STEPHEN STRAUSS

Head of Sales Enablement and Insights, PayPal

"This book cuts the fluff and arms business leaders with exactly the right foundational knowledge to lead successful AI initiatives at their companies. It's hands down the best playbook for executives starting on their automation journey."

JACK CHUA

Director of Data Science, Expedia

"As a deep learning researcher and educator, I'm alarmed by how much misinformation and misreporting occurs with AI. It's refreshing to see a practical guide written by experienced technologists which explains AI so well for a business audience. In particular, I'm glad to see this book addresses critical issues of AI safety and ethics and advocates for diversity and inclusion in the industry."

RACHEL THOMAS

Co-Founder, Fast.ai and Assistant Professor, USF Data Institute

“Full of valuable information and incredibly readable. This book is the perfect mix of practical and technical. If you’re an entrepreneur or business leader, you need this guide.”

JEFF PULVER

Co-Founder, Vonage and MoNage

“*Applied AI* is the perfect primer for anyone looking to understand the enterprise implications of emerging artificial intelligence technology—a must read for any business leader intending to stay ahead.”

ALEX STEIN

Sr. Director, Strategy & Business Development, Viacom International Media Networks

“Recent progress in AI will dramatically impact all aspects of a business. This excellent book provides a practical examination of how to harness disruptive technologies to achieve scalable and sustainable business success.”

STEVEN KUYAN

Managing Director, NYU Tandon Future Labs

Applied Artificial Intelligence

A HANDBOOK FOR BUSINESS LEADERS

For permission requests, contact the publisher "Attention: Permissions Coordinator," at the website address below.

Printed in the United States of America

First Printing December 2017

ISBN 978-0-9982890-2-1 (Paperback)
ISBN 978-0-9982890-5-2 (Kindle)

Edited by Natalia Zhang
Cover Illustration by Vanessa Maynard
Interior Design by Vanessa Maynard

Website: www.appliedaibook.com
Email: authors@appliedaibook.com

For all of you who build technology
to make tomorrow better than today.

TABLE OF CONTENTS

APPLIED ARTIFICIAL INTELLIGENCE

WHO THIS BOOK IS FOR

Applied Artificial Intelligence is a practical guide for business leaders who are passionate about leveraging machine intelligence to enhance the productivity of their organizations and the quality of life in their communities. If you love to drive innovation by combining data, technology, design, and people, and to solve real problems at an enterprise scale, this is your playbook.

There are plenty of technical tomes on the market for engineers and researchers who want to master the nitty-gritty details of modern algorithms and toolsets. You can also find plenty of general interest content for the public about the impact of AI on our society and the future of work.

This book is a balance between the two. We won't overload you with details on how to debug your code, but we also won't bore you with endless generalizations that don't help you make concrete business decisions. Instead, we teach you how to lead successful AI initiatives by prioritizing the right opportunities, building a diverse team of experts, conducting strategic experiments, and consciously designing your solutions to benefit both your organization and society as a whole.

How to Use This Book

The first part of this book, "What Business Leaders Need to Know," gives executives an essential education in the state of artificial intelligence today. We recommend reading this part in full before pursuing AI projects for your organization.

Chapters 1 and 2 provide a non-technical introduction to AI, the techniques used to power modern AI systems, and the functional differences between different levels of machine intelligence. While you do not need to memorize every detail, a passing familiarity with technical definitions will help you separate hype from reality when evaluating a project proposal for your own organization.

Chapters 3, 4, and 5 describe promising applications of AI in society as well as challenges that arise from biased or unethical algorithms. You'll learn how collaborative design is essential to ensuring that we build benevolent AI systems.

In the second part of our book, "How to Develop an Enterprise AI Strategy," we walk you through the strategic and methodological steps required to implement successful AI projects for your company. These chapters act as a reference guide as you are building your initiatives.

Read through them once to familiarize yourself with the content, and then refer back to specific sections as needed during your projects.

Chapters 6 and 7 teach you how to prepare your organization to succeed in AI projects. You will learn strategies to manage important stakeholders and attract technical talent.

In Chapter 8, we guide you through exercises that will help you to identify opportunities for AI adoption within your organization and develop a business plan for implementation and deployment. Chapters 9, 10, and 11 explain common technical challenges you will encounter in building AI and how to overcome them.

The last section of our book, "AI For Enterprise Functions," highlights popular AI applications for common business functions. Chapter 12 summarizes some of the challenges of adopting AI solutions for enterprises. Chapters 13 and 14 introduce common AI applications in essential administrative functions like finance, legal, and HR, while Chapters 15 and 16 describe how machine learning can dramatically improve business intelligence, analytics, and software development. Chapters 17, 18, and 19 focus on the revenue-generating functions of sales, marketing, and customer service.

Finally, Chapter 20 emphasizes the ethical responsibility that you, as business and technology leaders, have towards your workforce as well as towards ensuring that any technologies that you build have a benevolent impact on your customers, employees, and society as a whole.

Because AI technologies evolve very quickly, we created an educational website, **appliedaibook.com**, where we offer updated content and detailed case studies for specific industries. Supplemental content for this book can be found in our resources section at **appliedaibook.com/resources**.

We also created social communities and discussion forums for our readers to connect with us and each other, which you can join by visiting **appliedaibook.com/community**.

What Business Leaders Need to Know About Artificial Intelligence

1. BASIC TERMINOLOGY IN ARTIFICIAL INTELLIGENCE

Think about the most intelligent person you know. What about this person leads you to describe him or her this way?

Is she a quick thinker, able to internalize and apply new knowledge immediately? Is he highly creative, able to endlessly generate novel ideas that you'd never think of? Perhaps she's highly perceptive and hones in on the tiniest details of the world around her. Or maybe he's deeply empathetic and understands how you're feeling even before you do.

Human intelligence spans a wide spectrum of modalities, exhibiting abilities such as logical, spatial, and emotional cognition. Whether we are math geniuses or charismatic salesmen, we must utilize cognitive abilities like working memory, sustained attention, category formation, and pattern recognition to understand and succeed in the world every day.

Though computers trounce humans at large-scale computational tasks, their expertise is narrow, and machine capability lags behind human intelligence in other areas. The rest of this chapter will help you to understand the state of artificial intelligence today.

AI vs. AGI

Artificial intelligence, also known as AI, has been misused in pop culture to describe almost any kind of computerized analysis or automation. To avoid confusion, technical experts in the field of AI prefer to use the term Artificial General Intelligence (AGI) to refer to machines with human-level or higher intelligence, capable of abstracting concepts from limited experience and transferring knowledge between domains. AGI is also called “Strong AI” to differentiate from “Weak AI” or “Narrow AI," which refers to systems designed for one specific task and whose capabilities are not easily transferable to other systems. We go into more detail about the distinction between AI and AGI in our Machine Intelligence Continuum in Chapter 2.

Though Deep Blue, which beat the world champion in chess in 1997, and AlphaGo, which did the same for the game of Go in 2016, have achieved impressive results, all of the AI systems we have today are “Weak AI." Narrowly intelligent programs can defeat humans in specific tasks, but they can’t apply that expertise to other tasks, such as driving cars or creating art. Solving tasks outside of the program’s original parameters requires building additional programs that are similarly narrow.

"We're very far from having machines that can learn the most basic things about the world in the way humans and animals can," said Yann LeCun, head of AI at Facebook, in an interview with *The Verge*.[1] "In particular areas machines have superhuman performance, but in terms of general intelligence we're not even close to a rat." In addition, the path towards AGI is also unclear. Approaches that work well for solving narrow problems do not generalize well to tasks such as abstract reasoning, concept formulation, and strategic planning—capabilities that even human toddlers possess but our computers do not.

Modern AI Techniques

We are often asked to explain the key differences between machine learning, data science, AI, deep learning, etc. All of these are examples of machine intelligence, but they vary in their usage and potential impact. While engineers and researchers must master the subtle differences between various technical approaches, business and product leaders should focus on the ultimate goal and real-world results of machine learning models. This section is a guide to today's most popular techniques, but methodologies are constantly evolving. You don't need to memorize the guide, but you should try to gain a passing familiarity with the basic characteristics of each technique.

In general, most enterprise-scale technologies use a wide range of automation methodologies, but not all of them count as AI. Differentiating between methods that are AI and those that are not can be tricky, and there is often overlap. You will find that simpler approaches often outperform complex ones in the wild, even if they're intellectually less "advanced."

Though AI refers to a larger umbrella of computational techniques, the most successful modern AI solutions are powered by machine learning algorithms. For simplicity, we use AI and machine learning as interchangeable terms in this book.

STATISTICS AND DATA MINING

Statistics is the discipline concerned with the collection, analysis, description, visualization, and drawing of inferences from data. Its focus is on describing the properties of a dataset and the relationships that exist between data points. Statistics is generally not considered part of AI, but many statistical techniques form the foundation for more advanced machine learning techniques or are used in conjunction with them.

Descriptive statistics describes or visualizes the basic features of the data being studied. A simple application could be to find the best-selling retail item in a store in a specific period of time.

Inferential statistics is used to draw conclusions that apply to more than just the data being studied. This is necessary when analysis must be conducted on a smaller, representative dataset when the true population is too large or difficult to study. Because the analysis is done on a subset of the total data, the conclusions that can be reached with inferential statistics are never 100 percent accurate and are instead only probabilistic bets. Election polling, for example, relies on surveying a small percentage of citizens to gauge the sentiments of the entire population. As we saw during the 2016 US election cycle, conclusions drawn from samples may not reflect reality![2]

Data mining is the automation of exploratory statistical analysis on large-scale databases, though the term is often used to describe any kind of algorithmic data analysis and information processing, which may also include machine learning and deep learning techniques. The goal of data mining is to extract patterns and knowledge from large-scale datasets so that they can be reshaped into a more understandable structure for later analysis.

SYMBOLIC AND EXPERT SYSTEMS

Symbolic systems are programs that use human-understandable symbols to represent problems and reasoning.[3] The most successful form of symbolic systems is the expert system, which mimic the decision-making process of human experts. Expert systems are comprised of a series of production rules, similar to if-then statements, that govern how the program accesses a knowledge base and makes inferences.

Rule-based expert systems are most effective when applied to automated calculations and logical processes where rules and outcomes are relatively clear. As decision-making becomes more complex or nuanced, explicitly formalizing the full range of requisite knowledge and inference schemes required to make human-level decisions becomes impossible.

The rules engine and knowledge base for any expert system must be hand-engineered by domain experts. This is a huge drawback due to the limited number of experts who can perform the task and the time needed to program such a complicated system. The "completeness" of the knowledge base is questionable and will require continued maintenance (another huge drawback that requires enormous expenditures), and the accuracy of the system is overly-dependent on expert opinions that could be wrong. While symbolic systems are historically not scalable or adaptable, recent research has investigated combining them with newer methods like machine learning and deep learning to improve performance.

MACHINE LEARNING

What happens if you want to teach a computer to do a task, but you're not entirely sure how to do it yourself? What if the problem is so complex that it's impossible for you to encode all of the rules and knowledge upfront?

Machine learning enables computers to learn without being explicitly programmed. It is a field in computer science that builds on top of computational statistics and data mining. This book will focus primarily on discussing how machine learning is being applied in different industries across different functions, so you'll want to understand the broad categories in this field and how they are applied to business problems.

Supervised learning occurs when the computer is given labeled training data, which consists of paired inputs and outputs (e.g. an image of a cat correctly labeled as "cat"), and learns general rules that can map new inputs to the correct output. Supervised learning is commonly used for classification, where inputs are divided into discrete and unordered output categories, and for regression, where inputs are used to predict or estimate outputs that are numeric values. If you are trying to predict whether an image is of a cat or a dog, this is a classification problem with discrete classes. If you are trying to predict the numerical price of a stock or some other asset, this can be framed as a regression problem with continuous outputs.

Unsupervised learning occurs when computers are given unstructured rather than labeled data, i.e. no input-output pairs, and asked to discover inherent structures and patterns that lie within the data. One common application of unsupervised learning is clustering, where input data is divided into different groups based on a measure of "similarity." For example, you may want to cluster your LinkedIn or Facebook friends into social groups based on how connected they are to each other. Unlike supervised learning, the groups are not known in advance, and different measures of similarity will produce different results.

Semi-supervised learning lies between supervised and unsupervised learning. Many real-world datasets have noisy, incorrect labels or are missing labels entirely, meaning that inputs and outputs are paired incorrectly with each other or are not paired at all. Active learning, a special case of semi-supervised learning, occurs when an algorithm actively queries a user to discover the right output or label for a new input. Active learning is used to optimize recommendation systems, like the ones used to recommend movies on Netflix or products on Amazon.

Reinforcement learning is learning by trial-and-error, in which a computer program is instructed to achieve a stated goal in a dynamic environment. The program learns by repeatedly taking actions, measuring the feedback from those actions, and iteratively improving its behavioral policy. Reinforcement learning can be successfully applied to game-playing, robotic control, and other well-defined and contained problems. It is less effective with complex, ambiguous problems where rewards and environments are not well understood or quantified.

Chapter 10 discusses the mechanics of building machine learning models in more detail. You can also find updated technical resources on our book website, **appliedaibook.com**.

DEEP LEARNING

Deep learning is a subfield of machine learning that builds algorithms by using multi-layered artificial neural networks, which are mathematical structures loosely inspired by how biological neurons fire. Neural networks were invented in the 1950s, but recent advances in computational power and algorithm design—as well as the growth of big data—have enabled deep learning algorithms to approach human-level performance in tasks such as speech recognition and image classification. Deep learning, in combination with reinforcement learning, enabled Google DeepMind's AlphaGo to defeat human world champions of Go in 2016, a feat that many experts had considered to be computationally impossible.

Much media attention has been focused on deep learning, and an increasing number of sophisticated technology companies have successfully implemented deep learning for enterprise-scale products. Google replaced previous statistical methods for machine translation with neural networks to achieve superior performance.[4] Microsoft announced in 2017 that they had achieved human parity in conversational speech recognition.[5] Promising computer vision startups like Clarifai employ deep learning to achieve state-of-the-art results in recognizing objects in images and video for Fortune 500 brands.[6]

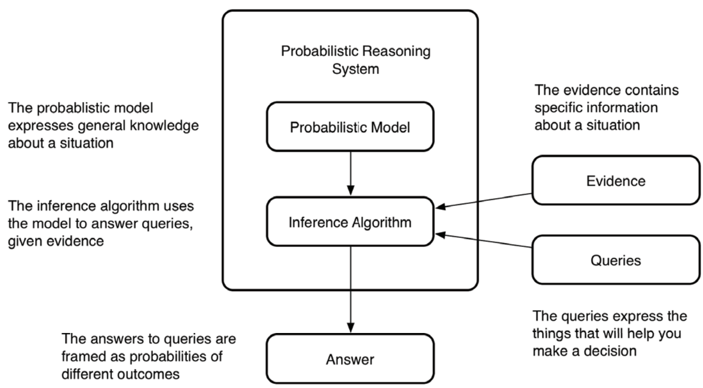
While deep learning models outperform older machine learning approaches to many problems, they are more difficult to develop because they require robust training of data sets and specialized expertise in optimization techniques. Operationalizing and productizing models for enterprise-scale usage also requires different but equally difficult-to-acquire technical expertise. In practice, using simpler AI approaches like older, non-deep-learning machine learning techniques can produce faster and better results than fancy neural nets can. Rather than building custom deep learning solutions, many enterprises opt for Machine Learning as a Service (MLaaS) solutions from Google, Amazon, IBM, Microsoft, or leading AI startups.

Deep learning also suffers from technical drawbacks. Successful models typically require a large volume of reliably-labeled data, which enterprises often lack. They also require significant and specialized computing power in the form of graphical processing units (GPUs) or GPU alternatives such as Google's tensor processing units (TPUs). After deployment, they also require constant training and updating to maintain performance.

Critics of deep learning point out that human toddlers only need to see a few examples of an object to form a mental concept, whereas deep learning algorithms need to see thousands of examples to achieve reasonable accuracy. Even then, they can still make laughable errors. Deep learning algorithms do not form abstractions or perform reasoning and planning in the same way that we humans do.

PROBABILISTIC PROGRAMMING

Probabilistic programming enables us to create learning systems that make decisions in the face of uncertainty by making inferences from prior knowledge. According to Avi Pfeffer in his book, *Practical Probabilistic Programming*, a model is first created to capture knowledge of a target domain in quantitative, probabilistic terms. Once trained, the model is then applied to specific evidence to generate an answer to a more specific query in a process called inference.



While the research and applications are in its early days, many experts see probabilistic programming as an alternative approach in areas where deep learning performs poorly, such as concept formulation using sparse or medium-sized data. Probabilistic programs have been used successfully in applications such as medical imaging, machine perception, financial predictions, and econometric and atmospheric forecasting.

Probabilistic programming is emerging as a hot area in technical research, but it has yet to be productized and operationalized for enterprise performance to the same degree that machine learning and deep learning have. We won't cover probabilistic programming in detail in this book, but you can check out the MIT Probabilistic Computing Project[7] for recommended readings and tutorials.[8]

OTHER AI APPROACHES

There are many other approaches to AI that can be used alone or in combination with machine learning and deep learning to improve performance. **Ensemble methods**, for example, combine different machine learning models or blend deep learning models with rule-based models. Most successful applications of machine learning to enterprise problems utilize ensemble approaches to produce results superior to any single model.

There are four broad categories of ensembling: bagging, boosting, stacking, and bucketing. Bagging entails training the same algorithm on different subsets of the data and includes popular algorithms like random forest. Boosting involves training a sequence of models, where each model prioritizes learning from the examples that the previous model failed on. In stacking, you pool the output of many models. In bucketing, you train multiple models for a given problem and dynamically choose the best one for each specific input.

Other techniques, such as evolutionary and genetic algorithms, are used in practice for generative design and in combination with neural networks to improve learning. Approaches like Whole Brain Uploading (WBE), also known as “mind uploading," seek to replicate human-level intelligence in machines by fully digitizing human brains. Yet other approaches seek to innovate at the hardware level by leveraging optical computing,[9] quantum computing,[10] or human-machine interfaces to accelerate or augment current methods.

The AI industry moves very quickly, and algorithms and approaches are constantly under development or being invented. To get an updated overview of modern AI technologies, download our latest guide on our book website at **appliedaibook.com/resources**.

2. THE MACHINE INTELLIGENCE CONTINUUM

If you're not an AI researcher or engineer, understanding the subtle differences and applications of various machine learning approaches can be challenging. Business problems can usually be solved in multiple ways by different algorithms, and determining the comparative merits of different methodologies can be frustrating without technical experience or practical experimentation.

To help business executives comprehend the **functional differences** between different AI approaches, we designed the Machine Intelligence Continuum (MIC) to present the different types of machine intelligence based on the complexity of their capabilities. While we've defined the continuum to contain seven levels, keep in mind that the distinction between levels is not a hard line and that many overlaps exist.

Systems That Act

The lowest level of the Machine Intelligence Continuum (MIC) contains Systems That Act, which we define as rule-based automata. These are systems that function according to some predefined script, often by following manually programmed if-then type of rules.

Examples include the fire alarm in your house and the cruise control in your car. A fire alarm contains a sensor that detects smoke levels. When smoke levels reach a predefined level, the device will play an alarm sound until it is turned off manually. Similarly, the cruise control in your car uses a powered mechanism to control the throttle position in order to maintain a constant speed.

You would never set your cruise control, take your hands off the wheel, and claim that you now have a self-driving car. Doing so would result in terrible outcomes. Yet most companies claiming to have AI are really just using Systems That Act, or rule-based mechanisms that are incapable of dynamic actions or decisions.

Systems That Predict

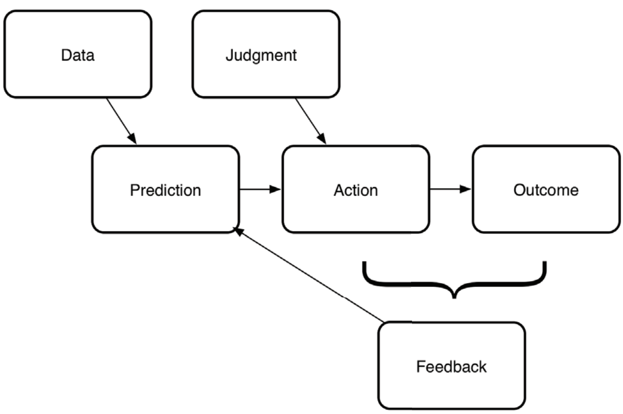
Systems That Predict are systems that are capable of analyzing data and using it to produce probabilistic predictions. Note that a "prediction" is a mapping of known information to unknown information and does not necessarily need to be a future event. Andrew Pole, a statistician for Target, explained to *The New York Times* that he was able to identify 25 products, including unscented lotion and calcium supplements, that can be used to predict the likelihood of a shopper being pregnant and even the stage of her pregnancy.[11] Target uses this information to serve eerily well-timed advertisements and coupons that encourage such customers to spend more money at the store.

Statistics power most Systems That Predict, but predictions are only as good as the data being used. If your data is flawed, or you choose sample data that does not sufficiently represent your target population, then you will get erroneous results. In business analysis, lack of data integrity and methodological mistakes are extremely common and often lead executives to the wrong conclusions.

Systems That Learn

While Systems That Learn also make predictions like statistical systems do, they require less hand-engineering and can learn to perform tasks without being explicitly programmed to do so. Machine learning and deep learning drive most of these systems, and they can function at human or better-than-human levels for many computational problems.

Learning can be automated at different levels of abstraction and for different components of a task. Completing a task requires first acquiring data that can be used to generate a prediction about the world. This prediction is combined with higher-level judgment to execute an action. The outcome from that action provides measurable feedback that can be reused at earlier decision points to improve task performance.



Many enterprise applications of statistics and machine learning focus on improving the prediction process. In sales, for example, machine learning approaches to lead scoring can perform better than rule-based or statistical methods. Once the machine has produced a prediction on the quality of a lead, the salesperson then applies human judgment to decide how to follow up.

More complex systems, such as self-driving cars and industrial robotics, handle everything from gathering the initial data to executing the action resulting from its analysis. For example, an autonomous vehicle must turn video and sensor feeds into accurate predictions of the surrounding world and adjust its driving accordingly.

Systems That Create

We humans like to think we're the only beings capable of creativity, but computers have been used for generative design and art for decades. Recent breakthroughs in neural network models have inspired a resurgence of computational creativity, with computers now capable of producing original writing, imagery, music, industrial designs, and even AI software![12]

Image from "Generating Stories from Images" by Samim Winiger, reprinted with permission

Engineer and creative storyteller Samim trained a neural network on 14 million lines of passages from romance novels and asked the model to generate original stories based on new images.[13] Flow Machines, a division of Sony, used an AI system trained on Beatles songs to generate their own hit, "Daddy's Car," which eerily resembles the musical style of the hit British rock group. They did the same with Bach music and were able to fool human evaluators, who had trouble differentiating between real Bach compositions and AI-generated imitations.

Autodesk, the leading producer of computer-aided design (CAD) software for industrial design, released Dreamcatcher, a program that generates thousands of possible design permutations based on initial constraints set by engineers. Dreamcatcher has produced bizarre yet highly effective designs that challenge traditional manufacturing assumptions and exceed what human designers can manually ideate.

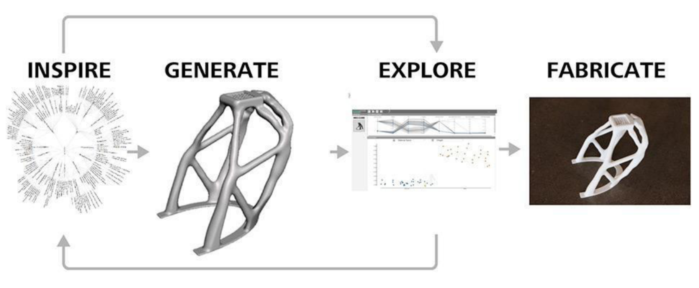


Image from Autodesk Dreamcatcher, reprinted with permission

AI is even outperforming some artists economically! In 2016, Google hosted an exhibition of AI-generated art that collectively sold for $97,605.[14]

Systems That Relate

Daniel Goleman, a psychologist and author of the book *Emotional Intelligence*, believes that our emotional intelligence quotient (EQ) is more important than our intelligence quotient (IQ) in determining our success and happiness.[15] As human employees increasingly collaborate with AI tools at work and digital assistants like Apple's Siri and Amazon Echo's Alexa permeate our personal lives, machines will also need emotional intelligence to succeed in our society.

Sentiment analysis, also known as opinion mining or emotion AI, extracts and quantifies emotional states from our text, voice, facial expressions, and body language.[16] Knowing a user's affective state enables computers to respond empathetically and dynamically, as our friends do. The applications to digital assistants are obvious, and companies like Amazon are already prioritizing emotional recognition for voice products like the Echo.[17]

Emotional awareness can also improve interpersonal business functions such as sales, marketing, and communications. Rana el Kaliouby, founder of Affectiva, a leading emotion AI company, helps advertisers improve the effectiveness of brand content by assessing and adapting to consumer reactions. Mental and behavioral health is also an area ripe for innovation. Affectiva originated from academic research at MIT that was designed to help autistic patients improve recognition of social and emotional cues.[18]

Systems That Master

A human toddler only needs to see a single tiger before developing a mental construct that can recognize other tigers. If humans needed to see thousands of tigers before learning to run away, our species would have died out from predation long ago. By contrast, a deep learning algorithm must process thousands of tiger images before it can recognize them in images and video. Even then, neural networks trained on tiger photos do not reliably recognize abstractions or representations of tigers, such as cartoons or costumes.

Because we are Systems That Master, humans have no trouble with this. A System That Masters is an intelligent agent capable of constructing abstract concepts and strategic plans from sparse data. By creating modular, conceptual representations of the world around us, we are able to transfer knowledge from one domain to another, a key feature of general intelligence.

As we discussed earlier, no modern AI system is an AGI, or artificial general intelligence. While humans are Systems That Master, current AI programs are not.

Systems That Evolve

This final category refers to systems that exhibit superhuman intelligence and capabilities, such as the ability to dynamically change their own design and architecture to adapt to changing conditions in their environment. As humans, we're limited in our intelligence by our biological brains, also known as "wetware." Instead of re-architecting our own biological infrastructure during our lifetime, we evolve through genetic mutations across generations. We cannot simply insert new RAM to augment our memory capacity or install a new processor if we wish to think faster.

While we continue to search for other intelligent life, we are not yet aware of any Systems That Evolve. Computers are currently constrained by both hardware and software availability, while humans and other biological organisms are constrained by wetware limitations. Some futurists hypothesize that we may be able to achieve superhuman intelligence by augmenting biological brains with synthesized technologies, but this research is currently more science fiction than science.

Once an upgradable intelligent agent does emerge, we will reach what many experts call the technological “singularity," when machine intelligence surpasses human intelligence.[19] Self-evolving agents will be capable of ever-faster iterations of self-improvements, leading to the eventual emergence of superintelligence.

To download a visual summary of the Machine Intelligence Continuum, visit the resources section of our book website at **appliedaibook.com/resources**.

How we build today’s Systems That Learn, Systems That Create, and Systems That Relate will affect how we build tomorrow’s Systems That Master and Systems That Evolve. While no one can predict what superintelligence will look like, we can take measures today to increase the likelihood that the intelligent systems we design are effective, ethical, and elevate human goals and values. The next few chapters tell the story of how modern AI can be used for the good of humanity, the immediate challenges that may cause AI to go awry, and the collaborative design principles we can uphold to build the best AI systems.

3. THE PROMISES OF ARTIFICIAL INTELLIGENCE

The promises of AI extend beyond the challenges of Silicon Valley and Wall Street. Emerging technologies like deep learning and conversational interfaces enable us to do far more than drive advertising clicks, streamline sales, and boost corporate profits. All around the world, entrepreneurs and executives leverage data combined with machine learning to fight social injustice and crime, address health and humanitarian crises, solve pressing community problems, and dramatically improve the quality of life for everyone.

Microfinance

When Sahil Singla joined FarmGuide, a social impact startup, he was shocked to discover that thousands of rural farmers in India commit suicide every year.[20] When harvests fail, desperate farmers are forced to borrow from microfinance loan sharks at crippling rates. Unable to pay back these predatory loans, victims kill themselves—often by grisly methods like swallowing pesticides—to escape reprisal from their debt holders.

Singla and his team are tackling this issue with deep learning. Recent growth of computational power and structured datasets has allowed deep learning algorithms to achieve better-than-human-level accuracy in a number of recognition and classification tasks. Computers can now recognize objects in images and video, transcribe speech to text, and translate languages nearly as well as humans can.

Using deep learning, FarmGuide analyzes satellite imagery to predict crop yields for individual farms. In the US, Stanford University researchers have shown machine-driven methods for crop yield analysis to be comparably accurate to physical surveys conducted by the USDA.[21] Armed with this previously unattainable information, Singla and his team can build better actuarial models for lending and insurance, thereby reducing the risk of loan sharks preying on at-risk farmers by providing them with lower and fairer interest rates for loans.

Social Justice

In Monrovia, the capital city of Liberia, fifteen-year-old Sarafina was being hounded by one of her teachers, who refused to give her a report card unless she had sex with him. Embarrassed, she kept the issue hidden from everyone, even her parents, until her father overheard a harassing phone call that the teacher made to their home. He confronted the teacher and successfully secured Sarafina's report card, but his daughter was reprimanded and forced to move to another school.[22]

Sarafina's experience is not unique. In Liberia, teachers enjoy high social status while children, especially young girls, are culturally trained not to speak out. While Sarafina's story sounds extreme to Westerners, her experience is painfully common and largely ignored in many developing countries.

Enter UNICEF's U-Report, a social reporting bot that enables young people in developing countries to report social injustice in their communities via SMS and other messaging platforms. "U-Report is not just about getting questions answered, but getting answers back out," explains Chris Fabian, Co-Lead of UNICEF's Innovation Unit. "We get responses in real-time to use the data for policy change."[23] By using a natural language interface to capture insights and performing statistical analysis on the aggregated results, the team leverages their more than 4.2 million users worldwide to identify and tackle challenging social issues like violence against children, public health policy, and climate change.

U-Report polled 13,000 users in Liberia to ask whether teachers at their schools were exchanging grades for sex. An astonishing 86 percent of reporters said yes.[24] Within a week of the U-Report on the "Sex 4 Grades" epidemic, hotlines around the country were inundated with reports of child abuse. Simply exposing a pervasive taboo inspired many more victims to speak up and reach out for help. The outpouring of responses provoked a government response and led UNICEF and Liberia's Minister of Education to collaborate on a plan to stop the abuse of authority.

In many parts of the world, citizens can't utilize the feature-rich but data-intensive mobile apps that many of us enjoy due to bandwidth limitations and limited access to phones with up-to-date features. Being limited to voice calls and SMS means that technologies like natural language processing (NLP), dialog systems, and conversational bots become critically important to delivering value.

Medical Diagnosis

AI can dramatically streamline and improve medical care and our overall health and wellbeing. The fields of pathology and radiology, both of which rely largely on trained human eyes to spot anomalies, are being revolutionized by advancements in computer vision. Pathology is especially subjective, with studies showing that two pathologists assessing the same slide of biopsied tissue will only agree about 60 percent of the time.[25] Researchers at Houston Methodist Research Institute in Texas announced an AI system for diagnosing breast cancer that utilizes computer vision techniques optimized for medical image recognition,[26] which interpreted patient records with a 99 percent accuracy rate.[27]

In radiology, 12.1 million mammograms are performed annually in the United States, but half yield false positive results, which means that one in two healthy women may be wrongly diagnosed with cancer. In these situations, the patients often undergo biopsies, an invasive procedure that removes tissue or fluid from a suspicious area for analysis. To reduce the number of unnecessary surgical interventions, researchers at MIT and Harvard Medical School have developed a diagnostic tool that uses machine learning to correctly identify 97 percent of malignant tumors. Since deployment, the technology has reduced the number of benign surgeries by 30 percent.[28]

Artificial intelligence technologies are already saving lives and transforming societies. If used wisely, AI can be used to tackle many of the world's greatest challenges. Used unwisely, however, AI can unintentionally amplify many of humanity's worst traits. We highlight the challenges that undermine benevolent AI in the next chapter.

4. THE CHALLENGES OF ARTIFICIAL INTELLIGENCE

"The future is already here—it's just not evenly distributed."
—William Gibson

When Timnit Gebru attended a prestigious AI research conference in 2016, she counted six black people in the audience out of an estimated 8,500 attendees. There was only one black woman: herself. As a PhD from Stanford University who has published a number of notable papers in the field of artificial intelligence, Gebru finds the lack of diversity in the industry to be extremely alarming.[29]

Data and technology are human inventions, ideally designed to reflect and advance human values. As our creations grow exponentially more powerful and their footprint ever larger on our society, we need to be increasingly mindful of the need to build them to be robust against adverse and unintended consequences.

We cannot blindly trust the output of automated systems without vetting the accuracy of both the input data and the decision-making process itself. Many machine learning algorithms already influence our daily decisions and actions, but bad data and methodological mistakes can easily lead to erroneous results. In California, a flight-risk algorithm in use by the San Francisco Superior Court mistakenly recommended a man for release before trial. Despite multiple previous probation violations and arrests for gun possession, the algorithm judged him to be a minimal flight risk because someone had mis-entered the number of days that he had already spent in jail. Five days after release, he and a partner shot a local photographer.[30]

More subtle and insidious is the danger that algorithms designed by an undiversified team of elites may overlook the needs and values of underrepresented groups and unintentionally amplify the discrimination against them. Amazon customers, for example, discovered that same-day delivery was unavailable in zip codes that contained predominantly black neighborhoods, while computer scientists at Carnegie Mellon found that women were less likely than men to be shown ads for high-paying jobs.[31] Even if characteristics such as race, religion, gender, or ethnicity are eliminated from models, other features that are highly correlated with those characteristics may be included and introduce the same bias.

The biases of technology creators trickle down to their creations. While AI researchers pride themselves on being rational and data-driven, they can be blind to issues such as racial or gender bias or ethical issues that aren't easily captured by numbers. With AI now used in high-stakes systems to identify terrorists, predict criminal recidivism, and triage medical cases, homogenous thinking in the technology industry has dangerous implications.

The Effects of Discrimination

To Latanya Sweeney, the first black woman to receive a PhD in computer science from MIT, the shortcomings of AI come as no surprise. Currently a professor at Harvard and the director of their Data Privacy Lab, Sweeney's research examines technological solutions to societal, political, and governance challenges. One of her important contributions illuminates discrimination in online advertising, where she discovered that online searches of names that are more associated with the black community are 25 percent more likely to be targeted by ads that implies the person being searched for has a criminal record.[32] Sweeney also uncovered SAT test prep services that charge zip codes with high proportions of Asian residents nearly double the average rate, regardless of their actual income.[33] While price discrimination based on race, religion, nationality, or gender is illegal in the United States, enforcement of existing law is challenging in e-commerce, where the evidence of differential pricing is obscured by opaque algorithms.

In healthcare, AI systems are at risk of producing unreliable insights even when algorithms are perfectly implemented, because the availability of medical data is affected by social inequality. Poorer communities lack access to digital healthcare, which leaves a gaping hole in the medical information that is fed into AI algorithms. Randomized control trials often exclude groups such as pregnant women, the elderly, or those suffering from other medical complications.[34] Such exclusions mean that the unique physical characteristics of these patients are not incorporated into studies, which in turn affects whether tested treatment will be effective on patients who don't share the characteristics of the original clinical volunteers. In the worst-case scenario, the treatment may actively harm the patient.

Advocacy for algorithmic fairness cannot solely be the responsibility of the disenfranchised. Lasting, fundamental changes can only happen when technology creators and the public at large awaken to the dangers of exclusion and make inclusion a true priority.

Malicious AI

We don't have to wait for AI to gain sentience and go rogue, because the probability of bad people taking advantage of intelligent automation for evil purposes is 100 percent. As machine intelligence becomes more powerful, pervasive, and connected, embedding AI in all of our personal and industrial computing devices increases the risk of attacks that can compromise the security infrastructures that protect our resources and communities.

Luminaries from the Future of Humanity Institute, OpenAI, Centre for the Study of Existential Risk, and leading universities in the US and UK issued a 100-page policy recommendation paper, "The Malicious Use of Artificial Intelligence,"[35] in which they described the fast-evolving threat landscape, identified key areas of security risk, and made high-level recommendations for preventative action that should be taken immediately.

The report was alarming, pointing out that existing threats will get worse while new threats of an unknown nature will almost certainly emerge. AI will be used to multiply the effects of a malicious campaign—augmenting "human labor, intelligence, and expertise" to make the process of attacking easier and faster—and to broaden the types and number of possible targets. Advances in neural network algorithms that can produce hyperrealistic audiovisual input may be hijacked to produce fake news that looks like it came from a credible source, or to circumvent security systems that use voiceprints or other identifying features. In addition, AI may fundamentally alter the arena of cyber attacks by increasing the efficacy, precision, and untraceability of such attacks. These attacks may even target and hijack supposedly secure AI systems by exploiting their vulnerabilities. One grim possibility is the deployment of autonomous weapons systems, such as a drone, using facial recognition technology to identify and attack individuals in a crowd.

Through wearables, standard computing devices, and the burgeoning Internet of Things (IoT), AI will inevitably permeate every corner of our existence. This means that our physical security, digital security, and even political security will be at risk of attack. While we spend much of our productive hours tethered to digital devices and roaming cyberspace, we still inhabit physical bodies and live in a material world. Nefarious AI can infect autonomous vehicles, connected appliances, and other devices to inflict bodily harm and property damage. Digital attacks may come as a coordinated and adversarial disruption of corporate data with the goal of compromising, devaluing, or altogether destroying an organization's data architecture. Finally, the use of technology—including AI, predictive analytics, automation, and social media bots—can have far-ranging social impact. AI can be used for illegal surveillance, propaganda, deception, and social manipulation.

5. DESIGNING SAFE AND ETHICAL AI

Ethics and Governance

AI systems can't simply be programmed to complete their core tasks. They must be designed to do so without unintentionally harming human society. As AI systems become more complex, the likelihood of facing ethical dilemmas also grows. Designing safe and ethical AI is a monumental challenge and a critical one to tackle now. To be effective, we must develop more sophisticated and nuanced policies that go far deeper and wider than simplistic, science fiction solutions like Asimov's Three Laws of Robotics.[36]

In a joint study, Google DeepMind and the Future of Humanity Institute explored fail-safe mechanisms for shutting down rogue AI.[37] In practical terms, these "big red buttons" will be signals that trick the machine to make an internal decision to stop, without registering the input as a shutdown signal by an external human operator. IEEE, the world's largest association of technical professionals, published *Ethically Aligned Design,* a set of standards for the ethical design of artificial intelligence and autonomous systems.[38] The publication lays out the chain of accountability for design and operation. It also emphasizes that to limit the possible extent of risks, such systems should not infringe on human rights, and their operations should be transparent to a wide range of stakeholders.

Hypothetical fail-safe mechanisms and hopeful manifestos are important, but they are insufficient for addressing the myriad of ways in which AI systems can go awry.

Homogeneous development teams, insular thinking, and lack of perspective lie at the root of many of the challenges already manifesting in AI development today. Luckily, as AI education and tools become more accessible, product designers and other domain experts are increasingly empowered to contribute to a field that was previously reserved for academics and a niche community of experts.

Education as Remedy

Tackling these challenges requires democratizing access to quality AI education and empowering collaborations between practitioners and multidisciplinary experts in order to gather missing data and build inclusive technology. Acquiring the requisite knowledge and resources to apply AI is a huge challenge for those who don't live in Silicon Valley or other major research hubs.

Many turn to massive open online courses (MOOCs) provided by companies such as Coursera, Udacity, and fast.ai as their only options.

Rachel Thomas, a deep learning researcher with a doctorate in math from Duke University, started fast.ai with Jeremy Howard, the former president of Kaggle, to advance the mission of making deep learning accessible to all. As passionate champions of diversity and inclusion, the two have taught over 50,000 students globally, including Sahil Singla of FarmGuide.

Fast.ai's non-stop efforts to democratize AI education are paying off. Students of its MOOC are using techniques taught in the class to treat Parkinson's disease, give visually impaired patients more independence, fight online hate speech, and end illegal logging and harmful human activity in endangered rainforests.

The work is not done, however. Even with MOOCs, students in developing countries face an uphill battle compared to their counterparts in developed countries. Some struggle with the lack of structured datasets available in their language or culture, others with the lack of reliable internet infrastructure and access. Still others face a lack of career opportunities. Finally, the lack of affordable access to computational resources, such as graphic processing units (GPU) and reliable power sources, presents a major obstacle for students who want to build their own models. Even with the right hardware, complex neural network models can take days, if not weeks, to train.

Even if computational resources were widely available, engineering education alone is insufficient to ensure that AI technologies are built safely and successfully. "Ethics training should be a mandatory part of engineering and computer science education," emphasizes Rana el Kaliouby, founder of Affectiva, a company that makes machines more emotionally intelligent.[39] El Kaliouby and her team regularly engage the public in open dialogue to uncover potential blind spots regarding transparency, privacy, security, and ethical concerns.

Improving access to tools and education will bring in new expertise and viewpoints that can help evolve a field traditionally driven by an elite few. With AI's exponential impact on all aspects of our lives, this collaboration will be essential to developing technology that works for everyone, every day.

Collaborative Design

As you embark on building your own AI technologies for your business or community, the following three principles of collaborative design will help you and your team approach AI development more holistically and successfully. Bringing in diverse expertise and thinking is critical to ensuring your technology is benevolent to all members of society and does not unconsciously reflect the biases of an elite minority.

BUILD USER-FRIENDLY PRODUCTS TO COLLECT BETTER DATA FOR AI

Data is a human construct, as are the tools that we design to gather it. Consumer-facing digital data is largely captured through the myriad of touch points that we have with our internet-connected devices and the complex ecosystem of apps, content, and networks that we access through them. If the products collecting requisite data to power AI systems do not encourage the right types of engagement, then the data generated from user interactions tend to be incomplete, incorrect, or compromised.

In designing a product, you are building a specific journey for your customers to experience, and you will invariably influence user behavior and the resulting data trail. Manipulative products like clickbait headlines and aggressive calls-to-action (CTAs) optimize for short-term gains in lieu of long-term relationships, and the data they collect may not serve your ultimate business goals. Even if you are intentional in both your data collection and your product's user experience (UX) design, remember that just because a user engaged with a button or clicked on an ad doesn't mean you know their motives or intentions.

The absence of experiential knowledge means that you cannot solely rely on data and algorithms to tell you which problems need solving. Machine learning and AI are not always the right solutions to a problem. Identifying the right problem and its solution requires tight integration and adaptation between your products and your users as well as a collaborative relationship between your team and your users. The UNICEF U-Report bot is a great example of this principle in action. Its key innovation was in designing the product to work over a single phone line for users who lacked smartphones and computers, not in its application of novel AI methodologies.

PRIORITIZE DOMAIN EXPERTISE AND BUSINESS VALUE OVER ALGORITHMS

When working with Fortune 500 companies looking to reinvent their workflows with automation and AI, we often hear this complaint about promising AI startups: "These guys seem really smart, and their product has a lot of bells and whistles. But they don't understand my business."

In most cases, having and using a fantastic machine learning algorithm is less important than deploying a well-designed user experience (UX) for your products. Thoughtful UX design that delights users will drive up engagement, which in turn increases the interactions you can capture for future data and analysis.

Thoughtful UX compensates for areas where AI capabilities may be lacking, such as in natural language processing (NLP) for open-domain conversation. In order to develop "thoughtful UX," you'll need both strong product development and engineering talent as well as partners who have domain expertise and business acumen. A common pattern observed in both academia and industry engineering teams is their propensity to optimize for tactical wins over strategic initiatives. While brilliant minds worry about achieving marginal improvements in competitive benchmarks, the nitty-gritty issues of productizing and operationalizing AI for real-world use cases are often ignored. Who cares if you can solve a problem with 99 percent accuracy if no one needs that problem solved? What's the utility of a tool whose purpose is so arcane that no one is sure what problem it was trying to solve in the first place?

EMPOWER HUMAN DESIGNERS WITH MACHINE INTELLIGENCE

"Tools are not meant to make our lives easier," says Patrick Hebron, author of *Machine Learning For Designers*, "[t]hey are meant to give us leverage so that we can push harder. Tools lift rocks. People build cathedrals."[40] Human designers can enhance their creations when they are supported by tools that use machine intelligence. The nascent field of AI design is one such area. While we are still figuring out which best practices should be preserved and which new ones need to be invented, many promising AI-driven creative tools are already in use.

Hebron insists that machine learning can be used to simplify design tools without limiting creativity or removing control from human designers. Machine learning can transform how people interact with design tools through emergent feature sets, design through exploration, design by description, process organization, and conversational interfaces. Hebron believes that these approaches can streamline the design process and enable human designers to focus on the creative and imaginative side of the process instead of on technical mastery of the design software. This way, "designers will lead the tool, not the other way around."