THE FUTURE OF HUMAN-ROBOT COLLABORATION

what to expect when you're expecting robots

LAURA MAJOR
and
JULIE SHAH

BASIC BOOKS
NEW YORK

Contents

Dedicated to our children, Luca, Lily, Lillian, and Vivien,
who inspire us to make a better world.

BASIC
BOOKS

Introduction

Imagine a world full of robots. It's a bit like the world of today, except these robots are not just expensive novelties. They aren't limited to a small handful of jobs and don't need you to tell them what to do. Instead, these robots are something like partners—they cooperate with you the same way teammates on the basketball court cooperate with each other. One sets a pick so the other can roll; one lobs the ball high above the basket and another swoops in to dunk. We call this *human-robot collaboration*, and it is likely to revolutionize our relationship to technology over the next several decades.

People seem to be concerned about whether robots will one day make us obsolete—whether they will become smarter, faster, better than their human creators. But the reality is that robots and humans will probably always be good at different things. And, as we intend to show here, it is possible that some of our most stubborn societal problems could be better addressed by the kind of collaboration we envision. The applications are vast. Through a symbiosis of human and artificial intelligence on the road, we can dramatically reduce fatalities due to car accidents and start to tackle the congestion problems that plague nearly every city in the world. Robots as personal augmentation systems can improve daily well-being and enable us to thrive independently far into old age and as our abilities change. Robotic orderlies can make emergency rooms safer and more efficient, shortening wait times and enhancing care. Robots will bring countless other small but meaningful improvements to our daily lives, and as working moms with an ever-growing to-do list, we personally look forward to these changes.

You may be thinking that robots like this are already here. After all, your Roomba can vacuum your living room on its own. But while your Roomba's ability to map your floor plan might seem impressive, it is not really much different from any other

household appliance. And this is true of most robots we currently encounter. We restrict their roles with simple rule-based behaviors and interact with them through taps on screens and other simple commands. They understand very little about us, and we request relatively little of them in return. Robots in factories work in cages. We turn on adaptive cruise control for our commute, but turn it off as soon as we hit traffic. We wake up in the morning and ask Siri about the weather, or tell Alexa to add milk to the shopping list, but ultimately, we dress ourselves and buy our own milk. We don't judge our Roomba too harshly when it gets stuck on a tuft of carpet or misses spots. It is a simple machine, not a very smart one. Most robots today have narrow functionality, can only operate in controlled environments, and require essentially constant human oversight. And given those three limitations, they perform beautifully.

But human-robot collaboration is something altogether more revolutionary. New types of intelligent robots are just now beginning to enter our cities and workplaces, and they are defined in large part by the way they transcend these limitations. Robots making package deliveries in our neighborhoods, or shopping for us at grocery stores—what we call *working robots*—can no longer be considered mere *tools*. They amount to new social entities. Let us be clear: whether these robots can be said to be conscious or as intelligent as humans is not really the point, and indeed, many working robots will be a far cry from sentient. What we need to understand about tomorrow's robots is that they are going to be something *different*, with roles mediated at all stages by the rules of social interaction. They will become more human in one specific way: whether they make our lives better or worse comes down to whether they know how to behave.

And there are many of them coming. If the picture we are painting feels like a far-off dream, it is because we are sleeping. There are 1.7 million industrial robots in operation around the world today.[1] That's the same number as the human population of Boston, Pittsburgh, and San Francisco combined. There are currently 30 million robots in our homes in the United States.[2] And that's not counting the Alexas, Siris, smart home devices, sidewalk delivery robots, grocery store robots, apartment security guard robots, and hospital service robots now making appearances

as we visit friends, run errands, and go shopping. Soon, your front yard and our neighborhoods may be swarming with drones. The National Aeronautics and Space Administration (NASA), entrepreneurs, and industry leaders are working quickly to open up our skies to urban air mobility—where drones deliver small packages and passengers zip across town in the air over roadway traffic.

IT IS TUESDAY MORNING. YOU STEP OUT OF YOUR HOUSE AND WALK toward your car, which is parked on the street. Meanwhile, a delivery robot is zipping down the sidewalk, attempting to deliver packages in time for the single-day shipping deadline. It detects you as a nearby obstacle and stops for safety, but not in time. You snag your foot on it and lurch forward, catching your fall. It's already a bad way to start a day. As you drive to work, you pause for a pedestrian to cross the street, and just as you proceed you spot a small assistant robot, probably carrying the person's laptop and lunch, trailing behind. It's low to the ground, like a puppy—and you almost didn't see it. You slam on the brakes, narrowly avoiding the robot with its cargo of laptop and sandwich, but the car behind you gently rear-ends you. After you thankfully confirm there was little damage to either of your bumpers, you get back on the road, but you feel the frustration grow when you realize you are now behind an autonomous vehicle doing a test run through a new neighborhood. It feels like time has slowed while you pace behind the vehicle at just below the speed limit, stopping for every object within ten feet regardless of whether the object is actually on the road. You're ready to pull your hair out by the time you arrive at the office, only to be stuck at the elevator bank by a delivery robot that you can't get to move out of the way of the buttons. It is only 9:00 a.m., and four different robots have already made your life more difficult, simply because they have not been designed to understand or care about you. You can't help but wonder who these robots are really helping.

How can you make a robot that understands strangers? The way to make this work is not to build arbitrarily "smarter" or "more powerful" robots, but to rethink what it is we expect from technology. Consider the search-and-rescue dog, for example. The

dog doesn't have to be commanded so much as it is guided by its human handler—with subtle hand gestures indicating areas for focus of attention. The dogs often act on their own. Their handlers depend on them, and the dogs have their own set of social norms for interacting with people. The vests they wear remind people not to touch or interact with them. In challenging spaces, they are put on a leash, so that their behavior might be more tightly controlled. They are still dogs, but because their roles—and their handlers' roles—have been carefully designed, search-and-rescue teams are able to do much more than either a dog or a person can do independently.

We envision human-robot collaboration in this way: people and robots buzzing around each other, sometimes working as individuals, and other times collaborating in groups. With a wave of a hand we will be able to offload a burdensome task to a robot, and people in need might call on multiple robot helpers to accomplish things they would not be able to accomplish on their own. But getting this right is a two-way street. Just as robots will need to be able to understand social norms, they will also require us to reconsider the place of technology in our everyday lives. We will have to make certain changes, individually and as a society, to incorporate robots into our world. This partnership will require new human and robot languages and norms. We will have to rethink our infrastructure. We will have to consider the implications of the fact that these robots will be commodities, available to some and not to others. And because all this is being set in motion by a handful of tech companies, we will need to be clear about industry's ethical responsibilities. It will take deliberate, collective action. That is the purpose of this book: to figure out what makes a robot socially and personally valuable, and then consider how we as a society can ensure that the robots we make have these characteristics. Robots will force us to rethink the role of technology in society—from the practical level, such as figuring out how robots will deal with bystanders as they make their way through our neighborhoods, to more philosophical issues, especially navigating the tension of how these technologies will differentially impact groups in society. The tech industry is currently leading this change, but the arrival of autonomous robots in society impacts all of us. In order to fully embrace this

future, it will require whole-society efforts.

Over the past few years, Julie's husband—a physician—has frequently texted her pictures of new robots he has encountered in the hospital where he works. One day a new one appeared, delivering medications floor to floor. A medical doctor entered an elevator to find one of these robots with a sign on it: "DO NOT GET IN ELEVATOR WITH ROBOT." The hospital staff was not supposed to come in close contact with the robot or interact with it in any way because it was still learning how to function safely and effectively in the human environment. Occasionally, the robot could lose its sense of where it was within the hospital, and would stop or start up again abruptly as it searched for clues. Naturally, this unpredictable behavior would be a concern in a confined space like an elevator. The robot was still basically a student, and just like a student driver, it needed to be treated with caution.

We know instinctively how to change our driving behavior when we see the "Student Driver" sign on a car. Good driving is not just about knowing the rules contained in a handbook; it involves the many informal rules and behaviors that drivers can only really learn with experience. That drivers know and abide by these rules make driving (for the most part) predictable. Student drivers do not yet have these mental models. Their ignorance, and perhaps their nerves, too, make them somewhat unpredictable, which may create an unsafe environment. The sign means that every experienced driver around that car should expect the unexpected.

Imagine how exhausting and stressful it would be if you had to drive every day surrounded by student drivers. Now imagine how exhausting and stressful—and possibly dangerous—it's going to be to coexist with hundreds of robots in our lives every day, whether on roads, in the hallways of our office buildings, in our parking lots, in our restaurants or hospitals, or buzzing around overhead as we walk down the street, especially if they don't understand the rules we all follow that make those spaces navigable and safe. The fact of the matter is that, at some point soon, robots will not yet be like "people," but neither will they be strictly "tools," only moving when we command them. They will be something altogether new. But what will that be? We believe that tomorrow's societies will be run increasingly by a new kind of relationship with technology, a

human-machine partnership. The upside is tremendous. Consider that car accidents cause nearly 1.25 million deaths annually across the globe.[3] That's over 3,000 fatalities on our roadways every day, about 100 per day in the United States alone.[4] It is the ninth leading cause of death worldwide. And yet there were only 500 deaths due to commercial aviation accidents in 2018 and only 144 fatalities the year before.[5] This is in part because the aviation industry has already embraced the idea of a human-machine partnership. Over the past few decades, the industry has reconceptualized the relationship between the pilot and the aircraft, and as a result, each is able to compensate for the difficulties of the other, and the skies are safer. Automation offers us the opportunity to reach similar, seemingly unachievable safety results on our roadways. Imagine a world in which fewer than 100 people die annually from traffic accidents around the world. With human-robot collaboration, such a future is possible.

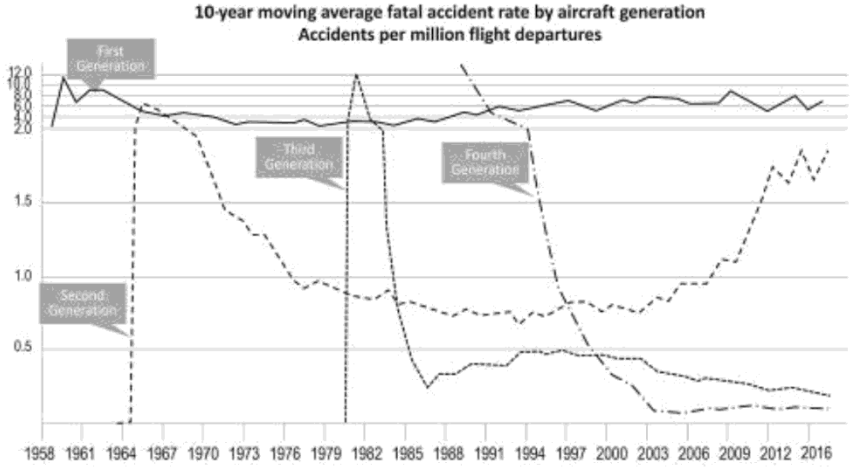

FIGURE 1: Accident rates for four levels of technological innovation in airplane technology: first generation (early commercial jets), second generation (more integrated auto-flight), third generation (glass cockpit and flight management system, or FMS), and fourth generation (fly-by-wire). *Source*: *A Statistical Analysis of Commercial Aviation Accidents, 1958–2016* (Blagnac Cedex, France: Airbus, 2017), https://flightsafety.org/wp-content/uploads/2017/07/Airbus-Commercial-Aviation-Accidents-1958-2016-14Jun17-1.pdf.

Achieving the safety standards that have made aviation so safe and reliable hasn't been easy. But the hard lessons learned offer some guidance as we begin to think about introducing robots into our lives that we can trust not to cause real harm. Because, make no mistake: although some independent robots will be doing work that is potentially only annoying or inconvenient, others will have the potential to hurt or even kill people. Any autonomous system unleashed into the chaotic realms of our modern lives without careful consideration of how it will impact us could be dangerous. In this book, we will look, by turn, at the kinds of decisions we will have to make when building a new social entity. Probably the most important lesson we can offer is this: people design robots, and people are imperfect. Each time automation systems were introduced into cockpits—autopilot, glass cockpit, and fly by wire—fatal accidents temporarily increased. Only after the initial problems were addressed were we able to reap the benefits. No matter how hard an engineering team tries to properly design a new system or how rigorously evaluators and regulators test it, it can never be perfect.

Through decades of hard work, experimentation, and refinement, engineers have optimized the complex human-technology partnership that makes our commercial air transportation system work for our benefit and well-being. But, much like parenthood, a human-technology partnership takes work. It is not something we can ever expect to be perfect right out of the gate. Think about the learning curve for pilots today. They still have to train extensively on the logic and behavior of the automation, and learn how to communicate with and rely on the automation. But they also train to maintain the skills to manually fly an aircraft in case the automation fails. The partnership between a pilot and the flight management system is honed through this training. The automation can't be preprogrammed to work perfectly with the pilot, in much the same way a person can't be programmed to live in a happy marriage with another human being. Developing automated systems takes time, expertise, and financial investment. But how many of these working robots can we reasonably invest in? There will be many robots performing many more tasks than we can imagine today; they will be navigating our everyday world as best they can; and we will have

to work with them as best we can without always having the luxury of extensive training or know-how. Still, the lessons from aviation offer insight into how robots and humans can work together, and we'll discuss many of these lessons at length. The main point remains: human-robot collaboration will force us to conceptualize new ways of integrating technology into society.

Determining what makes an effective human-machine partnership is trickier than it initially sounds, especially when safety matters. Imagine the dangers of negotiating throngs of robots zipping along our sidewalks as we walk down city streets. The problem is not just a matter of scale, of the increasing number of robots in our everyday lives, or of the physical proximity to more robots. Instead, the core of the challenge is a shift in the nature of consumer automation itself—from accessory technology to safety-critical systems.

Figure 2 illustrates what we mean. Here, we have attempted to capture the cost of failure across different applications and the amount of training that is required for their operation. Industrial applications, such as commercial flight, are highly complex: it is not possible to safely control these systems without robotics. Furthermore, the operators of these systems are highly trained, not only on the fundamentals of the applications (such as physics, aerodynamics, and electromechanical systems), but also on the robot. This training gives them the knowledge they need to manage the system even in the face of failures. Industrial applications are represented with squares in the figure.

Consumer products, in contrast, do not historically pose any serious safety risks and are typically designed to be used out of the box without training (beyond, perhaps, reading the instruction manual). Examples include Siri, Roombas, and Alexa. These are represented in the figure with diamonds.

Collaborative working robots represent a new class of consumer products that fall somewhere in between traditional consumer products and industrial applications. They introduce robotics into safety-critical activities and areas—self-driving cars on our streets, delivery drones on our sidewalks, robots monitoring inventory or cleaning up spills in grocery stores, and medication-delivery assistants in our hospitals. Whether such robots can survive in the real world is ultimately a question of

whether we can figure out how to interact with them. We already interact with consumer technologies, but how we interact now will not sufficiently account for all the new, potentially dangerous things these robots will be capable of doing in the future. Surely, they cannot be designed using exactly the same process as cockpit or spacecraft automation—we cannot afford to have all of us become experts about every robot we encounter, as pilots do with flight training, and anyway, everyday life is a lot harder to predict than a flight pattern. Getting the new hybrid design process right will be the key to unlocking an exciting and productive future where humans and machines truly partner to enhance our lives, and where we will each draw strengths from the other's capabilities. Getting it wrong is, for all the damage it will do, simply not an option.

FIGURE 2: A comparison of operator expertise, measured in hours of training and cost of failures, for three classes of applications: industrial applications (squares), commercial products (diamonds), and a new class of safety-critical commercial products (triangles). *Source*: Laura Major and Caroline Harriott, "Autonomous Agents in the Wild: Human Interaction Challenges," in *Robotics Research: The 18th International Symposium ISRR*, ed. Nancy M. Amato, Greg Hager, Shawna Thomas, and Miguel Torres-Torriti, Springer Proceedings in Advanced Robotics, vol. 10 (Cham, Switzerland: Springer, 2020).

The second major challenge is that this new class of safety-critical consumer products cannot be designed without careful consideration of social norms, just as aircraft cannot reasonably be designed without consideration of the air transportation system infrastructure and constraints. Aircraft are designed with special equipment to communicate with air traffic control and other aircraft. Pilots must have their flight paths approved before takeoff, and they must request a change and receive approval before modifying that path during flight. Their navigation solution is based on acceptable options defined by regulatory bodies, and different rules apply to aircraft with different capabilities. Some must stay at lower altitudes and only fly on clear days; others are

allowed to fly when visibility is extremely low and can get closer to other aircraft as they pack into smoother, faster tracks, reducing flight time and delays. Similarly, working robots will need to work within social norms and abide by rules and regulations that guide their safe use across societal situations. We must embrace this change and be thoughtful about how to design robots as partners, creating the necessary infrastructure and support for these new entities rather than having to relearn the hard lessons we encountered in the early decades of air transportation. In other words, we must begin to understand robots as *sociotechnical* systems.

The central goal of this book is to explore how the unprecedented design challenges of tomorrow's working robots force us to confront a central question about the role of technology in society: What can we expect from machines, and what can they expect from us? Out of this central question come several ideas that will structure our tour: How can we harness the relative strengths of humans and robots? Can autonomous systems be too independent for their own good? How do you plan for the bystanders who will be subjected to other people's systems working in public spaces? We provide design frameworks and solutions for how humans and machines can better predict each other's behavior. As a key to our approach, we introduce the notion of *automation affordances*, that is, design features that provide clues to the user or bystander as to how they might be able to influence or adjust a robot's behavior. More than natural language capabilities, automation affordances will provide for a basic language for robots and people to communicate with each other in any encounter. The introduction of working robots, as a social concern, will require a level of transparency and collaboration that is currently mostly absent from the tech industry. We will discuss new methods of evaluating and testing intelligent machines in order to ensure that profit motives do not put public safety at risk. We will also discuss the limitations of conceiving of these challenges as solely technological problems, and the ways in which we will need to co-design our society for working robots.

While there is much we can learn from this broader perspective—taken from aerospace, industrial systems, and

human-systems engineering—we also address the ways in which we are now grappling with challenges of human-robot collaboration that have no precedent in other industries. Working robots are safety-critical consumer products operating within environments that are less controllable and predictable than the applications in these other arenas, and they will be interacting with people who can be expected to have little or no training with them before they arrive on the scene in large numbers.

In this new setting, our old paradigms for using and working with technology quickly fall away. It feels quite comfortable to us to command our simple robots today, because we are clearly in control. But the world is changing rapidly around us. Robots are quickly evolving: instead of mere tools that we command and query, they are becoming intelligent partners that we will work with. We know this because it's our job in academia and industry to drive the advancement in artificial intelligence and robotics to make this vision a reality. Julie leads an artificial intelligence (AI) and robotics research lab at the Massachusetts Institute of Technology that focuses on the future of work and the potential for reverse-engineering the human mind to make robots better teammates. She has pioneered new forms of human-robot teaming in manufacturing, transportation, and health care. Laura has been leading teams in industry to design and develop new autonomous systems for our skies and roadways, revolutionizing digital assistants on our battlefields and bringing autonomous cars into reality. We are aided in this undertaking by unprecedented levels of private and public investment in new intelligent robot technologies, with visions of smart cities, schools, and workplaces on the horizon.

Along with these visions of possible futures, many of us feel great unease about the potential economic, workforce, and societal implications of these new technologies. We see billboards on our roads by insurance companies urging us to prepare for retirement because the robots are coming. Many of us worry about automation taking over our jobs, or about the *singularity*: the moment in the future when we enter an era of exponential growth in technology due to advances in AI.

It will surely require a collective effort to ensure that the new intelligent robots are harnessed to the task of enhancing human

well-being. But here is the central contention of this book: robots need not be superintelligent, bent on world domination, or capable of making the entire human workforce obsolete to pose a threat to human prosperity. If they don't know how to behave in public, that will be enough. A robot that knows how to ride the subway could be truly revolutionary. One that doesn't, and tries to catch a train anyway, could do more than just make a few people late for work.

There is a big difference between functional use of technology and acceptance of robots en masse: one sidewalk delivery robot may be a novelty, whereas hundreds of them blanketing your city could be a dangerous prospect. We are sitting at the precipice of this phase transition, and we must begin to shift our conversation from one of fear to one of solutions. Those solutions live at the intersection of society and technology. It takes a village to raise a child to be a well-adjusted member of society, capable of realizing his or her full potential. So, too, a robot.

CHAPTER ONE

The Automation Invasion

FOR AS LONG AS ROBOTS HAVE BEEN IMAGINED, WE'VE WONDERED not only what they can do, but what they should do. Yet such debates have always seemed a bit academic, because in our daily lives they have started popping up around us without much notice. It feels as if we are always playing catch-up, never with enough lead time to think deliberately about the technology and its role in our lives. Autonomous cars seemed to just appear on our roads one day. We crossed the line to a tipping point in technology where automated driving suddenly seemed within reach—so industry, government, and venture investors went for it, taking the rest of us with them. Robots have started to appear in grocery stores, gliding up and down the aisles looking for spills and other safety hazards. The customers in the Stop and Shop we visit in suburban Boston take it all in stride, for now at least. But supermarket employees wonder how quickly the roles of these robots will expand.

The emergence of new technologies can often feel abrupt, almost magical. This is because most of us are unaware of the many years of often incremental, commercially uninteresting technological innovations that make those breakthroughs possible. Occasionally a headline will herald some new conceptual development, which will quickly recede into the background of most people's lives. It may take decades before a robot appears that turns those technical innovations into a marketable machine.

Self-driving cars are our bellwether for the emergence of a new class of intelligent robots to be unleashed on society. They mark one of our first opportunities for everyday people to grapple with questions of sharing decision-making authority and control with an autonomous system. But Mercedes-Benz developed one of the first self-driving cars as early as the 1980s, and it could navigate streets without traffic at speeds of up to thirty-nine miles

control the body.

Finally, the brain takes the sensory inputs, forms an understanding of the world, makes decisions based on that understanding, and implements actions through the other components of the nervous system. In some animals, the brain can be simple—like the first Roomba, which followed one basic rule: clean in a spiral pattern until it hits a wall, and then follow that wall. In others, it can be complex, like the autonomous vehicles we see on the road that are increasingly capable of navigating construction zones and traffic jams.

We first invested in automation for aviation and industrial applications—such as in the control centers of nuclear power plants—when the tasks required to work these new technologies proved too hard for humans to perform reliably. Today, many safety procedures for nuclear power plants are fully automated and require no human intervention. The systems are too complex for people to monitor or to control manually, and the consequences of a failure are too high. In other words, automation allowed us to reimagine what we were capable of after we had maxed out our natural capabilities.

Robots excel in these applications in aviation, power plants, and factory settings because the environment is tightly controlled and there are detailed task procedures for them to follow. And so robots first infiltrated these industrial worlds that very few people have access to, or were placed in spacecraft, which even fewer people have access to. In these isolated worlds, engineers developed robots' sensors, nervous systems, and brains piece by piece, then tested them, learned from their failures, and improved and hardened the new technologies. Today, by and large, complex industrial applications are heavily controlled by automation with minimal human supervision.

The question now before us concerns the most effective way of using robots to automate our daily lives. We are in the midst of an active debate among manufacturers, legal scholars, and engineers about whether to automate various aspects of driving, and if so, how. Is a hybrid human-car approach the best, or should the person be removed entirely?[9] In fact, we already had to answer these sorts of questions decades ago, when designing the first moon-landing spacecraft, as well as when introducing cockpit

automation for air transport. As shown in figure 3, history bears out that seemingly small design decisions can have repercussions for decades, and we'll walk through some examples.

The Apollo Lunar Landing Training Vehicle was the first fly-by-wire aircraft, essentially the first aircraft with a central nervous system. Other advancements brought the Apollo Guidance Computer, which served as the brain of the Apollo Command Module and the Apollo Lunar Module. Landing on the moon was certainly a complex problem, one that neither human nor machine could do alone. The technical debate started then over how much authority the computer should have and how much authority should remain in the hands of the astronauts.

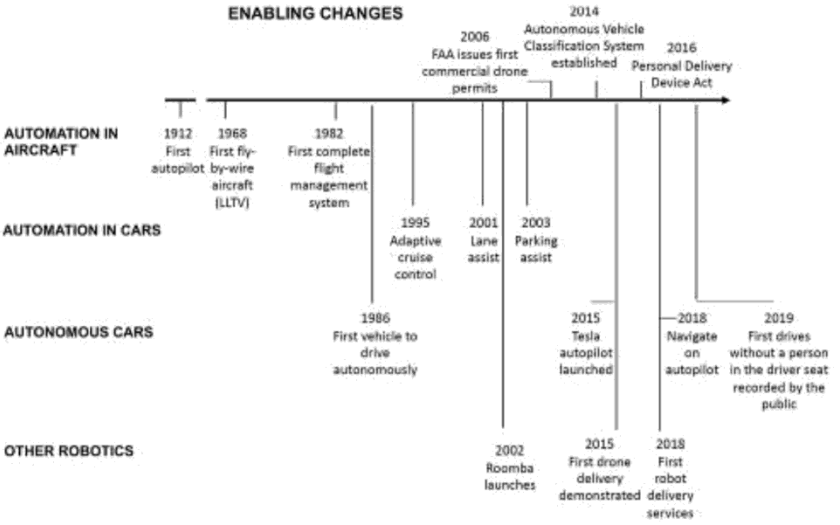

FIGURE 3: Timeline of key enabling changes across robotic applications to depict the technological trends that cross domains and have led to the phase transition that now brings us working robots.

In a seminal paper in the 1960s titled "The Role of Men and Instruments in Control and Guidance Systems for Spacecraft," designers of the Apollo Guidance Computer contemplated just this question.[10] They debated how much to trust the computer in moments of crisis or when time-critical decisions had to be made. The engineers described three types of events based on how decisions are made. Type I events involved foreseeable conditions with predetermined responses, such as the automatic cutoff of a rocket stage at a predetermined velocity. These situations, the engineers said, would be easy to automate. Type II events involved foreseeable situations for which "no appropriate . . . action could be programmed in advance due to the complexity of the general case . . . like landing on an arbitrary spot on the moon." The engineers concluded that Type II events were not straightforward enough to fully automate, but that human performance could be "enhanced with feedback on performance indicators and display of only relevant information." Finally, there were Type III events, defined as those that could not be anticipated, even by the designer or the pilot. The engineers determined that humans would far outstrip automation "in making decisions based on incomplete data in completely new situations."

All of this remains true today, except that technological advancements—in particular machine learning—have changed what we consider "the complexity of the general case" or "a completely new situation." Whereas in the past we had to craft decision-making rules for the automation for a variety of situations by hand, machines can now leverage data or demonstrations to learn an approximation of our human decision-making criteria that may be too complex or time-consuming for us to manually specify. Still, when the unforeseen happens, automation often falls short. In 2009, in what would come to be known as the "Miracle on the Hudson," the pilots of US Airways Flight 1549 successfully ditched an airplane in New York's Hudson River after the plane struck a flock of Canada geese and lost all engine power. Everyone on board survived. We can nearly fully automate the flying of a commercial airliner, but could an autonomous airliner have pulled this off? It is unlikely with today's technology. No artificial system today can replicate a human's

capacity for creative problem-solving.

Even in well-established industries, such as aviation, where automation is prevalent, we still debate what flight deck automation should control and what a pilot should ultimately still control. When you board an airplane, you may not think much about whether it is an Airbus or a Boeing. But decades ago, the two companies chose quite different paths representing different philosophies for how pilots and intelligent automation were to work together. When automation systems were first introduced, if you were riding in an Airbus, the plane's systems would have had authority over the pilot, but the opposite was true on a Boeing. So by and large, the automation could override the pilot in an Airbus, and the pilot could override the automation in a Boeing. These approaches resulted from a basic decision in the design of robotic systems that had to do with hard versus soft automation.[11] Hard automation has more protection against human error, essentially constraining the user from doing something that would put the vehicle in danger. Soft automation still employs safety constraints, but it considers the automation an aid: it provides an alert when the user is about to do something that may be dangerous, but allows the user to proceed and to override the warnings if they choose to do so. The latter allows for more human creative problem-solving. In other words, with soft automation, the user always has access to the full capabilities of the vehicle, whereas with hard automation, there are certain circumstances in which they do not have that access.

Of course, there are pros and cons to each style. For example, in 1985 a China Airlines Boeing 747 had an engine failure while cruising at forty-one thousand feet. It entered an uncontrollable dive and plunged more than thirty thousand feet. With the soft automation system, the pilots recovered control and passengers sustained few injuries. Analysis indicates that a hard automation protection system would have disallowed the pilots' inputs, the very thing that enabled them to successfully regain control of the aircraft. On the other hand, it could be argued that an Airbus flight control system, with its hard automation protections, would not have allowed the aircraft to enter the uncontrolled dive to begin with, because it would have prevented the pilot from taking the actions that got the system into the unstable state. Overall, it

FIGURE 4: This MIT Instrumentation Lab cartoon shows the impact of the extremes of automation. Even during the design of the Apollo Lunar Excursion Module, engineers were thinking about how much to automate and how much control to leave in the hands of the astronauts. If they automated too much, they feared it would leave the astronauts bored and unable to intervene if needed, but if they did not automate enough, giving the astronauts manual control over too many things, they feared they would overwhelm them. *Source:* NASA.

Getting it right will not be easy. You might even call it rocket science. But as we saw with the Apollo Guidance Computer, designing the partnership between robots and people will be fundamental to the success of automation in these applications. Our lunar landings were successful because the design focus went well beyond the underlying technology to include an analysis and understanding of human psychology and decision-making, in order to create systems that would enable people and robots to collaborate seamlessly at the right times and in the right ways.[13]

The cartoon in figure 4, from the 1960s, captures the thinking at the time. NASA needed to find the right balance between automation and manual control. The designers did not want to overwhelm the astronauts with too many tasks during the lunar descent, because they might not be able to keep up and perform adequately. But they also did not want to automate everything, lest the astronauts become disengaged, and fail to intervene if and when their intervention was needed.

We have a similar conundrum with working robots. In chapters 2 and 3, we will discuss hard-earned lessons from aerospace and industrial applications that cast doubt on the vision of a working robot that does everything independently, like Rosie from *The Jetsons*. This robot ideal actually isn't realizable or even desirable. Robots don't have the same capabilities as people, and they don't think like us. This is a strength, but to understand how to fit robots with humans, we first need to understand our own human limitations and strengths—including our propensity to trust inappropriately. The user-robot partnership must be designed from the beginning with this in mind to ensure that these new social entities are effective and responsible. Just because Rosie could flip pancakes doesn't mean you should leave her unattended

to put together Thanksgiving dinner.

To complicate matters further, the world that working robots must navigate is exponentially more complex than in a cockpit or factory. Public spaces are busy and change often. New roads are built, sidewalks are closed, storefronts change. What's more, many people will come into contact with a robot who have no idea what it is doing or how it may interfere with their activities. We accommodate people we don't know every day as we pass them on the street or at the grocery store, but this kind of task is very challenging for a robot. In chapter 4, we will discuss how our robots will have to be able to develop at least a minimal awareness of bystanders and accommodate them as they move along.

Of course, designing a lunar lander is quite different from designing a consumer product. There will be lay users instead of highly trained astronauts involved, and there has to be a positive business model. In fact, as we will see in chapter 5, the commercial world is not well positioned to guide the crossover of robots to society today. The way companies design products and what consumers seem to want from them are at odds with how autonomous social systems will have to be designed. Working robots are typically able to perform particular tasks with little or no interaction with people. Consumer products are designed essentially to delight and entertain their users. Moreover, there is little or no focus today on how the design of a product affects the user's ability to perform other tasks effectively. Does Facebook care about the impact on your work productivity? Consumer products like social media platforms are designed to be fun, which often comes at the expense of being productive, transparent, and robust. And they can afford that trade-off, because the stakes involved in our interactions with them are fairly low. If Twitter goes down, nobody gets hurt. Working robots, in contrast, are there for us to offload tasks to them that *must* be done, and that people do not want to do or cannot complete safely on their own. If those robots don't work well, or distract us at exactly the wrong time, there could be material consequences. And the design model must change to focus on the best approach to ensuring successful execution of the robot's task.

In fact, research has shown that a user's preference is often in direct conflict with the design that leads to the best performance,