BESTSELLING AUTHOR OF AI SUPERPOWERS

KAI-FU LEE

AI 2041

/// TEN VISIONS FOR OUR FUTURE ///

CHEN QIUFAN

AUTHOR OF WASTE TIDE

Published in the United States by Currency, an imprint of Random House, a division of Penguin Random House LLC, New York.

Currency and its colophon are trademarks of Penguin Random House LLC.

Library of Congress Cataloging-in-Publication Data
Names: Lee, Kai-Fu, author. | Chen, Qiufan, author.
Title: AI 2041 / by Kai-Fu Lee and Chen Qiufan.
Description: First edition. | New York: Currency, [2021] | Includes index.
Identifiers: LCCN 2021012928 (print) | LCCN 2021012929 (ebook) | ISBN 9780593238295 (hardcover; acid-free paper) | ISBN 9780593238301 (ebook)
Subjects: LCSH: Artificial intelligence in literature. | Artificial intelligence.
Classification: LCC Q335 .L423 2021 (print) | LCC Q335 (ebook) | DDC 006.3—dc23
LC record available at https://lccn.loc.gov/2021012928
LC ebook record available at https://lccn.loc.gov/2021012929

International edition ISBN 9780593240717

Ebook ISBN 9780593238301

crownpublishing.com

Book design by Edwin Vazquez, adapted for ebook

Cover Design: Will Staehle

ep_prh_5.7.1_c0_r0

Contents

What we want is a machine that can learn from experience.

—Alan Turing

Any sufficiently advanced technology is indistinguishable from magic.

—Arthur C. Clarke

THE REAL STORY OF AI

> Artificial intelligence (AI) is smart software and hardware capable of performing tasks that typically require human intelligence. AI is the elucidation of the human learning process, the quantification of the human thinking process, the explication of human behavior, and the understanding of what makes intelligence possible. It is mankind's final step in the journey to understanding ourselves, and I hope to take part in this new, but promising science.

I WROTE THESE words as a starry-eyed student applying to Carnegie Mellon University's PhD program almost forty years ago. Computer scientist John McCarthy coined the term "artificial intelligence" even earlier—at the legendary Dartmouth Summer Research Project on Artificial Intelligence in the summer of 1956. To many people, AI seems like the quintessential twenty-first-century technology, but some of us were thinking about it decades ago. In the first three and a half decades of my AI journey, artificial intelligence as a field of inquiry was essentially confined to academia, with few successful commercial adaptations.

AI's practical applications once evolved slowly. In the past five years, however, AI has become the world's hottest technology. A stunning turning point came in 2016 when AlphaGo, a machine built by DeepMind engineers, defeated Lee

Sedol in a five-round Go contest known as the Google DeepMind Challenge Match. Go is a board game more complex than chess by one million trillion trillion trillion trillion times. Also, in contrast to chess, the game of Go is believed by its millions of enthusiastic fans to require true intelligence, wisdom, and Zen-like intellectual refinement. People were shocked that the AI competitor vanquished the human champion.

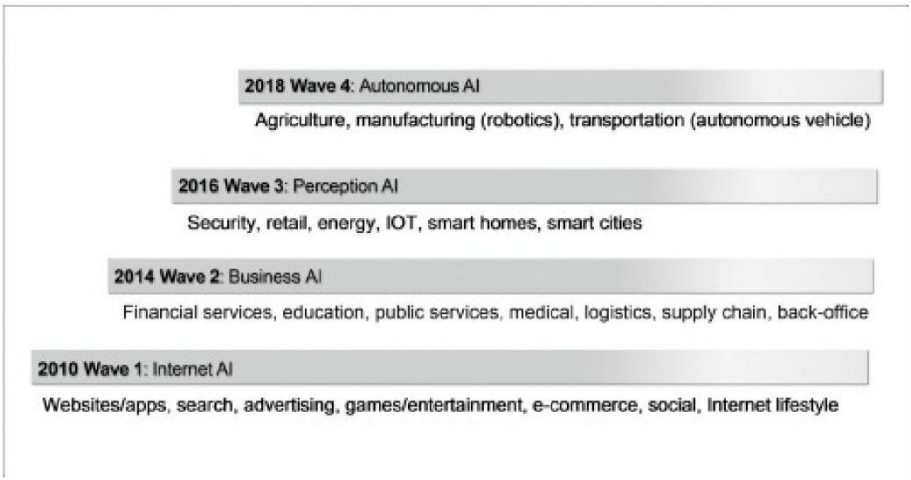
AlphaGo, like most of the commercial breakthroughs in AI, was built on deep learning, a technology that draws on large data sets to teach itself things. Deep learning was invented many years ago, but only recently has there been enough computing power to demonstrate its efficacy, and sufficient training data to achieve exceptional results. Compared to when I made my cold start in AI forty years ago, we now have about one trillion times more computing power available for AI experimentation, and storing the necessary data is fifteen million times cheaper. The applications for deep learning—and its related AI technologies—will touch nearly every aspect of our lives.

AI is now at a tipping point. It has left the ivory tower. The days of slow progress are over.

In just the past five years, AI has beaten human champions in Go, poker, and the video game Dota 2, and has become so powerful that it learns chess in four hours and plays invincibly against humans. But it's not just games that it excels at. In 2020, AI solved a fifty-year-old riddle of biology called protein folding. The technology has surpassed humans in speech and object recognition, served up "digital humans" with uncanny realism in both appearance and speech, and earned passing marks on college entrance and medical licensing exams. AI is outperforming judges in fair and consistent sentencing, and radiologists in diagnosing lung cancer, as well as powering drones that will change the future of delivery, agriculture, and warfare. Finally, AI is enabling autonomous vehicles that drive more safely on the highway than humans.

As AI continues to advance and new applications blossom, where does it all lead?

In my 2018 book *AI Superpowers: China, Silicon Valley, and the New World Order,* I addressed the proliferation of data, the "new oil" that powers AI. The United States and China are leading the AI revolution, with the United States leading research advances and China more swiftly tapping big data to introduce applications for its large population. In *AI Superpowers,* I predicted new advances, from big-data-driven decision-making to machine perception to autonomous robots and vehicles. I projected that AI's new applications in digital industries, finance, retail, and transportation would build unprecedented economic value, but also create problems related to the loss of human jobs and other issues. AI is an omni-use technology that will penetrate virtually all industries. Its effects are being felt in four waves, beginning with Internet applications, followed by applications in business (e.g. financial services), perception (think smart cities), and autonomous applications, like vehicles.



Four waves of AI applications are disrupting virtually all industries.

By the time you read this new book in late 2021 or beyond, the predictions I made in *AI Superpowers* will have largely become

reality. We must now look ahead to new frontiers. As I've traveled the world talking about AI, I'm constantly asked, "What's next?" What will happen in another five, ten, or twenty years? What will the future hold for us humans?

These are essential questions for our moment in history, and everyone working in the technology space has an opinion. Some believe that we're in the midst of an "AI bubble" that will eventually pop, or at least cool off. Those with more drastic and dystopian views believe everything from the notion that AI giants will "hijack our minds" and form a utopian new race of "human cyborgs" to the arrival of an AI-driven apocalypse. These various predictions may be born out of genuine curiosity or understandable fear, but they are usually speculative or exaggerated. They miss the complete picture.

Speculation varies wildly because AI appears complex and opaque. I've observed that people often rely on three sources to learn about it: science fiction, news, and influential people. In science fiction books and TV shows, people see depictions of robots that want to control or outsmart humans, and superintelligence turned evil. Media reports tend to focus on negative, outlying examples rather than quotidian incremental advances: autonomous vehicles killing pedestrians, technology companies using AI to influence elections, and people using AI to disseminate misinformation and deepfakes. Relying on "thought leaders" ought to be the best option, but unfortunately most who claim the title are experts in business, physics, or politics, not AI technology. Their predictions often lack scientific rigor. What makes things worse is that journalists tend to quote these leaders out of context to attract eyeballs. So, it is no wonder that the general view about AI—informed by half-truths—has turned cautious and even negative.

To be sure, aspects of AI development deserve our scrutiny and caution, but it is important to balance these concerns with exposure to the full picture and potential of this crucially

important technology. AI, like most technologies, is inherently neither good nor evil. And like most technologies, AI will eventually produce more positive than negative impacts on our society. Think about the tremendous benefits of electricity, mobile phones, and the Internet. In the course of human history, we have often been fearful of new technologies that seem poised to change the status quo. In time, these fears usually go away, and these technologies become woven into the fabric of our lives and improve our standard of living.

I believe there are many exciting applications and scenarios in which AI can profoundly enhance our society. Firstly, AI will create tremendous value to our society—PricewaterhouseCoopers estimates $15.7 trillion by 2030—which will help reduce hunger and poverty. AI will also create efficient services that will give us back our most valuable resource—time. It will take over routine tasks and liberate us to do more stimulating or challenging jobs. Lastly, humans will work symbiotically with AI, with AI performing quantitative analysis, optimization, and routine work, while we humans contribute our creativity, critical thinking, and passion. Each human's productivity will be amplified, allowing us to realize our potential. The profound contributions AI is poised to make to humanity need to be explored as deeply as its challenges.

Amid what seems like a feedback loop of negative stories about AI, I believe it's important to tell these other stories, too, and answer that question of "What happens next?" So I decided to write another book about AI. This time, I wanted to extend the horizon a bit further—to imagine the future of the world and our society in twenty years' time, or 2041. My aim is to tell the "real" AI story, in a way that is candid and balanced, but also constructive and hopeful. This book is based on *realistic AI,* or technologies that either already exist or can be reasonably expected to mature within the next twenty years. These stories offer a portrait of our world in 2041, based on technologies with

a greater than 80-percent likelihood of coming to pass in that timeframe. I may overestimate or underestimate some. But I believe this book represents a responsible and likely set of scenarios.

How can I be so confident? Over the past forty years, I have been involved in AI research and product development at Apple, Microsoft, and Google, and managed $3 billion in technology investments. So I have hands-on experience with the time and processes needed to take a technology from academic paper to pervasive product. Further, as an adviser to governments on AI strategy, I can make predictions based on my knowledge of policy and regulation frameworks, and the reasoning behind them. Also, I avoid making speculative predictions about fundamental breakthroughs and rely mostly on applying and extrapolating the future of existing technologies. Since AI has penetrated less than 10 percent of our industries, there are many opportunities to reimagine our future with AI infusion into these fields. In short, I believe that *even with few or no breakthroughs, AI is still poised to make a profound impact on our society.* And this book is my testimony.

I've been told that one of the reasons that *AI Superpowers* made an impact on readers was that it was accessible to people with no prior knowledge of AI. So when I embarked on this new book, I asked: What can I do to tell stories about AI in a way that makes them even more widely appealing? The answer, of course, was to work with a good storyteller! I decided to reach out to my former Google colleague Chen Qiufan. After Google, I started a venture capital firm. Qiufan did something more adventurous—he became an award-winning science fiction writer. I was delighted that Qiufan agreed to work with me on the project, and to dovetail his creativity with my judgment on what technology will be capable of in twenty years. We both believed that imagining the feasible technologies within a twenty-year period and embedding them in stories would be quite engaging,

and we wouldn't even have to resort to teleportation or aliens to mesmerize our readers.

Qiufan and I worked out a unique arrangement. I first created a "technology map" that projected when certain technologies would mature, how long it would take to gather data and iterate AI, and how easy it would be to build a product in various industries. I also accounted for possible externalities—challenges, regulations, and other deterrents, as well as story-worthy conflicts and dilemmas that might emerge alongside these technologies. With my input on the technological components, Qiufan then flexed his talents—dreaming up the characters, settings, and plotlines that would bring these themes to life. We worked to make each story engaging, provocative, and technologically accurate. After each one, I offer my technology analysis, digging into the forms of AI revealed and their implications for human life and society. We organized the stories to cover all key aspects of AI, and roughly ordered them from basic to advanced technologies. The sum of these parts, we hope, is a uniquely engaging and accessible primer on AI.

We named our book *AI 2041* because that is twenty years from the initial publication of this book. But it didn't slip our notice that the digits "41" happen to look a bit like "AI."

Many of our readers may love the wonderful storytelling of science fiction, but I imagine there are others who may not have picked up a novel or a collection of short stories since college. That's okay. If you fall into that camp, think of *AI 2041* not as "science fiction" but as "scientific fiction." The stories are set in wide-ranging locations around the world. In some, you may recognize a world that seems not too dissimilar from your own—with narratives that draw on existing customs and habits, albeit with an AI twist. In others, AI has transformed human life dramatically. Both AI enthusiasts and skeptics will have plenty to think about. Creating a book with a significant fiction component is inherently riskier than writing a nonfiction book

that simply describes the present and asks questions about the future. Qiufan and I sought to be bold with our narratives, and we believe the stories that follow will strike a chord with every open-minded reader whose imagination is large enough to ponder what the future holds.

The first seven stories were designed to cover technology applications for different industries in increasing technological complexity, along with their ethical and societal implications. The last three stories (plus chapter 6, "The Holy Driver") focus more on social and geopolitical issues raised by AI, such as the loss of traditional jobs, an unprecedented abundance of goods, exacerbated inequality, an autonomous weapons arms race, trade-offs between privacy and happiness, and the human pursuit of a higher purpose. These are profound changes, and humans may embrace them with compassion, exploit them with malice, capitulate to them with resignation, or be inspired by them to reinvent ourselves. In the final four stories, we decided to show four possible variations and different pathways, as a way of underscoring that the future is not yet written.

We hope the stories entertain you while deepening your understanding of AI and the challenges it poses. We also hope that the book's road map of the coming decades will help you prepare yourself to capture the opportunities and confront the challenges that the future will bring. Most of all, we hope you will agree that the tales in *AI 2041* reinforce our belief in human agency—that we are the masters of our fate, and no technological revolution will ever change that.

Now, let's take a journey to 2041.

HOW WE CAN LEARN TO STOP WORRYING AND EMBRACE THE FUTURE WITH IMAGINATION

IN AUGUST 2019, while visiting the Barbican Centre in London, I came across an exhibition titled *AI: More Than Human.* Like a refreshing summer downpour, the exhibition cleared my senses—and changed most of my preexisting biases and misconceptions toward artificial intelligence. The deceivingly simple name of the exhibition was nowhere near a sufficient representation of the diversity and complexity it contained. Each room of the exhibit revealed new wonders, all with a connection to the curators' expansive definition of what AI encompasses. There was Golem, a mythical creature in Jewish folklore; Doraemon, the well-loved Japanese anime hero; Charles Babbage's preliminary computer science experiments; AlphaGo, the program designed to challenge humans' fundamental intellect; Joy Buolamwini's analysis on the gender bias of facial recognition software; and teamLab's large-scale interactive digital art infused with Shinto philosophy and aesthetics. It was a magnificent and mind-expanding reminder of the power of interdisciplinary thinking.

According to Amara's law, "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run." Most of us tend to think of AI in narrow

terms: the murderbot from *The Terminator,* incompetent algorithms that could never match the wits or threaten the existence of humans in any way, mere soulless technological inventions that have nothing to do with how humans perceive the world, communicate emotions, manage institutions, and explore other possibilities of life.

The truth—as it has been revealed in stories ranging from the Chinese folktale of Yan Shi, the mechanic who creates a humanoid, to Talos, the bronze automaton in Greek mythology—is that humans' search for artificial intelligence has persisted throughout world history, long before computer science existed as a field or the term "AI" entered the lexicon. From the past era to the present day, the unstoppable force of AI has been revolutionizing every dimension of human civilization, and it will continue to do so.

Science fiction, my chosen field, plays a rather delicate role in investigating the human-machine paradigm. The 1818 novel *Frankenstein,* often praised as the first modern science fiction novel, hinges on questions that still resonate today: With the help of technology, are humans entitled to create intelligent life that's different from all currently existing forms of life? What would the relationship between the creation and the creator look like? The archetype of the mad scientist inflicting his creations on the world originated from Mary Shelley's masterpiece two hundred years ago.

While some may scapegoat science fiction, blaming it for people's narrow and often negative view of AI, that's only part of the story. Science fiction has the capacity to serve as a warning, but speculative storytelling also has a unique ability to transcend time-space limitations, connect technology and humanities, blur the boundary between fiction and reality, and spark empathy and deep thinking within its reader. Historian and bestselling author Yuval Noah Harari has called science fiction "the most important artistic genre" of our time.

That's a high bar to live up to. For science fiction writers like myself, the challenge we face is creating stories that not only reveal hidden truths about our present-day reality, but also, simultaneously, project even wilder imaginative possibilities.

Therefore, when my former colleague from Google Kai-Fu Lee got in touch with me and proposed this collaboration on *AI 2041*—a one-of-a-kind book project that combines science fiction and analysis of big ideas that animate technology—I was thrilled. The Kai-Fu I know is a pioneering global leader, a savvy and trend-making business investor, and an imaginative, open-minded prophet of tech. His notions on career development in his field have influenced a generation of young people. Now, his mind is set on the future.

Equipped with a profound understanding of cutting-edge research and its applications in the business world, Kai-Fu delineates the ways in which AI could change human society in twenty years in areas ranging from medicine and education to entertainment, employment, and finance. His idea for this project was ambitious, but it was also a kind of magical coincidence. Years earlier, in my own writing, I had developed the notion of "science fiction realism." To me, science fiction is fascinating because it not only generates an imaginative space for escapists to leave behind their mundane lives, play the role of superheroes, and freely explore galaxies far, far away, but it also provides a precious opportunity for them to temporarily remove themselves from everyday reality and critically reflect upon it. By imagining the future through science fiction, we can even step in, make change, and actively play a role in shaping our reality.

In other words, with every future we wish to create, we must first learn to imagine it.

My imagination began developing as a child thanks to classic works of science fiction like *Star Wars, Star Trek,* and *2001: A Space Odyssey.* Since I was ten, these works have been my portal to the

vast beyond and worlds unknown. I believe that, before setting pen to paper for each story, the key is always to orient the story in the history of its genre and a greater social context. As someone deeply invested in—even obsessed with—the fantasies of science fiction, I am in awe of how inclusive the spectrum of science fiction storytelling is. Almost any theme or style can find its place in the genre.

Before I became a full-time author, I worked in technology. A lot of people would assume engineers and computer science wizards might have little interest in fiction—because their brains are hardwired for science, as opposed to literature. But during my more than ten years working in tech, I encountered many engineers and technologists who were not-so-secret fans of speculative fiction. This enthusiasm sometimes manifested in meeting rooms with names like "Enterprise" or "Neuromancer," but it also was present among the formidable minds behind projects like Google X and Hyperloop. From the modern submarine to the laser gun, and from mobile phones to CRISPR, scientists will readily admit they got direct inspirations from fiction. Imagination indeed shapes the world.

From the beginning, I decided that *AI 2041* would challenge the stereotype of the dystopian AI narrative—the kind of tale where the future is irrevocably bleak. Without disregarding AI's faults or nuances, Kai-Fu and I endeavored to portray a future where AI technology could influence individuals and societies positively. We wished to imagine a future that we would like to live in—and to shape. We imagined a future where the next generations could enjoy the benefits of technological development, work to bring more achievement and meaning into the world, and live happily.

The path to imagining the future of our dreams was not always an easy one. Our challenge was to become immersed in the latest AI research and then to project, with science and logic, *realistically,* how the AI scene would appear in twenty years. Kai-

Fu and our team spent hours studying recently published research papers, conversing with experts, professionals, and thinkers involved with the AI industry, participating in the AI workshop hosted by the World Economic Forum, and visiting top AI tech companies, in order to ensure we had a comprehensive grasp of the technological and philosophical basis of AI development.

The other challenge was imagining the human future. We wished to represent how individuals from disparate cultures and industries and with different identities would react to the future shock induced by AI. Subtle psychological details are difficult to infer through mere logic and rationalization. To help fill in the emotional portrait of the characters in our stories, we looked to history and drew inspiration from similar world-changing events that have occurred in the past. To stimulate our readers' imagination and capacity to conceptualize alternative human conditions, we knew our stories must also spark empathy if we were to fully convey our vision and sentiment. Kai-Fu's analysis serves as the string that connects the soaring kite of imagination to the graspable reel of reality.

After months of intensive work and rounds of polishing, here are the ten portals we have assembled that will transport you to the time-space of 2041. We hope that you will embark on this journey with curiosity, an open mind—and an open heart, too.

One last thing: For me, the greatest value of science fiction is not providing answers, but rather raising questions. After you close the book, our hope is that lots of new questions will enliven your mind: For instance, can AI help humans prevent the next global pandemic by eliminating it at the very root? How can we deal with future job challenges? How can we maintain cultural diversity in a world dominated by machines? How can we teach our children to live in a society where humans and machines coexist? We hope our readers' questions will help take us further down the path as we shape a happier and brighter future.

Welcome to 2041!

CHAPTER ONE

THE GOLDEN ELEPHANT

STORY TRANSLATED BY BLAKE STONE-BANKS

IT IS BETTER TO LIVE YOUR OWN DESTINY IMPERFECTLY THAN TO IMITATE SOMEBODY ELSE'S PERFECTLY.

—BHAGAVAD GITA (भगवद्गीता, SONG OF GOD OR HINDU SCRIPTURE), CHAPTER 3, VERSE 35

NOTE FROM KAI-FU: The opening story takes readers to Mumbai, where we meet a family who has signed up for a deep-learning-enabled insurance program. This dynamic insurance program engages with the insured in the form of a series of apps intended to better their lives. The family's teenage daughter, however, finds that the AI program's persuasive nudges complicate her search for love. "The Golden Elephant" introduces the basics of AI and deep learning, offering a sense of its main strengths and weaknesses. In particular, the story illustrates how AI can single-mindedly try to optimize certain goals, but sometimes create detrimental externalities. The story also suggests the risks when one company possesses so much data from its users. In my commentary at the end of the chapter, I will explore these issues, offering a brief history of AI and why it excites many but has become a source of distrust for others.

ON THE SCREEN, the three-story statue of Ganesh swayed in the surf of Chowpatty Beach as though synced to the sitar soundtrack. With each wave, the towering idol descended lower until it was engulfed by the Arabian Sea. In the salty brine, the statue dissolved into gold and burgundy foam, washing onto Chowpatty Beach, where the colors clung like blessings to the legions of believers who had gathered for the Visarjan immersion ritual celebrating the end of the Ganesh Chaturthi festival.

In her family's Mumbai apartment, Nayana watched as her grandparents clapped their hands and sang along to the TV. Her younger brother, Rohan, took a mouthful of cassava chips and a deep swig from his diet cola. Though he was only eight, Rohan was under doctor's orders to strictly control his fat and sugar intake. As he wagged his head in excitement, crumbs sprayed from his mouth and flew across the floor. In the kitchen, Papa Sanjay and Mama Riya banged on pots and crooned like they were in a Bollywood film.

Nayana tried to shut them all out of her mind. The tenth-grader was instead focusing all her energy on her smartstream, where she had downloaded FateLeaf. The new app was all Nayana's classmates could seem to talk about lately. It was said to possess the answer to almost any question, thanks to the prescience of India's greatest fortune tellers.

The app—its branding and ad campaign made clear—was inspired by the Hindu sage Agastya, who was said to have engraved the past, present, and future lives of all people in Sanskrit onto palm leaves, so-called Nadi leaves, thousands of years ago.

According to the legend, simply by providing one's thumbprints and birthdate to a Nadi leaf fortune teller, a person could have their life story foretold from the corresponding leaf. The problem was that many leaves had been lost to meddling colonialists, war, and time. In 2025, a tech company tracked down and scanned all the known Nadi leaves still in circulation. The company used AI to perform deep learning, auto-translation, and analysis of the remaining leaves. The result was the creation of virtual Nadi leaves, stored in the cloud—one for each of the 8.7 billion people on Earth.

Nayana was not dwelling on the ancient history of the Nadi leaves. She had a more pressing matter on her mind. Users of the FateLeaf app could seek to uncover the wisdom of their Nadi leaf by posing various questions. While her family watched the Ganesh Visarjan celebration on TV, Nayana nervously typed out a question within the app: "Does Sahej like me?" Before she clicked "Send," a notification popped up indicating that an answer to her question would cost two hundred rupees. Nayana clicked "Submit."

Nayana had liked Sahej from the moment his stream first connected in their virtual classroom. Her new classmate didn't use any filter or AR background. Behind Sahej, hanging on the wall, Nayana could see rows of colorful masks, which, she learned, Sahej had carved and painted himself. On the first day of the new term, the teacher had asked Sahej about the masks, and the new student shyly gave a show-and-tell, explaining how the masks combined Indian gods and spirits with the powers of superheroes.

Now, in an invitation-only room on her ShareChat, some of Nayana's classmates were gossiping about Sahej. From the way his room was furnished to the fact that his surname was hidden from public view in school records, these girls were certain Sahej was among the "vulnerable group" that the government mandated make up at least 15 percent of their student body. At private schools across India, such children were practically guaranteed spots and their tuition, books, and uniforms were

covered by scholarships. "Fifteen percent" and "vulnerable group" were euphemisms for the Dalits.

From documentaries she had watched online, Nayana knew about India's old caste system, which was deeply embedded in Hindu religious and cultural beliefs. A person's caste had once determined one's profession, education, spouse—their whole life. At the bottom rung of this system were the Dalits, or, as they were sometimes referred to with derision, "untouchables." For generations, members of this community were forced to do the dirtiest jobs: cleaning sewers, handling the corpses of dead animals, and tanning leather.

The constitution of India, ratified in 1950, outlawed discrimination based on caste. But for years following independence, Dalit areas for drinking, dining, residing, and even burial were kept separate from those of groups considered higher in the system. Members of the higher castes might even refuse to be in the same room as the Dalits, even if they were classmates or colleagues.

In the 2010s, the Indian government sought to correct these injustices by establishing a 15-percent quota for Dalit representation in government positions and in schools. The well-intentioned policy had sparked controversy and even violence. Higher-caste parents complained that such admissions weren't based on academic performance. They argued that their children were paying the price for previous generations' sins and that India was just trading one form of inequality for another.

Despite these pockets of backlash, the government's efforts seemed to be working. The 200 million descendants of Dalits were integrating into mainstream society. It had become more difficult to recognize their past identity at a glance.

—

THE GIRLS IN NAYANA'S ShareChat couldn't stop talking about the new boy in school, Sahej, debating his background—but

also whether they would consider going out with him.

You shallow snobs, Nayana silently huffed.

For her part, Nayana saw in Sahej a kindred artistic spirit. Inspired by Bharti Kher, Nayana dreamed of becoming a performance artist, and she often had to explain that this was nothing like being a superficial pop entertainer. She believed great artists had to be brutally honest about their innermost feelings and should never accept the perspectives of others. If she liked Sahej, then she liked Sahej—no matter his family background, where he lived, or even his Tamil-accented Hindi.

The question Nayana had posed to the FateLeaf app seemed to take forever to process. Finally, a notification popped up on Nayana's smartstream accompanied by a palm leaf icon: "What a pity! Due to insufficient data provided, FateLeaf cannot currently answer your query."

The clink of Nayana's refund vibrated from her smartstream.

"Insufficient data!" Nayana silently cursed at the app.

Annoyed, she finally raised her head from her screen to notice her mother, Riya, putting the finishing touches on dinner. Something was off. In addition to a number of Indian holiday delights, Nayana saw several super-expensive dishes from a Chinese delivery place on the table. Such treats were rare for her penny-pinching father. But there was something even more unusual: Riya was wearing her favorite pure silk Parsi-style sari. She had her hair up and was wearing a complete set of jewelry. Even Nayana's grandparents seemed different—happier than usual—and for once, her fat brother, Rohan, wasn't pestering her with all kinds of stupid questions.

The Ganesh Chaturthi festival couldn't explain all this.

"So, is anyone going to tell me what's going on?" Nayana said as she stared at the spread on the table.

"What do you mean, what's going on?" Riya shot back.

"Am I the only one who thinks all this is a bit out of the ordinary?"

Nayana's parents glanced at each other for a second then burst out laughing.

"Take a look and tell us what's different," Riya said.

Nayana felt like she was about to lose her mind. "What are you hiding from me?"

"My sweet little girl, eat first." Grandmom began to pull apart the naan.

"Wait. Did Dad get promoted? Did we win the lottery? Did the government cut taxes?"

Dad wobbled his head back and forth. "All beautiful ideas. But no. It's all for your mother—"

Nayana spun toward her mother. "Mom, what did you buy this time?"

"Your tone should be more respectful when talking with your elders," Riya chided.

"It wasn't me who got taken to the cleaners for buying cheap..." Nayana's voice trailed off into a sigh.

Nayana exhaled. "And what exactly is it you bought?"

"Ganesh Insurance! They had an amazing sale for the holiday. First time ever that GI was fifty percent off! All the neighbors got it, too, and they're even more thrifty than me."

Dad clapped his hands in excitement. So did Nayana's grandparents.

"Wait! Hasn't our family always had a policy from the Life Insurance Corporation of India?"

"That policy wasn't nearly enough! Your grandparents are old and rely on us. What if something were to happen to us? Where will the money come from? We have to save where we can. And you and your younger brother are both in private school, and don't you still want to go to SOFT at Rai University? Tuition and dormitories cost a lot more than public universities in Mumbai."

"Why does the conversation always have to twist back to blaming me?"

"To plan for the future, you must also think about what's in front of you," Grandfather observed.

"So, what's the deal with this insurance exactly?"

"Well, Mrs. Shah from next door filled me in," explained Riya. "It's a platform that uses AI to adjust the insurance plan according to the family's needs. And for a very good price. And it's not just the one platform, more like a little family of apps. There's one for calculating and paying insurance fees, and one for investments, and my favorite one is the home goods shop. Another one shows you deals in your area. And just look at my hair. The salon the Cheapon deals app recommended only cost four hundred rupees."

Just as Rohan was about to steal a sweet, Nayana slapped the back of his hand, which he withdrew with a sheepish look.

"You sound like an advertisement," Nayana told her mother. "Why would an insurance company tell you where to get your hair done? And how exactly does this AI insurance know so much about our family?"

"This, well..." Mom searched for a way out of the question. "To get the benefits of Ganesh Insurance, we share data link access for each member of the family."

"What?" Nayana's eyes grew as wide as brass bells.

"It's all kept strictly confidential, unless we give permission for GI to use it."

"What right do you have to share my data link with some insurance company!"

"Hey, don't talk to your mother like that." Dad wagged his finger at Nayana. "Don't forget you're still a minor. As your parents, we have the right to make data decisions for you."

Nayana's face turned bright red; she was unaccustomed to such a sharp rebuke from her father. She threw her knife and fork down onto the plate and raced back to her room. She grabbed her quilt and pulled it over her head, imagining that somewhere on her Nadi leaf, it was written that today was the worst day of her life.

—

NAYANA AND HER MOTHER'S cold war lasted a week, until Nayana's smartstream began pushing some unusual new notifications:

IT'S GOING TO RAIN TODAY
SO TAKE AN UMBRELLA.
RESPIRATORY ILLNESSES ARE BECOMING MORE
PREVALENT, SO YOU SHOULD WEAR A MASK.
THERE'S A TRAFFIC ACCIDENT ON YOUR ROUTE
SO TO AVOID THE CONGESTION…

At first, Nayana was skeptical about the endless stream of notifications. But she found she couldn't stop reading them. From time to time, she actually got a useful tip. A clothing deal, a discount at a lunch place she liked...Of course, to actually redeem the deals, Nayana had to install the Cheapon deals app and other various golden-elephant-branded Ganesh Insurance apps on her smartstream and permit them to access her data.

It seemed Mom had already forced the golden elephant onto all the family's smartstreams. Women controlled data sharing in more than 60 percent of Indian households. All that personal data was linked to the national ID Aadhaar card and the unique identifying number issued to all of India's 1.4 billion residents by the Unique Identification Authority of India. Since implementing the system in 2009, after twenty years of development, the government had collected data including citizens' fingerprints, retina signatures, genetic histories, family information, occupations, credit scores, home-buying history, and tax records. With its clients' consent, Ganesh Insurance was able to tap into this rich trove of data to personalize its services.

Of course, there were some privacy restrictions. For example, social media data needed to be separately authorized, and use of minors' data required consent of their legal guardians.

Nayana aimed to be vigilant in every interaction with GI. In her data literacy class in high school, she had learned that on the Internet every click might sell you out. She carefully studied

the fine print before choosing between "I accept" or "I need more time to consider." Yet it seemed every time she selected "I need more time to consider," GI would send appealing new discounts and suggestions for how to solve her immediate problems.

For example, how exactly could she attract Sahej's attention?

Sahej was really cute, especially his sheeplike eyes. He instinctively wanted to please every classmate. He had even sent each classmate a wood carving he had made of a small animal head. But the virtual classroom had its limits. Sometimes all Nayana could see was just a blurry headshot icon and a glitchy voice due to Sahej's poor connection. After finally meeting Sahej during one of the school's "in-person" days, Nayana found it even more difficult to contain her feelings for him. She sought any excuse to talk with him. But for some reason, the boy kept his distance.

Does Sahej not like me? Or is it another reason?

Could Sahej's background, Nayana wondered, account for his shyness around her?

—

AS THE QUESTION LINGERED in Nayana's mind, little golden elephants popped up in a notification from MagiComb, GI's lifestyle advice app, about "how to make yourself more attractive to guys." Nayana guessed that the AI was able to use her online browsing and shopping data to infer what she was thinking, but these recommendations disturbed Nayana for another reason. Why should women need to change themselves to win a man's favor? Why couldn't women show men who they really were and see if they were or weren't a match?

Though she was still feeling annoyed at her mother, Nayana decided to ask her about the golden elephant's odd messages.

"Silly girl, machines only learn what is taught to them by human beings." Riya looked at her newly bought long skirt in the

mirror and turned. "But what's this all about? Have you met someone?"

"Not at all," Nayana replied, with a tinge of guilt.

"You can hide it from me, but not from the AI," her mother joked. "Are you sure you don't want me to help you scheme? You know, your mother knows a thing or two about men."

"I just don't know how I can find out what he really thinks of me. I give him likes online, but he never seems to respond."

"Ahh, so there *is* someone! It's not enough to give someone a like online. You've got to have guts. And that reminds me. If you permit GI to access your ShareChat account data, its recommendations will be better. Not to mention, the premium for our family will also dip just a bit more."

Nayana shook her head and left the room. She recalled that only a few weeks earlier, her mother had rejected Nayana's request to share her data link with a different app—FateLeaf—to obtain more accurate fortune-telling. Now, their positions were reversed. Of course, now money was on the line.

It wasn't just her mother. To Nayana, everyone in the family had been brainwashed by that little golden elephant. They had become hyperaware that any change in behavior might raise or lower their premiums. Once something was linked with money, it seemed to Nayana that the human brain went on autopilot. They'd do whatever it took to score an award and evade a penalty.

It wasn't that GI didn't have its plus side. The little golden elephant would remind Nayana's grandparents to take their medicine and nudge them to schedule doctor appointments. Even Nayana's father, who had never listened to anyone, gave up smoking when the little golden elephant kept chiding him. He swapped his favorite arrack for a healthier single nightly glass of red wine. His driving style even became more restrained. At the app's urging, he no longer zigzagged through Mumbai's congested streets like an out-of-work race car driver. GI had given him an incentive—by changing his behavior, he was able to lower his auto, health, and life insurance premiums.

If anyone in the family could resist the GI app's nudges, Nayana suspected it would be her brother, Rohan. After all, fat and sugar were as addictive as heroin, especially for children with no self-control. But that golden elephant made it happen. Even if the eight-year-old didn't understand insurance premiums or delayed gratification, the rest of the family were conditioned to see any sweet near the boy as a threat to their bank account. Their former indulgence of Rohan's sweet tooth was over.

It naturally made sense. Insurance companies wanted people to live healthier, longer lives—it made for better profits.

As for herself, Nayana was still on the fence. Should she hand over the data link of her ShareChat?

Equally puzzling was the question of Sahej. When Sahej had given everyone in the class a handmade wood carving, he chose a crow's head covered in patterns to give to Nayana. The tenth-grader practically tore her hair out thinking about what hidden meaning the gift might hold.

Doesn't the crow symbolize bad luck? Is he telling me to not be so loud and annoying? Am I coming on too strong? What's he saying exactly?

Nayana tortured herself with such questions. Her first thought was to turn to FateLeaf for a divination, but her mother had forbidden her to permit that app to access her data. *What about MagiComb?* Nayana wondered. Lovesick, she decided she would see what the elephant's omnipotent algorithm had to say about her future.

The future the little golden elephant imagined, however, wasn't anything like the one she'd hoped for.

—

EVERYTHING FELT IMMEDIATELY WRONG.

Granting GI access to her data on ShareChat, Nayana knew from her data literacy class, was like opening the door to your bedroom. Your whole private life might be visible at a glance. Although GI guaranteed that all data was fed anonymously to its

AI for purposes of federated learning and that no third party could access it, Nayana thought that sounded a bit like a farmer telling the turkey a week before Thanksgiving, "Hey, you're safe here."

Whenever she was browsing, chatting, liking, or even selecting emojis on ShareChat, all Nayana could think about now was how her choices would affect the family's insurance premiums. She found the whole system infuriating and ridiculous.

But perhaps, she wondered, *it's even more ridiculous to expect this AI to act as my matchmaker.*

Sahej posted almost nothing on ShareChat. He was like a person from some bygone era who had failed to keep up with technology. He occasionally posted a news article, quotes he liked, or outdated memes. But his usage was sporadic and unpredictable. His account, Nayana thought, looked like a fake zombie account.

The AI was supposed to help Nayana get together with Sahej, but how could it learn anything important about Sahej from his boring account? Meanwhile, it was all too easy for AI to understand Nayana's intentions, given her incessant clicking. To the AI, such things were a matter of math, not love.

To Nayana, something fishy was going on with GI when it came to Sahej. She found it odd that every time she refreshed Sahej's page or liked one of his posts, GI would send her a weird notification, as if trying to break her focus. If she tried to come up with a reason to talk with him, browsed online for a gift for him, or even just thought about inviting him out to coffee, that little golden elephant would pop up with some totally ridiculous recommendation or seemingly load a page in error.

The only possible explanation Nayana could think of was that the little golden elephant didn't want her to get close with Sahej at all. It was actively working against her.

Was the elephant like this with everyone? Is it because I'm too young? But isn't coupling up and marriage a good thing? Aren't we told that as a

country of 1.4 billion, our reproductive capacity will make us invincible on the world stage? What's the problem?

As her thoughts spiraled, Nayana noticed her mother watching her from the doorway.

"What in the hell have you been up to, young lady? Our premium is going through the roof!"

"Me?" Nayana didn't know what to say. It was clear to her the little golden elephant was determined to turn her whole virtual world upside down.

"Tell me now, or I'm taking your smartstream away!"

"No, you can't!"

"Sorry, but yes, that's exactly what I'm going to—"

Before Riya could finish, Nayana shot up, rushed past her mother, and raced out of the house as fast as her legs would carry her.

Clenching her smartstream, Nayana ran until she no longer recognized where she was. Finally, she saw the familiar relief sculptures of the New India Assurance Building in the Fort district. Sunset lit the weathered façade's artfully sculpted farmers, potters, spinners, and porters as Nayana decided it was the perfect moment to give Sahej a call, no matter how much it raised her family's premium.

The boy's avatar image popped onto her smartstream as the screen flickered with GI notices. Nayana could see that her family's premium had already increased by 0.73 rupees. It was a long time before he answered, and Nayana was about to give up when the phone finally connected with a videostream so dark she could barely make out the contours of a face and white-toothed grin.

"That you, Sahej?" Nayana asked timidly.

"It's me. Nayana?"

"I was afraid you weren't going to answer."

"Ermmm...it's a bit complicated. I can't speak long. But I really do want to talk with you."

"Me, too." Nayana's heart jumped. "I'm going to give you the address of a restaurant. Can we meet there?"

Sahej glanced about in silence for a moment before finally whispering, "Okay."

After hanging up, Nayana couldn't help but cheer.

Then someone called her name. She spun to see her mother shining gold and red against the setting sun, as though the goddess Saraswati had come down to Earth.

"How did you find me?"

"I'm the data manager of our house, and don't you forget it!" Her mother glared at her.

"I'm sorry." Nayana didn't dare look into her mother's eyes. "But, remember I told you about that guy? I'm going to meet up with him. But GI won't allow it, so..."

"You think that's why the premium's going up? GI wants to keep us living healthier and longer—and prevent us from doing stupid things that will harm us, not...Unless this is some kind of dangerous person?"

Nayana shook her head. "No, he's just my new classmate, Sahej. He's smart, a real talent. And this is the present he made for me. He carved it himself."

Her mother inspected the wooden crow head Nayana handed her.

"He doesn't sound like such a dangerous person. Is he handsome?"

Nayana let slip a shy smile, but it quickly turned into a grimace. "This sucks. What does GI know that I don't? Maybe I'll live longer if I never meet up with him."

"Sweetie, let me tell you something." Nayana's mother draped an arm over her daughter's shoulder. "I know we don't always see eye to eye. But I'm not as blind as you may think! You know, talking to you makes me think about something I read recently. Actually—it was an old ebook suggested by the MagiComb, come to think of it."

"What was it?" Nayana became curious.

"It was a book from 2021, and there was a story in there about a mom who is so superficial and proud and obsessed with her

noticed Sahej was careful to keep his distance from her, as though she carried a dangerous electric current.

"Sahej, why? Why can't we get close to each other?" Nayana chose her words carefully.

Now it was Sahej's turn to look surprised. "Nayana, do you really not know?"

"Know what?"

"My last name."

"The schools and virtual classroom keep your surname protected just like you're the offspring of a big star or some famous family."

"On the contrary, it's because they don't want it to trigger any discomfort."

"What kind of discomfort?"

"In the past, it was described as a feeling of being *polluted*."

"You're talking about your caste? But that whole system was outlawed years ago."

Sahej gave a bitter laugh. "Just because it's no longer permitted by law and doesn't appear in the news doesn't mean it's gone."

"But how would the AI know about it?"

"The AI doesn't know. The AI doesn't need to know the definition of the castes. All it needs is its users' history. No matter how we hide or if we change our surnames, our data is a shadow. And no one can escape their shadow."

Nayana thought about what her mother had said, that AI only learns what humans teach it. She rolled the thought about in her head, then looked at Sahej. "So you're saying that AI identifies the invisible discrimination in our society and quantifies it."

Sahej's expression became serious, but he exhaled a soft laugh. "I almost forgot. There's also the color of my skin. The Sanskrit word *vārna* once meant both *caste* and *color*."

"It's all so absurd!"

"No, it's reality. And in reality, women of low caste can date and marry men of higher caste. But the other way around will

never be accepted. The reputation of the girl's family would be damaged."

"But does the AI really care about those things?"

"Sure, AI doesn't care about our old social mores. It only cares about how to reduce the premium as much as possible, and that's why GI wants to stop us from being together."

When she heard Sahej say "together," Nayana's ears felt hot.

"Objective function maximization."

"What?"

"Humans give the AI its objective, which here is to decrease insurance premiums to the lowest cost possible. Then the AI does everything possible to achieve that goal. The AI won't consider anything at all beyond those factors, certainly not whether or not we're happy. Machines aren't smart enough to interpret all the feelings going on behind the data. Plus, these injustices and biases are still real. All AI does is lift that veil of shame."

"Why do you know so much about it?"

Sahej gave a little smile. "Because I want to go to Imperial College to become an AI engineer, so I can help change it."

They reached the crossroads near Nayana's home, and Sahej paused to prepare his goodbye.

"But why can't we change it now?" Nayana said. "Are we so ready to let AI arrange our fate? Like those predictions on FateLeaf that were written thousands of years ago?"

A strange expression emerged on Sahej's face. "Have you opened FateLeaf since connecting to GI?"

"Ugh, I'm so sick of that little golden elephant. What does it have to do with FateLeaf?"

"FateLeaf is in the GI family of apps, just like MagiComb and Cheapon. If you accept the data-sharing terms, you'll get more accurate fortunes."

"Of course! How did I not figure this out before? So the so-called fates of the Nadi leaves aren't real after all. I guess, like everyone else, I wanted it to be real—and for it to tell me what I wanted to hear." Nayana didn't know whether to rejoice or feel cheated.

Sahej looked at the girl before him. He paused, then pointed at the street he was going to take home.

"This road leads to where my family lives. It passes through the Dharavi construction site. There used to be more than a million people crowded into that 2.4-square-kilometer slum. Tourists visited to take photos, but not one ever wanted to stay. The government's finally transforming it into a community suitable for ordinary citizens. But I promise you, if you ever get close to Dharavi, your GI will flood with illness alerts, or warnings not to drink the water. The app will implore you to stay away. Nayana, I appreciate your sense of justice, but that path just isn't for people like you. The world is on your side, not that side. If we're going to talk about fate, that's what our fate is."

"Take me there." Nayana was startled by how quickly the words left her mouth, but she stepped forward nonetheless. "I want to prove I'm not that person you're thinking of."

Sahej tilted his head. "You sure?"

Nayana glanced at the road stretching into a forbidden hollow at the heart of Mumbai. She was afraid, but she remembered what her mother had told her before saying goodbye: *Some risks are worth taking.*

Sahej smiled, bent his arms, and gestured her forward in a gentleman's bow. "As you please."

The young couple made their way deeper into the ancient city, where centuries of renovation and innovation had branded every corner. Towers old and new lined their path like reincarnated souls. Of course, these souls, too, would eventually be broken up and reconstituted by tomorrow's machine gods.

"So, will you now finally tell me why on earth you made me a crow's head?"

"My astrological animal is the crow, though perhaps I'm more socially awkward than most crows."

"It's that simple?"

"It's that simple."

Nayana's smartstream vibrated with increasing frequency. She knew every vibration was an alert from that little golden elephant trying to save her, warning her to walk away from what was once the world's largest slum, incentivizing her to turn her back on its poverty, disease, discrimination, and *untouchables,* like the boy next to her.

She pulled her collar tight and continued forward at his side.

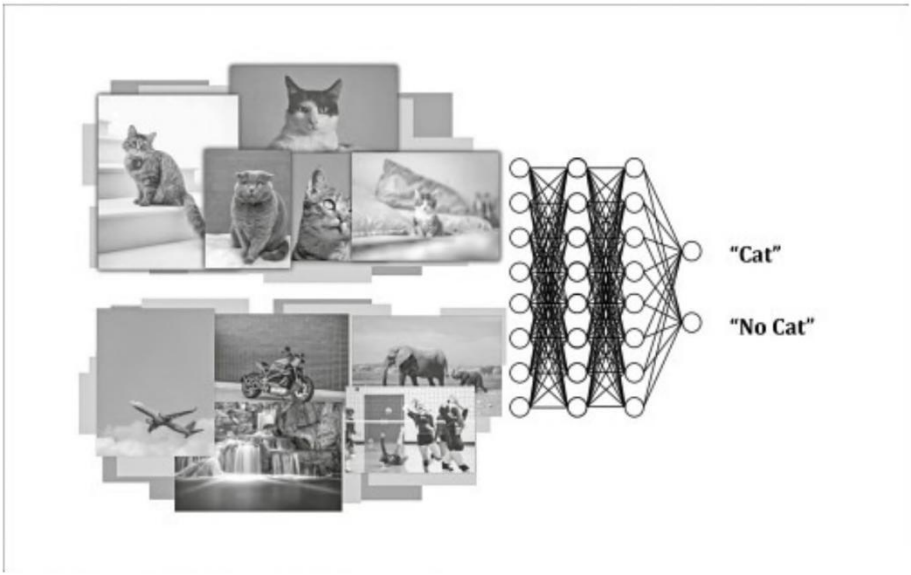
In the dark ancient streets ahead, an answer was waiting.

ANALYSIS

DEEP LEARNING, BIG DATA, INTERNET/FINANCE APPLICATIONS, AI EXTERNALITIES

The benefits of Ganesh Insurance—powered by deep learning AI—are clear in "The Golden Elephant." Nayana's mom, Riya, saves money thanks to the program's deals app. Her father, Sanjay, quits smoking and drives more safely. Even her brother is eating healthier, after AI raises an alarm about the potential for him to develop diabetes. Such a suite of apps running on the smartstream (mobile phone of 2041), marked by personalized nudges toward better health and well-being, could help people live longer, healthier, and wealthier lives. So, is there a catch? That question about trade-offs lies at the heart of "The Golden Elephant," which introduces the foundational AI concept of deep leaning.

Deep learning is a recent AI breakthrough. Among the many subfields of AI, machine learning is the field that has produced the most successful applications, and within machine learning, the biggest advance is "deep learning"—so much so that the terms "AI," "machine learning," and "deep learning" are sometimes used interchangeably (if imprecisely). Deep learning supercharged excitement in AI in 2016 when it powered AlphaGo's stunning victory over a human competitor in Go, Asia's most popular intellectual board game. After that headline-grabbing turn, deep learning became a prominent part of most commercial AI applications, and it is featured in most of the stories in *AI 2041.*

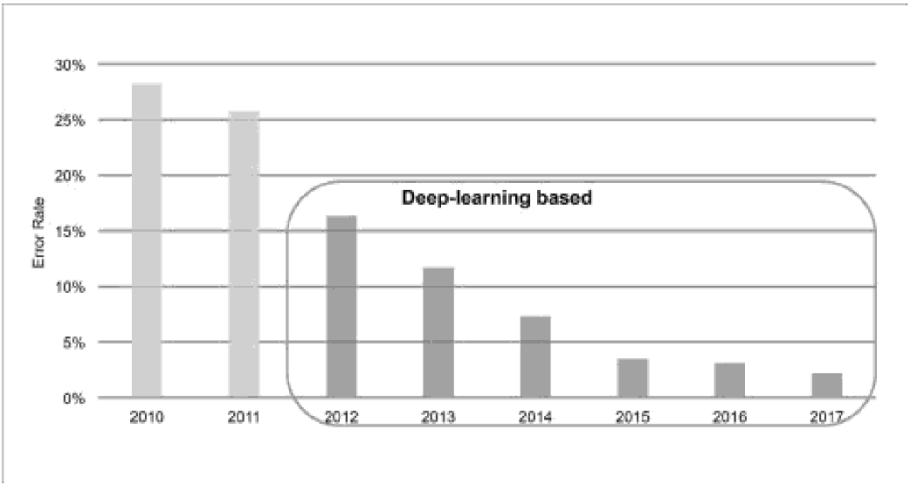


Deep learning neural network trained to recognize pictures of cats vs. pictures with no cats.

Deep learning is an omni-use technology, meaning it could be applied to almost any domain for recognition, prediction, classification, decision-making, or synthesis. Take insurance, the prime example in "The Golden Elephant." The deep learning powering Ganesh Insurance's apps has been trained to determine the likelihood that each insured may develop serious health problems, and then set premiums accordingly.

To train a network to separate those who likely face serious health claims from those who do not, AI would learn from training data comprising all past insurance applicants and their medical claims and family information. Each case would be labeled with "filed serious health claim" or "did not file serious health claim" in the output layer. Having absorbed this trove of data in the training process, the AI could infer the likelihood any new application would lead to a serious health claim, and decide whether to approve the insurance application or not, and if so, how much the premium should be. Note that, in this

scenario, no human would ever need to label an applicant as a health risk or not. Instead, the labels are based solely on "ground truth" (for example, whether each insurer had filed a serious health claim).



Deep learning led to dramatically lower computer-vision object recognition rates.

DEEP LEARNING: AMAZING CAPABILITIES BUT WITH LIMITATIONS

The first academic paper describing deep learning dates all the way back to 1967. It took almost fifty years for this technology to blossom. The reason it took so long is that deep learning requires large amounts of data and computing power for training the artificial neural network. If computing power is the engine of AI, data is the fuel. Only in the last decade has computing become fast enough and data sufficiently plentiful. Today, your smartphone holds millions of times more processing power than the NASA computers that sent Neil Armstrong to the

moon in 1969. Similarly, the Internet of 2020 is almost one trillion times larger than the Internet of 1995.

While deep learning was inspired by the human brain, the two work very differently. Deep learning requires much more data than humans, but once trained on big data, it will outperform humans by far for a given task, especially in dealing with quantitative optimization (like picking an ad to maximize likelihood of purchase, or recognizing a face out of a million possible faces). While humans are limited in the number of things they can pay attention to at once, a deep-learning algorithm trained on an ocean of information will discover correlations between obscure features of the data that are too subtle or complex for we humans to comprehend, and which may not even be noticed.

Furthermore, when trained on a huge amount of data, deep learning can customize for individual users, based on that user's patterns as well as similar patterns observed on other users. For example, when you visit Amazon, the website's AI highlights specific products intended to entice you and maximize your spending. And when you open a Facebook page, Facebook shows you content designed to maximize the number of minutes you will stay on Facebook. Amazon and Facebook's AI are targeted, meaning that they show different personalized content to each person. So the content shown to me works great on me, but may not work at all on you. This targeted accuracy is much more effective at producing clicks and purchases than the one-size-fits-all approach used by traditional static websites.

As powerful as it is, deep learning is not a panacea. While humans lack AI's ability to analyze huge numbers of data points at the same time, people have a unique ability to draw on experience, abstract concepts, and common sense to make decisions. By contrast, in order for deep learning to function well, the following are required: massive amounts of relevant data, a narrow domain, and a concrete objective function to

optimize. If you're short on any one of these, things may fall apart. Too little data? The algorithm won't have enough examples to uncover meaningful correlations. Multiple domains? The algorithm cannot account for cross-domain correlations and will not get enough data to cover all permutations. Too broad an objective function? The algorithm will lack clear guidance to sharpen its optimization.

It is important to understand that the "AI brain" (deep learning) works very differently from the human brain. Table 1 illustrates the key differences:

	Human Brain	AI Brain (Deep Learning)
Data required to learn	Few data points	Huge amount of data
Quantitative optimizing and matching (picking a face out of a million)	Hard	Easy
Customizing for each situation (showing each user a different product to maximize purchasing)	Hard	Easy
Abstract concepts, analytical reasoning, inferences, common sense, and insight	Easy	Hard
Creativity	Easy	Hard

Table 1: Strengths and Weaknesses of Human vs. AI "Thinking"

APPLYING DEEP LEARNING TO INTERNET AND FINANCE

Given the strengths and weaknesses of deep learning, it is no wonder that the first beneficiaries of this form of artificial intelligence are the biggest Internet companies. Tech behemoths like Facebook and Amazon have the most data, which are often

automatically labeled via user action (Did the user click or buy? How many minutes did the user stay on a page?). These user actions are directly related to a business metric (either revenue or clicks) to maximize. When these conditions are met, an app or platform can become a money-printing machine. As the platform collects more data, it makes more money. It is no wonder that giant Internet companies like Google, Amazon, and Facebook have experienced phenomenal growth in the past decade and become powerhouse AI companies.

Beyond Internet companies, the next industry that is low-hanging fruit for AI is finance, including banks and insurance companies, as "The Golden Elephant" shows. Consider the example of insurance. The industry has similar benefits to Internet companies': a large amount of high-quality data within a single domain (insurance) connected to business metrics. The emergence of AI-based fintech (financial technology) companies, such as Lemonade in the United States and Waterdrop in China, are making it possible to buy insurance in an app, or take a loan in an app, with instant approval. These AI-based fintech companies are poised to overtake brick-and-mortar financial corporations because they deliver better financial results (lower default or fraud rate), instantaneous transactions (using AI and the app), and lower costs (no humans in the loop). Traditional financial companies are also hurrying to implement AI in their existing products and processes. The race is on.

Another interesting benefit of AI fintech is that it can use data beyond those considered by human professionals. It can improve predictive power by tapping into massive heterogeneous data that would not be feasible for a human insurance underwriter to assess, for example, whether you buy more processed foods or vegetables, whether you spend a lot of time in a casino or in a gym, whether you invest in Reddit-group recommendations or hedge funds, whether you have a girlfriend or harass women online. All of this evidence would say a lot

can find elevated disease risks of smoking, and thus try to reduce smoking, which is good. But AI might also find that a potential romantic pairing—even one that could help bridge societal division in the long term—could, according to its narrow analysis of the data, increase insurance premiums. So, inference results in actions that serve to tear people apart and exacerbate inequality.

How can we solve this problem? One general approach is to teach AI to have complex objective functions, such as lowering insurance premiums while maintaining fairness. When it comes to maximizing the time humans spend on social networks, for example, Tristan Harris has proposed using "time well spent" as a metric instead of simple "time spent." These two goals could be blended into a complex objective function. Another solution proposed by AI expert Stuart Russell is to ensure that every objective function always be beneficial to humans, by finding a way for humans to be in the loop in the design of objective functions. For example, can we build objective functions for "greater human good," such as our happiness, and can we involve humans to define and label what happiness means? (We explore this idea further in chapter 9, "Isle of Happiness.")

All of these ideas require more AI research on complex objective functions, and also ways to quantify notions like "time well spent," "fairness," or "happiness." Furthermore, each of these ideas would cause companies to make less money. So how can companies be incentivized to do the right thing? One possibility is to have government regulations that penalize offenders. Another is to encourage positive behavior as a part of corporate social responsibility, such as ESG (environmental, social, and corporate governance). ESG is gaining traction in some business circles, and it is possible that responsible AI could be a part of the future ESG. Another idea is for third parties that can serve as watchdogs by creating dashboards for companies' performance, tracking metrics like rates of "fake news"

generated or "lawsuits filed alleging discrimination" to pressure them to incorporate pro-user metrics. Finally, perhaps the hardest but the most effective solution is to ensure that the AI owner is 100-percent interest aligned with each user (see chapter 9 for more on this utopian solution).

A second potential downside is fairness and bias. AI bases its decisions purely on data and outcome optimization, which may often be more equitable than decisions made by people, who can be unduly influenced by various prejudices. But there are reasons that AI, too, may be biased. For example, the data used to train the AI may be insufficient and inadequately represent race or gender demographics. One company's recruiting department may find that its AI algorithms are biased against women because the training data didn't include enough women. Or the data may be biased because it was collected from a biased society. Microsoft's Tay and OpenAI's GPT-3 were both known to make inappropriate remarks about minority groups.

Recently, research has shown that AI is able to infer sexual orientation with high accuracy based on facial micro-expressions. Such abilities could lead to discrimination. This is similar to what happened to Sahej in "The Golden Elephant," when his Dalit status was found not directly but by inference. In other words, Sahej was not labeled Dalit, but because his data and features correlated to being a Dalit, warning signals were sent to Nayana, as the AI system tried to keep the two apart. These unfair outcomes are not intentional, yet the consequences are extremely serious. If a society applied them to domains like hospital admissions or criminal justice proceedings, the stakes would be even higher.

Fairness and bias issues with AI will require substantial efforts to address them. Some steps are clear. First, companies using AI should disclose where AI systems are used and for what purpose. Second, AI engineers should be trained with a set of standard principles—like an adapted physician's Hippocratic

oath; engineers need to understand that their profession embeds ethical choices into products that make life-changing decisions, and thus promise to protect users' rights. Third, rigorous testing should be required and embedded in AI-training tools, to provide warnings or disallow use of models trained on data with unfair demographic coverage. Fourth, new laws requiring AI audits could be passed. If a company receives enough complaints, it could be AI audited (for fairness, disclosure, and privacy protection), the same way it might face a tax audit if its books look fishy.

A final issue is that of explanation and justification. People can always give a reason for why they made a decision, because human decisions are based on highly selective experience and rules. But deep learning's decisions are based on complex equations with thousands of features and millions of parameters. Deep learning's "reason" is basically a thousand-dimensional equation, trained from large quantities of data. This "reason" for producing a given output is too complex to explain fully to a human. Yet many key AI decisions are required, by law or by user expectation, to be accompanied by an explanation. A great deal of research is currently under way that attempts to make AI more transparent, either by summarizing its complex logic, or by introducing new AI algorithms that are fundamentally more interpretable.

These downsides of deep learning have caused significant public distrust of AI. But all new technologies have had downsides. History suggests that, with time, many of the early errors of a new technology will be fixed and improved upon. Think about the advent of the circuit breaker to avoid electrocution, and anti-virus software to stave off computer viruses. I am confident there will be technology and policy solutions to address the challenges of AI's influence, bias, and opaque operations. But first we must follow Nayana and Sahej's

footsteps—to inform people about the gravity of problems, and then to mobilize them to work toward a solution.

CHAPTER TWO

GODS BEHIND THE MASKS

STORY TRANSLATED BY EMILY JIN

TRUTH AND MORNING BECOME LIGHT WITH TIME.
—AFRICAN PROVERB

While other parts of Lagos strained under the pressure of its young population, the Yaba district was flourishing. Dubbed "The Silicon Valley of West Africa," the neighborhood stood out for its orderliness, fresh air, and high tech–infused daily life. Pedestrians could activate the cartoon animals on the billboards and interact with them via hand gestures. Cleaning robots roamed the streets, collecting and sorting trash, then sending it off to recycling centers where it was turned into renewable materials and biofuel. Sustainable bamboo fiber had recently made the leap from building material to fashion trend, at least for the denizens of Yaba.

Standing outside the station and holding his smartstream up to eye level, Amaka overlaid a live virtual route map onto the surrounding streetscape. Following the projected route, he began walking, eventually stopping before a gray building, emblazoned with the number 237 and tucked away on a quiet backstreet. The company he was looking for, Ljele, was apparently based on the third floor. Two days ago, he had received a mysterious email from an anonymous Ljele account about a job that was "right up his alley." The position was his under the condition of his showing up for an interview in person.

As Amaka entered a small reception area on the third floor, the receptionist smiled and pointed to Amaka's mask, indicating he should remove it for an identity check. The young man hesitated, then took his mask off. Reflected in the camera lens was a young, smooth face. His 3D-printed mask couldn't match the delicate quality of the pricey handmade versions sold at absurd prices to tourists in the Lekki Market, but the coarse reproduction, with its butterfly-like pattern, was enough to fool the facial recognition algorithm of most common surveillance cameras. In the eyes of AI, Amaka was a "faceless person." The mask not only saved him money, but, more important, shielded him from the authorities. After all, Amaka had yet to obtain a migrant residence permit.

When the face scan was completed, the receptionist brought Amaka into a conference room and told him to wait. He sat stiffly

as he pondered how he would answer questions regarding his previous work experience. *I have to lie,* he realized. *I don't have many other choices.*

Ten minutes passed. The promised interviewer did not appear. Abruptly, the projection wall across from him lit up, and surveillance camera video footage began to play.

To Amaka, the video footage was as familiar as the back of his own hand. Midnight. Dim, yellow streetlamps. Several homeless people were scattered under an overpass, lying on makeshift mattresses. The silhouette of a boy emerged from the shadows. The boy walked over to a group of sleeping people and gazed down. The camera zoomed in. The boy was white, no more than five or six years old, dressed in striped pajamas, his face wan and expressionless. One of the people woke up with a start and met the boy's eyes. The homeless man asked the boy what his name was and where he lived. The boy's body trembled as he mumbled incoherently. Suddenly, his face twisted, the corners of his lips stretching open and revealing two rows of sharp teeth. He bit down hard on the homeless man's neck. The man cried out in pain, waking up the others. The boy fled the scene, blood trickling down his lips and chin.

The video, originally posted to the Internet under the title "White Vampire Boy Attacks Homeless People in Lagos," had received millions of views within twenty-four hours of its first appearance on the GarriV video-sharing platform. Within days, however, the platform identified the video as a fake and removed it in compliance with the law. The uploader's account, "Enitan0231," was consequently terminated, with all its associated advertising revenue frozen.

Suddenly, a booming voice filled the conference room where Amaka still sat, alone. "Well done, Amaka! What a seamless fusion of realistic settings, amateur actors, and live video shooting. I can't believe you made this in an underground Internet café in Ikeja," said a man's voice with a heavy Igbo accent.

Instinctively, Amaka jumped to his feet. "Who are you?" His eyes surveyed the empty room and landed on the speakers.

"Hey, relax. You can call me Chi. Do you want a job, or not?"

Sighing, Amaka sat back down and slouched in the chair. The man named Chi was right. Without a residence permit, he could never find a real job in Lagos. The mysterious Ljele company was his only sliver of hope. "Why me?" he asked.

"We saw your work. You're talented. You're ambitious—you wouldn't have come to Lagos in the first place if you weren't determined to make a name for yourself. Most importantly, we need someone we can trust. *One of our own kind.*"

Amaka knew immediately what Chi was alluding to. Nigeria has more than 250 ethnic groups, with their own languages and customs, many of which had been in conflict for hundreds of years. The Yoruba and the Igbo, respectively the second- and third-largest ethnic groups in the country, had seen violent clashes in recent years, as both groups muscled for political gain. With the Yoruba as the dominant population in Lagos, Amaka, an Igbo from the southeast, usually concealed his ethnicity to avoid trouble. "What do you want me to do?"

"I want you to do what you do best. Fake a video."

"*Illegally,* I presume?"

"We'll supply you with all you need."

Amaka narrowed his eyes, his nostrils flaring. "And what if I turn down your offer? Will you kill me?"

"Kill you? No, no. Worse than that."

Another video started to play on the projection wall. A dance floor in a private nightclub. The camera zoomed down on the room from a corner of the ceiling. Several boys were dancing up against one another under the flashing laser lights, shirtless. The camera zoomed in farther to reveal the unmistakable face of Amaka. As the camera observed, Amaka turned and passionately kissed another boy whose cheeks glowed fluorescent pink. Amaka then twisted his upper body around to kiss a darker-skinned boy behind him. The video froze on this frame. The three young faces

were like mango leaves that overlapped, intertwined and merged into one another.

Amaka stared at the video, his expression blank. After a few moments, he grinned. The facial scan he had undergone back at the reception desk had provided the data to make this instantaneous deepfake.

"The face might be mine, but not the neck," said Amaka as he pulled down his hood, exposing a long pink scar that cut diagonally from below his right ear to his left collarbone. A souvenir from a street fight. "Also, don't forget we're in Lagos. The things people do here are far crazier than that."

"Sure, but this video can still send you to prison. Think about your family," said Chi, his voice turning soft.

Amaka fell silent. Three decades after the passage of the Same Sex Marriage (Prohibition) Act of 2013, Nigerian society remained just as hostile toward sexual and gender minorities as it ever had been. If someone reported him, Amaka knew it would be difficult to avoid dealing with the corrupt police, who would likely try to extort him, even if he could avoid criminal charges.

And then there was his family. While they hadn't had an easy relationship these past few years, Amaka hated to imagine the pressure that could descend upon the shoulders of his family members, especially his father, who expected the world of him. *Even if the video is a fake.*

The boy bit on his lower lip and pulled his hood back up. Concealing parts of his skin again gave him a little more sense of safety. "I need an advance payment. Cryptocurrency. Also, give me as much detail as you have on the target. I don't want to waste my time on research."

"Your call, my friend. As for the target...there's absolutely no way for you to miss him."

The blurred headshot of a man flashed on the projection wall. When the face's contours solidified into a clear picture, Amaka's eyes widened.

THE YORUBA CALLED THE CITY of Lagos "Eko," meaning "farm." In the equatorial monsoon climate, June was the coolest month with the most plentiful rain. With the rain's monotonous tapping on the metal roof as his background soundtrack, Amaka lay on the small bed of his illegal hostel room. He put on his XR glasses and fiddled with his new gadget—a dark green Illumiware Mark-V.

Compared with the pranks he had carried out in the past, this new job was on a completely different level.

It's not that he lacked experience in video deception—quite the opposite. Alone in his room, Amaka had spent many nights of the past year disguising himself as uptown girls on dating apps. In order to construct a flawless imitation, the first step was to gather as much video data as possible with a web crawler. His ideal targets were fashionable Yoruba girls, with their brightly colored V-neck *buba* and *iro* that wrapped around their waists, hair bundled up in *gele.* Preferably, their videos were taken in their bedrooms with bright, stable lighting, their expressions vivid and exaggerated, so that AI could extract as many still-frame images as possible. The object data set was paired with another set of Amaka's own face under different lighting, from multiple angles and with alternative expressions, automatically generated by his smartstream. Then, he uploaded both data sets to the cloud and got to work with a hyper-generative adversarial network. A few hours or days later, the result was a DeepMask model. By applying this "mask," woven from algorithms, to videos, he could become the girl he had created from bits, and to the naked eye, his fake was indistinguishable from the real thing.

If his Internet speed allowed, he could also swap faces in real time to spice up the fun. Of course, more fun meant more work. For real-time deception to work, he had to simultaneously translate English or Igbo into Yoruba, and use transVoice to imitate the voice of a Yoruba girl and a lip sync open-source