CONTENTS

Preface vii

1 The AI Paradox 1

2 The Agreement Paradox 20

3 The Intelligence Paradox 54

4 The Justice Paradox 81

5 The Regulation Paradox 97

6 The Power Paradox 127

7 The Superintelligence Paradox 159

8 The Solution Paradox 181

9 Epilogue 196

Notes 199

Bibliography 211

Index 217

1

The AI Paradox

THE IRREPLACEABLE HUMAN

ARTIFICIAL INTELLIGENCE (AI) is shaping our world in ways big and small, from helping us find the fastest route home to sparking debates about whether it could one day replace human workers or even outthink us altogether. Just think of the maps app on your phone. It quickly calculates the best route based on real-time traffic, something that would take a human much longer to figure out. Yet, if the app suggests a route through a dangerous area, it doesn't understand the broader context or consequences of sending you in that direction; it just follows the data about road and traffic conditions. This mix of impressive capabilities and clear limitations is what makes AI both exciting and deeply complex. As it evolves, it challenges us to consider not just what it can do but also what it cannot, and whether it could ever truly replicate the depth of human intelligence.

Throughout this book, we'll explore why, despite AI's potential to mimic or outperform certain abilities, the distinct qualities of human intelligence, such as our emotional depth, creativity, ethical insight, and capacity for complex reasoning,

remain beyond the reach of any machine. We will see that these uniquely human traits, which are woven together in intricate and evolving ways, cannot be fully replaced, no matter how advanced AI becomes. For instance, although AI can quickly analyze vast datasets, provide advice, generate text, and offer diagnostic insights, it cannot grasp the nuanced emotional cues in a conversation or generate or appreciate genuinely innovative artistic expressions.[1] For instance, Emily M. Bender, a linguist at the University of Washington, and colleagues, emphasize that large language models like ChatGPT replicate language patterns without true comprehension, leading to potential inaccuracies and a lack of genuine empathy.[2] Similarly, Shannon Vallor, a philosopher of technology, points out that although AI can mimic human traits, it lacks the capacity for virtues like courage, honesty, and empathy, which are fundamental to human experience. She warns that although AI might simulate emotions, it doesn't possess the genuine emotional depth that characterizes human interactions.[3]

Humans create timeless art, write evocative literature, and compose soul-stirring music, forms of creativity that AI can imitate but most likely not authentically replicate, raising the question of whether intelligence lies in the process, the outcome, or a combination of both. Additionally, humans possess moral and ethical discernment, that is, the ability to distinguish right from wrong, which allows us to navigate social complexities and make decisions based on empathy, values, and cultural context. This multidimensional form of intelligence enables humans to excel in areas like negotiation, leadership, and caregiving, where emotional intelligence and ethical considerations are crucial. In fact, the uniqueness of human intelligence is what enables the development of AI and its evolving capabilities

in the first place. It is the human capacity for abstract thinking, experimentation, and ethical reflection that drives the development and refinement of AI technologies. Thus, even as AI continues to evolve, the depth and breadth of human thought and experience remain unparalleled and foundational to technological advancement, as we will discuss in chapter 3. The fundamental paradox of AI is that, even in a machine-dominated era, human intelligence is still crucial.

The AI Paradox
The more AI can do, the more it highlights the irreplaceable nature of human intelligence.

Given this, on the one hand, it may be reassuring to know that we people will not be soon, or ever, replaced or taken over by AI; on the other hand, we do need to ask ourselves some fundamental and hard questions:

- How do we ensure safety and control over AI to prevent it from being used against human interests, including mitigating existential risks?
- Will all people benefit equally and equitably from the advances of AI technology, or will some of us be more "replaceable" than others? If so, who is replaceable and who will stay ahead?
- Will AI development contribute to even larger power imbalances and economic disruptions, such as widespread unemployment? How can we ensure that everybody can realize their full potential and unique capabilities?
- How can we address the unintended consequences that may arise from superintelligence? In particular, what

measures can prevent the weaponizing and misuse of AI for malicious purposes?
- Is there a risk that overreliance on AI could diminish human autonomy and critical thinking skills?

This book will explore these pressing questions one by one, uncovering the paradoxes at the core of each. Paradoxes challenge our assumptions, force us to think more deeply and reveal that reality is rarely as simple as it seems, leading us to realize that much of what we believe to be true is shaped by how we frame a question, not just by the answer itself. Central to the book is the core AI paradox, which we will explore in this chapter. It will help us get a better grip on AI's potential, its limits, and the key choices we face about its role in our lives. It is up to us to engage with these challenges, embrace the opportunities, and guide AI development toward a future that reflects our values and aspirations.

AI is challenging to define, but that does not hinder our ability to talk about it and to conceptualize it. In fact, we are all quite capable of engaging in meaningful discussions about it without being constrained by the lack of a formal definition. Our understanding of AI is shaped by its real or perceived capabilities, while we also continuously shape it through our conceptualizations and narratives. But because we need to start somewhere, a commonly accepted definition of AI comes from the OECD:[4] *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."*[5] Interestingly, this definition, updated in 2023, refines their original 2018 definition, which highlights the

inherent challenge of defining AI.[6] We will explore this issue further in chapter 2.

Independently of how we define it, for many of us, AI may feel like the weather, something that happens to us but is beyond our control. We adapt to the weather by carrying an umbrella when rain is expected or planning a picnic when the forecast is sunny. Yet AI is not the weather. Even though its behavior can sometimes seem unpredictable or difficult to understand, it is entirely a human creation. AI is more like a car engine. It is built by people, following specific instructions to achieve certain goals and meet particular requirements. It is a product of human creativity, shaped by our choices and intentions. Which means that we are not powerless. We have the ability to decide how AI is designed, developed, and applied to shape the outcomes we want. Ultimately, the results depend on the decisions we, as people, make.

This does not mean that we all have to become experts on the technology and the methods that make AI systems possible. For one, I don't really understand how a car engine works, nor am I able to build or repair a car engine. But I understand it sufficiently to know how to drive a car, and how to interact with other traffic on the road. In the same way, we need to be able to have some understanding of what AI can do, what it is and why it is developed, for whom and by whom it is developed, and what AI can do for us. In chapter 2 we will further discuss what AI is and what options we have.

For better or worse, AI is being defined by us, people. Then, shouldn't we all have something to say about it? To be part of this discussion, to be able to demand means by which to govern AI, to demand accountability for what is done using AI, and to define the boundaries of when and where AI can and should be

used, means that we all need to be able to understand what AI can and cannot do, and what its consequences are.

By accepting that AI "happens to us," we are accepting that others will determine what AI will do to all of us. As we will see in chapter 6, there are now some who benefit from this situation: the tiny group that currently has the power to decide about how to develop and use AI. But we cannot accept that we have no say on how it interacts with us, or passively let it be "dumped" on us. It is easy to let it be, but now is the time to determine how it should be. If we again take the automotive industry as an example, the continuous improvements we see in cars are not only the result of industry choices but are also guided by public opinion and regulation. Our responsibility starts with our ability and our willingness to be well informed and to contribute to the discussion; to use our voices as informed and concerned citizens of the world.

I've been working in the field of AI, both in industry and in academia, since the late 1980s. I developed my first AI system in 1986, an expert system to determine eligibility for social housing. Since then, I've not only witnessed but also directly experienced the field's ups and downs. Never before has there been such widespread excitement, and fear, across so many sectors as we have seen in the past decade, especially since the launch of large language models (LLMs) and other generative AI applications at the end of 2022.[7] The true and sustainable future of AI depends on our recognition that its potential is fundamentally rooted in human intelligence. I understand AI deeply and know how to build these systems, but the more I learn, the more I see that AI can never replace the essential ingenuity and creativity that only humans possess.

Take, for example, the use of AI in medical imagery. AI algorithms, particularly those based on machine learning, are

increasingly used to analyze medical images like X-rays, MRIs, and CT scans. Sometimes algorithms can detect patterns and anomalies with greater speed and accuracy than human radiologists. This led, some years ago, several experts, including 2024 Nobel Prize winner Geoff Hinton and the influential computer scientist Andrew Ng, to conclude that the job of radiologist is at risk of being replaced by AI. But they retracted that conclusion some years later, even though AI is very valuable to handle large volumes of diagnostic data, aiding in early detection of diseases such as cancer, which can significantly improve treatment outcomes. This advanced capability of AI, however, also highlights the indispensable role of human healthcare professionals.[8] Although AI can identify patterns and suggest diagnoses, it lacks the ability to consider not only the patient's full medical history but also their social context, lifestyle factors, and personal preferences.[9] AI also cannot provide the empathetic care and communication trust that are critical in the patient-doctor relationship.[10] So, even if AI platforms such as Google Health[11] or PathAI,[12] have demonstrated capabilities that can match or even exceed human performance in specific tasks, they are primarily designed to assist on very specific tasks and are not capable of fully replacing medical professionals. Doctors are essential for interpreting AI findings in the broader context of patient care and discussing treatment options with patients. Healthcare professionals are also responsible for the final diagnostic decisions, and are accountable for these. In fact, even if we think that machine-generated diagnostics may some day be more accurate than those of a medical professional, none of us would comfortably accept *"the computer said so"* as a justification for a medical decision, in particular if the computer cannot explain why. Human expertise and capability to take responsibility remain essential.

AI is often seen as more rational and objective than humans, leading to the belief that it will be able to make "*better*" decisions, and therefore soon replace us in many areas, including our workplaces. AI models interpret the world differently from humans. Current AI systems, particularly machine learning, excel at identifying patterns in data. For example, again in medical imaging, AI detects tumors by recognizing statistical correlations in images. In contrast, humans learn through relationships between concepts, often based on cause and effect. A doctor, for instance, relates lung cancer to symptoms like cough and shortness of breath, informed by a broader understanding of the patient's health, lifestyle, and context, elements that machine learning systems lack. AI models do not possess an ontology of the world or the ability to reason about relationships between concepts.

AI systems are designed to optimize results based on data-driven priorities, whereas human reasoning is rooted in understanding connections and meanings within a broader context. We don't just recognize patterns; we interpret them based on our knowledge of the world. Human cognition also involves abstract thinking, reasoning with limited data, and making inferences from incomplete patterns, skills that statistical methods used by AI cannot replicate, even with larger datasets. This leads to machine learning systems that encode the patterns and relationships in their training data but do not understand the objects those patterns represent. Their form of representation, though alien to us, can be valuable. For example, a system I once built for a national migration agency identified a seemingly bizarre pattern: a correlation between complex asylum cases and people born in January. The reason was simple: asylum seekers without documentation, who are most of the complex cases, are often assigned a "random" birth date, typically

January 1st. This pattern didn't imply causality but helped officials better understand the context of asylum applications. This example shows that correlations are valuable for identifying potential relationships worthy of further investigation, but do not, on their own, provide a basis for formal causal inference or for abstraction. On the whole, we have more to gain from the combined abilities of humans and machines. As we will discuss in chapter 3, human and artificial intelligence not only are different but also serve different purposes.

So, in a nutshell, it can be said that human reasoning uses abstraction and causation as basic processes whereas machines learn by correlation.[13] In *The Book of Why*,[14] computer scientist and philosopher Judea Pearl explains that humans are able to ask "what if" questions, imagine different scenarios, and reason beyond just recognizing the correlations in data that AI and machine learning rely on for finding patterns. AI, on the other hand, can't create such models; it processes the information it has seen without understanding the reasons behind events. Abstraction and causation, which are key to human thinking, go beyond what correlation alone can reveal. Seen as an evolutionary characteristic of humans,[15] abstraction enables us to survive in a complex and dynamic world. Thinking in causal relations allows us to make predictions on how it will evolve and how our actions will change it. Those abstractions and models are not 100 percent correct but are good enough for us to act on them. Whereas correlations help identify relationships, causality requires human interpretation and reasoning. Correlations can suggest associations but do not establish cause and effect. AI can help identify patterns, but humans are needed for causal inference and abstraction. The true potential lies in combining AI and human intelligence to enhance our abilities rather than replace them.

Human decision-making is often guided by intuition or a "gut feeling," a subconscious process machines cannot replicate. This instinctive aspect of human thought is crucial in everything from everyday choices to major life decisions, offering a depth of understanding that AI cannot achieve. Likewise, traits like empathy, self-awareness, and emotional intelligence are uniquely human and cannot be authentically emulated by AI. Although AI can mimic empathetic behavior, especially in therapeutic contexts, it lacks the genuine emotional experience inherent in humans. We humans also possess the ability for moral reasoning, making decisions based on ethical principles about what ought to be, whereas AI systems, driven by utility maximization and data-based logic, cannot distinguish the possible from the impossible or the useful from the useless—leading to phenomena like the "hallucinations" of large language models. Human intelligence is marked by flexibility and adaptability, allowing us to learn from limited information and apply knowledge across various contexts, an essential quality in an age where AI tackles increasingly complex tasks.

Our ability to operate across diverse contexts and our sense of empathy enable us to act not just for personal gain but out of selflessness, equity, and justice—even at the cost of our own interests. Guided by a sophisticated moral compass, we seek to prevent remorse and correct wrongs, aware of the impact our actions have on both ourselves and others. This moral dimension often leads us to "satisficing"—seeking solutions that are "good enough" in light of life's complexities. We live in a world full of shades of meaning, relational dynamics, and ethical ambiguities. The core difference lies not just in capabilities, but in the essence of being: AI calculates, while humans feel; AI iterates, while humans imagine.

A last important feature of human reasoning is that we do not glue social behavior on top of some intelligence. We behave socially before we behave intelligently. Our evolutionary history reveals that social behaviors like cooperation, communication, and group living were not just important for survival—they were the very foundation upon which our intelligence developed. As Frank Dignum, a well-known expert in social AI (and my husband), says: "*Intelligence is deeply grounded in our social nature; it's not merely that social abilities were added to a pre-existing intelligence, but rather that these abilities are the core of all our intelligence*." From infancy, our cognitive growth is intimately linked to social interactions, as our brains are wired to understand others, navigate social relationships, and work within groups. Our need to belong, recognize social cues, and cooperate with others has shaped our cognitive evolution. Thus, our intelligence is a consequence of our sociality, making social abilities the essence of what it means to be intelligent. In contrast, AI systems are more akin to "lone wolves." AI systems are designed to process information, solve problems, or perform specific tasks independent of social context. Although AI systems can mimic social interactions or collaborate with humans, these abilities are typically layered onto a framework that was not inherently designed for social engagement. Sociality is, at best, a capability that developers attempt to build on top of their existing functionalities. This difference highlights a significant distinction: for humans, social intelligence is intrinsic, whereas for AI, it is an extrinsic feature that requires deliberate engineering.

As AI progresses, we will increasingly value the unique human abilities of imagining and creating, not solely based on existing data or patterns but fundamentally emerging from our social traits. For instance, the 2023 Hollywood writers' strike

sought to establish regulations on the use of AI in content creation.[16] It turned out to be one of the longest strikes in that industry and underscored the tension between AI's analytical prowess and its lack of nuance in understanding context. This event was not merely a standoff to limit the capabilities of AI, and thereby a lost race, but a plea to use and develop AI differently. Stemming from concerns about unregulated AI usage potentially replacing human creativity and jobs in film and TV, the strike was a call for a balanced symbiosis between human ingenuity and machine efficiency. The outcome ensures that writers maintain control over how and when to use AI tools, and that studios cannot use AI or digital replications without the informed consent of the performer, thus preserving human roles in the creative process. This set a vital precedent for the industry, emphasizing the need to use AI as an augmentative tool rather than a replacement, thus enriching the creative landscape rather than impoverishing it. The strikes also challenged Hollywood power structures, securing significant gains for actors and writers. The paradox here is that by combating the use of AI in the industry, the strikes ultimately bolstered human creativity, underscoring the unique value of human imagination and securing greater rights and control for writers and performers in an increasingly AI-driven landscape.

The questions of what AI is, and how its capabilities differ from those of people and human organizations, become particularly relevant when we aim at governing or regulating it, as we will discuss in detail in chapter 5. What are we attempting to govern? What is the difference between AI and any other digital system? Is it the large amounts of data AI uses? But many other systems use big data, too. Is it the lack of transparency in AI? Even human organizations lack transparency. Or is our concern rooted in the fact that AI is often in the hands

of big private companies, which might operate beyond democratic oversight? Although market mechanisms are designed to handle the power of large corporations, AI presents unique challenges, in particular where it concerns the ability to act autonomously, bringing with it questions about accountability and unpredictability. AI systems, though efficient, don't have the moral and ethical reasoning humans do. They operate on algorithms and patterns, and lack the nuanced understanding humans bring to decision-making. This difference leads to public debates, regulatory efforts, and the formation of commissions and expert groups. Are we, perhaps, missing something else in this conversation? The essence of these concerns seems to go beyond just the technical aspects of AI and into the realm of its broader impact on society, ethics, and governance.

So far so good. The reflections above seem to suggest that although AI can replicate human intelligence in tasks like data analysis and decision-making, it cannot replace the essential human skills of critical thinking, empathy, and adaptability. In fact, throughout this book, I emphasize the idea that, although AI can process data and identify patterns at incredible speeds, it operates within deterministic frameworks, lacking the human abilities of intuition, moral judgment, and creativity. This inherent limitation reinforces the idea that AI should be seen as a complementary tool to human intelligence, not a replacement, as it cannot grasp the full depth of human reasoning, context, and ethical decision-making, abilities that are crucial for dealing with complex issues and maintaining meaningful human connections, especially in fields like customer service, healthcare, and creative industries. In these areas, the unique touch of human interaction remains irreplaceable by machines. Humans also play a pivotal role in creating AI itself, in tasks

like designing algorithms, curating and labeling data, setting objectives, and ensuring ethical standards. Currently, the complexity and creativity required in these tasks make it unlikely that AI can fully replace humans in AI development. As such, human oversight remains crucial for guiding AI toward beneficial and ethical outcomes, making the human role in AI creation irreplaceable at this stage.

This viewpoint raises an important question: Can it be sustained in the long term, especially when considering the potential rise of superintelligence? Superintelligence, defined as AI systems that surpass human intellect in nearly every domain, could fundamentally change our understanding of human intelligence's role. The advent of such AI systems, following the development of artificial general intelligence (AGI)[17] that operates at the level of human intelligence, would present unprecedented challenges and force us to reconsider the relevance and uniqueness of human cognitive abilities in a world where AI may exceed those capabilities.

The current debate is marked by a division between two camps: On one side, proponents of rapid AI development argue that it holds the key to solving global challenges such as climate change and healthcare, while driving economic growth and geopolitical advantage. They emphasize the importance of leading in AI to maintain a competitive edge and advocate for incremental improvements based on real-world deployment, seeing risks as overstated and manageable through innovation and adaptation. On the other side, those who prioritize caution focus on the existential risks advanced AI could pose to humanity, including the loss of control over AI systems. They highlight ethical concerns like bias, job displacement, and privacy issues, calling for strong governance and global cooperation. This camp stresses the need for precautions, believing

that once certain advancements are made, they may be irreversible, making it crucial to slow progress and ensure AI is developed safely and ethically.

Critics of existential risk arguments contend that these concerns are speculative and overestimate AI's future capabilities, focusing on unlikely worst-case scenarios. At the same time, they argue that this debate has distracted from pressing issues such as bias, privacy, and job displacement. Focusing on distant, improbable threats causes resources to be misallocated, developments to remain out of the public eye, and real-world challenges to be left unaddressed. As we will explore in later chapters, focusing on current risks and challenges will help us develop AI in line with principles of trustworthiness, responsibility, and governance, which will make catastrophic outcomes less likely.

In my view, the quest for superintelligence or AGI is not just a triumph of technology but, in many ways, a failure of governance and common sense. The real threat comes not from the power of superintelligent systems but rather from our inability to use technology responsibly. Solving our complex societal problems is not about perfect technology, but about using technology alongside better governance, deeper reflection, and greater participation. When it comes to the complex, wicked problems we face today, such as climate change, migration, and democracy—there are no perfect answers.[18] Every solution comes with trade-offs, and the goal is to understand what is at stake and the consequences of each proposed solution, rather than accepting the "perfect" answer an AGI system may offer. The risk lies not in the specific decisions AI will make, but in granting AI the power to make decisions in the first place. We all have a role to play—one that machines can never replace—because we are an integral part of society and

part of both the problem and the solution. In this book, we will explore the many options available to address these challenges. We have choices, and fortunately, not all of them rely on technology. But we must have the courage to make those choices and act on our ability to decide.

The Usefulness of Paradoxes

Paradoxes teach us to question and critically evaluate our assumptions, leading to a deeper understanding of the situations we encounter. Paradoxes highlight the complexity of the world around us. They remind us that simple, one-size-fits-all solutions are often inadequate for addressing complex problems, which require a more nuanced and contextual approach. In untangling this fundamental AI paradox we will encounter many more paradoxes. This book explores how these paradoxes can help us get a more informed understanding of the field, the vested interests and stakes involved, the opportunities and risks for individuals and society, and the implications for a sustainable and just future for all. I hope that it can help us get a deeper comprehension of our human role in an AI-mediated world, emphasizing our increasing responsibility for the technology we're creating and using to shape our world. The chapters of this book are arranged around the following paradoxes, challenging our intuitive assumptions about how AI works and what role it has in our society.

1. The AI Paradox: The more AI can do, the more it highlights the irreplaceable nature of human intelligence. (this chapter)
2. The Agreement Paradox: The more we explore AI, the harder it becomes to agree on its definition. (chapter 2)

3. The Intelligence Paradox: AI is what AI cannot do. (chapter 3)
4. The Justice Paradox: Less bias is not always more justice. (chapter 4)
5. The Regulation Paradox: Responsible innovation needs regulation. (chapter 5)
6. The Power Paradox: The more AI you get, the less control you have. (chapter 6)
7. The Superintelligence Paradox: The more we chase AGI, the more we discover that true superintelligence lies in human cooperation. (chapter 7)
8. The Solution Paradox: Solving problems with technology often creates more problems. (chapter 8)

Life is full of paradoxes, and they are not just silly or laughable; they often reveal flaws in how we understand a concept or situation. One famous example is Zeno's paradox: to reach a goal, you must first cover half the distance, then half of the remaining distance, and so on, suggesting you would never actually reach your destination. In everyday life, this reflects how breaking a larger goal into smaller tasks can sometimes feel overwhelming, preventing us from starting. Yet, dividing tasks into manageable steps is what allows us to make progress. Zeno's paradox teaches that overthinking can lead to inaction, and taking the first step is often better than being paralyzed by analysis.

This paradox also offers valuable insights for AI development. It aligns with the Intelligence Paradox discussed in chapter 3, where new challenges always arise just as we think we've achieved something. Instead of discouraging us, this highlights the importance of every incremental step in technological progress. Breakthroughs come from consistent effort

over time—neural networks in the 1970s, natural language processing in the 1950s, and the idea of intelligent machines dating back to ancient Greece. What seems like a dramatic leap today will likely be seen as a small step in the long term. Zeno's paradox reminds us that dividing complex goals into manageable tasks drives progress, and that overthinking can delay action. Similarly, AI innovation involves exploration, creative problem-solving, and often taking unconventional paths. It's not just about advancing technology but fostering social innovation—finding new ways to collaborate, include diverse voices, and regulate effectively. Each step, no matter how small, contributes to building more sophisticated and inclusive systems.

Paradoxes also show how our actions can lead to unexpected outcomes. For example, improving car fuel efficiency may seem environmentally friendly, but it can encourage more driving, which reduces the overall benefit. Similarly, focusing too much on making AI systems highly accurate can unintentionally result in systems that are harder to understand, less reliable, or more biased. This overemphasis on accuracy can lead to unintended consequences, such as AI systems that prioritize technical precision over fairness, transparency, or practical usefulness. To address this, we need a balanced approach that considers not just how accurate systems are, but how they affect people and society as a whole.

Key Takeaways and Reflections

Progress in AI is inseparable from human creativity and intellect. Every step forward in AI development relies on our ideas, imagination, and decisions. As AI evolves, it redefines tasks that require uniquely human traits such as emotional intelligence,

ethical reasoning, and creativity. The fundamental AI paradox, that the more AI achieves, the more it underscores the unique and irreplaceable nature of human intelligence, reminds us of our crucial role in shaping AI's impact. This calls for active engagement and adaptation as we prepare for the transformations AI will bring.

As Pablo Picasso once said, "Computers are useless. They can only give you answers." While AI provides answers to complex problems, it also pushes us to ask sharper, more meaningful questions. The AI paradox encourages us to anticipate unexpected outcomes and approach challenges from multiple perspectives. These reflections allow us to make thoughtful and responsible decisions that guide us toward a future where AI supports and enhances human values.

In summary, the fundamental AI paradox underscores the unique qualities of human intelligence that remain beyond AI's reach. While AI transforms the world around us, it relies on human creativity and ethical judgment to advance. To thrive in this evolving landscape, we must embrace critical thinking, adapt to new challenges, and consider the broader impacts of AI. By understanding and engaging with paradoxes like these, we can shape a future where AI aligns with our collective goals and aspirations.

INDEX

abstraction, 9, 58, 59, 79, 140, 193, 200
accountability, 5, 13, 27, 39, 41, 43, 46, 47, 52, 70, 82, 88, 96, 97, 100, 101, 103, 105–109, 113–115, 119, 120, 125, 128, 130–132, 134, 135, 137, 139, 142, 143, 147, 150, 153, 157, 158, 170, 175, 183, 193, 202, 206
agent, 29, 41, 43, 68, 69, 165, 177, 208; multiagent systems, 41, 42
agent-based modeling, 42
agentic AI, 42
AGI, 14, 15, 17, 76, 159–170, 172–175, 178, 188
AI; compositional, 48, 140; definition, 4; ethical, 70; ethics, 39, 82, 109, 146, 201, 202, 205; explainable, 142; generative, 32, 35, 40, 42, 155, 190, 199; human-centered, 38, 140, 179; humanlike, 45, 46, 77; hybrid, 167; narrow, 159, 166; neurosymbolic, 35; paradox, 3, 4, 16, 19, 75, 196, 198; race, 190; responsible, 51, 105, 191, 192; rule-based, 32; safety, 191, 192, 195; social, 11; superintelligence, 160, 169; symbolic, 35, 167
AI-solutionism, 184, 185, 187
algorithm, 6, 13, 14, 29, 32, 36–38, 40, 47, 50, 65, 66, 75, 84, 86–88, 92–95, 114, 130, 137, 142, 149, 161, 166, 175–177, 184, 193, 202, 204; deterministic, 65, 66; evolutionary, 167; nondeterministic, 65, 66; probabilistic, 65; search, 28; stochastic, 65
anthropomorphizing, 55, 162, 165, 166
arms race, 151, 153, 188–191, 194
art, 2
Artificial General Intelligence, 159, 166
autonomy, 4, 13, 23, 29, 38, 42, 43, 47, 51, 82, 119, 152, 154, 169, 170, 172, 173, 175–177, 200, 201

bias, 14, 15, 18, 29, 40, 49, 68, 71, 72, 74, 81, 83, 84, 86–94, 96, 97, 115, 116, 132, 134, 147, 149, 152, 174, 175, 185–187, 193, 195, 202, 204; social, 91, 92; statistical, 91, 92
biology, 59, 161, 171

causality, 9, 193
chatbots, 75–77, 131
climate change, 14, 15, 79, 163, 164, 173, 174, 184
cognition, 8, 26, 28, 58, 68, 70, 76, 160, 165; social, 64
consciousness, 62, 170, 171
correlation, 8, 9, 32, 76, 79
cryptography, 65

data, 8; analysis, 150, 174; analytics, 29, 71, 209; centers, 48, 119; dataset, 66, 72–74, 92, 93, 137, 149, 150, 167; datasets, 2, 8, 31, 33, 35, 40, 48; datification, 71, 73; synthetic, 71, 72, 203; training, 8
democracy, 15, 145, 148–150
digitalization, 22
discrimination, 81, 83, 87, 89, 91, 92, 97

economy, 99, 119, 155, 156, 201
emotion, 2, 10, 18, 59, 61–64, 70, 84, 131, 164, 169, 171, 178
ethics, 13, 30, 79, 85, 120, 123, 124, 153, 156, 176, 183, 191, 197, 202
ethnicity, 74, 86, 87
EU AI Act, 108
explainability, 7, 35, 40, 95, 129, 139, 140, 143, 193, 209

facial recognition, 146, 147, 152, 186
fairness, 18, 26, 37, 39, 40, 48, 50, 69, 81–91, 93–96, 99, 100, 107, 108, 122, 123, 125, 126, 139, 144, 150, 157, 192, 193, 195; equality, 86, 88, 90, 99, 100; equity, 10, 83–85, 88, 90, 108; inequality, 38, 87, 92, 132, 150, 155, 173, 176, 193, 197
foundational models, 32, 33, 35, 76, 190

gender, 83, 86, 89, 91, 93, 96, 100, 146, 186, 204
governance, 13–15, 20, 21, 30, 37, 51, 86, 101, 103, 106–109, 115, 118, 128, 134, 136, 141, 144, 145, 148, 151–154, 158, 178–180, 183, 189, 191–194, 201

hallucination, 10, 40, 77
human rights, 49, 82, 86, 102, 105, 106, 108, 151, 152, 154, 183, 187, 207

illegality, 49, 131
impossibility theorem, 91
injustice, 47, 83, 84, 86, 87, 100, 203
intelligence, 1, 2, 11, 23, 27, 28, 40, 54, 55, 57–60, 63, 64, 66, 67, 77, 79, 80, 154, 159–161, 166, 167, 170, 174, 176–178, 190, 196, 199; collective, 159–163, 168, 169, 176, 179; general, 165, 167; human, 3, 6, 9, 10, 13, 14, 16, 19, 23, 27, 28, 50, 54, 57–60, 63, 64, 66, 67, 69, 70, 75, 79, 159, 161, 162, 164, 166, 168, 171, 173, 174, 177, 178, 196, 202; humanlike, 69, 78, 176; hybrid, 209; machine, 57, 58, 63, 174; social, 11, 63, 164, 174; superintelligence, 3, 14, 15, 17, 159, 160, 162–164, 169–175, 179, 180

justice, 10, 17, 39, 47, 69, 81–86, 88, 90, 91, 93, 96, 99, 184, 194, 196; procedural, 90

language, 61, 67, 75–78, 87, 164–166, 171, 203; models, 199, 208; patterns, 2; processing, 164; translation, 50
Large Language Models (LLMs), 2, 6, 10, 28, 42, 67, 71, 75–78, 164–166, 188, 199
learning, 27, 28, 31–33, 35, 38, 40, 50, 59, 159, 164, 209; deep, 32, 67, 167, 190; lifelong, 156; machine, 6, 8, 9, 23, 24, 29, 32, 33, 40, 50, 52, 65, 67, 72, 87, 136, 137, 139, 193, 200, 204; reinforcement, 67; self-supervised, 76; strategies, 67
legality, 35, 47, 49, 74, 100, 106, 109, 115, 116, 120, 121, 130, 139, 168, 193, 201
linguistics, 78
literature, 2
logic programming, 28

migration, 8, 15
misinformation, 45, 81, 101, 130, 132, 145, 148–150
morality, 2, 10, 13, 49, 79, 88, 95, 120, 178, 196, 201
multidisciplinary, 51, 95, 122, 187, 194
music, 2, 59

natural language processing, 18, 28, 67
neural networks, 18, 28, 32, 35, 36, 66, 75, 199
neuron, 28, 61, 62
norms, 26, 52, 73, 88, 90, 92, 103, 120, 152, 177, 185

ontology, 8
optimization, 36, 65, 66, 95, 96, 119, 162, 176
oversight, 13, 14, 42, 47, 50, 52, 88, 96, 99, 102, 107, 114, 125, 127, 128, 134, 140, 146, 147, 151, 152, 157, 186, 188, 193, 194

paradox, 3, 4, 12, 16–19, 22, 48, 52, 55, 64, 67, 79, 82, 96, 98, 99, 125, 129, 140, 151, 153, 157, 159, 160, 179, 180, 182, 192, 195–198
pattern matching, 75, 200
policy, 20, 26, 39, 92, 107, 145, 148, 167, 191, 195, 199, 205; maker, 22, 25, 36, 39, 47, 50, 51, 70, 77, 91, 101, 104, 106, 143, 157, 191; making, 25, 36, 52, 115, 179, 184
politics, 132, 184
power, 3, 6, 12, 13, 15, 17, 25, 40, 41, 46, 73, 77, 83, 101, 102, 106, 114, 115, 127–129, 131–135, 139, 140, 144–147, 149–158, 160, 163, 167, 169, 170, 174, 175, 177, 178, 188, 190, 192, 194, 196, 197, 206; computational, 48, 135, 169, 190; computing, 27
privacy, 14, 15, 29, 37, 39, 68, 72, 82, 95, 97, 100–102, 115, 116, 118, 123, 125, 130, 134, 146, 147, 152, 186, 203; data, 144, 150

race, 74, 91, 93, 96
rationality, 63, 68, 69, 85
reasoning, 1, 8–10, 19, 28, 31, 35, 36, 52, 58, 59, 68, 76–79, 120, 121, 142, 161, 164, 165, 170, 190; cognitive, 67; commonsense, 43; ethical, 13, 80, 176, 196; human, 8, 9, 11, 13, 178; logical, 32, 35, 61, 63, 78; symbolic, 35
regulation, 6, 12, 17, 21, 25, 32, 44, 45, 51, 95, 97–105, 109, 111, 113–115, 121–126, 134, 135, 144–146, 150, 158, 177, 181, 183, 188, 192, 197; adaptive, 123; by design, 119; hard, 103; incentive-based, 113; rights-based, 105; risk-based, 105; self-regulation, 103, 104; soft, 103
regulation-free environment, 99
risk, 4, 14–16, 25, 39, 45, 51, 54, 57, 68, 71, 72, 78, 79, 83, 95, 98, 99, 101, 102, 104, 105, 107–109, 115–120, 124, 125, 127, 128, 131, 134, 135, 143, 153, 165, 168, 170–172, 175, 177, 182, 186, 191–194, 201, 208; existential, 3, 14, 15, 116, 172, 175, 177, 178; high-risk environments and threats, 83, 88, 93, 105, 109
robotics, 29, 36, 52, 63, 154
rule-based environments, systems, and decisions, 26, 33, 35

safety, 3, 57, 95, 97–99, 105–107, 109, 110, 113, 117, 120–125, 127, 128, 130, 131, 144, 147, 162, 170, 181, 188, 191, 192

security, 42, 68, 116, 125, 126, 142, 147, 152, 153, 163, 186, 188, 189, 192; cybersecurity, 29, 109, 110; data, 97
silicon vs. biology, 61
social behavior, 11
solutionism, 83, 184; techno-solutionism, 182, 184, 188, 208
sustainability, 6, 16, 43, 48, 112, 119, 151, 191
system; social credit, 146; socio-technical, 30, 38, 42, 51, 70, 177, 192

traits, 10, 86, 89, 160; human, 2, 18, 44, 46, 201; social, 11
transhumanism, 175, 176
transparency, 12, 18, 28, 32, 35, 37, 39–41, 43, 46, 47, 52, 53, 57, 72, 82, 90, 91, 96, 107–110, 112–114, 123, 125, 126, 129–131, 135, 137, 139–141, 143, 148–150, 152–154, 157, 158, 183, 191, 193; algorithmic, 141
trust, 7, 15, 20, 25–27, 35, 45, 46, 50, 51, 53, 57, 70, 81, 82, 86, 97, 99, 106, 107, 109, 110, 112, 121, 122, 124–126, 128, 132, 134, 139, 141–145, 148, 153–155, 183, 191, 194; public, 38, 106, 114, 120, 121, 125, 150, 151
Turing, 177, 201; machine, 65; test, 174

utility, 10, 31, 72, 178

values, 2, 4, 19, 37, 39, 41, 44, 49, 52, 63, 64, 69, 80, 89, 96, 97, 100, 106–109, 112, 114, 125, 126, 157, 158, 166–168, 174, 176, 179, 180, 183, 185, 187, 194–198, 201, 204
virtues, 2, 86

weaponization, 4, 149