

THE INTERNATIONAL BESTSELLER

'Important'
YUVAL NOAH HARARI

'Excellent'
BILL GATES

THE COMING WAVE

AI, Power and Our Future

'Astonishing'
STEPHEN FRY

'Stunning'
RORY STEWART

MUSTAFA SULEYMAN

CEO of MICROSOFT AI

with MICHAEL BHASKAR

WITH A NEW AFTERWORD

Praise for *The Coming Wave*

A *Sunday Times* and *New York Times* Bestseller
An *Economist*, *Financial Times*, *Guardian*,
Prospect and *Sunday Times* Book of the Year
Shortlisted for the *Financial Times* and Schroders
Business Book of the Year 2023

‘*The Coming Wave* is a fascinating, well-written and important book. It explores the existential dangers that AI and biotechnology pose to humankind, and offers practical solutions for how we can contain the threat. The coming technological wave promises to provide humanity with godlike powers of creation, but if we fail to manage it wisely, it may destroy us’

Yuval Noah Harari, author of *Sapiens*

‘This important book is a vivid wake-up call. It carefully outlines the threats and opportunities associated with the exhilarating scientific advances of recent years. *The Coming Wave* is rich with interesting facts, arresting arguments and compelling observations; it is essential reading’

Daniel Kahneman, Nobel Prize winner and
author of *Thinking, Fast and Slow*

‘An excellent guide for navigating our unprecedented times’
Bill Gates, co-founder of Microsoft

‘Deeply rewarding and consistently astonishing’
Stephen Fry, actor, broadcaster and author of *Mythos*

‘A stunning book by a man at the very centre
of the AI revolution’

Rory Stewart, former politician
and co-host of *The Rest Is Politics*

Copyrighted Material

‘An extraordinary and necessary book . . . There are terrifying moments, as there should be when one realises that most of what is familiar is about to be transformed.

But, ultimately one leaves energised and thrilled
to be alive right now’

Alain de Botton, philosopher and author
of *The Course of Love*

‘An erudite, clear-eyed guide both to the history of radical technological change and to the deep political challenges that lie ahead’

Anne Applebaum, author of *Twilight of Democracy*

‘This vital book is inspiring and terrifying at the same time . . . [It] is about the future for all of us: we need to read it and act on it’

David Miliband, former UK Foreign Secretary

‘Truly remarkable, ambitious and impossible to ignore, this book is a persuasively argued tour de force from a leading industry expert that will shape your view of the future – and rewrite your understanding of the present’

Nouriel Roubini, author of *Megathreats*

‘Mustafa Suleyman’s insight as a technologist, entrepreneur and visionary is essential. Deeply researched and highly relevant, this book provides gripping insight into some of the most important challenges of our time’

Al Gore, former Vice President of the United States

‘A practical and optimistic road map for action on the most important issue of our time: how to retain power over entities far more powerful than ourselves’

Stuart Russell, author of *Human Compatible*

‘Confused by the current furore about AI? This book is a good place to start . . . *The Coming Wave* is a sweeping account of the latest advances in AI and synthetic biology’

Financial Times

‘The rise of AI has led to a rise in books about AI, among
the most insightful of which is *The Coming Wave*’
Prospect

‘A cogent look at technology’s potential to transform the economy
and society, along with the risks of misuse and surveillance’
Economist

‘Heartfelt and candid . . . a serious exploration of
what the future might hold for us all’
Observer

‘A genuinely mind-boggling read, setting out the ineluctable
forces soon to completely transform politics, society and even
the fabric of life itself over the next decade or two’
Guardian

‘Suleyman is the perfect guide to a frequently
overwhelming subject’
i Paper

‘[A] sweeping look at the future of artificial intelligence
and other transformative technologies . . . Suleyman is
intimately familiar with the technologies, companies and
personalities at the heart of the AI revolution’
New York Times

‘Convincing . . . Politicians and lawmakers should listen to
the warnings of people such as Mustafa Suleyman’
New Statesman

‘Unusually thoughtful, expansive, historically rooted
and engagingly written’
Washington Post

‘Dazzling . . . You have by now read a great deal of both hype and
doom-mongering on the subject. But Suleyman’s is the book you
cannot afford not to read . . . brilliant’
Bloomberg

‘Calm, pragmatic and deeply ethical . . . enthralling
reading for everyone’
Angela Kane, former UN Under-Secretary-General

‘Nobody has been closer to the unfolding AI revolution
than Mustafa Suleyman, and nobody is better placed
to outline the risks and rewards of the huge technological
changes happening right now. This is an extraordinary
and utterly unmissable guide to this unique
moment in human history’
Eric Schmidt, former CEO of Google

‘It’s more important than ever to see someone within the
technology industry write with such honesty and rigour . . . This
book is a panoramic survey and a clarion call to action
impossible to ignore. Everyone should read it’
Fei-Fei Li, Co-Director of Stanford’s Institute
for Human-Centered AI

‘The best analysis yet of what AI means for the
future of humanity . . . [Suleyman] is a profoundly
talented entrepreneur, a deep thinker and one of the most
important voices on the coming wave of technologies
that will shape our world’
Reid Hoffman, co-founder of LinkedIn

‘Honest, passionate and unafraid to confront what is clearly
one of the great challenges our species will face this century.
Thanks to Suleyman we know what the situation is and what
our options are. Now it’s up to us to act’
Andrew McAfee, author of *The Geek Way*

‘A much-needed dose of specificity, realism and clarity . . . This
important book is a vivid and persuasive road map for
how human beings might guide technological innovations
rather than be controlled by them’
Martha Minow, Harvard Professor and former
Dean of Harvard Law School

Copyrighted Material

‘Offers a powerful argument that today’s explosive technological revolution is poised to be uniquely disruptive. Read this essential book’

Ian Bremmer, author of *The Power of Crisis*

‘Presenting a stark assessment of the dangers as well as the wonders of AI, Mustafa Suleyman proposes an urgent agenda of actions governments must take now to constrain the most potentially catastrophic applications of this revolutionary challenge’

Graham Allison, author of *Destined for War*

‘An incredibly compelling window into the current developments and exponential future of AI – from the ultimate insider . . . If you really want to understand how society can safely navigate this world-changing technology, read this book’

Bruce Schneier, author of *A Hacker’s Mind*

‘The coming wave of AI and synthetic biology will make the next decade the best in human history. Or the worst . . .

Thought-provoking, urgent and written in powerful, highly accessible prose, this is a must-read book’

Erik Brynjolfsson, Professor at Stanford’s
Institute for Human-Centered AI

‘Meticulously researched and packed with original insights and constructive recommendations’

Jason Matheny, CEO of RAND

‘Fantastically clear, energetic, well-researched and readable’
Sir Geoff Mulgan, Professor at University College London

‘Can we control these new technologies before they control us? A world leader in artificial intelligence and a longtime advocate for governments, big tech and civil society to act for the common good, Mustafa Suleyman is the ideal guide to this crucial question’

Jeffrey D. Sachs, Professor at Columbia University

Copyrighted Material

‘Sharp, compassionate and uncompromising . . . a resolute call to action for all of us to participate in this most consequential discourse’

Qi Lu, CEO of MiraclePlus

‘[Suleyman] has thought deeply about what needs to be done to ensure that emerging technologies are used for human good . . . A must-read’

Meghan L. O’Sullivan, Director of the Belfer Center for Science and International Affairs at the Harvard Kennedy School of Government

‘A brave wake-up call that we all need to answer – before it’s too late . . .’

Tristan Harris, co-founder and Executive Director of the Center for Humane Technology

‘Realistic, deeply informed and highly accessible . . . Suleyman’s remarkable and in some senses frightening book shows what must be done to contain these seemingly uncontrollable technologies’

Jack Goldsmith, Learned Hand Professor of Law at Harvard Law School

‘Brilliant and inviting, complex and clear, urgent and calm . . . Everyone who cares about the future should read this book’

Eric Lander, Founding Director of the Broad Institute of MIT and Harvard

‘A strikingly lucid and refreshingly balanced account of our current technological predicament, *The Coming Wave* articulates the defining challenge of our era . . . With every page turned, our odds improve’

Kevin Esvelt, Associate Professor at MIT Media Lab

Copyrighted Material

MUSTAFA SULEYMAN

Mustafa Suleyman is the CEO of Microsoft AI. Previously he was CEO and co-founder of Inflection AI. Before that he co-founded DeepMind, one of the world's leading AI companies, and after a decade there, he became vice president of AI product management and AI policy at Google. When he was an undergraduate at Oxford, Suleyman dropped out to help start a non-profit telephone counselling service. He lives in Palo Alto, California.

MICHAEL BHASKAR

Michael Bhaskar is a writer and publisher based in the UK. He is the author of *The Content Machine*, *Curation* and *Human Frontiers*.

Copyrighted Material

Copyrighted Material

MUSTAFA SULEYMAN

WITH

MICHAEL BHASKAR

The Coming Wave

AI, Power and Our Future

VINTAGE
Copyrighted Material

1 3 5 7 9 10 8 6 4 2

Vintage is part of the Penguin Random House group of companies

Vintage, Penguin Random House UK, One Embassy Gardens,
8 Viaduct Gardens, London SW11 7BW

penguin.co.uk/vintage
global.penguinrandomhouse.com



Penguin
Random House
UK

The edition first published in Vintage in 2025

First published in Vintage in 2024

First published in hardback in Great Britain by The Bodley Head in 2023

First published in the United States of America by Crown in 2023

Copyright © Mustafa Suleyman and Michael Bhaskar 2023

The moral right of the authors has been asserted

Book design by Barbara M. Bachman

Penguin Random House values and supports copyright. Copyright fuels creativity, encourages diverse voices, promotes freedom of expression and supports a vibrant culture. Thank you for purchasing an authorised edition of this book and for respecting intellectual property laws by not reproducing, scanning or distributing any part of it by any means without permission.

You are supporting authors and enabling Penguin Random House to continue to publish books for everyone. No part of this book may be used or reproduced in any manner for the purpose of training artificial intelligence technologies or systems. In accordance with Article 4(3) of the DSM Directive 2019/790, Penguin Random House expressly reserves this work from the text and data mining exception.

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

The authorised representative in the EEA is Penguin Random House Ireland,
Morrison Chambers, 32 Nassau Street, Dublin D02 YH68

A CIP catalogue record for this book is available from the British Library

ISBN 9781529923834

Penguin Random House is committed to a sustainable future for
our business, our readers and our planet. This book is made from
Forest Stewardship Council® certified paper.



Copyright Material

CONTENTS

GLOSSARY OF KEY TERMS	xiii
PROLOGUE	3
CHAPTER 1: CONTAINMENT IS NOT POSSIBLE	5
PART I: <i>HOMO TECHNOLOGICUS</i>	
CHAPTER 2: ENDLESS PROLIFERATION	23
CHAPTER 3: THE CONTAINMENT PROBLEM	35
PART II: THE NEXT WAVE	
CHAPTER 4: THE TECHNOLOGY OF INTELLIGENCE	51
CHAPTER 5: THE TECHNOLOGY OF LIFE	79
CHAPTER 6: THE WIDER WAVE	92
CHAPTER 7: FOUR FEATURES OF THE COMING WAVE	103
CHAPTER 8: UNSTOPPABLE INCENTIVES	117
PART III: STATES OF FAILURE	
CHAPTER 9: THE GRAND BARGAIN	147
CHAPTER 10: FRAGILITY AMPLIFIERS	160
CHAPTER 11: THE FUTURE OF NATIONS	183
CHAPTER 12: THE DILEMMA	205
PART IV: THROUGH THE WAVE	
CHAPTER 13: CONTAINMENT MUST BE POSSIBLE	225
CHAPTER 14: TEN STEPS TOWARD CONTAINMENT	239
LIFE AFTER THE ANTHROPOCENE	281
AFTERWORD FOR THE PAPERBACK EDITION	287
ACKNOWLEDGMENTS	295
NOTES	297
INDEX	326

Copyrighted Material

GLOSSARY OF KEY TERMS

AI, AGI, AND ACI: Artificial intelligence (AI) is the science of teaching machines to learn humanlike capabilities. Artificial general intelligence (AGI) is the point at which an AI can perform all human cognitive skills better than the smartest humans. ACI, or artificial capable intelligence, is a fast-approaching point between AI and AGI: ACI can achieve a wide range of complex tasks but is still a long way from being fully general.

THE COMING WAVE: An emerging cluster of related technologies centered on AI and synthetic biology whose transformative applications will both empower humankind and present unprecedented risks.

CONTAINMENT: The ability to monitor, curtail, control, and potentially even close down technologies.

THE CONTAINMENT PROBLEM: Technology's predisposition to diffuse widely in waves and to have emergent impacts that are impossible to predict or control, including negative and unforeseen consequences.

THE DILEMMA: The growing likelihood that both new technologies and even their absence might lead to catastrophic and/or dystopian outcomes.

FOUR FEATURES: The unique characteristics of the coming wave that exacerbate the challenge of containment. They are asymmetry, hyper-evolution, omni-use, and autonomy.

FRAGILITY AMPLIFIERS: Applications and impacts of coming-wave technologies that will shake the already brittle foundations of the nation-state.

Copyrighted Material

THE GRAND BARGAIN: In exchange for a monopoly over the right to use force, citizens expect nation-states to preserve order and provide public services, including by harnessing new technologies while minimizing the harmful side effects.

THE NARROW PATH: The potential for humankind to strike a balance between openness and closure when it comes to containing the technologies in the coming wave that avoids catastrophic or dystopian outcomes.

PESSIMISM AVERSION: The tendency for people, particularly elites, to ignore, downplay, or reject narratives they see as overly negative. A variant of optimism bias, it colors much of the debate around the future, especially in technology circles.

SYNTHETIC BIOLOGY: The ability to design and engineer new organisms or redesign existing biological systems.

TECHNOLOGY: The application of scientific knowledge (in the broadest possible sense) to produce tools or practical outcomes.

WAVES: The global diffusion or proliferation of a generation of technology anchored in a new general-purpose technology.

**T H E
C O M I N G
W A V E**

Copyrighted Material

Copyrighted Material

PROLOGUE

THIS IS HOW AN AI SEES IT.

QUESTION: What does the coming wave of technology mean for humanity?

In the annals of human history, there are moments that stand out as turning points, where the fate of humanity hangs in the balance. The discovery of fire, the invention of the wheel, the harnessing of electricity—all of these were moments that transformed human civilization, altering the course of history forever.

And now we stand at the brink of another such moment as we face the rise of a coming wave of technology that includes both advanced AI and biotechnology. Never before have we witnessed technologies with such transformative potential, promising to reshape our world in ways that are both awe-inspiring and daunting.

On the one hand, the potential benefits of these technologies are vast and profound. With AI, we could unlock the secrets of the universe, cure diseases that have long eluded us, and create new forms of art and culture that stretch the bounds of imagination. With biotechnology, we could engineer life to tackle diseases and transform agriculture, creating a world that is healthier and more sustainable.

But on the other hand, the potential dangers of these technologies are equally vast and profound. With AI, we could create systems that are beyond our control and find ourselves at the mercy of algorithms

Copyrighted Material

that we don't understand. With biotechnology, we could manipulate the very building blocks of life, potentially creating unintended consequences for both individuals and entire ecosystems.

As we stand at this turning point, we are faced with a choice—a choice between a future of unparalleled possibility and a future of unimaginable peril. The fate of humanity hangs in the balance, and the decisions we make in the coming years and decades will determine whether we rise to the challenge of these technologies or fall victim to their dangers.

But in this moment of uncertainty, one thing is certain: the age of advanced technology is upon us, and we must be ready to face its challenges head-on.

THE ABOVE WAS WRITTEN by an AI. The rest is not, although it soon could be. This is what's coming.

CONTINENT IS NOT POSSIBLE

THE WAVE

ALMOST EVERY CULTURE HAS A FLOOD MYTH.

In ancient Hindu texts, the first man in our universe, Manu, is warned of an impending deluge and becomes its sole survivor. The *Epic of Gilgamesh* records the god Enlil as destroying the world in a giant flood, a story that will resonate with anyone familiar with the Old Testament story of Noah's ark. Plato talked of the lost city of Atlantis, washed away in an immense torrent. Permeating humanity's oral traditions and ancient writings is the idea of a giant wave sweeping everything in its path, leaving the world remade and reborn.

Floods also mark history in a literal sense—the seasonal flooding of the world's great rivers, the rising of the oceans after the end of the Ice Age, the rare shock of a tsunami appearing without warning on the horizon. The asteroid that killed the dinosaurs created a towering mile-high wave, altering the course of evolution. The sheer power of these swells has seared itself into our collective consciousness: walls of water, unstoppable, uncontrollable, uncontainable. These are some of the most powerful forces on the planet. They shape continents, irrigate the world's crops, and nurture the growth of civilization.

Other kinds of waves have been just as transformative. Look again at history and you can see it marked by a series of metaphorical waves: the rise and fall of empires and religions, and bursts of commerce. Think of Christianity or Islam, religions that began as small ripples

before building and crashing over huge stretches of the earth. Waves like this are a recurrent motif, framing the ebb and flow of history, great power struggles, and economic booms and busts.

The rise and spread of technologies has also taken the form of world-changing waves. A single overriding trend has stood the test of time since the discovery of fire and stone tools, the first technologies harnessed by our species. Almost every foundational technology ever invented, from pickaxes to plows, pottery to photography, phones to planes, and everything in between, follows a single, seemingly immutable law: it gets cheaper and easier to use, and ultimately it proliferates, far and wide.

This proliferation of technology in waves is the story of *Homo technologicus*—of the technological animal. Humanity's quest to improve—ourselves, our lot, our abilities, and our influence over our environment—has powered a relentless evolution of ideas and creation. Invention is an unfolding, sprawling, emergent process driven by self-organizing and highly competitive inventors, academics, entrepreneurs, and leaders, each surging forward with their own motivations. This ecosystem of invention defaults to expansion. It is the inherent nature of technology.

The question is, what happens from here? In the pages that follow, I will tell you the story of history's next great wave.

LOOK AROUND YOU.

What do you see? Furniture? Buildings? Phones? Food? A landscaped park? Almost every object in your line of sight has, in all likelihood, been created or altered by human intelligence. Language—the foundation of our social interactions, of our cultures, of our political organizations, and perhaps of what it means to be human—is another product, and driver, of our intelligence. Every principle and abstract concept, every small creative endeavor or project, every encounter in your life, has been mediated by our species' unique and endlessly complex capacity for imagination, creativity, and reason. Human ingenuity is an astonishing thing.

Only one other force is so omnipresent in this picture: biological life itself. Before the modern age, aside from a few rocks and minerals,

most human artifacts—from wooden houses to cotton clothes to coal fires—came from things that were once alive. Everything that has entered the world since then flows from us, flows from the fact that we are biological beings.

It's no exaggeration to say the entirety of the human world depends on either living systems or our intelligence. And yet both are now in an unprecedented moment of exponential innovation and upheaval, an unparalleled augmentation that will leave little unchanged. Starting to crash around us is a new wave of technology. This wave is unleashing the power to engineer these two universal foundations: a wave of nothing less than intelligence and life.

The coming wave is defined by two core technologies: artificial intelligence (AI) and synthetic biology. Together they will usher in a new dawn for humanity, creating wealth and surplus unlike anything ever seen. And yet their rapid proliferation also threatens to empower a diverse array of bad actors to unleash disruption, instability, and even catastrophe on an unimaginable scale. This wave creates an immense challenge that will define the twenty-first century: our future both depends on these technologies and is imperiled by them.

From where we stand today, it appears that containing this wave—that is, controlling, curbing, or even stopping it—is not possible. This book asks why that might be true and what it means if it is. The implications of these questions will ultimately affect everyone alive and every generation that follows us.

I believe this coming wave of technology is bringing human history to a turning point. If containing it is impossible, the consequences for our species are dramatic, potentially dire. Equally, without its fruits we are exposed and precarious. This is an argument I have made many times over the last decade behind closed doors, but as the impacts become ever more unignorable, it's time that I make the case publicly.

THE DILEMMA

Contemplating the profound power of human intelligence led me to ask a simple question, one that has consumed my life ever since: What if we could distill the essence of what makes us humans so productive

and capable into software, into an algorithm? Finding the answer might unlock unimaginably powerful tools to help tackle our most intractable problems. Here might be a tool, an impossible but extraordinary tool, to help us get through the awesome challenges of the decades ahead, from climate change to aging populations to sustainable food.

With this in mind, in a quaint Regency-era office overlooking London's Russell Square, I co-founded a company called DeepMind with two friends, Demis Hassabis and Shane Legg, in the summer of 2010. This was our goal, one that in retrospect still feels as ambitious and crazy and hopeful as it did back then: replicate the very thing that makes us unique as a species, our intelligence.

To achieve this objective, we would need to create a system that could imitate and then eventually outperform all human cognitive abilities, from vision and speech to planning and imagination, and ultimately empathy and creativity. Since such a system would benefit from the massively parallel processing of supercomputers and the explosion of vast new sources of data from across the open web, we knew that even modest progress toward this goal would have profound societal implications.

It certainly felt pretty far-out at the time. Back then, widespread adoption of artificial intelligence was the stuff of daydreams, more fantasy than fact, the province of a few cloistered academics and wild-eyed science fiction fans. But, as I write this and think back over the last decade, progress in AI has been nothing short of staggering. DeepMind became one of the world's leading AI companies, achieving a string of breakthroughs. The speed and power of this new revolution have been surprising even to those of us closest to its cutting edge. Over the writing of this book, the pace of progress in AI has been breathtaking, with new models and new products coming out every week, sometimes every day. It's clear this wave is accelerating.

Today, AI systems can almost perfectly recognize faces and objects. We take speech-to-text transcription and instant language translation for granted. AI can navigate roads and traffic well enough to drive autonomously in some settings. Based on a few simple prompts, a new generation of AI models can generate novel images and compose text

with extraordinary levels of detail and coherence. AI systems can produce synthetic voices with uncanny realism and compose music of stunning beauty. Even in more challenging domains, ones long thought to be uniquely suited to human capabilities like long-term planning, imagination, and simulation of complex ideas, progress leaps forward.

AI has been climbing the ladder of cognitive abilities for decades, and it now looks set to reach human-level performance across a very wide range of tasks within the next three years. That is a big claim, but if I'm even close to right, the implications are truly profound. What had, when we founded DeepMind, felt quixotic has become not just plausible but seemingly inevitable.

From the start, it was clear to me that AI would be a powerful tool for extraordinary good but, like most forms of power, one fraught with immense dangers and ethical dilemmas, too. I have long worried about not just the consequences of advancing AI but where the entire technological ecosystem was heading. Beyond AI, a wider revolution was underway, with AI feeding a powerful, emerging generation of genetic technologies and robotics. Further progress in one area accelerates the others in a chaotic and cross-catalyzing process beyond anyone's direct control. It was clear that if we or others were successful in replicating human intelligence, this wasn't just profitable business as usual but a seismic shift for humanity, inaugurating an era when unprecedented opportunities would be matched by unprecedented risks.

As the technology has progressed over the years, my concerns have grown. What if the wave is actually a tsunami?

IN 2010 ALMOST NO one was talking seriously about AI. Yet what had once seemed a niche mission for a small group of researchers and entrepreneurs has now become a vast global endeavor. AI is everywhere, on the news and in your smartphone, trading stocks and building websites. Many of the world's largest companies and wealthiest nations barrel forward, developing cutting-edge AI models and genetic engineering techniques, fueled by tens of billions of dollars in investment.

Once matured, these emerging technologies will spread rapidly, becoming cheaper, more accessible, and widely diffused throughout soci-

Copyrighted Material

ety. They will offer extraordinary new medical advances and clean energy breakthroughs, creating not just new businesses but new industries and quality of life improvements in almost every imaginable area.

And yet alongside these benefits, AI, synthetic biology, and other advanced forms of technology produce tail risks on a deeply concerning scale. They could present an existential threat to nation-states—risks so profound they might disrupt or even overturn the current geopolitical order. They open pathways to immense AI-empowered cyberattacks, automated wars that could devastate countries, engineered pandemics, and a world subject to unexplainable and yet seemingly omnipotent forces. The likelihood of each may be small, but the possible consequences are huge. Even a slim chance of outcomes like these requires urgent attention.

Some countries will react to the possibility of such catastrophic risks with a form of technologically charged authoritarianism to slow the spread of these new powers. This will require huge levels of surveillance along with massive intrusions into our private lives. Keeping a tight rein on technology could become part of a drift to everything and everyone being watched, all the time, in a dystopian global surveillance system justified by a desire to guard against the most extreme possible outcomes.

Equally plausible is a Luddite reaction. Bans, boycotts, and moratoriums will ensue. Is it even possible to step away from developing new technologies and introduce a series of moratoriums? Unlikely. With their enormous geostrategic and commercial value, it's difficult to see how nation-states or corporations will be persuaded to unilaterally give up the transformative powers unleashed by these breakthroughs. Moreover, attempting to ban development of new technologies is itself a risk: technologically stagnant societies are historically unstable and prone to collapse. Eventually, they lose the capacity to solve problems, to progress.

Both pursuing and not pursuing new technologies is, from here, fraught with risk. The chances of muddling through a "narrow path" and avoiding one or the other outcome—techno-authoritarian dystopia on the one hand, openness-induced catastrophe on the other—grow smaller over time as the technology becomes cheaper, more

powerful, and more pervasive and the risks accumulate. And yet stepping away is no option either. Even as we worry about their risks, we need the incredible benefits of the technologies of the coming wave more than ever before. This is the core dilemma: that, sooner or later, a powerful generation of technology leads humanity toward either catastrophic or dystopian outcomes. I believe this is the great meta-problem of the twenty-first century.

This book outlines exactly why this terrible bind is becoming inevitable and explores how we might confront it. Somehow we need to get the best out of technology, something essential to facing a daunting set of global challenges, and also get out of the dilemma. The current discourse around technology ethics and safety is inadequate. Despite the many books, debates, blog posts, and tweetstorms about technology, you rarely hear anything about *containing* it. I see this as an interlocking set of technical, social, and legal mechanisms constraining and controlling technology working at every possible level: a means, in theory, of evading the dilemma. Yet even technology's harshest critics tend to dodge this language of hard containment.

That needs to change; I hope this book shows why, and hints at how.

THE TRAP

A few years after we founded DeepMind, I created a slide deck about AI's potential long-term economic and social impacts. Presenting to a dozen of the tech industry's most influential founders, CEOs, and technologists in a sleek West Coast boardroom, I argued that AI introduced a host of threats requiring proactive responses. It might lead to massive invasions of privacy or ignite a misinformation apocalypse. It might be weaponized, creating a lethal suite of new cyberweapons, introducing new vulnerabilities into our networked world.

I also underscored AI's potential to put large numbers of people out of work. I asked the room to consider automation and mechanization's long history of displacing labor. First come more efficient ways of doing specific tasks, and then entire roles become redundant, and soon entire sectors require orders of magnitude fewer workers. Over the next few decades, I argued, AI systems would replace "intellectual

manual labor” in much the same way, and certainly long before robots replace physical labor. In the past, new jobs were created at the same time as old ones were made obsolete, but what if AI could simply do most of those as well? There was, I suggested, little precedent for the new forms of concentrated power that were coming. Even though they felt distant, potentially grave threats were hurtling toward society.

In the concluding slide I showed a still from *The Simpsons*. In the scene, the townspeople of Springfield have risen up, and the cast of familiar characters charges forward carrying clubs and torches. The message was clear, but I spelled it out anyway. “The pitchforks are coming,” I said. Coming for us, the makers of technology. It was up to us to ensure the future was better than this.

Around the table, I was met with blank stares. The room was unmoved. The message didn’t land. Dismissals came thick and fast. Why didn’t economic indicators show any sign of what I was saying? AI would spur new demand, which would create new jobs. It would augment and empower people to be even more productive. Maybe there were some risks, they conceded, but they weren’t too bad. People were smart. Solutions have always been found. *No worries*, they seemed to think, *on to the next presentation*.

Some years later, a short time before the onset of the COVID-19 pandemic, I attended a seminar on technology risks at a well-known university. The setup was similar: another large table, another high-minded discussion. Over the course of the day a series of hair-raising risks were floated over the coffees, biscuits, and PowerPoints.

One stood out. The presenter showed how the price of DNA synthesizers, which can print bespoke strands of DNA, was falling rapidly. Costing a few tens of thousands of dollars, they are small enough to sit on a bench in your garage and let people synthesize—that is, *manufacture*—DNA. And all this is now possible for anyone with graduate-level training in biology or an enthusiasm for self-directed learning online.

Given the increasing availability of the tools, the presenter painted a harrowing vision: Someone could soon create novel pathogens far more transmissible and lethal than anything found in nature. These

synthetic pathogens could evade known countermeasures, spread asymptotically, or have built-in resistance to treatments. If needed, someone could supplement homemade experiments with DNA ordered online and reassembled at home. The apocalypse, mail ordered.

This was not science fiction, argued the presenter, a respected professor with more than two decades of experience; it was a live risk, now. They finished with an alarming thought: a single person today likely “has the capacity to kill a billion people.” All it takes is motivation.

The attendees shuffled uneasily. People squirmed and coughed. Then the griping and hedging started. No one wanted to believe this was possible. Surely it wasn’t the case, surely there had to be some effective mechanisms for control, surely the diseases were difficult to create, surely the databases could be locked down, surely the hardware could be secured. And so on.

The collective response in the seminar was more than just dismissive. People simply refused to accept the presenter’s vision. No one wanted to confront the implications of the hard facts and cold probabilities they’d heard. I stayed silent, frankly shaken. Soon the seminar was done. That evening we all went out for dinner and carried on chatting as normal. We’d just had a day of talking about the end of the world, but there was still pizza to eat, jokes to tell, an office to get back to, and besides, something would turn up, or some part of the argument was bound to be wrong. I joined in.

But the presentation gnawed at me for months afterward. Why wasn’t I, why weren’t we all, taking it more seriously? Why do we awkwardly sidestep further discussion? Why do some get snarky and accuse people who raise these questions of catastrophizing or of “overlooking the amazing good” of technology? This widespread emotional reaction I was observing is something I have come to call the pessimism-aversion trap: the misguided analysis that arises when you are overwhelmed by a fear of confronting potentially dark realities, and the resulting tendency to look the other way.

Pretty much everyone has some version of this reaction, and the consequence is that it’s leading us to overlook a number of critical

trends unfolding right before our eyes. It's almost an innate physiological response. Our species is not wired to truly grapple with transformation at this scale, let alone the potential that technology might fail us in this way. I've experienced this feeling throughout my career, and I've seen many, many others have the same visceral response. Confronting this feeling is one of the purposes of this book. To take a cold hard look at the facts, however uncomfortable.

Properly addressing this wave, containing technology, and ensuring that it always serves humanity means overcoming pessimism aversion. It means facing head-on the reality of what's coming.

THIS BOOK IS MY attempt to do that. To acknowledge and illuminate the contours of the coming wave. To explore whether containment is possible. To put things in historical context, and see the wider picture by stepping back from the daily firehouse chatter around tech. My aim is to confront the dilemma and understand the underlying processes that drive the emergence of science and technology. I want to present these ideas as clearly as I can to the widest possible audience. I've written it in a spirit of openness and inquiry: make observations, follow their implications, but also remain open to refutation and better interpretations. There is nothing I want more than to be proven wrong here, than for containment to be readily possible.

Some people may understandably expect a more techno-utopian book from someone like me, a founder of two AI companies. As a technologist and entrepreneur, I am, by default, an optimist. As a young teenager, I remember being totally captivated after installing Netscape for the first time on my Packard Bell 486 PC. I was entranced by the whirring fans and the distorted whistling of my 56 Kbps dial-up modem reaching its hand out to the World Wide Web and connecting me to forums and chat rooms that gave me freedom and taught me so much. I love technology. It's been *the* engine of progress and a cause for us to be proud and excited about humanity's achievements.

But I also believe that those of us driving technology's creation must have the courage to predict—and take responsibility for—where it might take us in decades to come. We must begin to suggest what to

do if it looks like there is a real risk that technology fails us. What's required is a societal and political response, not merely individual efforts, but it needs to begin with my peers and me.

Some will argue this is all overblown. That change is far more incremental. That it is just another turn of the hype cycle. That systems for coping with crises and change are actually quite robust. That my view of human nature is far too dark. That humanity's record is, well, so far, so good. History is full of false prophets and doomsayers proven wrong. Why should this time be different?

Pessimism aversion is an emotional response, an ingrained gut refusal to accept the possibility of seriously destabilizing outcomes. It tends to come from those in secure and powerful positions with entrenched worldviews, people who can superficially cope with change but struggle to accept any real challenge to their world order. Many of those whom I accuse of being stuck in the pessimism-aversion trap fully embrace the growing critiques of technology. But they nod along without actually taking any action. We'll manage, we always do, they say.

Spend time in tech or policy circles, and it quickly becomes obvious that head-in-the-sand is the default ideology. To believe and act otherwise risks becoming so crippled by fear of and outrage against enormous, inexorable forces that everything feels futile. So the strange intellectual half-world of pessimism aversion rumbles on. I should know, I was stuck in it for too long.

In the years since we founded DeepMind and since those presentations, the discourse has changed—to some extent. The job automation debate has been rehearsed countless times. A global pandemic showcased both the risks and the potency of synthetic biology. A “techlash” of sorts emerged, with critics railing against tech and tech companies in op-eds and books, in the regulatory capitals of Washington, Brussels, and Beijing. Previously niche fears around technology exploded into the mainstream, public skepticism of technology increased, and criticisms from academia, civil society, and politics sharpened.

And yet in the face of the coming wave and the great dilemma, and in the face of a pessimism-averse techno-elite, none of this is enough.

Copyrighted Material

THE ARGUMENT

Waves are everywhere in human life. This one is just the latest. Often people seem to think it's still far off, so futuristic and absurd-sounding that it's just the province of a few nerds and fringe thinkers, more hyperbole, more technobabble, more boosterism. That's a mistake. This is real, as real as the tsunami that comes out of the open blue ocean.

This isn't just fantasy or a chin-stroking intellectual exercise. Even if you disagree with my framing and think none of this is likely, I urge you to read on. Yes, I come with an AI background and am primed to view the world through a technological lens. I *am* biased when it comes to the question of whether this matters. Nonetheless, having been up close to this unfurling revolution over the last decade and a half, I am convinced we're on the cusp of the most important transformation of our lifetimes.

As a builder of these technologies, I believe they can deliver an extraordinary amount of good, change countless lives for the better, and address fundamental challenges, from helping unlock the next generation of clean energy to producing cheap and effective treatments for our most intractable medical conditions. Technologies can and should enrich our lives; historically, it bears repeating, the inventors and entrepreneurs behind them have been powerful drivers of progress, improving living standards for billions of us.

But without containment, every other aspect of technology, every discussion of its ethical shortcomings, or the benefits it could bring, is inconsequential. We urgently need watertight answers for how the coming wave can be controlled and contained, how the safeguards and affordances of the democratic nation-state can be maintained, but right now no one has such a plan. This is a future that none of us want, but it's one I fear is increasingly likely, and I will explain why in the chapters that follow.

In part 1, we look at the long history of technology and how it spreads—waves building over millennia. What drives them? What makes them truly general? We also ask if there are examples of societies consciously saying no to a new technology. Instead of turning away

from technologies, the past is marked by a pronounced pattern of proliferation, resulting in sprawling chains of both intended and unintended consequences.

I call this “the containment problem.” How do we keep a grip on the most valuable technologies ever invented as they get cheaper and spread faster than any in history?

Part 2 gets into the details of the coming wave itself. At its heart lie two general-purpose technologies of immense promise, power, and peril: artificial intelligence and synthetic biology. Both have been long heralded, and yet, if anything, I believe the scope of their impact is still often understated. Around them grow a host of associated technologies like robotics and quantum computing whose development will intersect in complex and turbulent ways.

In this section, we look at not only how they all emerged and what they can do but also why they are so hard to contain. The various technologies I’m speaking of share four key features that explain why this isn’t business as usual: they are inherently general and therefore omniscient, they hyper-evolve, they have asymmetric impacts, and, in some respects, they are increasingly autonomous.

Their creation is driven by powerful incentives: geopolitical competition, massive financial rewards, and an open, distributed culture of research. Scores of state and non-state actors will race ahead to develop them regardless of efforts to regulate and control what’s coming, taking risks that affect everyone, whether we like it or not.

Part 3 explores the political implications of a colossal redistribution of power engendered by an uncontained wave. The foundation of our present political order—and the most important actor in the containment of technologies—is the nation-state. Already rocked by crises, it will be further weakened by a series of shocks amplified by the wave: the potential for new forms of violence, a flood of misinformation, disappearing jobs, and the prospect of catastrophic accidents.

Further out, the wave will force a set of tectonic shifts in power, both centralizing and decentralizing at the same time. This will create vast new enterprises, buttress authoritarianism, and yet also empower groups and movements to live outside traditional social structures. The

delicate bargain of the nation-state will be placed under immense strain just when we need institutions like it most. This is how we end up in the dilemma.

In part 4 the discussion moves to what we can do about it. Is there even a slim chance for containment, for wriggling out of the dilemma? If so, how? In this section we outline ten steps, working out from the level of code and DNA to the level of international treaties, forming a hard, nested set of constraints, an outline plan for containment.

THIS IS A BOOK about confronting failure. Technologies can fail in the mundane sense of not working: the engine doesn't start; the bridge falls down. But they can also fail in a wider sense. If technology damages human lives, or produces societies filled with harm, or renders them ungovernable because we empower a chaotic long tail of bad (or unintentionally dangerous) actors—if, in the aggregate, technology is damaging—then it can be said to have failed in another, deeper sense, failing to live up to its promise. Failure in this sense isn't intrinsic to technology; it is about the context within which it operates, the governance structures it is subject to, the networks of power and uses to which it is put.

That impressive ingenuity giving rise to so much now means we are better at avoiding the first kind of failure. Fewer planes crash, cars are cleaner and safer, computers are more powerful and yet more secure. Our great challenge is that we still haven't reckoned with the latter mode of failure.

Over centuries, technology has dramatically increased the well-being of billions of people. We are immeasurably healthier thanks to modern medicine, the majority of the world lives in food abundance, people have never been more educated, more peaceful, or more materially comfortable. These are defining achievements produced in part by that great motor of humanity: science and the creation of technology. It's why I have devoted my life to safely developing these tools.

But any optimism we take from this extraordinary history must be grounded in blunt reality. Guarding against failure means understanding and ultimately confronting what can go wrong. We need to follow the chain of reasoning to its logical end point, without fear of where

that might lead, and, as we get there, do something about it. The coming wave of technologies threatens to fail faster and on a wider scale than anything witnessed before. This situation needs worldwide, popular attention. It needs answers, answers that no one yet has.

Containment is not, on the face of it, possible. And yet for all our sakes, containment *must* be possible.

Copyrighted Material