Superintelligence: Paths, Dangers, Strategies

Anthropic Bias explores how to reason when you suspect that your evidence is biased by "observation selection effects"--that is, evidence that has been filtered by the precondition that there be some suitably positioned observer to "have" the evidence. This conundrum--sometimes alluded to as "the anthropic principle," "self-locating belief," or "indexical information"--turns out to be a surprisingly perplexing and intellectually stimulating challenge, one abounding with important implications for many areas in science and philosophy. There are the philosophical thought experiments and paradoxes: the Doomsday Argument; Sleeping Beauty; the Presumptuous Philosopher; Adam & Eve; the Absent-Minded Driver; the Shooting Room. And there are the applications in contemporary science: cosmology ("How many universes are there?", "Why does the universe appear fine-tuned for life?"); evolutionary theory ("How improbable was the evolution of intelligent life on our planet?"); the problem of time's arrow ("Can it be given a thermodynamic explanation?"); quantum physics ("How can the many-worlds theory be tested?"); game-theory problems with imperfect recall ("How to model them?"); even traffic analysis ("Why is the 'next lane' faster?"). Anthropic Bias argues that the same principles are at work across all these domains. And it offers a synthesis: a mathematically explicit theory of observation selection effects that attempts to meet scientific needs while steering clear of philosophical paradox.

As we approach a great turning point in history when technology is poised to redefine what it means to be human, The Fourth Age offers fascinating insight into AI, robotics, and their extraordinary implications for our species. "If you only read just one book about the AI revolution, make it this one" (John Mackey, cofounder and CEO, Whole Foods Market). In The Fourth Age, Byron Reese makes the case that technology has reshaped humanity just three times in history: 100,000 years ago, we harnessed fire, which led to language; 10,000 years ago, we developed agriculture, which led to cities and warfare; 5,000 years ago, we invented the wheel and writing, which lead to the nation state. We are now on the doorstep of a fourth change brought about by two technologies: AI and robotics. "Timely, highly informative, and certainly optimistic" (Booklist), The Fourth Age provides an essential background on how we got to this point, and how—rather than what—we should think about the topics we'll soon all be facing: machine consciousness, automation, changes in employment, creative computers, radical life extension, artificial life, AI ethics, the future of warfare, superintelligence, and the implications of extreme prosperity. By asking questions like "Are you a machine?" and "Could a computer feel anything?", Reese leads you through a discussion along the cutting edge in robotics and AI, and provides a framework by which we can all understand, discuss, and act on the issues of the Fourth Age and how they'll transform humanity.

A leading artificial intelligence researcher lays out a new approach to AI that will enable us to coexist successfully with increasingly intelligent machines In the popular imagination, superhuman artificial intelligence is an approaching tidal wave that threatens not just jobs and human relationships, but civilization itself. Conflict between humans and machines is seen as inevitable and its outcome all too predictable. In this groundbreaking book, distinguished AI researcher Stuart Russell argues that this scenario can be avoided, but only if we rethink AI from the ground up. Russell begins by exploring the idea of intelligence in humans and in machines. He describes the near-term benefits we can expect, from intelligent personal assistants to vastly accelerated scientific research, and outlines the AI breakthroughs that still have to happen before we reach superhuman AI. He also spells out the ways humans are already finding to misuse AI, from lethal autonomous weapons to viral sabotage. If the predicted breakthroughs occur and superhuman AI emerges, we will have created entities far more powerful than ourselves. How can we ensure they never, ever, have power over us? Russell suggests that we can rebuild AI on a new foundation, according to which machines are designed to be inherently uncertain about the human preferences they are required to satisfy. Such machines would be humble, altruistic, and committed to pursue our objectives, not theirs. This new foundation would allow us to create machines that are provably deferential and provably beneficial.

A day does not go by without a news article reporting some amazing breakthrough in artificial intelligence (AI). Many philosophers, futurists, and AI researchers have conjectured that human-level AI will be developed in the next 20 to 200 years. If these predictions are correct, it raises new and sinister issues related to our future in the age of intelligent machines. Artificial Superintelligence: A Futuristic Approach directly addresses these issues and consolidates research aimed at making sure that emerging superintelligence is beneficial to humanity. While specific predictions regarding the consequences of superintelligent AI vary from potential economic hardship to the complete extinction of humankind, many researchers agree that the issue is of utmost importance and needs to be seriously addressed. Artificial Superintelligence: A Futuristic Approach discusses key topics such as: AI-Completeness theory and how it can be used to see if an artificial intelligent agent has attained human level intelligence Methods for safeguarding the invention of a superintelligent system that could theoretically be worth trillions of dollars Self-improving AI systems: definition, types, and limits The science of AI safety engineering, including machine ethics and robot rights Solutions for ensuring safe and secure confinement of superintelligent systems The future of superintelligence and why long-term prospects for humanity to remain as the dominant species on Earth are not great Artificial Superintelligence: A Futuristic Approach is designed to become a foundational text for the new science of AI safety engineering. AI researchers and students, computer security researchers, futurists, and philosophers should find this an invaluable resource.

Superintelligence

Hey, Cyba

Smart Robots, Conscious Computers, and the Future of Humanity

Artificial Superintelligence

A Rough Ride to the Future

Paths, Dangers, Strategies

The truth about AI from the people building it

Reveals how AI works and provides insight into what we can expect of it now and in the future.

A provocative attempt to think about what was previously considered unthinkable: a serious philosophical case for the rights of robots. We are in the midst of a robot invasion, as devices of different configurations and capabilities slowly but surely come to take up increasingly important positions in everyday social reality—self-driving vehicles, recommendation algorithms, machine learning decision making systems, and social robots of various forms and functions. Although considerable attention has already been devoted to the subject of robots and responsibility, the question concerning the social status of these artifacts has been largely overlooked. In this book, David Gunkel offers a provocative attempt to think about what has been previously regarded as unthinkable: whether and to what extent robots and other technological artifacts of our own making can and should have any claim to moral and legal standing. In his analysis, Gunkel invokes the philosophical distinction (developed by David Hume) between "is" and "ought" in order to evaluate and analyze the different arguments regarding the question of robot rights. In the course of his examination, Gunkel finds that none of the existing positions or proposals hold up under scrutiny. In response to this, he then offers an innovative alternative proposal that effectively flips the script on the is/ought problem by introducing another, altogether different way to conceptualize the social situation of robots and the opportunities and challenges they present to existing moral and legal systems.

'if AI is outside your field, or you know something of the subject and would like to know more then Artificial Intelligence: The Basics is a

brilliant primer.' - Nick Smith, Engineering and Technology Magazine November 2011 Artificial Intelligence: The Basics is a concise and cutting-edge introduction to the fast moving world of AI. The author Kevin Warwick, a pioneer in the field, examines issues of what it means to be man or machine and looks at advances in robotics which have blurred the boundaries. Topics covered include: how intelligence can be defined whether machines can 'think' sensory input in machine systems the nature of consciousness the controversial culturing of human neurons. Exploring issues at the heart of the subject, this book is suitable for anyone interested in AI, and provides an illuminating and accessible introduction to this fascinating subject.

SuperintelligencePaths, Dangers, StrategiesOxford University Press (UK)

The Big Picture

Life 3.0

Applied Artificial Intelligence

Building Artificial Intelligence We Can Trust

How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines

Robot Rights

The Fourth Age

Singularity Hypotheses: A Scientific and Philosophical Assessment offers authoritative, jargon-free essays and critical commentaries on accelerating technological progress and the notion of technological singularity. It focuses on conjectures about the intelligence explosion, transhumanism, and whole brain emulation. Recent years have seen a plethora of forecasts about the profound, disruptive impact that is likely to result from further progress in these areas. Many commentators however doubt the scientific rigor of these forecasts, rejecting them as speculative and unfounded. We therefore invited prominent computer scientists, physicists, philosophers, biologists, economists and other thinkers to assess the singularity hypotheses. Their contributions go beyond speculation, providing deep insights into the main issues and a balanced picture of the debate.

Two leaders in the field offer a compelling analysis of the current state of the art and reveal the steps we must take to achieve a truly robust artificial intelligence. Despite the hype surrounding AI, creating an intelligence that rivals or exceeds human levels is far more complicated than we have been led to believe. Professors Gary Marcus and Ernest Davis have spent their careers at the forefront of AI research and have witnessed some of the greatest milestones in the field, but they argue that a computer beating a human in Jeopardy! does not signal that we are on the doorstep of fully autonomous cars or superintelligent machines. The achievements in the field thus far have occurred in closed systems with fixed sets of rules, and these approaches are too narrow to achieve genuine intelligence. The real world, in contrast, is wildly complex and open-ended. How can we bridge this gap? What will the consequences be when we do? Taking inspiration from the human mind, Marcus and Davis explain what we need to advance AI to the next level, and suggest that if we are wise along the way, we won't need to worry about a future of machine overlords. If we focus on endowing machines with common sense and deep understanding, rather than simply focusing on statistical analysis and gathering ever larger collections of data, we will be able to create an AI we can trust—in our homes, our cars, and our doctors' offices. Rebooting AI provides a lucid, clear-eyed assessment of the current science and offers an inspiring vision of how a new generation of AI can make our lives better.

In The Ultimate Situational Survival Guide, survival expert Robert Richardson, founder of offgridsurvival.com, gives you real world advice on how to survive the very real dangers present in today's society. From surviving natural disasters, man-made disasters and disease outbreaks, to essential tactics and step-by-step instructions for surviving urban disasters, crime, violence and terrorist attacks, readers will learn the self-reliance strategies they need to survive in just about any situation.This is not your typical survival manual or the same old tired material that's been regurgitated time and time again. This is a self-reliance guide to surviving 21st century threats, including ways to survive cyber-attacks, infrastructure shutdowns and communication grid failures.

Book Description How will AI evolve and what major innovations are on the horizon? What will its impact be on the job market, economy, and society? What is the path toward human-level machine intelligence? What should we be concerned about as artificial intelligence advances? Architects of Intelligence contains a series of in-depth, one-to-one interviews where New York Times bestselling author, Martin Ford, uncovers the truth behind these questions from some of the brightest minds in the Artificial Intelligence community. Martin has wide-ranging conversations with twenty-three of the world's foremost researchers and entrepreneurs working in AI and robotics: Demis Hassabis (DeepMind), Ray Kurzweil (Google), Geoffrey Hinton (Univ. of Toronto and Google), Rodney Brooks (Rethink Robotics), Yann LeCun (Facebook) , Fei-Fei Li (Stanford and Google), Yoshua Bengio (Univ. of Montreal), Andrew Ng (AI Fund), Daphne Koller (Stanford), Stuart Russell (UC Berkeley), Nick Bostrom (Univ. of Oxford), Barbara Grosz (Harvard), David Ferrucci (Elemental Cognition), James Manyika (McKinsey), Judea Pearl (UCLA), Josh Tenenbaum (MIT), Rana el Kaliouby (Affectiva), Daniela Rus (MIT), Jeff Dean (Google), Cynthia Breazeal (MIT), Oren Etzioni (Allen Institute for AI), Gary Marcus (NYU), and Bryan Johnson (Kernel). Martin Ford is a prominent futurist, and author of Financial Times Business Book of the Year, Rise of the Robots. He speaks at conferences and companies around the world on what AI and automation might mean for the future.

When Humans Transcend Biology

The AI Does Not Hate You

The Cambridge Handbook of Artificial Intelligence

Rebooting AI

The Ultimate Situational Survival Guide

Human Enhancement

A Global Catastrophic Risk is one that has the potential to inflict serious damage to human well-being on a global scale. This book focuses on such risks arising from natural catastrophes (Earth-based or beyond), nuclear war, terrorism, biological weapons, totalitarianism, advanced nanotechnology, artificial intelligence and social collapse.

The idea of technological singularity, and what it would mean if ordinary human intelligence were enhanced or overtaken by artificial intelligence. The idea that human history is approaching a "singularity"—that ordinary humans will someday be overtaken by artificially intelligent machines or cognitively enhanced biological intelligence, or both—has moved from the realm of science fiction to

serious debate. Some singularity theorists predict that if the field of artificial intelligence (AI) continues to develop at its current dizzying rate, the singularity could come about in the middle of the present century. Murray Shanahan offers an introduction to the idea of the singularity and considers the ramifications of such a potentially seismic event. Shanahan's aim is not to make predictions but rather to investigate a range of scenarios. Whether we believe that singularity is near or far, likely or impossible, apocalypse or utopia, the very idea raises crucial philosophical and pragmatic questions, forcing us to think seriously about what we want as a species. Shanahan describes technological advances in AI, both biologically inspired and engineered from scratch. Once human-level AI—theoretically possible, but difficult to accomplish—has been achieved, he explains, the transition to superintelligent AI could be very rapid. Shanahan considers what the existence of superintelligent machines could mean for such matters as personhood, responsibility, rights, and identity. Some superhuman AI agents might be created to benefit humankind; some might go rogue. (Is Siri the template, or HAL?) The singularity presents both an existential threat to humanity and an existential opportunity for humanity to transcend its limitations. Shanahan makes it clear that we need to imagine both possibilities if we want to bring about the better outcome.

'Beautifully written, and with wonderful humour, this is a thrilling adventure story of our own future' Lewis Dartnell, author of The Knowledge and Origins 'The AI does not hate you, nor does it love you, but you are made of atoms which it can use for something else' This is a book about AI and AI risk. But it's also more importantly about a community of people who are trying to think rationally about intelligence, and the places that these thoughts are taking them, and what insight they can and can't give us about the future of the human race over the next few years. It explains why these people are worried, why they might be right, and why they might be wrong. It is a book about the cutting edge of our thinking on intelligence and rationality right now by the people who stay up all night worrying about it. Along the way, we discover why we probably don't need to worry about a future AI resurrecting a perfect copy of our minds and torturing us for not inventing it sooner, but we perhaps should be concerned about paperclips destroying life as we know it; how Mickey Mouse can teach us an important lesson about how to program AI; and how a more rational approach to life could be what saves us all.

Attention in the AI safety community has increasingly started to include strategic considerations of coordination between relevant actors in the field of AI and AI safety, in addition to the steadily growing work on the technical considerations of building safe AI systems. This shift has several reasons: Multiplier effects, pragmatism, and urgency. Given the benefits of coordination between those working towards safe superintelligence, this book surveys promising research in this emerging field regarding AI safety. On a meta-level, the hope is that this book can serve as a map to inform those working in the field of AI coordination about other promising efforts. While this book focuses on AI safety coordination, coordination is important to most other known existential risks (e.g., biotechnology risks), and future, human-made existential risks. Thus, while most coordination strategies in this book are specific to superintelligence, we hope that some insights yield "collateral benefits" for the reduction of other existential risks, by creating an overall civilizational framework that increases robustness, resiliency, and antifragility.

Coordination & Strategy
Artificial Intelligence and the End of the Human Era
Singularity Hypotheses
Entering Space
The Quest for Artificial Intelligence
A Scientific and Philosophical Assessment
The Power to Predict Who Will Click, Buy, Lie, or Die

The volume advances research in the philosophy of technology by introducing contributors who have an acute sense of how to get beyond or reframe the epistemic, ontological and normative limitations that currently limit the fields of philosophy of technology and science and technology studies.

'Fascinating' – Brian Cox, Mail on Sunday Books of the Year Where are we? Who are we? Do our beliefs, hopes and dreams hold any significance out there in the void? Can human purpose and meaning ever fit into a scientific worldview? Award-winning author Sean Carroll brings his extraordinary intellect to bear on the realms of knowledge, the laws of nature and the most profound questions about life, death and our place in it all. From Darwin and Einstein to the origins of life, consciousness and the universe itself, Carroll combines cosmos-sprawling science and profound thought in a quest to explain our world. Destined to sit alongside the works of our greatest thinkers, The Big Picture demonstrates that while our lives may be forever dwarfed by the immensity of the universe, they can be redeemed by our capacity to comprehend it and give it meaning.

A Huffington Post Definitive Tech Book of 2013 In as little as a decade, artificial intelligence could match and then surpass human intelligence. Corporations and government agencies around the world are pouring billions into achieving AI's Holy Grail--human-level intelligence. Once AI has attained it, scientists argue, it will have survival drives much like our own. We may be forced to compete with a rival more cunning, more powerful, and more alien than we can imagine. Through profiles of tech visionaries, industry watchdogs, and groundbreaking AI systems, James Barrat's Our Final Invention explores the perils of the heedless pursuit of advanced AI. Until now, human intelligence has had no rival. Can we coexist with beings whose intelligence dwarfs our own? And will they allow us to?

This bestselling book gives business leaders and executives a foundational education on how to leverage artificial intelligence and machine learning solutions to deliver ROI for your business.

How Can Machines Learn Human Values?
Anthropic Bias
Artificial Intelligence and the Problem of Control

On Intelligence
Architects of Intelligence
Twenty-Five Ways of Looking at AI
Our Final Invention

'This is the most important conversation of our time, and Tegmark's thought-provoking book will help you join it' Stephen Hawking THE INTERNATIONAL BESTSELLER. DAILY TELEGRAPH AND THE TIMES BOOKS OF THE YEAR AI is the future - but what will that future look like? Will superhuman intelligence be our slave, or become our god? Taking us to the heart of the latest thinking about AI, Max Tegmark, the MIT professor whose work has helped mainstream research on how to keep AI beneficial, separates myths from reality, utopias from dystopias, to explore the next phase of our existence. How can we grow our prosperity through automation, without leaving people lacking income or purpose? How can we ensure that future AI systems do what we want without crashing, malfunctioning or getting hacked? Should we fear an arms race in lethal autonomous weapons? Will AI help life flourish as never before, or will machines eventually outsmart us at all tasks, and even, perhaps, replace us altogether? 'This is a rich and visionary book and everyone should read it' The Times
Learn About The Future Of Artificial Intelligence In A Fraction Of The Time It Takes To Read The Actual Book!!! Today only, get this 1# Amazon bestseller for just $2.99. Regularly priced at $9.99. Read on your PC, Mac, smart phone, tablet or Kindle device Inside your cranium is the thing that allows you to read, your brain. Animals have other abilities like knifelike claws and powerful muscles. But our brain has let us create a system for verbal communication, science, electronics, and intimate public arrangement. Each generation has done better and progressed farther than the previous generation.We have the dominance, because we can build the things. We could build a superintelligence that could safeguard human values. But we'd only get one chance, because if the superintelligence became unfriendly, getting rid of it or changing it would be next to impossible.It seems possible that sometime soon there could be an artificial intelligence advancement. And a couple chapters of this book are devoted to possible pathways to that. But the majority of the book is devoted to what happens next. The powers of the superintelligence, the decisive choices available. Then how do we mold the conditions to get a survivable and favorable outcome. Towards the end we look at the big picture and how to avoid catastrophe.There may be things in this book Bostrom fails to take into account, and he may draw some wrong conclusions. There is uncertainty and it is expressed when necessary.Here Is A Preview Of What You'll Learn When You Download Your Copy Today * How Artificial Intelligence Works And The Way It Will Change The Future * The Reason Why It Would Be Difficult For One Organization To Dominate The Artificial Intelligence Industry * Learn How The World Needs To Work Together In Order To Create A Safe And Responsible Form Of Artificial Intelligence Download Your Copy Today! The contents of this book are easily worth over $9.99, but for a limited time you can download the summary of Nick Bostrom's "Superintelligence" by for a special discounted price of only $2.99
Global capital markets have undergone fundamental transformations in recent years and, as a result, have become extraordinarily complex and opaque. Trading space is no longer measured in minutes or seconds but in time units beyond human perception: milliseconds, microseconds, and even nanoseconds. Technological advances have thus scaled up imperceptible and previously irrelevant time differences into operationally manageable and enormously profitable business opportunities for those with the proper high-tech trading tools. These tools include the fastest private communication and trading lines, the most powerful computers and sophisticated algorithms capable of speedily analysing incoming news and trading data and determining optimal trading strategies in microseconds, as well as the possession of gigantic collections of historic and real-time market data. Fragmented capital markets are also becoming a rapidly growing reality in Europe and Asia, and are an established feature of U.S. trading. This raises urgent market governance issues that have largely been overlooked. Global Algorithmic Capital Markets seeks to understand how recent market transformations are affecting core public policy objectives such as investor protection and reduction of systemic risk, as well as fairness, efficiency, and transparency. The operation and health of capital markets affect all of us and have profound implications for equality and justice in society. This unique set of chapters by leading scholars, industry insiders, and regulators discusses ways to strengthen market governance for the benefit of society at whole.
"Mesmerizing & fascinating..." —The Seattle Post-Intelligencer "The Freakonomics of big data." —Stein Kretsinger, founding executive of Advertising.com Award-winning | Used by over 30 universities | Translated into 9 languages An introduction for everyone. In this rich, fascinating — surprisingly accessible — introduction, leading expert Eric Siegel reveals how predictive analytics (aka machine learning) works, and how it affects everyone every day. Rather than a "how to" for hands-on techies, the book serves lay readers and experts alike by covering new case studies and the latest state-of-the-art techniques. Prediction is booming. It reinvents industries and runs the world. Companies, governments, law enforcement, hospitals, and universities are seizing upon the power. These institutions predict whether you're going to click, buy, lie, or die. Why? For good reason: predicting human behavior combats risk, boosts

sales, fortifies healthcare, streamlines manufacturing, conquers spam, optimizes social networks, toughens crime fighting, and wins elections. How? Prediction is powered by the world's most potent, flourishing unnatural resource: data. Accumulated in large part as the by-product of routine tasks, data is the unsalted, flavorless residue deposited en masse as organizations churn away. Surprise! This heap of refuse is a gold mine. Big data embodies an extraordinary wealth of experience from which to learn. Predictive analytics (aka machine learning) unleashes the power of data. With this technology, the computer literally learns from data how to predict the future behavior of individuals. Perfect prediction is not possible, but putting odds on the future drives millions of decisions more effectively, determining whom to call, mail, investigate, incarcerate, set up on a date, or medicate. In this lucid, captivating introduction — now in its Revised and Updated edition — former Columbia University professor and Predictive Analytics World founder Eric Siegel reveals the power and perils of prediction: What type of mortgage risk Chase Bank predicted before the recession. Predicting which people will drop out of school, cancel a subscription, or get divorced before they even know it themselves. Why early retirement predicts a shorter life expectancy and vegetarians miss fewer flights. Five reasons why organizations predict death — including one health insurance company. How U.S. Bank and Obama for America calculated the way to most strongly persuade each individual. Why the NSA wants all your data: machine learning supercomputers to fight terrorism. How IBM's Watson computer used predictive modeling to answer questions and beat the human champs on TV's Jeopardy! How companies ascertain untold, private truths — how Target figures out you're pregnant and Hewlett-Packard deduces you're about to quit your job. How judges and parole boards rely on crime-predicting computers to decide how long convicts remain in prison. 182 examples from Airbnb, the BBC, Citibank, ConEd, Facebook, Ford, Google, the IRS, LinkedIn, Match.com, MTV, Netflix, PayPal, Pfizer, Spotify, Uber, UPS, Wikipedia, and more. How does predictive analytics work? This jam-packed book satisfies by demystifying the intriguing science under the hood. For future hands-on practitioners pursuing a career in the field, it sets a strong foundation, delivers the prerequisite knowledge, and whets your appetite for more. A truly omnipresent science, predictive analytics constantly affects our daily lives. Whether you are a consumer of it — or consumed by it — get a handle on the power of Predictive Analytics.

Being Human in the Age of Artificial Intelligence
Global Algorithmic Capital Markets
Superintelligence, Rationality and the Race to Save the World
The Alignment Problem
The Technological Singularity
Human Compatible
A Futuristic Approach

The essays in this book, written by researchers from both humanities and science, describe various theoretical and experimental approaches to adding medical ethics to a machine, what design features are necessary in order to achieve this, philosophical and practical questions concerning justice, rights, decision-making and responsibility in medical contexts, and accurately modeling essential physician-machine-patient relationships. In medical settings, machines are in close proximity with human beings: with patients who are in vulnerable states of health, who have disabilities of various kinds, with the very young or very old and with medical professionals. Machines in these contexts are undertaking important medical tasks that require emotional sensitivity, knowledge of medical codes, human dignity and privacy. As machine technology advances, ethical concerns become more urgent: should medical machines be programmed to follow a code of medical ethics? What theory or theories should constrain medical machine conduct? What design features are required? Should machines share responsibility with humans for the ethical consequences of medical actions? How ought clinical relationships involving machines to be modeled? Is a capacity for empathy and emotion detection necessary? What about consciousness? This collection is the first book that addresses these 21st-century concerns.

An authoritative, up-to-date survey of the state of the art in artificial intelligence, written for non-specialists.

'Vital reading. This is the book on artificial intelligence we need right now.' Mike Krieger, cofounder of Instagram Artificial intelligence is rapidly dominating every aspect of our modern lives influencing the news we consume, whether we get a mortgage, and even which friends wish us happy birthday. But as algorithms make ever more decisions on our behalf, how do we ensure they do what we want? And fairly? This conundrum - dubbed 'The Alignment Problem' by experts - is the subject of this timely and important book. From the AI program which cheats at computer games to the sexist algorithm behind Google Translate, bestselling author Brian Christian explains how, as AI develops, we rapidly approach a collision between artificial intelligence and ethics. If we stand by, we face a future with unregulated algorithms that propagate our biases - and worse - violate our most sacred values. Urgent and fascinating, this is an accessible primer to the most important issue facing AI researchers today.

In A Rough Ride to the Future, James Lovelock - the great scientific visionary of our age - presents a radical vision of humanity's future as the thinking brain of our Earth-system James Lovelock, who has been hailed as 'the man who conceived the first wholly new way of looking at life on earth since Charles Darwin' (Independent) and 'the most profound scientific thinker of our time' (Literary Review) continues, in his 95th year, to be the great scientific visionary of our age. This book introduces two new Lovelockian ideas. The first is that three hundred years ago, when Thomas Newcomen invented the steam engine, he was unknowingly beginning what Lovelock calls 'accelerated evolution', a process which is bringing about change on our planet roughly a million times faster than Darwinian evolution. The second is that as part of this process, humanity has the capacity to become the intelligent part of Gaia, the self-regulating Earth system whose discovery Lovelock first announced nearly 50 years ago. In addition, Lovelock gives his reflections on how scientific advances are made, and his own remarkable life as a lone scientist. The contribution of human beings to our planet is, Lovelock contends, similar to that of the early photosynthesisers around 3.4 billion years ago, which made the Earth's atmosphere what it was until very recently. By our domination and our invention, we are now changing the atmosphere again. There is little that can be done about this, but instead of feeling guilty about it we should recognise what is happening, prepare for change, and ensure that we survive as a species so we can contribute to -

perhaps even guide - the next evolution of Gaia. The road will be rough, but if we are smart enough life will continue on Earth in some form far into the future. Elected a Fellow of the Royal Society in 1974, JAMES LOVELOCK is the author of more than 200 scientific papers and the originator of the Gaia Hypothesis (now Gaia Theory). His many books on the subject include Gaia: A New Look at Life on Earth (1979), The Revenge of Gaia (2006), and The Vanishing Face of Gaia (2009). In 2003 he was made a Companion of Honour by Her Majesty the Queen, in 2005 Prospect magazine named him one of the world's top 100 public intellectuals, and in 2006 he received the Wollaston Medal, the highest Award of the UK Geological Society.

Summary and Analysis of Nick Bostroms "Superintelligence: Paths, Dangers, Strategies"
Possible Minds
Films from the Future
Global Catastrophic Risks
Self-Reliance Strategies for a Dangerous World
A Guide for Thinking Humans
The Inner Workings of a Virtual Personal Assistant

The human brain has some capabilities that the brains of other animals lack. It is to these distinctive capabilities that our species owes its dominant position. Other animals have stronger muscles or sharper claws, but we have cleverer brains. If machine brains one day come to surpass human brains in general intelligence, then this new superintelligence could become very powerful. As the fate of the gorillas now depends more on us humans than on the gorillas themselves, so the fate of our species then would come to depend on the actions of the machine superintelligence. But we have one advantage: we get to make the first move. Will it be possible to construct a seed AI or otherwise to engineer initial conditions so as to make an intelligence explosion survivable? How could one achieve a controlled detonation? To get closer to an answer to this question, we must make our way through a fascinating landscape of topics and considerations. Read the book and learn about oracles, genies, singletons; about boxing methods, tripwires, and mind crime; about humanity's cosmic endowment and differential technological development; indirect normativity, instrumental convergence, whole brain emulation and technology couplings; Malthusian economics and dystopian evolution; artificial intelligence, and biological cognitive enhancement; and collective intelligence.

The author of The Case for Mars provides an insider's look at the future of space exploration and travel, examining the true potential for human expeditions into outer space, the prospects for colonization of the outer planets of the solar system, and their implications for the future of humankind. Reprint.

"Startling in scope and bravado." —Janet Maslin, The New York Times "Artfully envisions a breathtakingly better world." —Los Angeles Times "Elaborate, smart and persuasive." —The Boston Globe "A pleasure to read." —The Wall Street Journal One of CBS News's Best Fall Books of 2005 • Among St Louis Post-Dispatch's Best Nonfiction Books of 2005 • One of Amazon.com's Best Science Books of 2005 A radical and optimistic view of the future course of human development from the bestselling author of How to Create a Mind and The Singularity is Nearer who Bill Gates calls "the best person I know at predicting the future of artificial intelligence" For over three decades, Ray Kurzweil has been one of the most respected and provocative advocates of the role of technology in our future. In his classic The Age of Spiritual Machines, he argued that computers would soon rival the full range of human intelligence at its best. Now he examines the next step in this inexorable evolutionary process: the union of human and machine, in which the knowledge and skills embedded in our brains will be combined with the vastly greater capacity, speed, and knowledge-sharing ability of our creations.

Artificial intelligence (AI) is a field within computer science that is attempting to build enhanced intelligence into computer systems. This book traces the history of the subject, from the early dreams of eighteenth-century (and earlier) pioneers to the more successful work of today's AI engineers. AI is becoming more and more a part of everyone's life. The technology is already embedded in face-recognizing cameras, speech-recognition software, Internet search engines, and health-care robots, among other applications. The book's many diagrams and easy-to-understand descriptions of AI programs will help the casual reader gain an understanding of how these and other AI systems actually work. Its thorough (but unobtrusive) end-of-chapter notes containing citations to important source materials will be of great use to AI scholars and researchers. This book promises to be the definitive history of a field that has captivated the imaginations of scientists, philosophers, and writers for centuries.

Artificial Intelligence: The Basics
New Waves in Philosophy of Technology
The Technology and Morality of Sci-Fi Movies
Predictive Analytics
Machine Medical Ethics
Fundamental Issues of Artificial Intelligence
Observation Selection Effects in Science and Philosophy

Science world luminary John Brockman assembles twenty-five of the most important scientific minds, people who been thinking about the field artificial intelligence for most of their careers, for an unparalleled round-table examina about mind, thinking, intelligence and what it means to be human. "Artificial intelligence is today's story--the story all other stories. It is the Second Coming and the Apocalypse at the same time: Good AI versus evil AI." --John Brockman More than sixty years ago, mathematician-philosopher Norbert Wiener published a book on the place of machines in society that ended with a warning: "we shall never receive the right answers to our questions unless the right questions.... The hour is very late, and the choice of good and evil knocks at our door." In the wake of adv in unsupervised, self-improving machine learning, a small but influential community of thinkers is considering Wiene words again. In Possible Minds, John Brockman gathers their disparate visions of where AI might be taking us. The of the long history of Brockman's profound engagement with the most important scientific minds who have been t about AI--from Alison Gopnik and David Deutsch to Frank Wilczek and Stephen Wolfram--Possible Minds is an ideal introduction to the landscape of crucial issues AI presents. The collision between opposing perspectives is salutary exhilarating; some of these figures, such as computer scientist Stuart Russell, Skype co-founder Jaan Tallinn, and physicist Max Tegmark, are deeply concerned with the threat of AI, including the existential one, while others, nota robotics entrepreneur Rodney Brooks, philosopher Daniel Dennett, and bestselling author Steven Pinker, have a very

different view. Serious, searching and authoritative, Possible Minds lays out the intellectual landscape of one of the important topics of our time.

From the inventor of the PalmPilot comes a new and compelling theory of intelligence, brain function, and the future intelligent machines Jeff Hawkins, the man who created the PalmPilot, Treo smart phone, and other handheld device has reshaped our relationship to computers. Now he stands ready to revolutionize both neuroscience and computing one stroke, with a new understanding of intelligence itself. Hawkins develops a powerful theory of how the human works, explaining why computers are not intelligent and how, based on this new theory, we can finally build intellig machines. The brain is not a computer, but a memory system that stores experiences in a way that reflects the tru structure of the world, remembering sequences of events and their nested relationships and making predictions ba those memories. It is this memory-prediction system that forms the basis of intelligence, perception, creativity, and consciousness. In an engaging style that will captivate audiences from the merely curious to the professional scien Hawkins shows how a clear understanding of how the brain works will make it possible for us to build intelligent machines, in silicon, that will exceed our human ability in surprising ways. Written with acclaimed science writer Sa Blakeslee, On Intelligence promises to completely transfigure the possibilities of the technology age. It is a landmar book in its scope and clarity.

The human brain has some capabilities that the brains of other animals lack. It is to these distinctive capabilities th species owes its dominant position. Other animals have stronger muscles or sharper claws, but we have cleverer b If machine brains one day come to surpass human brains in general intelligence, then this new superintelligence cou become very powerful. As the fate of the gorillas now depends more on us humans than on the gorillas themselves the fate of our species then would come to depend on the actions of the machine superintelligence. But we have advantage: we get to make the first move. Will it be possible to construct a seed AI or otherwise to engineer initia conditions so as to make an intelligence explosion survivable? How could one achieve a controlled detonation? To g closer to an answer to this question, we must make our way through a fascinating landscape of topics and considerations. Read the book and learn about oracles, genies, singletons; about boxing methods, tripwires, and min crime; about humanity's cosmic endowment and differential technological development; indirect normativity, instrum convergence, whole brain emulation and technology couplings; Malthusian economics and dystopian evolution; artific intelligence, and biological cognitive enhancement, and collective intelligence. This profoundly ambitious and original book picks its way carefully through a vast tract of forbiddingly difficult intellectual terrain. Yet the writing is so lu somehow makes it all seem easy. After an utterly engrossing journey that takes us to the frontiers of thinking abo human condition and the future of intelligent life, we find in Nick Bostrom's work nothing less than a reconceptuali of the essential task of our time.

This new book, by one of the most respected researchers in Artificial Intelligence, features a radical new 'evolutiona organization that begins with low level intelligent behavior and develops complex intelligence as the book progresse

Creating a Spacefaring Civilization

On the Origins of Life, Meaning, and the Universe Itself

The Singularity Is Near

A Handbook for Business Leaders

A New Synthesis

High Frequency Trading, Dark Pools, and Regulatory Challenges

Artificial Intelligence

This volume offers a look at the fundamental issues of present and future AI, especially from cognitive science, computer science, neuroscience and philosophy. This work examines the conditions for artificial intelligence, how these relate to the conditions for intelligence in humans and other natural agents, as well as ethical and societal problems that artificial intelligence raises or will raise. The key issues this volume investigates include the relation of AI and cognitive science, ethics of AI and robotics, brain emulation and simulation, hybrid systems and cyborgs, intelligence and intelligence testing, interactive systems, multi-agent systems, and super intelligence. Based on the 2nd conference on "Theory and Philosophy of Artificial Intelligence" held in Oxford, the volume includes prominent researchers within the field from around the world.

Hard Science Fiction Films that Predict the Future "As the breakneck advance of technology takes us into a world that is both exciting and menacing, sci-fi films give us an inkling of what is to come, and what we should avoid." —Seth Shostak, senior astronomer at the SETI Institute, and host of Big Picture Science #1 Best Seller in Nanotechnology and Computers & Technology Dr. Andrew Maynard, physicist and leading expert on socially responsible development of emerging and converging technologies, examines science fiction movies and brings them to life. Advances in science and technology are radically changing our world. Films from the Future is an essential guide to navigating a future dominated by complex and powerful new technologies. The jump from room-filling processors to pocket-size super computers is just the beginning. Artificial intelligence, gene manipulation, cloning, and inter-planet travel are all ideas that seemed like fairy tales but a few years ago. And now their possibility is very much here. But are we ready to handle these advances? As Maynard explains, "Viewed in the right way?and with a good dose of critical thinking?science fiction movies can help us think about and prepare for the social consequences of technologies we don't yet have, but that are coming faster than we imagine." Films from the Future looks at twelve movies that take readers on a journey through the worlds of biological and genetic manipulation, human enhancement, cyber technologies, and nanotechnology. Gain a broader understanding of the complex relationship between science and society. The movies include old and new, and the familiar and unfamiliar, to provide a unique, entertaining, and ultimately transformative take on the power and responsibilities of emerging technologies. If you have read books such as The Book of Why, The Science of Interstellar, or The Future of Humanity, you will love Films from the Future.

'If you think you understand AI and all of the related issues, you don't. By the time you finish this exceptionally lucid and riveting book you will breathe more easily and wisely' - Michael Gazzaniga A leading computer scientist brings human sense to the AI bubble No recent scientific enterprise has been so alluring, terrifying and filled with extravagant promise and frustrating setbacks as artificial intelligence. Writing with clarity and passion, leading AI researcher Melanie Mitchell offers a captivating account of modern-day artificial intelligence. Flavoured with personal stories and a twist of humour, Artificial Intelligence illuminates the workings of machines that mimic human learning, perception, language, creativity and common sense. Weaving together advances in AI with cognitive science and philosophy, Mitchell probes the extent to which today's 'smart' machines can actually think or understand, and whether AI even requires such elusive human qualities at all. Artificial Intelligence: A Guide for Thinking Humans provides readers with an accessible and clear-eyed view of the AI landscape, what the field has actually accomplished, how much further it has to go and what it means for all of

our futures.

To what extent should we use technological advances to try to make better human beings? Leading philosophers debate the possibility of enhancing human cognition, mood, personality, and physical performance, and controlling aging. Would this take us beyond the bounds of human nature? These are questions that need to be answered now.