AI-Driven Science: Speed versus Safety

AI Lab for Book-Lovers

xynapse traces is an imprint of Nimble Books LLC.

Ann Arbor, Michigan, USA
http://NimbleBooks.com
Inquiries: xynapse@nimblebooks.com

ISBN (tk): TK-ISBN-978-X-XXXXXX-XX-X

version: 0.1.0-202506162022

Contents

AI-Driven Science: Speed versus Safety

Publisher's Note

At **xynapse traces**, an imprint of Nimble Books LLX, we champion 'transcriptive transformation.' Through the contemplative practice of the Korean practice of _pilsa_–hand-copying texts–readers can forge a deeper connection with ideas crucial to humanity's understanding, capabilities, and contributions to the universe.

Our Artificial Intelligence stream explores a domain rapidly reshaping our reality. This volume, *AI-Driven Science: Speed versus Safety*, tackles a pivotal tension at the heart of progress. AI promises to turbocharge scientific discovery: imagine designing life-saving drugs in days, deciphering cosmic mysteries with unprecedented clarity, or crafting sustainable solutions for our planet at lightning speed. The speed of advance is exhilarating, a testament to human ingenuity amplified.

However, velocity without guidance is dangerous. AI safety is not about slowing innovation, but about avoiding disastrous collisions. When, if ever, does the race for breakthroughs eclipse the need for ethical deliberation or foresight into unintended consequences? This question is no longer academic; it' s at our doorstep.

Engaging with these complex issues through the Korean practice of *pilsa* allows for a unique form of internalization. Transcribing the arguments herein encourages a nuanced grappling with the stakes involved—moving beyond headlines to the core of the debate. This book doesn't offer easy answers. Instead, it seeks to illuminate the intricate balance required to harness AI's immense power for science. Understanding this inherent tension is essential for understanding the advent of AI.

Executive Summary: Quantified Benefits of Pilsa

Pilsa (필사), the Korean practice of mindful transcription, offers multiple documented benefits:

- **Boosts brain activity and fine motor skills**—fMRI scans show up to **3× more activation** in key language areas during handwriting vs. typing, with EEG studies confirming **widespread brain connectivity** that typing cannot achieve.

Transcription Note

As you engage with the quotes in this book, remember that the act of transcription is both a practice in mindfulness and a way to deeply connect with the text. Take your time with each passage, feeling the rhythm of the words as you write them. There's no right or wrong way to transcribe—what matters is your personal connection to the text and the moments of reflection that arise during the process.

Quotations for Transcription

AI can dramatically accelerate scientific discovery, for example by helping us cure diseases, create new materials, explore the universe and understand ourselves better.

If we put the wrong objective into a superintelligent machine, it will achieve that objective, and we will get exactly what we asked for, not what we wanted.

Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (2019)

Before the prospect of an intelligence explosion, we humans are like small children playing with a bomb. Such is the mismatch between the power of our plaything and the immaturity of our conduct.

Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (2014)

AI could greatly accelerate progress in biotechnology, for example by designing novel proteins or entire synthetic organisms. This could be used for great good, but also to create more powerful bioweapons, or for other dangerous misuses.

Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (2020)

The power of AI necessitates a new era of reflection and foresight. We must consider not only what AI can do but also what it should do, especially when it reshapes the frontiers of science.

Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (2021)

The problem of value alignment…is the problem of how to ensure that AI systems pursue goals that are beneficial to humans, even as those systems become far more intelligent than their creators.

Brian Christian, *The Alignment Problem: Machine Learning and Human Values* (2020)

As artificial systems become more autonomous and more powerful, the potential for both benefit and harm increases. Ensuring they are beneficial requires us to address the challenge of machine morality.

Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (2009)

Future AIs, if they achieve superhuman capabilities, might make conceptual breakthroughs that are opaque to human intellects. The challenge then is not just control but comprehension of the risks involved.

Martin Rees, *On the Future: Prospects for Humanity* (2018)

Algorithms are making increasingly important decisions about us, but they can be inscrutable black boxes. We need to demand transparency, or at least serious attempts at interpretability.

David Spiegelhalter, *The Art of Statistics: How to Learn from Data* (2019)

Development of safe AI is a prerequisite to development of any other type of advanced AI. It is not an optional feature or an add-on, but a fundamental requirement for survival.

Roman Yampolskiy, *Artificial Superintelligence: A Futuristic Approach* (2015)

The speed and scale of AI-driven discovery will challenge existing regulatory frameworks and ethical norms. We must proactively consider how to adapt these systems to ensure safety and accountability.

Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (2021)

AI's power to accelerate discovery is undeniable. But this acceleration must be coupled with a commitment to responsible innovation, ensuring that we are not just moving fast but moving in the right direction.

Fei-Fei Li, *The Worlds I See: Curiosity, Exploration, and Discovery at the Dawn of AI* (2023)

If we are to use AI in high-stakes domains like science and medicine, we need systems that are not just powerful pattern recognizers but also robust, reliable, and capable of genuine reasoning. Speed without reliability is a recipe for disaster.

Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019)

Responsible development and use of AI in science requires a holistic approach, integrating ethical principles and societal values throughout the lifecycle, not just as a final check before deployment. Speed must not compromise this integration.

Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way* (2019)

While AI can accelerate discovery by identifying patterns humans miss, the opacity of some models can be a barrier. If we don't understand *why* an AI makes a prediction, trusting it for critical scientific applications becomes a safety concern.

Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence (Revised and Updated Edition)* (2022)

AI systems used in scientific research can inherit biases from their training data or design, leading to skewed results. This isn't just an issue of accuracy; it's an issue of safety and justice in how scientific knowledge is produced and applied.

Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021)

The global nature of AI development and its impact on science means that safety and ethical standards cannot be effectively addressed by any single nation alone. International collaboration is essential to navigate these complex challenges.

Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (2021)

Given the transformative potential of AI in science, a robust application of the precautionary principle is warranted. This means prioritizing safety and thorough risk assessment even if it moderately slows the pace of discovery.

John Zerilli, John Danaher, James Maclaurin, Colin Gavaghan, Alistair Knott, Joy Liddicoat, and Merel Noorman, *A Citizen's Guide to Artificial Intelligence* (2021)

The problem is that we don’ t know how to specify what we want with the level of exactitude that computers require. This is the alignment problem, and it becomes more urgent as AI’s scientific capabilities grow.

Brian Christian, *The Alignment Problem: Machine Learning and Human Values* (2020)

As AI systems take on more significant roles in scientific discovery, the need for 'moral' or ethical control systems becomes acute. Unfettered speed in discovery without such controls poses unacceptable risks.

Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (2009)

The pace of technological advancement is exponential, not linear. This means that the 21st century will see 20,000 years of progress at today's rate, a scale of change that demands profound foresight.

Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (1999)

The Master Algorithm by itself is not dangerous, but it can be combined with data and goals to do harm. The more powerful our algorithms, the more care we need to take in how we use them.

Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015)

When an algorithm makes a discovery that no human can understand, it challenges the very nature of scientific endeavor. Is it still science if we can't explain the 'why'?

Hannah Fry, *Hello World: Being Human in the Age of Algorithms* (2018)

AI-driven scientific breakthroughs, particularly in fields like synthetic biology or materials science, could be weaponized faster than we can develop defenses or ethical guidelines to control them.

Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (2018)

The drive for rapid discovery with AI must be tempered by ethical reflection, ensuring that the direction of scientific progress aligns with human values and societal well-being, not just computational efficiency.

Mark Coeckelbergh, *AI Ethics* (2020)

The allure of AI-driven speed in science is undeniable, but we must be vigilant. An AI confidently spewing nonsense, or subtly biased results, could derail research on an unprecedented scale if not carefully validated.

Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019)

The challenge with AI in science is that its progress is so rapid. Regulatory and ethical frameworks struggle to keep pace, creating a dangerous lag where risks can emerge unchecked before society can react.

Martin Ford, *Rule of the Robots: How Artificial Intelligence Will Transform Everything* (2021)

AI reduces the cost of prediction, accelerating scientific discovery. However, this speed necessitates a renewed focus on distinguishing correlation from causation, a task where human scientific judgment remains crucial.

Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (2018)

The coming wave of technologies, including AI in science, will be characterized by its speed and scale. Containing the risks requires us to move just as fast in developing safeguards and global cooperation.

Mustafa Suleyman with Michael Bhaskar, *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma* (2023)

The new AIs, if they are to help us, will need to be partners in the quest for scientific understanding, not just incredibly fast tools. Their development must prioritize our long-term survival and well-being.

James Lovelock, *Novacene: The Coming Age of Hyperintelligence* (2019)

When AI systems assist in scientific discovery, their inscrutability can be a barrier. For science to maintain its integrity and public trust, we need methods for understanding and auditing these powerful, fast tools.

Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (2019)

My greatest fear is that the sheer speed and cutthroat nature of this AI race will lead companies and countries to deploy AI applications before they are truly safe and fair.

Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (2018)

The development of powerful new technologies, such as advanced AI and synthetic biology, has given humanity unprecedented power to shape the world. But this power comes with unprecedented risks.

Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (2020)

The development of AI for scientific research also raises complex ethical questions. How do we ensure that AI-driven discoveries are used responsibly? How do we guard against algorithmic bias in scientific inquiry?

Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (2021)

As AI becomes a core tool for scientific discovery, we will need new processes for validation and a heightened awareness of potential pitfalls like spurious correlations or biases embedded in the data.

Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (2018)

The solution is to design AI systems whose goals are, from the outset, uncertain—they know that they don’ t know what humans want. This uncertainty is crucial for safety.

Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (2019)

> *Algorithms can also go wrong. They can produce unfair or discriminatory outcomes, often because they were trained on biased data or because their creators didn’ t fully understand their implications.*

Erik Brynjolfsson and Andrew McAfee, *Machine, Platform, Crowd: Harnessing Our Digital Future* (2017)

The design of our information environment, including the algorithms that shape it, is an ethical task. We must ensure it serves human flourishing, not just efficiency or profit.

Luciano Floridi, *The Ethics of Information* (2013)

A commitment to intelligibility—to understanding how important decisions are made—should be a bedrock principle of a fair society. This applies with special force to automated systems that are increasingly shaping our lives.

Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2015)

Who knows? Who decides? Who decides who decides? These are the bedrock questions of the twenty-first-century polis, just as they were for Aristotle.

Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (2019)

The problem is not that high-tech tools are inherently biased, but that they are implemented in ways that reproduce and amplify existing inequalities.

Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (2018)

To understand the real costs of AI, we need to look beyond the algorithmic abstractions to the material realities of extraction, labor, and energy that underpin these systems.

Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021)

Precautionary principles suggest that the burden of proof should be on developers to show that a technology is safe, or that its potential benefits outweigh its potential harms, before it is widely deployed.

Wendell Wallach, *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control* (2015)

The project of alignment, then, is not merely to build machines that do what we say, but to build machines that help us clarify what it is we ought to say.

Brian Christian, *The Alignment Problem: Machine Learning and Human Values* (2020)

Our own values and desires influence our choices, from the data we choose to collect to the questions we ask. Models are opinions embedded in mathematics.

Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (2016)

Wisdom is needed more than ever when science offers us such rapidly expanding powers. But wisdom isn’ t advancing as fast as science and technology.

Martin Rees, *On the Future: Prospects for Humanity* (2018)

What we need most of all is AI that we can trust. AI that is safe, reliable, and fair. AI that has common sense, and that can explain its choices.

Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019)

The same AI breakthroughs that can be used to cure diseases can also be used to create bioweapons of terrifying power.

Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (2017)

A world of inscrutable technologies, then, is not only a world of diminished accountability but also one of diminished understanding, where we cannot learn from our machines or correct their mistakes.

Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2015)

The allure of automation is that it will grant us more control over our world. But the more we automate, the more we lose control.

Nicholas Carr, *The Glass Cage: Automation and Us* (2014)

Technologies are not merely tools; they are forms of life, and their invention is necessarily an ethical and political act.

Sheila Jasanoff, *The Ethics of Invention: Technology and the Human Future* (2016)

If a research program is of critical importance for humanity’ s future, then its governance should reflect that fact. This may call for broad international collaboration and oversight.

Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (2014)

When reward is tied to measured performance, the measures themselves are apt to be manipulated or distorted, especially when the stakes are high.

Jerry Z. Muller, *The Tyranny of Metrics* (2018)

The off-switch problem is a good illustration of why we need a new model for AI, one based on uncertainty about objectives and on learning human preferences.

Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (2019)

Our Earth is forty-five million centuries old. But this century is special. It is the first when one species—ours—can determine the biosphere's future.

Martin Rees, *On the Future: Prospects for Humanity* (2018)

The challenge is to infuse machines with the values and ethical reasoning capabilities that will enable them to make appropriate choices when confronted with novel situations, rather than simply programming them for every eventuality.

Wendell Wallach, *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control* (2015)

The problem is that we don’t know how to specify what we want with sufficient precision. If we get it wrong, the results can be catastrophic.

Brian Christian, *The Alignment Problem: Machine Learning and Human Values* (2020)

Artificial intelligence is neither artificial nor intelligent. Rather, artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications.

Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021)

If we are to retain control over entities far more intelligent than ourselves, we will need to learn how to build them so that their goals, if achieved, are beneficial to us.

Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (2019)

The main thing that has changed is that we can now run the same logical experiments on a million-dollar machine in a few minutes that would have taken a lifetime to run by hand.

George Dyson, *Turing's Cathedral: The Origins of the Digital Universe* (2012)

The allure of automated judgment is its promise of greater efficiency and insight. Yet, when the inscrutable dictates of black boxes govern important decisions, we risk sacrificing fairness, transparency, and accountability.

Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2015)

The real challenge posed by AI is not a matter of quantity of intelligence, but of quality of autonomy and adaptability, and this is what needs to be regulated, not feared.

Luciano Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (2014)

Rather than being a shortcut to innovation, high-tech tools too often become instruments of containment, deepening social and economic divides. They promise neutrality but often reproduce and even amplify existing biases.

Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (2018)

The genie is out of the bottle. We now have to work out how to live with these powerful technologies, to ensure they improve our lives and don't end them.

Toby Walsh, *2062: The World that AI Made* (2018)

We need to invest a lot more in the safety side, in the ethics side, in the societal impact side. The purely technical race for more powerful AI is not enough and could be dangerous.

Yoshua Bengio, *Architects of Intelligence: The truth about AI from the people building it* (2018)

The risk is that we may become unable to validate, verify, or even falsify the new theories and hypotheses generated by ever-more complex AI systems, leading to a crisis of intelligibility in science.

Luciano Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (2014)

AI systems are not objective or neutral; they are artifacts shaped by specific interests and values. In science, this means AI can amplify existing biases, skewing research outcomes and potentially harming vulnerable groups.

Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021)

The potential for AI to accelerate scientific discovery is immense, but this acceleration also amplifies the need for rigorous safety protocols and a deep understanding of the systems we are building to prevent unintended consequences.

Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (2019)

As AI systems become more capable of autonomous scientific discovery, we face a profound challenge: ensuring that this power is used wisely and ethically, not to create new harms or exacerbate existing inequalities.

Toby Walsh, *Machines Behaving Badly: The Morality of AI* (2022)

The race to develop powerful AI, including for scientific breakthroughs, must be paralleled by an equally determined race to develop robust safety measures and ethical guidelines. Progress without precaution is a dangerous gamble.

Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (2017)

We must resist the allure of technological solutionism, asking critical questions about power and equity before implementing automated systems in any domain, including scientific research where impacts can be profound and far-reaching.

Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (2018)

For AI to genuinely advance science for public good, its algorithmic processes must be transparent and contestable, not 'black boxes' that obscure biases or errors, thereby undermining trust and research integrity.

Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2015)

The acceleration of science via AI must be scrutinized for how it might entrench new forms of power and surveillance, potentially leading to a future where knowledge discovery is controlled by unaccountable forces.

Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (2019)

AI offers astounding possibilities for scientific advancement, but we must proactively shape its development and integration, ensuring that rapid discovery doesn't outpace our ethical frameworks or societal capacity to adapt.

Erik Brynjolfsson and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (2014)

'Models are opinions embedded in mathematics.' In science, this means AI-driven models, if not scrutinized for bias, can perpetuate systemic unfairness or lead to flawed conclusions despite a veneer of objectivity.

Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (2016)

Current AI often lacks true understanding. Relying on it for high-stakes scientific discovery without extreme caution and robust verification could lead to significant errors and misdirected research efforts, hindering genuine progress.

Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019)

When AI is used in scientific research, the stakes are incredibly high. A flawed algorithm or biased dataset can lead to incorrect discoveries, misallocated resources, and even direct harm if not developed and deployed responsibly.

Reid Blackman, *Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI* (2022)

The goal is to automate the scientific method. But this doesn’t mean scientists will be out of a job. On the contrary, they’ll be able to ask much bigger questions, and the Master Algorithm will be their indispensable assistant.

Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015)

AI will accelerate the pace of discovery, potentially solving some of humanity's greatest challenges. But this acceleration also compresses the time available to consider consequences, demanding new frameworks for responsible innovation.

Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (2021)

The code might be able to identify interesting new patterns, new hypotheses to test. But it is still the human who is absolutely essential in designing the experiment to test the hypothesis and to understand the results.

Marcus du Sautoy, *The Creativity Code: Art and Innovation in the Age of AI* (2019)

Without a capacity for genuine understanding, AI systems, no matter how statistically impressive, will remain prone to egregious errors. In science, such errors could derail progress or, worse, lead to harmful conclusions.

Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019)

The history of technology races is often a history of corners cut. In the context of AGI, the stakes are too high for that. We need to foster a culture of safety and cooperation, not reckless competition.

Brian Christian, *The Alignment Problem: Machine Learning and Human Values* (2020)

Human doctors and researchers would transition into roles as 'AI trainers' or 'AI explainers,' or focus on the 'last mile' of patient interaction and complex ethical choices.

Kai-Fu Lee and Chen Qiufan, *AI 2041: Ten Visions for Our Future* (2021)

Intelligence is a process, not a product. If we build machines that are more intelligent than we are, they will explore avenues of understanding that are closed to us. This is the price of admission to a larger world.

George Dyson (essay 'The Third Law'), *Possible Minds: Twenty-Five Ways of Looking at AI* (2019)

Our ability to create intelligent machines will soon outstrip our ability to control them. This is not a problem that can be deferred; it is one that we must address now, with urgency and foresight.

Amir Husain, *The Sentient Machine: The Coming Age of Artificial Intelligence* (2017)

AI will undoubtedly accelerate scientific discovery. However, we must be careful. If an AI proposes a new theory or discovers a new drug, we will need to understand how it reached its conclusions. We cannot blindly trust intelligent machines.

Jeff Hawkins, *A Thousand Brains: A New Theory of Intelligence* (2021)

Scientists and technologists must now also become ethicists and philosophers, or at least engage deeply with them. The questions raised by AI are too profound to be left to specialists in any single domain.

Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (2021)

Technologies are not merely aids to human activity, but also powerful forces acting to reshape that activity and its meaning.

Langdon Winner, *The Whale and the Reactor: A Search for Limits in an Age of High Technology* (1986)

Current AI is narrow; it works for specific tasks, but it breaks, often catastrophically, when it encounters situations that depart even slightly from its training data. This is not the path to trustworthy AI, in science or anywhere else.

Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (2019)

With powerful new technologies, the race to be first can overshadow the need to be careful. Ensuring that safety and ethics keep pace with innovation is one of the paramount challenges of our time.

Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (2020)