# ARTIFICIAL INTEGRITY

## THE PATHS TO LEADING AI TOWARD A HUMAN-CENTERED FUTURE

HAMILTON MANN

WILEY

## Praise for *Artificial Integrity*

"In *Artificial Integrity,* Hamilton Mann extends his groundbreaking ideas around "Digital for Good" to Artificial Intelligence. Clearly, we are at interesting crossroads as a society where the extensive opportunities from AI need to be weighed against the disruptions it can potentially create. Mann masterfully lays out how adapting AI to diverse cultural contexts, value models, and situational nuances can counter biases and lead us toward a more human-centered AI. Mann continues to make valuable contributions on how technology and society work together to create better outcomes for all."

**—Nilanjan Adhya**
*Chief Digital Officer, BlackRock*

"In *Artificial Integrity,* Hamilton Mann presents a unique perspective on the future of AI. His vision transcends mere technological progress, recognizing that the true challenge of AI lies in fostering systems that operate with unwavering integrity and align with human values. Drawing on a rich understanding of AI's origins and current trajectory, Mann introduces the transformative concept of '*artificial integrity*.' This paradigm shift provides a much-needed framework for developing AI systems that comply with ethical standards and actively reinforce and uphold human-centered beneficial societal norms. Mann's work is particularly timely, addressing the complex interplay between AI and society, from its impact on the job market to its potential to exacerbate or mitigate societal inequalities. He offers practical insights and strategies for business leaders, policymakers, and technologists to navigate the transitions ahead, ensuring that AI augments rather than replaces human capabilities. '*Artificial Integrity*' is a must-read for anyone involved in or affected by AI development. In today's world, where AI's influence is pervasive, Mann's vision of a future where human insight and artificial intelligence synergize to multiply value while adhering to core human values is both inspiring and crucial. This book sets a new standard for how we should approach the development and integration of AI in our society, making it an indispensable guide for shaping a human-centered AI-driven future."

**—Seth Dobrin**
*Former Global Chief AI Officer, IBM*

"Finally, we have a thoughtful, practical book on ensuring AI is used in a responsible "high integrity" way. Hamilton Mann, who is always so great about keeping our eye on the bigger picture, not only lays out the challenges ahead (including many that haven't gotten adequate coverage), but he also lays out a realistic path toward shaping the AI-future that we would want to have. A great contribution!"

**—David C. Edelman**
*Senior Lecturer of Business Administration, Harvard Business School and co-author of Personalized: Customer Strategy in the Age of AI, Harvard Business Review Press*

"Some people are nice, smart, and keep surprising you in a good way. Hamilton Mann is one of them. When it comes to artificial intelligence and integrity, he is at the right place to be able to talk and write about it. So I'm excited about Hamilton Mann's new book about artificial intelligence and integrity, and I strongly recommend it."

**—Jean Ferré**
*CEO Sinequa*

"In *Artificial Integrity,* Hamilton Mann has provided the seminal revelation that user trustworthiness of any Artificial Intelligence system does not equate to the accountable integrity of its designers. Lack of genuine consent for user data, anthropomorphism by design, and the harmful extraction of water and energy for data centers powering GenAI demand the integrity by design Mann elucidates that is anything but artificial. A really compelling and unique work in the space."

**—John C. Havens**
*Author of Heartificial Intelligence: Embracing our Humanity to Maximize Machines and Founding Executive Director of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.*

"Yes, with a cool technology like AI, the temptation to focus entirely on the tech is strong. But technology itself is agnostic; humans have considerable choice. As AI technologies become pervasive, what better time than now for Hamilton Mann to focus us on the paramount role of human agency? Smarter machines won't be enough—humans need to guide savvier machines from myopic optimization of narrow objectives to the complex, multi-criteria domain of complex decision-making."

**—Sam Ransbotham, PhD.**

*Professor of Business Analytics. Machine Learning and Artificial Intelligence at Boston College, Senior editor at Information Systems Research, Associate Editor at Management Science, Academic contributing Editor on AI at MIT Sloan Management Review and Host of Me, Myself and AI podcast*

"*Artificial Integrity* by Hamilton Mann is a groundbreaking work that redefines how we should approach AI development. Mann provides a compelling and essential read on the importance of prioritizing integrity over mere intelligence in AI systems. He challenges traditional ethical frameworks offering actionable advice to ensure AI aligns with human values and societal norms. This book helps to crystalize what many of us have been thinking and working on for quite some time.

Mann's insights are both incisive and practical, guiding us through the intricacies of creating AI that enhances rather than replaces human capabilities, a philosophy I am very passionate about. This book is a must-read!"

**—Noelle Russell**

*Chief AI Officer, AI Leadership Institute, Microsoft MVP in AI*

"This unique and impactful book is a call for humans to take up their responsibility to ensure that together with AI technology they work to enhance human wellbeing through creative, inclusive, and innovative interaction. Hamilton Mann specifically redefines humancentric AI in this engaging book to have a focus on the integrity of AI systems in

decision-making. The significance of the book lies in its not only demonstrating that technology is not neutral, but also unpacking how humans can and should interact with AI technology to build a shared value system that will benefit all of humanity."

**—Emma Ruttkamp-Bloem**

*Professor and head of the Department of Philosophy, Faculty of Humanities at the University of Pretoria, Member of the UN SG AI Advisory Body, Chair of UNESCO World Commission on the Ethics of Scientific Knowledge and Technology, AI Ethics Lead of the Centre for AI Research, Expert of the Global Commission on Responsible AI in the Military Domain, Associate Editor of the Science and Engineering Ethics, Member of the Women4EthicalAI (UNESCO), Member of the AI Ethics Experts without Borders (UNESCO), Member of the Editorial Board of AI and Law*

"Hamilton Mann provides a new framework, *Artificial Integrity*, for leaders and organizations to guide responsible AI development, balancing benefits and risks for customers, employees, and society more broadly as we navigate the uncertain path toward AI super-intelligence. This is a must read for anyone developing new AI solutions, shaping corporate strategy, or creating public policy for the new AI era."

**—Scott A. Snyder, PhD.**

*Senior Fellow, Management Department, The Wharton School, Adjunct Faculty, The Moore School of Engineering, University of Pennsylvania, Chief Digital Officer, Eversana, Co-author Goliath's Revenge*

"The thing about integrity is that it's innately human. And as John Wooden famously said, 'your character is what you really are.' In an era of generative, predictive, and general AI, humanity is by default automated. But with integrity, AI becomes more human and ultimately, augmented. Al integrity becomes a superpower. That's what Hamilton Mann is here to show us."

**—Brian Solis**

*Futurist, best-selling author of Mindshift, Transform Leadership, Drive Innovation, and Shape the Future*

"AI has immense power to do good, to contribute positively to individuals, organizations, and society. It also has a dark side. With *Artificial Integrity*, Hamilton Mann provides a roadmap to create and deliver AI that is responsible, ethical, and sustainable."

**—Michael Wade**

*TONOMUS Professor of Strategy, Digital, and Director of the TONOMUS Global Center for Digital and AI Transformation at IMD Business School and co-author of Hacking DigitaI, McGraw-Hill Education*

"*Artificial integrity* is a reminder that we must aspire to the highest standards in life and living, both for ourselves and the technology that we permit as part of it. We cannot expect the technology of tomorrow to reflect values that the humans of today do not manifest in practice. Values in, values out—or Garbage in, garbage out—the choice is ours. Hamilton Mann masterfully highlights the areas of concern as we are navigating this new era of all pervasive artificial intelligence. *Artificial integrity* starts with human integrity, and it is our responsibility, individually and as a species to strive for it."

**—Cornelia C. Walther, PhD.**

*Senior Visiting Fellow at the Wharton initiative for Neuroscience (WiN) / Wharton AI and Analytics; and the Center for Integrated Oral Health (CIGOH). Director POZE@ezop and a humanitarian practitioner for two decades with the United Nations in emergencies in West Africa, Asia, and Latin America.*

"*Artificial Integrity* is more than a book; it's the dawn of a new interdisciplinary field of science for the future. Bridging computational science, sociology, economics, and sustainability, it urges leaders, technologists, academics, policymakers and innovators to envision a world where AI elevates humanity. Hamilton Mann's concept of *Artificial Integrity* is a call to action with sharp guidelines and well-articulated framework for executing what AI can and should be."

**—Soumitra Dutta, PhD.**

*Peter Moores Dean and Professor of Management, Saïd Business School, University of Oxford, and a Fellow of Balliol College, Oxford.*

"*Artificial Integrity* offers both a thought-provoking and guiding approach to the complex relationship between artificial intelligence and our human value models in a society where advanced co-intelligence between humans and machines will increasingly coexist. This insightful concept and framework of '*artificial integrity*' coined by Hamilton Mann represents a paradigm shift and an essential playbook for leaders wishing to steer AI development towards a human-centered future. Read this book to understand, but more importantly, to act for the development of AI designed not simply to produce more artificial intelligence, but to uphold more integrity."

**—Luc Julia, PhD.**
*Chief Scientific Officer, Renault Group, co-creator of Siri, member of the French Academy of Technologies.*

"The world is racing headlong into a very near future radically transformed by Artificial Intelligence. This technology offers enormous opportunities for advances in medicine, science, innovation, and efficiency, but potentially threatens jobs, personal autonomy, and national security. In this important and timely book, Hamilton Mann proposes the radical new concept of *Artificial Integrity*—the purposeful alignment of Artificial Intelligence with human moral codes and social norms—and argues it should be baked into the code at the very heart of this new technology. Mann provides practical strategies for business leaders, managers, and policymakers to lead through this technological revolution. I urge everyone who cares about our collective future as humans to read this book."

**—Michael Platt, PhD.**
*Director of the Wharton Neuroscience Initiative and Professor of Marketing, Neuroscience, and Psychology at the University of Pennsylvania*

"As someone wary of AI hype, I found Hamilton Mann's "*Artificial Integrity*" refreshingly practical. Unlike works focused on AGI or broad ethical principles, Mann provides concrete strategies for embedding integrity into current AI systems. Drawing on global perspectives and bridging academia and business, he offers invaluable insights on '*artificial integrity*,' emphasizing

context-specific application and alignment with local norms. Mann's focus on augmenting human capabilities rather than replacing them is genuinely refreshing, showing how AI can enhance our lives by working alongside us. This essential read for technologists, business leaders, and policymakers offers a forward-thinking perspective on ethical AI implementation, will significantly influence responsible AI development and deployment."

**—Anand S. Rao, PhD.**
*Distinguished Service Professor of*
*Applied Data Science and AI,*
*Carnegie Mellon University*

# Artificial Integrity

## The Paths to Leading AI Toward a Human-Centered Future

Hamilton Mann

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada and the United Kingdom.

ISBNs: 9781394297986 (Paperback), 9781394298006 (ePDF), 9781394297993 (ePub)



Library of Congress Cataloging in Publication data available on request.

Cover image: © TAW4Getty Images

Author Photo: D-light Production

Cover design: Hamilton Mann

*For my R. and my A.—for always being there.*

# Contents

# Preface

Warren Buffett once said, 'In looking for people to hire, look for three qualities: integrity, intelligence, and energy. And if they don't have the first, the other two will kill you.' This wisdom begs the question: as we begin to 'hire' powerful intelligent machines to perform tasks traditionally done by humans?

*Artificial Integrity* aims to be the first comprehensive and authoritative guide to one of the most important socio-economico-technological developments of our time—the rise of artificial intelligence as a society model. Yet, artificial intelligence is nothing new.

The concept of creating machines that can emulate human intelligence has been a part of human imagination for centuries. From the mythological constructs to the mechanical automata of the Renaissance, humanity has long been fascinated by the idea of artificial beings capable of thought and action.

The term *artificial intelligence* itself was coined in the mid-20th century, but the foundational ideas can be traced back much further. Throughout the 20th century, significant milestones in artificial intelligence (AI) development continued to capture the public's imagination. Alan Turing's seminal 1950 paper, "Computing Machinery and Intelligence," introduced the concept of the Turing test to evaluate a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. This period also saw the creation of the first AI programs, such as the Logic Theorist and the General Problem Solver, which attempted to mimic human problem-solving processes.

From a nonorganic point of view, AI is fundamentally artificial by nature because it is constructed from nonliving materials and operates through programmed algorithms. However, from a societal point of view, AI is anything but artificial and yet, inextricably human by design as it deeply integrates into human social structures, influencing and being influenced by societal norms, behaviors, and expectations. In fact, the term *artificial* masks the dual nature of AI—both in its artificial origins and its profound societal integration inherited from vast datasets on which AI algorithms are trained that include human behaviors, language, and cultural artifacts, encompassing social media interactions, online content, and real-world activities, embedding cultural and societal values into the AI's functionalities and reflecting societal priorities, biases, and norms.

When considering the intelligence often assigned to AI, a different duality comes to light. At their core, AI systems are designed to mimic specific aspects of human intelligence through advanced computational capabilities. They excel and already surpass human ability at tasks involving data analysis, pattern recognition, and logical reasoning, showcasing a form of cognitive intelligence. While some are striving to develop artificial general intelligence (AGI)—or any form of superintelligence that could mimic and surpass various aspects of human intelligence—it is essential to remember that human intelligence encompasses dimensions beyond rationality.

Among other various traits, human intelligence, by definition, includes humans at its core. This involves embracing irrationality through our ability to recognize, feel, understand, intuit, manage, and live one's own emotions and those of others. It also involves the very social nature of who we are, through our capacity to navigate complex social environments and build meaningful relationships. Moreover, it implies a critical dimension—that of evaluating the consequences of actions, understanding societal norms, and making choices that align with individual and shared values.

These traits, more subjective than objective, more irrational than rational, constitute conditions that lead to what one may perceive as intelligent. In fact, the term *intelligent* masks the dual nature of intelligence—both in its input characteristics referring to capabilities as a means and its outcome perspectives, which ultimately determine what is acknowledged as intelligent or not, as an end.

The polysemic nature of the word *intelligence* creates its paradox. Intelligent inputs do not prevent unintelligent outcomes, which typically include outcomes that:

- Fail to consider the broader context or nuances of a situation such as a language translation AI system producing grammatically correct but contextually inappropriate translations.
- Reflect and perpetuate existing biases present in the training data, such as facial recognition systems that have higher error rates for certain demographic groups.
- Do not align with ethical standards or societal values such as an AI algorithm prioritizing profit over patient care in healthcare settings.
- Provide immediate benefits but cause long-term negative consequences such as an AI in finance optimizing for short-term profits at the expense of financial stability.
- Are opaque and cannot be easily understood or explained such as black-box algorithms where the decision-making process is not clear, leading to mistrust and accountability issues.
- Pose risks to safety and security due to vulnerabilities or misuse such as autonomous vehicles making critical errors that lead to accidents or that cannot cope with failure that may lead to harmful consequences for humans.
- And so on and so forth.

What creates the intelligence of a result lies in a particular characteristic: its integrity, in the sense of its coherence and its adherence to the values of the human society in which it occurs. Integrity, more than intelligence, is what precedes truly intelligent results.

Intelligence is like the horsepower of a car—it tells you how powerful the engine is and how fast it can potentially go. Integrity, on the other hand, is like the steering system—it allows one to determine the direction the car goes and whether it stays on the road, adhering to a path that avoids harm to others and obeys the rules of the road.

Today, AI is embedded in countless applications, from voice assistants and recommendation systems to autonomous vehicles and predictive analytics. These systems have demonstrated remarkable capabilities, yet they also highlight the need for a framework that ensures their development

aligns with human values principles and societal goals. This is where the concept of artificial integrity becomes paramount.

Despite the advancements in AI technology over the decades, the fundamental challenge has remained the same: how to create machines that not only perform tasks but do so in a manner consistent with human values and societal norms. This challenge has become even more critical as AI systems have grown more powerful and pervasive. From the potential for job displacement to issues of privacy and surveillance, the integration of AI into our daily lives raises critical questions about the kind of future we want to build. Will AI systems be designed to enhance human capabilities and promote fairness, or will they exacerbate existing inequalities and undermine our fundamental rights? The answers to these questions will shape the trajectory of AI development and its role in society.

As AI continues to permeate various facets of our lives, from healthcare and finance to education and entertainment, the question of how to ensure that these systems operate, not just with mere intelligence but with integrity, becomes ever more pressing. Therefore, the development of AI is not just a technical endeavor but a profound societal transformation that requires careful consideration of its long-term impacts on human values, rights, and well-being.

The concept of artificial integrity goes beyond traditional AI development, which often prioritizes technical prowess and functional performance of the system without the systemic interplay of the ecosystem where it takes place. Instead, it emphasizes the integration of AI models with our human-values models and our co-intelligence human–AI models into the very fabric of AI systems. This approach seeks to ensure that AI not only performs tasks efficiently but also aligns with the broader goals of human society, fostering trust and ensuring that technological progress contributes to the common good.

In the chapters that follow, we will delve deeper into examining how to develop artificial integrity and shape this vision. We will explore the societal impacts of AI, providing a roadmap for developing technologies that uphold the principles of artificial integrity. Through this journey, the aim is to equip you with the insights, approaches, and framework necessary to navigate the complex interplay between AI and human society, ensuring a future where technology, AI in particular, serves the greater good.

Through this book, my intent is to address some of the trickiest questions related to advancing AI systems in delivering integrity-led outcomes. These questions include the following:

- What are the challenges in creating AI systems that are both intelligent and integrity aligned?
- How can AI systems be designed to anticipate and mitigate unintended consequences, ensuring they operate within integrity-led boundaries?
- How can we address biases in AI systems to ensure fairness and equity in their outcomes?
- How can AI's role in data-driven decision-making be designed with the need for human oversight and the preservation of individual autonomy?
- In what ways can AI be designed to enhance human decision-making rather than replace it?
- How can we assess the artificial integrity of AI systems in real-world applications?
- How can we create a fair and transparent data economy that respects individual rights and fosters innovation?
- How can businesses leverage AI to drive sustainable innovation while ensuring that their AI systems operate with integrity and respect for societal values?
- What are the long-term implications of AI on job markets and workforce dynamics, on economics, and more globally on society, and how can we prepare for them?

Artificial integrity is the next AI frontier. This concept serves as a deep-dive reference into the intricacies of AI's influence in society and its undeniable partnership with humanity with the purpose of establishing what it takes to make artificial intelligence reach a form of integrity.

The inspiration for this book came from my firm belief that technology, including AI, can be a force for good. The concept of artificial integrity has been occupying my reflections for a while, as I navigate the way forward in orchestrating the implementation of AI in a leading aerospace and defense international company, in my role as a corporate executive, in charge of global digital transformation initiatives.

A pivotal moment that led me to write about this concept—putting together thoughts and frameworks based on experience and research—happened during an intervention by Yoshua Bengio, who introduced a panel discussion in which I was participating at the MIT Platform Summit in 2023, co-chaired by Marshall Van Alstyne, Geoffrey Parker, and Peter C. Evans. Yoshua was emphasizing that he was worried about the harmful consequences that AI may cause if put in the wrong hands and left insufficiently regulated. It resonated with my belief in making technology a force for good and in the power we have to make that vision a reality and shape the society we desire.

The debate on AI's role in society, its potential, and its pitfalls left me pondering not just the future of AI but the integrity with which we must approach its development. It convinced me that this is a field that needs to be a field in itself, uniting all the propositions from the collective intelligence we are capable of as humans, to make technology, and AI in particular, an enabler for sustainable progress.

Dive into a journey that promises a human-centered perspective on AI's transformative role in modern society and emerge with a renewed sense of purpose and an enriched understanding of what is intrinsically complementary to any form of intelligence: that of integrity that ensures coherence and harmony, allowing AI to function in symbiosis with the whole.

# Introduction

Some contemporary theories of intelligence have broadened the concept beyond language or any other cognitive abilities to include emotional and social aspects, such as in the theory of multiple intelligences by Howard Gardner or the concept of emotional intelligence by Daniel Goleman. Even within these frameworks, integrity isn't considered.

Just as is the case for humans, systems capable of exhibiting or imitating a form of intelligence are not a guarantee that their outcomes will be beneficial for society. Whatever it may be, any system capable of artificially mimicking a kind of intelligence should be able to do so while being bound to preserve the integrity of the ecosystem in which this said intelligence is a part. Systems equipped with artificial intelligence (AI) designed and deployed considering their ecosystem integrity are those that will be capable not just of artificial intelligence but also of artificial integrity.

Artificial integrity is the new "AI." This book, an eponym of this concept, serves as a deep-dive reference into the intricacies of AI's influence ethical conundrums and its undeniable partnership with humanity with the purpose of establishing what it takes to make artificial intelligence reach a form of integrity.

In *Artificial Integrity*, you'll find more than just insights. You'll find a vision—to shape a future where AI and humanity coexist, understand, and enhance one another. Meticulously curated not just to inform but to delve into deeper waters by examining AI from diverse global perspectives and advocating for an AI literacy that goes beyond code and algorithms,

*Artificial Integrity* covers a spectrum of concerns but, more importantly, presents solutions and fosters an inclusive dialogue around balanced AI (for artificial integrity).

One of the standout features of this book is its human life–centric approach. It uniquely presents the symbiotic relationship between AI and humanity, illustrating how they can harmoniously coexist and amplify each other, holistically.

In the following chapters, we will embark on a comprehensive journey to understand and implement artificial integrity within AI systems. Each chapter is designed to build upon the last, guiding you through the complexities, challenges, and opportunities that come with implementing artificial integrity into AI. Here is a breakdown of what you can expect to learn in each chapter:

**Chapter 1: The Stakes for Building Artificial Integrity** This chapter explores the critical importance of embedding artificial integrity into AI systems. It first delves into the concept of "Digital for Good," highlights the transformative potential of AI in societies, the exclusivity inherent in AI design, and the complex impacts of AI on the workforce. The origins of AI are also examined to understand its current trajectory and future implications.

**Chapter 2: Unpacking What Artificial Integrity Is** You will discover a comprehensive definition of artificial integrity, understanding it as a force sustaining societal values. The chapter discusses the dynamics of artificial integrity's metamodel and how it differs from AI ethics. It emphasizes the co-development implications of human and artificial intelligence.

**Chapter 3: What It Takes to Envision Human and Artificial Co-intelligence** This chapter outlines the modes of interaction between human and artificial intelligence. It presents the Marginal Mode for basic discernment, the Human-First Mode as a guide for consciousness, the AI-First Mode as a frontier for intelligent solution-driven actions, and the Fusion Mode as a novel paradigm for co-intelligence.

**Chapter 4: What Navigating Artificial Integrity Transitions Implies** Here, the focus is on the transitions necessary to achieve artificial integrity. It emphasizes the importance of mindset, data over procedures,

and ethical stands. The chapter also explores the transition in human capital values and introduces pioneering AI-driven operating models.

**Chapter 5: How to Thrive Through Navigating Algorithmic Boost** This chapter addresses how to leverage the algorithmic boost provided by AI. It guides you through navigating from Marginal mode to AI-First mode, navigating from Human-First mode to Fusion mode, and navigating along other transitional paths, explaining the benefits and challenges of each transition while articulating guiding posts to address such transitions.

**Chapter 6: How to Thrive Through Navigating Humanistic Reinforcement** Emphasizing human agency, this chapter discusses how to navigate transitions, such as from AI-First mode to Human-First mode and from Fusion mode to Human-First mode, as well as other key transitions. It highlights the importance of reinforcing humanistic values in the age of AI and provides practical guidelines for operationalizing these transitions effectively.

**Chapter 7: How to Thrive Through Navigating Algorithmic Recalibration** You will learn about the necessity of recalibrating algorithms to adapt to changing contexts. The chapter covers transitions such as Fusion mode to Marginal mode and Human-First mode to AI-First mode, providing strategies and concrete implementation approaches to ensure adaptive and responsible AI use, as context evolves.

**Chapter 8: What Change to Envision in Economic AIquity and Societal Values** The final chapter envisions a future where AI aligns with economic equity and societal values. It discusses data as a common currency, the paradoxes of impartial data, and the dual nature of biases. The chapter concludes by reflecting on the value placed on human life.

Designed for a broad audience—from executives, industry leaders, and entrepreneurs to policymakers, academics, tech enthusiasts, and curious general readers—*Artificial Integrity* stands apart with its multidimensional perspectives, its multidisciplinary approaches, its in-depth analysis, and its actionable insights. It bridges technical prowess with human values, societal structures, and cultural narratives to build artificial integrity that draws the path to leading AI with integrity toward a human-centered future.

# 1 The Stakes for Building Artificial Integrity

As we stand on the brink of a new era in artificial intelligence (AI), the discourse around its development has expanded beyond the technical and into the societal arena. Central to this is the concept of *artificial integrity*. This term might initially appear oxymoronic. And rightly so. After all, can integrity, a fundamentally human trait characterized by honesty and moral uprightness, be artificial?

As a matter of fact, the same goes for the term *artificial intelligence*, to which we are now accustomed. Can intelligence, another fundamentally human trait, be artificial? So-called "artificial" intelligence has often been and is still commonly compared to human intelligence. Yet, they are inherently different, as their origins differ, even though the same term is used to name both.

Indeed, AI, engineered from codes and algorithms, operates without consciousness or emotional depth, aspects typically intertwined with human intelligence. This distinction is crucial, as it shapes our expectations as AI systems become more integrated into societal functions. Similarly, so-called

"artificial" integrity is not the same as that which concerns human behavior or quality, even though here again we use the same word to name both. Artificial integrity reflects a deeper query into the essence of AI and the values frameworks we expect these systems to uphold.

## "Digital for Good" Matters

AI systems are not merely tools but active agents capable of decisions that affect lives and livelihoods. The development of AI systems that can perform tasks traditionally requiring human intelligence raises a significant question: how can these systems also embody the kind of decision-making integrity analogous to what we expect of humans in similar roles?

The world is increasingly driven by data and algorithms. As AI technologies progress and take on tasks that were once the sole preserve of humans, from driving cars to making medical diagnoses, the need for them to operate within defined value boundaries and demonstrate a form of integrity—also artificial and thus different from that of humans—becomes essential.

In other words, inventing a new form of intelligence implies the mirroring need for a new form of integrity that this new intelligence must also be capable of exhibiting. This form of integrity must be tailored to the unique characteristics of artificial intelligent systems, differing fundamentally from human integrity but paralleling its importance in guiding operations and decisions. This means programming AI not just with intelligence for task completion but also with frameworks and mechanisms that guide artificially, integrity-led interactions and decisions—a form of integrity that, while artificial, is indispensable for ensuring artificial intelligent systems contribute positively to society without causing harm.

This also means continually refining the frameworks and mechanisms that guide these systems toward integrity-led interactions and decisions as our societal norms and values understandings evolve. Consequently, creating AI systems that demonstrate integrity over intelligence involves more than just technical expertise; it requires a multidisciplinary approach incorporating value standards.

For instance, as an autonomous vehicle must decide in real time how to act in an emergency situation where human safety is at risk, the guidelines governing such decisions must be rooted in a framework that reflects

societal values on life and safety, be encoded into the vehicle's decision-making algorithms, and be auditable. As we give birth to a new form of intelligence through technology, the ongoing development of AI should be a commitment to integrity, ensuring that these systems do more than perform tasks efficiently—they must also act as integrity-led entities within society.

Just as we expect human intelligence to be guided by integrity, so too must artificial intelligence be steered by artificial integrity.

Technologies, particularly artificial intelligence, must be designed and deployed in a way that respects the rights and self-determination of individuals. This means avoiding systems that manipulate, deceive, or limit people's freedom of choice and decision-making abilities.

The collection, processing, and sharing of data must be carried out with the informed consent of individuals while ensuring the security and confidentiality of their personal information.

These technologies must be designed to promote equality and prevent discrimination. This includes fighting biases in AI algorithms, which can perpetuate or amplify social inequalities.

The principle of integrity includes the principle of responsibility. Developers and users of technology must be responsible for their creations and their impacts. This means that AI systems must be transparent, explainable, and subject to oversight and accountability mechanisms.

The challenge of safety must be proportional, if not superior, to the extraordinary capabilities that AI systems allow. They must be safe and secure, protecting users from harm or malicious use. This includes preventing physical and cyber risks associated with their use.

Furthermore, they must be in harmony with the preservation of the environment and the sustainability of resources. This means minimizing their ecological footprint.

Finally, they must serve human progress, contributing to the well-being, education, and development of individuals' capacities. This implies designing AI systems that enrich human life and promote personal and collective flourishing.

Overall, integrity in the coexistence of humankind and machine requires a holistic approach, centered on the respect for human dignity, the promotion of justice, and well-being, while ensuring the responsibility and sustainability of technological developments.

Such an approach is to be pursued when it comes to the development of technology, including but not limited to AI, in society.

Therefore, the primary stake, in laying the foundations for the creation of artificial integrity, is the establishment of basic principles concerning the inclusion of technology in the broader ecosystem, thus beyond just its operating system: the Digital for Good principles.

To grasp the essence of what these principles are, it is essential to acknowledge that technology, even when created with the best intentions for the greater good, is not inherently immune from producing adverse effects in society.

Technology itself is not inherently sustainable or positive. A comprehensive understanding of Digital for Good starts by thinking against oneself, especially as a technologist, with the acknowledgment of some critical paradoxes, paving the way for a more mindful and responsible approach to technological development in society.

**There's No Technology Without Environmental Debt.** If technology is of great help and promise in the fight against climate change, the production and disposal of technology itself often contribute significantly to environmental degradation. It is primarily due to the use of natural resources and the impact of the manufacturing process on the environment (Freitag C. et al., 2021). The shift to electric vehicles is a concrete illustration. While the aim is to reduce emissions, it requires a significant increase in lithium, cobalt, and other materials, potentially leading to ecological damage and geopolitical tensions over resource control. Moreover, the end-of-life disposal of these technologies poses challenges in terms of recycling and waste management.

**The Chaos Theory Applies to Technology.** While technology is rooted in scientific discoveries and principles, the broader impacts and consequences of its implementation may not always be entirely predictable. Technological interventions in complex ecological and social systems can have unpredictable and nonlinear outcomes due to the interconnected nature of these systems. Scientifically, this is understood through chaos theory (Curry D. M., 2012). For example, introducing a new technology in agriculture might unexpectedly affect local biodiversity or social dynamics in ways that are difficult to predict and control.