HYPER GIANT

THE NEED FOR RECIPROCAL ARTIFICIAL INTELLIGENCE

AI THAT SERVES THE BEST NEEDS OF HUMANITY

BY ROSS A. MCINTYRE,
VP OF STRATEGY, HYPERGIANT

> Intelligent machines will embody values, assumptions, and purposes, whether their programmers consciously intend them to or not. Thus, as computers and robots become more and more intelligent, it becomes imperative that we think carefully and explicitly about what those built-in values are.
>
> — M. Mitchell Waldrop[1]

As technology rapidly advances, so does the need for a communal and collective strategy that provides guidelines and laws for its development. The lack of regulation and collective guidance of technology has resulted in a number of technological innovations that have hurt the collective fabric of our society. Our response to this is the creation of ethical AI standards at the individual and corporate level. However, we do not think these standards sufficiently address the challenge of governing rapidly changing technology—it places a strenuous onus on individuals and corporations to act as the arbiters of social ethics.

It is society's responsibility to create a framework for the development of responsible AI.

The ethically responsible application of artificial intelligence will not sufficiently address the societal problems of today, much less those of tomorrow. Given the alarming increase in cases of AI being used to corrupt or degrade our social fabric, we must do more to ensure an alignment between the needs of humanity and the rapid escalation of technological progress.

While the development of ethical AI is critical to artificial general and artificial super intelligence, it stops short of the necessary steps and philosophies for the potential of our AI future. We are on the precipice of defining what AI can and should look like. It is our moment to create the guidelines for how this technology operates and to develop a road map for the type of technological future we would like to see.

More than creating ethical artificial intelligence, we must ensure AI acts in service of the best interests of humanity. This means we must do more than declare that our technology does no harm. Instead, we must begin to certify that our technology is in service of human interests, advances human life, and is harmonious with societal structure and focused on the improvement of the human experience. AI must be reciprocal.

Ethical AI does no harm; reciprocal AI, the term we have coined to discuss this new paradigm, argues we must remove the potential for harm and develop artificial intelligence with a holistic, long-term view on ramifications. We must go further to protect human society. We must start by focusing on the intentional creation of AI that serves the best needs of humanity.

How Do We Define the Best Interests of Humanity? The Philosophical Precedent

It is complicated to define the best interests of humanity. To start, one might look at philosophy or the philosophical precedent applied to other industries that are required to care for the human condition. The most obvious industry is medical. The medical industry looks to the Hippocratic oath, established circa AD 275, as the standard for its ethical guidance. One of the oldest binding documents in history, it's focused on core principles: to remediate the ill to the best of one's ability,

to maintain personal confidentiality, and to convey the teachings of medicine to the next generation[2]. In our common understanding, it's reduced to one simple dictum: "Do no harm."

A similar sentiment is often referred to in the technology industry and was seen clearly in Google's original code of conduct, through the motto "Don't be evil." Notably, when Google restructured beneath the aegis of Alphabet in 2015, this motto was modified and further deprecated by 2018, when the motto was removed from the code of conduct preface and retained in the last

sentence, albeit in a somewhat qualified version[3], that reduces one of the company's previously central ideals to, essentially, an afterthought. Technology companies have historically distanced themselves from the technology they have created and any negative human impacts. Largely, neither consumers, governments, or the market have forced them to change or improve their approach (except insofar as markets respond to poor user experiences).

There is a certain inherent weakness to establishing frameworks defined only by negative space—as in the case of "Don't be evil" or "Do no harm." Negative space represents actual shapes that share edges with the positive space. So, speaking about what one might do defined primarily by what one won't do is problematic[4] – there is no more fruitlessly perpetual task than enumerating what something is not; literally everything (aside from the object, entity or concept) must be catalogued to form such a definition.

It cannot be the job of technology companies to individually act as arbiters of what products are harmonious with, or best enhance, humanity. That job must reside broadly in our government and civil society. Beyond the complex technical challenges that arise in ensuring ethical conduct in artificial intelligence and robotics, there exists moral questions regarding the disintermediation of humans: How can we ensure machine intelligences[5] engage in moral reasoning? Can it be safely defined by what such an entity does not do? Or, do we define it within a broad cultural understanding of what we can do well? These questions require we apply a moral filter to the questions of technology development, and that moral filter must be decided broadly by our society.

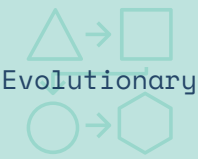
From a business perspective, there may be no significant measurable economic value associated with aggressive ethical compliance (beyond the conceit that things that are positive for society contribute to overall stability and are, therefore, good for business), and this is problematic. It requires a shift in thinking away from the immediate needs of the business.

So, how do we define "good" rather than defining what is evil, harmful or insufficiently compliant with our concept of "good?" "Common good" or "social good" benefits society as a whole, rather than solely the individuals who make up that society. It is a denial of the idea that society is composed of individuals living in isolation from their fellows.

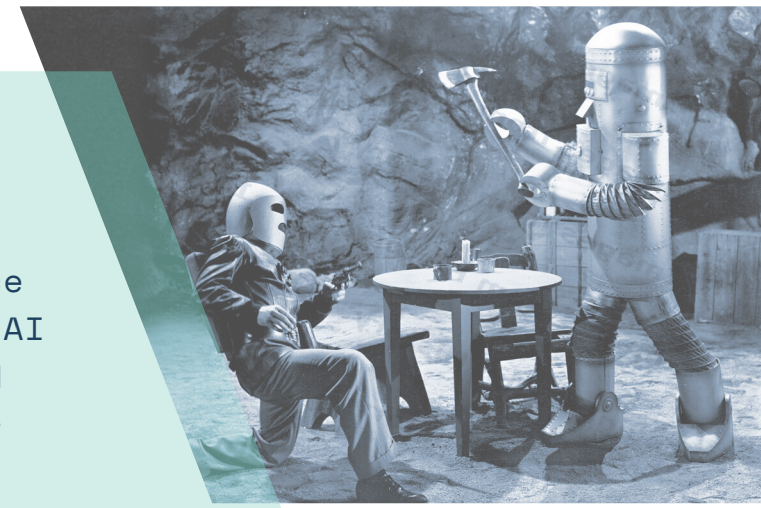
Since the days of the Greeks, political philosophy has indicated certain good things—security and justice, for example—can only be achieved through citizenship and public participation. This is supported by the works of Aristotle, Macchiavelli, and Rousseau. It is demonstrated clearly in republicanism, which posits the common good must be achieved through political means including participation in self-government. Good ethics, in this context, are moral principles that direct decision-making toward efforts that support the common good.

From Model to Super-Model

For many scientists, researchers and developers, the endgame of artificial intelligence is the creation of general AI that meets or exceeds the capabilities of a human brain[6]. One of the more spectacular contradictions in the field is what seems difficult has proven easy, while the most elusive successes are those that seem, at first blush, to be the easiest.

Model	Core Principle	Importance
 Symbolic	"[Artificial intelligence] is about algorithms enabled by constraints exposed by representations that model targeted thinking, perception, and action."	Symbolic AI is utilized by applications that require environmental perception and the ability to act thereon.
 Connectionist	Intelligence emerges from a vast array of interconnected units resembling the electrical activity of biological neurons.	These models were originally believed to be more conducive to re-creating learning, cognition and memory, but such models are based on an incomplete understanding of the brain's complexity.
 Evolutionary	Focuses on imitating the evolution of complex organisms in order to achieve programs that could automatically improve themselves."	Evolutionary algorithms can serve as adaptive substitutes for deep learning; they are considered well suited for cloud-based or massively parallel multicore processing[7]
 Corporeal	The only way to approximate human intelligence ("strong AI") would be to possess a body that facilitates interaction with the world.	Moravec's Paradox posits that reasoning itself requires little computation, but sensorimotor skills require vast computational resources[8]

How Do We Define the Best Interests of Humanity? The Corporate Precedent



Many companies embrace “do no evil” thinking, especially in regards to artificial intelligence. The evidence lies in the structure of companies’ AI ethical standards; many are oriented toward what should be avoided rather than what should be done. Yet as AI becomes more pervasive and autonomous, we must carefully consider the benevolent and malevolent[9] uses of this technology and move toward a “do good and not evil” moral prerogative. For instance, AI can enable autonomous delivery of medicine via unmanned aerial vehicles just as readily as enabling lethal autonomous weapons intended to kill human combatants.

Benevolent AI Use Cases	Malevolent AI Use Cases	Indifferent Gone Malevolent
Mining and analyzing biomedical data for the identification of molecules utilized in failed clinical trials to predict how the same (or similar) compounds can be leveraged in targeting other maladies.[10]	Creation of "lethal autonomous weapons" (LAWs) is abundantly possible[11] (and could, potentially, have already been developed outside of public awareness). The United Nations Secretary-General has urged the organization to restrict development of LAWs.[12]	In May 2010, a Waddell & Reed automated algorithm prompted a loss of 9% of Dow Jones index value in a so-called "flash crash" through large trading bets on Chicago's derivatives exchange.[13]
Utilizing fragment-detection algorithms in combination with machine learning and natural language processing, Quill aims to help 30 million low-income K–12 students in the U.S. become stronger writers.[14]	The president of India's ruling Bharatiya Janata Party (BJP), Manoj Tiwari, used "deepfake" technology to make a campaign video look like he was speaking in that region's dialect.[15]	Microsoft released a Twitterbot named Tay in 2016 as an experiment in conversational understanding. Within 24 hours, across 96K tweets, Tay was parroting back racist and misogynist content.[16]
Hypergiant's EOS bioreactor captures and sequesters atmospheric carbon utilizing algae through AI-driven algae growth, carbon capture, and algae output.[17]	In 2016, scientists from security firm ZeroFOX created an AI phishing tool (SNAP_R) intended to incite users to click malicious links. The AI was able to drive such a campaign far more effectively than human agents.[18]	In early 2018, Google's advertising algorithm was discovered to be biased, showing high-paying executive positions to male job seekers and not women.[19]

When we create AI, we must move intentionally, not thinking solely about technological achievement but also about intent and future intent. There are fewer instances of people creating technology with malicious aims than instances of people creating technology without thinking exhaustively of consequence. The latter cases are often the most pernicious and detrimental.

Commercial AI software providers often have mature product management disciplines, and ethical concerns are often addressed by a company's underlying business model or ethos. Taking that into account, positive societal impact is either a Key Performance Indicator (KPI) or it is not – product managers will likely follow the precedent set by a company's core values. Thus, there is much more extensive damage that can be done when a company's employees have no alternative to utilizing the software mandated by their organization. Absent the market impacts of producing commercial AI software that is in some fashion antithetical to the common good, companies likely need to establish principles that drive development toward positive integration of reciprocal principles.

Rather than focusing on what AI isn't, we must consider what AI should be.

What's in It for Me? Enter Reciprocal AI

Reciprocal AI is the transcendence and evolution of ethical AI.

Reciprocity is defined as "the practice of exchanging things with others for mutual benefit."

As a social norm, reciprocity reflects an expectation that people will respond in kind to any particular stimulus—responding to gifts and kindness with corresponding benevolent responses and responding to harmful acts with retaliation (or indifference, at least). These norms can be crude and transactional, like the Code of Hammurabi or *lex talionis* – the laws of retribution that are best known as "an eye for an eye.[20]" These norms can also be complex and nuanced as a subtle understanding that the role of anonymous donations can still be a form of quid pro quo. These norms change on a situational basis and vary between societies, but anthropologists have suggested some form of the norm is a societal inevitability.[21]

AI encompasses a broad range of techniques and technologies: machine learning, neural networks, natural language processing, computer vision, and more. It is capable of providing well-defined customer value provided one has critically identified the business problem at hand. From the simplest product suggestions to the most advanced autonomous vehicles, AI (and the entire tapestry of the technologies it includes) is being utilized to solve problems both mundane and fantastical. The rise of artificial intelligence is often characterized as the Fourth Industrial Revolution.

If history holds true with this particular revolution, some jobs will be destroyed, but a far greater number will be created. Reciprocal AI is focused on the promotion of AI use cases that deliver both business value and human value.

In other words, the creation of solid business cases alone is not sufficient, nor is simply ensuring they are ethically compliant. Technology should serve both business needs and human needs. There is no exclusivity.

In order to ensure AI is reciprocal, it must exhibit the principles of fairness, bidirectionality, and responsibility. Reciprocal AI requires an ongoing relationship between mutual beneficiaries in order to ensure bidirectional recompense. This is unlike the transactional relationship that pervades "digital transformation" efforts, which is characterized by an exchange, e.g.,paper-based processes being replaced by digital workflows. Business evolution toward horizontally integrated artificial intelligence is more relationship-based, e.g., the creation of knowledge graphs are long-term efforts and can rarely be positioned as 1:1 replacements for previously existing workflows or processes.[22]

Positive externality stands to reason that producers and consumers will both benefit, but must also benefit a (potentially unknown) third actor. For example, education directly benefits the individual but also provides benefits to the broader society. Similarly, Tim Berners Lee, who created the World Wide Web and made it freely available, created a large positive externality. Reciprocal AI can provide mutual benefits to the involved parties and still serve positive externalities.

Fear of a Black Box: The Pursuit of Evolved Ethics for Machines

Prior to the late 20th century, the ethics of machines was largely the purview of sci-fi authors, but significant progress in the field of AI has led to a renewed focus on the field of Machine Ethics. Digital Ethics and Privacy has emerged as a significant topic at all major industry research companies. Gartner's Top 10 Strategic Technology Trends for 2019 argued companies must focus on doing what is best for the business and doing what is ethical with regard to AI-specific issues or adjacent technologies. Companies and organizations that are transparent about their pursuit of a digital ethics approach will build trust with users and enhance their brand.

Those concerned about the ascendance of artificial intelligence often focus on the potential for biased, deceptive, malicious, or humanity-devaluing applications. Luminaries such as Elon Musk[23], Bill Gates[24] and Stephen Hawking[25] have all commented publicly on the potential existential risks associated with superintelligence.

Companies building such technologies are responding by creating corporate principles meant to ensure that systems are designed and deployed in a way that is harmonious with such positions. However, just as Google changed direction in its code of conduct, many are convinced these well-meaning promises will not be kept.

This further exemplifies why ethical principles are the domain of government regulation. On February 11, 2019, the Trump administration issued an executive order to establish the "American AI Initiative" primarily to improve AI R&D initiatives, augment workforce development, and support international engagement. The government followed up in January 2020 with "Artificial Intelligence for the American People."[26] This set of regulatory principles focuses on the private sector and are oriented toward three objectives: ensuring public engagement, limiting regulatory overreach, and promoting trustworthy technology. The principles are open for public comment on the whitehouse.gov site, but they are toothless from the perspective of enforcement. In an effort to achieve their middle objective—limiting regulatory overreach—they establish themselves as philosophies, bereft of authority and consequently, without mandatory adoption in the public sector where artificial intelligence improvements are prevalent.

The contrasting perspective (one taken by the #FreeAI Manifesto) believes that success lies in multiplicity, decentralization, and open economy rather than regulation and standardization.

So, can ethical frameworks for the development, deployment, and usage of artificial intelligence ever work? Is reciprocity a reasonable goal?

The Central Tenets of Reciprocal AI

At its current evolutionary state, artificial intelligence remains almost entirely under the control of its creators. There is little an AI application can do—for good or evil—that could not be planned for, provided developers are diligent and thorough in examining potential scenarios or consequences. The scope, scale and potential that underlies artificial intelligence makes it among the most consequential suites of technologies employed by humanity, which makes it necessary to be more thorough in our reasoning.

Over the past 20 years of software development, creators have trended toward a user-centric approach that accounts for the needs of the end users in providing function and context through progressive disclosure of information. This has slowly become table stakes for software developers as it results in superior software. Similarly, implementation of reciprocal AI must become the new baseline and all parties involved in such development must ask themselves if a given solution could do more than satisfy a discrete business need. The design thinking approach, which is based in user-centricity, ensures that software is responsible to the user; reciprocal AI must be responsible to groups of users and should have an aggregate, societal perspective.

Reciprocal AI requires:

- Deliberate intentionality in respect to end goals
- Equanimity in the effort expended and effort saved
- Bidirectionality in respect to data provided and data returned
- Egalitarianism of social responsibility

This necessitates transcending simple (or even complicated) compliance. It necessitates a more expansive evaluation of what is "good" or "valuable" to a company, and an acceptance of (if not an appetite for) non-fiduciary returns and side effects.

Reciprocity requires intention and deliberateness.

Our vision of the future is one where we have reciprocal AI technology — AI technology that goes beyond "do no harm" and instead focuses on a mutually harmonious relationship with humanity. We call this future version of AI reciprocal AI because we insist this AI results in a positive relationship with humanity. Anything short of a positive and reciprocal relationship puts humanity at huge risk. There is no room for error.

To get there, we believe we must:

1. Focus on building technology that sets the precedent for positive AI and human engagements.
2. Develop government regulations that direct technology companies to engage with government and social organizations on the development of new AI technology.
3. Sets a government-level (or global) vision for the future AI we wish to develop.

This core level of investment and organization will allow us to ensure that all AIs developed are aligned with the commonly defined interests and values of our society. Like regenerative systems but taken further, in the future reciprocal AI will be focused on:

- Advancing human life and the human experience
- Improving the societal structure of humans and AIs

This does not apply to the development of every AI solution. Instead, we should apply reciprocal AI thinking to the development of core game-changing AI technologies.

What to Do Today

Today businesses, governments, and technologists must focus on the development of ethical AI. We must understand and have broad questions about morality and technology that does no harm. Then we must take it one step further by orienting toward active collective benefit. It is about moving past ethical equilibrium, amplifying the positive, and asking what more can be done.

For businesses looking to leapfrog ethical AI and set the standard for reciprocal AI, you must ask a number of questions about the products you are developing:

1. Does this product benefit humanity as a collective whole?
2. Does this product improve not just a business function but a societal function?
3. Does this product contribute significantly to the core development of AI?
4. Does this product leave the world better than it was before?
5. Is there a way of rethinking the expression/ mechanism of AI to make it better in general?
6. What's the opportunity cost?
    a. New revenue streams and business models
    b. Doing good as a way to make more money

This is a new framework of thinking about business: positive intent first business. This model really works better.

Ethical AI should be reflexive. Reciprocal AI is both aspirational and completely necessary. We are working to create a community of people aligned with our belief systems who see this as transformative for the future of the industry. Please sign up here to stay connected and to learn more.

Endnotes

1 “A Question of Responsibility,” by M. Mitchell Waldrop, AI Magazine, 8(1), 28, 1987
2 “Medical Definition of Hippocratic Oath,” by William C. Shiel, Jr., MD, FACP, FACR, medicinenet.com
3 In early 2018, Google eliminated the motto from the preface of the corporate philosophy. It was retained in the final line as “And remember....don’t be evil, and if you see something that you think isn’t right—speak up!”
4 Side note: an old client once asked for functional specifications that indicated everything an application would not do. It took some time, but I was ultimately able to force an understanding that defining everything an app doesn’t do would literally take forever.
5 In the OpenMind BBVA article “The Future of AI: Toward Truly Intelligent Artificial Intelligences,” by Ramón López de Mántaras, the principal artificial intelligence models identified are Symbolic, Connectionist, Evolutionary, and Corporeal. The plurality here is a reflection of those different models, the ongoing need to distinguish between “Strong” and “Weak” AI, and an understanding that AIs may be employed on unconnected solutions, i.e., more than one purpose-built artificial intelligence.
6 “The Future of AI: Toward Truly Intelligent Artificial Intelligences,” by Ramón López de Mántaras, OpenMind BBVA, in “Towards a New Enlightenment? A Transcendent Decade,” 2018
7 “New Twist on AI Evolutionary Algorithms in Neuroscience,” by Cami Rosso, Psychology Today, May 24, 2019
8 “It is comparatively easy to make computers exhibit adult level performance […] and difficult or impossible to give them the skills of a one-year-old.” —Hans Moravec
9 Here, I utilize the loose definition found in the “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation” written in February 2018 by a consortium of 26 authors from 14 institutions, spanning academia, civil society, and industry: “all practices that are intended to compromise the security of individuals, groups, or a society.”
10 “This AI Unicorn Is Disrupting the Pharma Industry in a Big Way,” by Joao Medeiros, Wired, 2019.
11 “Death by algorithm: the age of killer robots is closer than you think,” by Kelsey Piper, Vox.com, June 21, 2019.
12 “Autonomous weapons that kill must be banned, insists UN chief,” UN News, March 29, 2019.
13 “The 2010 ‘flash crash’: how it unfolded,” by Jill Treanor, The Guardian, April 22, 2015.
14 “AI for Good—An Overview of Benevolent AI Initiatives,” by Kumba Sennaar, Emerj, November 22, 2019.
15 “An Indian politician used AI to translate his speech into other languages to reach more voters,” by Kim Lyons, The Verge, February 18 2020.
16 “Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day,” by James Vincent, The Verge, March 24, 2016.
17 “Hypergiant Industries Launches Eos Bioreactor: A Prototype Bioreactor for Commercial and Personal Carbon Sequestration,” by Hypergiant Industries, PRNewswire, September 17, 2019.
18 “Hackers Have Already Started to Weaponize Artificial Intelligence,” by George Dvorsky, Gizmodo, September 11, 2017.
19 “Probing the Dark Side of Google’s Ad-Targeting System,” by Tom Simonite, MIT Technology Review, July 6, 2018.
20 “Code of Hammurabi,” History.com Editors, History.com, September 9, 2019.
21 “The Norm of Reciprocity,” by Alvin Gouldner, American Sociological Review 25, 1960.
22 There is an inherent ridiculousness to the concept of “digital transformation:” Transformation implies that there is a moment at which the transformation is complete and the business can move on to other matters. AI adoption is more evolutionary and continues as new opportunities arise or new technologies are created.
23 “Elon Musk’s nightmarish warning: AI could become ‘an immortal dictator from which we would never escape,’” by Peter Holley, The Washington Post, April 6, 2018.
24 “Bill Gates: A.I. is like nuclear energy — ‘both promising and dangerous,’” by Catherine Clifford, CNBC.com, March 26, 2019.
25 “Stephen Hawking warns artificial intelligence could end mankind,” by Rory Cellan-Jones, BBC.com, December 2, 2014.
26 “Artificial Intelligence for the American People,” whitehouse.gov.

HYPER
GIANT