CEI Center for Ethical Innovation

Ethical Innovation in Artificial Intelligence – A Five-Dimensional Perspective

Rowland Chen[1], M.S., M.B.A., Chesley M. Chen[2], M.S., M.B.A., Jennifer Crakow[3], M.A.,

and Robert E. Tench[4], Ph.D.

The conversation around ethical innovation in artificial intelligence (AI) has evolved significantly as capabilities have grown, centering not only on what AI can achieve but also ensuring actions align with beneficial and moral standards. Ethics in artificial intelligence involves at least five critical dimensions:

1. Human-centered ethical design for AI
2. Embedding ethics within AI technologies
3. Ethical use of AI
4. Change in people's attitudes and mindsets
5. Motivation for ethical behavior

A *status* quo exists which is a default position where people already make ethical choices regarding the development, deployment, and use of AI without specific regard to the ethical decisions they are making. Whether people are aware or not, they currently make ethical decisions when choosing to develop and use AI. These unconscious decisions represent *de facto* choices and result in *de facto* consequences, running the risk of causing harm rather than delivering beneficial outcomes.

The conscious and deliberate implementation of the five dimensions begins to address the ethical decisions and choices people make about AI's design and use. Implementing these dimensions is transformative in nature requiring a comprehensive effort spanning human behavior, conscience, technology, and regulation. Collaborative work lies ahead. Understanding the five-dimensional perspective of ethics in artificial intelligence and their underlying concepts requires the exploration of their meanings, implications, and the challenges they present.

Dimension 1: Human-Centered Ethical Design

Human-centered ethical design (HCED) is a multi-step method for deeply architecting strategy, process, organizations, products, and services in alignment with genuine customer needs and aspirations. In the case of applying HCED to AI technology, it represents a radical change in how decisions are made regarding which AI systems to develop and deploy and for what use cases. The steps in HCED are

- Research
- Divergent thinking
- Convergent thinking
- Generative decision-making
- Pen to paper

The major difference between HCED and traditional design thinking is the incorporation of ethics in every step of the approach. It is not an afterthought added at the final step for justification. Details of each of the five HCED steps are described below. "Red teams" are in use, such as at OpenAI with its Sora product. A red team enters the innovation picture to test for ethics and bias after a product has been designed and

developed. It is a *post facto* stop-gap measure. A more effective approach is to track against ethics and morality from the start of innovation.

Research

In the realm of human-centered ethical design, research guides every subsequent decision and action. When applying HCED to AI technology, research involves a thorough understanding of the end users' needs, behaviors, environments, and challenges. Research in HCED goes well beyond data collection. It seeks to uncover the human contexts in which AI will operate.

Ethical considerations are at the forefront, ensuring that AI solutions are developed with an awareness of potential biases, privacy concerns, and broader societal impacts. Engaging with a diverse group of stakeholders, including those who might be marginalized or adversely affected by AI, is crucial to identify and address ethical dilemmas early in the development of AI innovations. The comprehensive approach to research ensures that AI technologies are not only technically sound but also socially responsible and aligned with human moral values.

Divergent Thinking

Divergent thinking in the context of HCED for AI involves exploring a wide array of potential solutions without being confined by current constraints and existing technologies. Divergence encourages creativity and innovation, allowing designers, developers, and business people to brainstorm a multitude of ideas that address the users' needs identified during the research phase. The goal is to think broadly about possible applications of AI, considering various scenarios, contexts, and impacts. Ethical considerations are integral, as each idea is evaluated not just for its technical feasibility, innovativeness, and commercial viability, but also for its potential consequences on people and society. Expansive thinking helps to uncover unique, inclusive, and often unconventional solutions that prioritize human well-being and ethical standards.

Convergent Thinking

Following the phase of divergent thinking, convergent thinking brings focus and clarity to the HCED approach by narrowing down the broad array of ideas to the most viable, impactful, and ethical options. Convergent thinking involves critical analysis and synthesis of the generated ideas, considering risks and ethical concerns. Convergence is crucial for aligning AI development with human-centric values, as it requires a careful evaluation of how each solution serves the end users' needs and adheres to moral principles. Decisions made during convergence are informed by a deep understanding of the research findings, ensuring that the selected AI applications are not only innovative but also responsible and beneficial to society.

Generative Decision-Making

Generative decision-making (GDM) in the HCED process for AI is about creating pathways that enable the chosen ideas to evolve into tangible solutions. GDM is characterized by iterative and rapid development, where generative AI itself is used to support decisions about design and development of prototypes and decisions inherent in product development. The emphasis is on co-creation, involving users in design decisions to ensure that the AI solutions are accessible, intuitive, and truly meet their needs. Ethical considerations are embedded in each iteration, with ongoing assessments of how the AI impacts user privacy, autonomy, and trust.

Pen to Paper

Pen to paper (PTP) signifies the transition from conceptual design to actual implementation. PTP requires carefully selected, ethically vetted, and user-validated ideas be documented and translated into actionable design and development plans. Specific deliverables of PTP, detailed specifications, design guidelines, and ethical frameworks, are articulated, providing a clear roadmap for developers, designers, technologists, philosophers, cultural anthropologists, and other subject matter experts. The documentation process ensures that the human-centric and ethical considerations remain at the forefront throughout the development lifecycle of an AI system. With documentation the team commits to creating AI solutions that respect and enhance human dignity, foster inclusivity, and are accountable to the individuals and communities they serve. The deliverables embody the commitment to not only imagine but also realize AI that contributes positively to society. In other words, HCED forms a social contract for ethical innovation in AI.

Dimension 2: Ethics Embedded within Artificial Intelligence

Ethics embedded within AI involves making the technologies themselves intrinsically ethical. Required are the design of AI systems capable of understanding and adhering to ethical principles autonomously. To date, embedded human traits into AI has proven a daunting technical challenge. Machine learning is possible. True machine creativity is elusive. Embedding ethics can be represented by a set of AI design principles for researchers, developers, and product managers to follow in their innovative new applications of AI. The following are required for embedded ethics:

- Unbiased training data and accessed information
- Algorithmic fairness of software designers and developers
- Moral decision-making in a global context
- Value alignment with cultural norms
- Ethical reasoning
- Autonomy and consent of humans

Unbiased Training Data and Accessed Information

The importance of unbiased training data and accessed information in embedding ethical innovation within artificial intelligence is an essential requirement for the development of AI systems that are fair, equitable, and reflective of a diverse society. Unbiased data ensures that AI algorithms generate decisions based on a balanced representation of the real world, avoiding the perpetuation of historical biases and discrimination that can arise from skewed datasets. By prioritizing the removal of biases in training data and ensuring the information accessed by AI systems is comprehensive and diverse, designers and developers can create AI technologies that are more inclusive and ethically aligned with societal values. This approach not only enhances the reliability and trustworthiness of AI but also ensures that the benefits of innovation are accessible to all segments of society ensuring a technological landscape that champions diversity and equity. As AI continues to influence every aspect of our lives, the commitment to preventing bias becomes not just a technical necessity but a moral imperative to ensure ethical innovation is embedded in the heart of AI design and development.

Algorithmic Fairness of Software Developers

Creating algorithms that can make unbiased and non-discriminatory decisions is an intricate process that necessitates advanced methods to both identify and mitigate biases present in training data and the decision-

making processes themselves. Equitable systems ensure that all individuals and groups receive equal treatment. The journey toward algorithmic fairness involves deploying a variety of techniques, such as auditing datasets for bias, employing fairness-aware machine learning models, and continuously monitoring outcomes to safeguard against unintended discriminatory effects. The goal is to develop AI that not only perform their intended tasks efficiently but do so in a manner that is just and unbiased, reflecting a commitment to fairness.

Moral Design Decision-Making in a Global Context

For AI systems tasked with scenarios that necessitate moral judgments, embedding ethics in decisions concerning design is a critical step, integrating moral frameworks that enable these systems to analyze various values and make decisions. These decisions must align with accepted and established ethical and moral principles. The complexity of human moral reasoning appears at the individual human, family, community, and societal levels. Achieving this level of sophistication in AI requires a multidisciplinary approach, drawing from philosophy, cognitive science, and technology to design algorithms that can navigate ethical dilemmas. These systems must be capable of considering the broader implications of their actions, ensuring that decisions are made with a conscientious understanding of right and wrong, thereby embodying the essence of moral decision-making in the digital realm.

Value Alignment with Cultural Norms

Ensuring that AI objectives are in harmony with human values is an essential element for ethical integrity. Value alignment encompasses the establishment of goals that adhere to ethical standards while devising strategies to achieve these goals without causing unintended adverse effects. Value alignment is a challenge that necessitates ongoing dialogue among technologists, ethicists, and the broader public. Alignment calls for a concerted effort to embed human values into the fabric of AI development as exemplified by HCED, ensuring that as AI becomes more integrated into our lives, it behaves in ways that are beneficial and non-harmful which truly reflect human ethical standards.

Ethical Reasoning

Envisioning AI systems capable of ethical reasoning extends beyond merely programming decisions based on fixed ethical guidelines. The capability involves the development of AI that can assess various actions in new and complex situations to identify the most ethical path forward. These ambitious goals require AI to be endowed with a deep understanding of ethical principles and the ability to apply these principles across a spectrum of scenarios that are pre-trained as well as untrained. Crafting such systems demands a blend of technology, philosophy, and practical ethics, aiming to create AI that knows what is right and can discern with artificial conscience the ethically best course of action in circumstances that are ambiguous, unprecedented, and untrained.

Autonomy and Consent of Humans

As AI systems evolve to operate with greater autonomy, it becomes imperative to ensure they respect human autonomy and the principle of consent. Designing AI that actively seeks and honors consent, particularly in applications where personal sensitivity and privacy are at stake, is crucial. Artificial consent involves creating mechanisms within AI solutions that prevent manipulation, deceit, or coercion of users. Ensuring respect for human autonomy and consent requires a deliberate design philosophy that places these values

at the core of AI development, reflecting a commitment to safeguarding human dignity and freedom in an increasingly automated world.

Dimension 3: Ethical Use of Artificial Intelligence

The ethical use of AI revolves around human decision-making concerning the design, development, deployment, and governance of AI systems. This use encompasses a broad range of ethical considerations influencing how technology impacts individuals and societies. These considerations include:

- Beneficence
- Privacy and data protection
- Transparency and explainability
- Accountability and responsibility

Beneficence

The principle of beneficence in ethical AI underscores the imperative that AI technologies be deployed with the best intentions of yielding positive societal impacts. The beneficence principle calls for AI applications to enhance productivity, tackle intricate challenges, and elevate the overall quality of life, while striving to minimize potential harm. It necessitates a thorough evaluation of the implications that AI technologies may have across various sectors, such as healthcare, education, security, and finance, to ensure their contributions are beneficial and align with the broader social good. The commitment to beneficence demands vigilance against the risks of unintended negative outcomes, advocating for a proactive stance in leveraging AI as a force for societal improvement and human welfare.

Privacy and Data Protection

In the era of data-driven technology, AI systems' reliance on extensive datasets introduces profound privacy and data protection concerns. Ethical deployment of AI requires the adoption of stringent safeguards to prevent the misuse or unauthorized access to personal information. Implementing comprehensive data protection measures, securing informed consent, and empowering individuals with the ability to manage their own data necessitate a deep integration of ethics and technology. The aim is to forge a framework within which AI operates not only with efficiency but with an unwavering respect for the privacy rights and data sovereignty of individuals, thus fostering a trust-based relationship between technology and people.

Transparency and Explainability

Transparency in AI demands that the operations of AI be accessible and comprehensible to users and stakeholders, promoting a culture of openness. Explainability goes a step further, insisting that the reasoning behind AI-driven decisions be conveyable and justifiable in terms that humans can understand. AI can know longer be treated as a “black box” which has unknown inner workings. Clarity is essential for fostering trust, particularly in sensitive applications such as medical diagnosis or judicial decision-making. Ensuring that individuals can grasp how AI solutions reach their output enables a more informed public conversations around AI, reinforcing the accountability of AI developers and deployers, and building confidence in AI technologies as tools for public good.

Accountability and Responsibility

Accountability and responsibility in the context of AI revolve around the crucial need to identify and attribute responsibility when AI systems malfunction or cause harm. The assignment of accountability—whether it falls to the designers, developers, the entities deploying the AI, or potentially the AI systems themselves—is fundamental to the ethical use of technology. Establishing clear mechanisms for accountability ensures that AI is used responsibly, with well-defined protocols for addressing and autonomously correcting errors. This framework not only supports ethical standards but also underpins the trust between AI technologies and the societies they serve, ensuring a path for recourse and correction in the face of mistakes.

Dimension 4: Changes in Attitude and Mindset

The seamless integration of artificial intelligence into everyday life necessitates a significant transformation in how we perceive and engage with technology. As AI increasingly influences critical areas of human activity, such as healthcare, financial services, and criminal justice, the need for a shift in societal attitudes and mindsets towards AI's ethical implications becomes essential. The shift hinges on five critical factors each playing a significant role in managing change effectively:

- Awareness
- Multidisciplinary collaboration
- Continuous education
- Regulation and governance
- Ethical leadership

Awareness

Despite the growing prevalence of AI, there remains a significant gap in understanding among designers, technologists, philosophers, political scientists, policymakers, and the public about how AI decisions can exacerbate discrimination or undermine privacy and autonomy. Elevating awareness entails educating all stakeholders on the critical importance of incorporating ethical considerations from the ideation stage through to deployment to adoption to utilization, while also spotlighting instances where AI has failed in order to illustrate the tangible consequences of overlooking ethics. Such awareness cultivates a culture of accountability and vigilance, ensuring ethical challenges are preemptively recognized and addressed.

Multidisciplinary Collaboration

Addressing ethics in AI is not a task for designers, technologists and AI developers alone. It demands contributions from a wide array of disciplines, including ethics, law, social sciences, and humanities. Multidisciplinary collaboration introduces varied perspectives, enabling a more comprehensive approach to identifying and addressing ethical issues. By weaving ethical, cultural, and societal considerations into AI development, technologists can engineer systems that are not just technologically advanced but also socially conscientious. A collaborative endeavor ensures AI technologies are crafted with a profound respect for human values, leading to outcomes that are equitable and beneficial.

Continuous Education

Given AI technology's rapid advancement, ethical guidelines and standards must evolve accordingly to stay relevant. Ongoing education is vital for designers, developers, users, and regulators to keep abreast of new developments and ensure ethical considerations are woven into AI innovation. Education includes

continuous training in ethical design principles, awareness of new ethical challenges, and updates on best practices for reducing biases and promoting fairness. Cultivating a culture of lifelong learning among AI stakeholders allows for the ethical norms and practices to adapt to emerging challenges, ensuring AI remains a positive force.

Regulation and Governance

The creation and enforcement of regulatory and governance frameworks influence people, how they perceive AI systems, and set the tone of an ethical environment. Regulations have the potential to be guideposts for people to operate ethically and how product managers, designers, developers, and senior leaders can be held accountable for AI system behavior. These frameworks involve setting clear guidelines, standards, and oversight mechanisms that delineate ethical AI usage.

Additionally, governance outlines responsibilities for AI generating entities be they corporations, universities, government agencies, or other organizations with people at the forefront of AI advances. Regulation should balance the promotion of transparency, fairness, and privacy protection with the encouragement of innovation in the people and groups driving AI innovations. Effective governance also requires international collaboration to tackle the global implications of AI technology. By implementing robust regulatory structures, societies can protect ethical principles in AI, thus preventing harm and fostering trust in AI applications.

Visible Ethical Leadership

Instilling ethics in AI necessitates visible ethical leadership at all levels, from AI engineering teams to corporate leaders to policymakers to heads of nations. Leaders must embody ethical values, demonstrating their commitment to responsible AI development, regulation, and usage. Leadership involves role modeling, prioritizing ethical considerations in decision-making, dedicating resources to ethical AI programs, and promoting ethical practices within their ecosystems. Ethical leaders are instrumental in cultivating a culture that prizes high ethical standards, encouraging open discussion, and rapid resolution of ethical dilemmas.

Dimension 5: Motivation for Ethical Innovation

People need incentives to do either the “right” thing or the “wrong” thing when it comes to ethical innovation in AI. As John Steinbeck wrote in “East of Eden”, a choice between right and wrong exists (*timshel*). *Timshel* is a Hebrew word that means "Thou mayest" and implies free will and choice between good and evil. Among others, the factors below influence business leaders’, AI designers’, and developers’ choices:

- Profit
- Power
- Love
- Altruism
- Conscience

Profit

Profit, the lifeblood of innovation, can be a double-edged sword. It drives research, fuels development, and incentivizes groundbreaking solutions. Medical advancements, cleaner technologies, and life-saving drugs

often blossom from the fertile ground of profitability. However, the pursuit of profit can become an insatiable beast, prioritizing shareholder gain over human well-being. Unethical practices like data exploitation, environmental disregard, and cornering markets emerge from this twisted perspective. Innovation, once a beacon of progress, becomes shrouded in ethical dilemmas. Striking the right balance between financial sustainability and ethical responsibility is crucial to ensuring that profit remains the engine of innovation, not its abyss.

Power

Power, another alluring force in innovation, can be a catalyst for positive change. It provides the resources to tackle complex challenges and champion groundbreaking ideas. Leaders with vision, armed with the power to enact change, can drive AI innovation towards a better future. However, power unchecked can become a seductive siren song, leading to manipulation, control, and silencing dissenting voices. Innovation becomes a tool for self-preservation, prioritizing the agendas of the powerful over the needs of the many. Ethical considerations get tossed aside, leading to technologies designed for surveillance, control, and manipulation. Ensuring that power serves as a guiding light, not a blinding force, is essential to harness its potential for good in the realm of ethical innovation.

Love

Love, in its multifaceted forms, emerges as a powerful counterpoint to profit and power. Love ignites a passion within us, inspiring solutions that address human suffering and cater to our deepest needs. From the innovator driven by the love for a child to the scientist seeking to cure a spouse's illness, love's tender touch steers innovation towards a path of betterment. Love presents itself as a potent motivator in ethical innovation, driving creators to solve pressing human issues and fulfill deep-seated needs. Whether it's an inventor motivated by love for a family member or a scientist aiming to cure a loved one's disease, love guides innovation towards enhancing human welfare.

Altruism

But the human spirit is not a monolith. Alongside love stands altruism, the selfless act of giving. It fuels innovation that benefits others without personal gain, a testament to the inherent human desire to contribute to the greater good. From the volunteer programmer creating open-source software to the doctor developing vaccines for neglected diseases, altruism paints innovation with the colors of compassion and shared humanity.

Conscience

Even the most noble intentions can stray from the path of righteousness. Conscience, our internal moral compass of right and wrong, plays a critical role in ensuring AI innovation aligns with ethical principles. It compels humans to question, to challenge, and to ensure that their creations do not harm the very people they are meant to serve. From the whistleblower exposing unethical practices to the engineer refusing to build weapons of mass destruction to the programmer leaving a company that inadvertently supports the creation of biased algorithms, conscience serves as the guardian angel of innovation.

Navigating the intricate environment of potentially conflicting motivators requires a delicate touch, a constant balancing act. People must weigh the potential benefits of innovation against the ethical considerations that safeguard human well-being. Only through careful consideration of these guiding

principles can we ensure that innovation serves as a force for good, a symphony of progress that uplifts humanity and protects our shared world.

Implications and Future Direction

The five dimensions of ethical innovation in artificial intelligence provide a high-level framework for moving forward with ethical product ideation, research, design, development, engineering, marketing, deployment, and use of innovative new technologies. The holistic perspective, that includes human-centered ethical design, embedding ethics within AI technologies, ethical use of AI, change in people's attitudes, and motivation, enables a firmer grasp of moral standards, requirements, aspirations, and issues.

A primary task required for ethical innovation in AI is a global agreement on what is ethical and what is not. Much research and study about ethics among philosophers already span centuries. What is different now is an urgent need to solve for ethical innovation when the establishment of universally moral standards remains elusive. A need might exist to start with the golden rule – "Do unto others as you would have them do unto you." Other historic guideposts in the Western World include the ten commandments and the seven deadly sins. In Eastern cultures, equivalents to the Western golden rule exist. In Buddhism, "Hurt not others with what pains yourself." In Taoism, "Regard your neighbor's gain as your gain, and regard your neighbor's loss as your loss." Many other analogs exist across cultures throughout time.

The journey towards ethical innovation in AI faces challenges that span both human and machine realms. Among these, the task of establishing universal ethical principles that hold true across various cultural and social landscapes stands as a formidable challenge. Difficulty stems from the diverse interpretations of ethics and morality across different societies and even within societies. This diversity can complicate the creation of a universally accepted ethical framework for AI. Moreover, the brisk pace at which AI technologies evolve often surpasses our capacity to fully comprehend and effectively govern their societal impacts. This discrepancy raises concerns about people's ability to ensure that AI development aligns with ethical standards, further complicated by technical hurdles in designing AI systems that can genuinely comprehend and embody moral principles.

Barriers to a state of ethical AI exist in technical and non-technical form. Challenges in technology and human nature are enormous and cannot be underestimated or ignored. Nor can the challenges be overstated and taken as existential threats without evidence of malicious nature in AI systems that, as of now, is purely a human trait.

Despite the apparent obstacles, the commitment to advancing ethical AI is undiminished and recognized as a breakthrough for current and future applications. Advancing AI demands a concerted effort from a wide array of stakeholders, including academic researchers, industry leaders, designers, technologists, ethicists, and policymakers, among others. The collaboration across these groups is vital for fostering an environment where ethical considerations are integrated into the fabric of AI development and deployment.

Through such multidisciplinary engagement, there exists a significant opportunity to craft standards, regulations, and best practices that guide ethical innovation in AI. The collective approach not only aids in navigating the complex ethical landscape but also ensures that AI technologies serve the greater good, respecting human values and norms specific to communities be they groups of technology users, local communities, societies, or in some cases, the world.

Looking ahead, the implementation of ethical innovation in AI will require continuous dialogue and collaboration. This forever engagement is key to addressing ever-evolving challenges and ensuring that ethical considerations integrate with technological advances. By fostering an environment where ethical discussions and decisions are prioritized, it is possible to develop AI that not only enhances human capabilities but does so in a manner that is ethically sound and socially responsible. The future of ethics in AI hinges on our ability to adapt, innovate, and collaborate, ensuring that as AI technologies become increasingly a value-adding part of people's lives and they do so in ways that uphold and promote accepted moral standards.

The Authors

1. Rowland Chen, M.S., M.B.A,, Chief Executive Officer of The Silicon Valley Laboratory, Adjunct Professor of Business at De Anza College, formerly Visting Scientist in artificial intelligence at Carnegie Mellon University
2. Chesley M. Chen, M.S., M.B.A., Chief Executive Officer of Burma Road Ventures and Adjunct Professor of Entrepreneurship and Innovation at University of Massachusetts Lowell and University of Denver
3. Jennifer Crakow, M.A., Chief Design Officer of The Silicon Valley Laboratory
4. Robert E. Tench, Ph.D., Science Advisor to The Silicon Valley Laboratory

About the Center for Ethical Innovation

A joint initiative of The Silicon Valley Laboratory and the University of California, San Diego, the Center for Ethical Innovation (CEI) provides a multi-disciplinary focal point for filling the gaps between industry and academia and between domains such as technology, philosophy, cultural anthropology, and others required to take a truly holistic approach to thought leadership and the execution of educational and business programs. The CEI's vision is simple:

> *A world where all innovation is ethical innovation, and the importance of ethics in innovation is never questioned.*

And the CEI mission is straightforward:

> *To succeed at the necessary steps towards the Center for Ethical Innovation vision:*
>
> - *To identify and develop frameworks and practices for all stages of ethical innovation*
> - *To drive awareness of the need for ethical innovation*
> - *To create a sense of urgency among all stakeholders*
> - *To deliver ethical innovation programs, including both education and implementation*
> - *To develop the next generations of innovators who place ethics at the forefront of their thinking*

For more information, please contact the CEI via email: rchen@centerforethicalinnovation.com