Working Title: Person-centered AI for executive business ethics

Authors:

Helen M. J. Alford, Ph.D., O.P.
Pontifical University of Saint Thomas Aquinas
Rome, Italy
alford@pust.it

William J. Mea, Ph.D. (corresponding author)
Georgetown University
Washington, DC, USA
wm547@georgetown.edu
(703) 789-0582

FILE NAME: AI ethics_ Alford n Mea (Double-space line numbers) 2023 11 24 Fri

I: Traditions and Frameworks

Introduction

With artificial intelligence (AI), every major business decision has a technical aspect that needs an ethical dimension to put people first. This article covers academic traditions that help us frame the ethics of artificial intelligence (AI). It also develops integrated models for corporate decision-making in light of personalism and virtue traditions. It contributes to the literature by adding a layer to our understanding of the challenges of preserving human dignity as firms grapple with AI. It also provides a framework for executive wisdom when it comes to corporate strategy. The premise of this article is that person-centered traditions offer a sound foundation for AI ethics. We develop models for business executive decision making that incorporates the strength of prior concepts and frameworks. Our goals are to build from the *what* of generally agreed ethical principles, to the *how* of emerging tools. The new family of transformational general-purpose technologies has the potential to disrupt firm management practices. Executives need an approach to discerning what is right. We offer some recommendations on what will help

fill the gap between the drive to build algorithms that mark economic progress while preserving human dignity and serving society at large. To that end, we concentrate on strategy and trust.

We answer the following questions: What is AI and how does the technology work? Why does it matter, and what are some misconceptions? What contributions do allied studies – a faith perspective, philosophy, psychology, and management – have for ethical AI? Most importantly, how can we help firm decision makers engage with this technology in an ethical way that centers on human dignity?

Defining AI & How It Works

Before going further, we should define what the field of AI is (and is not). AI is not the seemingly sentient computer Hal in the 1968 Kubrick film 2001: A Space Odyssey. As is the case with philosophy and basic sciences, we should provide operational definitions with concrete precision. This is a quality often missing from many discussions in the field because we need metaphors to describe AI systems.

Defining Core Terms

In brief, the term *artificial intelligence* "is (defined as) the study of ideas that enable computers to" (Winston, 1984, p.1) exhibit human-like intelligence. Artificial refers to man-made, while intelligence refers to thinking power. It is a field of computer science devoted to developing machines that can mimic human intelligence – thinking, learning, problem-solving

and other brain-like activities - the outcome of which is used for developing software and systems (Shirkin, 2020). On the other hand, AI cannot intuit. It is difficult to fully understand human intelligence let alone have a computer mimic it. AI can only identify solutions from previously known associations and make a prediction. It is the processing of symbolic representations. Computers using AI have the ability to recognize patterns and make predictions based on patterns. The term was coined at a small workshop sponsored by The Defense Advanced Research Projects Agency at Darthmouth University in 1956. Not all of the few participants liked the use of the term intelligence. The goal was to develop systems that replicate human intelligence and solve knowledge-intense functions. An *algorithm* is a process AI uses to solve a problem in finite steps, often in a way that engages a repeated operation. Algorithms are the code of a model, and not a formula. In other words, it is "a step-by-step procedure for solving a problem or accomplishing some end" (Meriam-Webster, n.d.)

AI applications are currently *narrow AI*, that is, meant to handle a single or very small number of cases. *Intelligence systems* are machines so technically sophisticated that they can interact with a dynamic environment. Although AI is a branch of computer science, it is also the domain of mathematics, philosophy, psychology and neuroscience, biology, statistics, sociology, economics, and political science. In this paper, we also apply AI to management.

Because we use computers so frequently without thinking, it is a good reminder that computers "can be defined as a *machine* that stores and manipulates information under the control of a changeable" (Zelle, 2017, p. 2) via instructions telling it in a precise sequence what to do. Computers can carry out any process that can be described, but there are limitations, and

some problems are just too complicated or vague to be analyzed. Thus, some problems in AI are intractable. Computers are not biological brains, and they are certainly not minds. Machines act in the sense of carrying out instructions on data, they do not think. AI has relatively little to do with human intelligence. It involves modelling to simulate intelligence via algorithms. Computers do not have self-awareness, self-directed goals, a personal history, feelings, or consciousness. AI often appears brilliant in what it can accomplish, but most of those things are mundane and involve narrowly defined data-intense computation.

Machine Learning (ML) is the term for a rapidly expanding subarea of AI and computer science that provides systems with the capacity to be trained from experiential practice without explicit programming (Shirkin, 2020). The term first appeared in 1959 for a program designed to play checkers and learn to improve its performance. While the term "learning" is used, it is important to be clear that humans and computers "learn" in different ways. In ML the program logic is reversed. Instead of providing rules (algorithms) for what to do with the data, the program describes the desired output, characterized in a dataset, and the machine develops the rules. Smartphones and online shopping recommender systems are ML examples.

How AI Technology Works

How does AI technology work? To use an analogy, it is like a smart rat in a Skinner box. Software instructions defined by a human programmer prescribe clearly in detail how to carry out an operation that searches for an answer using a set of clean data. Basically, it is a simulation of intelligence that follows an algorithmic process (operations sequence) to find the solution to a

well-defined problem in a finite amount of time. The algorithm begins from a present state, determines possible next actions via search, while exploring ways to reach a goal state (answer or no answer). The data used is like a tree in the physical world, with each item included being a *node*. *Branch nodes* are levels, and leaf nodes are the ends with *parent-child* connections. The search also includes *graphs*, a kind of tree extension that can have multiple attributes (e.g., weight of relationship, distance, time required to a route, etc.) between nodes. No machine is capable of elaborating all potentialities in a given situation, so search heuristics (educated guesses for a solution) have been developed to aid in solving problems. These could be described as tribes or families of algorithms, each with its own pros and cons. For a thorough treatment of how AI works, see Russell & Norvig (2022) or Winston (1984) and Harlick and Elliot (1980) for search. For a summary version, see Shirkin (2020).

There is a significant difference between *explainable* and *non-explainable AI*. Explainable AI (XAI) builds models with a clear explanation for choices and outputs (e.g., decision trees, regression, etc.) of a *network*. An input layer might have just one layer between its nodes and output layer. There is only one connection per node from layer 1 to 2, and from 2 to 3. Each node would have a conditional probability to the succeeding layer. Thus, the output would provide a clear explanation for prediction. For example, in logistics regression one would know the predicted change in a model given a change to any one value.

By contrast in *Deep Learning*, there would be several hidden layers of nodes, each connected to every other in the network. The complexity of the calculations could not possibly explain the prediction. In this case, the scientist does not create the network, but it is created

through training. The relationships might even change given a different start location for the search. This would be somewhat like trying to solve a Rubik's cube starting from a different starting point.

Large language models (LLMs), like commonly used ChatGPT, are a form of ML that uses deep learning with vast amounts of text data during training. Like databases, they only produce a result from the data on which they are trained and struggle to "imagine" anything else. There is no evidence they generalize. To generalize means to apply the underlying patterns well to new situations in real-world scenarios. Being non-deterministic, unlike conventional databases, the resulting outputs will change with every query (Azhar, 2013).

The AI Technology Hype Cycle

AI is getting so much attention that an American comedian (Seinfeld) recently commented, "So let me get this right; we're smart enough to invent AI, dumb enough to need it, and so stupid we're not sure we did the right thing." AI is full of hype, as has been the case with the emergence of earlier general-purpose technologies. There is much excitement, but results are inconclusive even if promising (or foreboding to many). AI may lead to a potential misallocation of investment capital in the economy and resources within firms. In most cases, the technology adoption curve is not as rapid as imagined. Rather, the curve tends to be better described as a hype cycle: slow at first, takes off rapidly, diffuses more broadly, and then dissipates as a new technology replaces it. It is not like the first part of an ascending s-curve that takes off

dramatically but never decreases. Most AI applications are boring and narrow (e.g., email spam filters, ecommerce recommenders, games, etc.).

Social progress and prospering are only possible when there is an urge to create new solutions to human problems. Innovation introduces new ideas, creates products and services in the marketplace that benefit people's lives, and forces firms to reorganize. Innovation is the practical application of an invention or new concept that leads to improvements in a good or a way of producing it (Hubbard & O'Brien, 2021). In our modern society, technology is a key enabler of progress, yet the economist Joseph Schumpeter correctly noted that it results in a creative disruption. It dislodges firms and often renders employee redundancies. These disruptions are the focus of AI public policy debates.

Regulations may be helpful when guided by standards such as IEEE's (2019) for intelligent systems and create a common platform by experts. Regulations that limit can be problematic. The history of resisting technological change only impedes prosperity. For instance, Britain's Red Flag Law (Locomotives Act of 1865) impeded United Kingdom industrial development of the automobile. Yandle (Yandle, 1987; Yandle, 1999; Smith & Yandle, 2014) emphasizes the risks to competitive markets from regulatory capture by the largest tech firms at the expense of the smaller and nimbler firms without lobbyists. Rogers (2003) emphasized that technology diffusion is a form of social change whereby a technology is communicated as new ideas are exchanged through new knowledge and persuasion.

AI Misunderstandings

AI involves human factors engineering, and unfortunately, the implementation has violated a number of fundamental engineering principles. This can be minor in the case of purchasing nudges, but it could be disastrous in safety-critical automation systems, such as in aviation. It is a problem that there is no general theoretical basis for determining which functions can be automated and predict automation capabilities (Kaliadros, 2022). Our engineering assumptions are often wrong because we employ cognitive heuristics. There is no agreed way to model the ways that humans and automation work together. The complex multivariate tradeoffs needed are infinite. We should not confuse the idea of "automation" with "autonomous." It leads to an oversimplified view of complex variables. Autonomous is charged with subjective attributes and a distraction from defining functions. Functions describe what is to be automated. To automate is a binary decision, automation describes the detailed level at which a function may be automated. To automate, the function must be repeatable. Automation that considers human factors is:

> "a complex design decision that involves many considerations, assumptions and uncertainties of hardware/software, operational environments, operator training, etc., perhaps most important of which is the function being automated. Even if we limit considerations to just functional performance, versus broader considerations like ethics, jobs, and human factors, the details, and nuances of a function within its operating environment are unique, and the possible algorithms (means of automation) infinite" (Kaliadros, 2022, p. 2).

There are no simple answers to these human factors' issues. However, part of the solution is to emphasize the importance of functional detail for who or what has control of the output. "(F)unctions "can" be automated, but whether automation will be valuable or appropriate is generally not answerable" (Kaliadros, 2022, p. 1).

The Broader Challenge

AI, an emerging tool, is so useful that it can decrease the time it takes to get things done, while simultaneously removing drudgery and increasing productivity. However, futurists and technology monopolists are screaming that AI is dangerous for humanity as we know it. That depends on what we delegate to full automation. For instance, Soviet officer Stanislav Petrov defied orders to activate nuclear launch sequence in 1983 when he applied human logic that the USSR's systems erroneously reported a US attack.

New research (Martinez, 2023; DeStefano, Kellog, Menietti, & Vendraminelli, in press) shows that there are many trust issues associated with the use of AI. Under conditions of complexity and stress, experts are apt to trust a "Black box" AI system over one with which they have expertise. Counterintuitively, people are sometime just too trusting of technology they do not understand. Thus far, developers of this new general-purpose technology (including for nuclear use) have been able to avoid catastrophe. Less considered is the loss of human capacity when surrendering to algorithms (Kissinger, 2022).

An example of the somewhat preposterous lengths to which a firm can go to develop AI applications is reported with a Canadian firm that has developed an AI-enabled pram (Hupke, 2023). While the test data may be trusted, what if the pram was fully deployed on icy sidewalks in Toronto during the morning rush hour? Perhaps the father, without his first fortifying cup of coffee, is texting as he walks well behind. Would the AI be robust enough to compensate for the different situation? What if he is also a father of five and walking the family dog? To add complexity, how would an AI-driven car interact if both baby and father cross the street at a crosswalk, and the car cannot stop in time? In the latter situation, would the car select to run down the dog, baby, or dad?

Materialism Versus Hope

AI and technology advances most notably take place in the imperative of high-risk conditions. Dark forces in social developments of recent centuries might lead us to believe that the future of humanity with AI is no better that of the dystopian world Benson (1907) depicted in *Lord of the World* just before Europe's second 30 Years War (WWs I and II). In philosophy this is reflected in Marx materialism, radical behaviorism in psychology, and reductionism in science.

Regardless of the monumental task ahead, our view is one of hope for the future of AI. One need not be a chirpy, naïve techno-optimist to point out the mounting evidence. Tech innovation has been responsible for lifting humanity to better economic conditions and ambition. Technology continues to build possibilities for civilization. Virtuous firm executives need a deep philosophical rudder to steer the ship in an era of disruption.

The core output of AI is the domain of math and computer engineering, but other intellectual traditions also have a critical input – philosophy, psychology and neuroscience, sociology, economics, management, anthropology, etc. It is our belief that faith, philosophy, positive psychology, and management theory have many good answers for the challenges of businesses working with AI. There is a danger in anthropomorphizing technology tools like AI, or even using the term "intelligence." It may be useful to simply think of it as a particularly useful tool.

II: Perspectives on Human AI

A Faith Perspective on Human Work

The following is a short review of recent faith-based approaches to the ethics of work. Hirsi Ali (2023) notes that the Western "legacy consists of an elaborate set of ideas and institutions designed to safeguard human life, freedom, and dignity — from the nation state and the rule of law to the institutions of science, health, and learning… find their roots in Christianity." We will focus on Catholic Social Teaching (CST) as a primary source. That is not to say that other faith traditions do not have their own answers, e.g., Muʿtazilah rationality in Islam (Reilly, 2011), centrality of character in Africa, and others in Confucianism and Buddhism. For people whose faith is not Catholic or of no faith, the appeal of CST may not be immediately apparent. We concentrate on CST because it is a comprehensive and cohesive system that traces back to influences such as Aristotle with emphasis on universal values and ethical guidance.

Papal encyclicals communicate important teachings with authoritative (though not necessarily infallible) moral guidance to Church members. The encyclical, *Laborem Exercens* (John Paul II, 1981), "On Human Work" extended Church teaching on labor on the 90th anniversary of Pope Leo XIII's encyclical, *Rerem Novarum* (Of Revolutionary Change). Cosden (1998) notes on *Rerem* that human work is transformative, with an interrelated threefold nature of relational, ontological (of being), and instrumental dimensions. In the encyclical, hierarchical priority is given to the relational over other dimensions. Work is accomplished in a social and relational context. It perfects the person and contributes to society. When work is treated as a fungible commodity to be sold, the right order is violated. Technology can add to this danger.

Laborem specifically addresses the anticipated future impact of technology on humanity. Properly understood, humans should be understood as the subject of work rather than as mere instruments in the economy. The early Industrial Revolution gave rise to the dangerous idea that labor is just one input of production. This is also a concerning issue in the modern era. Emerging information technologies (IT), such as AI, can potentially degrade the value of humans' work if is treated as a mere instrument in the economy. Technology is a tool, "an ally," to serve man. It "facilitates, perfects, accelerates, augments work; increases the quantity and often quality of its output (John Paul II, ibid, Section 5). The dignity of humans requires that the value of persons be given the right priority and that virtuous consideration guide business action. However, technology can "cease to be man's ally and can become almost his enemy" (John Paul II, ibid, Section 5), potentially reducing him to a slave when he should be its master (Kim, Fabrizio, Katherina, Sison, & Benito, 2021).

Philosophy: Personalism & Virtue

What is the relevance of philosophy to AI? Since AI is interdisciplinary, a philosophical perspective is the meeting point for the broader context of other disciplines and the impact of AI on the human condition. The point of philosophy is to make oneself better, and by extension, the lives of others as well. Having elaborated on key Church teaching as it relates to information technology (thus AI), we now turn to two helpful traditions – personalism and virtue ethics - whose genesis traces back to Western antiquity. The Western mind is inescapably shaped by concepts originating in Greek philosophy of reason: that truth is knowable and the free will of man. For Aristotle, experience tells us what is accurate, and the virtue of judgment guides reason. Christian revelation introduced a historical understanding of the universe. There is "cause and effect." For Janas (2021), CST gives expression to the personalist tradition and virtue ethics – "an integrated anthropological vision of the person and work (p. 9). It is incumbent to use both our reason and our free will to do what is good (just), which gives us the capacity to determine what is just and unjust. We do not address deontology (duty-based ethics) or utilitarianism (consequentialism-based ethics) but focus on two approaches closely aligned with Catholic tradition – personalism and virtue ethics.

Personalism

As a philosophy, personalism focuses on the priority of the individual and the person's distinctive experience, emotion, and value system. In this approach, humans are not objects for study or manipulation (as in behaviorism). Rather, they have inherent dignity. This philosophy is often associated with existentialism and humanism in business management literature. It developed in the 19th century in response to depersonalizing circumstances and theories – Enlightenment rationalism, collectivism, determinism, and materialism. In essence, personalism affirms the centrality of the person (personhood) and social dimension of persons as the beginning point of philosophical examination. Notable proponents (Max Scheller Martin Buber, Jacques Maritain, Dietrich von Hildebrand, Gabriel Marcel, Edith Stein, Emmanuel Mounier, Karol Józef Wojtyła) all share the view that persons have a decisive rational, spiritual dimension that distinguishes them.

Human dignity is emphasized in opposition to forces such as radical individualism and denial of a spiritual nature. Wojtyła (Holub, 2021) rejects the extremes of both radical individualism and collectivism. Personalists claim that "the departure point for understanding the world and in referring to all moral truth is the absolute value of the person" (Stanford Encyclopedia of Philosophy, 2022). Persons are distinguished from other non-humans in that they are a *someone* rather than a *something*. From an economic perspective, they cannot be exchanged as each one has his or her unique value and human dignity. Persons develop themselves in relation to others and freely create through their cognitions, emotions, and experience. They are subjects determined by autonomy and consciousness. They are not objects.

Alford (2010) extends the premises of personalism in the context of work and criticisms of management theory. Citing Ghoshal (2005), she notes that assumptions about humans matter, and bad management theories destroy good practice. Humans are both individuals and persons. They seek to sustain life for themselves (as individuals), but they do so in the context of relating to others (as persons in relatedness). Humans seek out ways to achieve some common good, and "business is a common project among stakeholders, through which they grow together, as well as through which they achieve individual objectives" (Alford, 2010, p. 700).

Virtue Ethics

Virtue ethics derives from Scholasticism (principally Aristotle and Aquinas), and like personalism, is rooted in Christian thought that affirms human dignity and the uniqueness of each person. Three concepts derive from ancient roots – excellence or virtue (*arête*), practical wisdom (*phronesis*), and happiness (*eudaimonia). Eudaimonia* (also translated as flourishing, is central to character development. Virtue is a disposition of habit of developing personality – "expect, value, feel, choose, desire, react" (Stanford Encyclopedia of Philosophy) - oriented to do what is good (e.g., to be courageous).

Novak (1996) suggests that in Aristotle's view, one should try to imagine the sort of person one hopes to be by the end of a virtuous life. Then, one must embark on self-discovery "-learn who we are, who we are made to be – and in another sense, we must *make* ourselves" (p. 107). This effort is less of what is the right thing to do in every circumstance than on finding the right balance (golden mean) between virtues in any particular circumstance. For example, for a

soldier to be courageous would be to find the balance between being rash or timid in the face of attack. A virtuous person seeks harmony with rational examination of alternative actions and consequences, but not without an internal struggle. Practical wisdom gained from training and life experience makes wise decisions possible.

Unlike utilitarianism that emphasizes good for the greatest number or deontology that emphasizes duties, virtue ethics is oriented to self-development and character. According to van Staveren (2009), it "was developed for individuals, not for organizations or society (p. 572). We believe it offers a fruitful vehicle for firm managers' prudential judgement on AI. Phronesis (i.e., intelligence applied to practical judgement) makes it feasible to unite an executive's intent to do what is right with the practical circumstances in which she finds herself. Aristotle employed the image of an archer, aiming the arrow to hit a target and understanding the circumstances to make hitting the objective most likely. Natural law establishes a practical guide for guiding actions. Aristotle links reason to virtue, whereas for Aquinas (Stanford Encyclopedia of Philosophy, 2022) the mind works jointly with the will. The will is a faculty of the soul where voluntary action originates.

It may seem paradoxical that Adam Smith, who inspired the concept of the benefits of division of labor also emphasized virtuous traits for market success. In *Theory of Moral Sentiments*, he begins with the virtue of sympathy for others: "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it" (Smith, 1790, p. 11). A communitarian form of virtue ethics by MacIntyre (1987) stresses cooperative

human activity, which is especially relevant in the work context. A somewhat different but related vein in feminist economics emphasizes the virtue of caring as a basic human motivation that supports labor productivity.

Domènec Melé figures prominently in academic leadership on virtue ethics of work. He (Melé & Fontrodona, 2017) draws a distinction between capitalism as now widely practiced in the Anglosphere and the Thomistic approach. Melé states, "Aquinas' view entails, therefore, a natural right to property, but regarding its use, he stresses the responsibility of property owners to pay attention to other people's economic needs" (Melé, 2016a, p. 296.). Rigid standards cannot be applied without consideration of the specific situation. Moral judgements from a Scholastic standpoint require deliberation of principles, mainly the virtue of justice. He states, "Scholastics' moral judgments sought justice, … the result of a careful deliberation of what justice entails in a business action in its specific context" (Melé, 2016a, p. 299).

Cognitive Risks Associated with Surrendering Human "Thought"

New technology offers extraordinary benefits to productivity. But those marvelous advances must be balanced against concerns about "corrosive habits of mind," i.e., loss of deeper understanding and skills (Kissinger, 2022, p. 405) that historically develop at major technological inflection points. We are now at such a point. Along those lines, Kissinger (et al., 2022) states that:

> “Erudition, learnedness, serious and independent thinking” [can be lost] ... in an age dominated by image. ... Just as earlier the transition from oral to written culture at once yielded benefits of literacy and diminished the art of spoken poetry and storytelling, the contemporary shift to visual culture brings both losses and gains." (p. 405)

In short, as we begin to rely more on technology, we lose some skills as a society. It is human nature to resort to the easy path to a goal rather than think comprehensively. For example, professors even at top academic institutions now implore students to go physically to the library building to “wander the stacks” and work with a research librarian for human assistance. It is easier, but often less productive or enriching, to go online.

Relevance of Psychology & Management Theory

Psychology and management theory are overlapping fields. Both are relevant to AI ethics in the sense that each seeks to understand the person in the context of work. Psychology is the empirical study of human cognition and behavior, while management theory seeks to best direct human activities in the context of an organization. As each developed in the late 1800s and early 1900s, the attempt to bring rigor of basic physical sciences has had the benefit of improving their methods and prediction. At times, however, that focus on methods and neglect of the mind has led to some mechanistic assumptions about workers that misses the deeper understanding of persons.

Psychology

We have already addressed the myth that computers can think and that AI systems work like the brain. It seems to make sense that since computer systems' use of a set of rules would be like human decision making. However, that is not how humans make decisions as many human decisions happen subconsciously and with intuition (Brown, 2023). Nevertheless, psychology is an essential contributor to the development of AI. Insights from the field are essential for interactive design, ethics, and cognitive computing.

We can thank the behaviorists (e.g., Pavlov, Thorndike, Watson, and Skinner) for bringing increased rigor to methods of scientific research and an understanding of the external environment influences on actions. Their work helped advance psychotherapy techniques that alleviate human suffering. Nevertheless, their limit of study to only what is directly observable has influenced a more materialistic conception of humans, one that has slipped into the assumptions about humans in some business practices and engineering.

There are many more theorists who serve as a counterbalance to a materialist view of humans. In the context of work with relevance to technology, positive psychology and advances in neuropsychology offer a more nuanced perspective on humans as ends in themselves. The psychiatrist, Victor Frankl, stated that the main goal in life is to find meaning (Frankl, 1985). In work, people not only meet their material needs, but they also experience fulfillment (Janas, 2021). Positive psychology, like personalism, focusses on human flourishing (Seligman, and

Csikszentmihalyi, 2000). Flow is the mental state of being fully absorbed that describes hyperfocus and loss of awareness when one is absorbed in a task, somewhat like the moments in which we enjoy uninterrupted work. Like personalism, self-determination theory (Ryan & Deci, 2000) focuses on motivation, values, and aspirations. Like virtue ethics theory, it also focuses on the relational nature of humans. AI can enhance a meaningful work life by removing many of the mundane and repetitive tasks. AI could increasingly make time for the more creative and fulfilling tasks of work and make joyful flow a more frequent experience.

Management Theory

Management theory within the broader context of business practices often takes its cues from psychology research. At times phenomena explored in clinical or industrial psychology research is applied more broadly than the intent or limitations of the original research. This can lead to the error of generalizing from the specific of a particular experimental situation to a broader circumstance and to problems of threats to external validity (Cook & Campbell, 1979). Quick fix, templated management lore and fashions "can sound quite compelling .. (with) beliefs, anecdotes that are so pervasive that they achieve the status of immutable fact" (Banasiewicz, 2019, p. 33).

Alford (2010) notes that "perverse effects arise from the distorted assumptions" (p. 697) of man and their misapplication. A failed effort to mimic the natural science model tends to exclude human values and intentionality. Taylor's (1911) scientific management presents us an example of applying economic efficiency in business operations that treats humans as just

another element in the machine of productive automation. More recently, the era of reengineering seemed to deny the essence of personhood in the work context.

Notwithstanding that management fads come and go, with many fading into obscurity, some have merit. A few offer a way for academics, consultants, and firm managers to talk with each other about underlying objectives and a way of operating. An enduring management theory tested in universal norms is needed. One such theory is humanistic management which aligns with personalism and virtue ethics. It is particularly relevant to AI ethics in its focus on participation and teamwork, dignity, meaningful work, and well-being in the workplace. It is relevant to AI across numerous dimensions (e.g., tech innovation, leadership decision-making, strategy, change management, allocation of resources, and program management).

Humanistic management was a reaction to a negativistic and mechanistic view of man among behaviorists and in scientific management. Humanistic management draws much from existentialism that concentrates on the individual and the idea of becoming. In Herzberg's (1959) *two-factor theory*, hygiene factors such as pay are extrinsic to work, whereas motivators are factors that provide status, a sense of accomplishment, and actualization. Motivators are intrinsic to work. Douglas McGregor (1960) stands out for his theory x and theory y of management. Theory x represents tight control of workers (the stick), whereas theory y managers provide a positive environment to motivate (the carrot) (van Biljon, 2022). Likert (1961) established the idea of participative management and the nature of highly effective groups.

A Human-Centered Technology

Having considered the philosophy and related fields as applied to work, we reflect now on what it means for a human-centered technology in the case of AI. Rosebrock (1981) points out that there is a duality to consider when developing technical systems: human-centered versus technocratic design. In the early Industrial Revolution technocratic design became the dominant standard. The technocratic design approach was to simplify a manufacturing task to its basic elements, for which humans would need little intelligence, skill, or training. From an input cost perspective this might appear to make sense where there is abundant cheap labor and total cost of production is the only goal. However, this is a departure from the intent of Adam Smith's theory of moral sentiments (Smith, 1759). It fails to act on one's natural sense of sympathy for others and neglects virtue amidst social relations.

We should distinguish between efficient processes and effective outcomes. Engineers typically strive to build the most efficient production system. Usually, design is focused on the machine (such as a robot), which sometimes ignores the full range of potentially contributing inputs, including the human inputs. Substitute a human for a robot with multiple capabilities and we would then be thinking more clearly about the fuller range of contributing that could lead to a more effective outcome. It was the paradigm of the newly developing engineering profession early in the last century to adopt the assumptions of technocratic design – build the machine and have humans adapt. "To spend effort and money on fragmenting jobs and reducing their content seems neither rational nor efficient" (Rosenbrock, 1981, p. 6) at this time. Rather, engineers should develop technology be developed to take advantage of the full range of human capacity. Thus, it makes the design work more interesting and a worker's role more stimulative.

In the case of information technology, when we code in a technocratic manner, we can lose valuable human inputs. Programs become a driving force for faster production rather than supporting creative alternatives. Unique human skills can become fossilized. For example, the unique human skill of navigation by using a sextant and the stars has become fossilized in the face of global positioning systems (GPS). The US Naval Academy has had to reintroduce the sextant's use. This is similar to the way people now depend on their phones for directions (until power is lost and they must either turn to a map or ask passersby).

This aligns with the idea from *Laborem Excercens* (John Paul II, 1981, 1982) that man is not just a means in the economy, but its end purpose. Technology should not supplant humans. We can extend this to AI as there are clearly roles in production that machines will never be capable of fulfilling. Cooley (1982) argued against compartmentalized engineering because of his experience developing computer-aided design. Why, he argues like Rosenbrock, should we waste so much human ability and effort. Scientific Management, when extended from manual to intellectual work is made worse when (Rosenhead, 1980)

> "Extracted and formalized… The open-ended problem-solving capability, the tacit knowledge of "ordinary people", is jettisoned in the search for a narrowly defined "efficiency" (p. 616).

Let us extend the argument along another dimension – averages. If we build code to the average worker, especially to intellectual tasks, much of the person's unique skills may be lost.

Humans are multi-dimensional with varieties of unique skills gained from the interaction of their genetics and life experience. When attacking an enemy across a stream, a general would not want to send the force across a depth of the average height of a soldier. Only half would make it across. Instead, he/she would build a bridge or ford at the depth all could cross. The same holds true for developing algorithms that support the unique characteristics (e.g., intellectual strengths, personality, preferences, etc.) of the operator.

AI Ethics

Having considered philosophical and ethical perspectives, what do these perspectives mean for a human-centered technology approach for AI and ML? How do some of the leading authors suggest we put these ideas into practice?

AI4People

Much of what we hear in the popular press is about a foreboding downside of AI. Floridi and colleagues (Floridi, Cowles, Beltametti, Chatila, Chazerand, Dignum, Luetge, Madelin, Pagallo, Rossi, Schafer, Valke, & Vagena, 2018) present a European consortium synthesis of opportunities and societal risks in AI's development and deployment. They cite high-level opportunities that correspond to its risks along with a helpful summary schematic. While there is the potential to devalue human abilities, it can also foster their potential. AI can be a great force for good, not just in new productivity and profit. It is a potential aid to help people realize greater personal actualization.

There is a risk when humans delegate too much responsibility to AI for outcomes. Used properly, algorithms can be tools in the hands of individuals to enhance agency and the quality of their lives. Examples include applications being developed to assist persons with mental health difficulties to employ cognitive behavioral techniques that challenge thought patterns which lead to anxiety or dysphoria.

Even informed people inappropriately surrender responsibility to AI when control algorithms are not in their ultimate interest. We have previously noted (DeStefano, 2022) that even experts under stress are apt to defer to uninterpretable AI judgments. Even when there is an opportunity for people to be "in the loop," they may choose incorrectly to default to trust an algorithm that poorly suits the situation. On the other hand, there is a great opportunity to increase individuals' and society's capabilities. Logistics, disease control, and medical diagnosis are all areas where AI enhances capacity. For example, the US Navy has achieved much progress in aviation predictive maintenance that lowers costs and increases safety.

Floridi (et al., 2018) cites the potential for erosion of self-determination. We would note that algorithms can nudge us in unhealthy and sometimes nefarious ways. Malevolent actors can nudge addictions (e.g., gaming, gambling, pornography) and political polarization. By contrast, put to positive social purposes, AI can be used to assist flourishing and human dignity.

At present, there is reasonable concern that we may become like slaves to AI as was the case of child labor in the early Industrial Revolution. On the other hand, AI is "a powerful force,

(it can be) a new form of smart agency" (Kim, Fabrizio, Katherina, Sison, & Benito, 2021, p. 689). As perhaps a future extension of Floridi's idea of self-realization, we can foresee a better future. Machines may not only work in tandem with their operators, but perhaps they could adjust to the skills and needs of the individual. Every human has their own personality and intelligences that develop in sometimes unpredictable ways over the lifespan. We now have human work teammates that adjust to us and help us develop creative insights together. Having a future digital twin with insight into our competencies could make work more rewarding and productive. A futuristic machine that appreciates individual traits would be an extraordinary step toward self-actualization.

AI & ML Tool Development.

Morley and colleagues (Morley, Floridi, Kinsey, & Elhalal, 2020) conducted an extensive study (a review of publicly available ethics tools, methods, and research) to bridge the gap between AI/ML principles and actual practices. They reviewed over 70 AI guidelines documents (e.g., OECD, EU, Montreal Declaration, IEEE, etc.) and 425 references of tools, methods, and firms. They advocate that it is time to bridge the gap between the "what" of ethics principles to the "how" of technical methodology. They identified useful tools, but there are holes in some important areas. There is a fragile consensus on guiding principles across guidelines, and they narrow these down to five.

The first principle, beneficence, involves the idea that AI must respect human autonomy and be beneficial to humanity. The second, non-maleficence, is that development must not

infringe on security and prevent harm. The third, autonomy, represents the idea that AI should be fair and enhance human-centered values. The fourth, justice, means that it should promote prosperity and inclusivity. The fifth, explicability, represents the idea that AI systems must have transparency, be explainable, and be clear on accountability.

In Morley's (et al., 2020) typology these principles cross-reference with the seven standard stages of AI/ML development: from business use case to -> design to -> train and test data to -> deploy, and finally to -> monitor. Each phase has the need for specific ethics tools. We will return to elements of Morley's AI4People framework incorporating it into our own models for firm AI ethics in the next section.

III: Models for Executive AI Business Ethics

Personalism to Ethics for AI (P2e4AI)

What is the purpose of work? The answer to this question depends on the perspective one takes. Using a narrow perspective, one can view the individual as simply an economic entity where the purpose is exchanging one's labor for necessities and the occasional extravagance. Viewing the purpose of work through a broader lens, national economy for example, work means growing gross domestic product increasing at a sustainable rate with its components in balance. The CST perspective focuses on the dignity and worth of every human being. AI can enhance work from all of these perspectives. But enhancing the purpose of work for all these perspectives must first begin with a consideration of the person's autonomy and dignity.

We have developed a Personalism to Ethics for AI (P2e4AI) Model (see figure 1. below) that integrates personalism and virtue ethics from concepts previously discussed in section II into a more business-focused ethics. The statistician George Box said, “Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary, following William of Occam, he should seek an economical description of natural phenomena” (Box, 1976, p. 792). We agree that all models are wrong in some respects, and we also agree with him that some are useful. In the case of AI ethics, parsimony dictated we should provide a useful model for the busy executives.

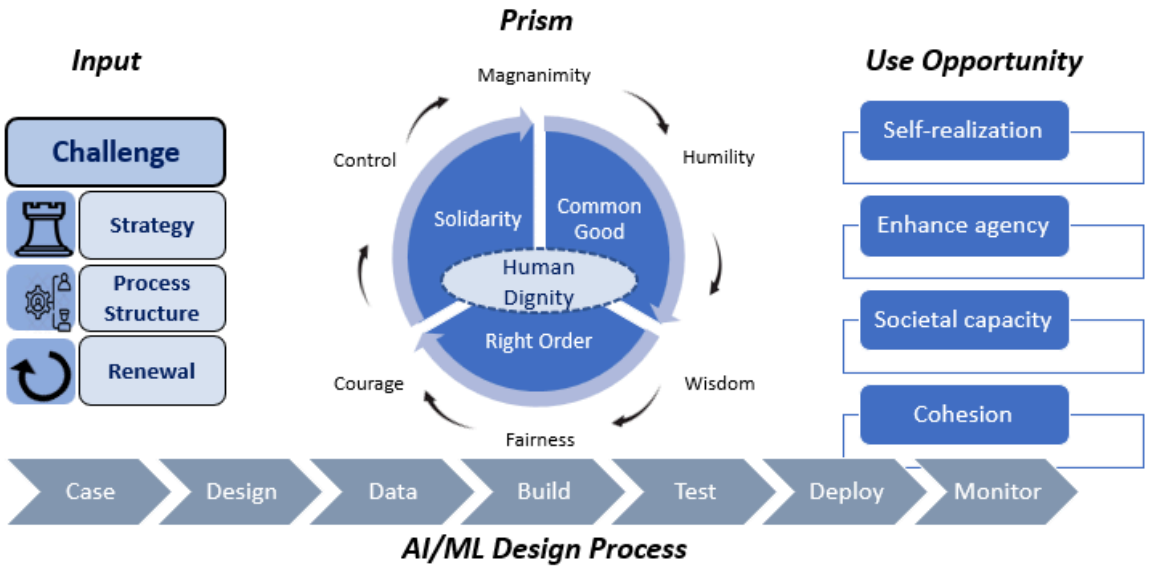

Figure 1. P2e4AI Model

The P2e4AI model draws in part from the Human-Dignity Centered (HDC) framework (Mea & Sims, 2019) cited in humanistic literature (e.g., Guitián & Sison, 2023; Guitián, 2023; Akrivou, & Scalzo, 2020). The HDC framework is also cited in articles on topics such as moral

psychology, entrepreneurship, labor conditions, consumer behavior, healthcare, computer cognition, and sustainability. This source framework, based on natural law and CST, provides a corporate outline by which managers apply ethics to improve humanistic practices and corporate performance. The P2e4AI also blends in concepts from Floridi's (Floridi et al., 2018) AI4 People framework and Morley's (Morley et al., 2020) AI/ML development typology. It captures in one place a graphical representation of business AI issues (inputs), human dignity considerations (prism), corporate outputs (use opportunity), and a corresponding pathway (design process) for AI development.

P2e4AI Model does not provide a formula but is more aimed at a reflective process. This dynamic should result in a virtuous cycle for a firm in which everyday actions build a self-reinforcing culture of human dignity. The issue of trust has relevance for AI. In its entirety, the model should foster corporate excellence. To summarize the model in a few words, "Be educated about AI and do it morally."

Business Challenge Inputs

The left section of the figure represents *Inputs*, that is, some of the most urgent challenges for executives considering whether and how to implement AI in their firms. The first challenge is *Strategy* because it sets direction for the future. Managerial effort needs to be applied to ongoing operations and profit while simultaneously planning for a future of AI disruption. Executives need to decide if the firm should be on the leading edge of risk in an environment that may be unsettling to employees or whether a safer approach makes sense.

The second challenge involves *Process & Structure*. With the exception of leading-edge and well-financed technology firms, AI will drive major changes in process and structure. A firm's processes and structure develop from its early beginnings as the founders resolve the challenge of the core problem the organization addresses in the market. As actions coalesce around habits, these develop into agreed-upon processes and a structure for organizing culturally accepted ways of doing things (Schein, 2010). In the face of AI adoption, executives will face the necessity of changing their workflow and thus how they organize to stay profitable. Other firms will race to do the same. This will be no small feat, since organizing for innovation during times of evolving general-purpose technology (such as AI) is a trial. New technologies (e.g., steam, electricity, transistors, computing, the internet) have always been disruptive for entire industries, the people who work in them, and the communities in which they live.

The third challenge, *Renewal* in the lefthand portion of figure 1, represents the way in which executives must align the interests of the workforce with the need to operate in a future that includes AI. In the face of uncertainty, employees will need confidence and new skills. Employees are a firm's most valuable "asset." Their current deep knowledge of how the firm succeeds will need to be augmented in a way that is least disruptive.

Ethical Prism

The center of the diagram in figure 1 represents an ethical *Prism* through which to view ethical leadership. Philosophy and virtue ethics can give guidance on how to sort through AI challenges. Managers can reference ethical principles and virtues to sort through issues that

support human dignity. This section portrays an order aligned with personalism that places people, not things, at the center of what matters. Human proximate needs and long-term prosperity find a home in organizations that attend to dignity while pleasing investors with good returns. Employees also develop their own skills and character.

The middle ring in figure 1 features principles drawn from CST that surround *Human Dignity* at its focus. Human dignity is a principle in both personalism and virtue ethics. Human dignity can mean many things, but the essence is that humans are ends in themselves, with rights that dwell in them. No AI decision should violate humans' rights. In the context of sorting through issues, AI decisions should reflect the need to treat people as ends in themselves and not as means to an end. *Common good* means to consider workers' needs and client welfare. While there might be some easy ways to implement AI actions, executives need to reflect on the long-term outlook: how can they retain talent and serve future customer needs. Put another way, short-term financial gains at the expense of persons would be objectionable. *Right Order* is a translation of the principle, *subsidiarity*. Subsidiarity is the idea that the power to make decisions should reside at the lowest level. Rather than centralizing decisions, leaders need to give employees maximum freedom to judiciously make decisions for issues they best understand as they have the most competent understanding. Finally, *Solidarity* means to treat others as oneself, i.e., abide by the Golden Rule. Leaders have a central role in making AI deployment successful. The technology will fail without planning to include effective communications, training, and staff involvement in requirements.

The outer ring surrounding the central section of figure 1, *Ethical Prism*, represents *key virtues* relevant for managers when applied to AI decisions. Leaders' virtues need to be developed to direct reason and the will. The P2e4AI model adopts six virtues from the HDC framework. Three virtues stand out as critical for corporate leadership for the future of AI – magnanimity, humility, and wisdom. *Magnanimity* is the virtue of being great of mind and heart, displaying a noble generosity of spirit. It is to fully appreciate one's gifts and to have a commitment to do good with those gifts. Executives need magnanimity to remind them that their vocation calls them to a higher obligation to stakeholders. To whom much is given, much is expected. *Humility* means to have a cleared-eyed sense of oneself – both one's weaknesses and strengths. Temptation toward pride is the opposite of humility.

Wisdom, classically labeled prudence, means to have the insight to choose the right action and most practical action given particular circumstances. A glaring example of failure of wisdom and to appreciate weaknesses in understanding was made in oncology research. IBM Watson teamed with a renowned cancer center in 2018, and because the AI team used hypothetical patients' information rather than real data, it sometimes gave glaringly unsafe treatment recommendations (Dilmegani, 2023). Failing to prescribe clear management business objectives, poor data quality, and failing to coordinate among internal firm teams can lead to technical AI disappointments that tarnish corporate reputation.

Although leaders may have prior experience in implementing technology, new learning may be required. For AI to be successful, humility and wisdom may orient executives to update their technical knowledge. Even experienced tech firms can fail. Amazon aimed to develop a fair

recruitment tool, but its use of internal data was problematic (Drage & Mackereth, 2022). Amazon's effort to create a merit-based system were laudable, but the way it went about its study would have led to discrimination against women when the benchmarks used were from Amazon men.

Fairness, or justice, involves justly giving to others what is justly due to them, justice. In the context of AI progress, this requires the need to move forward as quickly as prudent with customer-safe AI. At the same time, managers need to engage employee talent and train as needed. *Courage* in the context of AI decision making is the balance of rashness and avoidance. Aquinas states that courage is needed to endure challenges with sufficient confidence to overcome obstacles (Floyd, n.d.). The futurist, Marr (2020), states that leaders need the qualities of vision, courage, and authenticity to deal with the potential ethical pitfalls and people-related challenges that will be part of an organization's managing AI.

Control can be understood as temperate self-restraint, modifying one's own personal desires in order to be in harmony with others for social goods. A balance of these virtues is needed to handle current operations and future successes. One would expect better of Microsoft, but it had to shutter a Twitter chatbot heralded as able to speak as young people do. The AI tool had to be quickly withdrawn when a vulnerability led to its chatting racist and holocaust-denying material (Lee, 2016).

Ethical Design Process

The bottom section of the figure 1 diagram, *AI/ML Design Process*, represents a simplified version of concepts from Morley and colleagues' (Morley et al., 2020, table 6 on page 2154). The steps address common ethics problems of AI design (e.g., inconclusive evidence, inscrutability, traceability, etc.), from inception to final deployment and monitoring. The phases are: 1) *Use Case* – business problem elaborated (description of how the system will interact with users to accomplish a goal); 2) *Design* – case turned into requirements; 3) *Train* Data -input information for AI model learning; 4) *Build* – algorithms built; 5) *test* – test the system; 6) *deploy* – go live; and 7) *monitor* – assess ongoing performance. Since AI is more like a dim-witted though strong and hard-working employee than an evil genius, human oversight is needed. Each stage of the AI/ML Design Process needs to ensure that outputs are accurate and serve the right human purpose. Building ethical AI needs greatest attention at the outset. Morley (et al., 2020) lists some of the tools that are available for each step that build in features such as beneficence (i.e., beneficial to people) and explicability (i.e., explainable, and accountable).

Use Opportunity

The right section of the diagram in figure 1, *Use Opportunity*, depicts desirable high-level opportunities for a good AI society as developed by Floridi and colleagues (Floridi et al., 2018). Each of the four boxes represents overcoming a risk. These would translate well to desirable outcomes in the case of decision-making by firm executives. *Enabling human self-realization* overcomes devaluing workers' skills. *Emphasizing human agency* prevents misuse of removing human responsibility. *Increasing societal capabilities* obviates reduced human control. *Cultivating cohesion* prevents erosion of self-determination.

In the next section, we develop a model that organizes executives' attention to the shifts required to build a successful and ethical business that focuses on building strategic AI capabilities for the future while managing to execute tactics in the present.

An AI Cognitive Enhancement Strategy Model

Figure 2 below depicts an *AI Cognitive Enhancement Strategy Model*. The goal of this strategy model is to help executives assess the value of different options for future business success. Ethics and strategy may be thought of as two sides of the same coin that asks, "How should I act to prepare for the future success of my firm?" Executives have an ethical responsibility to guide their firms through market turbulence that may be the norm as AI diffuses. A recent study by Cisco (2023, Nov 14) of over 8,000 executives showed that nearly all felt a sense of urgency with AI. However, only 14 percent felt ready to face the challenge and only nine percent have the kind of corporate culture needed to support a transition.

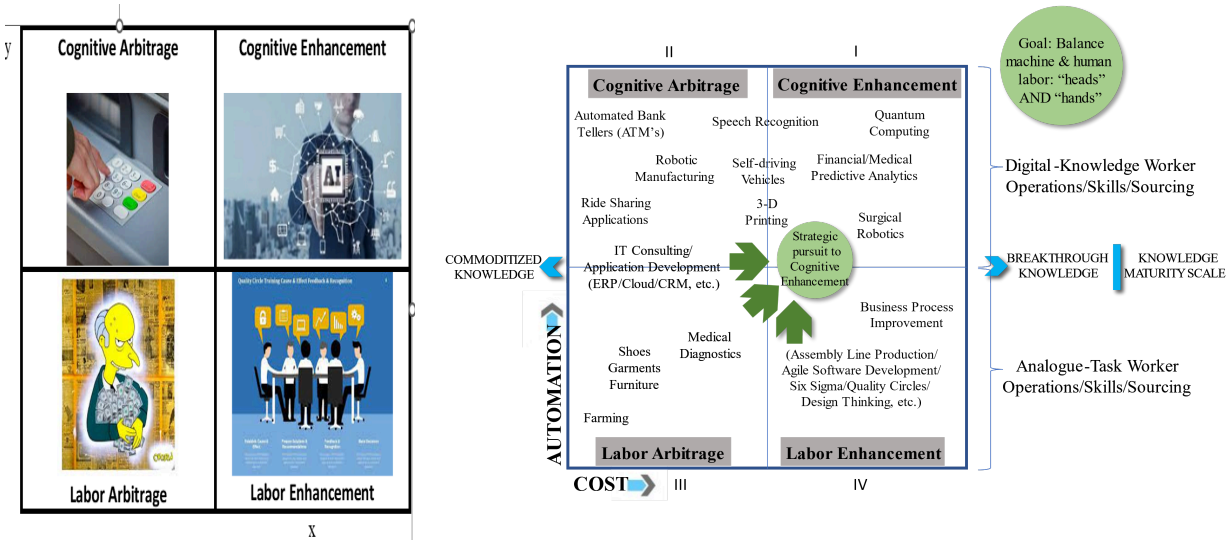

Figure 2. AI Cognitive Enhancement Strategy Model

[replace Simpsons' Mr. Burns]

One fundamental question for executives to ask is, what can AI do for people when we see it as a tool and not just a technology concept (Borrow, 2023). Another critical question is how to manage the balance of currently profitable competencies while simultaneously pivoting to future AI opportunities. The issue is one of strategic orientation. For comprehensive reviews of business strategies see Aurik, Jonk, & Fabel, (2014), Ghemawat (2016, 2002), and Grant (2019). In the context of strategy that employs humanistic principles, decision-making:

> "starts with the perception of a need to define the nature of the issue or problem which should be addressed or solved and the goal to be achieved with the decision. The human dimension is present in this first stage as the consideration of how the needs, issue or problem affects people" (Melé, 2016b, p. 51).

AI Cognitive Enhancement Strategy Model in figure 2 above accounts for the idea that executives need to take advantage of full human capacities. Leaders need to commit to a future where AI tools are used to the best advantage to serve customers, employees, and investors.

The concepts for the AI Cognitive Enhancement Strategy Model come in part from what is commonly called the GE-BCG 9-box matrix and the McKinsey's *3 Growth Horizons* (Baghai, Coley & White, 1999). Although the 3 Growth Horizons is an older model, the US National Security Council still considers it a quality approach for technology innovation. According to the 3 Growth Horizons approach, it makes sense for firms to balance strategic interests across three areas of innovation. The first horizon in McKinsey's 3 Growth Horizons is *current core business operations* to ensure present profit to continue as an ongoing concern. The second horizon is *emerging opportunities* that promise future growth. The third horizon is composed of *promising ideas* not fully formed but which could potentially lead to new disruptive growth capabilities.

The AI Cognitive Enhancement Strategy Model takes a further step. It translates the 3-horizon framework into a two-by-two matrix to reflect on the value of humans to a firm's success. It classifies approaches that firms could take toward their most important asset, people. The x axis represents potential cost and uncertainty, while the y axis represents potential return. The left matrix illustrates a simplified version, and the right one represents an elaboration.

Each of the quadrants in the AI Cognitive Enhancement Strategy Model represents a different option and orientation for future AI direction.

- Quadrant III, *Labor Arbitrage*, represents perhaps the least ethically desirable direction. The example of offshoring to low-cost labor nations would be an example of simply

"giving up" when confronted with new technology threats (Guitián & Sison, 2023). This was a common solution in the era of reengineering, and many firms are finding that fragile supply chains make this a less trustworthy solution.

- Quadrant II *Cognitive Arbitrage* involves using AI to remove repetitive tasks from humans to make way for more and richer jobs. For example, ATMs historically replaced many teller roles but vastly expanded opportunities for specialized banking services.
- Quadrant IV *Labor Enhancement* is representative of organizational development techniques that stem from action research and the human relations movement in industry. Thus, we find that consulting services, such as change management, have greatly expanded.
- Quadrant I *Cognitive Enhancement* represents greater high-level machine and human collaboration. Investments in this area (like *Horizon 3*), have higher and uncertainty. However, projects in this area have greater potential profitability and the promise to expand human capabilities. Examples are human-guided robotic surgery that decreases medical complications and radiological diagnostics. Where possible, it makes sense for corporate strategists to look for opportunities in this quadrant that will create profits and serve society better while increasing humans' intellectual engagement.

IV: Reflection on Personalism, Virtue, & AI Executive Ethics

At the outset, we proposed that executives need to do two things to get AI right – become knowledgeable about the technology and implement it in a way that is human-centric. In this section we re-examine personalism and other themes in light of the P2e4AI and AI Cognitive

Enhancement Strategy models. We specifically address personalism and trust. In the appendix we also offer questions firm managers might consider in light of the models which can help them aim for ethical AI.

A Personalist Human-Centric AI

Is an AI Future Bleak or Hopeful?

Use of the term intelligence for computer systems has created much confusion. Its use clouds the purpose for what AI should serve – humanity and people. It encourages the false notion that computer brains will replace the role of people at work, trigger mass redundancies, and someday rule the world. Some pessimists (e.g., Frey & Osborne, 2017) forecast that with AI up to 47 percent of US jobs could be at risk, with the majority being in lower-skill roles. Kissinger and colleagues (Kissinger, Schmidt & Huttenlocher, 2023) remark, "Enlightenment science accumulated certainties; the new AI generates cumulative ambiguities."

Fire is perhaps an apt comparison to a general-purpose technology such as AI. It has been serving humans for 17 million years, and yet we have continued interventions to make it safe. In the recent past we were confronted with risks that came about with the discovery of nuclear fission. More recently, we have not used the deadly force of nuclear fission since World War II. But for the unfortunate souls in Hiroshima and Nagasaki, we have somehow prevented the use of nuclear weapons. Like fission, the consequences of AI misuse by malevolent actors would be

disastrous. This argues that controls should not be in the hands of a few decision makers alone as was the case in August 1945.

We would agree that AI will replace some tasks that create a skills mismatch, but this is the norm in the history of new general-purpose technologies such as steam power and electrification. Productivity improvements create unseen new jobs, individual wealth, and capital for new investment. Resistance and regulatory capture only impede growth and opportunity. We take the more hopeful position that AI will introduce new human labor- and cognitive-intensive tasks. Acemoglu and Restrepo, (2019) note,

> "Recent technological change has been biased towards automation, with insufficient focus on creating new tasks where labor can be productively employed ... but this might mean missing out on the promise of the "right" kind of AI with better economic and social outcomes" (p. 2).

Humans Are Not Replaceable - AI Serves

Human beings are not speedbumps on their way to some greater material utopia, and there is a danger in anthropomorphizing technology tools like AI. Mechanistic assumptions about how to manage work can be helpful, but not scientific when applied to humans. When describing usefulness of information for organizational decision-making, Haekel (1997) notes that, "science and math have nothing to say about intelligence, knowledge, and wisdom" (p. 1).

The sort of knowledge needed to direct a firm needs to be transformed from the more objective data (such as AI outputs) to knowledge and wisdom.

Many fear that machines will replace humans. It is true that algorithms can prevail over humans in a few limited areas. Their processing speed is faster and memory greater for use of objective evidence than humans. Human processing is qualitatively different and irreplaceable. Human computation is largely non-explicit and multidimensional. Machine learning systems by contrast are confined to processing data that has happened in the past. Human cognitive processing is shaped by factors such as emotion and culture in a brain with neural connections constantly rewiring themselves (Banasiewicz, 2019). Only humans can focus on the bigger managerial questions that focus on "the why" something has happened and what they will do about it in the future. Only the wisdom that comes with a repository of life experience can make responsible decisions in these firm circumstances.

No matter how fast a machine processes information, it will never replace a sensory apparatus of 126 trillion unique, integrated synapses per person. In short, AI provides speed analysis of objective data, but management needs to direct both the input and the usefulness of the output to serve the human condition. People are the repositories of wisdom. Acemoglu and Restrepo note (2018)

> "There are still many tasks that we cannot automate, including complex reasoning, judgement, analogy-based learning, abstract problem-solving, and a mixture of physical activity, empathy, and communication" (p. 12).

Basically, humans are not fungible. They have unique capabilities that machines cannot replace. The created world has a purpose, and every person has a role within it. Humans alone create AI and serving humanity is its ultimate purpose.

A Personalist & Humanistic Perspective on AI Ethics

AI will have an immense impact on employment, the economy, and as yet unforeseen risks and opportunities for humanity. Bertaloso and Rocchi (2018) assert that a reductionist idea of human work limits it to a task-performing activity. It creates an unnecessary tension when envisioning a future where AI could handle more tasks. Taylorism, which focuses on the task and engineering methods, ignores the social nature of firms. Bertaloso and Rocchi (2018) state, "Real human work is never an isolated episode; it is always part of a larger individual and social narrative" (p. 8). Overcoming the mechanistic narrative when implementing AI would require a managerial theory and business ethics of work that recover the collaborative nature of individual persons in a larger whole. Personalism, virtue ethics, and humanistic management have overlapping themes that we have blended into the P2e4AI and AI Cognitive Enhancement Strategy models.

Hasnas (1998) notes that the field of business ethics can be charged with being impractical. Ethicists employ abstract terms that fail to communicate with business managers in a language they understand. Normative theories (describing oughts of human behavior) should derive intermediate principles from the abstract, and that has been one of our objectives, "to

provide human beings with (practical) guidance" (p. 20). Marin and colleagues (Marin, Boanță, Hadăr, Badea, Vlăduț, Bucur, Ciocănel, & Ivan, 2015) note that business models need to reflect that firms are living entities. Their technology reflects the characteristics of a firm's leadership. It has been our intent to link ethical theory to its application using our two models as a tool to translate principles into ethical AI implementation that firm leaders can employ.

Drucker (1999) wrote, "Management in the new Network Society will require different behavior, different attitudes (p. 795). Humanistic psychology and management introduced the idea of the subjective meaning of people and work. Human experience is uniquely internal and subjective, and that work involves a contribution in the context of social connections. Humanistic management is a person-oriented management in which profits serve human ends and not as mere resources to serve a profit goal. Humanistic management seeks to incorporate a comprehensive understanding of all human capacities as the end of firm actions. It recognizes the whole of the human being – intention, autonomy, development and becoming. It emphasizes sociability and flourishing.

Melé (2016b) states that one of the basic functions of management in strategy

> "practices entail establishing (top managers) or following (middle managers) the goals to achieve frequently under conditions of uncertainty, foreseeing what could happen in the future and planning whatever might be require to attain the desired ends" (p. 49).

When discussing the future of work, Fontrodona and Melé (2022) underscore the importance of dignity and flourishing in a future that includes major technological changes, such as AI. "Workers – the subjects of work – not only produce objects but develop themselves, acquiring good habits or virtues, which are central to human flourishing" (p. 183). If management is guided by virtue, it contributes to a more meaningful work experience. They also emphasize the idea of business as a community of persons, noting that solidarity contributes to effective results. Likewise, Hayek (1945) emphasized the unique value that every person brings to the firm. Other than specific knowledge of scientific rules, he says,

> "practically every individual has some advantage over all others in that he possesses unique information of which beneficial use might be made, but of which use can be made only if the decisions depending on it are left to him or are made with his active cooperation" (p. 521-522).

This knowledge specific to each individual speaks both to their unique value and also to the need for management to foster subsidiarity.

Melé (2009) joins virtue ethics to personalism's principles while emphasizing the virtue of practical wisdom; the latter which we have highlighted is essential for firm managers who will be guiding strategy on AI deployments. Aristotle (Nicomachean Ethics, II, 6) said that only prudent persons with practical wisdom discover what is good: "Every action and purpose seems to aim at some good (Pakaluk, 2005, p. 48-9). Aquinas followed Aristotle's lead in integrating virtue ethics with norms and principles. Personalists are diverse but all agree on the unique

nature of persons and their human dignity. Wojtyla's Realistic Personalism underscores social reciprocity in the Golden Rule. In a business context, this implies that there needs to be a disposition to contribute not only to one's own wellbeing but to others as well. People flourish in relationships with each other. The common good is another important principle. The common good implied here is not one of general interest or the good for the most people; rather the common good here means common aspects of flourishing. Firms are a community of common flourishing that are part of a larger social purpose.

Strategy, Uncertainty & Ambidexterity

Strategy

Organizational decision-making and strategy normally take place in a macro-environment that cannot be controlled. The result of executive decisions is shaped not only by their choices but also by unpredictable factors – competitor actions, supply and demand, and innovations. Unlike some other professions, there are no accepted standards for managers on how to act in the present and plan for the future. The dominating strategy these days is a zero-sum game. For example, a simplified Porter strategy model focused on market capture tends to dominate some corporate minds. Though it has intuitive appeal, its "grab what you can" orientation has a narrow assumption about social reality. A more positive and enlightened approach would focus on the reality of social value (Ghousal, Bartlett, & Morgan,1999).

From a human-centric perspective, we need to shift strategy model metaphors. A personalist, human-centric perspective strategy is a "critical bridge between business ethics and business practice" (Hansen & Smith, 206, p. 202). This shifts the unit of analysis from "value margin capture" to human interaction (with customers, employees, and investors).). The modern corporation amplifies human effort across boundaries. Firm growth depends on the quality of management decision-making, and a firm's foundation is a sort of moral contract with employees and society.

Our two models offer executives broader options to steer their AI strategy in an ethical direction - to increase cognitive engagement. Firms' AI policy ought to focus on the value of intrinsic human dignity. Strategy needs to balance current operations and future planning. "Capitalism is not a system solely about things but about the human spirit" (Novak, 1996, p. 120).

Types of Risk

Most mature firms are well-oiled machines designed to run with maximum efficiency and reject risk. Once an organization grows to a certain size and success, the cultural DNA resists change. The disruptive nature of AI requires a new sort of courageous leader, one who can manage in the face of uncertainty. Not all of what people think of as risk is the same. There are many types. The first form of risk, *statistical risk* is quantifiable with probability distributions that can be incorporated into planning (e.g., serial throws of dice). The second form of risk,

uncertainty, on the other hand as in AI strategy, involves unique circumstances (Knight, 1921). Managing uncertainty due to unknown instances needs to be guided by wise and resolute leaders.

To those two types of risk, we can add two others. A third type of risk, *Black swans*, involves unforeseen disasters that are later rationalized (Taleb, 2015) (e.g., 9-11 attack, the 2007-8 financial crisis). A fourth type of risk, *ignorance*, is the most hazardous form of risk where one cannot fathom an unknown unknown (Zeckhauser, 2006). There are a number of ways to handle and counteract risk. In the case of ignorance, leaders would have three options implied by our two models: project portfolio optimization, strategic inference and having had experience in similar situations – in short, wisdom.

Organizational Ambidexterity

Achieving balance under conditions of AI disruption requires virtuous leaders with ambidextrous skills. Organizational ambidexterity involves the ability to explore and exploit. Exploitation involves developing present capabilities with new efficiencies and incremental improvement. Exploitation involves competing with new technologies such as AI where flexibility and experimentation are required (O'Reilly & Tushman, 2013). This is the difference between evolutionary and revolutionary change. Managing evolutionary change requires a full measure of the virtue of subsidiarity since small innovations need the expertise of people closest to the business issue.

In the case of revolutionary change, greater trust is needed, and executives need to build up a reservoir of faith that major changes can be achieved. Rosing (et al., 2011) describes the executive challenge as one of directing opening versus closing activities. Exploration requires creativity and experimentation, where searching for new solutions and greater risk is accepted. The reality of implementing AI will take more than knowledgeable blocking and tackling of hidden back-office systems. It will take executives who think with clarity about the virtues needed to shape the processes and organization to the benefit of people first.

Corporate Results & Metrics

When it comes to measuring corporate results, it is important to be reminded of the purpose of one's measuring system in the first place. In the blizzard of firm data, managers often get focused on financial metrics alone when the real purpose of the firm (and technology systems such as AI) is to serve people. There is an important difference in the way terms are defined when leaders talk about results. Leaders will sometimes say that they seek what is *effective* (desired results), but often they only measure what is *efficient* (productivity, speed). In the context of AI, efficiency would mean a system achieves its task with a minimum of resources such as time, power, or memory. Effectiveness in the context of AI would mean that an algorithm achieves the goal of producing a valuable solution. Examples would be an algorithm that quickly searches messages versus one that removes spam and identifies critical emails on which to devote attention.

Outcomes are ultimate objectives that firm leaders seek to achieve, but these should not be confused with *outputs* – the measurable actions that contribute to the outcomes and that are only steps in a process toward achieving a result (Mills-Scofield, 2012). An output in AI would be a model-generated results, such as a prediction or classification. An outcome would go beyond that to the actual use, such as a more meaningful user experience or improvement in business performance.

Over time, firms often accept a one best-way of doing things (core business processes) because it is efficient. If firm leaders adopt an AI Cognitive Enhancement Strategy Model *Quadrant I Cognitive Enhancement*, or *Quadrant IV Labor Enhancement* approach (staff engagement), it opens the aperture to successes other than cost-cutting alone. Adopting a balance scorecard (Kaplan & Norton, 1996) can help augment a strategy by tracking progress toward common efforts (solidarity, common good) beyond finances to include customer, learning, and business process.

Human Augmentation

Policymakers are reasonably concerned about the socioeconomic impacts of AI's potential disproportionate impact on low-skill workers, but perhaps an argument can be made that there are promising options to augment and expand human abilities. We have previously noted how Quadrant I Cognitive Enhancement expands options (e.g., in medical diagnosis and surgery). The same can be true in other ways.

Agrawal and colleagues (Agrawal, Gans, & Goldfarb, 2023) note that the standard interpretation of the future is that low-skilled job will be replaced. They argue that low-skill workers may instead be able, with the help of AI, to expand opportunities for more complex roles. "Workers with generic skills are helped when AI is adopted to (allow those workers to) be able to participate in jobs previously only available to those with specialized skills" (p. 3). This would expand the number of available people to take on expanded new tasks, increase incomes, and potentially ease the imbalance of wages. In the case of already highly skilled professionals (e.g., physicians, mental health specialists, college professors), it would free them to perform even more complex tasks. The best targets for AI development may come not from building systems for skills that people lack, but instead building algorithms so more people benefit from the skills.

One could imagine several such neuropsychological examples. For instance, there are already genetic tests that forecast response to psychotropic medications and contraindicated therapy treatments. Similarly, there are brief AI-enhanced psychometric tests primary care physicians can use that capture Diagnostic and Statistical Manual of Mental Disorders diagnoses and preview the best forms of medication and talk therapy. Thus, physicians can save time and increase the quality of care. Imagine a future in which ML automatically detects the unique Big 5 personality factors (Barrick & Mount, 1991) and aspects of intelligence (Lubinski, 2004) of the operator. Over time the machine could adjust the way it interacts and perhaps cue the operator in ways that most complement the human's skills and abilities for personal growth.

Trust

Trust needs to be established in AI development. Trust that people will do what they say, and social ties are what makes commerce possible in the first place. Contracts, laws, and regulations are only a backstop. Coase (1937) argues that price mechanisms account only in part for the success of a firm. Inside the firm the coordinating function is largely relationships. Trust between participants inside the firm is magnified in the larger socioeconomic ecosystem. Trust within a firm and with external agents reduces the cost of each transaction by knowing there is less risk. Given the newness of AI and its associated uncertainty, the P2e4AI and the AI Cognitive Enhancement Strategy models underscore the factors that increase trust – virtuous and informed leaders – needed to establish lasting relations based on long-term interests of all parties. Trust inside the corporation and externally with suppliers, customers, and communities yields success.

Challenges to AI Trust

While many of the claims of danger with AI seem excessive, there is still reason to be cautious and exercise due diligence. For example, Kissinger (et al., 2023) notes, generative AI "appears superhuman or at least greatly enhanced cognitively ... at the same time, it possesses a capability to misinform its human users with incorrect statements and outright fabrications"

Human acceptance of algorithmic models can yield perplexing decisions. It would seem that most people would prefer an interpretable model, but that is not what DeStefano (et al., 2023) found. In a fashion industry study, the authors compared two models, one interpretable

(clear explanations for its predictions, e.g., rule based) versus another uninterpretable algorithmic model ("Black box" whose inner workings are hard to understand). A new uninterpretable model was designed to help experts decide the number and type of products to ship to each of its stores. Counterintuitively, under uncertain conditions, experts showed a preference for uninterpretable algorithmic models.

DeStefano (et al., 2023) developed two concepts related to the emerging theory of interpretable algorithms. First, *overconfident troubleshooting*—rejection of recommendations from interpretable algorithms happens due to the belief that decision makers already understand complex processes better than they actually do. Second, *social proofing the algorithm*—having included respected peers' data in AI development makes it more likely to accept what comes from an uninterpretable algorithm in situations where there is uncertainty. They overestimate their understanding of simpler models, yet they choose to select hidden models.

It seems on the surface to be a safer bet. When implementing AI, it follows that executives should alert managers and other staff to this phenomenon. In higher-risk situations people should prefer explainable AI. In a related way Banasiewicz (2019) contends that corporate reliance on industry benchmarks and best-practices can stifle innovation. Benchmarks are "easy to communicate" (p. 39), but if a firm is trying to be a top performer it will need to shoot for higher results.

Trustworthy AI

In most cases of technological innovation there is a tendency toward rapid prototyping, experimentation, and observing where things break. In the case of AI, we instead need to take a different tack. Building ethical AI requires that the deployment process is improved by the sort of reflection and design considerations in the P2e4AI Model. A proper balance of balance of deliberation and sense of urgency to build AI from which society can benefit is needed. "Think and (then) code" in AI/ML development says Morley (Morley et al., 2020, p. 2158). The classic phrase Roman emperors employed, *"festina lente"* (translated literally, make haste slowly, Morley et al, 2020, p. 2158) applies. In the AI business context this would mean that managers should proceed expeditiously, but with a prudent balance of urgency and diligence. Whereas cowboy coding with "all gun's blazing" may make sense in low impact situations, it can lead to adverse outcomes when it comes to algorithms that impact dignity. If one is programming AI for high-risk situations where safety is an issue (e.g., aviation), development should proceed with caution.

Confusion due to anthropomorphism is especially problematic in the context of discussions on human and machine collaboration. "Human-automation teaming" (HAT) implies that AI would act like a proactive team member and less like a human-controlled tool (with qualities such as shared goals, mutual coordination, awareness, and bi-directional understanding). In effective AI system design, however, it is necessary that all functions are described in unambiguous terms at the right level of detail (Kaliadros, 2022). Some AI functions can be automated, but they need explicit human control or human monitoring. Automation levels can be classified. The first, Human-within-the Loop (HWTL) is where the human plays a continuous supervisory control over the machine. The second, Human-over-the-Loop (HOTL),

involves human supervisory control with active monitoring to take control when needed. The third, Human-over-the-Loop (HOVTL), is where the automated systems can take independent action, informing the human by exception when it is necessary for him/her to engage.

We earlier noted that automation is a binary decision, as in aviation systems for example, when it comes to safety. Such decisions have moral consequences (as in AI/ML) and should not be left to engineers alone. Executives and managers need to be able to grasp how human-machine collaboration works. A basic question for AI is how much freedom it should have when executing a task. The answer requires that those in charge of AI implementation to clearly determine two factors: 1) automation - to what extent should a task be automated; and 2) autonomy – how autonomous the AI system should operate when performing the task (Simmler & Frisschnecht (2021)

Simmler and Frisschnecht (2021) developed a taxonomy that helps capture the basic structure of human-machine collaboration which helpfully augments our two models (especially the AI/LM Design Process section of the P2e4AI Model). They developed a five-by-five matrix tool that makes clear, regardless of technical knowhow, that categorizes the two aspects of collaboration – automation and autonomy.

Technical autonomy can range from: a) openness to expanding inputs; to b) adaptably learning from experience; to c) indetermination of not always leading to the same outputs; to d) non-transparency of not being able to determine how the system arrived at its output. Simmler's (et al., (2021, adapted from p. 245) levels of autonomy for systems are:

1) Deterministic – determined (every possible condition of the system unambiguously results in a subsequent condition of $n + 1$), closed (i.e., has no ability to learn), unadaptable, and transparent.
2) Non-transparent – determined, closed, unadaptable, and non-transparent.
3) Indeterminate – indetermined, closed, unadaptable, and non-transparent.
4) Adaptable – indetermined, closed, adaptable (systems that can alter their behavior), and non-transparent.
5) Open – indetermined, open, adaptable, and non-transparent

Automation on axis y levels are as follows: level 1 offers decisions; level 2 executes with human approval; level 3 executes if no human veto; level 4 executes and then informs the human; and level 5 allows the technology full independence without informing a human.

V: Summary and Conclusion

We set out in this article to explain how AI can contribute to human well-being and how firm managers can implement it in an ethical manner. At the start, we defined core AI terms, touching on how it works in order to explain the ways in which this technology innovation will contribute to what we believe can be a hopeful future. There is much in the way of confusion about AI. Major new general-purpose technologies have always had the capacity to displace workers, but prudent development and deployment can serve humanity in yet unseen ways.

In the case of meaningful work, it should be understood that firms are a socio-economic collaborative whose efforts ultimately serve people. A materialist *homo economicus* view of humans in their role in work ignores biological evidence that humans are social and seeking shared actualization as opposed to being narrowly self-interested rational individuals. People thrive when they cooperate intelligently with virtue in shared business objectives.

The philosophy of personalism developed in the 19th Century as a reaction to a dark materialist view of mankind. The central focus of all personalism's theorists is the concept of human dignity, the absolute moral value of each human person as ends in themselves and not as objects of some other purpose. Personalism and allied traditions (humanistic psychology and management) provide a rich source of principles that can be applied to practical AI ethics.

Likewise, the overlapping concepts that come from virtue ethics provide time-tested norms worth applying to firms' efforts to implement ethical AI. Virtues are dispositions of habit that build character while seeking flourishing (*eudaimonia*). Applying virtue does not involve application of rigid norms; rather it involves seeking the balance between possible goods (and avoidance of vice) to apply in particular circumstances. The most prominent virtue for executives and firm managers to apply is wisdom.

Human-technical systems in the firm need to be effective and orient toward the ultimate purpose, serving persons. AI systems should be built around human skills rather than seeking shortsighted efficiency stripped of the highest outcome – humans prospering.

We developed two models as tools to help guide firm managers in their search to apply ethics that acknowledge the primacy of intrinsic human dignity while working through the challenges of AI. The P2e4AI Model provides a graphical representation of a pathway to guide the blocking and tackling of decision making for AI development and deployment. The AI Cognitive-Enhancement Model specifically addresses strategy and how to orient the portfolio of opportunities and investments in a way that maximizes human flourishing with the help of AI tools.

Limitations

One potential criticism of our models for executive AI decision-making would be that even identical moral questions could lead to different outcomes. What a firm executive or subordinate team manager decides would depend on the experience and intent of the individual decision-maker. This is also the case more broadly for virtue ethics. The core issue is that the executive should make their best *eudemonic* (conducive to happiness) decision that balances virtues and can be applied to their situation, especially the virtue of wisdom.

Being virtuous as advised in our models is no guarantee of business success. Principles governing moral human behavior have persisted successfully since the time of Aristotle (and outlasted many management fads).

Hasnas' (1998) broad criticism of the state of business ethics would apply to our study. While our models attempt to simplify philosophical principles in plain language, there is room to derive greater specificity. More intermediate normative principles are needed, which we try to overcome in part by providing the questions in the appendix. In real-life application, it would require that each firm develop its own set of guiding questions and then train implementors on the principles. Business Ethics in a Box (Brennan, English, Hasnas, & Jawarski, 2023) developed at the Georgetown Institute for Markets and Ethics offers a host of materials for developing practical tools.

The open-ended nature of virtues and principles makes it difficult to distinguish one from another, or to even count the number of virtues and vices. Given that each person deciding virtue in a particular situation needs to decide between competing ends of a virtue makes that even more complicated. Ethics education is not about rules but more about developing character, and more can be done to examine what really works when it comes to the college classroom.

A number of the limitations Morley (et al., 2020) cites for their study also applies to our own. The distinctions between the AI/ML Design Process phases in our P2e4AI Model do not reflect the reality of everyone's design process. How exactly to govern ML is an open question, even if there is agreement that something needs to be done about it. Since the field is developing rapidly, it is also hard to keep track of applicable ethical AI tools to test.

Future Research

Going from the broad principle to the specific ethics in code is no small feat, but it is doable and a worthy goal to which virtue ethics can contribute as new algorithm tools. There continues to be a gap between the broad principles and clearly directed actions that needs to be filled. Greater collaboration between engineering teams and other professionals continues to be needed to fill that gap. Though AI would not exist without computer science, no single academic field has a lock on ethical AI. Math, engineering, psychology, sociology, public policy all have important roles to play in successful firm development and deployment of ethical AI. Bringing together multidisciplinary researchers into the development process of pro-ethical design tools and methodologies will be essential.

The opportunities for research on AI ethics are indeed wide open. As Brennan and colleagues (Brennan et al., 2023) note, well-intended people sometimes do bad (or misinformed) things. One avenue worth exploring along the lines of DeStefano (et al., 2023) is the heuristics people use when making decisions about the use of algorithmic models under conditions of stress. Diving deeper, it would be useful to look into the characteristics of those persons making decisions. As for experimental research using the models we developed, these would perhaps lend themselves to college student simulations comparing the effectiveness of decisions by major or in combined teams of mixed majors (e.g., engineering, psychology, philosophy, etc.). Experimental studies using different materials from the Brennan tools would also be interesting and extend our understanding of how people make decisions with regard to AI ethics.

Where Are We Now and Where Do We Need to Go?

Four decades plus after John Paul's encyclical *Laborem Exercens*, much progress has been made toward recognizing the inherent dignity of humans and firms focusing on the primacy of people over tools and capital. Crises drive us to look for deeper answers to economic problems. AI raises urgent ethical questions. However, we still need to see the person at the center of technology. In some ways, we find a greater openness to person-centered technology. This era of AI disruption offers an opportunity for firm leaders to heroically put human interests at the center of their decisions.

Closing Thoughts

In developing AI, it is imperative that firm executives and managers do it so an ethical way that preserves human dignity and contributes to long-term prosperity. Choosing what is good in ever-changing circumstances is a monumental challenge. Discerning what is good in an era of disruption requires leaders with the sort of character that uses right reason and wisdom. Self-mastery and clear thinking on technology business issues is the remedy for disorder (Smith, 2004).

A logical, human-centered approach to AI is an antidote to mindlessly pushing ahead on AI implementation that ignores the reality of firms as living cooperatives that lead to actualized lives. We believe our models contribute to that purpose. As Norvig (Lynch, 2021) says:

"One way to think of AI is as a process of optimization — finding the course of action, in an uncertain world, that will result in the maximum expected utility. In the past, the interesting questions were around what algorithm is best for doing this optimization. Now that we have a great set of algorithms and tools, the more pressing questions are human-centered: Exactly what do you want to optimize? Whose interests are you serving? Are you being fair to everyone? Is anyone being left out? Is the data you collected inclusive, or is it biased?"

references

Acemoglu, D. & Restrepo, P. (2018). *Artificial Intelligence, Automation and Work*. National Bureau of Economic Research. Working Paper 24196 http://www.nber.org/papers/w24196

Acemoglu, D. & Restrepo, (2019). *The wrong kind of AI? Artificial intelligence and the future of labor demand* (2019). Cambridge: National Bureau of Economic Research, Inc. Working paper series, 25682 - doi: https://doi.org/10.3386/w25682

Agrawal, A., Gans, J.S., & Goldfarb, A. (2023, The Turing transformation: Artificial intelligence, intelligence augmentation, and the skill premium. Brookings. https://www.brookings.edu/articles/the-turing-transformation-artificial-intelligence-intelligence-augmentation-and-skill-premiums/

Akrivou, K., & Scalzo, G. (2020). In search of a fitting moral psychology for practical wisdom: Exploring a missing link in virtuous management. *Business Ethics: A European Review*, *29*, 33-44.

Alford, H. (2022). The practical wisdom of personalism. *The Journal of Management Development*, vol. 29, no. 7/8, 2010, pp. 697–705, https://doi.org/10.1108/02621711011059130.

Aurik, J, Jonk, G. & Fabel, M. (2014). The history of strategy and its future prospects. AT Kearney. https://www.kearney.com/documents/291362523/291367011/History+of+Strategy+and+Its+Future+Prospects.pdf/73992a5e-afee-6240-8e19-e9a4af987470

Azhar, A. (2023, Nov 13). GPTs, GPTs everywhere, but AGI nowhere. *LinkedIn*. https://www.linkedin.com/pulse/gpts-everywhere-agi-nowhere-azeem-azhar-ehlre/?trackingId=buTIw51pQ2uSWBaCJ9635Q%3D%3D

Baghai, M., Coley, S., & White, D. (1999). *The alchemy of growth: Practical insights for building the enduring enterprise*. New York: Perseus Books.

Banasiewicz, A.D. (2019). *Evidence-based decision-making: How to leverage available data and avoid cognitive biases.* New York: Routledge.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Benson, R. H. (1926). *Lord of the world*. New York: Dodd, Mead & Co.

Bertolaso, M., & Rocchi, M. (2022). Specifically human: Human work and care in the age of machines. *Business Ethics, the Environment & Responsibility (Print)*, *31*(3), 888–898. https://doi.org/10.1111/beer.12281

Brennan, J., English, W.E., Hasnas, J. Jawarski P.M. (2023). Business Ethics in a Box. Washington, DC: Georgetown Institute for Markets and Ethics. https://www.businessethicsinabox.com/

Brown, B.R. (2023) *Engineering Intelligent Systems*: *Systems Engineering and Design with Artificial Intelligence, Visual Modeling, and Systems Thinking.* Hoboken, NJ: John Wiley & Sons.

Borrow, E. (2023, Nov 4). Fei-Fei Li: A tech pioneer focused on making AI a force for good. *Wall Street Journal*. Page B14. https://www.wsj.com/tech/ai/a-tech-pioneer-focused-on-making-ai-a-force-for-good-d60f8f10?mod=Searchresults_pos1&page=1

Box, G.E.P. (1976). Science and statistics, *Journal of the American Statistical Association*, 71:356, 791-799, DOI: 10.1080/01621459.1976.10480949

Cisco (2023, Nov 14) Global AI readiness index: Intentions outpacing. San Jose, CA. Cisco Systems. https://www.cisco.com/c/m/en_us/solutions/ai/readiness-index.html

Coase, R. H. (1937). The nature of the firm. *Economica*, *4*(16), 386–405. https://doi.org/10.2307/2626876

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Cosden, D. T. (1998). *The heavenly good of earthly work: The nature of work in its instrumental, relational, and ontological dimensions* (Order No. 10166612). Available from ProQuest Dissertations & Theses Global. (1826320868). Retrieved from https://www.proquest.com/dissertations-theses/heavenly-good-earthly-work-nature-instrumental/docview/1826320868/se-2

Cooley, M.J. (1982). Architect or Bee: The Human Technology Relationship. Boston, MA: South End Press. ISBN: 0896081311

DeStefano, T. Kellog, K.C., Menietti, M., & Vendraminelli, L. (2022, Oct 13, in R&R). Why providing humans with interpretable algorithms may, counterintuitively, lead to lower decision-making performance. *MIT Sloan Research Paper No. 6797-22*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4246077

Dilmegani, C. (2023, Jul 10) 4 Reasons for artificial intelligence (AI) project failure in 2023. https://research.aimultiple.com/ai-fail/

Drage, E. & Mackereth, K. (2022) Does AI debias recruitment? Race, gender, and AI's "Eradication of Difference. *Philosophy and Technology.* 35, 89 https://doi.org/10.1007/s13347-022-00543-1

Drucker, P. F. (1999). Peter Prucker on the new business realities. *Antitrust Bulletin, 44*(4), 795-793. Retrieved from https://www.proquest.com/scholarly-journals/peter-drucker-on-new-business-realities/docview/201020567/se-2

Floridi, L., Cowles, J., Beltametti, M., Chatila, R., Chazerand, P, Dignum, V, Luetge, C, Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valke, R. & Vagena, E. (2018). "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines (Dordrecht)*, vol. 28, no. 4, pp. 689–707, https://doi.org/10.1007/s11023-018-9482-5.

Floyd, S. (n.d.) Feiser, J. & Dowden, B. (eds.). Thomas Aquinas, Moral philosophy. Internet Encyclopedia of Philosophy. https://iep.utm.edu/thomasaquinas-moral-philosophy/

Fontrodona, J., & Melé, D. (2022). Thinking About the Future of work: Promoting Dignity and Human Flourishing. *Humanistic Management Journal*, *7*(2), 181–188. https://doi.org/10.1007/s41463-022-00136-2

Frankl, V. E. (1985). *Man's search for meaning*. Washington Square Press / Pocket Books.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting & Social Change*, *114*(January), 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Ghoshal, S. (2005). Bad management theories are destroying good management practices. *Academy of Management Learning & Education*, *4*(1), 75–91. https://doi.org/10.5465/AMLE.2005.16132558

Ghemawat, P. (2016). Evolving Ideas about Business Strategy. *Business History Review*, *90*(4), 727–749. https://doi.org/10.1017/S0007680516000702

Ghemawat, P. (2002). Competition and Business Strategy in Historical Perspective. *Business History Review*, *76*(1), 37–74. https://doi.org/10.2307/4127751

Ghousal, S., Bartlett, C., & Morgan, P. (1999, Spring). A new manifesto for management. *Sloan Management Review*. 40(3). 9-20.

Grant, R. M. (2019). *Contemporary strategy analysis* (10th edition.). New York: Wiley.

Guitián, G. (2023). How financial institutions can serve the common good of society: Insights from Catholic Social Teaching. *Business Ethics, the Environment & Responsibility*, *32*, 84-95.

Guitián, G., & Sison, A. J. G. (2023). Offshore outsourcing from a Catholic social teaching perspective. *Journal of Business Ethics*, *185*(3), 595-609.

Haekel, S.H. (1997). The development and application of organizational knowledge. IBM Advanced Business Institute

Hansen, F. and Smith, M. (2006). The ethics of business strategy. in Handbook of Business Strategy. Vol. 7 No. 1, pp. 201-206. https://doi.org/10.1108/10775730610618828

Hasnas, J. (1998). The Normative Theories of Business Ethics: A Guide for the Perplexed. *Business Ethics Quarterly*, *8*(1), 19–42. https://doi.org/10.2307/3857520

Harlick, R.M. & Elliot, G.L. (1980). Increasing tree search efficiency for constraint satisfaction problems. *Artificial Intelligence.* (14) 263-313.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, *35*(4), 519–530.

Herzberg, Frederick. (1959). *The motivation to work* (2d ed.). Wiley.

Hirsi Ali, A. (2023) Why I am now a Christian. *Unheard.* https://unherd.com/2023/11/why-i-am-now-a-christian/

Hołub, G. (2021). *Understanding the person: Essays on the personalism of Karol Wojtyła*. Bern: Peter Lang.

Hubbard, R. G., & O'Brien, A. P. (2021). *Economics* (8th edition.). Boston: Pearson.

Hühn, M.P., & Mandray, S. (2023). Is rationality reasonable? How ancient logos changes management theory. *Journal of Business Ethics*. https://doi-org.proxy.library.georgetown.edu/10.1007/s10551-023-05487-w

Hupke, H. (2023, Jan 10). Self-driving strollers and hands-free pickup trucks: What could possibly go wrong? *USA Today*. https://www.usatoday.com/story/opinion/columnist/2023/01/10/self-driving-baby-stroller-cars-child-safety/11019775002/

IEEE (2019) Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. Version 2. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (eds.) https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Janas, P. (2021). Doctoral dissertation: Framing meaningful work with personalism and virtue ethics. Pontifical University of St. Thomas, Rome.

John Paul II (1981). On human work: Encyclical Laborem exercens. Washington, D.C.: Office of Publishing Services, US Catholic Conference. https://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091981_laborem-exercens.html

John Paul II. (1982). *Encyclical On Human Work: Laborem Exercens: on the Ninetieth Anniversary of Rerum Novarum*. St. Paul's Editions.

Kaliardos, W.N. (2022). Enough fluff: Returning to meaningful perspectives on automation. IEEE/AIAA 41st Digital Avionics Systems Conference (DASC). Norfolk, VA.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Harvard Business Review Press.

Kim, T. W., Fabrizio, M., Katherina, P., Sison, A. J., & Benito, T. (2021). Master and slave: The dialectic of human-artificial intelligence engagement. *Humanistic Management Journal, 6*(3), 355-371. doi: https://doi.org/10.1007/s41463-021-00118-w

Kissinger, H.A. (2022). *Leadership: Six Studies in World Strategy*, New York: Penguin Press. ISBN: 9780593489444

Kissinger, H.A, Schmidt, E., & Huttenlocher, D. (2023, Feb 28). ChatGPT Heralds an Intellectual Revolution. *The Wall Street Journal. Eastern Edition.*

https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6

Knight, F. H. (Frank H. (1921). *Risk, uncertainty, and profit*. Houghton Mifflin Company.

Kosslyn, S.M. (2019) Are you developing skills that won't be automated? Harvard Business Review. https://hbr.org/2019/09/are-you-developing-skills-that-wont-be-automated

Lee, D. (2016, Mar 16). Tay: Microsoft issues apology over racist chatbot fiasco. *BBC News*. https://www.bbc.com/news/technology-35902104

Likert, R. (1961). *New patterns of management.* McGraw-Hill.

Lynch, S. (2021, Oct). Peter Norvig: Today's most pressing questions in AI are human-centered. Stanford University Human-Centered AI. https://hai.stanford.edu/news/peter-norvig-todays-most-pressing-questions-ai-are-human-centered

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 Years After Spearman's 1904 "'General intelligence,' objectively determined and measured." *Journal of Personality and Social Psychology*, *86*(1), 96–111. https://doi.org/10.1037/0022-3514.86.1.96

MacIntyre, A. C. (1994). How can we learn what Veritatis Splendor has to teach? *The Thomist*, *58*(2), 171–195. https://doi.org/10.1353/tho.1994.0026

Marin, A., Boanță, L., Hadăr, A., Badea, D. M., Vlăduț, G., Bucur, D., Ciocănel, B., & Ivan, I. (2015). Business models and competitive advantage for technology transfer entities. *Romanian Review Precision Mechanics, Optics & Mechatronics*, *2015*(48), 103–109.

Marr, B. (2020). 10 essential leadership qualities for the age of artificial intelligence. *Forbes.* https://www.forbes.com/sites/bernardmarr/2020/10/12/10-essential-leadership-qualities-for-the-age-of-artificial-intelligence/?sh=564894497f79

Martinez, J. (2023, Sep/Oct). People may be more trusting of AI when they can't see how it works. *Harvard Business Review*. p. 1-5. https://hbr.org/2023/09/people-may-be-more-trusting-of-ai-when-they-cant-see-how-it-works

McGregor, D.M. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.

Mills-Scofield, D. (2012, Oct). It's not just semantics: Managing outcomes vs. outputs. Harvard Business Review. https://hbr.org/2012/11/its-not-just-semantics-managing-outcomes

Melé, D. (2009). Integrating personalism into virtue-based business ethics: The Personalist and the common good principles. *Journal of Business Ethics*, *88*(1), 227–244.

https://doi.org/10.1007/s10551-009-0108-y

Melé, D. (2016a). Re-thinking capitalism: What we can learn from scholasticism?: JBE. *Journal of Business Ethics, 133*(2), 293-304. doi:https://doi.org/10.1007/s10551-014-2368-4

Melé, D. (2016b) Understanding humanistic management. *Humanist Management Journal* **1**, 33–55 https://doi.org/10.1007/s41463-016-0011-5

Melé, D., & Fontrodona, J. (2017). Christian ethics and spirituality in leading business organizations: Editorial introduction: JBE. *Journal of Business Ethics, 145*(4), 671-679. doi:https://doi.org/10.1007/s10551-016-3323-3

Morley, J., Floridi, L., Kinsey, L. & Elhaial, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods, and research to translate principles into practices." *Science and Engineering Ethics*, vol. 26, no. 4, pp. 2141–68, https://doi.org/10.1007/s11948-019-00165-5.

Novak, M.J. (1996). *Business as a calling: Work and the examined life*. New York: The Free Press.

O'Reilly, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *Academy of Management Perspectives*, *27*(4), 324–338. https://doi.org/10.5465/amp.2013.0025

Pakaluk, M. (2005). *Aristotle's Nicomachean ethics: An introduction*. Cambridge University Press.

Reilly, R.R. (2011). The Closing of the Muslim Mind: How Intellectual Suicide Created the Modern Islamist Crisis. Washington, DC: Robert Gateway.

Rogers, E.M. (2003). Diffusion of Innovations. (5th ed.) New Yok: Free Press.

Rosenbrock, H. H. (1982). Engineers and the work that people do." *Measurement and Control (London)*, vol. 15, no. 10, 1982, pp. 387–91, https://doi.org/10.1177/002029408201501004.

Rosenhead, J. (1981). [Review of *Architect or Bee: The Human/Technology Relationship; Living Think Work: Where Do Labour Processes Come From?*, by M. Cooley & M. Hales]. *The Journal of the Operational Research Society*, *32*(7), 615–617. https://doi.org/10.2307/2580933

Rosing, K., Frese, M. & Bausch, A. (2011). Explaining the heterogeneity of the leadership-innovation relationship: Ambidextrous leadership. *The Leadership Quarterly*, *22*(5), 956–974. https://doi.org/10.1016/j.leaqua.2011.07.014

Russell, S.J. & Norvig, P. (2022). *Artificial Intelligence: A Modern Approach (Pearson Series in Artificial Intelligence)* (4th ed.). London: Pearson. ISBN: 978-0134610993

Schein, E. H. (2010). *Organizational culture and leadership* (4th ed.). New York: Jossey-Bass.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *The American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Shirkin, R. (2020). Artificial Intelligence: The Complete Beginners' Guide to Artificial Intelligence. Amazon KDP Publishing. ISBN-13 979-8609154415

Simmler, M. & Frischknecht, R. (2021). A taxonomy of human–machine collaboration: capturing automation and technical autonomy. *AI and Society* 36 (1):239-250
- *Artificial Intelligence and Law* 31 (2):213-237.

Smith, A. (1759, 2010 edition). The Theory of Moral Sentiments. Ryan Patrick Hanley, Editor. London: Penguin Classics.

Smith, A. (1790, 2005). The Theory of Moral Sentiments. 6th Edition. Sao Paulo: MetaLibri.

Smith, A. C., & Yandle, B. (2014). *Bootleggers and Baptists: How economics forces and moral persuasion interact to shape regulatory politics*. Washington, DC: Cato Institute.

Smith, S. B. (2004). An exemplary life: The case of René Descartes. *The Review of Metaphysics*, *57*(3), 571–597. http://www.jstor.org/stable/20130344

Stanford Encyclopedia of Philosophy. (2022). Personalism. https://plato.stanford.edu/entries/personalism/

Taylor, F. W. (1911). *The principles of scientific management*. Harper & Brothers.

Taleb, N.N. (2015). Silent Risk: Lectures on Fat Tails, (Anti)Fragility, and Asymmetric Exposures. SSRN Electronic Journal. DOI: 10.2139/ssrn.2392310

Van Biljon, P. (2022). *Innovation for value and mission : an introduction to innovation management and policy*. De Gruyter.

van Staveren, I.P. (2009). Virtue ethics. In Eds., Handbook of Economics and Ethics. Cheltenham, UK: Edward Elgar. ISBN 978 1 94542 926 1

Winston, P.H. (1984). Artificial Intelligence. (2nd Ed.). Reading, MA: Addison-Wesley.

Yandle, B. (1983, May). Bootleggers and Baptists - The education of a regulatory economist. *Regulation, 7*, 12. Retrieved from

https://www.proquest.com/magazines/bootleggers-baptists-education-regulatory/docview/210511343/se-2

Yandle, B. (1999). Bootleggers and Baptists in retrospect. *Regulation (Washington. 1977)*, *22*(3), 5–.

Zeckhauser, R. (2006). Investing in the unknown and unknowable. *Capitalism and Society*, 1(2), 5-. https://doi.org/10.2202/1932-0213.1009

Appendix: Sample High-level Questions for AI Corporate Ethics

Theme	Question
Strategy	● Are our metrics focused on the right things, and what do we really measure? Is our scorecard truly balanced for the current and future situation? ● Should we seek incremental or radical innovation for our business with AI? ● Are we realistic in our expectations for the AI results we can achieve, and how might we need to augment our workforce to face the challenge?
Process	● As we enter a period of change, what will be the major changes in the business process? ● Is there one way of doing business or are there multiple paths and projects that impact organizational structure? ● Should AI development proceed within a smaller unit that advises others, or should it take place simultaneously across the organization and geographies?
Renewal	● Have we appropriately anticipated the impacts of changes that may take place, and have we put in place mechanisms to reduce the unknowns? ● Is our organization ready for the change? What do we need to put in place to ensure success? ● In what ways might the corporate culture not be ready, and in what ways can we get people ready?
Common good	● Have we instituted a Risk Evaluation Board that reports directly to the CEO?

	● What are the distinctive differences we provide to customers through our AI initiatives, and how does that contribute to society?
Right order	● Are we delegating decisions to the right level and giving people the freedom to make decisions where the expertise is greatest?
Solidarity	● Are we upskilling the workforce to be ready for and minimize potential disruptions? ● Am I willing to listen to dissenting voices and fully engaging their ideas and logic? ● Would it make sense to have an "AI immersion week" in which people throughout the firm try to handle as many tasks using AI as possible to give them a sense of greater comfort?
Magnanimity	● Are we communicating the broad vision for AI in a way that people understand, and am I committing to excellence in its use?
Humility	● Have I considered my blind spots, and do I ask for internal and external feedback? ● Where I lack the skills, have I engaged with the right people that will complement my strengths and weaknesses? ● Am I working on my own personal professional development and helping those around do the me?
Wisdom	● Is the board proactively informed so that we speak with one voice? ● Are we managing the balance between current competencies that can sustain the business for the near term while managing risk to explore and invest in new AI opportunities? ● What is our prior experience with major technology implementations, and how can we make this one work?
Fairness	● Do we have the right incentives in place, financial and non-financial, to make our AI initiatives benefit stakeholders appropriately? ● How will our initiatives serve our employees, customers, suppliers, stockholders, the communities in which operate, and society?

Courage	● Are we reaching for the stars with AI in a way that can realistically propel the business forward, or on the other hand, or are we playing it so safe that it puts the long-term at risk?
Control	● Do the algorithms we use provide adequate insight and understanding given the type of business issue? ● Have we managed to answer with "value alignment problem," that is, achieved agreement between our true preferences and the goals we put into the machine?
Use Case	● Is everyone on the design team in agreement with how the system should interact with users to accomplish the goal? Is it the right goal? ● Does the goal also support human agency in addition to other considerations? Are we developing business goals that put people first?
Design	● Who is the human responsible for the outcomes at each level of development? ● Are we using the right algorithm tribe for the problem at hand? What are the pros and cons of the alternatives? ● Are we leaving development to our engineers alone, or do we have an integrated team that can offer expertise from other professions?
Data	● Do we have the right data? Is the data appropriately cleaned and structured? ● What are the missing pieces, and could that lead to bias?
Build	● What ethical tools are we using to evaluate consequences at each stage of development? ● Does the functional decomposition clearly delineate that appropriate control as each level and step? ● Are we sure that the purpose we put into the machine is the purpose we really desire?
Test	● Where an ethical tool is not immediately available, how do we fill in the gap to make management and team decisions? ● Do we have an adequate understanding of the task environment (performance, environment, actuators, sensors) in which the agent will operate?

	● Do our performance measures reflect what we are trying to achieve?
Deploy	● Are we truly ready throughout the firm – technically and organizationally – ready to deploy the AI? ● Do we have adequate safeguards against bias and other mishaps? ● Do the results of our AI deployment achieve the human-centered objectives we desired, and how can we refine our future deployments? ● As we deploy our AI tools, are we looking for non-expected results that will help us refine our assumptions about how people really use AI?
Monitor	● What are the mechanisms in place to keep track of the impact our AI has on our products and services? How does it impact our customers, society, and the future of our organization? ● Do we conduct independent audits of the integrity of our systems? ● How do we safeguard data? ● How do we ensure that what we have developed is not used for malevolent purposes?
Self-realization	● Does the data we use to make AI-aided decisions also include provision for context, emotion, and relationships that play a role in human choice? ● Are we investing in developing roles that increase cognitive enhancement?
Enhance agency	● Are our AI tools provably beneficial to the humans impacted (i.e., value aligned)? ● How do our AI tools contribute to better use of human capacities? ● Are we prioritizing human wellbeing in our AI system design? ● As we delegate some tasks to machines, do we maintain oversight at the proper level?
Societal capacity	● Are our AI efforts replacing the value of people, or on the other hand, do they enhance workers' strengths and contribute to their actualization? ● Are we guarding privacy in use of AI, and do we have the proper security safeguards in place?

Cohesion	● Are we seeking to develop AI in the firm that answers economic and efficiency questions without considering the effects on persons? ● Are we collaborating with standards bodies and regulatory agencies in providing feedback on what is effective and safeguards society and individuals?