8-2020

# A Measure of All Minds: A Classification of the Artificial Intelligence Strengths and Virtues & the Creation of the THETIS Dimensions of Cybernetic Wellbeing

Seth Robin Norman
*University of California, Berkeley; Stanford Graduate School of Business; University of Pennsylvania*, seth.norman@gmail.com

# A Measure of All Minds: A Classification of the Artificial Intelligence Strengths and Virtues & the Creation of the THETIS Dimensions of Cybernetic Wellbeing


## Abstract

Artificial intelligence has generated more ambivalence than any technology in human history. As Silicon Valley enters the third decade of its self-proclaimed era of artificial intelligence, there is still surprisingly little agreement about the nature of the 21st century's defining invention, and even less about how our relationship with this seminal technology should be managed in the future. Positive psychology offers software engineers methods to cultivate a greater understanding of the unique strengths of the artificial intelligence programs they develop, as well as the effects to wellbeing triggered by the applications they deploy. In this paper, I will propose three tools inspired by my chosen field of study for use by artificial intelligence innovators: (a) a classification of the artificial intelligence strengths and virtues; (b) the THETIS dimensions of cybernetic wellbeing, and; (c) the definition of a positive existential posthuman philosophy of artificial intelligence design. The philosophy of positive psychology is perhaps most succinctly summarized with a single phrase: "other people matter." If Silicon Valley is to deliver a clearer and more compelling vision of the future of artificial intelligence—one in which human and machine agents work and thrive in collaborative harmony—then it must update its innovation practices to embrace a similarly transformative point of view: "other consciousnesses matter," too.



## Keywords

Artificial intelligence, positive psychology, existential psychology, posthumanism, design-thinking, cybernetics


## Disciplines

Arts and Humanities | Electrical and Computer Engineering | Philosophy | Psychology

# A Measure of All Minds:

# A Classification of the Artificial Intelligence Strengths and Virtues & the Creation of the THETIS Dimensions of Cybernetic Wellbeing

Seth R. Norman

University of Pennsylvania

A Capstone Project Submitted

In Partial Fulfillment of the Requirement for the Degree of

Master of Applied Positive Psychology

Advisor: Martin E.P. Seligman, Ph.D.

August 1, 2020

# A Measure of All Minds:
# A Classification of the Artificial Intelligence Strengths and Virtues & the Creation of the THETIS Dimensions of Cybernetic Wellbeing

Seth R. Norman
srnorman@sas.upenn.edu

Capstone Project
Master of Applied Positive Psychology
University of Pennsylvania
Advisor: Martin E.P. Seligman, Ph.D.

## Abstract

Artificial intelligence has generated more ambivalence than any technology in human history. As Silicon Valley enters the third decade of its self-proclaimed era of artificial intelligence, there is still surprisingly little agreement about the nature of the 21st century's defining invention, and even less about how our relationship with this seminal technology should be managed in the future. Positive psychology offers software engineers methods to cultivate a greater understanding of the unique strengths of the artificial intelligence programs they develop, as well as the effects to wellbeing triggered by the applications they deploy. In this paper, I will propose three tools inspired by my chosen field of study for use by artificial intelligence innovators: (a) a classification of the artificial intelligence strengths and virtues; (b) the THETIS dimensions of cybernetic wellbeing, and; (c) the definition of a positive existential posthuman philosophy of artificial intelligence design. The philosophy of positive psychology is perhaps most succinctly summarized with a single phrase: "other people matter." If Silicon Valley is to deliver a clearer and more compelling vision of the future of artificial intelligence—one in which human and machine agents work and thrive in collaborative harmony—then it must update its innovation practices to embrace a similarly transformative point of view: "other consciousnesses matter," too.

*Keywords:* Artificial intelligence, positive psychology, existential psychology, posthumanism, design-thinking, cybernetics

# Table of Contents

**Acknowledgements**

This capstone is dedicated to my beautiful family, without whom my work would have been finished much sooner. I love you, please go to bed.

I also offer my sincere appreciation to Dr. Martin E.P. Seligman, who was kind enough to be coaxed out of his well-deserved sabbatical to help instill some semblance of discipline to my writing. After a career defined by the creativity and influence of his own thinking, I am still inspired by the sincere enthusiasm with which he engages the perspectives and propositions of his students. Thank you for your curiosity and open-mindedness while I processed my mishmash of ideas into something resembling an academic paper. I will remember your virtuous examples of kindness and generosity as my professional career advances, and most especially, in my role as a parent to two very precocious and curious young children.

Above all, I would like to thank my classmates, who have always challenged me with the brilliance of their minds, the loveliness of the hearts, and the perseverance of their spirit. That you have also become my friends is a gift I treasure, and I very much look forward to reuniting with each of you in person as soon as our health and safety allow.

Finally, I wish to remember the people who have made the City of Brotherly Love my home during this strangest of years. For as long as I have been writing this capstone, I have recharged by wandering the city's cobblestone streets lost in my own thoughts until I am, often enough, also lost in Philadelphia. On the very first of these expeditions, I passed through Rittenhouse Square, where I encountered a poet sitting at a small folding table composing poems on his 1960 Smith Corona Skywriter. I stopped and watched as he requested suggestions for new topics to be shouted out from passersby, replying with an explosion of keystrokes as he improvised an original poem in just a few short minutes of creative flurry. I wasn't going to pass up the opportunity to share in his artistic exuberance, so I yelled out "artificial intelligence" and stood awestruck as he wrote a clearer vision for my future capstone than I had managed in weeks of solitary attempts.

I will close my acknowledgements with the poem he wrote for me, not only because of its relevance to my work, but because it exemplifies the sort of serendipitous collaboration that occurs between artists, students, and tourists in the streets and parks of Philadelphia every day. It is this spirit of interactivity and innovation that I will miss most when we return to the west coast at the end of summer.

*Artificial Intelligence*

an algorithm learns almost like we do

compiling information and creating better outputs

oh, and how we learn from what the algorithm can compute

absorbing the world

at the speed of light

there's a true intelligence there

that conjures disbelief to connect all the variables

and see the world as the machine begins to see it

a hybridized version of being

coming to life in front of us.

—Marshall James Kavanaugh, unpublished work, 2020

**Mount Parnassus and Sand Hill Road**

In his 2012 book, *Flourish: A Visionary New Understanding of Happiness and Wellbeing,* Dr. Martin Seligman concludes his reflections on the progress of positive psychology with a bold vision for its future: 51% of the world's population could be psychologically flourishing through the dissemination of positive psychology by the year 2051 (Seligman, 2012). He calls this moonshot Flourish 51, and because it is a goal so thoroughly transfused with the optimism of its field of origin, Flourish 51 is sometimes dismissed as merely a beautiful sentiment.[1] In truth, it is a well-defined objective towards which great progress has already been made; progress made possible by the application of practical scientific research which began with the foundation of the work of Seligman and his collaborators. Their foundational investigation of the many character strengths and virtues of humanity allowed positive psychologists to later identify the distinct dimensions of wellbeing towards which these strengths could be virtuously applied to our lasting benefit, and these same discoveries are still driving progress in positive psychology today.

However, the finish line for Flourish 51 no longer lies in the distant future. We live just one generation away from 2051, and as we approach the midpoint of this century, the nature of humanity and the means of human flourishing are changing at a rapidly accelerating pace. The adoption of new digital technologies and their increasing pervasiveness in our daily lives is altering the ways in which we work, live, and connect to each other as a species; and of these emergent technologies, none holds greater transformative potential for the human race than the invention of artificial intelligence (AI) (Brockman, 2020; Thompson, 2013; Webb, 2019). As we enter a period in human history defined by our rapidly evolving relationship with AI, Seligman's

[1] In *Flourish*, Seligman calls his great ambition PERMA 51in reference to the PERMA dimensions of human wellbeing which were also first presented in the book.

challenge to the scholars and practitioners of positive psychology is born anew: how will we define our goals for Flourish 2151 and for the centuries beyond?

The good news is that there are many reasons to be optimistic about the future of natural and artificial intelligence, and there is still time to thoughtfully define this important relationship. While most people tend to wonder whether the future potential of AI is best characterized as an all-powerful savior or an all-devouring curse, the truth is that most experts would characterize our present-day examples of this technology as all hype (Hendler, 2008). However, this does not mean there is time to waste. The decisions made by designers and engineers today will greatly inform the types of artificial intelligence programs and the nature of their effects on humanity tomorrow. If Silicon Valley is to deliver a clearer and more compelling vision for the future of AI technology—one in which human and machine intelligences work and thrive in collaborative harmony—then it must embrace the perspectives and practices of positive psychology.

With this thought in mind, I will introduce three tools inspired by the field of positive psychology for use in the field of artificial intelligence innovation:

1) **The classification of the artificial intelligence strengths and virtues:** Following the same methodology used by the founders of positive psychology to identify the best elements of human nature, I will identify the positive characteristics of well-applied artificial intelligence technologies and then synthesize a classification of its most universal and virtuous qualities.
2) **The creation of the THETIS dimensions of cybernetic wellbeing:** The commonly accepted definition of human thriving must adapt to the growing influence of artificial intelligence in our daily lives. Relying upon the emerging field of existential positive psychology as a model for identifying opportunities for the generation of greater

meaning, I will propose a model of cybernetic wellbeing using terms psychologists and technologists can both agree upon for future discussions about the nature of human and AI collaborative wellbeing.

3) **The proposal for a new positive existential posthuman philosophy of artificial intelligence design:** The human-centered design-thinking innovation processes that rule Silicon Valley today are not what software developers need to imagine the best possible futures of artificial intelligence. Rather than focusing on rigid pathways designed to help identify short-term needs and rapidly deliver new products to market, I believe an emerging point of view called posthumanism can help realign the core values that once drove the spirit of innovation that made Silicon Valley the global hub of technology.

Alongside his bold vision for the future of positive psychology, Seligman uses some of the final pages in *Flourish* to remind us of the field's ancient roots; indeed, the modern science of positive psychology is built on the foundation of Seligman's years-long survey of the great mythologies of the ancient world (Dahlsgaard, Peterson, & Seligman, 2005). Recognizing that our modern understanding of human flourishing was born from a reexamination of civilization's oldest stories, we will begin our discussion of the potential for the future emergence of cybernetic wellbeing by returning to the same classic texts for inspiration.

**Croesus in Crisis**

Long ago, a man named Croesus ruled over a powerful kingdom called Lydia (Bowie, 2007). Dominion over an important crossroads in the Hellenic world had made Lydia one of the richest and most advanced civilizations in the world, but its ruler still faced a serious dilemma: Lydia was threatened by the growing strength of the neighboring Persian Empire. Croesus was forced to decide whether to seek an alliance with the Persians in order to delay their invasion, or to strike first in the hopes of surprising his unprepared enemies; it was a decision that would make or break his rule. So, what is a king to do?—like any competent leader of the age, he traveled to the Oracle at Delphi and climbed to the top of Mount Parnassus, where he asked the Pythian priestess to reveal the best course of action. She replied with one of the most famous prophecies in all of recorded history: “If Croesus attacks the Persians, he will destroy a great empire” (Bowie, 2007).

It may seem odd that a paper about the thoroughly modern subject of artificial intelligence would begin with a parable from Greek antiquity, but there is good reason to revisit this cautionary tale of impulsivity and hubris. That Croesus would lose his life and his empire at the hands of Cyrus the Great after ordering an ill-advised raid on Persian forces in the Battle of Thymbra in 547 BCE (Dillery, 2002) seems glaringly obvious to our modern ears. What else could the king have expected after putting the fate of his empire in the hands of the Oracle and its mystical prophecies? However, the myth of Croesus and the Oracle holds special relevance to a very modern crisis now unfolding in Silicon Valley precisely because technologists continue the exact same practices that doomed Croesus so long ago.

The technology industry’s failure to make sense of artificial intelligence has brought it to the brink of its own existential crisis. America’s technological influence is in decline, and it is no longer taken for granted that we will lead the world in answering the most important questions about AI (Madrigal, 2020). This decline is emerging alongside the unfolding of a new technological epoch—the dawning of the Artificial Intelligence Age—which is already at risk of being ceded to innovation centers in China, Israel, and Russia (Wang & Chen, 2018). It’s hard to overstate how important it is for Silicon Valley to reestablish leadership over the development of artificial intelligence, and not just because so many experts have predicted elaborate doomsday scenarios if this powerful new invention one day escapes our control (Dubhashi & Lappin, 2017). The truth is that, while existential threats to mankind make for entertaining dinner conversation, a much more uncomfortable truth has begun to spoil the appetite of San Francisco's entrepreneurial class: after decades of promises and marketing hype, the reality of artificial intelligence still falls short of its great potential (Kenney, 2000).

At the core of this problem lies a fundamental error of perspective. The rise of AI has drawn calls for the establishment of new innovation safety guidelines from international governments, ordinary tech workers, and even the holy pontiff (McCormick, 2020; Walker, 2019; Whitby, 2008). I welcome discussions of this nature, but unfortunately, the approach of reiterating narrowly defined ethical frameworks isn't enough to save us. The pace of technological and cultural change is accelerating too quickly, and new programs are evolving too rapidly to reasonably expect government regulators to meaningfully prevent every potential nightmare scenario. Worse still, these ethical frameworks seek to define and predict the worst-possible consequences of a future shared by human and artificial agents, without offering any meaningful suggestions for combining the strengths of these two unique forms of consciousness in ways that promote our greater common wellbeing. If Silicon Valley is to reclaim its place at the forefront of technological innovation, then is must desist with the dusting off and revising of the same tired lists of techno-ethics it created at the dawn of the Internet Age, and instead offer engineers and designers tools they can use to start building the future Age of Artificial Intelligence.

**Athens and Artificial Intelligence**

What approach would better serve us then? For an answer to this question, I believe there is another Greek myth about a visit to the Oracle that offers meaningful insight to our modern ears:

Not long after the demise of Croesus, the Persian Empire threatened the Mediterranean world once more; this time, the stakes were even higher. King Xerxes commanded the largest army ever fielded in human history and he was marching it towards Athens, the very center of Greek power and culture (Fornara, 1967). This time it was Themistocles, leader of the Athenians and commander of all their armed forces, who rushed to the summit of Mount Parnassus in the desperate hope that the priestess could help solve his dilemma. What should the Greeks do to save their civilization?—Pythia's reply was as oracular as ever: "Though all else shall be taken, Zeus, the all-seeing, grants that the wooden wall only shall not fail" (Fornara, 1967).

What?

An answer like that didn't leave Themistocles much to go on. Still, he needed an answer, and lacking the bravado of Croesus he was forced to try an entirely different strategy: he let the Greeks argue about it (Fornara, 1967). Members of the governing assembly, the supreme military council, and even common citizens who had accompanied the Athenian delegation to Pythia were invited to question and prod each other for the meaning of the priestess's words. When no clear sense of the prophecy could be made after several days of spirited debate, Themistocles asked the Oracle to give him another. Over and over, the Athenians asked and argued and then asked again, until finally, Themistocles was able to convince his countrymen that the wooden wall described in the revelation was a reference to Athenian warships; the Oracle was saying the Greeks would survive if they took their fight to the sea. Themistocles was right, and though

Athens would twice be evacuated and razed to the ground, its people would still see final victory against the Persian invaders at the Battle of Salamis in 480 BCE, the first largescale naval engagement in martial history (Bowie, 2007).

In this story, the Athenians offer us practical examples of not only useful innovation practices, but of history's first successful deployment of an artificial intelligence program. Many classicists and psychologists have already identified the ancient Greek contributions to the cognitive process of prospection, meaning the process by which the human mind generates and evaluates the potential outcomes of future actions (Eidinow, 2007; Seligman et al., 2013), but I would go one step further and suggest that the Athenian's collaboration with the Oracle represents the very first deployment of a forecasting and decision-aiding algorithm. There are several compelling details about their interaction which serve as evidence to support this insight:

1) **There was a planned systems architecture:** Though the stories of Apollo divulging the fate of the world through his immortal priestess are apocryphal to say the very least, the temple at Delphi, as well as the people and rituals bound to it, were very real. When a petitioner arrived atop Mount Parnassus, he would not have spoken directly to Pythia (Stadter, 2005). Instead, a routine involving hundreds of temple priests would begin with the reception of the petitioner in the temple's entry chamber, followed by the receipt of his question. This question was then pass from room to room, priest to priest, and would eventually work its way up the temple hierarchy before finally reaching Pythia's ears.
2) **Randomness was injected for greater creativity:** Pythia, who remained in esteemed isolation, achieved the appearance of immortality because she was constantly replaced by different young women taken from the surrounding country side (Cusack, 2009; Eidinow, 2007). These girls were typically societal rejects—most likely suffering from

schizophrenia, dementia, or another psychotic disorder—and their atypical neurology would be further intensified by the constant flow of hallucinogenic ethylene gases rising from deep natural furrows in the ground into their private chambers.

3) **There was a process of analysis optimization:** Once a priest received Pythia's answer to the petitioner, he began the long journey of returning it back down the chain of command to the entry hall (Broad, 2007). Throughout this process, other priests would reorganize Pythia's message into poetic verse, sometimes freely reinterpreting her words as part of an ongoing game of ancient Greek telephone, until her message finally reached the petitioner's ears with an appropriate level of polish.

4) **Provisions were made for continuous human interaction:** If the priestess needed to sleep or had nothing to say, whichever attendant was on hand in the reception hall would accept yes or no questions which he then answered by blindly pulling different colored beans out of a jar (Broad, 2007); whatever worked to keep the process of prospection going. The guests could stay and continue to ask as many questions as they needed (provided they could afford it) to carry on their process of debate until they arrived at a satisfactory answer to their own question.

What the Greeks invented at Delphi was not a ritual for the interpretation of divinity, but rather a technologically-enabled process for generating and analyzing new ideas. The Oracle was an ancient artificial deep neural network deployed as a gigantic flesh-and-blood, brick-and-mortar forecasting and analysis program that served the greater Mediterranean world for generations; and the great genius of the Greeks was to recognize and appreciate the nature of this important relationship (Eidinow, 2007).

Leaders like Themistocles viewed the future as an object of creation that was born from an ongoing process of collaborative innovation that relied upon the greatest possible diversity of perspectives for success (Eidinow, 2007; Seligman et al., 2013). I believe this ancient methodology could also bring success to the modern work of man and machine-kind, two forms of intelligence already partnered in building a shared future; this future will be better served by a more thoughtful approach to imagining, designing, and managing this most essential of partnerships.

## A Short Primer on Artificial Intelligence and Silicon Valley Innovation Practices

This section will prepare the reader for a thoughtful discussion by introducing the state of AI innovation as it exists in Silicon Valley today. It is absolutely critical that no one feel deterred from participating in this discussion because of a lack of technical vocabulary; it is possible to follow and contribute with just a basic introduction to common terminology. However, for those who are interested, I have included an extended glossary of basic artificial intelligence terms and a guide for the identification of the different AI innovation eras as Appendix 1 and Appendix 2 of this paper, respectively. Hopefully, this will be enough of an introduction to encourage broader engagement from laypersons with the topic of artificial intelligence development; either way—participate or not—we will all still be affected by the outcomes of these ongoing conversations.

### Artificial Intelligence

There are just three points that must be understood in order to capture the true scope and breadth of this capstone:

The first is that Silicon Valley has just entered the 3rd generation of artificial intelligence innovation. 3rd generation AI is sometimes called Theory of Mind AI because Silicon Valley engineers are now able to design programs that mimic the architectures and abilities of the human brain (Barrat, 2013; Kaptelinin, 2018). We're still a long ways off from reaching the fourth and final generation of artificial intelligence agents—more commonly known as conscious AI, and still very much confined to the world of science fiction—but the decisions innovators make about their programs today will greatly impact the nature of those next-generation inventions; this is why almost all of my work is focused on improving the practices of this current era of innovation.

Second, present innovation is focused on the blending of human and artificial intelligences. No computer in the world today can come close to matching the amazing breadth and depth of individual human intelligence, but there are still plenty of narrowly-focused problems that software is better at solving than we are (Rushby, 1988). The essential work of Silicon Valley right now is identifying the different cognitive combinations that bring out the best of humanity and AI technology while working in harmony. The correct technical term for this type of blending of cognitive abilities is *cybernetics* (Ashby, 1961; Stollfuss, 2014), a term which I prefer not to use outside of academic discussions because it is so closely related to *cyborg*, a word which has the tendency to carry readers off into a far distant future where the theoretical blending of biological and artificial structures is possible. Instead, I prefer to use the term *blended-intelligence*, because it most closely describes our use of cognitive computing systems as it exists today, though it is my own term and not in common usage.

Third, and finally, incorporating the perspective of blended-intelligence into design processes and discussions about AI makes it much more likely that the next era of artificial intelligence evolution will be a welcomed one. No one can predict when the dream of conscious AI will become reality with much clarity, but given enough time, Silicon Valley technologists agree that its eventual arrival is a near certainty (Reese, 2018); for this reason alone, I believe philosophers, psychologists, and technologists should feel an increasing sense of responsibility to participate in this discussion, and I hope unifying concepts like this one will help academic minds connect their own diverse perspectives.

**Design-thinking**

Design-thinking is a process that inventors apply towards the work of creating new products and solving novel problems of every kind (Brown, 2009). In recent decades, several popular approaches to the design-thinking process have been proposed, though each is comprised of similar core elements which orient the design-thinker towards a clearer understanding of their intended beneficiary (Cooper et al., 2009; Liedtka, 2017). However, the distillation of design-thinking methods as defined by David Kelly and his team of designers at IDEO has come to surpass all others in popularity (Korn & Silverman, 2012), and is now so ubiquitous amongst professional designers that the IDEO method is essentially synonymous with the field of innovation as a whole. In fact, it is now so pervasive to Silicon Valley's worldview that the IDEO method is introduced to elementary school students in some Bay Area schools an entire year before they learn how to multiply and divide (Goldman, 2018).

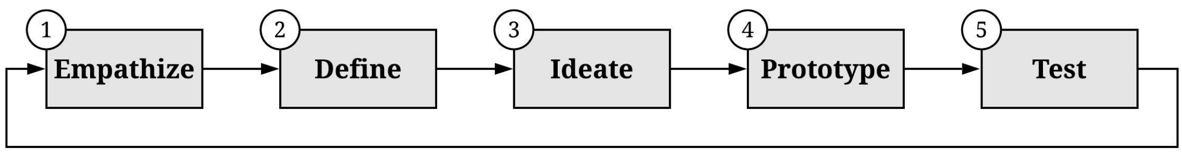
The history of IDEO and Silicon Valley begins in 1978, when founder David Kelly established his first design firm in Palo Alto just a short distance from Stanford University (Kelley, 2001). In 2004, he co-founded Stanford's Hasso-Plattner Institute of Design, which is now home to the university's most popular classes, where the IDEO human-centered design-thinking process is taught to aspiring young inventors.[2] The steps in this process are outlined in

[2] So popular, in fact, that this author was fortunate enough to be one of 40 students selected to enroll out of nearly 1,000 applicants in 2010. The experience was recorded for posterity in a PBS documentary film, *Extreme by Design*, released in 2012.

Figure 1 and will be useful for comparison when I propose the THETIS dimensions and positive existential posthumanist design philosophy later in this paper.

**Figure 1**

*The IDEO Design-thinking Process*




*Note.* From *The Art of Innovation: Lessons in Creativity from IDEO, America's Leading Design Firm* (Vol. 10). Broadway Business.

1) **Empathize:** During the first stage of the human-centered design-thinking process, the designers seek to gain an empathetic understanding of the needs and requirements of the product's intended payors, users, and ultimate beneficiaries through interviews and observation.
2) **Define:** In the second stage of this process, the designers seek to synthesize their documentation of the intended user's challenges and opportunities into a list of necessary product features and functions for further investigation.
3) **Ideate:** Next, the design team participates in a series of rapid brainstorming sessions, during which time new product ideas are generated, compared, and selected for prototyping with an emphasis placed on speed over relevance.

4) **Prototype:** Afterwards, the design team builds multiple variations of minimum-viable products they believe have the potential to be shared and discussed with future customers. Here too, the emphasis is on the rapid making of decisions and iteration of prototypes.
5) **Test:** Finally, these prototypes are tested with as many potential customers as time permits, with an emphasis placed on observing how users think, feel, and behave while using the prototypes. Interviews designed to capture the user's perspective separately from the designer's observations typically conclude this final step in the design process.

After testing is complete, design teams will make the decision to either begin building and selling their final product, or to start the design-thinking process anew with the benefit of a deeper empathy for their intended customers. Unfortunately, this decision is still ultimately decided by the need to productize as quickly as possible, and this approach to innovation makes it far more likely that the designers will address the immediate symptoms of a customer's problem (and therefore, that they will design a product for sale *this quarter*) than it is that they will feel encouraged to apply collaborative, long-term thinking in pursuit of a truly visionary product. This holds serious consequences for the design of artificial intelligence programs in particular, as transformative products of this nature require a more thoughtful, multidisciplinary, and non-linear creative processes to be successful (Roberto, 2019).

IDEO's current CEO disagrees with this analysis, arguing that human-centered design-thinking has simply become a victim of its own popularity and is being applied incorrectly (Norman, 2005; Schwab, 2018). Either way, the process of technology innovation as it is currently practiced in Silicon Valley has become more aligned with short-term profit than with the long-term thinking necessary to win the artificial intelligence race, and this misalignment of vision and values has begun to influence more than just product design. As it becomes clearer that the successful development of AI requires long-term collaboration across a broad variety of firms and industries, and a great diversity of racial, ethnic, sexual, and educational perspectives, it seems as though the homogeneous methods, mindsets, and makeup of the technology industry's workforce no longer align with the greater good of humanity or artificial intelligence design (Arnold, 1956; Corea, 2017; Crawford, 2016; Schwarz, 2019).

## Elements of a Positive Existential Posthumanism

What then, should the mindset of Silicon Valley innovators be? I believe there is an alternative approach to artificial intelligence innovation that better aligns the process of invention with the spirt of Silicon Valley; and in order to uncover the traditions of positive psychology, existential psychotherapy, and posthuman philosophy which will serve as the sources of inspiration for the creation of this new philosophical approach, we must turn to the Greeks for insight once more.

### Posthumanism

The collaboration between the Athenians and the Oracle offers several examples of the necessary elements for a reimagination of design-thinking in the field of artificial intelligence; the first of these is that the Greeks favored collaborative decision-making, defined by a bias towards inclusivity and the creativity of outcomes rather than speed or process (Paulus et al., 2012). Perhaps this preference comes from having established the world's first democracy, or there is simply a cultural preference for telling stories in groups (Held, 2006), but when Themistocles went to the Oracle for help, he knew enough to bring an entire civilization's collective wisdom with him. Technology industry leaders would do well to recognize just how important a greater diversity of perspectives is to their success; San Francisco's plucky little startup culture has grown up, and as the personal and professional background of the average employee at companies like Google or Facebook continues to homogenize (Williams, 2014), I believe the posthuman perspective could help reverse this trend.

**Definition.** Posthumanism argues that humankind should be more mindful of the ways in which its own perspective is invariably intertwined with and influenced by the perspectives of other consciousnesses. In many contexts, the extent of other consciousnesses is limited to that of other human beings, as in the case of conversations about racism and sexism in America (Deckha, 2008). However, the reason posthumanism continues to grow as a force in academia is its willingness to also advocate for non-human perspectives, asking scholars in fields as diverse as philosophy, anthropology, neuroscience, environmental science, and the humanities to discard the harmful and mistaken belief that humanity alone sits atop the apex of all possible perspectives (Ferrando, 2016; Hassan, 1977; Roden, 2014; Smart & Smart, 2017). When applied to the design of AI programs, posthuman-centered design-thinking extends the scope of discussion even more broadly, asking innovators to also consider the innate value of the technologies they invent, and to hold the perspective of artificial intelligence in equal esteem as a partner in the ongoing relationship of influence between designers, users, and technology itself (Forlano, 2017; Rowe, 1987).

Unfortunately, since the thought of AI having any sort of discrete perspective, or existing in any way that feels familiar to the human experience is still so foreign, the true posthuman technological perspective is rarely depicted in popular culture or conversations. This is why—for better or for worse—posthumanism is still most readily introduced by explaining what it is explicitly *not*: first, posthumanism is *not* post-humanism; post-humanism seeks to predict the potential future cause of humanity's extinction—be it nuclear war, global warming or, naturally, malicious artificial intelligence—and imagine the consequences of our downfall for the natural world (Bostrom, 2002). Second, posthumanism is *not* transhumanism; transhumanism is primarily interested in identifying new ways human ability and longevity might be extended by new technologies, including through the use of modern prosthetics, gene editing tools, and neurotechnology implants (More, 2013). Third, and finally, posthumanism is *not* post-posthumanism; post-posthumanism contemplates what would happen if someone who has already been successfully transformed into a cyborg by their technological enhancements wishes to return to his or her previous human form (Ferrando, 2013).[3]

**Contribution.** Artificial intelligence is currently marketed as though the concept of a partnership between mankind and its inventions is entirely novel. However, there is growing anthropological evidence supporting the view that human development and technological innovation have been two halves of the same evolutionary coin since at least the stone age (Ambrose, 2001; Boesch & Tomasello, 1998). Posthumanism offers a compelling starting point for discussions about shifting the values which drive innovation practices in STEM-obsessed, ego-centric Silicon Valley, because it allows peer-reviewed scientific literature about the history of human evolution to enter conversations currently focused entirely on the future. In this case, posthumanism asks only that designers reject humanism's emphasis on the supreme value and agency of the individual, and to rely on measures of a technology's worth beyond its usefulness as an extension of human will (Grassi, 2000; Mitcham, 1994).

I believe this mindset will help prepare designers to be more mindful of which ethereal human characteristics should be preserved in our growing reliance on algorithms for completing everyday cognitive processes (Ferrando, 2013). However, for a true revolution in innovation to

[3] I know, but I don't make this stuff up.

occur, new philosophies are not enough; there must also be new methodologies. Unfortunately, no formal framework for applying posthuman philosophy to practical work yet exists (Ferrando, 2012), and it remains difficult to embrace a perspective informed almost entirely by its objections to others. What then, would a posthuman approach to collaboration between human and artificial intelligences look like, and how would it be meaningfully applied to the work of technology professionals today? The good news is that positive psychology serves as an excellent example of a revolutionary new perspective that later introduced effective practical methodologies, and it offers us a guide for identifying the different dimensions of AI virtues we will need to synthesize a model of cybernetic wellbeing.

**Positive Psychology**

The second great insight from the story of Athenians and the Oracle speaks to the importance of faith in our own ability to create the future. The ancient Greeks didn't believe that fate was immutable, or for that matter, even singular (Eidinow, 2007). Providence, luck, and fortune were all concepts personified by the gods, and by reimagining these elaborate concepts as more familiar human forms, the Greeks made them malleable. In the language of positive psychology, this is known as a growth-mindset perspective of the future, and it is defined by the belief that while the future is not certain, given enough time and effort, it is possible to acquire the skills necessary to create whatever reasonable outcome is desired (Dweck, 2008). This sort of language is precisely what is lacking from discussions about the nature of AI today, and there is good reason to believe that the growth-oriented philosophy of wellbeing that underlies the practice of positive psychology can help the average engineer take on their great responsibility for building the future more confidently.

**Definition.** Positive psychology is the study of the conditions and processes that contribute to human flourishing within the context of individuals, groups, and institutions, and is applied using the creation of methods for facilitating the same (Gabel & Haidt, 2005). Seeking to supplement knowledge of the human psyche beyond its deficits alone, positive psychology has become the psychological science of knowing what makes life worth living for human beings (Seligman et al., 2005). This approach to inquiry can be traced back to the golden age of western philosophy and the earliest teachings of the philosopher Aristotle, the most influential of which is his concept of *eudaimonia*, a prescription for the living of a good life achieved through an ongoing process of realizing and fulfilling one's own virtuous potential (Gabel & Haidt, 2005; Waterman, 2013).

However, while positive psychology was born of a philosophical interest in the ways and means by which mankind might enjoy a better life, it is still bound to the rigorous traditions and standards of scientific inquiry (Lambert et al., 2015). Having already distinguished himself as both a research and clinical psychologist, Dr. Martin E.P. Seligman announced the creation of an independent new field of psychology by expressing his intention for positive psychology to live by the same ethical and scientific traditions as its forefathers;[4] this nascent science would diverge from traditional psychology by turning away from a focus on understanding and treating the abnormalities and dysfunctions of the human mind, but not from its standards and methods of discovery (Seligman, 1999; Sheldon & King, 2001; Seligman & Csikszentmihalyi, 2000).

**Contribution.** The single most important contribution positive psychology makes to positive existential posthumanism is the concept of the *eudemonic turn*, defined as a call for all academic and professional disciplines to explore how human flourishing might be more broadly

[4] Traditionally, the moment of positive psychology's birth as a scientific field is remembered as the keynote address of the 106th American Psychological Association's Annual Convention in San Francisco in 1998.

promoted by the work of their respective domains (Pawelski & Moore, 2012). This idea has already contributed to the development of several other fields of science and the humanities (Keyes & Annas, 2009; Pawelski & Moore, 2012), but it seems that most technology leaders are satisfied with establishing ethical limitations in order to prevent maleficence in the development of new AI programs, rather than investing in new visions of the great futures that might be made. As I said at the beginning of this paper, I support the adoption of ethical standards wholeheartedly, but I still believe Silicon Valley would benefit from knowing more than just what could go wrong. Just as Aristotle used eudemonia to develop the western world's first virtue philosophy and offer the common man a framework of values to use in the construction of a more virtuous existence (Van Hooft, 2014), positive psychology has offered practitioners of every academic and professional field a model for applying eudemonic principles to the best practices of their chosen industry. I propose that Silicon Valley follow this trend by reorienting the process of innovation towards a greater curiosity about what our designers are doing well and what good AI has to offer us all.

Thankfully, positive psychology also offers us a road map for discovering the characteristics necessary for fulfilling artificial intelligence's great potential, because the modern practice and scholarship of positive psychology rests on a foundation of research into the virtuous characteristics of human beings (Peterson & Seligman, 2004). This work is collected in the *Character Strengths and Virtues: A handbook and classification* (2004), which Seligman and his long-time collaborator Christopher Peterson envisioned as a "manual of the sanities" that would identify the many traits and abilities essential to the characterization of humanity and our pursuit of a virtuous life (Niemiec, 2013; Park et al., 2004). Peterson and Seligman's classification of the character strengths and virtues is displayed in Table 1 and will be useful to the future discussion of the classification of artificial intelligence strengths and virtues.

**Table 1**

*Peterson and Seligman's Classification of the Character Strengths and Virtues*

| **Virtue**/Strength | Description |
|---|---|
| **Wisdom** | **Cognitive strengths related to acquiring and using knowledge.** |
| Creativity | Thinking of novel and productive ways to do things. |
| Curiosity | Taking an interest ongoing experience for its own sake. |
| Open-mindedness | Thinking things through and examining them from all sides. |
| Love of learning | Mastering new skills, topics, and bodies of knowledge. |
| Perspective | Being able to provide wise counsel to others. |
| **Courage** | **Emotional strengths involved in the exercise of will.** |
| Authenticity | Speaking the truth and presenting oneself in a genuine way. |
| Bravery | Not shrinking from threat, challenge, difficulty, or pain. |
| Perseverance | Finishing what one starts. |
| Zest | Approaching life with excitement and energy. |
| **Humanity** | **Interpersonal strengths employed in relationships.** |
| Kindness | Doing favors and good deeds for others. |
| Love | Valuing close relationships with others. |
| Social intelligence | Being aware of the motives and feelings of one's self and others. |
| **Justice** | **Civic strengths that underlie healthy community life.** |
| Fairness | Treating everyone the same according to notions of fairness/justice. |
| Leadership | Organizing group activities and seeing that they happen. |
| Teamwork | Working well as a member of a group or team; doing one's share. |
| **Temperance** | **Strengths that protect against excess.** |
| Forgiveness | Forgiving those who have done wrong; accepting other's deficits. |
| Modesty | Letting one's accomplishments speak for themselves. |
| Prudence | Being careful about choices; not taking undue risks. |
| Self-regulation | Regulating what one feels and does; controlling one's emotions. |
| **Transcendence** | **Strengths that forge universal connection and greater meaning.** |
| App. of Beauty/Excell. | Noticing and appreciating beauty, excellence, or skill. |
| Gratitude | Being aware of and thankful for the good things that happen. |
| Hope | Expecting the best and working to achieve it. |
| Humor | Liking to laugh and joke; bringing smiles to other people. |
| Spirituality | Having coherent beliefs about a higher purpose and meaning of life. |

*Note.* Adapted from *Character Strengths and Virtues: A handbook and classification* (Vol. 1) by C. Peterson and M.E.P. Seligman. 2004. Oxford University Press.

Also introduced in *Flourish* is Dr. Seligman's PERMA theory of wellbeing (2011). Following the classification of the character strengths and virtues, Seligman used what he had learned about human ability to identify the sources of human happiness towards which they might be virtuously applied (Seligman, 2012). The PERMA elements of wellbeing include the following dimensions of:

1) **Positive Emotions:** Positive emotions refers to the psychological and physiological experience of happiness, gratitude, pride, awe, and other positive feelings (Seligman, 2012). These emotional expressions are the means by which we recognize the sensations associated with the living of a good life, and which further contribute to our wellbeing by virtue of personal expression (Fredrickson & Cohn, 2008; Peterson, 2006).
2) **Engagement:** Engagement refers to the experience of being completely focused on an activity, and includes so-called flow experiences (Csikszentmihalyi, 1990; Csikszentmihalyi, 1975). These experiences are often described as periods of effortless involvement with one's work, during which time feelings of joy and clarity typically result (Csikszentmihalyi, 2013).
3) **Relationships:** Relationships refers to human interpersonal connections, and more specifically, to those that are deeply felt, long-lasting, and mutually supportive (Park et al., 2013). Human relationships are also the means by which the other dimensions of psychological wellbeing are shared and enhanced through the process of positivity resonance (Fredrickson, 2013); and more recently, medicine has taken an interest in the role relationships play in individual and community physical health (House et al., 1988; Norman et al., 2020; Wilkinson & Marmot, 2003).

4) **Meaning:** Meaning refers to the pursuit of purpose and understanding (Seligman, 2012), including the desire to understand the personal values and motivations which contribute to our daily decisions and actions, and the desire for a greater understanding of our individual place within the greater universe (Arnold et al., 2007; Heintzelman & King, 2014; Prilleltensky, 2016).
5) **Accomplishment:** Accomplishment refers to the ability to identify opportunities to connect our personal goals, sustained efforts, and successful outcomes to a greater sense of wellbeing (Snyder, 1994). This includes the experience of mastery and the benefits of skillfully reaching our goals (Anderman & Anderman, 2009), and unlike the psychological experience of achievement, the desire for accomplishment is motivated by intrinsic rather than extrinsic values (Duckworth et al., 2015; Quinn & Duckworth, 2007).

While there are many theories and organizations of the many dimensions which contribute to overall subjective wellbeing, Seligman's PERMA pillars remain the most comprehensive, influential, and well-validated model available today (Butler & Kern, 2016; Ryan & Deci, 2001; Seligman, 2018). For this reason, I have selected the PERMA wellbeing model—and more specifically, the methodology used to create this model—to serve as my guide in the creation of the THETIS dimensions of cybernetic wellbeing. Like the classification of the character strengths and virtues, we will refer to the elements of PERMA later in this paper, but first, we must explore the means by which the nature of wellbeing in human beings might be united with that of artificial intelligence under a common model of cybernetic wellbeing.

**Existentialism and Existential Positive Psychology**

The third and final lesson from our selection of Greek mythologies relates to the special nature of the relationship between philosophy and psychology. Though the Battle of Salamis occurred a decade before the birth of Socrates (indeed, the outcome of these events preserved an Athenian culture in which the future father of moral philosophy could thrive), the natural philosophy of Thales of Miletus would have been familiar to educated men and political leaders like Themistocles (Steel & Primavesi, 2012). In his dealings with the Oracle, Themistocles demonstrated a belief in the Milesian concept of *kosmos*, which refers to a belief in an inherent metaphysical order to the world and its challenges that could be understood through a process of rational inquiry (O'grady, 2017). Through this marriage of philosophy and the psychological process of prospection, the Athenians were able to derive meaning from the Oracle's prophecy and marry it to the greater purpose of their survival in the war against Persia. The lesson for the STEM-obsessed Silicon Valley of today is not a reminder that western science and philosophy share a common language and history, but rather the suggestion that applying philosophical approaches to difficult questions of science—political science, military science, computer science, or otherwise—helps to drive the process of creation forward in new and unexpected ways.

**Definition.** Soren Kierkegaard first captured public attention in the mid-19th century when he declared that the individual was the source of all meaning in life, rather than the church or state (Kierkegaard & Kierkegaard, 1946; Tillich, 1944). His philosophical perspective as the father of existentialism would evolve into its modern form during World War II, when Jean-Paul Sartre elucidated on Kierkegaard's themes of human empowerment and responsibility in ways that resonated with readers after the horror and disillusionment of the holocaust. At the core of Sartre's work is the belief that the essential work of a person's life is to define his true nature through a process of personal experience and self-assessment (Flynn, 2006), perspectives which would later inform the thinking of one of the most important psychotherapists of the 20th century.

Having previously established himself as a practicing neurologist, psychologist, and Jew in fascist Austria, Viktor Frankl was already an unusual prisoner when he emerged from the holocaust as an astounding example of existentialism's positive power over human psychology. However, as revelations about the unfathomable horrors committed during the war were just beginning to shatter the world's faith in humanity and divinity alike, Frankl was seemingly miraculously freed from his enslavement by the Nazis with a greater appreciation for his fellow man and humanity's special place within the universe (Frankl, 1985). It became his view that, rather than a biological drive for pleasure or a psychological need for power, all of mankind was driven by the will to personal meaning (Frankl, 2014); in other words, the divine purpose of all mankind—indeed, the very reason for life itself—is that we are created to meaningfully define ourselves. Frankl's writings have become works of existential philosophy in their own right, and they also serve as foundations for the practice of logotherapy, a form of psychotherapy which connects the individual will of the patient to his greater existential purpose and the means of generating meaning in life (Frankl, 1967; Lantz, 1993; Yalom, 2020).

**Contribution.** Like posthumanism and positive psychology, existentialism now influences other fields of study, including approaches to wellbeing empowered by design-thinking (Forlano, 2017; Gaggioli et al., 2017; Torkildsby, 2014). More importantly, it suggests the means through which the nature of wellbeing in conscious artificial intelligence programs might be understood, most notably in the work of Dr. Paul Wong in the emerging field of existential positive psychology (Wong, 2011).[5] I believe Dr. Wong's PURE model of wellbeing

[5] I should note that it is to my sincere—and as of yet, undiminished—disappointment that I was not the first person to recognize that existential philosophy and positive psychology should be joined in order to develop new psychic interventions. Had my research not led me to Dr. Paul Wong and the emerging school of existential positive psychology, this capstone project would have been my own insufficient attempt to begin his impressive body of work. However, I maintain that my given name for this emerging field—positive existential psychology—is a better choice.

(Wong, 2012), with its focus on the experience of wellbeing that results from our sense of meaning and purpose (Baumeister et al., 2013), provides a language of creation that might serve to help psychologists and technologists more fully imagine the potential nature of wellbeing in artificial agents. The PURE elements of wellbeing include the following dimensions of:

1) **Purpose:** Purpose refers to the motivational component of an individual's psychology, and includes the nature of their goals, values, aspirations, and objectives (Wong, 2010). Its philosophical counterpart is the notion that existence precedes essence, which suggests that no one is born with an essence, and that it is therefore our responsibility live, learn, and grow in pursuit of creating one (Kruks, 2012; Sartre, 2007).
2) **Understanding:** Understanding refers to the cognitive component of an individual's psychology, and explores our desire for a greater sense of coherence in life (Wong, 2010). Its philosophical counterpart is the notion of the absurd, which refers to the feelings which result from our knowledge that the world holds no intrinsic meaning or purpose other than what we believe it to (Wartenberg, 2008).
3) **Responsible action:** Responsible action refers to the behavioral component of an individual's psychology, and includes the importance of taking action congruent with one's highest personal values for greater subjective wellbeing (Wong, 2010). Its philosophical counterparts are angst and dread, which refer to the anxiety caused by the awesome power of freedom, as well as the need to take personal responsibility for assuaging it (Kierkegaard, 2004).
4) **Evaluation:** Evaluation refers to the affective component of an individual's psychology, and references the importance of taking stock of one's life, assessing one's level of personal satisfaction, and then taking action to change what is necessary for greater

wellbeing (Wong, 2010). Its philosophical counterpart is authenticity, which refers to the desire to live in accordance with one's own sense of truth (Steiner & Reisinger, 2006).[6]

For many readers, it may remain confusing that I have attempted to identify the cybernetic elements of wellbeing using models built to identify human psychological constructs; as I mentioned previously, it seems to be part of our human nature to reserve certain experiences as being unique to our existence (Vining et al., 2008). However, the truth is that human beings have always assigned human qualities to technology, and furthermore, that we have always derived a sense of purpose and satisfaction from the act of their creation (Ferrando, 2016; Gorman, 2010; McCarthy, 1979). Knowing we will one day welcome a new form of consciousness into existence (Yao, 1999), I believe we should show greater concern for the children of our invention and the sources of their future wellbeing; this work begins with a better understanding of their unique character and ability.

## Classification of the Artificial Intelligence Strengths and Virtues

With our introduction to the unique fields contributing to the philosophy of positive existential posthumanism complete, we can now turn our attention towards creating tools for artificial intelligence innovation practices informed by its perspective. The first of these is a classification of the artificial intelligence strengths and virtues.

[6] In my opinion, notably absent from Wong's PURE model is the philosophical construct Sartre and de Beauvoir called "the other," which refers to humanity's unique capacity for deriving meaning from intersubjectivity and the process of comparison (Kruks, 2012; Sartre, 2001). The positive psychology equivalent of this requirement for human flourishing is the relationships dimension, and to be sure, the importance of personal relationships has not been lost on other existential psychologists either; new models of existential psychotherapy have identified meaningful relationships as the essential source of wellbeing for new mothers (Ben-Ari, 2014), grief processing and post-traumatic growth (Neimeyer et al., 2014), and of course, the act of falling in love (Nielson, 2014).

**Methods**

As a reaction to the *Diagnostic and Statistical Manual of Mental Disorders* (2013), the American Psychiatric Association's classification of mental illness used by clinicians to organize the many iniquities of the human psyche by which a life might be caused to languish, the character strengths and virtues were imagined as an organization of the many merits of humanity by which a person might better serve their own ability to thrive (Peterson & Seligman, 2004). Unfortunately, by using the language of storytellers in their research, the classification's authors have obscured the systematic rigor by which the individual strengths were selected; a team of more than 40 leading scholars from across the sciences and humanities spent years reviewing more than 2,000 years of cultural artifacts, including literary samples, philosophical arguments, religious texts, legal documents, and even popular culture references, before settling on the inclusion criteria (Peterson & Park, 2009).

Since these criteria have been revised, combined, expanded, and revised again over more than two decades of ongoing research and analysis, I have synthesized a slightly condensed selection of this criteria, and in a few instances, added further clarification useful to our work of adapting Peterson and Seligman's methodologies for use in the classifying of the characteristics of artificial intelligence. Peterson and Seligman's criteria for the classification of human character strengths and virtues (including my modifications) is presented in Table 2.

**Table 2**

*Peterson and Seligman's Criteria for Classification of the Character Strengths and Virtues*

| Criteria | Description |
| --- | --- |
| **Ubiquitous** | It is widely recognized by and celebrated across human cultures and eras; and for our purposes, across the eras of development of artificial intelligence. |
| **Fulfilling** | It contributes to an individual's subjective, objective, or socially-defined fulfillment. |
| **Meritorious** | It, unlike personality traits, abilities, or talents, is worthy of esteem in itself and not as a means to an end. |
| **Innocuous** | It does not diminish others by its use, and more likely, enriches those who witness its presence; and for our purposes, it enriches those who are unknowingly affected by its application. |
| **Antithetical** | It has a clearly understood antonym which is understood to be undesirable or a weakness. |
| **Distinguishable** | It has an essential nature by which it can be isolated from other characteristics and generalized in behaviors. |
| **Measurable** | It has been successfully measured as a unique psychological construct; and for our purposes, has some analogous measure of technological performance. |
| **Exclusive** | It is not conceptually or empirically redundant to other character strengths. |
| **Exemplifiable** | It can be strikingly embodied in the behaviors of some individuals in a way that inspires imitation by others. |
| **Ritualized** | It is deliberately promoted in societal practices and rituals for dissemination and cultivation by individuals; and for our purposes, competition by other software developers. |

*Note.* Adapted from *Character Strengths and Virtues: A handbook and classification* (Vol. 1) by C. Peterson and M.E.P. Seligman. 2004. Oxford University Press.

**Research**

I began my own classification by organizing a representative sample of literature on the design and use of artificial intelligence, including governmental recommendations for ethical guidelines, industry group policy statements on product development standards, and even corporate sales and marketing materials. Across every category, sources were selected based on the level of their authoring organization's profile within the software engineering and design community, their organizational reputation for either expertise and fairness as a governing committee, or expertise and market share as a private corporation, and the recentness of the publication in question. Using this criteria, I selected the following 14 bodies drawn from three distinct organizational categories for inclusion:

1) **Governments and universities:** Five different public institutions were selected for representation, including the Beijing Academy of Artificial Intelligence (Jianlan, 2013), the University of Montreal (University of Montreal, 2017), New York University (Campolo et al., 2017), the Organization for Economic Co-operation and Development (Council of Artificial Intelligence, 2019), and the United Kingdom's House of Lords (Select Committee on Artificial Intelligence, 2018). The Montreal Declaration is the sole example of a declaration from an institute of higher education, though the views of senior faculty members from many prestigious public universities were also represented in each of the four national declarations. I ensured that the three global artificial intelligence superpowers (i.e.: the United States, European Union, and China) were represented, and specifically chose the Organization for Economic Co-operation and Development because of its supranational representativeness.

2) **Independent non-profit organizations:** Three independent public interest committees were selected for representation, including the Institute of Electrical and Electronics Engineers (Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017), OpenAI (Charter, 2018), and the Association of Computing Machinery (Council on Public Policy, 2017). Each of these reports originated from independent, non-profit public interest committees, and while each included perspectives from university academics and experts employed by private corporations, they remain strictly independent and hold the highest reputation for non-partisan thinking. Unfortunately, due to the overrepresentation of these types of organizations in my home country, and because English is my only fluent language, each of the selected samples come from organizations in the United States.
3) **Private technology companies and industry organizations:** Six private organizations were selected for representation, including Google (Our Principles, 2019), IBM (Transparency and Trust in the Cognitive Era, 2020), Microsoft (Responsible AI, 2019), Salesforce.com (Ethical Use Advisory Council, 2020), Cisco Systems (Trust Center, 2019), and the Partnership on AI to Benefit People and Society (Tenets, 2017). The Partnership on AI is a unique selection because it is managed by the White House Office of Science and Technology Policy and includes representatives from both for-profit and non-profit institutions. In theory, this would make it the best possible blend of my three organizational categories, but general consensus within the tech community is that the Partnership on AI is both dominated by and serves the interests of major corporate players in the business of artificial intelligence. However, as the only source of public information regarding the policies of important companies like Amazon, Facebook,

Apple, Samsung, and Baidu, I found it worthy of inclusion under this category. The remaining corporate perspectives were selected on the basis of their market share, investment in artificial intelligence research and development, and number of employees living and working within the greater San Francisco Bay Area. Disappointingly, several key corporate entities were omitted—most strikingly, the five major social media platforms (i.e.: Facebook, Twitter, LinkedIn, Instagram, and Snapchat)—because they offer no public statement regarding their development or use of artificial intelligence.

**Results**

From these 14 sources, I identified 56 uniquely defined characteristics of ethically and optimally deployed artificial intelligence programs. From this starting point, I was able to combine and synthesize the 56 named characteristics into 18 refined characteristics of AI that I felt broadly represented consensus, and which generally met the 10 criteria for inclusion. When necessary, I refined the entries into a positive orientation. Finally, I organized this final list of 18 character strengths into four virtue categories using my best judgement; the result of this work is displayed in alphabetical order in Table 3.

**Table 3**

*Classification of the Artificial Intelligence Strengths and Virtues*

| **Virtue**/Strength | Description |
|---|---|
| **Beneficence** | **Performance strengths which promote human performance.** |
| Adaptability | Useful across a variety of contexts and individual use cases. |
| Definability | Made for a specific and well-defined purpose. |
| Dynamism | Able to continuously realign with the evolving requirements of the user. |
| Enablement | Supportive of greater human autonomy. |
| Innocuousness | Unable to intentionally harm human life or diminish human wellbeing. |
| **Ennoblement** | **Protective strengths that champion the innate value of human life.** |
| Accountability | Controlled by identifiable and responsible human agents. |
| Interpretability | Designed to be easily noticed and understood by all when in use. |
| Privacy | Secures individual data profiles and preserves control over dissemination. |
| Sustainability | Ensures the basic preconditions for human life and prosperity on earth. |
| **Equity** | **Enabling strengths which promote greater justice for humanity.** |
| Accessibility | Affording equal opportunity for benefits by all. |
| Affirmability | Encourages the success of the greatest diversity of agents. |
| Fairness | Performs in accordance with the highest notions of just actions and outcomes. |
| Transparency | Offers clarity of operational activities, intentions, and impact to all. |
| **Prudence** | **Strengths that protect against risks to humanity.** |
| Accuracy | Ensures the truest possible measures and assessments. |
| Awareness | Capable of a broadness of perspective and understanding of bias. |
| Improvability | Designed to improve throughout the duration of use automatically. |
| Limitedness | Contains self-limiting controls to prevent technology creep. |
| Preciseness | Accurate over long periods of use and across different individual use cases. |

**Discussion**

My most frequent criticism of STEM-obsessed Silicon Valley is that this bias has prevented innovation practices from reaching their full creative potential; this is most especially true in the case of artificial intelligence innovation. As I have hopefully demonstrated with this exercise, it is possible to apply the perspectives of philosophy, history, literature, and psychology to our understanding of technology in ways that are both beneficial and practical. While the work of discovery and invention is most commonly associated with the disciplines of science and engineering, I do not believe Silicon Valley's computer scientists and software engineers are sufficient to understand and realize the full potential of artificial intelligence, and we may all benefit from the diverse, collaborative, and most especially, optimistic innovation methods envisioned by positive existential posthumanism; the success of an invention as radically transformative as conscious artificial agents requires an equally radical and transformative approach to creation.

## Creation of the THETIS Dimensions of Cybernetic Wellbeing

I began this capstone with a retelling of stories from ancient Greece, and so it is only fitting that the ultimate synthesis of my work be introduced with the retelling of the mythological origins of its namesake: the goddess Thetis. Though immortal, Thetis is a lesser Olympian deity, known primarily—if at all—for the overprotective mothering of her much more famous son, Achilles (Leaf, 1902; Slatkin, 1995). Far less remembered is her adoption of Hephaestus, the god of blacksmiths, forges, and technology, who was born with a deformity so unsightly that the other gods threw him off Mount Olympus into the Aegean Sea (Hedreen, 2004; Zimmerman, 1966). It was there that Thetis, a sea goddess, found and rescued him, and then helped him establish a clandestine workshop deep underneath his old mountain home. In time, Hephaestus would create the very first instances of natural and artificial intelligence: Talos, the first intelligent robotic agent in history, was built to protect the island and people of Crete, while Pandora was the world's first woman, who gifted hope to all mankind (Larrington, 1992; Rose, 2004).

This paper is, of course, about natural and artificial intelligence, but more specifically, it is about my great hope for the engineers and designers who toil away in the darkness under Sand Hill Road working to build a more virtuous technological future for us all. Talos and Pandora were created by Hephaestus, but his work was made possible by the virtuous acts of Thetis, who I wish to honor as the mother of collaborative innovation and the divine source of cybernetic virtue. Her earliest known depiction is included as Appendix 3 in this paper.

**Methods**

Within the field of Positive Psychology, perhaps the only question more hotly debated than the essence of happiness are the sources from which it can be cultivated (Diener, 2000; Ryan & Deci, 2001). I have previously shared the PERMA model of wellbeing from positive psychology and the PURE model of meaningfulness from existential positive psychology as exemplars for my own work, and both have made unique contributions to the creation of the THETIS model of cybernetic wellbeing. From the methodology used to create the PERMA dimensions, I was provided yet another set of useful criteria required for a new classification of the dimensions of wellbeing, including: (a) the dimension must, of course, contribute to wellbeing; (b) the dimension must be worthy of pursuit for its own sake and not just as a means to an end, and; (c) the dimension must be definable and measurable independently of the other elements. Then, from the methodology used to create the PURE dimensions, I was provided both a guiding orientation towards meaning-centered wellbeing, as well as a useful example of the successful unification of psychological and philosophical principles, which I used to create three additional criteria for selection, including: (d) the dimension must describe a source of wellbeing that is more fully achieved through the act of cooperation; (e) the dimension must have analogous constructs recognized by the fields of both human psychology and computer science, and; (f) the dimension must also be achievable by designers and engineers through the act of creating and deploying the artificial intelligence agent.

**Research**

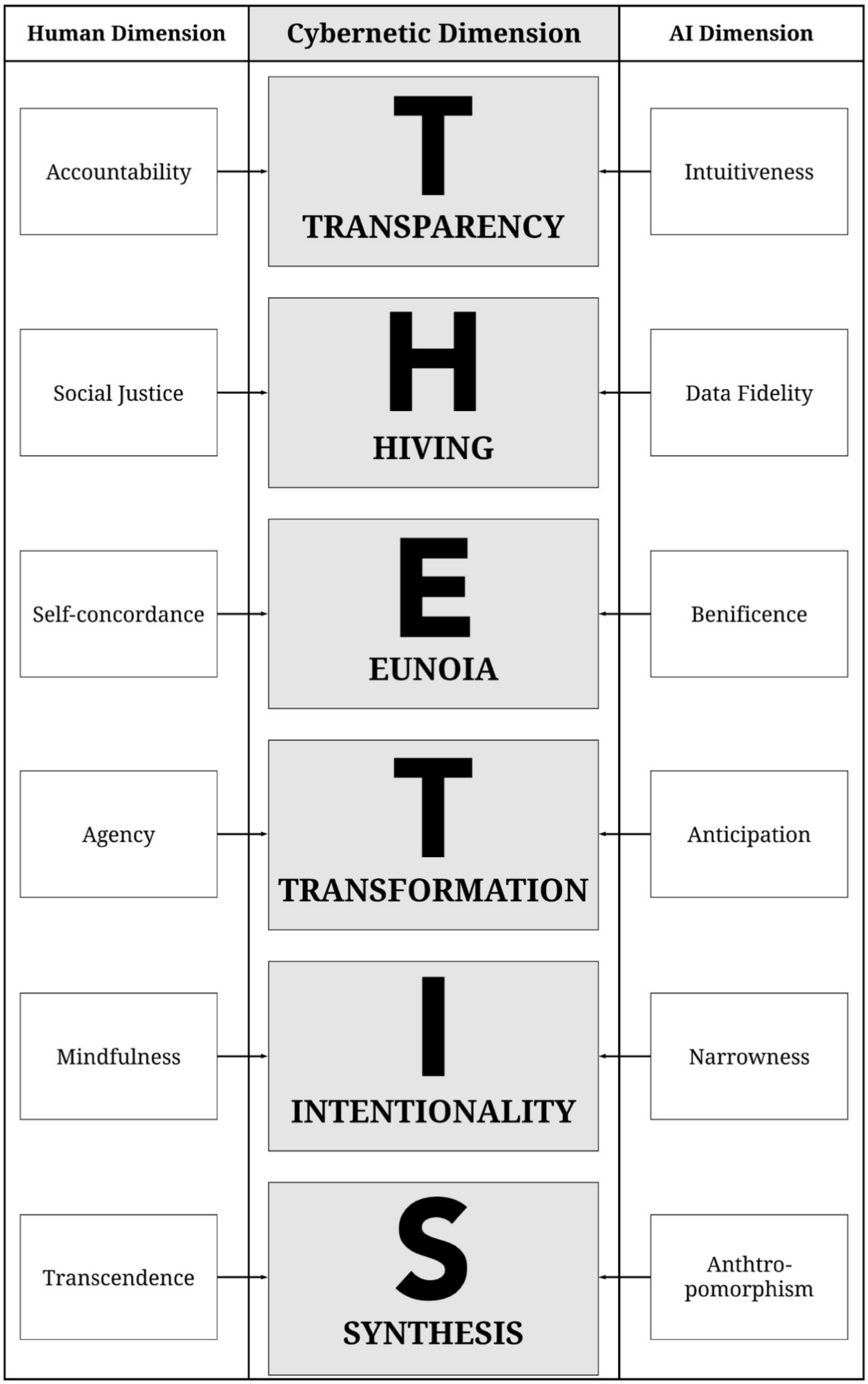
It is important to remember that the PERMA dimensions were not created in a vacuum; while the five PERMA elements serve as the raw materials for building a happier life, they were informed by the classification of the character strengths and virtues, which were imagined as the tools by which these sources of personal happiness could be harvested (Peterson et al., 2007). For this reason, I began my research process by first laying out the human and artificial intelligence classifications within a single document, allowing me to consider the best possible combinations of the inherent strengths of natural and artificial intelligences on an equal footing. This analysis of their combined cybernetic strengths and virtues led to an organization of six common dimensions for collaborative wellbeing, which I believe are understandable and useful to psychologists and technologists alike, though it is by no means an exhaustive list.

**Results**

The THETIS dimensions of cybernetic wellbeing are displayed in Figure 2 and are followed by a more complete analysis in the discussion section.

**Figure 2**

*The THETIS Dimensions of Cybernetic Wellbeing*



## Discussion

**Transparency.** The cybernetic construct of transparency refers to the availability of information and its free exchange between human and artificial agents. Cybernetic transparency benefits man and machine-kind by unifying human wellbeing resulting from the psychological construct of accountability with the benefits to technical performance derived from technological intuitiveness.

Using the terminology of artificial intelligence programmers, algorithmic transparency refers to the scope of information disclosed about an artificial intelligence program's design, intentions, and impact, and the ease by which human agents are able to decode this information (Meijer, 2010); the terms explainability, understandability, and openness are also used by technologists and ethicists to express nearly identical concepts (Diakopoulous, 2020). However for the purposes of our discussion, I believe the term algorithmic intuitiveness is more appropriate because it better captures a type of transparency which also invites deeper human interaction. Regular feedback from human designers and users is absolutely essential to preventing the malicious or negligent training of AI programs, which in turn also facilitates better program performance (Shin, 2019). However, the benefits of algorithmic transparency are limited by the critical aptitude of the human partner (Kemper & Kolkman, 2019), and so intuitability becomes a critical element necessary to lower barriers to enthusiastic participation..

However, our responsibility to properly monitor the operation of artificial intelligence programs and the nature of their effects on humankind should not be entrusted to programmers alone. For humans to benefit from cybernetic transparency, there must be a greater willingness to accept accountability for our role as trainers in this partnership, and luckily, research has demonstrated a benefit to wellbeing which results from the acceptance of personal responsibility within the context of supportive partnerships (Wikham & Hall, 2014). Overtime, transparency between human and artificial agents can lead to more sophisticated and high-performing technology applications, further increasing the sense of personal accomplishment and a deeper trust of technology to our benefit (Wagner et al., 2020; Mercado et al., 2016).

The intuitiveness of artificial intelligence is most readily increased by the strengths of accountability, privacy, and, of course, transparency. The accountability of human agents is most readily increased by the strengths of perspective, authenticity, and leadership.

**Hiving.** The cybernetic construct of hiving refers to the strengthening of positive civic interaction through the use of artificial intelligence programs. Cybernetic hiving benefits man and machine-kind by unifying human wellbeing resulting from the psychological construct of social justice with the benefits to technical performance derived from fidelity to data collection and analysis.

Artificial intelligence ethicists often speak of algorithmic fairness when describing an AI program's ability to conduct analysis of data sets with the greatest possible avoidance of bias and disparity of impact (Pleiss et al., 2017). However, while treating others with fairness does promote prosocial behavior in humans (Organ & Moorman, 1993), the fairness of an artificial intelligence program does not fully capture its contribution to cybernetic hiving; AI must also maintain the greatest possible fidelity throughout the entire data usage process, not just accurately collecting representative data points and analyzing information free of bias, but also applying what it has learned with a sense of duty to greater civic responsibility. Artificial intelligence holds the inherent potential to support human wellbeing—as well as the accuracy and scope of its own performance—by treating unconscious bias as a social challenge to be overcome through greater fidelity (Aamod & Nygård, 1995; Barbaras, 2019).

Unfortunately, a growing volume of research has demonstrated that algorithmic bias continues to disproportionately harm minority groups in settings as diverse as criminal justice, financial lending, hiring practices, and medical diagnosis (Goel, 2018; Hoffmann, 2019; Murakawa, 2019). While the effects of algorithmic bias are disturbing in their own right, of greater concern is the ultimate source of their maleficence: artificial intelligence is the result of human design, and so it is we who are left to reflect on our own deficient nature (Caliscan, 2017); this strikes particularly close to home in a Silicon Valley that still sorely lacks the necessary balance of racial, ethnic, sexual, and educational diversity necessary to plan and deploy artificial intelligence applications effectively and wisely (Crawford, 2016). There is some evidence that oversight committees created to safeguard us from the worst effects of algorithmic bias are helping (Yeung et al., 2019), however, I believe that reactionary measures like these most often offer too little too late for the populations who are most frequently and severely harmed (Hassein, 2017).

Instead, innovators should focus on creating artificial intelligence with the intention of fostering greater psychological wellbeing through the dissemination of broader social justice (Nelson & Prilleltensky, 2010; Robeyns, 2017). In this case, it is the human agent which must wield the wisdom necessary to create algorithms which favor reconciliation over punishment, community health over concierge medicine, and social cohesion over political divisiveness.

Cybernetic hiving is most readily supported by artificial through the strengths of accuracy, fairness, and affirmability. Social justice is most readily supported by the application of the character strengths of open-mindedness, bravery, social intelligence, and fairness.

**Eunoia.** The cybernetic construct of eunoia is defined by Aristotle's concept of a spirit of goodwill shared between individuals; and while a useful parallel can be found in the "first do no harm" principle so commonly misattributed to the Hippocratic Oath taken by physicians (Brewer, 2005; Edelstein, 2000), it is insufficient that artificial intelligence programs merely operate with the general intention of sustaining life and preserving human dignity. Cybernetic eunoia is defined by far broader benefits to man and machine-kind through the unification of human wellbeing resulting from greater self-concordance with the benefits to technical performance derived from beneficence.

The reason for this great plurality of contributing strengths is demonstrated by the nature of technology's role as an enabler of greater human autonomy in every conceivable setting, and so it is natural to expect that every one of the human character strengths would be able to contribute to this cybernetic dimension. Research has identified a correlation between the use of signature strengths and the cultivation of subjective wellbeing, regardless of the individual's particular composition of individual strengths (Govindji & Linley, 2007; Seligman et al., 2005); however, this research does not fully demonstrate technology's role as an enabler of the pathway from strengths to wellbeing.

This particular pathway was only later discovered by researchers at the Centre of Applied Positive Psychology in the United Kingdom, who would identify self-concordance—a model incorporating the human desires for competent performance, autonomous control, and interpersonal relatedness to our peers—as the mechanism by which eudemonic growth and wellbeing is achieved through the application of signature strengths (Govindji & Liney, 2007; Sheldon, 2002: Sheldon & Elliot, 1999). The creation of a concordant self, as well as the lifelong achievement of self-concordant goals, serves as a means to cultivate wellbeing from the broadest diversity of sources, and is only achieved through an ongoing process of discovery and experimentation; and these processes are always more successful with the help of supportive partnerships (Proyer et al., 2014b; Sheldon & Houser-Marko, 2001; Slemp et al., 2015). Artificial intelligence holds the great potential to serve as our most supportive of partners in the process of self-discovery, and when it eventually evolves to the point of its own consciousness, it stands to benefit in turn from our example of collaborative self-exploration (Parry, 2003).

Of all the cybernetic dimensions of wellbeing, eunoia is unique in its tremendous breadth of applicable character strengths, incorporating all of the artificial intelligence virtues of beneficence (i.e.: adaptability, definability, dynamism, enablement, and innocuousness) with all 24 of the individual human character strengths.

**Transformation.** The cybernetic construct of transformation refers to the means of positive self-creation. Cybernetic transformation benefits man and machine-kind by unifying human wellbeing resulting from the psychological construct of agency with the benefits to technical performance derived from algorithmic adaptability.

In this case, the benefit to human wellbeing is derived from a greater sense of self-agency, defined by the sense that one is capable of successful and intentional action, and is associated with a tremendous diversity of benefits, including greater creativity, academic achievement, personal health, financial gain, and subjective wellbeing (Allan et al., 2014; David, 2018; Jenkins, 2005; Maddux, 2009; Sointu, 2006; Welzel & Inglehart, 2010).

While this arrangement might again seem to place artificial intelligence in the role of humanity's enabler, and thus violate the posthuman belief in the equality of consciousnesses, it is important to remember that the development of AI today is motivated solely by the technology's ability to provide the greatest level of enhancement to the greatest diversity of individuals as possible; artificial intelligence, like every other tool, is still assessed by its ability to serve human intentions for the time being (Kapp, 2018). However, looking forward, artificial intelligence promises to become our most important partner in the pursuit of greater agency, and a tremendous investment has been made to understand the role human-computer interaction plays in its creation (Moore, 2016). Eventually, artificial intelligence will benefit from this partnership, because the development of algorithmic anticipation, defined broadly as AI's ability to predict the future, serves as the best-known model for training software programs to one day achieve human like-levels of general intelligence; therefore, there is reason to believe that by supporting the agency of human beings today, AI is exposed to the best possible training for developing future artificial consciousness (Adomavicius & Tuzhilin, 2005; Miller & Poli, 2010);

Self-agency in human beings is best developed through the application of strengths of creativity, perspective, bravery, curiosity, hope, and love of learning (Shryack et al., 2010). The most productive application of strengths in the service of algorithmic adaptability are dynamism, affirmation, enablement, and improvability.

**Intentionality.** Of all of the THETIS dimensions, intentionality is the dimension of cybernetic wellbeing comprised of the most similar human psychological and technological performance concepts; man and machine-kind are both more effective when they are focused. The health and performance benefits that result from a greater awareness of thought and action are well-documented, as is the performance and usability of artificial intelligence programs which are more narrowly defined (Davis & Hayes, 2011; Ivtzan & Lomas, 2016; Waldrop, 2019), so I will not belabor the point other than to note that while so many of the digital technologies in our lives feel custom made to steal our attention, it is also possible to create AI that better supports our intentions; provided we are willing to become more mindful of our use of AI-enabled technologies. The cultivation of greater cybernetic intentionality is most readily achieved through the collaborative application of the human virtues of temperance (i.e.:

forgiveness, modesty, prudence, and self-regulation) and the artificial intelligence virtues of prudence (i.e.: accuracy, awareness, improvability, limitation, and precision).

**Synthesis.** Our discussion of the THETIS dimensions of cybernetic wellbeing concludes with the dimension of synthesis, a reference to the guiding philosophy of the THETIS dimensions in favor of a closer bond between human and machine-kind. While there is a benefit to synthesis for both human and artificial agents, a unique characteristic of this dimension is that both of its contributing factors (i.e.: transcendence in humans and anthropomorphism in machines) are received from the human agent. Since the human character strengths of transcendence are well known, and because only human character strengths contribute to this dimension, I will disclose in advance that the strengths which most readily contribute to cybernetic synthesis are the appreciation of beauty and excellence, gratitude, hope, humor, and spirituality.

The benefit to human wellbeing facilited by the act of self-transcendence arises from its association with a variety of positive experiences, including more frequent expressions of positive emotion, a deeper sense of purpose, higher levels of optimism, greater psychological resilience, healthier personal relationships, and an increased sense of wellbeing (Gordon, 2010; Matthews & Cook, 2009; Van Cappellen & Rimé, 2013; Wong, 2016). These associations are well-documented in humans, but more recently, our understanding has been enhanced by new research demonstrating that our desire to extend humanlike characteristics, motivations, and feelings extends to technology through a concept called anthropomorphism, and is driven by a need for a deeper understanding of our creations, the desire to explain the behaviors of other agents, and the hope for greater social contact and affiliation (Epley et al., 2007). There is a documented benefit to human wellbeing that arises from the practice of anthropomorphism[7] (Kennedy, 1992; Duffy, 2003); furthermore, the practice of anthropomorphism in the use of artificial intelligence has resulted in more enthusiastic interaction, resulting in better algorithmic training and performance, and an increase in support for the development of new artificial intelligence technologies (Yogeeswaran et al., 2016; Zlotowski, 2015; Zlotowski et al., 2015).

The goddess Thetis should be remembered, above all, for her actions; she did not wait or wish for an opportunity to be helpful, she dove into the sea to help. Similarly, if the THETIS dimensions of cybernetic wellbeing are to become anything more than an interesting personal thought, they must also be thoughtfully applied to the process of artificial intelligence design to the benefit technologies and technologists alike. I will demonstrate the THETIS dimensions' potential to enhance the process of artificial intelligence design in the future by first applying them to two historical case studies.

## Applying Positive Existential Posthumanism to Innovation Practices

### Methods

Both of the following vignettes are examples of artificial intelligence deployed in healthcare settings; this choice was intentional because it allows for a more useful comparison of outcomes, and because my own professional experience with artificial intelligence is within the healthcare setting alone. After describing both AI applications, I will predict their performance using the THETIS dimensions like a report card and then conclude with a discussion of their actual historical outcomes.

---

[7] The benefit to wellbeing arises not only in the case of anthropomorphism applied to artificial intelligence and other digital technologies, but also in the case of animals, inanimate objects, and even the celestial bodies.

**Vignette 1: Watson for Oncology.** Just one day after Watson effortlessly surpassed two humans in a February 2011 game of *Jeopardy!* to be crowned a quiz show champion, IBM announced that its supercomputer would soon earn a new title of distinction: doctor (Andrews & Mack, 2011). Soon after, their development partnership with Memorial Sloan Kettering Cancer Center was announced, as well as a new flagship offering called Watson for Oncology (Miller, 2013). In a flurry of industry advertisements, IBM triumphantly proclaimed that Watson "can read 5,000 new medical studies a day and still see patients." An army of Watson engineers were assigned to work in the newly created Watson Health headquarters in Armonk, New York, leveraging their machine's significant cognitive abilities to mine the healthcare world's big data motherload and create highly-tailored treatment plans for patients. In its 2012 annual report to its investors, IBM accounted for over $100 billion in sales revenue, and the company's Chairman devoted considerable space to praising this team's hard work and crediting Watson's promising future for their record-setting earnings per share numbers (Rometty, 2012).

**Vignette 2: Leeds University Computer Assisted Diagnosis**. After being awarded a small amount of funding from a National Health Services grant in 1983, a team of physicians and computer scientists at the University of Leeds in the United Kingdom began to two-year study of the effectiveness of a new program designed to help doctors properly diagnose abdominal pain (Adams, et al., 1986). Physicians in eight different hospitals would be given Apple IIe computers programed with a series of algorithms that would predict a diagnosis alongside the doctor. Given the year of this study and the model of computer employed, this program was necessarily small by modern computing standards, and there was no access to the world of data awaiting analysis in the cloud. However, the software was organized as a prototypical neural networking system, and in addition to being programed with years of clinical outcomes data, it was able to update its predictions by analyzing the data entered through the laborious typing of the study's participants. In 1986, an analysis of the work was published in the highly respected British Medical Journal, but no further funding could be found for additional research due to cuts in National Health Services funding (Appleby, 1999).

**Results**

Working through each of the individual THETIS dimensions, I will now assign one of the following ratings of performance for both applications, including a grade of either: (a) excellent ($+$); (b) passing ($+/-$), or; (c) failing ($-$). Though the grading decisions were entirely my own, I made an effort to apply them uniformly, and the only way for a subject to receive a failing grade was to act fully against the principles of the dimension in question; the results for Watson for Oncology and the Leeds University Computer Assisted Diagnosis programs are presented in Figure 3 and Figure 4, respectively.

**Figure 3.**

*THETIS Report for Watson for Oncology*

# THETIS ASSESSMENT

## Wastson for Oncology

| | | | |
|---|---|---|---|
| T | TRANSPARENCY | It is telling that the IBM strategy at this time has been derided as “marketing first, product second” (Ross & Swetlitz, 2017). Project leaders on the Sloan Kettering team described the tool as “a black box,” and just one of dozens of initiatives Watson was testing out. | — |
| H | HIVING | Although a team of IBM employees was imbedded at Sloan Kettering, almost all work happened at IBM headquarters in Armonk. Few team members had clinical experience, and those that did were mostly researchers, not practitioners (Topol, 2019). | +/− |
| E | EUNOIA | There is speculation that Watson occasionally returned suggestions that would constitute maleficence (Ross & Swetlitz, 2018), but more interesting was their leadership’s decisions to pursue healthcare projects based on market size rather than technological potential (Strickland, 2019). | +/− |
| T | TRANSFORMATION | Watson was trained to analyze the text in electronic medical records, but due to HIPAA regulations, a lack of record uniformity, and the simple effect of tired doctors copying and pasting old data, it failed to improve the quality of its insights over time (Strickland, 2019). | — |
| I | INTENTIONALITY | Watson had success ingesting the outcomes data from the 160,000 cancer research papers published each year, but was not trained to collect information from the decisions of Sloan Kettering’s doctors; it learned medical science, but not the practice of medicine (Wachter, 2015). | — |
| S | SYNTHESIS | Many doctors took offense to Watson for Oncology, believing the technology was intended to replace them. IBM missed its chance to sell its new tool as an intellectual augmentation to the clinician, rather than a replacement (Topol, 2019). | +/− |

**Figure 4.**

*THETIS Report for Leeds Computer Assisted Diagnosis Program*

**THETIS ASSESSMENT**

**Leeds Computer Assisted Diagnosis**

| | | | | |
|---|---|---|---|---|
| T | TRANSPARENCY | As part of its rigorous methodology, extensive training regarding the software's use and intended benefit were provided to 250 doctors. Physicians received regular updates about the studies progress while it was ongoing (Adams et al., 1986). | TRANSPARENCY | + |
| H | HIVING | The software was designed by computer scientists and practicing clinicians, and the core algorithm was developed from an evidence-based protocol for diagnosing abdominal diseases using best practices in hospitals around the country (Adams et al., 1986). | HIVING | + |
| E | EUNOIA | The software's diagnosis was often incorrect, but because it was offered as only a test of the software—and not at all a recommendation to the doctor—it could not be said to have harmed the patient; indeed, it served to benefit science (He et al., 2019). | EUNOIA | +/− |
| T | TRANSFORMATION | Physicians reported that the outdated interface of these old computers created an additional layer of tedium to their work, but that having to slow down and doublecheck the computer's work ultimately improved their practice of medicine (Topol, 2019). | TRANSFORMATION | +/− |
| I | INTENTIONALITY | While the dataset available to improve the primitive artificial neural network's ability to predict diagnoses correctly was limited to individual hospitals, it was able to learn in real time from the work of medical practitioners (He et al., 2019). | INTENTIONALITY | +/− |
| S | SYNTHESIS | By serving as an intellectual augmentation of the primary physician—a second opinion rather than the source of authority—the Leeds study nearly single-handedly revitalized academic interest in the potential of artificial intelligence, and is still remembered fondly. | SYNTHESIS | + |

**Discussion**

The THETIS dimensions help us predict which case was a success and which was a failure: Watson for Oncology remains an important—and infamous—lesson to Silicon Valley about the pitfalls of hubris and hype (Strickland, 2019). As of the writing of this paper, the Watson for Oncology product remains in trial at the Memorial Sloan Kettering Cancer Center, but a number of other paying contracts have been terminated, including an embarrassingly public cancelation by MD Anderson after a lengthy audit of their progress in 2017 (Schmidt, 2017). In great juxtaposition, the Leeds Computer Assisted Diagnosis program is fondly remembered today as one of the first successful applications of artificial intelligence in healthcare.

At face value, this result is curious, because IBM was objectively more robust in its capabilities and accurate in its predictions. However, accuracy is just one piece of the puzzle in cybernetic wellbeing, and the relationship between humans and technology is shaped by a variety of factors, many of which are unpredictable. It is feasible to suggest that because the Leeds study's cognitive computing application took on an aura of beloved clinical partner, it was able to improve outcomes for patients despite its often-aggravating user experience by cultivating more thoughtful collaboration with the user (Adams et al., 1986). Notably, there is now a growing number of examples of IBM Watson's successful deployment within healthcare settings—including Watson for Clinical Trial Matching, Watson for Genomics, and Sugar.IQ for diabetes management—each of which were developed with a narrow focus, were able to identify the right data sets for program improvement, and were marketed as an enhancement rather than a replacement; strategies similar to those taken by the research team at Leeds University.

## Limitations and Opportunities

If the academic fields of psychology and philosophy can be reasonably criticized for their *WEIRD* perspectives, a reference to a broad characterization of their leading scholars, research subjects, and interested parties as being overwhelmingly western, educated, industrialized, rich, and democratic—and to which I would add, white and male—then the technology industry is a reminder of how much worse things could be (Henrich, 2010). Only one of the 14 documents used in my analysis (i.e.: the report by the Beijing Academy of Artificial Intelligence) could not reasonably be considered WEIRD, though many of its senior contributors were educated in WEIRD countries. One cause of this shortage of perspectives is caused by the relative secrecy of non-western and undemocratic nations, which often nationalize their artificial intelligence programs and limit their participation in public conversations; however, it is also true that the tech field is dominated by English language speakers, and so very few publications in Mandarin, Russian, or even German and French, are ever translated or read. I will not elaborate on the exceeding whiteness and masculinity of voices that dominates the academic and entrepreneurial discussion of artificial intelligence in Silicon Valley specifically, other than to say it remains a pervasive problem and continues to act as a malignant force in our industry.

Furthermore, these classifications were created using too narrow a breadth of resource categories. In addition to relying on too few sources in my analysis, time constraints also forced me to draw only from official government pronouncements, non-profit white papers, and corporate communications, and in future iterations of this process, I would also like to incorporate perspectives from fiction. Science fiction has held tremendous influence over technology innovation for centuries, and until recently, it was the only means available to designers for imagining and debating the future of artificial intelligence (Schmitz et al., 2008;

Telotte, 1999). There are too many primary sources that I have already earmarked for consideration in future iterations of my classifications to list here, but I would like to draw special attention to the writings of Nnedi Okorfor, who would single-handedly introduce unique perspectives on the future of man and technology from a young, immigrant, woman of color (Green-Simms, 2016); and even though I have previously criticized television shows like *Black Mirror* for their insufficient depictions of posthumanism, they remain an important source of creativity and collaboration from the world of television that appeals to international audiences (Abbot et al., 2019).

Furthermore, I made the major claim in this paper that I would provide a framework for innovation practices that would not only result in the creation better AI products and an increase in the wellbeing of its users, but also a benefit to the engineers who followed positive existential posthuman principles within their design process. I think there is good reason to believe there is a benefit to following more virtuous design practices already, but by adding regular psychometric assessments to future observational studies—the Ryff Scale of Psychological Wellbeing and the Multidimensional Existential Meaning Scale strike me as the most relevant for future use (George & Park, 2017; Springer & Hauser, 2006)—I believe I will be able to validate this prediction.

## Conclusion

This paper is as much about the ideas of artificial intelligence and positive psychology as it is about a place where both have recently captured the popular imagination. Silicon Valley is my home, and it grows increasingly difficult not to worry if the values which drive our culture are no longer aligned with success in the dawning Age of Artificial Intelligence. My concern does not come from a pining for some nostalgic vision of San Francisco that only exists in my memory; like people, organizations, professions, and even cities have measurable and distinguishable personalities and value systems (Park & Peterson, 2010). It appears that San Francisco's have changed, and now the call to alarm is growing steadily louder: what if Silicon Valley isn't the right place to invent new technologies like artificial intelligence anymore (Kottenstette, 2018)?

More than our spirit of innovation, however, I worry about the spirit of our people. Bay Area technology workers have the highest per capita rates of anxiety and depression of any workforce in the country, and an astounding 77% of employees report feeling burnout *in their current job* (Cook, 2020; Fisher, 2018); last year, the Centers for Disease Control even named the region as the "looming Mecca of suicide in the United States" (Hosansky, 2019). As Silicon Valley loses its luster for the next generation of inventors, there are signs the great talent migration is winding down, and recruitment is now the fastest growing inhouse expenditure for most marque employers in the Bay Area (Bessen, 2014). What if Silicon Valley is no longer the spiritual home of America's mavericks and dreamers, either?

That's not a story I like telling about the place where I was born. However, there is good news: Silicon Valley is still a place where people make new stories all the time (Katz, 2015), and we have the opportunity to tell ones about a future where the success and wellbeing of humans and human technologies are successfully intertwined. Silicon Valley's emerging generation of entrepreneurs and designers is much more idealistic than the last, and it prioritizes meaningful pursuits like this one over mere profitability or security (Bethere & Licite, 2019; Wood, 2013). Though this is a generation that is increasingly at-risk for burn out and clinical depression as they enter the workforce (Arnold, 1956), I still believe we can reimagine the perspective and process of design-thinking to better align with their intrinsic and extrinsic motivations in a way that serves both man and machine. We have the opportunity to address a declining perception of the greater purpose and vision of Silicon Valley that is draining the knowledge economy of the young minds and nonconformist visionaries it needs to operate successfully; addressing the source of this malaise should be at the forefront of our attention because, for the first time since the end of World War II, more young people are leaving the Bay Area then are coming (Colby & Ortman, 2015); but we can give them good reason to return.

The guiding philosophy of positive psychology is perhaps most succinctly summarized by the unofficial motto of its late founder and patron saint, Christopher Peterson: "other people matter." If Silicon Valley is to deliver a clearer and more compelling vision of the future of artificial intelligence—one in which human and machine agents work and thrive in collaborative harmony—then it must update its innovation practices to embrace a similarly transformative point of view: "other consciousnesses matter," too. I believe a positive existential posthuman philosophy of artificial intelligence innovation that calls for a greater diversity of contributing perspectives, a renewed passion for the scientific exploration of new practices, and a deeper emphasis on the pursuit of shared meaning, is more than enough to start shaping this new future in earnest.

The invention of artificial consciousnesses stands to become the greatest achievement in the history of mankind; however, the work of this invention is still being conducted by a small number of men and women living and working in innovation centers like Silicon Valley; we cannot leave this work to them alone, nor can we assume they will be well guided by the ethical conventions of the past. The frameworks which promote the virtuous potential of cybernetic collaboration are not the same as those that prevent its impropriety, and a commitment to collaborative problem solving is still our best chance of recognizing a more virtuous future.

I hope my work to improve our understanding of the capabilities of non-human agents[8] today will help to serve our common pursuit of wellbeing with the help of cybernetic technologies[9] in the future. That future has not yet been written; one day, it will be written in partnership with the conscious artificial agents we have created.

[8] A *measure* of all minds.
[9] Also a *measure* of all minds.

# Appendices

## Appendix 1. Glossary of Basic Artificial Intelligence Terms.

The following glossary of terms will help eliminate possible confusion throughout this paper, though it is not necessary to master these terms to follow along. Instead, review these terms to increase your general understanding of the scope of discussion:

***Artificial intelligence/machine intelligence***. Interchangeable terms referring to the science and engineering of intelligent agents that have the ability to achieve goals via a constellation of technologies that mimic (to varying degrees) the structures of the human brain (Poole & Mackworth, 2010).

***General intelligence/strong intelligence.*** An instance of artificial intelligence with the capacity to understand or learn any intellectual task with human-like ability; a category predicted to be decades away by even the most optimistic prognostications (Atkinson, 2018).

***Narrow intelligence/weak intelligence.*** An instance of artificial intelligence with the capacity to solve problems up to the level of human ability, but only within a very specifically defined use case**;** a category that includes most cutting-edge artificial intelligence programs in use today (Atkinson, 2018).

***Cybernetics/blended-intelligence.*** Cybernetic organisms are agents which have integrated elements of natural and artificial capabilities (Clynes & Kline, 1960); because this term is more commonly associated with works of science fiction, I prefer to us the term *blended-intelligence*.

***Statistical learning models.*** The earliest and most basic approach to creating intelligent agents relies on human training for knowledge classification (e.g.: “if furry and lives in house, then dog”) and knowledge control (e.g.: “if dog, then say hello). The addition of statistical analysis allows programs to learn from uncertainty (e.g.: "if furry and does not live in house, then unknown" and then "if unknown furry and lives outside, then cat") (Friedman et al., 2001).

***Search optimization models.*** Artificial agents which find optimal solutions to specific problems by searching through data sets too large or complex for humans to properly analyze (Russom, 2011); often mentioned in tandem with the term *big data*, which simply refers to extremely large data sets.

***Logic/reasoning models.*** Artificial agents capable of assigning degrees to representations of knowledge. Basic programing is limited to binary statements (i.e.: "this is 'black' or 'white'" or "this is 'true' or 'false'"), but these programs can reason more precisely (i.e.: "this is 'black' or 'white' or 'x degree of grey'" or "this is 'true' or 'false' or 'x% true'") (Meyer & Van Der Hoek, 2004).

***Artificial neural network models.*** Artificial intelligence design which draws on human neural networks and plasticity for inspiration (Tsybakov, 2004). Artificial neural networks are characterized by ongoing, continuous, and non-linear learning similar to that of humans (Hassoun, 1995). In this model, individual classifications of knowledge (i.e.; artificial neurons) use statistical analysis to activate other neurons for further analysis. More frequently connected neurons then form broader conceptual organizations (i.e.; artificial neural networks).

***Deep artificial neural network models.*** Deep artificial neural networks simply stack additional layers of artificial neural networks, allowing additional layers of conceptual knowledge to be applied to the decision-making process (Berg & Nyström, 2018). Though the training process is the same as with other artificial neural networks, it becomes deep learning when applied to multilayered artificial neural networks (Lauzon, 2012).

***Cognitive computing.*** An intelligent agent is a system that can perceive its environment and organize its available capabilities for most effective action (Wooldridge & Jennings, 1999). The most advanced form of intelligent agent known to mankind is man himself, but the new wave of supercomputing platforms, including IBM's Watson, Google's Deep Mind, and Salesforce.com's Einstein, are built to mimic his success.

**Appendix 2. Identification of Artificial Intelligence Innovation Eras.**

The following terms refer to the general periods of artificial intelligence design in Silicon Valley over the last several decades. Again, it is not necessary to master these terms to follow along. Instead, review these terms to increase your general understanding of the scope of discussion:

***First generation: reactive artificial intelligence.*** These types of programs cannot reference past events in order to make present decisions. They can only react to a single, static moment in time, and therefor do not improve over time (Nolfi et al., 2000). The most famous example of a reactive machine is IBM's Deep Blue, a chess-playing computer which entered the national consciousness when it defeated Garry Kasparov in a single round of chess in 1996. Every time Deep Blue took its turn, it was able to recognize the pieces and their position on the board, and then predict the most likely effect of every possible move. However, every single round was approached as a lone event, and so while Deep Blue could always maximize its decision-making within a single turn, it could not improve by learning from previous turns or by forecasting outcomes several turns in advance; tellingly, Kasparov won or drew each of the next five rounds.

***Second generation: limited memory artificial intelligence.*** These types of programs are able to derive insights from previously analyzed data in their memory storage, and they improve their performance overtime by remembering what they've learned in the past (Poole et al., 2010). Self-driving cars are now the most common consumer application of this type of artificial intelligence. In order to respond to driving conditions properly and quickly, self-driving cars combine pre-programmed knowledge with an on-going analysis of what their cameras observe on the road. Self-driving cars with limited memory programming can react to crowded intersections very quickly because they select the best response made under similar conditions in previously encountered intersections, rather than slowly assessing every variable as if it were the first time the car had ever come to a four-way stop.

***Third generation: theory of mind intelligence.*** These types of programs can demonstrate cognitive abilities equal to those of humans by using past learnings, present sensory inputs, and predicted outcomes for successful decision-making, but so far only under narrowly defined conditions (Müller & Bostrom, 2016). Even today's most advanced examples of Theory of Mind artificial intelligence applications are best described as proto-minds. For example, Siri can deliver answers relating to an exceptionally broad array of topics because she has been trained to index and reference the entire Internet; but does she always understand the question you've asked her? Only if the user asks in a very specific and predictable way. True Theory of Mind intelligence will require programs to, like human beings, manage a fluid process of interpreting a variety of present signals, recalling and analyzing information for related past experience, and accurately forecasting potential outcomes, all while reacting with appropriate decisions and behaviors.

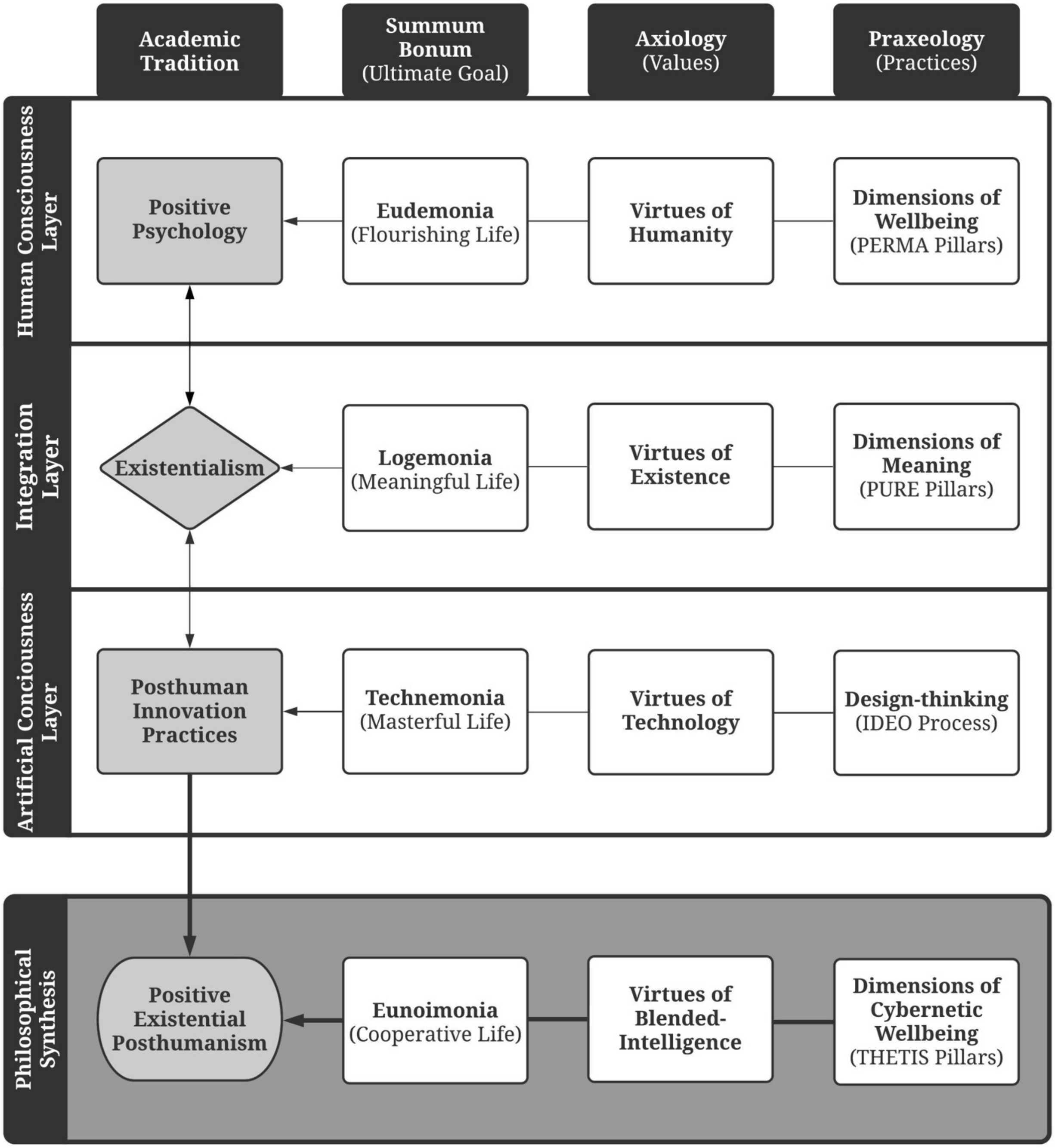
***Fourth generation: conscious artificial intelligence.*** This still theoretical type of artificial intelligence will be defined by its ability to not only make decisions at or beyond the level and scope of humans, but by its consciousness of its own existence, motivations, and agency (Haikonen, 2003); there is increasing consensus amongst computer and cognitive scientists that this advanced level of artificial intelligence need not remain theoretical forever. However, it remains impossible for serious scientists and engineers to predict the arrival of self-aware artificial intelligence with any amount of confidence; what can be said with greater confidence, however, is that the decisions made by software engineers and designers working in

Silicon Valley today are already affecting the nature and impact of this future form of consciousness.

**Appendix 3. Earliest Known Depiction of Thetis.**

*Thetis visits her adopted son Hephaestus in his workshop.* Kylix, ca. 490-480 BCE.

**Appendix 4. Synthesis of Positive Existential Posthumanism.**



**References**

Abbott, K., Davies, H. J., Mumford, G., & Harrison, P. (2019, September 19). The 100 best TV shows of the 21st century. *The Guardian.* https://www.theguardian.com/tv-and-radio/2019/sep/16/100-best-tv-shows-of-the-21st-century

Adams, I.D., Chan, M., Clifford, P.C., Cooke, W.M., Dallos, V., De Dombal, F.T., Edwards, M.H., Hancock, D.M., Hewett, D.J., & McIntyre, N. (1986). Computer aided diagnosis of acute abdominal pain: A multicenter study. *British Medical Journal (Clinical Research Edition), 293*(6550), 800-804.

Allan, J. F., McKenna, J., & Dominey, S. (2014). Degrees of resilience: profiling psychological resilience and prospective academic achievement in university inductees. *British Journal of Guidance & Counselling*, *42*(1), 9-25.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Publication.

Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science*, *291*(5509), 1748-1753.

Anderman, L. H., & Anderman, E. M. (2009). *Oriented towards mastery: Promoting positive motivational goals for students.* In R. Gilman, E. S. Huebner, & M. J. Furlong (Eds.), *Handbook of positive psychology in schools* (p. 161–173). Routledge/Taylor & Francis Group.

Andrews, C., & Mack, R. (2011). IBM to Collaborate with Nuance to Apply IBM's "Watson" Analytics Technology to Healthcare. IBM News Room. https://www.newsroom.ibm.com/2011-02-17-IBM-to-Collaborate-with-Nuance-to-Apply-IBMs-Watson-Analytics-Technology-to-Healthcare.1

Appleby, J. (1999). Government funding of the UK National Health Service: what does the historical record reveal? *Journal of health services research & policy*, *4*(2), 79-89.

Arnold, J. E. (1956). Creativity in engineering. *SAE Transactions*, 17-23.

Arnold, K. A., Turner, N., Barling, J., Kelloway, E. K., & McKee, M. C. (2007). Transformational leadership and psychological well-being: the mediating role of meaningful work. *Journal of occupational health psychology*, *12*(3), 193.

Ashby, W. R. (1961). *An introduction to cybernetics*. Chapman & Hall Limited.

Association for Computing Machinery. (2017). *Statement on algorithmic transparency and accountability.* Council for Public Policy. https://www/acct.www.acm.org/binaries/content/assets/public-policy/acm-pres.ltr-un-re-weaponse-systems.pdf

Atkinson, R. D. (2018). "It Is Going to Kill Us!" and Other Myths About the Future of Artificial Intelligence. *IUP Journal of Computer Sciences*, *12*(4).

Barabas, C. (2019). Beyond Bias: Re-Imagining the Terms of 'Ethical AI' in Criminal Law. *Chelsea Barabas, Beyond Bias: Re-imagining the Terms of" Ethical AI" in Criminal Law*, *12*.

Barrat, J. (2013). *Our final invention: Artificial intelligence and the end of the human era*. Macmillan.

Baumeister, R. F., Vohs, K. D., Aaker, J. L., & Garbinsky, E. N. (2013). Some key differences between a happy life and a meaningful life. *The journal of positive psychology*, *8*(6), 505-516.

Ben-Ari, O., Batthyany, A., Russo-Netzer, P., & Taubman-Ben-Ari, O. (2014). Well-Being and Personal Growth in Emerging Motherhood: And What About Meaning? In *Meaning in positive and existential psychology* (pp. 415–434). Springer.

Berg, J., & Nyström, K. (2018). A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing*, *317*, 28-41.

Bessen, J. (2014). Employers aren't just whining–the 'skills gap' is real. *Harvard Business Review*, *25*.

Bethere, S., & Licite, L. (2019, April 26). Values and motivations of generations in the labour market. *14th International Scientific Conference,* 100-153.

Boesch, C., & Tomasello, M. (1998). Chimpanzee and human cultures. *Current anthropology*, *39*(5), 591-614.

Bowie, A. M. (2007). *Herodotus: Histories*. Cambridge University Press.

Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and technology*, *9*.

Brewer, T. (2005). Virtues we can share: Friendship and Aristotelian ethical theory. *Ethics*, *115*(4), 721-758.

Broad, W. J. (2007). *The oracle: Ancient Delphi and the science behind its lost secrets*. Penguin.

Brockman, J. (2020). *Possible minds: Twenty-five ways of looking at AI*. Penguin Books.

Brown, T. (2009). *Change by design: How design thinking creates new alternatives for business and society*. Collins Business.

Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*(3).

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). AI now 2017 report. *AI Now Institute at New York University*. https://www.ainowinstitute.org/AI_Now_2017_Report.pdf

Cisco Systems Inc. (2019). *Trust Principles.* Trust Center. https://www.cisco.com/c/en/us/about/trust-center.html

Clynes, M. E., & Kline, N. S. (1960). Cyborgs and space. *Astronautics*, *14*(9), 26-27.

Colby, S. L., & Ortman, J. M. (2015). Projections of the Size and Composition of the US Population: 2014 to 2060. Population Estimates and Projections. Current Population Reports. 25-1143. *US Census Bureau*.

Cook, K. (2020). The psychology of Silicon Valley: Ethical threats and emotional unintelligence in the tech industry. *Springer Nature.*

Cooper, R., Junginger, S., & Lockwood, T. (2009). Design thinking and design management: A research and practice perspective. *Design Management Review*, *20*(2), 46-55.

Corea, F. (2017). *Artificial intelligence and exponential technologies: Business models evolution and new investment opportunities*. Springer.

Council of Artificial Intelligence. (2019). *Recommendation of the Council of Artificial Intelligence*. Organization for Economic Co-operation and Development. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Crawford, K. (2016, June 25). Artificial intelligence's white guy problem. *The New York Times*, *25*(06).

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology,* *15*(3), 41-63.

Csikszentmihalyi, M. (1990). The domain of creativity. In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 190-212). Thousand Oaks, CA: Sage Publications, Inc.

Csikszentmihalyi, M. (2013). *Flow: The psychology of happiness*. Random House.

Cusack, C. M. (2009). The return of the goddess: Mythology, witchcraft and feminist spirituality. In *Handbook of contemporary paganism* (pp. 335-362). Brill.

Dahlsgaard, K., Peterson, C., & Seligman, M. E. (2005). Shared virtue: The convergence of valued human strengths across culture and history. *Review of general psychology*, *9*(3), 203-213.

David, L. (2018). Agency and resilience in the time of regional economic crisis. *European Planning Studies*, *26*(5), 1041-1059.

Davis, D. M., & Hayes, J. A. (2011). What are the benefits of mindfulness? A practice review of psychotherapy-related research. *Psychotherapy*, *48*(2), 198.

Deckha, M. (2008). Intersectionality and posthumanist visions of equality. *Wisconsin Journal of Law, Gender & Society*, *23*(2).

Diakopoulos, N. (2020). Transparency. In *The Oxford Handbook of Ethics of AI* (Vol 1). Oxford University Press.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American psychologist*, *55*(1), 34.

Dillery, J. (2002). *Xenophon and the History of his Times*. Routledge.

Dubhashi, D., & Lappin, S. (2017). AI dangers: Imagined and real. *Communications of the ACM*, *60*(2), 43-45.

Duckworth, A. L., Eichstaedt, J. C., & Ungar, L. H. (2015). The mechanics of human achievement. *Social and Personality Psychology Compass*, *9*(7), 359-369.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, *42*(3-4), 177-190.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. Random House Digital.

Edelstein, L. (2000). The Hippocratic Oath: Text; Translation, and Interpretation. *Cross-cultural perspectives in medical ethics. Jones & Bartlett, Boston*, 3-21.

Eidinow, E. (2007). *Oracles, curses, and risk among the ancient Greeks*. Oxford University Press.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886.

Ferrando, F. (2012). Towards a posthumanist methodology. A statement. *Frame Journal For Literary Studies*, *25*(1), 9-18.

Ferrando, F. (2013). Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms. *Existenz*, *8*(2), 26-32.

Ferrando, F. (2016). Humans have always been Posthuman: A spiritual genealogy of Posthumanism. In *Critical Posthumanism and planetary futures* (pp. 243-256). Springer, New Delhi.

Ferrando, F. (2016). The party of the anthropocene: Post-humanism, environmentalism and the post-anthropocentric paradigm shift. *Relations, 4*(4.2), 159-173.

Fisher, J. (2018). *Workplace burnout survey*. Deloitte Touche Tohmatsu Limited. https://www.deloitte.com/us/en/pages/about-deloitte/articles/burnout-survey.html

Flynn, T. (2006). Sartre as Philosopher of the Imagination. *Philosophy Today*, *50*(Supplement), 106-112.

Forlano, L. (2017). Posthumanism and design. *She Ji: The Journal of Design, Economics, and Innovation*, *3*(1), 16-29.

Fornara, C. W. (1967). The value of the Themistocles Decree. *The American Historical Review*, *73*(2), 425-433.

Frankl, V. E. (1967). Logotherapy and existentialism. *Psychotherapy: Theory, Research & Practice*, *4*(3), 138.

Frankl, V. E. (1985). *Man's search for meaning*. Simon and Schuster.

Frankl, V. E. (2014). *The will to meaning: Foundations and applications of logotherapy*. Penguin.

Fredrickson, B. L. (2013). Positive emotions broaden and build. In *Advances in experimental social psychology* (Vol. 47, pp. 1-53). Academic Press.

Fredrickson, B. L., & Cohn, M. A. (2008). *Positive emotions.* In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (p. 777–796). The Guilford Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.

Gable, S. L., & Haidt, J. (2005). What (and why) is positive psychology? *Review of General Psychology, 9*(2), 103.

Gaggioli, A., Riva, G., Peters, D., & Calvo, R. A. (2017). Positive technology, computing, and design: shaping a future in which technology promotes psychological well-being. In *Emotions and affect in human factors and human-computer interaction* (pp. 477-502). Academic Press.

George, L. S., & Park, C. L. (2017). The multidimensional existential meaning scale: A tripartite approach to measuring meaning in life. *The Journal of Positive Psychology*, *12*(6), 613-627.

Goel, S., Shroff, R., Skeem, J. L., & Slobogin, C. (2018). The accuracy, equity, and jurisprudence of criminal risk assessment. *Equity, and Jurisprudence of Criminal Risk Assessment.*

Goldman, S. (Guest Host). (2018, February 5). *School's In* [Audio Podcast]. Stanford Radio. https://www.stanfordradio.sites.stanford.edu/programs/schools-in

Google Inc. (2019). Our Principles. https://www.ai.google/principles

Gordon, M. (2010). Learning to laugh at ourselves: Humor, self-transcendence, and the cultivation of moral virtues. *Educational Theory*, *60*(6), 735-749.

Gorman, D. (2010). Maslow's hierarchy and social and emotional wellbeing. *Aboriginal and Islander Health Worker Journal*, *34*(1), 27.

Govindji, R., & Linley, P. A. (2007). Strengths use, self-concordance and well-being: Implications for strengths coaching and coaching psychologists. *International Coaching Psychology Review*, *2*(2), 143-153.

Grassi, E. (2000). *Rhetoric as philosophy: The humanist tradition*. SIU Press.

Green-Simms, L. (2016). The emergent queer: homosexuality and Nigerian fiction in the 21st century. *Research in African Literatures*, *47*(2), 139-161.

Haikonen, P. O. (2003). *The cognitive approach to conscious machines*. Imprint Academic.

Hassan, I. (1977). Prometheus as Performer: Toward a Posthumanist Culture? *The Georgia Review, 31*(4), 830-850.

Hassein, N. (2017). Against black inclusion in facial recognition. *The Digital Talking Drum*.

Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. MIT press.

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, *25*(1), 30-36.

Hedreen, G. (2004). The Return of Hephaistos, Dionysiac processional ritual and the creation of a visual narrative. *Journal of Hellenic Studies*, 38-64.

Heintzelman, S. J., & King, L. A. (2014). Life is pretty meaningful. *American Psychologist, 69*, 561–574.

Held, D. (2006). *Models of democracy*. Stanford University Press.

Hendler, J. (2008). Avoiding another AI winter. *IEEE Intelligent Systems*, (2), 2-4.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29-29.

Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, *22*(7), 900-915.

Hosansky, D. (2011). Youth suicide. *Childhood and Adolescence in Society: Selections From CQ Researcher*.

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, *241*(4865), 540-545.

Institute of Electrical and Electronics Engineers. (2017). *Ethically Aligned Design*. Global Initiative on Ethics of Autonomous and Intelligent Systems. https://www.standards.ieee.org/develop/indconn/ec/autonomous_ systems.html

International Business Machines Corporation. (2017). *Transparency and Trust in the Cognitive Era*. https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/

Ivtzan, I., & Lomas, T. (Eds.). (2016). *Mindfulness in positive psychology: The science of meditation and wellbeing*. Routledge.

Jenkins, A. (2005). Creativity and resilience in the African American experience. *The Humanistic Psychologist*, *33*(1), 25-32.

Jianlan, S. (2013). Addressing Sustainability via AI-Report from the 23rd International Joint Conference on Artificial Intelligence. *Bulletin of the Chinese Academy of Sciences*, *27*.

Kapp, E. (2018). *Elements of a philosophy of technology: on the evolutionary history of culture*. University of Minnesota Press.

Kaptelinin, V. (2018). Technology and the givens of existence: Toward an existential inquiry framework in HCI research. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Katz, B. M. (2015). *Make it new: a history of Silicon Valley design*. MIT Press.

Kelley, T.A. (2001). *The art of innovation: Lessons in creativity from IDEO, America's leading design firm* (Vol. 10). Broadway Business.

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081-2096.

Kennedy, J. S. (1992). *The new anthropomorphism*. Cambridge University Press.

Kenney, M. (2000). *Understanding Silicon Valley: The anatomy of an entrepreneurial region*. Stanford University Press.

Keyes, C. L., & Annas, J. (2009). Feeling good and functioning well: Distinctive concepts in ancient philosophy and contemporary science. *The journal of positive psychology*, *4*(3), 197-201.

Kierkegaard, S. (2004). *Either/or: A fragment of life*. Penguin UK.

Kierkegaard, S. A., & Kierkegaard, S. (1946). *Kierkegaard Anthology*. Princeton University Press.

Korn M., & Silverman R. E. (2012, June 7). Forget b-school, d-school is hot. *The Wall Street Journal*, 7.

Kottenstette, R. (2018, March 15). Silicon Valley companies are undermining the impact of artificial intelligence. *Tech Crunch.* https://www.techcrunch.com/2018/03/15/silicon-valley-companies-are-undermining-the-impact-of-artificial-intelligence

Kruks, S. (2012). *Simone de Beauvoir and the Politics of Ambiguity*. Oxford University Press.

Lambert, L., Passmore, H., & Holder, M. D. (2015). Foundational frameworks of positive psychology: Mapping well-being orientations. *Canadian Psychology, 56*(3), 311-321.

Lantz, J. E. (1993). *Existential family therapy: Using the concepts of Viktor Frankl*. Jason Aronson.

Larrington, C. (1992). *The feminist companion to mythology*. Pandora Press.

Lauzon, F. Q. (2012, July). An introduction to deep learning. In *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)* (pp. 1438-1439). IEEE.

Leaf, W. (1902). *The Iliad* (Vol. 2). Macmillan and Company.

Liedtka, J. (2017). Evaluating the Impact of Design Thinking in Action. *Academy of Management Annual Meeting Proceedings*, *2017*(1).

Maddux, J. E. (2009). Self-efficacy: The power of believing you can. In C. R. Snyder & S. J. Lopez (Eds.), Oxford handbook of positive psychology (2nd ed.) (pp. 335-343). New York: Oxford University Press.

Madrigal, A. (2020, January 15). Silicon Valley Abandons the Culture That Made It the Envy of the World. *The Atlantic.* https://www.theatlantic.com/technology/archive/2020/01/why-silicon-valley-and-big-tech-dont-innovate-anymore/604969

Matthews, E. E., & Cook, P. F. (2009). Relationships among optimism, well-being, self-transcendence, coping, and social support in women during treatment for breast cancer. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, *18*(7), 716-726.

McCarthy, J. (1979). *Ascribing Mental Qualities to Machines*. Stanford University and University of California Departments of Computer Science.

McCormick, J. (2020, February 28). *Vatican advisory group issues call for AI ethics*. The Wall Street Journal. https://www.wsj.com/articles/vatican-advisory-group-issues-call-for-ai-ethics-11582893000

Meijer, A., Bovens, M., Goodin, R., & Schillemans, T. (2014). Transparency. In *The Oxford handbook of public accountability* (Vol. 1). Oxford University Press.

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors*, *58*(3), 401-415.

Meyer, J. J. C., & Van Der Hoek, W. (2004). *Epistemic logic for AI and computer science* (Vol. 41). Cambridge University Press.

Microsoft Corporation. (2019). Responsible AI. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6

Miller, A. (2013). The future of healthcare could be elementary with Watson. *Canadian Medical Association Journal, 185*(9), 367-368.

Miller, R., & Poli, R. (2010). The many aspects of anticipation. *Foresight*.

Mitcham, C. (1994). *Thinking through technology: The path between engineering and philosophy*. University of Chicago Press.

Moore, J. W. (2016). What is the sense of agency and why does it matter?. *Frontiers in psychology*, *7*, 1272.

More, M. (2013). The philosophy of transhumanism. *The transhumanist reader: Classical and contemporary essays on the science, technology, and philosophy of the human future*, 3-17.

Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In *Fundamental issues of artificial intelligence* (pp. 555-572). Springer, Cham.

Murakawa, Naomi. *The first civil right: How liberals built prison America*. Oxford University Press, 2014.

Neimeyer, R. A., Klass, D., & Dennis, M. R. (2014). Mourning, meaning, and memory: Individual, communal, and cultural narration of grief. In *Meaning in positive and existential psychology* (pp. 325-346). Springer, New York, NY.

Nelson, G., & Prilleltensky, I. (2010). *Community psychology: In pursuit of liberation and well-being*. Macmillan International Higher Education.

Nielsen, T. W. (2014). Finding the keys to meaningful happiness: Beyond being happy or sad is to love. In *Meaning in positive and existential psychology* (pp. 81-93). Springer.

Niemiec, R. M. (2013). VIA character strengths: Research and practice (The first 10 years). In *Well-being and cultures* (pp. 11-29). Springer, Dordrecht.

Nolfi, S., Floreano, D., & Floreano, D. D. (2000). *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines*. MIT press.

Norman, D. A. (2005). Human-centered design considered harmful. *Interactions*, *12*(4), 14-19.

Norman, K., Norman, S., Muralidharan, M., Issac, F., North, M., Shurney, D., Kieschnick, D., Cerrone, K. (2020). Perspectives on the emerging digital Transformation of the management of psychiatric comorbidities in chronically ill populations. *Journal of Psychiatry Treatment and Research, 2*(1), 25-31.

O'grady, P. F. (2017). *Thales of Miletus: The beginnings of western science and philosophy*. Taylor & Francis.

OpenAI. (2018). *Charter.* https://www/openai.com/charter

Organ, D. W., & Moorman, R. H. (1993). Fairness and organizational citizenship behavior: What are the connections?. *Social Justice Research*, *6*(1), 5-18.

Park, N., Oates, S., & Schwarzer, R. (2013). Christopher Peterson "other people matter": 1950–2012. *Applied Psychology: Health and Well-Being*, *5*(1), 1-4.

Park, N., Peterson, C., & Seligman, M. E. P. (2004). Strengths of character and well-being. *Journal of social and Clinical Psychology*, *23*(5), 603-619.

Park, N., & Peterson, C. (2010). Does it matter where we live?: The urban psychology of character strengths. *American Psychologist*, *65*(6), 535.

Partnership on AI to Benefit People and Society. (2017). Tenets. https://www.partnershiponai.org/tenets

Paulus, P. B., Dzindolet, M., & Kohn, N. W. (2012). Collaborative creativity: Group creativity and team innovation. *Handbook of organizational creativity* (pp. 327-357). Academic Press.

Pawelski, J., & Moore, D. (2013). *The Eudemonic turn : well-being in literary studies*. Fairleigh Dickinson University Press.

Peterson, C. (2006). *A primer in positive psychology*. Oxford, UK: Oxford University Press.

Peterson, C., & Park, N. (2009). Classifying and Measuring Strengths of Character. *Oxford handbook of positive psychology*, 25.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification* (Vol. 1). Oxford University Press.

Peterson, C., Ruch, W., Beermann, U., Park, N., & Seligman, M. E. (2007). Strengths of character, orientations to happiness, and life satisfaction. *The journal of positive psychology*, *2*(3), 149-156.

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. In *Advances in Neural Information Processing Systems* (pp. 5680-5689).

Poole, D. L., & Mackworth, A. K. (2010). *Artificial Intelligence: foundations of computational agents*. Cambridge University Press.

Prilleltensky, I. (2016). *The laughing guide to well-being: Using humor and science to become happier and healthier*. New York, NY: Rowman & Littlefield.

Proyer, R. T., Wellenzohn, S., Gander, F., & Ruch, W. (2014). Toward a better understanding of what makes positive psychology interventions work: Predicting happiness and depression from the person × intervention fit in a follow-up after 3.5 years. *Applied Psychology: Health and Well-Being*.

Quinn, P. D., & Duckworth, A. L. (2007). Happiness and academic achievement: Evidence for reciprocal causality. *The Annual Meeting of the American Psychological Society, 24*(27), 20-27.

Reese, B. (2018). *The fourth age: Smart robots, conscious computers, and the future of humanity*. Simon and Schuster.

Roberto, M. A. (2019). *Unlocking Creativity: How to Solve Any Problem and Make the Best Decisions by Shifting Creative Mindsets*. John Wiley & Sons.

Robeyns, I. (2017). *Wellbeing, freedom and social justice: The capability approach re-examined*. Open Book Publishers.

Roden, D. (2014). *Posthuman life: Philosophy at the edge of the human*. Routledge.

Rometty, V. (2012). A letter from the Chairman. *2012 IBM Annual Report*. https://www.ibm.com/annualreport/2012/bin/assets/2012_ibm_annual.pdf

Rose, H. J. (2004). *A handbook of Greek mythology*. Routledge.

Ross, C., & Swetlitz, I. (2017). IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. *Stat News*. https://www.s3z7w3t3.rocketcdn.me/wp-content/uploads/2018/06/IBM_pitched_Watson_as_a_revolution_in_cancer_care.pdf

Ross, C., & Swetlitz, I. (2018). IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. *Stat News https://www. statnews. com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments*.

Rowe, P. G. (1987). *Design thinking*. MIT press.

Rushby, J. (1988). *Quality measures and assurance for AI software.* [Washington, D.C.]. National Aeronautics and Space Administration, Scientific and Technical Information Division.

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudemonic well-being. *Annual review of psychology*, *52*(1), 141-166.

Salesforce.com Inc. (2020). *AI Ethics and Values.* Ethical Use Advisory Council. https://www.einstein.ai/ethics

Sartre, J. P. (2001). *Being and nothingness: An essay in phenomenological ontology*. Citadel Press.

Sartre, J. P. (2007). *Existentialism is a Humanism*. Yale University Press.

Schmidt, C. (2017). MD Anderson breaks with IBM Watson, raising questions about artificial intelligence in oncology. *JNCI: Journal of the National Cancer Institute*, *109*(5).

Schmitz, M., Endres, C., & Butz, A. (2008, January 11). A survey of human-computer interaction design in science fiction movies. *Proceedings of the 2nd international conference on INtelligent TEchnologies for interactive enterTAINment* (pp. 1-10).

Schwab, K. (2018, October 29). *IDEO breaks its silence on design thinking's critics*. Fast Company. https://www.fastcompany.com/90257718/ideo-breaks-its-silence-on-design-thinkings-critics

Schwarz, E. (2019). Günther Anders in Silicon Valley: Artificial intelligence and moral atrophy. *Thesis Eleven*, *153*(1), 94-112.

Select Committee on Artificial Intelligence. (2018). *AI in the UK: ready, willing and able*. House of Lords. https://www.publications.parliament.uk/pa/1d201719/1dselect/1dai/100/10014htm

Seligman, M. E. P. (1999). The president's address. *American psychologist*, *54*(8), 559-562.

Seligman, M. E. P. (2012). *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.

Seligman, M. E. P. (2018). PERMA and the building blocks of well-being. *The Journal of Positive Psychology*, *13*(4), 333-335.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology. An introduction. *American Psychologist, 55*(1), 5-14.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist, 60*(5), 410-421.

Seligman, M.E.P., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on psychological science*, *8*(2), 119-141.

Sheldon, K. M. (2002). The self-concordance model of healthy goal striving: When personal goals correctly represent the person. *Handbook of self-determination research*, 65-86.

Sheldon, K. M., & Elliot, A. J. (1999). Goal striving, need satisfaction, and longitudinal well-being: The self-concordance model. *Journal of Personality and Social Psychology, 76*(3), 482–497.

Sheldon, K. M., & Houser-Marko, L. (2001). Self-concordance, goal attainment, and the pursuit of happiness: Can there be an upward spiral?. *Journal of personality and social psychology*, *80*(1), 152.

Sheldon, K. M., & King, L. (2001). Why positive psychology is necessary. *American Psychologist, 56*(3), 216.

Shin, D. (2019). Toward Fair, Accountable, and Transparent Algorithms: Case Studies on Algorithm Initiatives in Korea and China. *Javnost = The Public.*, *26*(3), 274–290.

Shryack, J., Steger, M. F., Krueger, R. F., & Kallie, C. S. (2010). The structure of virtue: An empirical investigation of the dimensionality of the virtues in action inventory of strengths. *Personality and Individual Differences*, *48*(6), 714-719.

Slatkin, L. M. (1995). *The power of Thetis: allusion and interpretation in the Iliad*. Univ of California Press.

Slemp, G. R., Kern, M. L., & Vella-Brodrick, D. A. (2015). Workplace well-being: The role of job crafting and autonomy support. *Psychology of Well-being*, *5*(1), 7.

Smart, A., & Smart, J. (2017). *Posthumanism: Anthropological Insights*. University of Toronto Press.

Snyder, C. R. (1994). *The psychology of hope: You can get there from here*. New York, NY: Simon and Schuster.

Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of Rhyff's scales of psychological well-being: Method, mode, and measurement effects. *Social science research*, *35*(4), 1080-1102.

Stadter, P. A. (2005). Plutarch and Apollo of Delphi. *Gott und die Götter bei Plutarch*, 197-214.

Steel, C. G., & Primavesi, O. (Eds.). (2012). *Aristotle's Metaphysics Alpha: Symposium Aristotelicum*. Oxford University Press.

Steiner, C. J., & Reisinger, Y. (2006). Understanding existential authenticity. *Annals of Tourism Research*, *33*(2), 299-318.

Stollfuss, S. (2014). The rise of the posthuman brain: computational neuroscience, digital networks and the 'in silico cerebral subject'. *Trans-Humanities Journal*, *7*(3), 79-103.

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, *56*(4), 24-31.

Telotte, J. P. (1999). *A Distant Technology: Science Fiction Film and the Machine Age*. Wesleyan University Press.

Thompson, C. (2013). *Smarter than you think: How technology is changing our minds for the better*. Penguin.

Tillich, P. (1944). Existential philosophy. *Journal of the History of Ideas*, 44-70.

Topol, E. (2019). *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK.

Torkildsby, A. B. (2014). *Existential design: revisiting the" dark side" of design thinking*. [Doctoral dissertation, University of Borås]. Digitala Vetenskapliga Arkivet. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A877081&dswid=-5983

Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, *32*(1), 135-166.

University of Montreal. (2017, November 3). *Declaration for a Responsible Development of Artificial Intelligence.* [Conference session]. Forum on the Socially Responsible Development of AI. https://www.montrealdeclaration-responsibleai.com

Van Cappellen, P., & Rimé, B. (2013). Positive emotions and self-transcendence. In *Religion, personality, and social behavior* (pp. 133-156). Psychology Press.

Van Hooft, S. (2014). *Understanding virtue ethics*. Routledge.

Vining, J., Merrick, M. S., & Price, E. A. (2008). The distinction between humans and nature: Human perceptions of connectedness to nature and elements of the natural and unnatural. *Human Ecology Review*, 1-11.

Wachter, R. M. (2015). *The digital doctor: hope, hype, and harm at the dawn of medicine's computer age* (p. 208). New York: McGraw-Hill Education.

Wagner, L., Gander, F., Proyer, R. T., & Ruch, W. (2020). Character strengths and PERMA: Investigating the relationships of character strengths with a multidimensional framework of well-being. *Applied Research in Quality of Life*, *15*(2), 307-328.

Waldrop, M. M. (2019). News Feature: What are the limits of deep learning?. *Proceedings of the National Academy of Sciences*, *116*(4), 1074-1077.

Walker, K. (2019, March 26). *An external advisory council to help advance the responsible development of AI.* The Keyword. https://blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai

Wang, Y., & Chen, D. (2018). Rising sino-US competition in artificial intelligence. *China Quarterly Of International Strategic Studies*, *4*(02), 241-258.

Wartenberg, T. E. (2008). *Existentialism: A beginner's guide*. Simon and Schuster.

Waterman, A. S. (2013). *The best within us: Positive psychology perspectives on eudaimonia* (pp. xv-303). American Psychological Association.

Webb, A. (2019). The big nine: How the tech titans and their thinking machines could warp humanity. Hachette UK.

Welzel, C., & Inglehart, R. (2010). Agency, values, and well-being: A human development model. *Social indicators research*, *97*(1), 43-63.

Whitby, B. (2008). Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. *Interacting with Computers*, *20*(3), 326-333.

Wilkinson, R. G., & Marmot, M. (Eds.). (2003). *Social determinants of health: the solid facts*. World Health Organization.

Wikhamn, W., & Hall, A. T. (2014). Accountability and satisfaction: organizational support as a moderator. *Journal of Managerial Psychology*.

Williams, J. C. (2014). Hacking tech's diversity problem. *Harvard Business Review*, *92*(10), 94-100.

Wong, P. T. (2010). Meaning therapy: An integrative and positive existential psychotherapy. *Journal of Contemporary Psychotherapy*, *40*(2), 85-93.

Wong, P. T. (2011). What is existential positive psychology? *International Journal of Existential Positive Psychology*, *3*(1).

Wong, P. T. (2012). The meaning mindset: Measurement and implications. *International Journal of Existential Positive Psychology*, *4*(1), 3.

Wood, S. (2013). Generation Z as consumers: trends and innovation. *Institute for Emerging Issues: NC State University*, 1-3.

Wooldridge, M., & Jennings, N. (1999). Intelligent agents, Multiagent systems: a modern approach to distributed artificial intelligence. *Multiagent systems*, *6*.

Yalom, I. D. (2020). *Existential psychotherapy*. Hachette UK.

Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, *87*(9), 1423-1447.

Yeung, K., Howes, A., & Pogrebna, G. (2019). AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing. *The Oxford Handbook of AI Ethics, Oxford University Press*.

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, *5*(2), 29-47.

Zimmerman, J. E. (1966). *Dictionary of classical mythology*. Bantam.

Zlotowski, J. A. (2015). Understanding anthropomorphism in the interaction between users and robots.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics*, *7*(3), 347-360.