

Article

A Daoist-Inspired Critique of AI's Promises: Patterns, Predictions, Control

Paul D'Ambrosio ^{1,2} 

¹ Institute of Modern Chinese Thought and Culture, Philosophy Department, East China Normal University, Shanghai 200241, China; pauljdambrosio@hotmail.com

² Ma Yifu Humanistic Research Center, Zhejiang University, Hangzhou 310058, China

Abstract

Contemporary discussions of AI are often framed according to generally held assumptions which have largely escaped serious critical analysis. For instance, those who promote AI tout its predictive prowess: powerful algorithms fed massive amounts of data are able to discover knowable patterns that can accurately forecast the behaviors in anything from individual movie preferences to financial markets. Armed with this type of knowledge we can then use AI, the hope goes, to be more objective in our ethical practices. And most seriously, we must extend this to the way we develop AI, not only do we want AI to function ethically, but we caution ourselves that if Artificial General Intelligence (AGI), superintelligence, or anything like the “singularity” is ever developed, it should be positively aligned with human values. Reflecting on these positions from the perspective of classical Daoism gives us reason to pause. While Daoist texts also assume there are patterns in the world which one can successfully go along with, they are not enthusiastic about the rational or knowable nature of these patterns—rather, they encourage us to appreciate them as fundamentally complex and mysterious. In this article, some Daoist attitudes are also concretely applied to ethical considerations, which cannot easily be controlled or known, much less put into code. Inspired by Daoist texts, we might cultivate an attitude less filled with hubris than humility, where we are allowed more space from which we can reflect on how we think about AI. Many of the most pressing issues associated with AI could, in fact, be significantly alleviated simply by shifting the way we think about, use, and develop these technologies.

Keywords: AI; AI ethics; Chinese philosophy; Daoism; alignment; AGI; singularity; prediction



Academic Editors: Kaspars Klavins and Lihua Li

Received: 17 August 2025

Revised: 15 September 2025

Accepted: 24 September 2025

Published: 29 September 2025

Citation: D'Ambrosio, Paul. 2025. A Daoist-Inspired Critique of AI's Promises: Patterns, Predictions, Control. *Religions* 16: 1247. <https://doi.org/10.3390/rel16101247>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It has become fashionable in certain English language academic circles where research on early Chinese thought prevails—and this is very different in the Chinese tradition, until today—to attempt to be exceedingly precise about what one means by “Confucianism” or “Daoism”. In these spheres it is safest to avoid these terms all together. To speak instead of, for example, the *Analects* or *Laozi*. But even then, qualifications are expected, which books of the *Analects*, what version of the *Laozi*—the author is often supposed to apologetically acknowledge that they are only speaking of one very tiny slice of whatever tradition or text. (As if the opposite would ever be possible anyways.)

Another fashion expects quite a conflicting approach. When speaking of contemporary themes, such as environmentalism or democracy, scholars suddenly don a different

dress. Here it is not uncommon to see versions of “Daoism and Environmentalism” or “Confucian Democracy”. In style and content, many of these projects speak as if they have a comprehensive view on Confucianism and democracy, or have mastered both Daoism and environmentalism. Of course, not all studies are like this; some simply involve ideas from Daoism or Confucianism to reflect on some aspects of environmentalism or democracy. But (in this author’s experience at least) the majority are not like these and assume a rather arrogant tone. Scholarly debates expose this attitude most explicitly, as if my reading of how the *Zhuangzi* would react to contemporary environmental problems is somehow better than yours. Not only are the *Zhuangzi* and environmentalism themselves far too complex, but their relationship is exponentially more complex than either. And if there is anything scholars of the *Analects* or *Zhuangzi* should know about these texts, it is that they are rich, and that assuming a definite answer as to how they would deal with any issue not directly discussed in these texts is already wrong-headed. One way to approach these texts, and religion and philosophy in general, is as presenting ways for people to reflect on problems themselves. Nothing else. This is the general spirit taken in the current article.

In this paper I aim to construct reflections on some contemporary discussions of AI based on inspirations from Daoist texts. Fully acknowledging the above two fashions, I will qualify my discussion as such: considerations of Daoism in this essay will be limited to what *I have learned from Daoism about the reflecting on the world* (including myself, and relationships with others), as opposed to what I have learned from Daoism about Daoism, or, strictly speaking, what I think Daoism itself has to offer in terms of thinking about issues related to AI. Of course, I think that my understanding of Daoism is, to some degree, an accurate reading, and that what I have learned about the world is also telling (again, to some extent) of how the world is and how it operates. However, I seriously do not think that this is *the* way to read Daoism or *the* way to interpret the world. The most obvious limitations are that I am American, that I have grown up in this time period, with certain values and perspectives on the world, and the whole host of other contingencies which make me who I am—or, as some Daoist thinkers might put it, which make my perspective what it is (because that is all I am).¹ Nevertheless, I think that this paper can contribute to both understanding of AI and to readings of Daoism in productive ways.

Reflecting on AI from the perspective of a certain way of understanding Daoist texts offers a somewhat unique view—at least when compared with most academic discussions. Much of the published discussions of AI, AI ethics, or alignment put the world, humans, and human values in mathematical terms. In other words, everything becomes translated into clear algorithmic terms, so that AI can fully understand. This is unsurprising in some sense—the tendency to make things objective and mathematical is attractive for many reasons. A Daoist-inspired reflection differs in that it does not share this desire, and instead respects a high degree of unknowability, unpredictability, uncontrollability, and unengineerability of the world, of humans, and of values. Rather than trying to parse the world into parts which are knowable and controllable, Daoist thinkers aim to go along well with things, or with *dao* 道 or the way,² which is not something that can be fully grasped or manipulated. This essay takes inspiration from this way of reflecting on humans and the world.

Given the general readership of this journal I am writing for scholars whom I assume have a good foundation in philosophy, perhaps some understanding of early Chinese thought, and cursory knowledge of AI. The “Daoism” I reference mainly revolves around understandings of *Daodejing* and *Zhuangzi*, though again this paper is not solely about how to read those texts, but importantly also about how to extend ideas from them to develop reflections on AI today. (And this extension always means leaving somethings out, and manipulating ideas in certain ways—after all, the *Daodejing* and *Zhuangzi* have absolutely

nothing to say about AI, algorithms, or associated technologies.) For those unfamiliar with AI, I give a brief discussion in the first section of this paper. The general structure is as follows, in the first section I outline how AI operates, with a special focus on its concentration on prediction and patterns. In the second section I look at some general similarities with the way we find the world described in Daoism, i.e., *Daodejing* and *Zhuangzi*, before providing concrete critical reflection points based on what one might learn therefrom. Section three is a sketch of ethical considerations and AI, and in the fourth section some Daoist critiques of ethics, and of categorizing the world generally, are given. In the conclusion I will summarize some of the main points and rehearse how we might use observations from Daoism to inform the way we approach, use, and perhaps even develop AI technologies.

2. AI: Patterns, Prediction, “Objectivity”

Nearly all of the technologies commonly referred to as “AI” are made up almost entirely of algorithms. These algorithms are sets of rules or calculations which aim at optimizing for an objective function: a goal. They are basically complex mathematical computations geared towards predicting certain results. (The same is true of large language models [LLMs], such as ChatGPT, which predict what words would come next, one by one, in response to certain prompts.) The power of algorithms or AI today comes from their complexity. The “deep learning”³ revolution in AI is behind practically all the AI that makes headlines today. However, none of this would be possible without the “Web 2.0” or the “participatory web” and the behavioral data it produces. To put it simply, algorithms or AI cannot function without a massive amount of data, so-called “big data”. Without big data, everyday AI, including spam filters and resume readers, would not function, nor would the more amazing versions, such as ChatGPT or self-driving cars, be possible. The algorithms that are behind them are useless without big data—which is by and large human produced behavioral data.

Many assume that the computations done by AI are more or less similar to the ones we consciously experience in our minds; namely, that algorithms function according to cause-effect thinking. According to this way of thinking, we might assume that Netflix algorithm sees that you watched *Lady Chatterley’s Lover* and *Madame Bovary* and then recommends *Emma* because, like a human, it thinks you like period pieces, or historical romances. Or, you watch *Emma* and *The Menu* so it recommends *The Queen’s Gambit* because, like a human, it thinks you like Anya Taylor-Joy. While these scenarios are quite likely—i.e., the Netflix algorithm would probably recommend the respective third movie based on the first two—the “reasoning” we attribute to the AI that makes these recommendations is flat-out wrong. Despite their appearance of functioning according to cause-effect thinking or similar calculations, algorithms actually derive their predictions (or in this case recommendation) based on specific correlations. People who watched *Lady Chatterley’s Lover* and *Madame Bovary* tended to also watch *Emma*. Likewise, there is a high degree of correlation between watching *Emma*, *The Menu*, and *The Queen’s Gambit*. The algorithm does not “think” you like period pieces or have favorite actresses. It simply recognizes patterns based on behavioral data collected from many other people and makes a prediction (recommendation) accordingly.

Correlation is the underlying drive behind not just movie recommendations, but all AI. Predictions about the market, about recidivism, or about what is a pedestrian or a shopping cart, or what is a human face versus a gorilla’s face, are all based on correlations. This was a key element in the breakthrough with “deep learning”—images could be recognized as, for example, a cat versus a dog, when programmers ceased trying to define cat and dog for the program. Instead, they simply fed the algorithms tons and tons of images labeled “cat” or “dog,” and let the algorithm decide how it would predict “cat” or “dog” in images.

This is essentially what “deep learning” means (in many contexts); the algorithm teaching itself based on training sets. Algorithms identify data points and find correlations or patterns that allow it to effectively fulfill the objective function, i.e., the goal.

Importantly, the algorithm itself is part of its own prediction. When Netflix recommends that you will watch *Emma* what it actually does is predict that you *might* click on the “watch” button for *Emma*. This is not, however, a prediction that you might yourself simply find your own way to *Emma* (and given the number of movies on Netflix, that would be a ridiculously absurd prediction to make), the prediction is that if “recommended” to watch *Emma*, you will likely choose to play the movie. In other words, it predicts that when you are given *Emma* as a choice, displayed on your screen and presented by the algorithm, you will pick it. This active role which algorithms have in their own “discovery” of patterns in data is one of the most damning problems with AI today.

In addition to embedded biases, in the program itself, in training sets, and in data⁴ (the latter being perhaps the largest and most difficult problem to tackle), AI also creates (and this is in part due to the biased data) self-fulfilling prophecies. Describing predictive policing Sarah Brayne does an excellent job of summarizing both of these issues:

Predictive algorithms hold up a mirror to the past, and project into the future. If historical biases and police practices inform where and whom the police surveilled in the past, they will also shape where and from whom cops collect the crime data that is fed into policing algorithms to generate those risk scores. It is a self-fulfilling statistical prophesy. (Brayne 2021a, 16:50)

Brayne also looks at programs that suggest high crime risk. Unsurprisingly, she says, police go to those recommended areas and “find crime”. Like Netflix, the argument is that the algorithm predicts that if the police go there looking for crime, they will find it. The prediction is not that crimes will happen in one place or another. Just as Netflix does not predict that you will watch *Emma* without being recommended it, the algorithms informing police are not suggesting that a crime will happen in one area and not another. Again, the Netflix recommendation is that when presented with *Emma*, a film you might not otherwise find, you might click on it. Police algorithms similarly recommend an area where crimes might be found, but perhaps if and *only if* police go looking for them there (Brayne 2021b). The result can be, as Brayne argues, that old biases become cooked into new tools. Problematically, these tools appear to have mathematical certainty, and be, somehow, outside of the influence of bias. This is particularly insidious as assumptions about algorithms and math suggest objectivity and are seldom subject to skepticism or critical reflections. So the biases AI reinforces are both harder to identify and more damning for victims. Combating these issues is also more complex than it may first appear. There are no simple solutions. AI is multifaceted, and its ability to train itself, while relying on amounts of data so large that they are hard to even imagine, means that there are no simple solutions to addressing these issues.⁵

When academics and culture critics demand more “transparency” about the “black boxes” that AI are, they often misunderstand that to some extent this is both impossible and even undesirable. Programmers themselves do not actually know why or how a movie is recommended, a person is deemed likely to be a criminal, or one word is predicted to follow another (as in ChatGPT and other LLMs). Of course there can and should be more transparency about those aspects of the algorithm that can be known, the training sets (and their biases), and, most importantly, how the AI is utilized. But the algorithms themselves cannot be, as Elena Esposito states, made stupid enough for humans to understand (Esposito 2022, p. 17). And as Max Tegmark argues, if they could, they would no longer function how they function (cf. Tegmark 2017). Importantly, this is not because they are “more intelligent” —unless you define intelligence as the ability to find effective patterns in huge

amounts of data. Since, as Esposito notes, they are effective at communicating with us, but not necessarily “intelligent,” it comes as little surprise that we might never be able to understand exactly how they function.

One way to approach the problems with AI can begin from recognizing that, given the multifaceted nature, what deep learning actually is, and the importance of big data, biases and self-fulfilling prophecies can be minimalized to some degree but never solved. As Esposito writes,

algorithmic bias is only one component of the problem. Deeper, and more difficult to manage is what is often labeled as “data bias,” which does not depend on the values of the programmers. Instead, it depends on the underlying source of the algorithms’ efficiency: the access to the big data they find on the web, which frequently builds upon the uncoordinated input of billions of participants, sensors, and other digital sources. Machines participate in a communication that is neither neutral nor egalitarian, and they learn to work correspondingly, in ways that can be biased very differently from the preferences of their designers. In pursuing the goal of algorithmic justice, then, the most difficult problems are communicative, not cognitive. (Esposito 2022, p. 109)

Cathy O’Neill, whose book *Weapons of Math Destruction* pioneered research on the prejudices built into algorithms which had been widely regarded as “mathematical” and therefore objective, makes a similar point. According to O’Neil, “Big Data processes codify the past. They do not invent the future. Doing that requires moral imagination, and that’s something only humans can provide.” (O’Neil 2016, p. 123). In other words, there are inherent limitations to what we can do with big data, or AI. We are not going to find data that is free of human biases; data is, after all, human, and humans have biases. Unfortunately, the same is true of the programs themselves, they have biases cooked into their very core. However, what O’Neill is pointing to is recognizing these limitations and moving beyond them—while simultaneously acknowledging we cannot entirely eliminate them.

Almost all of the discussion about algorithms or AI and their predictions is centered on making them more efficient (for whatever purpose) and extinguishing (certain) biases. However, as hinted at by O’Neill, and taken up more concretely by Brayne, Nicholas Carr (cf. Carr 2010) and others, the way we use this technology is perhaps the most pressing issue. Today judges in some parts of the world are required to look at algorithmically generated sentencing suggestions. We might want to push to make these algorithms as transparent as possible, to fetter out many biases, and overall make them fairer and more just. Additionally, we can ask that judges are well informed about these suggestions, and especially about the algorithms that produce them. Perhaps we want judges to work with algorithms in making sentences for those convicted of crimes—or even in deciding if someone committed a crime. But we might also step back and ask why we want this in the first place. Instead of focusing narrowly on the predictions AI makes, and ensuring that they are objective in finding patterns and analyzing them, we can reflect on the assumptions we make about these predictions and patterns before we even see them. How we approach AI is a question often overlooked. Considering Daoist appreciations of patterns and predictions can help shake off some of our assumptions and perhaps reevaluate the overall approach we take towards any prediction tool.

At the risk of being somewhat oversimplified, we may consider the way we think about the predictive dimension of AI, and in some regards of AI in general, as providing an accurate model of the world. It is treated as and is programmed to capture some essential patterns which elucidate critical features from which we can make predictions. The hope is that we can come to know, predict, and control more about the world, ourselves, and

our interactions with the help of these algorithms. Undoubtably, AI feeds our Promethean hubris and various promises of mastery.

3. Dao: Pattern, Prediction, “Unknowability”

Daoist texts offer interesting resources to develop reflections on the approach many people have to AI, and associated desires for objectivity and mastery over the world. In Daoist classics we do not find a denial of the possibility of finding and engaging in crucial patterns of the world, nor do they reject predictions wholesale. However, what they do offer are patterns which are not so much models of the world; they are better understood as loose guides which are acutely attuned to an appreciation of the world as highly complex, and only unknowable and uncontrollable in limited scopes. Daoist texts point to a type of paradoxical anti-prediction prediction. The promises proffered by Daoist thinkers are largely negative: the easing of stress and anxiety, dismantling the tightly held desires and senses of self, calling into question views of how the world is or ought to be, as well as not interfering with otherwise natural happenings. In this way they are highly critical of the types of patterns and predictions that bolster human desires to manage and to mold things according to their own relatively context-independent intentions and ideas.⁶ Along this trajectory common notions of virtue, morality, and productive ways of living roles are often criticized.⁷

These critiques of virtues or moral and ethical thinking in general are well-known features of early Daoism.⁸ Chapters 2, 5, 18, 19, and 38 of the *Laozi* are among the most obvious and famous discussions of how moral thinking⁹ can be problematic. We can focus in on the first lines of read chapter 2, for example, as representative of a logic which undermines all fixation on valuation; these lines read

When all under heaven know beautiful as beautiful, then there is already the ugly; when all know good as good, then there is already not good. [...] Therefore, the sage takes on [the task of] not acting-for (*wuwei*) and performs “no speech” instruction. (Chen 2020, p. 56; translation modified)

Determining something as valuable automatically necessitates that whatever is not contained within that category or partaking within that essential “positiveness” is not valuable and perhaps even warrants disregard or disdain. There can never be any way of including everything as “beautiful” or “good” and so these categories simultaneously manifest their opposites. The sage, who is representative of that which the *Laozi* wishes to advocate, recognizes this and thereby does not “act-for” anything in particular. This person neither orients themselves towards what is “beautiful” or what is “good”. Of course what they do end up doing might later be judged as a valuation of the “beautiful” or the “good,” but this was not part of their original intention—they did not have a prior fixation on anything in particular as “beautiful” or “good”. Moreover, what they end up doing is largely determined by their environment, which can be contrasted with the relatively context-independent intentions and ideas of people who are attached to certain virtues or models, and then attempt to impose them on the world.

Relatedly, chapter 5 of the *Laozi*, which states that heaven and earth are not humane, and neither too are sages, can be read as being in-line with the “logic” of chapter 2. The sage should not “act-for” abstract notions, and rather forget about “humaneness”. This is seen even more explicitly in chapters 18 and 19. Chapter 18 can be translated:

When the great Dao¹⁰ is abandoned, there is humaneness (*ren* 仁) and duty (*yi* 義). When three family relations and six roles (六親) lack harmony, there is filial piety and parental care. When the state and families are thrown into confusion, then loyal servants arise. (Chen 2020, p. 139; translation modified)

Specific virtues and patterns of relationships are instituted not to bring the world back to “the great Dao”—which is usually understood as something like harmony and is the preferred state of the world—rather, they only signal the denigration of a better world.

Establishing virtues and setting up frameworks for family relations and social roles are also ways of making the world more knowable, predictable and controllable. Much of their validity is accounted for by a claim that they have been derived from the “natural” ways of human beings, or from “human nature,” “natural human dispositions” and other naturalistic claims. The predictive aspects predict not only how others will react, or how the world will become, but also, as highlighted with the idea of “self-cultivation,” about how the person themselves will transform. For example, in the *Analects* 1.2 we find an explicit discussion of this thinking:

Master You said, “There are few who, being filial (xiao 孝) and fraternal (di 弟), are fond of offending their superiors. There has been none, who, not liking to offend against his superiors, has been fond of stirring up chaos. Exemplary persons cultivate the root, for having the root established, the Way will grow. Filial piety and fraternal love—they are the root of humanheartedness (ren 仁), are they not?” (Ni 2017, p. 81)

Daoists reject this way of understanding. Asking people to develop themselves in certain ways based on past models, or picking out some aspects of human behavior, thoughts, and emotions as the prioritized representations of “natural,” and grasping tightly to them, ignores just how incongruous and multifarious people are, how multifaceted and conflicting interactions can be, and how complex and unknowable the world is as well.¹¹

Thus, models, virtues, and self-cultivation promise a world that is more knowable, predictable, and controllable—and if we are convinced by them, then we also believe that they can all be improved in terms of accuracy, resonance, and efficiency. In other words, they are, from a rather broad perspective, functionally akin to AI. Accordingly, the Daoist response to models, virtues, and self-cultivation can be used to help us reflect on our assumptions about AI as well. But it is useful to incorporate the more positive dimensions of Daoism as well. Afterall, Daoist texts are not only critical of certain desires to make the world more knowable, predictable, and controllable; they also offer alternative ways to think about and interact with oneself, others, and the world, namely through “self-so” (*ziran*) and “not acting-for” (*wuwei*). In these ways people aim to get some sense of *dao* or the way, and act in accord with it.

Chapter 57 of the *Laozi* clearly demonstrates how the productive patterns of “self-so” and “not acting-for” are useful guidelines for thinking about how the world might operate—though we should be cautious about thinking too mechanically or concretely about what happens:

Govern the state with uprightness (*zheng*), employ military means with ingenuity, take control of the world by not undertaking anything. How do I know this to be so? Because of this: The more prohibitions and taboos there are in the world, the more impoverished the people will become; if people have many sharp instruments, the state and families will become more and more shrouded in darkness; if people use many clever techniques, strange things will arise ever more often; the more conspicuous laws and decrees become, the more bandits and robbers there will be. And so the sage says: “I do not act-for, and the people self-transform (自化); I delight in tranquility, and the people self-align (*zheng*); I do not undertake anything, and the people self-thrive; I am without desires, and the people self-simplify.” (Chen 2020, p. 316; translation modified)

Here the *Laozi* steers us away from thinking about the world in terms of knowability, predictability, and controllability. Of course, one could object, and argue that the *Laozi* is saying that the world is knowable, predictable, and controllable. For example, we can know what kinds of attitudes are efficient, and what kinds are problematic. There is plenty of prediction and control here as well; “not acting-for” results in people acting in spontaneous and natural ways—it gives rise to “self-so” action in people.¹² However, all of these knowable, predictable, and controllable elements are, at best, paradoxical. The *Laozi* does not really claim to know that much about the world, except that certain orientations are often problematic and some are usually better. Those which bring us closer to *dao* (way) or which help us align ourselves well with *dao* are helpful, those which do not are problematic.¹³ Similarly, the prediction and control we find here are not about optimizing for what one predetermines, or what one intends in a concrete sense. (This is what the *Laozi* critiques above.) Rather, it is about allowing things to follow their own natural dispositions, as they interact with all other natural dispositions around them. In this way the prediction and control we find here are more about not predicting or controlling (in most understandings of these concepts). Or, we can understand the *Laozi* as emphasizing that one cannot predict much, or control themselves or others in a significant sense.

The *Laozi* thereby presents us not necessarily with a complete rejection of finding efficient patterns which will help us know, predict, and control the world or the way of the world (and humans)—namely, *dao*. The world (and the way of the world—*dao*) is knowable, predictable and controllable in small degrees, and in certain narrow ways, but only if we are relatively “open,” tolerant, and do not actually try to gain too much precision in our attempts to know, predict, and control. Paradoxically, what this really means is that we come to realize that there is much of the world that we do not know, we can never predict, or control, and through recognizing this the *Laozi* is actually suggesting we become satisfied therein. We thus acknowledge that sometimes the more we try to resist certain things, or steer them in specific ways, the more the exact opposite comes about. Our best predictions are actually related to letting things follow their own natural course, and we bring this about by acting in this way ourselves—i.e., when we follow our own natural inclinations (i.e., when we are *wuwei* and *ziran*).

Generally speaking, when compared to most approaches to AI, including those who develop it and theorize about it, Daoist texts express much less of a belief that the world can be known, mathematically modeled, and operated in effectively using precise calculations. Many of the titans of AI are highly invested in finding patterns which can help us know, predict, and control the world. This is the promise of AI, and it should come ever closer to total knowledge, prediction, and control. For example, in conversation with Stephen Hawking, Mark Zuckerberg outlined his approach to human relationships:

“I’m most interested in questions about people. I’m also curious about whether there is a fundamental mathematical law underlying human social relationships that governs the balance of who and what we all care about. I bet there is.” (Fisher 2022, p. 122)

Speaking about Amazon, and in response to what he thinks of human judgment, Jeff Bezos has reflected a similar attitude:

“There is a right answer or a wrong answer, a better answer or a worse answer, and math tells us which is which.” (Fisher 2022, p. 115)

Proceedings from the 1956 Dartmouth Conference are where the term “Artificial intelligence” was coined. There, we find the following:

“The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.” (Fisher 2022, p. 123)

And in practice, everything from finding romantic partners and parenting children, to coming up with sentences for criminals and developing effective trade routes are increasingly subject not only to AI, but also to the algorithmic thinking which underlies it.¹⁴

Daoism is then best understood as an alternative *attitude* to some of the more all-encompassing, “cause-effect,” and abstract views of knowability, knowability, predictability, and controllability. Some approach AI thinking in the same way ethics or virtues are often treated, as if they are accurate models of persons and the world, and with the underlying assumption that we can manipulate the things according to the suggestions of these models. While there may indeed be some desired results in this type of practice, Daoism warns that perhaps there are limitations, and suggests being less fixated on the accuracy of models, and less reliant upon their wide-spread implementation.

4. AI Ethics

In his humorous and half-joking (and half-not joking) book *Parentonomics* (Gans 2008) the economist Joshua Gans refers parenting as “one big economic management problem”. Gans applies principles from economics in his role as father with the hopes of optimizing the development of his children. Incentive and rewards structures play a pivotal role. So, for his older daughter, who loves food, he uses it as an incentive to optimize for desired behavior. Unfortunately, the results were sometimes not exactly what Gans hoped for. Case in point: enlisting his older daughter to assist potty training her younger brother, Gans rewarded her with a piece of candy every time she helped him use the bathroom. The daughter, however, promptly found a loop-hole: the more she forced her brother to drink water the more he would go to the bathroom—and the more candy she would receive. She too began to optimize her younger brother’s pee for rewards.

From an ethical perspective we can say that the daughter reacted to her father’s mechanistic incentive structure for good behavior by perfectly following suit. She took a mechanistic approach to being an older sister in the same way her Gans took a mechanistic approach to being a father. Brian Christian, who believes algorithms can be used to help people live better, suggests that Gans scold his daughter “in precisely equal measure...[for instance taking a candy away when she forces her brother to drink water] so that the net gain of further repetitions is zero” (Christian 2020, p. 170). In other words, when the mechanisms do not produce the desired result, use other mechanisms to fix them.

Indeed, this is how we often seek to fix issues associated with algorithms and AI. Problems with mathematical models require mathematical solutions. So when facial recognition technology does a poor job of recognizing faces of darker skinned persons, or of female faces, one effective solution is to provide it with a more diverse training set. Or, if search engines skew towards undesirable associations, we have to adjust algorithms so that they optimize for what we consider more appropriate. At the core of all AI development, and especially those technologies used to tackle moral issues, is an algorithmic understanding of the issues at play. Even concepts like “privacy” need to be translated into mathematical language so that algorithms or AI can express them. Otherwise they remain too opaque, and do not communicate within or through the technology.

For example, computer scientists Michael Kearns and Aaron Roth claim to have proven that AI “understand,” and thereby express, a certain type of privacy. They offer “differential privacy” as “a mathematical formalization of the [idea] that we should be comparing what someone might learn from an analysis if any particular person’s data was included in the dataset with what someone might learn if it was not” (Kearns and Roth

2020, p. 36). If this is the model for privacy, then we already have proven methods to program privacy into AI systems. Exactly what this type of privacy means for people, and what it has to do with other areas where we want to keep certain information confidential, is not part of Kearns and Roth's equation. They expressly celebrate that their understanding does not "ponders deep and existential questions,"; moreover, "the definition of harm [that which differential privacy avoids] can be anything you want it to be" (Kearns and Roth 2019, 8:30). What matters in this context is being able to explain "differential privacy" to machines. What exactly "harm" might mean, and the gap between "differential privacy" and the type of "privacy" most people think of, is less important.

When we do want to "ponder deep and existential questions" or perhaps even reflect on what it means to be a good person, family member, or citizen, learning from philosophy, and particularly from "ethics," can be a good place to start. Today, North American introductory colleges courses on "ethics" usually include at least the following four ethical theories: virtue ethics, deontology, consequentialism (utilitarianism) and care ethics (or feminist ethics). In a nutshell these approaches can be summarized as such: (1) Virtue ethics identifies different character virtues and asks people to develop them. When analyzing a situation, the virtue ethicists looks at the expression of particular virtues within a certain frame (often a fictional crisis), which often incorporates analysis of concrete situations and communities. (2) Deontology says that acting according to duty is best. The approach has most famously been associated with the categorical imperative (one formulation of which is: "I should never act in such a way that I could not also will that my maxim should be a universal law" (cf. Kant 2002). From this perspective determining the ethical value of an action has nothing to do with virtues or consequences; whether or not something good results from acting out of duty is not important. Acting out of duty is itself the only thing that counts. (3) The consequentialist says that only consequences of actions count. One oversimplified formulation can be "maximize the greatest good for the greatest number" (cf. Bentham 1977, p. 393). Of course we need to determine what counts as good and for whom, but we do not think too much about motivations or other factors. Consequences are all that really matter. (4) Care ethics, sometimes called feminist ethics or the ethics of care, criticizes the above three positions for being overly rational and overly focused on individuals as abstract "rational" agents. Some versions say that relationships matter most, and feelings need to be central for any ethical assessment. Determining the ethical value of a situation means looking at the relationships of the people involved, and placing primacy on their emotions.

Each of these ethical approaches view the world in a different way. They are methods of organizing our attention and structuring what counts and how or why. The same situation might be deemed highly ethical from a deontologist's perspective and highly unethical by from a consequentialist's viewpoint. The value comes from the ethical theory itself. An ethical theory is an *approach* insofar as it determines what is a value, how to weigh that value relative to other values. It tells us *how to approach* situations, and what counts in these situations and what does not. Each theory already predetermines value, and even tells us what we should consider meaningful or not. When we apply a specific theory to a situation, we already know how it will view the situation. In fact, even before we know what the situation is, we already know from our approach what counts and what does not. If, for example, we start from deontology, we already know that the concrete consequences do not matter. If care ethics is our model, interrelatedness and emotions will be primary.

When disputing which ethical theory is best, there is little recourse outside of preference. Of course one can say "deontology is clearly better than virtue ethics because virtues can be used by thieves and mobsters". In turn, the virtue ethicists can retort "well, that is not using my theory correctly, and deontology doesn't work because it undermines com-

mitments to communities and the individual's cultivation of character". There is clearly no way to somehow ultimately solve these issues. When one argues that one ethical theory is "better" than another, this involves either an appeal to one ethical theory as primary, or (relatedly) resting on ethical intuitions which are situated in a way that skew towards the preferred theory or drawing on examples that neatly fit into a particular theory. When the deontologist says, "look, robbers can also exhibit bravery; therefore, virtue ethicists and wrong and deontology is the obvious alternative" they rely on our (supposed) shared agreement that robbers are bad, that bravery can be exhibited by robbers, and, most importantly, that no exceptions to virtue ethics can be admitted into the approach. These assumptions are based on neither virtue ethics nor deontology. Instead, deontology and virtue ethics are meant to respond to, or develop models for, our initial intuitions (e.g., robbers are bad, etc.). From here ethical theories promise to teach higher-order ethical thinking, which is not as ready-made as familiar observations such as "robbers are bad". In other words, the deontologist will give a few examples showing that our ethical institutions agree with deontology (or that they should), and from there explain what this entails when we develop deontology as a fuller and more abstract model of ethics.

An "ethic" in this sense is then a particular way of looking at the world. It is a model which tells what is valuable, what is meaningful, and expounds on how we should use this thinking in unfamiliar ways. So, while being "honest" might be an obvious ethical virtue, the theory of virtue ethics can tell us what more about this virtue, and how to use it. Or, while many people might agree that in some cases sacrificing a few people to save many people is a good idea, consequentialism suggests that this basic logic can be true in more situations than we might readily admit.

Current research on AI ethics is often centered around a number of different approaches, all of which are related to the depiction of ethics given above. For example, skewing AI or algorithms towards particular ethical theories, having it operate according to more specific standards or values, or towards human alignment in general—i.e., having AI be fundamentally and exclusively aligned with human values. Specifically, this means that, for example, those who work in virtue ethics argue that AI needs to exhibit certain virtues (such as justice, honesty, responsibility and the like). Or they might say that AI needs to contribute to "human flourishing". Indeed, many researchers try to incorporate one or many ethical approaches when discussing how to make AI more ethical. AI privacy has been conceived of along deontological lines, and dilemmas self-driving cars might face have been discussed with recourse to consequentialism. To some extent, however, this entire discourse has overlooked the most foundational aspect of AI: AI itself is an ethic.

Algorithms, which are the basic building blocks of AI systems, are mathematical models. Most seek to optimize for an objective function, and do so through making predications based on big data. Even if we develop definitions of specific virtues, such as privacy, or honesty, having an AI exhibit these virtues means taking a mechanistic and mathematical approach to them. In other words, there is no way to develop an AI system that does not see the world as a complex math problem, because that is simply all that algorithms are, and the only way they can (thus far) be developed. (In fact, many scholars of philosophy, especially in Anglophone discourses, including so-called analytic philosophy, argue that philosophy too should be mathematical.)

Recent work focused on human alignment has sought to overcome associated issues with developing ethical AI. Since algorithms maximize for a given objective, many developers have pointed out that humans need to be extremely careful in coming up with goals. If Gans' daughter is able to find loopholes in straightforward instructions, imagine what a technology several times more "intelligent" than a person might come up with. Loopholes always exist, and can have disastrous results, in ways and magnitudes that the people cre-

ating the technology might never fathom.¹⁵ Accordingly, many argue that we either need to pause our development of AI and perhaps proceed with extreme caution, or we need to come up with an entirely new approach.¹⁶

Another approach to solving these issues has been pioneered by Stuart Russell, who has developed methods of “inverse reinforcement learning” for training AI. Instead of focusing on specific goals or actions, this type of approach prioritizes states, and hopes to thereby (perhaps) completely eschew the very possibility of loopholes. Russell outlines how “inverse reinforcement learning” training works:

The machine’s only objective is to maximize the realization of human preferences.

The machine is initially uncertain about what those preferences are.

The ultimate source of information about human preferences is human behavior.
(Russell 2019, p. 173)

Since AI does not start with any goal to maximize, it cannot find any loopholes. Rather it focuses instead on preferences that humans have, maybe even some that they themselves do not know, and seeks only to aid humans in achieving their preferences. Moreover, since the system(s) remain unsure about human preference, they have to consistently update themselves. One major issue that still remains, however, is of having “bad” preferences, and other problems that can arise when considering conflicting preferences, either within the same person, or between groups. For example, when someone’s preference includes harming other people, or one’s self.

Regardless of how AI develops, whether thinking about maximizing a goal revolving around a virtue, a consequence, or even maximizing the realization of human preferences, the approach is always based on a very particular and narrow view of the world. A view which can only see, suggest, and interact in mathematical and mechanistic ways. This is how AI sees the world and it is how it sees humans. AI also sees the interrelations between humans and interactions between humans and the world in the same way as well. Here again looking to Daoist resources can help us consider our attitude towards these technologies. Early Daoism does not offer specific critiques on this technology—it was of course not something imaginable—but Daoist texts can help us reflect on the way AI parses the world, and the attitude humans take towards that parsing.

5. A Daoist Ethos

All moral technologies break the world apart, and locate value in some areas, and thereby imply that other areas are without value. This is just as true of deontology, consequentialism, virtue or care ethics as it is of AI technologies. Daoism proffers an approach where the parsing of the world is understood as a necessary process, yet it is not praised. Whereas moral perspectives and AI prioritize some features of the world, person, or relationships over others, Daoist classics remind us that while this may be a useful way to think at certain times or for certain reasons, we should never forget that the world, persons, and relationships are far more complex than we can ever imagine. Moreover, there is much about the world that we do not, and cannot know. And even the things we do know necessarily blind us to other factors.

The opening stories of the *Zhuangzi* tell of a huge fish that transforms into a bird. As this absurdly large bird takes flight, a smaller bird criticizes it, laughingly stating that flying a few meters is good enough. There are many ways to interpret this story, and it is certainly making more than one point, for instance, many scholars bring up the difference in perspectives as important, or, as Guo Xiang 郭象 (d. 312) highlights, we also find an illustration which tells that each thing has its own environment where it fits best. One of the less noticed parts of this story is that the *Zhuangzi* does not necessarily favor the larger

perspective or environment of the big bird over the smaller one.¹⁷ Additionally, while we might think that the larger perspective (of the big bird) is better than the smaller one, we are actually told that the big bird sees not much more than the blue sky. In other words, the story expresses the inherent limitations of any creature, or any perspective.

Of course, from a broader perspective one might see a “bigger picture,” but if the big bird soars high in the air, then its view of the world is probably not that different than our own from an airplane. The details of the world, specific distinctions, concrete interactions, and particulars, are all lost to the perspective soaring a mile high. Similarly, the *Zhuangzi* says that when you assert certain things or think some things are right (both referred to as *shi* 是) then you necessarily reject other things, and think some things are wrong (both referred to as *fei* 非). Rather than trying to find the most right way to view the world—and trying to make the world fit that view, as is often the case, the *Zhuangzi* suggests being able to take on various different (*shi-fei*) perspectives. Or, it says, one can realize, and affirm, that they temporarily lodge in one perspective or another. But if they do not grasp too tightly to it, if one does not identify or attach too much to any particular perspective, then they will go through the world more smoothly.

Above we already saw a similar observation in chapter two of the *Daodejing*. Calling something beautiful makes other things ugly, calling something good makes other things not-good, and when we categorize the world thus, we have a less nuanced view of things—the world, people, or interactions and relationships—than we would without the use of general categories. This is not to say we should abandon all categories, that would be completely absurd, but we should, these Daoist texts suggest, realize their inherent limitations. This logic is found throughout the *Daodejing* and *Zhuangzi*. Discussions of usefulness and uselessness in the *Zhuangzi* operate on the foundation of this type of observation. However, many of the stories related to the usefulness and uselessness make even more nuanced points.

One of the most quoted sections in the *Zhuangzi* related to use is about a special balm that could be helpful for keeping hands from chapping. A family had their own special balm, the ingredients and making of which was passed down for generations. For this family, it meant they could eke out a living bleaching silk. When someone with political aspirations saw this, he offered to buy it for a pretty sum. The family sold the method, reasoning that they would make more money from this one sale than their ancestors had made for generations. The man who then obtained this knowledge made the balm for an army, who were able to conquer their opponents, and the man was then enfeoffed lands and given titles by the ruler. For one group the balm meant making a meager living bleaching clothing and getting a modest sum of silver; for another it meant winning battles and being given great generational wealth.

This story follows from one where Zhuangzi’s best friend Huizi speaks of gourds given to him that turned out to be useless. Zhuangzi tells the story of the balm, and then says that however useless the gourds may have been in terms of conventional practices, since they were huge, Huizi could have cut holes in them and used them to float down a river. This is often taken to be a celebration of finding useless things useful. Though we might retort that floating down a river in a gourd is not only not really “useful” in any sense—what use could this have?—but also probably not desirable, even if one is just relaxing and enjoying the river. Read together, and the two stories are interwoven, we can also take the point to be that uses can depend on contexts, and on how one views things. And of course, whether something is deemed useful or not is entirely dependent upon who is using it, and why. Most importantly, when one holds a narrow view of how something should be used, and when they are not open to other possibilities, then whatever is

being looked at, the person themselves, and the interactions involved, will all be narrowly restricted as well.

This is all related to putting the world into categories—which are always categories one makes use of when viewing the world and operating as part of it. If they are loosely held, and a person does not attach too much to them, make them too rigid, or try to conform the world accordingly, then they can be relatively unproblematic. Of course there is no way to do anything without relying on categories to some extent. The *Zhuangzi* is particularly interested in the attitude one takes towards these categories. We find in it the story of a gardener who labors rather needlessly to water his plants. One day Zigong, a disciple of Confucius, sees the gardener and suggests a rather simple fix to the back breaking work of lugging buckets of water up to the plants. A simple water lever device could be constructed Zigong says, and this would not only save the gardener from chiropractic pain, but more importantly allow him to grow even more plants. A bigger harvest could mean selling his surplus and making a profit. The gardener scoffs. He wants no part in the machine. This machine would only transform his heart and mind into something mechanical as well. In other words, he might start thinking and feeling according to precise calculations about how much water he can pump, how many extra plants he can have, and how much he can make.

The claim here is not against selling plants for profit, nor is it a rejection of using technologies—although both readings have been put forth as well. Reading the *Zhuangzi* as a whole, and for the purposes of this article, we might simply note that the gardener is worried about being/becoming overly mechanistic. As shown with Zhuangzi's criticisms of Huizi's not floating around in a giant gourd, the *Zhuangzi* is not against relaxing, but nor is it advocating that people are always relaxing and working as little as possible for the most amount of profit. It simply provides a space where people might reflect on their knee-jerk reactions, and call into question some of their most familiar ways of seeing and interaction with the world, others, and even themselves. The old gardener simply might like carrying the pails, it gives him something to do, perhaps he finds it meditative, and he probably has little use for extra cash. We should not, however, attempt to find a specific point or some other goal which the gardener might be working towards. It is the constant search for such ends which edges one into mechanistic thinking, and eventually being.

The suggestions we find in Daoism often revolve largely upon being simple, or unadorned, and following what is “self-so” or “natural” without “acting-for”. The *dao* 道 of Daoism is normally used to indicate a path, way, or method, and has been read, at least since Wang Bi 王弼 (d. 249) as something like an ontological substance as well. Following *dao* is one shorthand way of thinking about what it means to be simple, unadorned, “self-so” and not “acting-for”. In several places the *Daodejing* and *Zhuangzi* say that *dao* is the undifferentiated everything. Anytime we make distinctions, including the use of language and ideas in the most basic sense, we automatically break *dao* apart, and move from it. The opening lines of the *Daodejing* quite plainly state that any *dao* which can be identified is not the *dao* it is speaking about, and both the *Daodejing* and *Zhuangzi* are suspicious about how much we tend to rely on names (terms, concepts, or language and ideas in general).

Another way to note the discussion given in these early Daoist texts about the relationship between names or ideas and the things they seek to speak about is that they are always incongruous. There is always a mis-match between what a name says, or what an idea conceives, or even how someone sees something, and that thing itself.

AI and algorithms break the world up into rather narrow, rigid, and clear-cut categories. There is little, if any, room for ambiguity in these technologies. The first point we might note from a Daoist perspective is that this leaves little room for appreciat-

ing the ambiguity of the world, and trying to develop are responses without assuming much certainty.

Furthermore, as we increasingly rely on AI, algorithms, and related thinking, we become more mechanical ourselves. Our thoughts, feelings, the ways we treat ourselves and others, develop so that there are specific means and ends, high degrees of knowledge, prediction, and control are assumed, and progressively desired.

Third, dipping into Daoist thinking helps us realize that there is too little focus on the attitude we take towards AI, algorithms, and other technologies. There would be less problems with bias and all sorts of unfairness if we did not rush to adopt these technologies so quickly, and thought more deeply about what they actually *do* do, and the spaces we might want them to be part of—and of course those spaces where we do not want them.

6. Conclusions

Our promethean drive, and the hubris associated with it, has long been part of the human experience. It may, as the sociologist Hartmut Rosa argues, have been multiplied in modernity, and grown even more in recent years (Rosa 2019). Our ability to know, predict, control, and engineer the world seems to be increasing almost daily. However, as Rosa also notes, even if we know when the first snow of the year might fall, we cannot know how that will make us feel. And, paradoxically, it seems that the more we know, predict, control, and engineer the world, the less we feel in touch with it. As our grasp on the world widens and deepens, our connection to it seems to fade. The feeling of snow suddenly falling—unexpectedly—and other similar experiences are things that many might consider exceedingly important to being human. A mechanical approach to the world can not only take away from magical experiences but also often leaves us blind to nuance. We can easily become deaf to meaningful particulars when we think we already know what we want and can reliably predict or control the world, others, or ourselves accordingly. This is clear when we think about mechanistic approaches to ethics.

In *What Money Can't Buy*, Michael Sandel considers the influence money has on our practices (Sandel 2012). He asks, for example, if when a school system offers to pay elementary students for reading books over the summer, they might not be encouraging the wrong type of thinking. Perhaps the students will cultivate an attitude towards learning or reading which will, in the long run, be more detrimental to their growth. We can extend this type of thinking to other even more foundational practices. Teaching a child to say “sorry” or “thank you,” for example, is supposed to develop something in or about them. We make them use these words, even when they do not mean them, so as to mature their sense of apology and gratitude. The hope is not that they will mechanically say them, and only in the situations taught. Rather, we hope to help children cultivate themselves to not only mean these words but also feel them and use them in novel ways as they approach a world of nuance and particulars.

Reading early Daoist texts can also help us gain a greater appreciation for the unknowability, unpredictability, uncontrollability, and unengineerability of the world. Like the gardener, we might find that there are things we are happy being oblivious to, or knowing only partly. Sometimes walking through the mud with a heavy bucket of water is exactly what we want or perhaps need. And often times selling everything we can, or all the (spare) time we have, is not a very nice life. Accordingly, we might reimagine our assumptions about technology, what it can do, and the roles we give it. This does not necessarily solve many serious problems, but it could dissolve quite a few.

More concretely, it might help enhance the type of design of AI and algorithms related to Russell’s “inverse reinforcement learning”. For example, instead of attempting to assist in maximizing the realization of human preferences, it might be more open to simply

enhancing them. And rather than being initially uncertain, it might remain uncertain, and there could be an allowance for a high degree of uncertainty. But most importantly, what is being explored in this essay is ways of changing our attitude towards AI, which would influence not only how it is designed, but also how we use it.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: There are no conflicts of interest.

Notes

- ¹ Some readers of the *Zhuangzi*, for example Ziporyn (2003), argue that what a thing is, or who someone is, is nothing more than a perspective. This is also a very informative way to think about the works of Guo Xiang (see Yang 2010).
- ² A more detailed discussion of *dao* is given below.
- ³ “Deep learning” refers to the ability algorithms have to train themselves and improve without direct intervention from humans (unsupervised learning).
- ⁴ It is important not to downplay the significance of these problems. Indeed, these issues have been some of the most contentious surrounding AI ever since it became successful in identifying photos and in a host of other areas. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (2016) by Cathy O’Neil was a pathbreaking study which opened the floodgates for rich discussions of all sorts of derisive biases embedded in these technologies. Below we will look at this in a bit more detail.
- ⁵ Partly for the above reasons, as well as others, Elena Esposito has proposed referring to these technologies as “artificial communication” instead of “artificial intelligence”. Algorithms and what we have termed AI are, she argues, are competent in communicating with humans. They are able to generate outputs which make sense to us, and can be useful in all sorts of ways. But this does not necessarily mean these technologies are “intelligent”. They rely on human data, and are restricted in all sorts of ways. See her book *Artificial Communication: How Algorithms Produce Social Intelligence* (2022).
- ⁶ This of course does not mean that Daoism rejects all human desires, intentions, or goals. As pointed out by reviewer #1, Chapter 36 of the *Laozi* includes a discussion of how one might achieve their goals. And indeed, if we read the *Laozi* as including advice for political rulership, then much of it can be read as strategic. This is how, for example, the *Hanfeizi* 韓非子 uses ideas from the *Laozi*. Unfortunately it is outside the scope of this article to consider how a more strategic understanding of Daoism could be discussed in the context of AI.
- ⁷ Criticisms in Daoist texts sometimes overtly criticize Confucianism, though the relationship is complex, and certainly cannot be accurately understood as directly rejecting Confucianism. See (D’Ambrosio 2020), for a discussion of how to approach these issues.
- ⁸ Commenting on an early draft of this paper, Michael Nylan suggests that much of this section is more relevant for the *Laozi* than the *Zhuangzi*. I will not rehearse all her comments here, but the point is well taken, and while I think there is room to develop a similar understanding from the *Zhuangzi*, this is certainly not the only way to read the text—in fact one could read the *Laozi* differently as well. For the purposes of this article I will simply use this way of reading the *Laozi* as a guiding inspiration for developing a particular way of thinking about AI.
- ⁹ It should be noted that the “moral thinking” the *Laozi* mainly targets in these chapters is Confucian. However, we can also note that some scholars have found inspiration from the *Laozi* to criticize morality in general (see Moeller 2009).
- ¹⁰ Michael Nylan suggests that “*dao*” here might be referring to “the great way of governing” (private communication).
- ¹¹ This is, of course, a complex issue. Daoist texts can also be read as suggesting the people act in certain ways to bring about specific results. The major difference is of degrees. The *Analekts* assumes a much higher degree of knowability, controllability, and predictability than the *Laozi* or *Zhuangzi*.
- ¹² This is not to say that these are states which are achieved and then simply become a person’s static way of being. The content and style of the *Laozi* (as well as the *Zhuangzi*) are better understood as suggesting that people can become better at becoming these ways, but that they constantly are them in a matter of degrees, and go in and out of these ways of being.
- ¹³ Depending on which text one focuses on, and the specific reading they are after, we can understand *dao* as something more singular, and something like “the way of the world and all things (including humans)” or as ways of doing specific things, such as “*dao* of archery”. Regardless of which type of *dao* we are speaking about, Daoist texts claim that they are, or it is, not fully

understandable and that we cannot hope to control *dao* or *daos*, even if we can act with them/it well. A more detailed discussion of *dao* is given below.

- 14 There are tons of examples that could be added to the four given here. To add just one interesting one, Rana El Kaliouby, who is especially vocal about biases and keeping human emotion in AI, speaks of the promise of “engineering serendipity”. (Fridman 2022, 1:23:00).
- 15 One of the most popular examples is the “paperclip maximizer”. This hypothetical example, which is highly metaphorical, worries that if humans asked a super intelligent AI or AGI system to make as many paperclips as possible, they might quickly find a world completely full of paperclips—and perhaps even humans have been destroyed to make more paperclips.
- 16 Some of the more prominent voices in this area include Elon Musk, Eliezer Yudkowsky, and Max Tegemark. One example of this movement, which includes a huge list of supporters, is the “Pause Giant AI experiments: An Open Letter” published in 2023, (see <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>) (accessed on 3 May 2025).
- 17 Guo also makes this argument, both regarding this passage and others. Many scholars, especially in contemporary academia, disagree with Guo’s assessment, or simply do not know about it, and ignore this possibility. For a more detailed discussion of this issue, see (Jiang 2019).

References

- Bentham, Jeremy. 1977. *Comment on the Commentaries and a Fragment on Government*. Edited by J. H. Burns and H. L. A. Hart. London: The Athlone Press.
- Brayne, Sarah. 2021a. Predict and Surveil: Data, Discretion, and the Future of Policing. Available online: <https://www.youtube.com/watch?v=Jo-3vRTPTDw> (accessed on 15 August 2025).
- Brayne, Sarah. 2021b. *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford: Oxford University Press.
- Carr, Nicholas. 2010. *The Shallows: What the Internet is Doing to Our Brains*. New York: W. W. Norton & Company.
- Chen, Guying 陳鼓應. 2020. *Laozi jinzhu jinyi 老子今注今譯 [The Annotated and Critical Laozi]*. Translated by Paul D’Ambrosio, Xiao Ouyang, Ady van Den Stock, R. A. Carleo, III, Joanna Guzowska, Lidia Tammara, and Dimitra Amarantidou. Boston: Brill.
- Christian, Brian. 2020. *The Alignment Problem: Machine Learning and Human Values*. New York: W.W. Norton and Company.
- D’Ambrosio, Paul. 2020. Confucianism and Daoism: On the Relationship Between the *Analects*, *Laozi*, and *Zhuangzi*, Part I. *Philosophy Compass* 15: 1–12. [CrossRef] [PubMed]
- Esposito, Elena. 2022. *Artificial Communication: How Algorithms Produce Social Intelligence*. Cambridge: MIT Press.
- Fisher, Max. 2022. *The Chaos Machine: The Inside Story of How Social Media Rewired Our Minds and Our World*. New York: Little, Brown and Company.
- Fridman, Lex. 2022. Rana el Kaliouby: Emotion AI, Social Robots, and Self-Driving Cars|Lex Fridman Podcast #322. Available online: https://www.youtube.com/watch?v=36_rM7wpN5A (accessed on 11 November 2022).
- Gans, Joshua. 2008. *Parentonomics: An Economist Dad Looks at Parenting*. Cambridge: MIT Press.
- Jiang, Limei. 2019. The Big and the Great: A Reconstruction of Zhuangzi’s Philosophy of Transcendence. *Religions* 10: 30. [CrossRef]
- Kant, Immanuel. 2002. *Groundwork for the Metaphysics of Morals*. Translated by Allen W. Wood. New Haven: Yale University Press.
- Kearns, Michael, and Aaron Roth. 2019. The Ethical Algorithm|Michael Kearns & Aaron Roth|Talks at Google. Available online: <https://www.youtube.com/watch?v=tmC9JdKc3sA> (accessed on 17 August 2025).
- Kearns, Michael, and Aaron Roth. 2020. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford: Oxford University Press.
- Moeller, Hans-Georg. 2009. *The Moral Fool: A Case for Amoralism*. New York: Columbia University Press.
- Ni, Peimin. 2017. *Understanding the Analects of Confucius: A New Translation of Lunyu with Annotations*. Albany: SUNY.
- O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.
- Rosa, Hartmut. 2019. *Resonance: A Sociology of Our Relationship to the World*. Translated by James C. Wagner. New York: Polity Press.
- Russell, Stuart. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Penguin Books.
- Sandel, Michael. 2012. *What Money Can’t Buy: The Moral Limits of Markets*. New York: Farrar, Straus and Giroux.
- Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf Publishing.
- Yang, Lihua 楊立華. 2010. *Guo Xiang Zhuangzi Zhu Yanjiu 郭象《庄子注》研究 [A Study of Guo Xiang’s Commentary to the Zhuangzi]*. Beijing 北京: Beijing Daxue Chubanshe 北京大學出版社.
- Ziporyn, Brook. 2003. How many are the ten thousand things and I? Relativism, mysticism, and privileging the oneness in the ‘inner chapters’. In *Hiding the World in the World*. Edited by Scott Cook. Albany: State University of New York Press, pp. 33–63.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.