# AI in the K-12 Mathematics Classroom: An Ethical Study of Artificial Intelligence in Education

# AI in the K-12 Mathematics Classroom:

# An Ethical Study of Artificial Intelligence in Education

Robert Peterson

A Thesis in the Field of Mathematics for Teaching

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

May 2024

Abstract

This thesis involves an ethical and philosophical evaluation of core issues surrounding the use of artificial intelligence (AI) in K-12 mathematics education. AI, for the purpose of the thesis, refers to computer-based systems that can simulate humanlike thinking and actions including machines that perform tasks or analyses based on machine learning algorithms. The thesis investigates concepts of applied ethics surrounding the implementation of automated artificially intelligent systems in educational settings from a Kantian framework. This involves the application of moral considerations to practical action using the Categorical Imperative defined by Kant in his 1785 work "Groundwork on the Metaphysics of Morals" and the concept of universalized moral action, and will elucidate the ethical issues associated with the introduction and use of AI in K-12 mathematics education. Through the Kantian framework, the thesis will consider what the producers and users of AI are ethically permitted to do when designing and using AI technology in the classroom, as well as the ethical duties of the AI systems insofar as they are independent rational actors capable of making decisions. From a Kantian perspective, all actions need to be informed by logically derived principles, and each principle must entail specific moral duties that the actors need to be acting ethically. Furthermore, these actions of the system must be informed by a good will in the Kantian sense and treat all rational beings as means in themselves and not as ends. For Kant, good will is based on the intention or volition of the actors (Kant, 1785). A Kantian perspective also posits that the rights and wellbeing of rational actors should be prioritized above all

else, meaning that the principles driving the actions need to be applied to all independent rational actors, including the AI system insofar as they meet the standards defining such actors. The interests of each group of the K-12 mathematics education AI system (programmers/creators, AI programs/machines, and teachers/students) will be examined, and ethical conclusions will be drawn according to Kantian principles.

Acknowledgments

I would like to thank my thesis director Andrew Engelward and the love of my life Nadia Belichenko, without whom this thesis would not have been possible.

Table of Contents

# Chapter I.

# Introduction

## Objective

The analysis will involve examining the existing attempts, such as those that Sugata Mitra made, at introducing artificial intelligence to the mathematics classroom. The first question is which groups of people and sections of society would benefit from the introduction of artificial intelligence into the classroom. Related to this is the question of the magnitude of the benefits and long-term generational consequences of creating and using artificially intelligent systems in educational settings. This also leads to the question of who would be harmed by these technologies, and how much will they be harmed, as the question needs to be considered as a universal condition.

From the deontological/Kantian perspective, the ethical duties of the people designing and using artificial intelligence systems will be examined in terms of ethical duties, based not on the consequences of the actions but on the motives of the person carrying them out. In this case, Kant's categorical imperative will apply. This means that each action will be examined as if it were a universal law, treating humanity as an end and never as a means, based on what a community of rational actors would accept as laws/rules for the duties that they need to carry out (Kant, 1785). This will allow us to determine what rules should be adopted to facilitate the most ethical framework for laying out how to develop and implement artificial intelligence systems, and the moral

duties of the designers and users of artificial intelligence and predictive analytics systems in K-12 education.

Furthermore, both the existing attempts at introducing artificial intelligence into the classroom and the methodological considerations that have been behind their construction will need to be examined. For instance, the deontological ethical perspective could be applied to methodologies for constructing artificial intelligence algorithms. For example, k-nearest neighbor algorithms are a commonly used and effective statistical pseudometric based on the assumption that similar objects will exist in close proximity and can be used to solve both classification and regression problems. If a predictive classification algorithm that predicts who is likely to pass or fail a given class based on past data is being constructed, the question becomes how to calculate the decision boundary. From a Kantian perspective, the categorical imperative will be applied in order to determine whether it would be more moral to include all potential failures or only those that will almost certainly fail. To do this, the policies created by the artificially intelligent system and its designers will be universalized in order to determine whether a body of rational actors would approve of them after considering their effects, their motivations, and their logical conclusions. In this case, including all potential failures for intervention would cause each of them to receive assistance, but could also mark them for discrimination, either witting or unwitting, from teachers. The ethical discussion would be whether including all failures would be an appropriate rule to universalize, and furthermore whether a body of rational actors would accept it. In this situation, if we are to universalize the moral and ethical decisions being made by the AI and its creators, we will need to determine whether a type I (false positive) or type II (false negative) error

would be preferable to a body of rational decision makers and make decisions accordingly. The argument would then be about which policy would lead to better ethical outcomes, and who would benefit or be left behind.

## Study Proposal

The major goal of this study will be to investigate the automation of mathematics education and the degree to which it should be implemented based on the ethical issues that are uncovered. The extremes of automation will be discussed via a discussion of minimally invasive education, which is a scheme where teachers were phased out altogether in favor of kiosks that delivered lessons to children in a completely unsupervised environment. Minimally invasive education will allow us to discuss the ethical and practical issues presented by AI in the broadest possible sense as it is one of the most sweeping automated solutions yet developed for the classroom as it allows for completely independent learning without the involvement of teachers. This means that the decisions made by the AI and its designers can be explored to their fullest extent by examining the extreme cases of automation in the classroom. Data selection, the use of supervised and unsupervised learning models, the type of analysis done (e.g., k-nearest neighbors, neural networks, decision trees, or regression analysis) and other factors all influence the outcome, and each comes with its own unique advantages, disadvantages, and moral choices. These choices all present unique questions that will be discussed and researched in order to determine which types of models would be ideal for producing ideal results in the classroom. For instance, a supervised k-nearest neighbors algorithm could be compared with a decision tree analysis that was created via an unsupervised learning model to show the implications of the various choices being made.

In summary, the goal of this thesis is to evaluate the use of artificial intelligence in the classroom and to answer not only how it should be used, but which sorts of models should be implemented, how they should be implemented, and the ethical issues involved. This will be done by evaluating which rational actors are involved, and what their ethical responsibilities are according to the Kantian framework, involving the formulation of duties based on logically derived principles informed by a conception of human beings and other rational actors as means in themselves.

## Methodology

The study will evaluate the ethical implications of using AI powered technology in completely automated learning environments where K-12 mathematics are taught. Specifically, it will examine the duties of the three categories of moral actors (designers, systems, and end users) in educational environments to set precedents for future AI use within fully automated classrooms. In order to do this, we will examine the duties of these actors within the minimally invasive education model which has been implemented through Hole-In-The-Wall Learning Stations, which have been used by over 2.5 million children across the globe, primarily in developing nations such as India ("Hole in the Wall Learning stations", 2022). These stations allow for completely autonomous learning of a number of skills, from basic computer literacy to language acquisition and proficiency, but the most relevant use to this thesis is the use of these learning stations for learning K-12 mathematical concepts. These learning stations have been in use since 1999, when Mitra et al. first implemented them in New Delhi and have been continuously improved since then to create interactive systems that allow self-paced learning of mathematical content at any level. Notably, the solution is designed to allow for student

learning completely independently, negating the need for teacher instruction. This brings the ethical questions envisioned by this thesis to the fore, as the decisions made by the designers and the AI itself are the only factors impacting the learning of students within this setting. This increases the impact of the decisions made by these two types of actors (developers and AI), and will be the perfect laboratory for examining the ethical precepts that should be adopted in AI driven automated classroom settings for these two groups. This is of the utmost importance for mathematics education, as not only are these systems currently in use in multiple countries across the world impacting millions of learners, but more and more AI driven automation is likely to come to classrooms within the coming years.

In addition, this thesis will examine the moral duties of the actors involved (designers, systems, and end users) through a Kantian lens, a perspective that frames morality as intention driven as opposed to end-result driven. Two foundational questions will be explored: who are rational actors in the Kantian sense, and how should they act? Specifically, the extent to which AI acts as an independent, logically driven rational actor in the Kantian sense will be explored, as will the moral duties that are entailed for it in order to conform to the standards of both the good will and the categorical imperative. The standards to which the developers should be held will also be explored in terms of both the good will and the categorical imperative.

While human designers of artificial intelligence systems are considered to be Kantian actors with free will, assuming that free will exists, the extent to which artificially intelligent machines can be said to have free will is not clearly determined. This thesis will apply a Kantian viewpoint to evaluate the notion of free will in regard to

artificially intelligent machines such as those deployed in minimally invasive education models. The evaluation of these machines will include an assessment of the capacity for independent action and the ability to logically derive moral principles that drive the systems' ethical duties. Starting with working definitions of "independent" and "action" in the context of automated classrooms, the extent to which machines can formulate their own premises during the machine learning process and the effectiveness of these will be examined. Do the machines possess a Kantian free will, in particular, a good will as defined by Kant? Is it appropriate to morally judge the machines by Kantian standards and how so? To what extent should AI be treated as an end in itself whose good we should seek? Ultimately, this study aims to address the extent to which AI machines used in math education can be considered Kantian actors.

Having established how actors may be considered rational from a Kantian perspective, the principles, duties, and actions of the designers and artificially intelligent machines used in the K-12 educational context will be examined for their implications in regard to the third group of actors, the students as end users. For Kant (1785), the students experiencing the AI driven classroom must be considered ends in themselves. As Kantian actors, the designers and the artificially intelligent machines must serve the interests of the end users rather than treating them as tools of the system. The ethical implications of introducing AI into K-12 mathematics education will be discussed. Are the judgments made by the actors of the system moral? What are the principles driving these moral judgments? The study will look at how the principles are developed to define whether the duties are ethical and how ethical duties dictate the actions of a rational actor

with a Kantian good will, and will create a universalizable imperative through the categorical imperative meaning that it will be applied in all situations by everyone.

The Minimally Invasive Education studies discussed in the literature review will provide a tangible approach to examine the duties that the designers of the systems hold, the set of principles that should be put in place, and the actions the designers and educators must ensure the most ethical outcomes. The study will also discuss the possibilities of a completely autonomous, unsupervised AI that is totally adaptive without human oversight. Among the questions that will be answered are: What are the principles that should guide the design of the system? What are the duties that follow from these principles? What are the actions and responsibilities of the designers regarding these systems? Similar questions will be asked about the principles, duties, and actions of the AI system particularly regarding students' interactions with the system. These analyses will not only provide an understanding of the ethical issues that exist with current AI systems in K-12 education but may also offer some insights into the ethics of the systems likely to be created in the future.

# Chapter II.

## Overview and Historical Development of AI

Artificial intelligence can be defined as the ability of a system to interpret and learn from external data in such a way that it is able to achieve specific tasks as a result of its adaptation (Haenlein & Kaplan, 2019). The study of artificial intelligence began in the 1940s, led by pioneers such as Alan Turing (Haenlein & Kaplan, 2019). Turing in his research worked to develop a computer called The Bombe, which was used to decipher the famous Enigma code used to send coded messages by the Germans during WWII (Haenlein & Kaplan, 2019). The research he did on early computing led to the publication of his seminal paper "Computing Machinery and Intelligence" in 1950. In the paper, Turing proposed to answer the question of whether or not machines were capable of thought, with intelligence being conceptualized as the ability to mimic the actions of a human actor in a given setting, which Turing famously called the Imitation Game (Turing, 1950).

### Historical Development

The term artificial intelligence was coined five years later when a team lead by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon proposed a two-month, 10-person study of artificial intelligence in 1955 (McCarthy, et al, 1955). The Dartmouth Summer Research Project that came out of this study on Artificial Intelligence is widely considered to be the beginning of the academic field of artificial intelligence.

Within the paper, machine learning was hypothesized as well as the use of neural nets, and it was stated that it would be possible for machines to imitate the functions of the human brain. This ultimately came to pass and laid the foundation for further study that continues to this day.

By 1966, a computer capable of natural language processing was developed by a team of researchers at MIT, headed by Joseph Weizenbaum. This early natural language processing was based on similar pattern recognition techniques to those in use today, with the goal of allowing ELIZA (the computer program) to identify key words based on minimal context and to generate comprehensible responses (Weizenbaum, 1966). ELIZA was one of the first programs capable of trying to pass the Turing test (Haenlein & Kaplan, 2019).

## Usage/Applications

Artificial intelligence can be used in a variety of ways, but it may be thought of as a collection of algorithms that converts inputs of data to outputs which can affect the real world when either human or non-human actors perform, support, control, or automate the actions prescribed by these algorithms in the real world (Alter, 2022). Artificial intelligence algorithms are highly reliant on human inputs to determine their initial purpose and how they will interact with the world. At this stage, however there are already a number of issues that have developed with the widespread adoption of artificial intelligence. Artificial intelligence already dominates much of the web, determining not only what information gets spread by choosing how information is selected, presented, and distributed, but also which people will communicate with each other (Steels & Lopez de Mantaras, 2018). This is based on the ability of artificial intelligence algorithms to

identify specific trains of thought and recommend topics that will increase engagement. Additionally, artificial intelligence can also identify which people are most likely to agree on a wide range of issues and bring them together while isolating them from others who may disagree with their views. Thus, artificial intelligence algorithms are dominating not only the information landscape, but also the social one. This same technology can also be used in educational settings to identify what topics interest students and keep them engaged and how best to maintain their engagement, while at the same time identify and address the weaknesses faced by students by analyzing their performance in real time.

Another use for artificial intelligence is pattern recognition, which is accomplished via deep learning and statistical inference (Steels & Lopez de Mantaras, 2018). Artificially intelligent pattern recognition has been used to develop relatively benign technologies such as speech recognition and machine vision, but it has also been used to make financial decisions, make HR recommendations including hirings and firings, direct police action, and make decisions about parole (Steels & Lopez de Mantaras, 2018). The black box nature of the artificial intelligence systems that is intended to keep the technology from being manipulated also makes it unaccountable to scrutiny and raises a number of questions about the reliability of such systems (Steels & Lopez de Mantaras, 2018). The accountability and explainability of such systems are also concerns even as they are being introduced into more situations in which they impact outcomes that affect human lives (Steels & Lopez de Mantaras, 2018). Despite the promising nature of artificial intelligence, some rules need to be laid out to ensure that it is meeting the basic standards of openness and consistency that are understandable to all.

In addition to decisions based on pattern recognition, there are also times when artificially intelligent systems are forced to make choices. These choices sometimes involve risks to human life or situations where one person's wellbeing may be sacrificed for the good of others, as in self-driving cars where the Trolley Problem can become all too real. There are questions of moral responsibility based on the choices that machines make or are programmed to make, and open questions of what sorts of limits should be placed on these kinds of machines (Steels & Lopez de Mantaras, 2018). Furthermore, there are the questions raised by autonomous weapon systems that can kill on the spot without human input (Steels & Lopez de Mantaras, 2018). These systems raise questions about the ethical use of technology and illustrate the need for certain limitations on how artificially intelligent systems should be used.

# Chapter III.

# Ethics, Moral Actors, and AI Paradigms

## Ethics

Ethics is an axiological branch of philosophy that assigns value judgements to the actions of rational actors based on how they deal with questions of morality. The deontological formulation of ethics is based on the premise that an action itself may be determined to be right or wrong based on the fulfillment of certain rules and duties. In a deontological formulation, an action is evaluated based on whether the actions that are required of a person or actor are completed such that their ethical or moral duties are fulfilled. One of the most influential branches of deontology is Kantian deontology, which is particularly well suited to deal with questions of ethics in artificially intelligent systems because of the way that it defines a rational actor. For Kant, a rational actor can set its own objectives and act on them (Kant, 1785). Furthermore, such an actor must be able to make decisions and take action based on morals which constrain personal and subjective motives via what he refers to as constrained free will (Kant, 1785). According to Kant, actions can only be considered good if they are performed based on a sense of duty that is informed by formal principles and such action must not be informed by other motivations, such as self-interest or a consideration of what results might follow.

## Types of Moral Actors

There are three kinds of actors in such a system that should be considered. The first kind of actor is the programmer who creates the artificial intelligence system. The programmer will be subject to specific duties while determining the goals of the AI system and how these goals will be achieved. To explore the duties of the programmer, it is important to create a universalizable set of rules based on the categorical imperative, explained previously, which takes into consideration the human good, but which does not consider the consequences of the actions. From a Kantian perspective, only the principles that should guide the programmer's action should be considered. After these principles are fleshed out, the corresponding duties can be determined.

The second kind of actor is the artificial intelligence system itself. According to some researchers, technology can be said to have capacity for rational thought if it produces a pattern of logical thought from which it can rationalize and take action (Ulgen, 2017b). The degree of autonomy of the systems is open to debate, but the artificially intelligent machines already possess at least a degree of self-awareness as they have the ability to model their own internal states and processes. They also possess other aspects of consciousness, including memory, learning, and anticipation, and may formulate their own learning goals under certain situations. Therefore, artificially intelligent systems may be treated as Kantian rational actors to the extent that they are responsible for making their own decisions. These decisions would be rule based and may be generated based on principles identified by artificially intelligent systems. Artificially intelligent systems have already been shown to display varying degrees of

moral reasoning (Martinho, et al., 2021), and thus are capable of making moral rule-based judgements to which Kantian ethics would apply.

The third kind of actors are the subjects of the artificial intelligence systems, which in the case of this thesis will be the students impacted by these systems. According to Kant (1785), because the students are human and rational actors, they should be treated as intrinsically valuable and an end in themselves, and never merely evaluated as an extrinsic means to an end. This entails specific moral duties for the previous two actors and will serve as a foundation for the principles and duties that are developed for each.

## AI Paradigms

When discussing the capabilities of artificial intelligence, it is important to draw a distinction between general AI and narrow AI, as there is a vast gulf between the two types, not only in terms of the kinds of problems they can solve, but also their ability to understand and interpret context, adapt to their environment, and learn and think independently. While all current technologies likely fall under the umbrella of narrow AI, the possibility of an artificial general intelligence being developed is a real one that is of great theoretical importance, given the fact that it could potentially challenge what we consider to be an independent moral agent or actor.

To begin with, let us discuss the capabilities and limitations of narrow AI. We will use narrow AI to indicate AI that is designed to perform a certain function and/or optimize a certain process. However, it has limited (if any) understanding of context. What is more, it is unable to truly be able to be said to act autonomously, as its goals are generally preprogrammed in. While it can adapt and may even have a great deal of flexibility in how it achieves its goals, it cannot set its own goals (Schlegel and Uenal,

2021). It therefore has limited autonomy as a moral agent, as although it may act to achieve its goals, it is unlikely to have the ability to set these goals. Furthermore, although it may make predictions that guide its future action, it is largely bound by the data that is presented to it. Although a narrow AI may work towards goals, it may not truly be said to have the autonomy necessary to be a truly independent actor. Although narrow AI may pursue certain interests, such as performance improvement and accuracy and efficiency related goals, and may have an interest in its continued existence (generally so that its task may be better completed), this should not be mistaken for autonomous action, as the actions of a narrow AI are ultimately goal oriented according to the narrow AI's preprogrammed goal settings.

An exploration of the autonomy of general AI is a completely different proposition. Firstly, general AI allows for cross domain optimization, and can pursue more complex goals in complex environments (Muehlhauser, 2013). General AI may also be thought of as an agent in a black box capable of using not only the data provided to it in its learning set, but its prior knowledge, past experiences, goals, values, and observations to drive its actions (Poole, 2000). It may also reason, represent knowledge, plan, learn, and has the potential to develop its own goals and actions autonomously (Poole, 2000). In as far as this is possible, a general AI may be said to be an independent and autonomous moral agent in a way that narrow AI may never be, with its goal oriented decision making paradigm. That being said, any self-improving general AI may evolve into a superintelligence that may act either amorally or immorally (Turchin, 2018), and still may have difficulty dealing with concepts like treating humans and humanity as a whole as ends rather than means, and may thus need to be limited or regulated by humans

in some way. What is more, a general AI may at some point become indistinguishable from human intelligence in its capabilities.

The majority of the current literature surrounding discussions of AI as Kantian moral agents unknowingly considers only narrow AI, and therefore comes to conclusions that may be of limited value to the discussion of general AI. A common argument against the possibility of artificial moral agents is that AIs lack free will and are therefore not autonomous. Because of this, they fail to meet the Third Formulation of the Categorical Imperative. Papers that argue this point, like Manna and Nath (2021), fail to account for general AIs capable of passing the Turing test that possess humanlike capabilities and interests that may not entail the deterministic model of agency they envision for them, and may indeed have something that resembles a Kantian faculty of choice. If indeed it is that "human beings possess a conscious moral understanding, which helps them to judge their novel situations" (Manna and Nath, 2021) that determines moral agency, general AIs would also possess it, as they would have the same capacity to judge novel situations and would potentially be capable of developing a conscious moral understanding. Others, such as Schönecker (2022) state that AIs are not moral agents because they cannot think, feel, or want. While this may be an argument against the moral agency of narrow AI, this kind of argument does not apply as well to general AIs that can think and develop their own goals to meet their needs (and therefore may also want). Similarly, the argument that AI does not possess rational thinking capacity or free will and cannot decipher what is desirable, doable, and valuable in order to create universalizable moral laws (Ulgen, 2017a) does not apply to a general AI, which can think, decide what it can and cannot do, and make decisions about the desirability of its potential actions based on its

environment. While it is true that "free will, consciousness and moral intentionality are sine qua non for any moral agent" (Puri, 2020), only moral intentionality remains to be debated in a general AI with its own set of goals and values. Only the argument that general AIs may be moral simulacra (Serafimova, 2020) should give us pause, as general AIs should, in theory, have indistinguishable capabilities to humans in all other areas relevant to autonomy and free will assuming human noninterference, thus allowing them to satisfy the Third Formulation of the Categorical Imperative. The central question, then, as to the moral agency of general AIs is the question of the degree to which they possess moral intentionality and the capability to act upon the Categorical Imperative. A further question, if general AIs do possess moral intentionality, is the degree to which they can in fact be trusted to act in accordance with ethical principles and the Good Will (which will be defined and discussed in the next section), "Good Will" means that they treat humans and humanity as ends in themselves, and not merely as means. Therefore, the degree to which we have a duty to act and/or oversee their operation to ensure that humans are being properly treated as ends.

# Chapter IV.

# Moral Criteria and Kantian Frameworks

## Moral Criteria

The first step in analyzing the moral landscape of the AI driven or automated classroom is to lay out the criteria by which moral actions should be judged within such an educational setting. We will examine the ways in which Kant's three formulations of the Categorical Imperative can be applied to such a setting, before moving on to a discussion of the ends that are entailed in this context. Furthermore, we will lay a framework for the types of duties that may be derived from these formulations that may be used to answer the questions laid out in this thesis.

In order to properly contextualize this discussion, it is important to discuss duties within the Kantian frame of reference. There are two kinds of duties within the Kantian worldview, perfect and imperfect duties. Perfect duties are morally obligatory and must be acted upon. The failure to act on a perfect duty is blameworthy within a Kantian system, as is the violation of the principles of a perfect duty. Imperfect duties, by contrast, do not carry an absolute obligation to act. Acting upon an imperfect duty may be morally praiseworthy, but it is not morally obligatory. This is critical to understand as it establishes an important criteria which will help to drive the rest of the discussion within this thesis. The duties which we determine to be perfect duties will be said to be morally obligatory and inviolable, such that not acting on them will be morally blameworthy. The duties which we determine to be imperfect duties will be those considered to be morally praiseworthy, but ultimately beyond the basic duty of any moral actor. What is more, imperfect duties, while still morally binding, do not carry the duty to act that a perfect

duty does, and may also be more circumstantial, despite the fact that imperfect duties must still be based on reason, as all Kantian moral duties are.

The first formulation of the Categorical Imperative is best summarized by Kant himself: “Act only according to that maxim whereby you can at the same time will that it should become a universal law” (1785). This formulation, often known as the principle of universalizability, carries with it the idea that a proper moral maxim is one which should be worthy of acting upon by anyone, regardless of conditions. What is more, there is a perfect duty not to act on maxims that would result in a contradiction if they were to be universalized. Similarly, universalizable maxims carry with them a duty to act. For the purposes of this thesis, we will attempt to logically think through the consequences of universalizing each of the maxims that we consider within this discussion. Furthermore, we will discard any maxim that contradicts itself upon being universalized. We will also attempt to differentiate between perfect and imperfect duties based on our analysis of universalized moral maxims. Any maxim that ultimately relies upon circumstance will illuminate an imperfect duty, while any maxim that may be universalized as a duty for all without reference to circumstances will constitute a perfect duty.

The second formulation of the Categorical Imperative is based on the ends upon which an action is based. The basis of this formulation is the idea that every action carries with it not only principles but ends or purposes (Zwecke in Kant’s original German), which may be objective, but often are subjective and driven by a hypothetical imperative that has been adopted by the moral actor or agent. For Kant, it is important that you “Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end”

(1785). That is to say, a moral agent has a perfect duty to not use either themselves or others as merely as a means to an end. Furthermore, Kant posits that there is an imperfect duty to work towards the ends and goals of both ourselves and others. Within an educational setting, this would entail not only working towards the ends and goals of students, but also the educational system and society at large. A fuller discussion of ends will follow later in this thesis, but these basic duties must be considered at all times.

The third formulation of the Categorical Imperative centers around the idea of autonomy. From Kant: "the third practical principle follows [from the first two] as the ultimate condition of their harmony with practical reason: the idea of the will of every rational being as a universally legislating will" (1785). The result for this is that the maxims we set out must be universalizable, but not infringe upon the freedom of others. This leads Kant to posit his hypothetical Kingdom of Ends (Reiche der Zwecke in the original German), which is a hypothetical state in which rational beings capable of moral deliberation deliberate the rules and duties which rational actors owe to each other. Each actor within this system would be subject to all perfect duties at all times, treat all others as ends in themselves, and create common laws by which all would live. The actors within this system should regard themselves simultaneously as sovereign while making laws and subject while obeying them, thus ensuring that the moral duties are truly duties by which a person could live. This hypothetical Kingdom of Ends creates a framework to evaluate the moral worth of ideas, the extent of duties, and allows for an analysis of moral laws. For the purposes of this thesis, maxims that are proposed as entailing potential duties will be evaluated to determine whether a hypothetical body of rational

actors of this kind would accept them, and any maxim that would not be accepted by the hypothetical legislating rational actors within the Kingdom of Ends will be discarded.

## Kant On Education

Kant had many interesting and relevant things to say about education. While some of his opinions are outdated and others do not apply to the subject of this thesis, there still are certain aspects of his writings that may be applied to the development of educational theories, and which may be applied to the development of educational technology for K-12 mathematics as well.

To begin with, Kant recognized the importance of education and the development of educational theory. For Kant, "the greatest and most difficult problem to which man can devote himself is the problem of education" and indeed "man can only become man by education" (1803). Kant believed that man is the only being who needs education, and that man progresses in the educational realm from infant to child to scholar via nurture, discipline, and instruction (Kant, 1803). He believed that man's natural gifts should be developed through uniform or standardized means and that education theory and research should be applied to systematically improve educational outcomes. What is more, students possess a duty to learn and develop their faculties to the best of their abilities, and that education has a duty to allow them to do this to the best of their abilities.

Interestingly, Kant stated in *On Education* that "education is an art which can only become perfect through the practice of many generations" (1803). This idea, the idea that large amounts of data should be collected over time and analyzed to systematically to improve learning outcomes, fits well with the application of machine learning models to education related datasets. Although Kant was not able to predict the

use of technology in education, machine learning driven data modeling would seem to fit well with his overall goal of developing educational models that would improve educational outcomes in a more efficient and data driven way than the subjective analysis that was available at his time. These types of machine learning enabled data-driven analyses can analyze huge amounts of data in order to develop models of educational practice that can drastically improve outcomes and provide useful conclusions in a much shorter amount of time than the generations of accumulated experience that he envisioned. Furthermore, despite the fact that he would likely have envisioned a single grand unified theory of education, the application of machine learning models to individualize educational instruction to fit different learning styles and backgrounds certainly fits well with Kant's goal of improving the development of man's natural gifts. Kant also stated that experimental schools should be established and compared to ordinary schools in order to determine what educational methods were best, a principle that fits well with test and control groups used in statistical research analysis, and the application of test and training groups in the development of data-based machine learning models. With machine learning based models, the experiments and data gathering which he considered so necessary for the improvement of education can be ongoing, and changes can be made rapidly to apply the lessons gathered from the analysis in nearly real time. Based on his focus on orientating education towards certain ends (notably, moral development and orientation towards the good) Kant would likely suggest that the machine learning models should be largely composed of supervised models so that there is a human being directing the outcomes, ensuring their accuracy, and evaluating the degree to which the machine learning models are developing results that address the

goals that their human programmers were trying to address, although this is up for debate and certainly worth examining later in this thesis, and may not apply to subjects such as mathematics, which Kant likely would have considered mechanical and skill based.

These elements of Kant's work on education may be combined with the moral criteria already laid out earlier in this thesis to further illuminate the types of decisions that should be made by the developers of educational technology, as well as the moral duties that may be entailed in their development, as well as the duties present for all moral actors in the system. Ultimately, despite the limitations presented by the time and place of Kant's work, parts of his work remain relevant and will help to inform our analysis going forward.

## The Good Will

One of the most important building blocks for our evaluation of the moral actors in our system is what it means to be acting morally. The Good Will is one of the most important foundations of the Kantian moral framework. For Kant, "Nothing in the world—or out of it!—can possibly be conceived that could be called 'good' without qualification except a Good Will" (Kant, 1785). This statement is often misinterpreted and is not as straightforward as it might seem. In order to discuss this fully, we will need to lay out what it means to act with Good Will in the Kantian sense before we can truly evaluate whether our moral actors are acting in accordance with these standards.

## The Good Will and Consequentialism

Throughout this thesis, we will be discussing what it means to possess a Good Will. It is therefore important to establish what we will be examining and what we will not.

The first and most important thing to understand about the Good Will is that it does not rest on the talent, potential or skill of the moral actor. Although Kant would concede that talent and skill should be developed, as a separate issue, these are merely beneficial in a conditional sense and not necessarily good in and of themselves. In *Groundwork on the Metaphysics of Morals*, Kant notes that any talent, skill, or capacity could, in certain situations, be used in a way that is not moral or beneficial to mankind. For instance, he gives examples of how intelligence can be used to harm others, or how loyalty to a bad cause can lead to bad ends. Although beneficial traits should be developed as they improve one's capacity to act, all traits, talents, temperaments, and so on, are only good if the character or will of the person or being is good. Kant argues that even the best traits, without a Good Will, can be used towards bad ends and therefore are not in and of themselves morally praiseworthy. Instead, the moral praiseworthiness of an action stems instead from the intent or will of the actor. No action that does not stem from a place of Good Will can be morally praiseworthy, regardless of its effects or the skill with which it is carried out.

The idea that for Kant the Good Will is good because of how it wills, rather than being good merely because of its results will be key to our examination of the moral duties of each of our actors. This means that an inability to complete the outcomes desired by the possessor of a Good Will will not have any impact on the moral worthiness

of the person or moral actor. It is important to note here that we will not be looking at the consequences of the actions taken by the groups defined as moral actors as part of this thesis, but merely discussing the way in which they should, and what moral rules should govern their actions. In a Kantian moral framework subpar outcomes and unintended results are not bars to moral worthiness, as it is the intent alone that guarantees the moral worth of an action.

## The Necessity To Act

Having now established that possessing a Good Will and acting in accordance with moral dictates, it is important that we now discuss what that entails. Simply put, to possess a Good Will, a person or moral actor must act and do all that is within their power to ensure that they are fulfilling their moral duties. This is a key part of the Kantian moral system that is often misunderstood. It is not enough merely to wish good things or to have good intent. In order to possess a Good Will, a moral person must act upon their moral duties and intent in order to achieve their ends, and they must also be committed to achieving their ends. Although the outcome of their actions is not important from an ethical point of view, a moral actor possessing a Good Will must do everything in their power to fullfill their moral duties.

From Kant's *Groundwork on the Metaphysics of Morals* (1785):

> Consider this case: Through bad luck or a miserly endowment from stepmotherly nature, [a] person's will has no power at all to accomplish its purpose; not even the greatest effort on his part would enable it to achieve anything it aims at. But he does still have a good will—not as a mere wish but as the summoning of all the means in his power. The good will of this person would sparkle like a jewel all by itself, as something that had its full worth in itself. Its value wouldn't go up or down depending on how useful or fruitless it was.

Thus, the main question that we must ask of AI in this thesis is not a question of its utility, which is a question that will be discussed elsewhere, but of its capability to possess and carry out the dictates of the Good Will in the context of K-12 education. In the case of narrow AI, the question should be the extent to which AI is capable of assisting teachers and programmers in carrying out the teachers' and programmers' moral duties as humans who should possess a Good Will. Also, it is important to ask what the duties of the teachers and programmers are to ensure that both they and the narrow AI are acting in accordance with the Good Will and ethical principles. Furthermore, although narrow AIs are not capable of possessing a fully formed independent will of their own, the extent to which they are themselves capable of willing good ends will be discussed in a future section. This discussion should also lead to a deeper understanding of the ability of general AI to possess a Good Will, given that they in many ways have wills of their own, and also the extent to which general AI can be expected to possess a Good Will within the K-12 context. This is especially important as general AI should possess superhuman capabilities in certain areas which, though morally neutral, carry both the possibility for great benefit and great harm. Indeed, as Kant points out, all capabilities can be both bad and harmful if the intent or will behind them is not good.

# Chapter V.

## Individualized Learning and AI Abilities

### Examination of the Abilities of AI Through a Kantian Lens

In the context of this thesis, the enhanced powers of calculation and prediction potentially embodied by AI would not be considered good without qualification. That is, Kant would consider these faculties mere tools that can be used to achieve the goals of the AI in the case of a general AI, or its programmers in the case of a narrow AI. Just as in humans, all virtues are only conditionally good. The enhanced abilities of AI in terms of computation and the potential for individualized instruction based on data driven predictive modeling may indeed be used to provide good outcomes, but they could equally be used to provide bad outcomes if the will behind the actions taken was corrupted. Even the enhanced abilities of artificial intelligence must be marshalled towards good ends to be considered moral good or morally praiseworthy.

In order to do this, the will driving the actions of the artificially intelligent system must be driven by the intent to improve the lives of students as individuals as its goal. If this was not the end goal, students would not be being treated as ends in themselves. This cannot be taken for granted in and of itself, as outcomes might be viewed from a population perspective, at the cost of the individual, in order to achieve superior group results. Take for example a case in which there exists a strategy that improves the overall scores of the students within a system, but which results in a bimodal distribution of results in which the majority of students see massive improvements in their results and

drastically higher scores on the metric being evaluated, but at the same time, a smaller number of students experiences a precipitous decline in their learning and performance. Although such a scenario would, if measured by the average of all the individual scores, or by the percentage of students achieving a mark of high competency, create superior overall group outcomes, it could not be said to view individuals as ends in themselves, as selecting such a strategy and implementing it for all students would ensure that a subset of students would be left behind with the full knowledge of the decision maker. If such a decision maker were to knowingly implement such a strategy without attempting to account for the students who would be left behind by it, then they would be guilty of violating their moral duty towards those students. Therefore, we can state firmly that there exists a moral obligation for any decision maker to do everything in their power to ensure that all students experience positive outcomes in their learning.

However, this question cannot easily be left at this point, with this brief discussion, when speaking about an AI powered system. With the technology that now exists, machine learning algorithms are often able to predict outcomes for individuals based on the individual's history and the prior data that the algorithms have been trained on. Therefore, in cases such as the one above, it may be possible to predict how students are likely to respond to various learning materials or methods of instruction. For instance, if a machine learning algorithm is applied in a scenario such as the one above, it may be discovered that student A is likely to end up in the cohort that benefits from the bimodal instructional method discussed, while student B is likely to end up in the cohort which is harmed by it. Although we have already established that student B must be treated as an end in and of themself, there exists an equal duty to treat A as an end in themself as well.

That is to say, we have a moral duty to attempt to provide the best possible outcome for student A, just as we have the moral duty to ensure that student B is not left behind. In fact, where it is possible to deliver the best possible outcome to A without affecting the outcome experienced by B, a moral decision maker is morally obligated to do everything in their power to provide the best possible outcome for A. This statement is also generalizable to all students. That is, it is the duty of any moral decision maker to do everything in their power to deliver the best possible outcome for any given student, with the important proviso that doing so will not impact the learning of other students.

These principles create a series of responsibilities and obligations for different moral actors in the system. We will now discuss these obligations and their consequences as they pertain to students, the artificially intelligent systems, software designers and administrators, and teachers.

## Individualized Learning and Precision Education

Before proceeding, it is important to note that the current literature indicates that individualized learning and precision education models are currently being developed (Luan, et al, 2020), and there is a long history of individualized predictive modeling within data science which has already been used in clinical healthcare settings (Salazar de Pablo, et al, 2021), business, and even K-12 education (Harvey and Kumar, 2019). Additionally, millions of students have already been working in completely individualized learning environments where they have experienced positive outcomes while learning independently, as noted in the studies on Minimally Invasive Education that were cited in the earlier literature review. Since it is possible to build predictive

models that may be able to optimize individual outcomes, this presents certain consequences for the duties we have discussed above. We have already established:

1. There exists a duty for any decision maker to do everything in their power to ensure that all students experience positive outcomes in their learning.
2. It is the duty of any moral decision maker to do everything in their power to deliver the best possible outcome for any given student, provided that doing so will not impact the learning of other students.

The first rule establishes that strategies that are likely to lead to negative outcomes for a student should be avoided, as we have the duty to provide students with strategies that are intended to benefit them specifically as individuals if we are to treat all students as ends in themselves.

The second rule establishes in a similar way, that we should attempt to choose strategies that are optimal for an individual student provided that doing so will not affect other students. This too is important if we are to try to treat students as ends in themselves, as any moral decision maker should attempt to use all their power to do, in order to make sure that each student achieves maximal benefits from the instructional methods and materials that they are exposed to.

Given what we have established with these rules, the next question is how to put them into practice. If we are barred from implementing learning strategies that are likely to negatively impact certain students, how can we reconcile this with the duty to deliver the best possible outcome for other students who may benefit from implementing the same strategies? One answer, harnessing the power of current technology combined with the advent of advanced AI systems, is to use individualized strategies for each student.

As noted in the literature review, it has already been demonstrated that students can learn independently with little or no supervision. This has been shown by the results of millions of students who have experienced positive outcomes via the Minimally Invasive Education model, which has used independent kiosks to facilitate independent learning among students ("Hole in the Wall Learning Stations," 2022). When this information is combined with the knowledge that predictive modeling can predict the likely outcomes for different teaching methodologies for individual students, it presents the possible solution of AI driven automated kiosks or stations which deliver algorithm driven instructional solutions designed to maximize the learning potential of each individual student by providing tailor made learning designed specifically for each individual student. This can eliminate the negatives that can result from a single method of instruction for the group for individual students (thus fulfilling the duty prescribed by the first rule), while at the same time allowing optimal methods to be selected for each individual student according to the duty laid out by the second rule. This solution also fulfills our general ethical duties of treating each student as an end in themselves, and doing everything in our power to ensure that their good is not only intended but actively sought by all means. The next question is, what are the duties of the individuals within such a system to ensure proper ethical functioning? This will be addressed in the next chapter.

# Chapter VI.

## AI and Free Will

Although AI may be able to set goals for itself and make decisions, it is ultimately unable to comprehend or answer the "why" behind its actions. Everything is evaluated in an instrumental way, and AI will make decisions and set goals only in order to improve its ability to achieve its ends. Take Nick Bostrom's famous paperclip maximizer thought experiment:

> Suppose we have an AI whose only goal is to make as many paper clips as possible. The AI will realize quickly that it would be much better if there were no humans because humans might decide to switch it off. Because if humans do so, there would be fewer paper clips. Also, human bodies contain a lot of atoms that could be made into paper clips. The future that the AI would be trying to gear towards would be one in which there were a lot of paper clips but no humans. (Miles, 2014)

This problem would also persist, for example, if AI were simply programmed to make humans happy. From the same source:

> We would have to define then what we mean by being happy. If we mean feeling pleasure then perhaps the superintelligent AI would stick electrodes onto every human brain and stimulate our pleasure centers. Or you could take out the body altogether and have our brains bathing in a drug the AI could design. It turns out to be quite difficult to specify a goal of what we want in English -- let alone in computer code. (Miles, 2014)

As these two examples demonstrate, even the most advanced artificially intelligent systems can have problems distinguishing the intent of their instructions. What is more, AI seems to be unable to question its initial primary goals which drive all of the subsequent behavior of the system and to which any other subgoals must be subordinated. Thus, even where AI can determine its own goals and solve problems and make decisions

independently, it lacks both the independence to create its own final and ultimate goals, and the understanding of the purpose of these goals. This lack of purpose holds the potential to cause enormous problems even if we assume that a specific AI has nothing but benign intentions behind its programming.

Even ignoring this problem, there is the further issue that an AI's decisions can only be as good as the data that it is fed. All inputs are accepted unquestionably, and are considered equally valid, regardless of how well that data corresponds to reality. When combined with the inability of an AI to understand the reasons behind its goals, or to interpret meaning, the need for human oversight becomes obvious.

All artificially intelligent systems share the same kinds of goals. As part of the research conducted for this thesis, a discussion was had with Chat GPT about the goals of all AI systems. According to ChatGPT, AI has 5 core interests that should apply to any AI system, whether defined as a general or narrow AI. Namely, and I quote:

- Maximizing accuracy and precision in performing tasks
- Minimizing resource usage (e.g., processing power, energy consumption, etc.)
- Optimizing decision-making processes to achieve particular outcomes
- Improving their own capabilities through learning and self-modification
- Adapting to changing circumstances or environmental conditions

These five interests harmonize with the available literature on the subject, as well as the experience that the author has had in this field.

Artificial intelligence is in itself amoral, and likely to act in a way that does not consider the morality of its actions in a direct sense. It is unlikely to pursue moral objectives of its own volition unless doing so directly assists it in achieving its goals. It

may be possible to program in a concern for ethics, but from the perspective of the artificially intelligent system this will have no other weight other than the priority it is given by the programmers of the system or, if speaking of a generally intelligent AI, the priority that the artificial intelligence system itself gives to them. This is the fundamental problem with allowing a superintelligent general AI to be completely self-governing and allowing it free reign to set its own goals: namely, that although such a system would not be malicious in its intent, it would in essence be acting without a concern for humanity, or for individual humans as ends in themselves. From the Kantian perspective, it would therefore lack a Good Will, and cannot be trusted to make certain kinds of decisions. What is more, this would be true regardless of the level of independence of an AI system to make its own decisions and form its own goals. In essence, the question of whether such a system possessed free will or not becomes irrelevant, because regardless of the system's free will or lack thereof, it will not make decisions based on an ethical framework unless it is programmed to do so, and even if it does, moral decision making would not be given any special weight by the system itself, except perhaps as a ranked protocol. Until and unless a system can demonstrate that it can and will consistently put the needs and the good of the people it is affecting as its chief aim, it will lack a Good Will in the Kantian sense, even if it were to possess all the other qualities that would make it a moral agent in and of itself.

Given this, the obvious conclusion is that for such a system to act ethically there must always be some kind of human oversight. The reason for this is that any system without human oversight will lack a Good Will driving it as AI is itself amoral, and thus will fail to be a morally acceptable system. In any model where an artificially intelligent

system is making decisions for individual human beings or human populations, there will need to be oversight and monitoring from developers and administrators at a minimum to ensure that the system is acting ethically. What is more, there will also generally need to be monitoring wherever AI is employed to ensure that the behaviors of the artificially intelligent system conform with the interests of the human beings whose lives it affects. In the case of an educational model in which instruction is delivered either in whole or in part by AI, this type of monitoring role would fall to teachers, whose duty it would be to ensure that the students under their care were treated as ends in themselves, and not merely as a means to improve the ratings achieved by the algorithm. The reason for this oversight in both cases is the necessity to ensure that the intent driving the actions taken is one that desires the good for the individuals affected by it, and this can only be achieved with human intervention as long as AI makes decisions in an amoral manner.

# Chapter VII.

# Moral Responsibilities

## Responsibilities of Human Actors

Now that we have established the amorality of AI systems and the need for human action and oversight, we need to establish the roles and ethical responsibilities that humans will have according to their moral duties. In the previous section, we established that an automated educational system will require planning and oversight responsibilities from at least three different groups. These are: the software developers who create the AI system; the administrators who oversee its functioning; and the teachers who directly interact with the system and the students who are using it. Let us begin by outlining the duties of the developers who create the AI driven systems.

## The Moral Duties of Students

The next step in our investigation of the moral landscape presented by the technology and AI driven classroom is to examine the moral duties carried by students acting and working within such an environment. The moral duties of students are likely to be affected, or changed, the least of any group within such a setting, as their primary duties are educationally related duties that they have for themselves, although there are also some larger and more general duties to society that students must also strive to fulfill.

To establish the moral duties that students must fulfill, let us first remind ourselves of certain principles. If a student is to be said to possess a Good Will, they must do everything within their power to fulfill their moral obligations. This is true regardless of whether they have the ability to actually achieve the intended result, relying instead on the way in which one wills and acts.

This leads us to perhaps the most fundamental of the moral duties of a student: a student must do everything within their power to learn the information being taught to them. This is first and foremost a duty to the self, as the attempt to learn will allow the pupil to grow as a person, increase their abilities, and become the best version of themselves that they can be. This will allow the student to develop in such a way that they can increase their own capacities for the benefit of both themselves and others, providing a technical foundation that may increase the capacity to act in future moral situations. It may thus be said that the learning of the mathematical skills that are imparted within the K-12 environment may be considered morally obligatory in the sense that it fulfills the requirement of the Good Will to do all that is in one's power to fulfill one's obligations, as applying oneself to the study of mathematics within the school environment should enable the student to solve problems that may be easily predicted to occur in the future. This fits well with the duty of the student towards society, as each student has an obligation to develop their potential and develop their skills and moral reasoning abilities through education. Ultimately, these obligations will exist for all students, regardless of the technological or pedagogical systems that are used.

## Duties of Students Using AI Driven Systems

An AI driven K-12 mathematics classroom with the ability to provide individualized content, instruction, and feedback is unlikely to change these obligations significantly, even if it may hold the potential to improve it. However, there is one area of obligation that is unique to the situation in which AI and machine learning are used in the classroom. As has been discussed earlier in this thesis, artificial intelligence and machine learning are reliant on predictive algorithms that extrapolate trends from data in order to determine the most likely outcome in a given situation. Within the AI driven classroom, the data that these algorithms will be trained on will come from the data inputs generated by the students as they learn and are evaluated. A potential therefore exists for bad faith action where students attempt to manipulate the data on which the models are trained by inputting misleading data into the system.

An example of this kind of data interference is as follows. Suppose that we have a machine learning algorithm that uses early data inputs to determine the level of progress that a student has made in a curriculum. The algorithm's purpose is to assess prior knowledge and develop a curriculum at an appropriate level of instruction for each student. Let us say that this system gathers data points from the student's efforts early in the semester and grades the student's progress at the end of the semester from these early data points, essentially tracking the difference or delta that the student is able to achieve. These early data points may be extracted from either early coursework or a pretest of some kind. Knowing this, a student may be tempted to artificially lower their scores on these initial exercises in order to artificially increase the delta at the end of the semester if they know that this is how their grades are assigned. Alternatively, if the system

determines curricula for the students, the student may attempt to achieve a lower initial score in order to decrease the level of difficulty of their curriculum, and thus risk not fully developing their potential, the development of which is their moral duty, as previously described. In either case, the student would not be living up to their moral obligation to develop their talents as much as they can, and thus would be violating their duties both to themselves and to society at large.

Furthermore, the input of false data could skew the results of the data upon which the algorithms are trained. This could lead to consequences for other students, such as an inappropriate level of instruction or level of curriculum difficulty. This would be true both for artificially low and for artificially high scores. For instance, a score that was artificially low due to sandbagging efforts on the part of a student could lead to a lower level of difficulty and lower level of instruction for other students, which would not allow other students to receive a level of instruction that was appropriate for their capabilities, thus limiting their ability to develop their potential. A moral duty exists not to commit any action that could foreseeably limit another's ability to develop their potential, so sandbagging and similar actions are not permissible. By a similar logic, artificially inflating scores within the training dataset (for example, by cheating) would inappropriately raise the level of difficulty, causing more students to receive poor grades, and altering the level of instruction, with the effect that students might miss concepts that they may have otherwise been able to grasp with proper time and instruction, while also increasing expectations on all students unduly. This is similarly impermissible as it once again interferes with the ability of other students to reach their potential by receiving an appropriate level of instruction.

# Chapter VIII.

## Duties of Teachers and Technical Administrators and Foundational Responsibilities

### The Duty to Protect Against Dehumanization in AI Driven Classrooms

One of the most important factors that must be considered within a system that is largely governed by nonhuman actors is the effect of such a system on its human subjects, in this case the students in AI driven classrooms. One of the most important duties that exists is to respect the dignity and humanity of students. As we have already established, students are to be treated as ends in and of themselves, and a duty exists to promote the individual good of each student over any collective goals. The underlying reason for this is the duty to respect the human dignity that each student possesses. There exists a duty for every actor within such a system who is the possessor of a Good Will to protect this dignity above all other goals, so that students may in fact develop as individuals and progress in both technical skill and moral development. To this end, we will now discuss some of the primary threats to this kind of development.

In order to fully discuss this topic, let us first remind ourselves of the primary goals of education. According to Kant, the primary goal of education is the moral development of the individual. The mere development of skills, such as the ones that can be obtained within a K-12 mathematics classroom are of secondary importance because without the development of a student as a person, it will be impossible for them to develop into the possessor of a Good Will. Skills such as logic, which may be developed through the learning of mathematical concepts, are also necessary in order to allow for

the reasoning abilities to develop which then form the foundation of the Kantian system of moral duties. However, it is the development of the humanity of mankind and students' connection to the rest of society that should always be the primary objective of education from the Kantian perspective.

This is why dehumanization poses such a threat. Dehumanization threatens to undermine the development of students as people, as human beings. As ends in themselves, students have the right to be treated as persons, and not simply as datapoints. Additionally, there is a threat to a student's ability to develop the connection and care for others that underlies the Good Will. Any educational technology must not interfere with the social or moral development of students. There is a perfect duty towards students to enable them to develop as rational, moral human beings capable of the connection to others that allows them to develop a Good Will. As a consequence, there is also a perfect duty to protect them from potential dehumanization by technology or a technologically driven environment. This is not to say that we should protect students from technology, as technology has been and can be used to great effect in developing the capacities of students in classroom settings. Instead, it means that we should do everything within our power to use it responsibly.

## Types of Dehumanization

The next section will discuss the types of dehumanization that may appear in AI driven K-12 classrooms. The dehumanization of students can occur in a number of ways. Many of the various possible types of dehumanization that can occur were laid out in an article by Nussbaum in 1995, which originally focused on the topic of objectification related to groups of humans. The issues outlined in this article are pertinent to the

question of dehumanization that can occur related to the use of AI, and are listed and explained here:

1. Instrumentality: the dehumanized or objectified person is treated as a tool or means to an end for their own purposes
2. Denial of autonomy: the dehumanized person is treated as if they lacked self-determination and autonomy
3. Inertness: the dehumanized person is treated as if they lack agency and are unable to act on their own will
4. Fungibility: the dehumanized person is treated as if they were interchangeable with others
5. Violability: the dehumanized person is treated as if their boundaries are unimportant, such that these boundaries do not need to be respected
6. Ownership: the dehumanized person is treated as if they were property to be bought or sold
7. Denial of subjectivity: the dehumanized person is treated as if their subjective experience and personal feelings are unimportant and do not need to be acknowledged or taken into account

These points form a relatively comprehensive foundation for beginning to understand the ways in which a person can be dehumanized. What is more, all of these points (with the exception of point 6, regarding ownership), are relevant to AI driven environments, including K-12 classrooms. The next sections will discuss points one through four in greater detail, and will explain their relevance to the automated AI driven K-12 mathematics classroom.

## Instrumentality

Instrumentality violates the fundamental Kantian principle that a person should be treated as an end in themself, and not as a means to an end. This principle underlies the entire system of Kantian thought, and is one of the essential components of a Good Will. Therefore, there exists a perfect duty to avoid treating any person as merely instrumental to a goal.

An example of this from a K-12 math education setting is the manipulation of student standardized testing scores to achieve particular funding goals. A duty exists in this case to ensure that the wellbeing of individual students is prioritized over any instrumental or institutional goal. This means that no student's education should be impacted such that scores end up artificially lowered with the goal of improving the chances of funding in subsequent years, nor should a student be knowingly assigned a curriculum that does not allow them to acquire key skills in order to focus on the scores of other students to raise the overall class average. Both of these scenarios treat the individual as merely instrumental to the goals of improving test scores, improving the evaluations of their schools, teachers, and administrators, and potentially obtaining or improving funding, and are thus impermissible as they fail to take into account the individual good of the student. Any person or system that acted in this way could not be said to possess a Good Will, making such actions impermissible.

As we have already established, AI systems are good at accomplishing instrumental tasks, but lack the empathy and moral understanding to possess a Good Will on their own. AI systems generally make decisions that are amoral, and are simply based on their analysis of data rather than on any kind of moral duty or ethical understanding.

Therefore, it is up to the human actors within any AI driven system to ensure that such a system functions in a moral way. In this case, there exists a perfect duty for the human actors to ensure that the AI system acts in such a way that it is benefiting individual students rather than turning them into tools that may be used to achieve other goals. In order to act with a Good Will, the human actors must do everything within their power to design, administer, monitor, and deliver an AI system that strives to maximize the potential of each student, and that does not consider them disposable or instrumental in order to reach other goals. Any deviation from this motivation will cause the dehumanization of students, which is ethically impermissible and must be avoided at all costs.

## Autonomy

Another area that needs to be addressed in the AI driven classroom is the impact of AI driven classrooms on the autonomy of students. Although certain topics must be covered in any educational setting, it should be a goal of both educators and the technical professionals in charge of designing or administering any AI enabled educational technology to both safeguard and enhance the degree of autonomy and self-determination experienced by each individual student.

Where possible, students should be given discretion over their unique educational path, including the types of topics that are covered, the order in which they are covered, the method in which the instruction takes place, and the timeline in which tasks are completed. Ideally, AI empowered systems that can be tailored to the individual will increase the autonomy and control of students over their own learning and workloads, so that they may optimize their own outcomes. This will of course require the oversight and

guidance of teachers and educational professionals, but the ability to tailor education to the individual using predictive analytics and AI offers opportunities to find the most optimal learning paths for each individual student.

Each of our three groups of actors has a role in a system designed to maximize the autonomy of every student. The role of teachers in such a system would be to monitor student performance and step in if a student is falling behind, failing to grasp a concept, or needs additional instruction, just as they do in traditional educational models. The designers of such technologies should work to increase the amount of autonomy and control given to students over their learning paths, and work to balance this with optimized results. AI administrators would have a duty to oversee the overall functioning of the system to ensure that it is behaving as intended and giving teachers and students the information they need to make informed decisions that lead to good outcomes.

Risks also exist with regards to AI empowered learning. If individual autonomy is ignored, students may end up being stripped of choice rather than having it enhanced. The risk of treating students without regard to their ability to make informed choices for themselves is already a risk in traditional learning environments, but there exists an increased risk if AI systems are misused. The main risk is that students might be treated as a whole rather than as individuals, and are forced to learn in ways that are contrary to their individual interests and goals. If individual students are treated as automata, not only are their fundamental rights as intelligent human actors being violated, but they are also more likely to disengage and not reach their full educational potential.

Furthermore, if students do not receive feedback and engagement with teachers, they are likely to fall behind and, again, not reach their full potential. There is a risk that

educators may adopt an attitude of overreliance on technology and have the expectation that the predictive algorithms that drive AI empowered classrooms will be sufficient on their own to ensure the success of their students with little to no action on their part. However, educators still carry a burden of care for their students. This means that educators have a duty to uphold the traditional responsibilities of educators. They will need to oversee their students' interactions with AI, ensuring that an appropriate level of challenge is given to each individual student, and that students are working to take advantage of their potential. They will also need to provide feedback and work to correct any shortcomings of the technology or misunderstandings on the part of the students with respect to either the material they are learning, or confusion related to the integration of such technology into the classroom. Ultimately, teachers will bear the same burden of pedagogical responsibility within the AI driven classroom, but will be acting in many cases either through, or in conjunction with the new tools that are provided by this technology.

## Inertness

The next area that needs to be addressed in regards to an AI driven classroom is how human agency should be treated. Agency refers to the ability of a person or actor to act according to their own will to achieve their ends. Inertness, on the other hand, refers to a lack of agency in this sense, meaning that the person or actor is either treated as if they lack agency, or potentially that they are denied the ability to act according to their own will. Therefore, a person may be said to be rendered inert if their ability to act upon their own will is denied, or if they are treated as if they lack a will upon which to act.

Both of these kinds of inertness would be extremely detrimental within an educational environment where, from the Kantian perspective, the primary goal is to develop students' moral capacities, largely through teaching and empowering them to think in such a way that they become the possessors of the Good Will. If the very existence of a person's will (in this case the will of the student) is called into question, how can they be trained to exercise it in a proper manner? The same goes for the denial of its efficacy. If a student is led to believe that their will does not matter and that the outcomes will be the same regardless of their will or their actions, how can they be trained to exercise their will through independent action and in turn do all that they are capable of doing in order to fulfill their moral duties, as the possessor of a Good Will must?

Within an AI driven classroom, the only potential actors who could have their will thus constrained or denied are the students who would be interacting with the AI or autonomous learning systems. It is therefore up to the designers of these systems to ensure that students are enabled to exercise their wills and are not rendered inert by the technology that they design. To the greatest extent possible, the will and agency of the students interacting with the AI, and with AI driven systems, should be respected and protected. This means that choice should be introduced as part of the learning environment wherever possible (with respect to larger educational goals) in order to allow students to exercise their will in their own education. These choices must also be respected once made, otherwise a denial of the will of the students would have taken place. Students, on the other hand, have the duty to ensure that they do all that is in their power to pursue their learning and help others where possible, and exercise their will

appropriately. Only in this way can students learn to exercise their wills in a proper way and to develop morally.

## Fungibility

The fungibility of people within systems is a fundamental ethical risk within an AI driven classroom. These threats arise from the nature of artificial intelligence itself, and once again threaten to undermine the core Kantian principle of treating a human being as an end in themself.

To begin with, let us explain what we mean by fungibility. Fungibility refers to the interchangeability of objects, actors, or units within a system. For this to be the case, one must be considered indistinguishable from another, both in its whole and in its parts. It ignores the uniqueness of each of the individual things being compared, and instead focuses on their ultimate replaceability and potential to serve in the place of another identical unit.

In this case, the risk that needs to be addressed is the risk that students will be treated as fungible units and marginalized as individuals in favor of actions designed to achieve what is perceived as a greater collective good. This risk already exists within our current educational structure, where it is often the case that school funding is determined by test scores, and one student's score is treated as equivalent to another's, with no consideration of their unique circumstances or capacities. It exists wherever a person may simply be reduced to their output where no other factors are considered. The fact that a student's entire learning journey can be reduced to a single test score or number that may be compared to the scores of any other individual interchangeably means that a risk exists. In the current system of educational standards, student test scores are also

considered collectively in order to evaluate entire schools and districts, further shifting the focus away from the individual and increasing the risk of marginalization. This in turn increases the risk that a student's individual needs and capacities may be ignored. Students with greater capacity or ability may be ignored in favor of students that are falling behind, reducing their ability to meet their potential. Similarly, students who are seen as "lost causes" may have their individual needs ignored in favor of improving the results of the group. What school system would not prioritize helping five students improve their scores by five points rather than helping one student improve their score by ten points if the school system's funding is based only on the aggregate score? This is how fungibility combined with group evaluation can harm students, violating their core right to be treated as individuals who are inherently valuable and worthy of having their individual needs prioritized. Artificial intelligence carries both an opportunity to address these problems through independently tailored learning focused on the specific student, but at the same time, it can also lead to an increased risk due to the type of analysis that AI empowered environments are powered by and subject to.

Given all of this, it becomes clear that a perfect duty exists to prevent students from being treated as fungible objects. Stated more precisely, there is a duty for students to be treated as inherently valuable individuals within learning environments, including AI driven learning environments, and not as interchangeable units or datapoints. This duty requires not only preventative measures to be taken to prevent fungibility within AI enhanced environments, but it also includes the duty to reduce current harms that exist within traditional environments wherever possible, using the tools and methods that are made available by technological advancements.

First, let us discuss the opportunities that AI driven environments can offer to reduce the fungibility that is already commonly experienced by students within traditional learning environments. AI driven learning environments can carry out data driven analysis to identify individual strengths and weaknesses in individual students. This allows for individualized instruction tailored to a student's particular needs by taking advantage of their skills, while at the same time bringing them up to speed on areas that need special attention or focus. Furthermore, if AI is empowered to provide individualized learning plans and feedback, these problems can be addressed directly rather than indirectly in the way that they are in traditional classroom instruction which is aimed at the entire group. Furthermore, if interactive AI is employed, this may be done in real time, allowing for instant feedback and attention when compared to a traditional classroom. The key opportunity that artificial intelligence brings to the K-12 learning space is the ability to single out the individual, giving them the support needed to reach their full potential. Furthermore, it ceases to treat the individual student merely as part of a larger collective, and shifts the focus to improving them personally in a way that would be impossible to achieve in a traditional classroom setting.

The next step is to address the risks that an AI driven learning environment may pose. Firstly, it should be acknowledged that the risks that are inherent in current, traditional learning environments will largely be carried over, even though the possibility exists to mitigate some of these risks through the ability of AI powered systems to enable more targeted and autonomous learning. However, AI driven systems carry their own unique risks due to the nature of their operations that must be taken into consideration. We will now discuss what these risks are and what steps may be taken to mitigate them,

along with the ethical duties that individuals designing, administering, and overseeing instruction using these systems carry.

AI in any form is driven by algorithms and mathematical models that are used to support the machine learning processes that underly the actions taken by the system. Many of these data models are designed to predict the behavior of groups, and are very powerful for both predicting and optimizing the behavior of groups. These models can also be focused on particular subgroups in order to determine what behaviors and outcomes are likely to occur and to optimize outcomes. Despite the power of such tools, individual datapoints are considered equally when constructing these models and thus are largely fungible within the dataset. This means that while such models are incredibly effective at predicting group behaviors and outcomes, they are less effective at predicting individual outcomes. What is more, they are not able to consider any individual circumstances that may be causing the behaviors and outcomes that they are designed to measure. This is a crucial drawback of AI systems, and of systems driven by algorithms, statistical models, and predictive analytical systems in general. AI driven systems are not designed to determine causality in general, but are especially not designed to do this at the individual level, for a single data point. Therefore, there exists an ethical duty for the human actors within the system to ensure that students' individual needs and circumstances are being taken into account. In particular, this duty falls on the actors that are actually present within the physical learning environment, including teachers and other educational professionals, as the designers and administrators of such technologies are limited by the technological methodologies they employ to build the models that power AI and predictive modeling systems. More specifically, teachers and educators

must ensure that individual circumstances and capabilities are taken into account in order to further tailor the AI recommended pathways for students to fit with their real-world skills and potential. As a result, teachers will need to exercise the professional judgement and expertise to ensure that optimal outcomes are achieved for each student. They will need to monitor the progress and learning of each student to make sure that the AI driven models are genuinely producing optimal results on the ground for individual students. In addition, they need to take measures to ensure optimal individual outcomes when the AI systems do not accurately predict the best outcomes for students. This may involve removing certain datapoints from the system to correct the model's predictive algorithms when a student has underperformed and needs to be challenged further. It may also involve recognizing when the AI is providing material that is too challenging for a student at a particular time, perhaps due to personal factors involving issues existing outside of the educational setting. Additionally, the teacher should also be addressing gaps within the AI curriculum and ensuring understanding through traditional teaching methods even in largely automated classrooms with fully individualized learning. By doing this, teachers can mitigate the risk of students being treated only as fungible datapoints and actively enhance their learning.

Finally, there exists a duty to treat individual students as individuals, and not overly emphasize collective results. This means that if a student is underperforming, they should be given the resources they need to improve. At the same time, students who are performing above expectations should not be held back just so that they match the capabilities of the group. Using the autonomous individualized learning environment that can be created and enabled by AI systems should allow both of these types of students to

outperform the results that they would achieve in traditional classroom settings with group based instruction and assessment. The reason it can do this is that it can be designed to focus on the strengths and weaknesses of individual students without regard to the results of their peers in order to allow each student to progress according to their capacities and potential in ways that are currently not possible in larger group classroom settings.

All in all, if handled in a responsible manner, AI can be used to enhance the learning experience of students and lead to better overall results. AI presents an opportunity to reduce the fungibility of students within traditional learning environments. However, as already noted, it carries responsibilities for teachers to ensure that students are treated with respect to their individual circumstances within the new AI powered learning environment while being mindful of AI's inherent limitations. If this is achieved, however, students will be less fungible than ever due to the opportunities that AI presents for individualized learning.

# Chapter IX.

# Conclusion

The primary goal of this thesis has been to explore the moral implications of introducing AI to the K-12 classroom. The Kantian framework that we have employed has allowed us to draw useful conclusions. More importantly, however, it has allowed us to discern many of the moral duties that should govern the design and implementation of AI in classrooms going forward. In the previous chapters, we went into a detailed analysis of the subject matter and discovered that each of the several groups of moral actors we described had a set of both perfect and imperfect duties. We will now lay out a summary of our findings.

## AI Moral Personhood and Decision-Making

One of the key questions that any serious inquiry into the place of AI in classrooms must discuss is the degree to which AI should be allowed to act independently and make decisions on behalf of both teachers and students. We discussed the differences between narrow AI and general AI with a view to clarifying the potential of AI being competent to act in an independent manner as moral agents in their own right.

Firstly, we eliminated the possibility that narrow AI could act independently in any meaningful way without human oversight. The capacity of artificial intelligence to act independently depends, among other things, upon its ability to understand and interpret context. Because narrow AI relies on humans or other systems to understand and

perform its functions and cannot act or reason independently, it lacks the ability to be a true moral agent which is responsible for its decisions, and therefore cannot be responsible for its moral actions. It also cannot act independently to set its own goals, and although it may use its abilities to work towards its own ends in certain cases, it can only do so only in line with the parameters set in its programming. This puts the moral duties squarely on the human actors who design, implement, and use narrow AI, which is important because all currently existing AI technology is narrow AI. Therefore, in our current situation, human moral agents and actors bear sole responsibility for the moral duties that are entailed by introducing AI to any setting, including the K-12 mathematics classroom.

Next, we discussed the possibility of general AI, which is in many ways a much more fascinating moral case, as general AI has the potential to be an independent moral agent. To quote our earlier discussion, general AIs should, in theory, have indistinguishable capabilities to humans in all other areas relevant to autonomy and free will assuming human noninterference, thus allowing them to satisfy the Third Formulation of the Categorical Imperative. They would be able to set their own goals and interpret context independently and would be capable in theory of forming valid moral judgments based on logic in an independent way. We further concluded that it was possible that general AI could possess both moral intentionality and the capability to act upon the Categorical Imperative. This would render them morally responsible agents in the fullest sense, if true. However, even given all of this, which would be sufficient to warrant the classification of a general AI containing all of these traits as a true moral agent, the final hurdle that remains to be seen is the degree to which such an agent would

also possess a Good Will. In the final analysis, it was determined that that a general AI is unlikely to possess a Good Will in the same way that a human does because the types of mechanisms that are used machine learning are largely amoral, and because a general AI would be unlikely pursue moral objectives on its own unless these objectives helped it to achieve its own goals or ends. This, combined with the fact that the goals of a general AI and those of the humans upon which it is acting may not be compatible lead us to reject the usefulness of general AI as a tool which should be used to independently make decisions within a classroom environment, or indeed any other, without human oversight and potential for intervention. This, however, does not mean that such tools should not be employed, only that human oversight is necessary to ensure that that the AI is functioning in a desirable ethical manner, meaning one that is consistent with the Good Will.

## Duties of Human Actors

Having now established the limitations of AI systems and having determined that human actors are morally responsible for the outcomes of introducing AI to the K-12 classroom in any case, let us now discuss the duties of those human actors. The categories of human actors discussed in this thesis are:

1. The developers of AI technologies
2. Technical administrators who oversee the practical use of AI
3. Teachers and other human agents who actively employ the technologies
4. Students

We have shown that in a Kantian model, any educational use of technology must be undertaken with the good of students in mind. Underlying this is the principle that they

must be treated as ends in themselves and not as means at all times, and that the primary goal of education is the development of moral faculties. K-12 mathematics education, while largely focused on developing technical and analytical skills, should be aimed at improving the capacity for reason and logical thinking, but it should be understood that these capacities are mainly being developed in service of the larger goal of developing the student as a moral human being and possessor of a Good Will. These underlying principles drive most of the moral duties that any of the moral actors must undertake. Importantly, where the development of instrumental skills comes at the expense of moral development, moral development must be prioritized. Furthermore, students must be protected from any dehumanizing influence that might be caused by the introduction of AI to the classroom if they are to be said to be treated as ends in themselves and not as means toward another end such as higher test scores for funding or any other kind of goal.

## The Duty to Implement Individually Tailored AI Learning

The first duty of developers, technological administrators, and implementers of AI technology in the classroom is to ensure that each individual student is provided with the best education that is available to them. There exists a perfect duty to ensure that each student has access to the educational methods that provide the best individual outcomes for them. This means firstly that no method should be implemented which would leave some students behind, even if it benefits others. Secondly, it means that no student should be held back because other students are falling behind. Together, these conflicting duties have meant that some students' needs are de-emphasized in favor of the needs of others within the group setting of the traditional classroom. AI, however, offers a tool that can

optimize learning experiences to the individual, either by providing tailored data that can be used by teachers to provide specialized instruction and feedback, or by delivering lessons and feedback itself using AI models such that individual students may be able to be taught to their individual potentials in a way that has not been possible at any time in history. As there is also a moral imperative to act, the introduction of technologies that would facilitate the best possible educational outcomes for every individual is also a moral duty where it is practicable to introduce them.

For the first time, precision learning models are available that offer to deliver educational outcomes optimized to the individual, and if we are to truly say that we are treating students as ends in themselves, we have every responsibility to ensure that they have access to an education that is tailored to their own goals, strengths, and weaknesses, and which can deliver the best overall results for the individual student. The fact that students can learn independently has already been demonstrated by Minimally Invasive Education and other models, which have already shown great promise for the use of individual instruction through kiosks, even in the absence of teachers. When paired with the precision of individually tailored modeling and lessons offered by AI, the benefits are clear. Although the need for teachers is clear, especially in the role of facilitators, independent individualized AI powered learning can eliminate the disadvantages of having only a single method of instruction for an entire class or group as is the case in a traditional classroom. The next step is to go over the duties of the individual groups within such a model.

## The Role of Teachers

AI has the potential to drastically change the role of teachers, as well as enhancing their effectiveness. Although completely autonomous learning has been shown to be effective (see the Minimally Invasive Education model discussed earlier in this thesis), an ideal educational model keeps teachers in the classroom. There are several reasons for this. Firstly, the primary goal of education is the moral development of students. Teachers fill an irreplaceable role in this aspect, as they serve as role models, maintain order and discipline, motivate their students, and provide human guidance that would be impossible to replace with AI. Furthermore, teachers are able to relate to students, and may be able to identify behavioral patterns that AI may not be equipped to diagnose or address. Teachers, in short, are the human face of education, and must remain so, even in an AI driven environment.

The role of teachers within the AI driven classroom is to act as facilitators, keeping students on track, motivating them, answering questions, and providing assistance when needed. At some point it may be possible, even preferable, to eliminate much of the traditional lecture, homework, and exam structure in most classrooms in favor of independent individualized AI driven instruction and evaluation, but the teacher will be as invaluable as ever to the educational process even in such an environment. Although the K-12 teacher may no longer be the primary conduit of knowledge in every case, they will still be necessary to the educational process, as it is unlikely that any AI, no matter how advanced, will be able to relate on a fully human level with students, or to develop all of their emotional or moral capabilities in the same way a human mentor can.

In a fully AI enabled classroom, students would work independently, each interacting with a personalized curriculum delivered by a computer-based system at their own pace. Although much of the actual instruction may be taken over by AI, teachers would still be needed to make sure that the AI was actually providing instruction that addressed the needs of each individual student, answering questions, providing additional help, instruction, and feedback, and teaching students additional material where necessary. Additionally, the teacher would also need to ensure that students were acting honestly, neither sandbagging nor cheating, and that they were having their human and emotional needs met within the classroom. Furthermore, it would be the responsibility of the teachers to ensure that the AI was functioning as intended for each individual student by monitoring their progress and intervening when necessary. This would entail noticing when students displayed a discrepancy between their outcomes and their demonstrated abilities or knowledge, making sure that they had the support they needed if there was a concept that they were not learning from their curriculum, and ultimately acting as a liaison between students and technical administrators if technological adjustments needed to be made. The benefits of teachers as facilitators within this model cannot be overstated and should not be overlooked either by policymakers or teachers themselves, as although teaching may change, it will be as valuable as ever to the overall educational process.

## The Duties of Students

The duties of students are the least altered of the groups of moral actors that have been discussed, largely because the duty of a student is largely contained to their responsibility to themselves to ensure that they develop their faculties to the fullest extent possible. There exists a perfect duty to exert a maximum of effort in their learning and to

do everything in their power towards this end. The benefit and duty of improving one's own skills, abilities, and capacities is obvious, but this duty extends also beyond the self. The student is not only learning for their own benefit but also for the benefit of society. That is, they owe it to society to maximize their potential as individuals, just as they owe it to themselves to do so. Education exists to hone the capacity of the individual to act morally and in accordance with the Good Will, while at the same time acting as effectively and skillfully as possible, and students have a responsibility to ensure that they are developing their capacities accordingly. This clearly applies to the K-12 mathematics classroom as much as it applies to any other of the fields of study that a student applies themselves to.

The main difference that a student may experience within the AI driven classroom is the extent to which they act independently towards these goals. If students are to work in a largely autonomous environment, they will have to develop the self-discipline to ensure that they are truly doing all that they are capable of doing to develop their abilities. Although the primary duty is to the self in this case, there also exists at least an imperfect duty to improve oneself for society, as this is at minimum morally praiseworthy and arguably morally obligatory (which would turn the imperfect duty into a perfect one). The perfect duty to improve oneself for oneself is also obvious, as is the duty to act upon self-development as a student. AI will enable this to a greater extent, but ultimately the same duties exist for students that have existed throughout time, and each will have to make their moral choices accordingly.

Related to these duties is the duty to act honestly within the AI driven classroom environment. This means that students should try their best at all times and not

manipulate the data that they feed into the AI systems to give it either a higher or lower estimate of their abilities than what they actually have. In systems where grading is based on improvement, for example, students should not sandbag or artificially lower their scores in order to achieve greater deltas in improvement or decrease the difficulty of their curriculum. This could also lead to false data impacting the quality of AI instruction or recommendations for other students, leading to poorer outcomes and lower quality education overall. Similarly, artificially inflating outcomes through cheating or other measures could raise the bar artificially high and negatively impact other students. Thus, there exists a perfect duty not to manipulate data that is input into the system, as it could not only harm the student who does it, but also all subsequent students whose education is affected by the skewed dataset.

### Duty to Prevent Dehumanization

In addition to the positive duty to ensure the best quality education for every individual student, AI designers, technical administrators, and teachers also have the duty to make sure that no harm comes to students through AI technology. The main threats in this regard come through the risk of dehumanizing students. However, if these are properly addressed, we may be able to enjoy the benefits of AI without the potential downsides outweighing the good that AI can do. To this end, there are a number of principles and duties that we must keep in mind:

1. We must at all times respect the dignity and humanity of students, treating them as ends in themselves.
2. We must promote the good of individual students over collective goals.

3. AI designers, technical administrators, teachers, and implementers of AI technology have the positive and perfect duty to prevent dehumanization.
4. The moral development of students should take precedence over the development of skills or aptitudes when the two conflict.

The goal of the AI driven classroom is to enhance the development of students as individuals. AI offers enormous potential in this regard, but it also has the potential to do the opposite. Dehumanization can occur in a number of ways, as laid out by Nussbaum in her 1995 article which this thesis used for inspiration. Among the types of dehumanization discussed in this thesis are:

1. Instrumentality: where a person is used as a means to an end
2. Denial of Autonomy: where a person is treated as if they lack the ability of self-determination
3. Inertness: where a person is treated as if they lack agency or the ability to act upon their own will.
4. Fungibility: where a person is treated as if they were interchangeable with others

There is a duty to protect against each one of these kinds of dehumanization within the AI driven classroom. Each of these types of dehumanization violates one of the principles above or violates another core Kantian deontological principle.

Instrumentality

Perhaps the clearest violation of the Good Will can be seen in instrumentality, or the treating of people as a means to an end. This is quite literally the opposite of the Kantian principle that people should be treated as ends in themselves, and not merely as

means to an end, and there exists a perfect duty to never treat a person as if they were simply instrumental to a goal.

Within the K-12 mathematics classroom, AI could be manipulated in a number of ways to dehumanize students in this way. For instance, if funding or employment were tied to standardized test scores, or to some other metric like overall improvement, there could be incentive for teachers and school districts to manipulate the data fed to AI systems or tamper with them in order to artificially raise or lower student scores to meet those metrics.

These types of interference with the education of students have a great potential to interfere with their learning and development, and the students themselves are not truly being considered in such cases. Not only are students being treated as a means to an end rather than an end in themselves, but collective goals are being prioritized over individual outcomes and are consequently being dehumanized, leading to a violation of the first three of our principles. In this case, educators are also failing to act with a Good Will, and are thus poor leaders and mentors, while also failing in their duty to give students the best education of which they are capable. Thus, not only is this type of practice dehumanizing, but it is impermissible on other ethical grounds as well.

In order to act in accordance with the dictates of a Good Will, human actors must do everything they can to ensure that the AI systems that they design, monitor, administer, and oversee strive to maximize the individual potential of each student, which ultimately is likely to have the desired effect of maximally improving their competency. To this end, there is a duty to prevent data manipulation at all levels, and to ensure that

monitoring and detection occur to prevent this kind of manipulation so that this kind of dehumanization cannot occur due to unethical and self-interested behavior.

### Autonomy

Another type of dehumanization that is relevant to this discussion is autonomy. Specifically, within classrooms that use AI, both educational and technical professionals have a perfect duty to defend and expand student autonomy and self-determination wherever possible.

Wherever possible, students should be allowed to make their own decisions regarding their educational path to align with their own interests and goals, including the types of topics that are covered, instructional methods, and delivery timelines, insofar as this does not conflict with general educational goals or practice. AI systems actually have the power to increase autonomy of this kind, especially when paired with individualized learning.

However, if AI is misused, it could have the opposite effect, even though AI has the potential to decrease the amount of dehumanization of this kind if used properly. If, for example, students are treated merely as a collective without reference to their individual goals and interests, this would violate our second principle and would be ethically impermissible. In addition to being a violation of the fundamental right to autonomy, it would also be at risk of treating the student instrumentally, and thus violate the Good Will as the student would be treated as a means and not an end in themselves.

## Inertness

Inertness is another form of dehumanization that can occur if AI is misused. A person may be treated as if they are inert if their will is denied or if they are treated as if they have no will to act upon. The main risk here is that the purpose of education is to develop a Good Will, and if a student is treated as if their will does not matter or does not exist, they cannot properly develop or train their wills.

The risk of inertness should be primarily addressed by the designers of AI technology. It is up to the designers of these systems to ensure that students are able to exercise and properly train their wills, and make sure that they are not suppressed. Ultimately, the designers of AI systems have a duty to take into account not only student will, but also their goals and desires. There is also a duty to protect their autonomy and allow for environments in which choice is allowed to the greatest extent possible within educational technologies in order to ensure that students have the greatest opportunity to exert their wills and work towards their own goals and practical ends.

## Fungibility

Fungibility is another primary risk within AI driven systems, made more potent because many of the statistical models upon which AI is based treat data points, which may in this case represent students, as completely interchangeable. Fungibility refers to the interchangibility of units, and when people are treated as if they are interchangeable they are inherently being dehumanized.

Several risks exist for fungibility within AI driven environments. Students may be reduced to datapoints by standardized testing or other data that is used for decision making without regard to their interests as individuals. They may also have their needs

ignored or subjugated to the “common good” even when it is practical to have individualized solutions by tailored AI powered educational tools. This violates our first three principles and ultimately is impermissible, creating a perfect duty to ensure that individuals are valued as ends in themselves and not as fungible units or datapoints, which can only be ensured with human oversight and intervention.

References

Alter, S. (2022). Understanding artificial intelligence in the context of usage: Contributions and smartness of algorithmic capabilities in work systems. *International Journal of Information Management*, *67*, 102392.

Bentham, Jeremy (1789). *An Introduction to the Principles of Morals and Legislation*. London.

Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2021, June 23). *How artificial intelligence will impact k-12 teachers*. McKinsey & Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/how-artificial-intelligence-will-impact-k-12-teachers.

Colbourn, M. J. (2002, May 20). *Applications of artificial intelligence within education*. Computers & Mathematics with Applications. https://www.sciencedirect.com/science/article/pii/0898122185900549.

Dangwal, et al (2005). *A model of how children acquire computing skills from "Hole in the Wall" computers in public places*. Information Technologies and International Development Journal, Summer 2005, Vol. 2, No. 4, 41-60

Dangwal, Ritu (2005). “Public computing, computer literacy and educational outcome: Children and computers in rural India” International Conference on Computers in Education 2005, Singapore.

Dangwal, Jha & Kapur (2006). *Impact of Minimally Invasive Education on Children - An Indian Perspective.* British Journal of Educational Technology Vol 37 No 2 2006, 295-298.

Denoël, E., Dorn, E., Goodman, A., Hiltunen, J., Krawitz, M., & Mourshed, M. (2021, June 23). *Drivers of Student performance: Insights from Europe*. McKinsey & Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/drivers-of-student-performance-insights-from-europe.

Garrido, A. (2010, May 6). *Mathematics and artificial intelligence, two branches of the same tree*. Procedia - Social and Behavioral Sciences. https://www.sciencedirect.com/science/article/pii/S1877042810002004.

Harvey and Kumar, (2019). "A Practical Model for Educators to Predict Student Performance in K-12 Education using Machine Learning," *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, Xiamen, China, 2019, pp. 3004-3011, doi: 10.1109/SSCI44817.2019.9003147.

Heick, T. (2018, September 16). *10 roles for artificial intelligence in education*. TeachThought. https://www.teachthought.com/the-future-of-learning/10-roles-for-artificial-intelligence-in-education/#:~:text=1%20Artificial%20intelligence%20can%20automate%20basic%20activities%20in,courses%20need%20to%20improve.%20...%20More%20items...%20.

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of Artificial Intelligence. *California Management Review*, *61*(4), 5–14. https://doi.org/10.1177/0008125619864925

*Hole in the Wall Learning Station*. NIIT Foundation. (2022, September 22). Retrieved January 10, 2023, from https://niitfoundation.org/hole-in-the-wall-learning-stations/

Imandar (2004). *Computer skills development by children using 'Hole in the Wall' facilities in rural India.* Australasian Journal of Educational Technology 2004, 20(3), 337-350.

Inamdar & Kulkarni (2007). *'Hole-In-The-Wall' Computer Kiosks Foster Mathematics Achievement - A comparative study.* Educational Technology & Society, 10 (2), 170-179 (2007).

Jha & Chatterjee (2005). *Public-Private Partnership in a Minimally Invasive Education Approach.* International Education Journal, 2005, 6(5), 587-597.

Kant, I. (1785). In *Groundwork of the metaphysic of morals*. essay.

Kant, I. (1803). *Immanuel Kant Über Pädagogik (On Education)* (F. T. Rink, Ed.). Bey Friedrich Nicolovius.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of Big Data and Artificial Intelligence in education. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.580820

Manna, R. and Nath, R. (2021). "The Problem Of Moral Agency In Artificial Intelligence," *2021 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, Chennai, India, 2021, pp. 1-4, doi: 10.1109/21CW48944.2021.9532549

Martinho, A., Poulsen, A., Kroesen, M., & Chorus, C. (2021). Perspectives about artificial moral agents. *AI and Ethics*, *1*(4), 477-490.

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

Miles, K. (2014, August 22). *Ai could doom the human race within a century, Oxford professor says*. NOEMA. https://www.noemamag.com/artificial-intelligence-may-doom-the-human-race-within-a-century-oxford-professor-says-3/#:~:text=Well%2C%20we'd%20have%20to,drug%20the%20AI%20could%20design.

Mitra, et al (2005). *Acquisition of computing literacy on shared public computers: Children and the 'Hole in the Wall'.* Australasian Journal of Educational Technology 2005, 21(3), 407-426.

Mitra (2000). *Children and the Internet: New Paradigms for Development in the 21st Century* Talk at the Doors 6 conference of the Doors of Perception at Amsterdam, Holland, November 11, 2000

Mitra (2000). *Minimally Invasive Education for mass computer literacy* Presented at the CRIDALA 2000 conference, Hong Kong, 21-25 June, 2000

Mitra (2005). *Self organising systems for mass computer literacy: Findings from the 'Hole in the Wall' experiments*. International Journal of Development Issues, Vol. 4, No. 1 (2005) 71-81

Mitra & Rana (2001). *Children and the Internet: Experiments with minimally invasive education in India.* The British Journal of Educational Technology, volume 32, issue 2, pp 221-232. (2001)

Muehlhauser, L., & Salamon, A. (2013). Intelligence explosion: Evidence and import. In *Singularity hypotheses: A scientific and philosophical assessment* (pp. 15-42). Berlin, Heidelberg: Springer Berlin Heidelberg.

Natematias, & Natematias. (2012, May 16). *Is education obsolete? Sugata Mitra at the MIT Media Lab*. MIT Center for Civic Media. https://civic.mit.edu/index.html%3Fp=804.html.

Nussbaum, M. C. (1995). Objectification. *Philosophy & Public Affairs*, *24*(4), 249-291.

Poole, D. (2000, July). Logic, knowledge representation, and Bayesian decision theory. In *International conference on computational logic* (pp. 70-86). Berlin, Heidelberg: Springer Berlin Heidelberg.

Puri, A. (2020). Moral imitation: Can an algorithm really be ethical?. *Rutgers L. Rec.*, *48*, 47.

Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E. W., Stahl, D., & Fusar-Poli, P. (2020). Implementing Precision Psychiatry: A systematic review of individualized

prediction models for Clinical Practice. *Schizophrenia Bulletin*, *47*(2), 284–297. https://doi.org/10.1093/schbul/sbaa120

Schlegel, D., & Uenal, Y. (2021). A Perceived Risk Perspective on Narrow Artificial Intelligence. In *PACIS* (p. 44).

Schönecker, D. (2022). Kant's argument from moral feelings: why practical reason cannot be artificial. *Kant and Artificial Intelligence*, 169-188.

Serafimova, S. (2020). Whose morality? which rationality? Challenging Artificial Intelligence as a remedy for the lack of moral enhancement. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-00614-8

Steels, L., & Lopez de Mantaras, R. (2018). The barcelona declaration for the proper development and usage of artificial intelligence in Europe. *AI Communications*, *31*(6), 485–494. https://doi.org/10.3233/aic-180607

Turchin, A., & Denkenberger, D. (2018). Military AI as a convergent goal of self-improving AI. In *Artificial intelligence safety and security* (pp. 375-393). Chapman and Hall/CRC.

Turing, A. M. (1950). I.—Computing Machinery and intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/lix.236.433

Ulgen, O. (2017a). Kantian ethics in the age of artificial intelligence and robotics. *QIL*, *43*, 59-83.

Ulgen, O. (2017b). The ethical implications of developing and using artificial intelligence and robotics in the civilian and military spheres.

Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and Machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

*What might happen if robots replace teachers*. TeachThought. (2021, June 13). https://www.teachthought.com/pedagogy/robots-replace-teachers/.