2023

# Data Ethics and the Dilemma Created by Turing's Learning Machines

Jacob Kuhn
*University of Northern Iowa*

# DATA ETHICS AND THE DILEMMA CREATED BY TURING'S LEARNING MACHINES

A Thesis Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors

Jacob Kuhn

University of Northern Iowa

May 2023

This Study by: Jacob Kuhn

Entitled: Data Ethics and the Dilemma created by Turing's Learning Machines

has been approved as meeting the thesis or project requirement for the Designation *University Honors.*

Approved by:

Dr. Eugene Wallingford, Honors Thesis Advisor

Dr. Jessica Moon Asa, Director, University Honors Program

## I. Introduction

The world is in the midst of the Fourth Industrial Revolution: The Data Revolution. The public is beginning to feel the ramifications of this revolution as all forms of media fill up with story after story about Artificial Intelligence. Users may have noticed that companies have begun to use the buzzwords "Artificial Intelligence" to market their platforms. At the same time, 67% of companies say they use machine learning today (Brown). One might wonder what Artificial Intelligence means in terms of everyday life. What it describes is the collection of a large amount of data, often referred to as Big Data, and the analysis of this data to make predictions.

For the most part, everyday individuals know data is being collected on them, but they may wonder how far data collection goes. To answer this question, one only needs to examine how they go about their daily lives. Every Google search, each liked Instagram post, any item added to a shopping cart, every reported crime, and even every individual's face recorded entering a building can create data. An important piece to remember is that if something can be represented in characters or quantities, it can be turned into data for a computer. This data can be used or sold to create Artificial Intelligence that can do a multitude of things. Individuals can not ignore the Data Revolution anymore because it has infiltrated their daily lives.

Big Data refers to the collection of large amounts of data on a population along with the computation done on this large amount of data to reveal patterns or predictions in different industries, human behaviors, and even medicine. A machine learning algorithm is applied to the data to make these predictions. The term "machine learning algorithm" is simply the process by which a computer goes about predicting an outcome on data. When they are applied to data, the term "Artificial Intelligence" is typically used. Given the fact that data is collected from everyday individuals, controversy arises from how it is collected and used. Data has given

corporations like Microsoft, Apple, Google, Meta, and Amazon equal amounts of political influence as some nations (Vallor and Rewak). At this point, one might be concerned about the data that is collected on them and what is being done to regulate it. The growing field of Data Ethics seeks to offer guidance on this question.

The main purpose of this research is to shed light on the good and bad that has come about from the interaction of Big Data and Artificial Intelligence in society. Transparency with the public is paramount for the future of Artificial Intelligence. Without awareness, the public is blind to the parts of the Data Revolution that could help them or hinder them. The key question is what AI advancements are being made and what ethical problems do they pose to the general population? To help answer this question, it is best to examine the founding of Artificial Intelligence and the views the founders had on the matter of ethics.

## II. The Founding of Artificial Intelligence

Best known for his work with decrypting the Enigma Machine in World War II thanks to the film *The Imitation Game*, Alan Turing is often referred to as the "Father of Modern Computer Science" ("About us - Alan Turing Institute"). One area Alan Turing explored toward the end of his short career was what would later be known as Artificial Intelligence. The foundational work of this field is in Alan Turing's paper, "Computing Machinery and Intelligence," where he opens with the question, "can machines think?" (Turing 433). To answer this question, Turing put forth the Turing Test as a theoretical method of scoring how intelligent a system was.

In this test, an interviewer passes questions behind two closed doors. On the other side of each door, there could be either a person or a computer. The two interviewees (the human or computer) answer the question and pass a response back to the interviewer. If the interviewer cannot identify the computer's response from each of the interviewees' responses, that computer would have passed the Turing Test. Of course, this test was not called the Turing Test by Turing himself. Instead, he called it "The Imitation Game."

In a part of this profoundly philosophical writing, Turing attempts to describe a learning machine by using an analogy of a child's mind. Turing starts with the initial state of the brain. For a child, this would be at birth. The child develops its brain through education but experiences other events outside of structured education that helps the child grow, too. This use of an initial state, a history of past decisions developed through learning, and an element of randomness are three pieces of machine learning that were foundational to the development of learning algorithms. Turing and other researchers of the time thought games like chess would be a good

place to start work on learning machines (Turing 460). These competitions display a man versus machine sentiment.

The idea of intelligent computers that could outperform humans was a novel thought in the mid-1900s. Computers were the names given to individuals who did computations before the invention of computers in the middle of the 20th century. The idea of a thinking machine that could pass as a human would have been otherworldly. Today, however, researchers question the importance and relevance of the Turing Test with modern AI research (Sharkey). Whether or not the Turing Test is relevant does not matter since the original "test" was intended to be a thought experiment to explore and innovate on a new idea.

A BBC article by Professor Noel Sharkey at the University of Sheffield argues that despite the Turing Test being somewhat unimportant today due to AI's focus on the practical instead of the theoretical, it had an incredible influence on Artificial Intelligence during its inception (Sharkey). Many individuals entered the field of Artificial Intelligence after Turing's paper was published. They were driven by the intrigue of creating a machine that could think like a human. For a few years in the late half of the 1900s, contests were held to see if a computer could pass the test. Artificial Intelligence branched away from "imitating" humans to competing with humans to do jobs faster and more efficiently. Examples of AI competitions with humans include the Watson program that competed on Jeopardy and the Deep Blue program that defeated a world champion in chess. In these examples, the computers were not trying to fool their competition; the competitors knew they were versing a computer. Instead, the focus of these competitions was to make a computer that could outsmart a human opponent.

Based only on Turing's article and these early examples of Artificial Intelligence, there seems to be little concern about the ethics of machine learning. The truth is, there was not much

concern at all about the ethical nature of machine intelligence. Turing was relatively nonaligned with machine learning applications. In his papers, he is silent on positive and negative uses for AI because computers were not anywhere close to doing what they could today. Additionally, Turing was focused on the theoretical and philosophical ideas of early Artificial Intelligence. He only thought about how a machine could think and not about the limits that should be imposed on a thinking machine.

This neutrality on the ethics of Artificial Intelligence and data usage extended to researchers immediately following Turing and to researchers today. A 2020 article published in the scientific journal, *Nature* by Pratyusha Kalluri states that AI researchers believe they are ethically neutral. The researchers' only objective is to experiment and therefore gain knowledge. However, Kalluri also states that by being impartial, AI can be taken advantage of by those with influence as a way to grow their power: “an indifferent field serves the powerful,” (Kalluri). Without considering the ethics of a given field, those in power will exploit those who are simply pursuing knowledge.

In the span of less than a hundred years, Turing’s Learning Machines evolved from an idea into a technological revolution. Data and Artificial Intelligence are multidisciplinary fields that can be used to make predictions in any field. With no limits to the scope of Artificial Intelligence and a sense of neutrality by those in the foundations of the field, growing pains are bound to emerge with how Big Data is collected and Artificial Intelligence is used.

## III. Modern Uses of Big Data and Artificial Intelligence

Today, Artificial Intelligence and data have the potential to work in any field. Those who use it seek to make predictions or define correlations between elements in a given field. However, what may be a favorable outcome to an individual through the use of AI may not be favorable to the population the individual is predicting on. It is also helpful to note whether Big Data as a whole is ethical or not. This may come down to who is using the data and what their motivations are. Areas of concern when it comes to data are data privacy, data collection, data bias, and data ownership.

Before diving into different implementations of Artificial Intelligence, it is important to understand how modern machine learning works at a high level. The first part of machine learning is data. As stated at the beginning of this thesis, data is anything that can be represented by characters or quantities. The data that is prepared for the machine is called the *training set*. Most of this data is passed onto the machine learning algorithm created by programmers to do whatever the programmer wants the machine to do. The machine learning algorithm can be descriptive and develop an explanation of the data for the programmer, it can be predictive and predict what is going to happen next, or it can be prescriptive and use the information gathered from the data to make suggestions (Brown). A machine learning algorithm is only limited by the programmer using it.

In this section, a variety of different applications of Artificial Intelligence and Big Data are explored. These include predictive policing, ChatGPT, deepfakes, medical predictions, and mass surveillance. It is important to note that this list is not exhaustive. AI is becoming increasingly ingrained in every part of the world. This list merely provides a range of different domains that are mentioned in today's media.

## A. Predictive Policing

Predictive Policing is a strategy employed by police that uses AI to detect where crime is likely to occur. Crime prediction can either target a general area of a crime or a person that will commit a crime (Lau). However, AI does not offer a guarantee when predicting outcomes. It is a percentage-based outcome of whether crime will occur or not. When predicting what area is likely to have crime, data can be gathered from crime reports. These would usually contain the street address of the crime. By using AI machine learning methods on this data, the police can see where crime tends to happen. With this information, police can put more patrols in crime-likely areas in an attempt to curb crime rates in a neighborhood. Person-based crime prediction looks at which individuals need to be watched in case they become re-offenders based on different risk factors (Lau). Since 2010, predictive policing has been implemented in urban areas like Los Angeles, Chicago, and New York City to predict gun violence on a neighborhood basis and an individual basis (Li). However, crime prediction comes with certain problems.

One problem that occurs in crime prediction is biased data. Some researchers call data used by predictive policing "dirty data" meaning it is "derived from or influenced by corrupt, biased, and unlawful practices," (Richardson et al.). Police data shows that Black people are arrested at a rate five times higher when compared to white people in 800 districts across the country. In 250 of those 800 districts, Black people are arrested ten times more often than white people (Thomas et al.). This is all while Black and white people use drugs at a similar rate (NAACP). Therefore, predictive policing in these areas tends to target neighborhoods of color at a disproportionate rate. Some claim that predictive policing is "punishing people for being poor," (Prison Policy Initiative). Biased data returns biased results.

This is because Artificial Intelligence has no sense of *truth*. Whatever data is fed to it, it will predict on. What this means is that if incorrect information is used to develop a machine learning model, that model will make incorrect predictions. If an individual does not understand that the machine is making incorrect predictions, those predictions could be taken as truth. For example, if an AI was trained on images of apples and bananas to guess which fruit was which, but the training data had the image's labels reversed meaning that apples were labeled as "bananas" and vice versa, the machine would then think apples were bananas. If an individual came along with no prior knowledge of apples or bananas and learned from the machine, their understanding of these fruits would be flawed.

Another problem with predictive policing is that targeting "hot-spots" with a police presence creates more crime in said location, which then asks for a higher police presence. This creates a never-ending cycle where police are sent to a location that is said to have high crime. The police report more crimes there. Then, the prediction model says that police need to go to the location because there is a high crime there.

Oakland, California is an area that implemented predictive policing in 2011. However, the data their model was trained on had Black people being arrested at a rate 1.5x higher than white people. This then has the model report that the police target Black people more than other races. Some researchers found that when an area was targeted as "high risk" and a higher police presence was sent there, the crime in that area increased by 20% (Lum et al. 16). This results in an over-policed, disproportionate population when compared to society as a whole.

Cities like Los Angeles have discontinued their predictive policing efforts, but others still use them to predict crime. Predictive policing encourages society to have an intense ethical discussion about the role of police and their crime stopping abilities. Ordinary citizens want

crime rates to be lower. The Big Data Revolution offers up new opportunities to do this. However, AI cannot account for the biases that come along with crime reports. Minorities have the most to lose from being targeted by predictive policing. It is up to the public to understand the limitations of Artificial Intelligence, and to not take data driven policing as Truth.

**B. ChatGPT and other LLMs**

ChatGPT is a topic that has currently captured the public's eye. It is described as a "quirky chatbot" by NBC reporters and an unsettling development in technology by educators (Rosenblatt; Munemo). As an event for Valentine's Day, the New York Times even made ChatGPT into a poet to generate love poems (The New York Times). ChatGPT has passed many different exams, too. These include an MBA exam, the US medical licensing exam, and some law school exams, but the BAR has not been passed yet. ChatGPT passed all of these exams with grades ranging from a 'B' to a 'C' (Varanasi). However, despite this coverage in the press, one might ask themselves what this software is and how it works.

ChatGPT is a type of Large Language Model (LLM), and it is certainly not the only LLM in existence. Online translators, website chatbots, and AI assistants like Siri and Alexa are also LLMs. These LLMs use enormous amounts of data to gain knowledge through machine learning to recognize different words, summarize long texts, translate languages, and generate text by predicting ensuing words (Lee - NVIDIA). LLMs have become very popular in the last few years. One of these is ChatGPT.

ChatGPT is developed by OpenAI, an investee of Microsoft, whose mission statement reads, "OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity," (OpenAI - About). The style that ChatGPT is meant to be

interacted with is a “conversational way” (OpenAI - Introducing ChatGPT). This means that the chatbot will interact in the format of a dialog and will remember what was previously said and have the autonomy to refuse to answer certain questions. For a user, ChatGPT looks like an interface with a textbox to enter a question or prompt. This could range from asking historical questions to asking the bot to write code that creates a website. ChatGPT will then respond to whatever prompt is given. The entire interface looks a lot like any messaging app used by most of the population.

The training process of the bot was incredibly intense but follows a style of machine learning called reinforcement learning where the Artificial Intelligence is rewarded for getting a correct answer. If the model outputs an incorrect or undesirable answer, the AI can be punished or simply not rewarded. This system is called a *reward system*, and the best way to imagine this process is to call it trial and error (Brown). The machine has a base score. By creating a good response, the score is increased. By creating a poor or mediocre response, the score might be decreased or left the same. The machine will make a decision based on a given environment and based on that decision. By remembering what actions lead to what outcomes, the machine can work towards an optimal outcome with the highest score whether that is a self-driving car or an AI that learns how to play Black Jack.

OpenAI trained a reward system for ChatGPT based on human input on what constitutes a good prompt response to a plethora of different prompts. After having the reward system, OpenAI could let ChatGPT start generating original responses. Then, the system could score itself on how well it did with the reward system. Using this score, the model would move on to the next prompt while eventually seeing patterns in what responses gave good scores and what

responses gave bad ones (OpenAI - ChatGPT). This process allows ChatGPT to get better and better.

OpenAI makes sure to communicate that there are certain limitations to ChatGPT's abilities. One of them is that ChatGPT sometimes writes believable responses that are in reality incorrect. The AI does not know what is and is not true; it only knows what it has been trained on. Another weakness is that ChatGPT likes to repeat words and phrases and will sometimes write very long sentences for what should be simple responses. OpenAI blames this on the training data where human scorers gave higher scores to responses that were longer. This leads ChatGPT to believe that long responses are good responses. Repeating words and phrases opens up the possibility of individuals knowing ChatGPT wrote any given piece of information. Additionally, some programs can detect whether a piece of writing was made by an LLM or not. Copyleaks is a company that has software that detects AI-generated text. Their "AI Content Detector" uses AI technology to detect if a given text was written by a chatbot or not. They claim that their system has a 99.12% accuracy rate (Copyleaks). This accuracy rate shows how well AI is at recognizing AI.

To test this, I generated text with ChatGPT and fed it to the software. Instantly, it recognized it as a 100% probability of AI-generated text. Secondly, I mixed in human-generated text with ChatGPT text and fed it to the software. The system's score decreased to around 80% likelihood of AI presence. What I discovered was that for the system to recognize a given text as "human-generated," I had to rephrase most of the text and reformat paragraph lengths. A conclusion that can be drawn from this is that, if someone wants to submit ChatGPT text as their own, they will have to change a lot of the phrasing to not have their text recognized by AI text detectors.

Those who face the most ethical questions about ChatGPT and other LLMs are teachers. Some educators are arguing to eliminate chatbots from education so that students learn from their mistakes instead of never learning in the first place by using a chatbot (Munemo). Others are embracing chatbots to come up with more engaging lessons. A teacher in Kentucky uses ChatGPT to play a game called “Find the Bot " where students have to summarize an article while ChatGPT does the same. The rest of the class then has to choose which summary was written by students and which was written by the “Bot.” Students also appreciate ChatGPT when they are looking for a spark of creativity (Gecker). A lot of value can come from using ChatGPT as an aid.

There is no clear black-and-white answer to what ChatGPT and other LLMs should be used for. It is an ethical dilemma that will have to be shaped by those in education and society. Should journalists be replaced in favor of AI-written news stories? Should students still be required to write papers the way they have been taught for decades? AI throws a wrinkle in all of that.

### C. Deepfakes

Another example of modern Artificial Intelligence is the use of deepfakes. Deepfakes are videos, photos, or audio where an artificial system can create realistic material that is hard to discern from materials that are actually from reality (Blauth et al.). The name ‘Deepfake’ has unethical origins. Meredith Somers, a news reporter for MIT Sloan, states that the term came from a Reddit user who had the same name. This user would create artificial pornographic material by face-swapping individuals (Somers).A deepfake does not necessarily have to have fraudulent use, though.

Individuals who are fans of the movie industry have noticed this technology in recent years. Robert De Niro was de-aged using deepfake technology in *The Irishman*, the late actor Peter Cushing's face was recreated in *Rogue One: A Star Wars Story* to fill the role of Admiral Tarkin, and most recently, Mark Hamill was face-swapped with an actor in the show *The Mandalorian* to be a younger Luke Skywalker. These examples outline the popular implementations of deepfakes. However, the limits of deepfakes are constantly being pushed. One group, Corridor Digital, has been experimenting with deepfakes from Keanu Reeves to their own childhood photos (Corridor Digital, "Deepfakes and AI"). The entertainment industry seems to be at the forefront of deepfake technology.

A question that someone might be wondering is how deepfakes are made. Researchers from the Georgia Institute of Technology, Yisorle Mirsky and Wenke Lee, describe the process of making a visual deepfake as a machine that copies the expression, mouth, gaze, pose, and body using Generative Neural Networks (GANS) which are machine learning techniques where two models are competing against each other to be the most accurate (Mirsky and Lee). This is known as a zero-sum game (Goodfellow et al.). The process has two parts: the Discriminator and the Generator. The process begins with pictures or videos of a person from many different points of view. Then, the algorithm detects faces and crops them. After cropping the faces, it extracts intermediate representations like the depth of the face, key landmarks, and skeletal features. With this information, the Generator generates an image from noise. Noise is simply a random variation in pixels. This generated image is fed to the Discriminator that must decide between a real image of said person and the fake image to determine what is fake and real. The outcomes of the Discriminator are fed back to the Generator so that the generator can adjust its methods and get better at creating images that replicate the *real* images. These two parts compete back and

forth. The Generator is trying to fool the Discriminator, and the Discriminator is trying to identify the Generator's images.

This technology has gotten more effective over the past few years. A survey in 2018 found that individuals presented with a deepfaked image could only tell about half of the time whether an image was deepfaked or not (Rössler et al.). That is the same odds as flipping a coin. With more time, this technology will only become better. Those interested in this area of technology may have seen MIT's deepfake of Richard Nixon in 2020 announcing that the Apollo 11 mission failed (Arts at MIT). Any individual who has gone through the educational system knows this to be false. The project was a proof of concept to show what deepfake technology could do. Additionally, it shows that deepfakes can mislead the public and lead to fraud if individuals are not aware of the facts.

Vocal deepfakes follow a similar process of creation. However, instead of training on visual examples, Artificial Intelligence is trained on audio examples. The process can be summarized in three stages: the encoding stage, the synthesizer stage, and the vocoder stage (Blue et al.). Below is a description of each stage.

The encoder stage is when the machine learns the specific elements that make up the voice of the individual being cloned. This is done through uploading sentences where the individual is enunciating different sentences. With more audio examples, the model can become better, but there are diminishing returns as the examples grow. What this means is that the percentage by which the model gets better grows at a smaller rate the more examples presented. Early on, the model might get better by a factor of 10 percent. As examples grow in number, however, the model might only get better by a fraction of a percent. The output of this stage is called the embedding. When it is passed onto the synthesizer stage, a Mel Spectrogram is

constructed based on the embedding that simulates the way human ears pick up on vocal frequencies. A Mel Spectrogram is a visualization of sound. The final stage, the vecoder, takes this spectrogram and converts it into an audio recreation that sounds like the original individual (Blue et al.). These three steps come together to form a vocal deepfake.

For this thesis, an experiment was conducted with a vocal deepfake where I used ElevenLabs' software to clone my voice. ElevenLabs is a startup created in 2022 by ex-Google machine learning engineer, Piotr Dabkowski, and ex-Palantir deployment strategist, Mati Staniszewski. The mission of their company states that they want to make "on-demand multilingual audio support a reality across education, streaming, audiobooks, gaming, movies, and even real-time conversation," (ElevenLabs). The process of cloning a voice costs $5 a month (as of February 2023) and only takes one minute of clear voice recordings to generate the cloned voice.

To experience the limits of the machine learning algorithm on my voice, I used voice recordings of myself saying various nursery rhymes and tongue twisters. After the minute of audio was uploaded, I was able to enter text into a box and generate an audio file of my deepfaked voice. Subjects who played the audio to test the audio's believability were friends and family. To present the experiment, I asked them to listen to an audio recording. I did not tell them it was AI-generated, but asked their opinion on the recording after they listened. This was done to keep the same spirit of the Turing Test where during the "game," players did not know if a computer contestant was an option. I revealed it was AI-generated only after they gave typically confused responses as to why they were made to listen to the recording. Every subject believed that the deepfaked audio was me speaking and not AI-generated. What this simple

experiment shows is that deepfaked audio has advanced to a point where humans cannot instantaneously tell what is real and what is not.

Many researchers, including those at the University of Florida, have programs that can correctly identify deepfaked audios with 99.5% accuracy (Blue et al). However, it seems humans can not tell the difference between deepfakes and reality at a certain level (Rössler et al.). An example of this deepfake fraud was when deceptive actors had funds moved from a company by impersonating a CEO (Stupp). There is no reason to think that politicians, celebrities, and even everyday individuals could not also be deepfaked as illustrated by the experiment in this section. Hypothetically, if a group or individual developed a deepfake of an individual's voice, they could contact that person's family demanding money without the family knowing anything was amiss.

Currently there are options for individuals who can no longer speak due to illness or death to "bank" their voice recordings (Costello). This process can only speak words that have been previously spoken, and cannot create new words. Deepfaked voices that can speak new words might be a place where this technology could help people in the future. People could hear lost loved ones. Those who can longer speak can say whatever they want instead of having a limited amount of words that they pre recorded before losing their voices.

There is only so much one can do to prevent themselves from being fraudulently deepfaked. While companies like ElevenLabs state that illegal use of their technology is against their Terms of Service, those who want to commit crimes will. To reduce the chances of being deepfaked, limit digital footprints. Photos and videos can be used to train these models to recreate anyone, and technology is improving to where even one minute of audio could lead to one's voice being deepfaked.

## D. Big Data in Medicine

Another area where Artificial Intelligence can help society is in medicine. As stated in the previous sections, Artificial Intelligence learns from prior data. In the instance of healthcare, if a medical group has thousands of mammogram pictures and labels for whether a given picture has cancer or not, a machine learning model can be developed to predict whether a new mammogram from a patient has cancer or not. In addition to looking at mammograms for cancer, other medical research has looked at AI prediction of skin cancer and diabetic retinopathy (Esteva). Research into cancer prediction with AI models is a hot topic currently.

A 2019 article published in the *Journal of Family Medicine and Primary Care* summarizes that artificial systems have not had a positive rate of diagnosing patients *with* breast cancers. However, it has been used effectively in determining images that a doctor should look at or images where the patients should not be concerned (Amisha et al.). A place where Artificial Intelligence has had major successes in medicine is with medical record keeping. Physicians spend 49.2% of their time entering records for patients. With the help of systems called medical scribe services, doctors can now spend more time face-to-face with patients instead of entering records (Amisha et al.). Artificial models have not removed the necessity of a doctor. AI is a tool that doctors can harness to provide better care to their patients.

Ethical concerns in medical AI use arise from three standards adopted by Institutional Review Boards. They are respecting patient autonomy, protecting patient privacy, and achieving equity among patients (Howe and Elenberg). Patient data is needed in Big Data medical research projects. In the past, every patient needed to give consent for their data to be used in a medical study. That created a large amount of overhead work for a study to happen. A new standard has

been adopted called the Revised Common Rule that seeks to help patient autonomy and help administrative work. This states that a patient does not need to give consent if deidentified data is used and only needs to give broad consent when the data being used is publicly available (Menikoff et al.). Broad consent means there is no limit to the number of studies patient data can be used for. However, patients may be unaware of which data is and is not publicly available. What these problems pose is a lack of patient autonomy with how their identifiable information is being used, stored, and maintained. Researchers put forth solutions to this problem. Informing students of all levels how their data could be used and putting warning labels on websites on what data would be collected and made publicly available are possible solutions (Howe and Elenburg). With these solutions, some ethical problems in the medical field could be addressed.

The question of equity arises in medical AI usage when predictions are made across different individuals' genetics. Some studies that use Big Data to make medical predictions have data that is made up of mostly European-descended, white data points (Lewis et al.). It is easy to see how this can be problematic when trying to make predictions with a patient who is neither European nor white. This ties into what was explained in Section III.A on Predictive Policing. Artificial Intelligence cannot fix mistakes in unrepresentative data. Therefore, unsatisfactory predictions may be made when the population being predicted on is not representative of the population the AI model is going to be applied to. Evidence of this is displayed in a study that found that when population data was adjusted to better represent Black individuals, Black patients were referred for care services by 30 percent more (Obermeyer et al.). Without this advancement, those people could have been turned away from care.

Big Data has a lot of potential to help society when it comes to the medical field. Diseases can be spotted sooner based on past information, and diagnoses can be made quickly

with machine learning models. However, ethical concerns arise with how the data used in these models are created and how the population of these datasets is relative to society as a whole. Nevertheless, AI will continue to become more involved with medicine, and humanity will benefit from its success while also suffering its consequences.

### E. Mass Surveillance

Privacy and Artificial Intelligence go hand in hand. Modern Artificial Intelligence uses Big Data to train on and develop this data, while organizations find any way they can to collect it. Ethics comes into play with how these organizations collect their data. Sometimes, individuals can opt into their data being used. An example of this is when signing up for some websites, users can opt in or opt out of the third-party distribution of their data. What if users do not know that their data is being collected? This quickly becomes a surveillance issue. Big Data could be made up of a person's face, records, location, actions, etcetera. With this, Big Data is capable of mass surveillance.

Big Data is a term related to the NSA (National Security Agency) leaks by Edward Snowden to the media in 2013. The first of these leaks showed that the NSA had gotten data from millions of Americans from their phone records. Verizon was one of the phone companies that gave data, and Snowden leaked that they were not allowed to inform the compromised individuals that the NSA had collected their data (Greenwald). Other leaks by Snowden showed that the NSA had gained direct access to many different websites' user information like Yahoo, Google, and Microsoft (Gellman and Poitras). Other compromised data came from telephone and internet companies. What this information allowed the NSA to do was monitor millions of individuals. In order to do this, the NSA had algorithms that would filter the data, analyze it, and draw conclusions from it. This information was then given to the NSA about those individuals.

Additionally, this practice was not targeted toward a few individuals; it monitored everyone (Lyon). It seems that under the guise of national security, the United States government decided to implement mass surveillance.

The implications of this revelation are that information like the number of phone calls one makes to specific individuals and the information one searches for on the internet could *trigger* the NSA's "anti-terrorism" checker and mark someone as high-risk. Big Data takes every element of a person's identity and combines it into a single profile that in the wrong hands could implicate that person. Whether or not this practice helped with the issue of national security is a moot point since no matter the outcome, the collection of Big Data on the personal lives of individuals means a Surveillance State has been created through mass surveillance.

A simple example of this mass surveillance and data collection misuse can be seen in a study where researchers created a babysitter "risk" tracker. In this study, researchers trained an AI by scanning the social media accounts of babysitters and rating the babysitter on how safe they are (Obermeyer et al.). These babysitters never gave consent for their social media pages to be analyzed in this way. It is easy to see all of the different ways this type of AI could be implemented on any population of people with this example. Imagine a society that scores citizens based on their actions or how they lean politically. That type of society would pose a lot of ethical concerns.

This extreme mass surveillance thought experiment is not hypothetical at all. One has to look no further than Chinese society where Big Data and Artificial Intelligence have been implemented to keep track of every citizen. This technology is known as "one person - one profile" and has been implemented over the past 5 years (Baptista). China has many different forms of surveillance that watch its citizens. Half of the world's surveillance cameras are located

in China; these cameras are located in places like hotels, security checkpoints, and restaurants (Qian et al.). These surveillance cameras gather the locations and faces of individuals in China. An important note though is that these cameras alone do not use AI. This part of the surveillance practice is only the data collection portion and is a perfect example of Big Data. Additional surveillance practices use citizens' phones to monitor locations and listen to individuals (Qian et al.). New ways of data collection of citizens appear every year here.

AI has recently been implemented on this data to create more sophisticated surveillance. With AI, data can be filtered to pick out specific individuals even if their face is partially covered (Baptista). This combination of Big Data with machine learning algorithms gives the Chinese government the ability to pick out specific citizens, see who they are associated with, and strike accordingly. Under the guise of protection from criminals, the Chinese government has created a personal database. The ethical question, of course, is who is or is not a criminal? This is a subjective opinion based on what the government says at any given time.

This recent implementation of Big Data with Artificial Intelligence in China ties directly to their Social Credit System. China's Social Credit System (SCP) is a system that scores citizens based on their loyalty to the state. It is a national incentive program. If a citizen engages in "trust-breaking" activities, they will be punished and put on a "black list" that is publicly broadcasted. On the other hand, if a citizen engages in "trust-keeping" activities, they will be rewarded and put on a "red list" that is also publicly broadcasted (Cho). This practice creates a state of self-policing where individuals act in one way to avoid consequences despite possibly not believing in the actions they are performing. With Big Data and AI, the SCP can keep track of individuals wherever they are and score them on whatever they are doing. The prospect brings about images of George Orwell's *1984*.

Mass surveillance is a topic that every individual should be aware of. With advanced technology comes advancements in surveillance. Governments tend to overreach citizens' privacy for the sake of security, and it is up to individuals to draw the line when it comes to surveillance. Awareness that these systems are in motion is the first step.

## IV. Progress in Data Ethics

Ethics within machine learning and Artificial Intelligence is an incredibly new topic which has surfaced in recent years. In 2014, the IEEE (the Institute for Electrical and Electronics Engineers), the largest organization of technical professionals, began holding ethics conferences to define ethical rules. Here, professionals discuss the theoretical and practical applications of ethics in Big Data and Artificial Intelligence. For ethics to be ingrained in this field, there need to be standards. The IEE created a standards organization (IEE SA) to act as a global community dedicated to creating standards for Artificial Intelligence. They state that, "as artificial intelligence, autonomous intelligent systems (AIS), machine learning, autonomous vehicles, and robotics advance at a rapid pace, careful considerations need to be made during development and implementation regarding humanity," (IEE SA). Additionally, the IEE SA helps with the education, certification, and standardization of AI systems and Data Ethics.

In 2019, the Turing Institute, a research institute based in the UK dedicated to machine learning and Artificial Intelligence, released a paper on ethics written by Dr. David Leslie. In this paper, Leslie calls for the standardization of Artificial Intelligence. The motivation for developing these standards was due to abuses occurring within Artificial Intelligence. These abuses include bias, denial of rights, and invasion of privacy. Many of these issues were apparent in the examples that were stated in Section III. To combat misuse, the author of this paper, Dr. David Leslie, implores that anyone using Big Data should follow his FAST Track Principles.

These principles hold firm to the belief that the fairness of a model's outcome should be questioned. What this means is that a given model should not be accepted as truth. One needs to analyze the fairness of the predictions before stating that the model is accurate. Next, the individual using data in a model should be held accountable for that data. Accountability is

necessary for the future of Big Data. Anyone can pull data from anywhere. If that data ends up being collected through misuse or is inaccurate in some way, the user of the data should be held accountable. Thirdly, the work should be sustainable and not have a negative social impact. It is easy to imagine a situation where someone uses Artificial Intelligence as a form of control over a given population. One only needs to look at how predictive policing affects minorities. Finally, Artificial Intelligence should be transparent in design, implementation, and evaluation (Leslie). This allows third parties to know what the data is, how it was collected, how it was used, and what conclusions can be made from the said model. Standards are necessary for Artificial Intelligence due to the potential for abuse as well as actual malicious exploitation.

The Association for Computing Machinery (ACM) is an international computing organization that has developed a general ethics code that computing professionals should follow. They state that, “computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good,” (ACM). The first ethical principle to follow is that computing should contribute to society’s well-being since everyone has a stake in the consequences of computing. Computing professionals should also “avoid harm,” “be trustworthy and honest,” and “be fair” (ACM). These three principles reflect on the ethical problems discussed earlier. Data has the power to cause harm through being biased in some way. Therefore, it is important for professionals to take steps to reflect on the work they are doing and call out unethical practices. The final three principles ACM says computing professionals should follow in their work are to credit the work of others, respect individual’s privacy, and “honor confidentiality” (ACM). These points can be summarized as not plagiarizing and not sharing information that should be private. The ACM’s

Code of Ethics and Professional Conduct is a good resource to understand how a *good* developer should act professionally.

The Data Science Association is a professional organization for data scientists. It is dedicated to improving the profession of data science. The organization has nine different rules that define the conduct by which a data science professional should operate. *Rule 8* specifically details how data scientists should ethically work with data. Within this rule, there are fifteen principles to abide by. A few of these principles are simple. One says that a data scientist should not "cherry-pick" data. This means that a professional should not point to a small group of data that confirms a specific position while ignoring a large amount of data that may come to a different conclusion. Another principle states that a data scientist should present accurate findings to clients even if the data concludes something that the client may not like. Additionally, data scientists should not knowingly use biased data, use false evidence to make conclusions, and/or claim that weak data or weak evidence is great data or evidence (Data Science Association). Creating a culture of ethics is the first step of solving these problems.

Sometimes, data scientists may need to be whistleblowers to criminal activities. The Data Science Association's Code of Ethics also speaks on the criminal side of data ethics when it says:

> (d) If a data scientist reasonably believes a client is misusing data science to communicate a false reality or promote an illusion of understanding, the data scientist shall take reasonable remedial measures, including disclosure to the client, and including, if necessary, disclosure to the proper authorities. The data scientist shall take reasonable measures to persuade the client to use data science appropriately.
>
> (e) If a data scientist knows that a client intends to engage, is engaging or has engaged in criminal or fraudulent conduct related to the data science provided, the data scientist shall

> take reasonable remedial measures, including, if necessary, disclosure to the proper authorities (Data Science Association).

These principles educate professionals on the very real criminal outcomes that could occur and their responsibility to being ethical with their work.

What these different professional organization's code of ethics display is a community that is aware of the ethical challenges that come with Artificial Intelligence and Big Data. Those in the role of a data scientist should understand the work that they are doing has real-world consequences. Data has the potential to lift everyone up, but if used unethically, it can bring everyone down.

When looking at the actions of countries around the world surrounding AI, ethics becomes relevant when looking at the benefits and drawbacks of Artificial Intelligence to the global society. Society needs to be safeguarded from malicious entities who would use Artificial Intelligence in poor ways. An article from 2017 published in *Science and Engineering Ethics* analyzed reports produced by the governments of the United States, the European Union, and the United Kingdom. The paper's purpose was to cross-analyze countries about their views on what governmental power should be allowed to do when regulating Artificial Intelligence. The authors considered their responses and how they could lead to a good AI society.

The response by the United States was determined to follow a free-market system for AI where companies are encouraged to innovate however they can. The report factored in how AI would create and destroy jobs, so they put forth that AI operations should help society and follow international humanitarian laws (Cath et al.). The European Union's statement called for a new regulatory body to help with regulating Artificial Intelligence. Their reasoning for this body was to protect the workforce by monitoring job markets and determining what educational

changes were needed in order for the workforce to adjust to new AI capabilities (Cath et al.). Finally, the UK government's release followed closely to the one released by the United States. They implored that AI should not be regulated too much, but international law and ethics should be followed. Their release called for the implementation of an international council that would focus on Artificial Intelligence (Cath et al.). The three releases had similarities and differences, but all three understood that Artificial Intelligence and machine learning are becoming more prevalent in society and may require regulation.

While it is encouraging to know that Data Ethics is a priority for important bodies in the field of Artificial Intelligence, it is up to the everyday individuals to hold companies, people, and governments accountable. Those same groups have misused these innovations in the past, and they will continue to do so in the future. Everyone needs to ask themselves what data they have, who has access to it, and what their data is being used for. Data is now an integral part of human existence. It is up to us to be aware and incorporate ways to protect it.

## V. Conclusion

The future can seem very scary when it comes to recent advancements in Artificial Intelligence and Big Data. The founders of the field, like Alan Turing, did not fully consider the ethical implications these new systems would play in society. It is not a fault on their part, though. They were more concerned with what could be done than with what should be done. Therefore, it is up to ordinary individuals today to know how these new technologies will affect their lives. Developers working with these new technologies must learn and implement the values of Data Ethics into their systems.

The Data Revolution is only beginning. AI is implemented into new fields every year. The revolution can add many benefits, such as predicting abnormalities in the field of medicine. Chatbots can increase productivity when it comes to communication. Deepfakes create the possibility for families to hear lost loved ones. With so much good and bad to be had, the ethics of AI are hard to define. Limiting factors of this thesis surround how fast this revolution is expanding. The thesis may become outdated quickly.

What started with a simple question of, “Can machines think?” has turned into a revolution that will define this era of human history. The legacy of Artificial Intelligence is still being written, but its purpose should not be evil. Those in the technoligical space and the general public do not have the choice to interact with AI or not. They will feel the ramifications of these advancements no matter what. That is why this issue is so pressing. AI should seek to benefit society as a whole. If the population understands Data Ethics and the lines that should be drawn, the full capabilities of machine learning and Big Data can be positive. A bright future is on the horizon if ordinary citizens can stand up for themselves and understand what malicious actors could do if allowed to exploit the population.

## VI. Works Cited

"About Us." *The Alan Turing Institute*, https://www.turing.ac.uk/about-us.

"About." *OpenAI*, https://openai.com/about.

"ACM Code of Ethics and Professional Conduct." *Association for Computing Machinery*, 2018, https://www.acm.org/code-of-ethics.

"Ai Content Detector: Ai Detector." *Copyleaks*, https://copyleaks.com/features/ai-content-detector.

Amisha, et al. "Overview of Artificial Intelligence in Medicine." *Journal of Family Medicine and Primary Care*, vol. 8, no. 7, 2019, pp. 2328–2331., https://doi.org/10.4103/jfmpc.jfmpc_440_19.

"Autonomous and Intelligent Systems (AIS)." *IEEE Standards Association*, https://standards.ieee.org/initiatives/autonomous-intelligence-systems/.

Baptista, Eduardo. "China Uses AI Software to Improve Its Surveillance Capabilities." *Reuters*, Thomson Reuters, 8 Apr. 2022, https://www.reuters.com/world/china/china-uses-ai-software-improve-its-surveillance-capabilities-2022-04-08/.

Blue, Logan, et al. "Who Are You (I Really Wanna Know)? Detecting Audio {Deepfakes} through Vocal Tract Reconstruction." *USENIX*, 2022, https://www.usenix.org/conference/usenixsecurity22/presentation/blue.

Brown, Sara. "Machine Learning, Explained." MIT Sloan, 21 Apr. 2021, https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

Cath, Corinne, et al. "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach." *Science and Engineering Ethics*, vol. 24, 28 Mar. 2017, pp. 505–528., https://doi.org/10.1007/s11948-017-9901-7.

Cho, Eunsun. “The Social Credit System: Not Just Another Chinese Idiosyncrasy | Journal of Public and International Affairs.” *Princeton University*, The Trustees of Princeton University, 1 May 2020, https://jpia.princeton.edu/news/social-credit-system-not-just-another-chinese-idiosyncrasy.

“Code of Conduct.” *Data Science Association*, http://www.datascienceassn.org/code-of-conduct.html.

“Criminal Justice Fact Sheet.” *NAACP*, 2022, https://naacp.org/resources/criminal-justice-fact-sheet.

“Deepfakes & A.i.” *Corridor Digital*, 2023, https://www.corridordigital.com/season/11.

Esteva, Andre, et al. “Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks.” *Nature*, vol. 542, no. 7639, 2017, pp. 115–118., https://doi.org/10.1038/nature21056.

Gecker, Jocelyn. “Some Educators Embrace CHATGPT as a New Teaching Tool.” *PBS*, Public Broadcasting Service, 14 Feb. 2023, https://www.pbs.org/newshour/education/some-educators-embrace-chatgpt-as-a-new-teaching-tool.

Gellman, Barton, and Laura Poitras. “U.S., British Intelligence Mining Data from Nine U.S. Internet Companies in Broad Secret Program.” *The Washington Post*, 7 June 2013, https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html. Accessed 28 Feb. 2023.

Goodfellow, Ian, et al. “Generative Adversarial Networks.” *Communications of the ACM*, vol. 63, no. 11, 2020, pp. 139–144., https://doi.org/10.1145/3422622. Accessed 28 Feb. 2023.

Greenwald, Glenn. “NSA Collecting Phone Records of Millions of Verizon Customers Daily.” *The Guardian* , 6 June 2013, https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order. Accessed 28 Feb. 2023.

Howe, Edumund G, and Falicia Elenburg. “Ethical Challenges Posed by Big Data.” *Innovations in Clinical Neuroscience*, vol. 17, 1 Oct. 2020, pp. 10–30.

“In Event of Moon Disaster.” *Arts at MIT*, 4 Feb. 2022, https://arts.mit.edu/in-event-of-moon-disaster/.

Kalluri, Pratyusha. “Don’t Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power.” *Nature*, vol. 583, no. 7815, 2020, pp. 169–169., https://doi.org/10.1038/d41586-020-02003-2.

Lau, Tim. “Predictive Policing Explained.” *Brennan Center for Justice*, 1 Apr. 2020, https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained.

Lee, Angie. “What Are Large Language Models Used for and Why Are They Important?” *NVIDIA Blog*, 30 Jan. 2023, https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-models-used-for/#new_tab.

Leslie, David. “Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector.” *The Alan Turing Institute*, 2019, https://doi.org/10.5281/zenodo.3240529.

Lewis, Cecil M., et al. “The Human Microbiome Project: Lessons from Human Genomics.” *Trends in Microbiology*, vol. 20, no. 1, 2012, pp. 1–4., https://doi.org/10.1016/j.tim.2011.10.004.

Li, Jonathan. “Pitfalls of Predictive Policing: An Ethical Analysis.” *Viterbi Conversations in Ethics,* 17 Feb. 2022, https://vce.usc.edu/volume-5-issue-3/pitfalls-of-predictive-policing-an-ethical-analysis/.

Lum, Kristian, and William Isaac. “To Predict and Serve?” *Significance,* vol. 13, no. 5, 2016, pp. 14–19., https://doi.org/10.1111/j.1740-9713.2016.00960.x.

Lyon, David. “Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique ...” *Sage Journals*, 2014, https://journals.sagepub.com/doi/full/10.1177/2053951714541861.

Menikoff, Jerry, et al. “The Common Rule, Updated.” *New England Journal of Medicine*, vol. 376, no. 7, 2017, pp. 613–615., https://doi.org/10.1056/nejmp1700736.

Mirsky, Yisroel, and Wenke Lee. “The Creation and Detection of Deepfakes.” *ACM Computing Surveys*, vol. 54, no. 1, 2021, pp. 1–41., https://doi.org/10.1145/3425780. Accessed 28 Feb. 2023.

Munemo, Julia McKenzie. “A Message to Students About ‘The Bot.’” *Inside Higher ED*, 23 Jan. 2023, https://www.insidehighered.com/views/2023/01/23/message-students-about-%E2%80%98-bot%E2%80%99-opinion?utm_source=Inside%2BHigher%2BEd&utm_campaign=f2c8692cd4-DNU_2021_COPY_02&utm_medium=email&utm_term=0_1fcbc04421-f2c8692cd4-197444505&mc_cid=f2c8692cd4&mc_eid=a23c9d7802.

Obermeyer, Ziad, et al. “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations.” *Science*, vol. 366, no. 6464, 2019, pp. 447–453., https://doi.org/10.1126/science.aax2342.

OpenAI. “Introducing Chatgpt.” *Introducing ChatGPT*, 30 Nov. 2022, https://openai.com/blog/chatgpt.

"Poverty and Debt." *Prison Policy Initiative*, https://www.prisonpolicy.org/poverty.html.

"Prime Voice Ai." *ElevenLabs*, https://beta.elevenlabs.io/about.

Qian, Isabelle, et al. "Four Takeaways From a Times Investigation Into China's Expanding Surveillance State." *The New York Times*, 21 June 2022, https://www.nytimes.com/2022/06/21/world/asia/china-surveillance-investigation.html. Accessed 28 Feb. 2023.

Richardson, Rashida, et al. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *94 N.Y.U. L. REV. ONLINE 192 (2019)*, 13 Feb. 2019, pp. 1–42., https://ssrn.com/abstract=3333423 . Accessed 21 Feb. 2023.

Rosenblatt, Kalhan. "CHATGPT, an AI Chatbot, Has Gone Viral. Some Say It's Better than Google; Others Worry It's Problematic." *NBCNews.com*, NBCUniversal News Group, 2 Dec. 2022, https://www.nbcnews.com/tech/tech-news/chatgpt-ai-chatbot-viral-rcna59628.

Rössler, Andreas, et al. "FaceForensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces." *ArXiv.org*, 24 Mar. 2018, https://arxiv.org/abs/1803.09179.

Sharkey, Prof Noel. "Alan Turing: The Experiment That Shaped Artificial Intelligence." BBC News, BBC, 20 June 2012, https://www.bbc.com/news/technology-18475646.

Somers, Meredith. "Deepfakes, Explained." *MIT Sloan*, 21 July 2020, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

Stupp, Catherine. “Fraudsters Used AI to Mimic CEO’s Voice in Unusual Cybercrime Case .” *Wall Street Journal*, 30 Aug. 2019, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402. Accessed 28 Feb. 2023.

Thomas, Pierre, et al. “ABC News Analysis of Police Arrests Nationwide Reveals Stark Racial Disparity.” *ABC News*, 11 June 2020, https://abcnews.go.com/US/abc-news-analysis-police-arrests-nationwide-reveals-stark/story?id=71188546. Accessed 17 Feb. 2023.

Turing, Alan M. “Computing Machinery and Intelligence.” *Mind*, LIX, no. 236, 1 Oct. 1950, pp. 433–460., https://doi.org/10.1093/mind/lix.236.433.

Vallor, Shannon, and William J Rewak. “An Introduction to Data Ethics .” *Ethics Center: Technology Ethics*, Santa Clara University, https://www.scu.edu/media/ethics-center/technology-ethics/IntroToDataEthics.pdf.

Varanasi, Lakshmi. “CHATGPT Could Be a Stanford Medical Student, a Lawyer, or a Financial Analyst. Here's a List of Advanced Exams the AI Bot Has Passed so Far.” *Business Insider*, Business Insider, 11 Feb. 2023, https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1.

“What Alan Turing Means to US.” *The Alan Turing Institute*, 23 June 2019, https://www.turing.ac.uk/blog/what-alan-turing-means-us.

“Write A (Bitter Valentine) to (Your Therapist) in the Style of (a Yelp Review).” *The New York Times*, The New York Times, 13 Feb. 2023, https://www.nytimes.com/interactive/2023/02/13/opinion/valentines-day-chatgpt.html.