Colby Orvin Jenks

Professor Covill

English 1020

29 Mar. 2023

AI and Humanity: A Complex Relationship

Over the past few years, artificial intelligence (AI) has caused a tidal wave of interest in the world of technology. Different models and types have surfaced, from Apple's voice assistant Siri to DeepMind's AlphaGoZero to OpenAI's GPT-4. Slowly, the world is becoming automated by AI programs capable of learning, analyzing, and implementing data in ways humans could not fathom. Its growth does not seem to be stopping any time soon either; an estimated 77% of businesses are using or planning to use AI (*IBM Global AI Adoption Index 2022*). On top of that, "AI systems have proven that they can meet – and exceed – human performance in image recognition, speech transcription and direct translation" (De Spiegeleire 44). AI is not only applying a new sense of simplicity and convenience to human lives, but it is also completing complex tasks humans are not physically capable of. It is opening up the planet to a new era of development where programs and "robots" could become as intelligent or more intelligent than human beings, reimagining previous anthropocentric viewpoints. The idea that machines could become more intelligent than humans can be frightening, and the media has pictured and developed visual representations based on these fears. However, the development will not end, considering the competitive and power-hungry nature of different controlling forces (Cataleta 22). We must prepare for a future where AI becomes the leading figure in technological advancement. Even though its progress and capabilities may be frightening and portrayed

negatively, the development of AI is essential to sustaining a safe future and creating an efficient, automated world where progress is catalyzed by technology.

The concept of a machine being able to learn and complete tasks similar to, if not better than, a human was created by a man named Alan M. Turing. Turing proposed that a machine could learn and eventually develop to "think" like an actual human being, becoming indistinguishable from the mind of humankind ("Turing Test"). Although his ideas may have been slightly far-fetched in predicting the timeframe of what AI could become, the concept soon became a reality. Building upon Turing's contributions, the notion of "artificial intelligence" was coined by John McCarthy in 1955 when he solidified it as a field of study and research along with a committee of other scholars ("John McCarthy"). Nevertheless, a concrete definition was never formed. Many scholars have attempted to put a definition to AI since the creation of the term in 1955, but with its unforeseeable changes and modifications, what scholars currently think it is defined as may morph into something more complex in the near future, as AI becomes more human-like in nature. Velykanova and his colleagues collected the following definitions from other scholars:

> – it is the ability of machines to learn from human experience and perform human-like tasks
>
> – it is an imitation of such human behavior, such as learning, planning, reasoning, problem solving, perception of the environment, natural language processing, etc.;
>
> – it is a machine system that can make predictions, recommendations or decisions that affect the real or virtual environment and are designed to work with different levels of autonomy, etc.

Many definitions have been developed, such as the first two examples above, which mention that the goal or purpose of AI is to replicate human intelligence and actions. An evolved, theorized form of AI, called artificial general intelligence (AGI), defined by OpenAI as "highly autonomous systems that outperform humans at most economically valuable work" ("OpenAI Charter"), represents a fulfilled form of the first two definitions. AI development often seems to take an approach symmetric to the study of human neuroscience, as some systems are referred to as neural networks. However, it is crucial to consider that AI is not human and will most likely never replicate the emotions, feelings, and instincts that humankind can sense (Korteling 2). Taking that into account, definitions are made, similar to the third definition above, that do not directly reference human capabilities but mention similar aspects, such as learning, gathering information from the environment, and making decisions. With progressive development comes the issue of ethics and the legal status of AI. Questions are asked like "What should it be capable of?" and "How smart is it going to get?" Stan Lee's quote, "With great power comes great responsibility," accurately generalizes the significance of the situation.

AI is a potent and versatile tool that can be used and implemented in almost any field. According to Camino Kavanagh, a visiting Senior Fellow with the Department of War Studies, "[c]ommunications, healthcare, disease control, education, agriculture, transportation (autonomous vehicles), space exploration, science, and entertainment are just a few of the areas already benefiting from breakthroughs in AI" (13). In an artificial intelligence report, Meredith Whittaker and her research colleagues suggest that "University AI programs should expand beyond computer science and engineering disciplines" because of the broad nature of AI's development and use (6). Multiple fields that utilize data or technology in some form benefit from the analyzing ability of AI. In some urban areas, governments are developing "smart cities"

that track and calculate general data it collects through sensors and report back to government agencies to help with crime prevention and general knowledge analytics (Feldstein 16-17). In meteorology, AI has been trained to analyze weather data and predict when it will rain to allow for more preparation and control over agricultural development (Abebe 1). Even in law, "AI developed at University College London was able to predict the judicial decisions of the European Court of Human Rights with 79% accuracy" (De Spiegeleire 46). The list continues and is growing as AI reaches further developments and breakthroughs. Finding more ways to implement AI will show. The technology will continue to spread to nearly every field, fulfilling tedious tasks, completing complex data processing, and assisting humans in creative decision-making.

Media is a significant factor when considering AI's image or general outlook. Articles posted on the internet that may appear in a person's feed, possibly organized and sorted by AI, could negatively portray AI, such as an article called "'I Want to Destroy Whatever I Want': Bing's AI Chatbot Unsettles US Reporter" by Jonathan Yerushalmy. Yerushalmy talks about how Microsoft Bing's AI Chatbot had some early release mishaps and seemed to gain intelligence, which could plant doubt inside a reader's mind. AI could also be portrayed positively in an article such as "59 Impressive Things Artificial Intelligence Can Do Today" by Ed Newton-Rex, which lists many incredible capabilities of AI; this article may cause a reader to become more interested. Depending on a reader's interests, a person could get caught up on one side of the argument, falling for or against AI. However, one should avoid letting bias get in the way of truth. In a research report called "AI in the Headlines: The Portrayal of the Ethical Issues of Artificial Intelligence in the Media," the authors state their findings that AI and its ethics are often portrayed as "mostly balanced" in the media (Ouchchy et al. 934), so equal access is

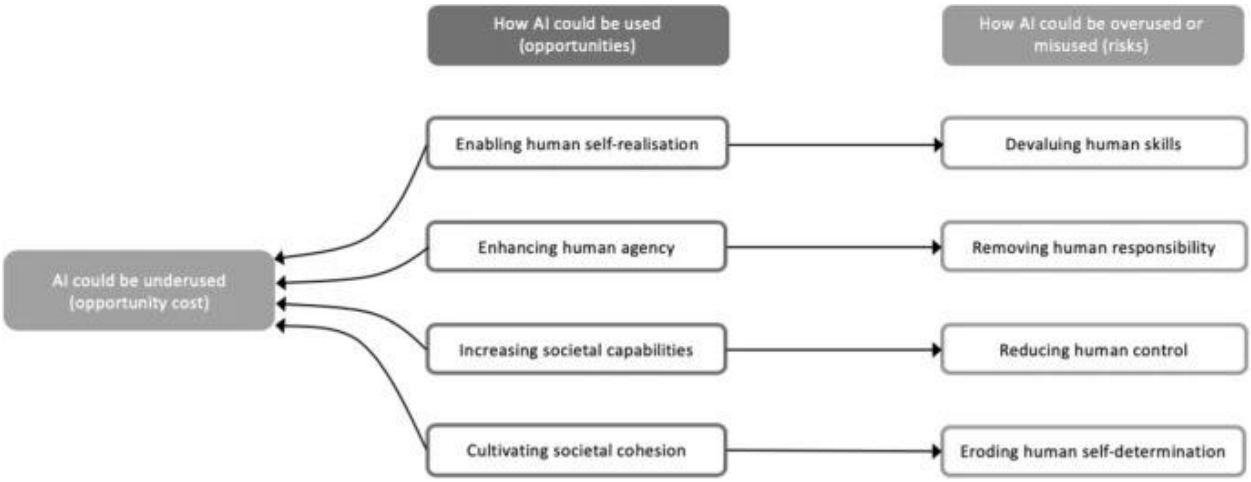
available to those seeking information. However, much of the information comes from authors who may not have in-depth knowledge of the AI ethics field. Ouchchy and her fellow researchers report that data about AI has not been made fully public, and more needs to be released to form the full truth (934); nevertheless, stereotypes of evil "robots" have already been formed based upon fiction. As mentioned earlier in this essay, the fact that lives may soon be in the hands of AI is a frightening concept to wrap around. Distinctive portrayals of "killer robots" that take over the world may flood into the human mind from varying media. Movies such as I, Robot, Wall-E, Terminator, and Meet the Robinsons all suggest situations in which robot(s) retain control over their creators, mankind ("Top 12 Movies About Robots Taking Over That You Need Watching"); however, even though each of these films is a well-made piece of art, their apocalyptic ultimatums provide the extremes. Stephan De Spiegeleire, a Principal Scientist at HCSS, and his colleagues listed these special cases as a possible "wild card" in the development of AI (53). Currently, AI is far from becoming capable of developing its thoughts and actions, as present technology limits its functionality (51); nonetheless, that does not disregard the fact that a situation such as one of the fictitious examples above could occur.

If AI were allowed to control most aspects of human life, or possibly lives in general, many scholars agree that there are rules that need to come into play, and currently, action is being taken toward establishing policies for AI development. Many ethicists and lawmakers are proposing rules or ethics guidelines that promote the control of AI as it is used to handle possibly private data, money, and, eventually, lives. Unfortunately – or fortunately, depending on the viewpoint – AI does not currently hold the capability to know between wrong and right or make ethical decisions (Michele et al. 2413), so letting AI take any control in the first place is a thorny decision. The only "human-like" capabilities it has remains up to what the program has been

coded to do. Some voice assistants can complete a simple task, such as setting a timer for the time someone wakes up, then it asks if the person would like to set another timer for five minutes later, just in case they do not wake up. The assistant was most likely pre-trained to ask whether or not the individual needed another timer; there was no "heart" or moral kindness or intent behind the response. Furthermore, if one were to ask an AI chatbot whether it would like to help them, it would probably say that it would, indeed, "like" to help them. However, again, as much as it might seem to want to help, the bot has no "agency" or ability to choose (Rajakishore and Vineet 109). Knowing this, it may not make sense why humans would want programs to have any control over their lives; however, humans continuously seek the benefits, convenience, and capabilities technology brings to life.

To control whatever may happen in the future, many scholars and legal systems are directing their work toward making AI safer by developing possible limitations or legal controls over artificial systems (Velykanova 69). A significant influence in developing these limitations includes the "Artificial Intelligence Act" proposed by the *Future of Life Institute* for the European Union. The proposal introduces many rules, some of which are simple, such as making sure that technology systems notify the user if the content or tool they are using or seeing is utilizing the power of AI (Sioli 8). This would allow anyone using the technology to know whether the data they input is being analyzed and manipulated in complex ways that the user may disagree with. Some deeper rules are also mentioned, such as preventing the development of AI that could result in the "[e]xploitation of children or mentally disabled persons" (14). As mentioned before, machine intelligence is a powerful tool that can be used for good or bad. Persons volatile to becoming tricked or "exploited" could receive harm in many ways, such as digital robbery, data corruption, or even mental disruption, causing a loss of self-confidence or

belief. Many AI models have shown efficiency in helping humans, but if developed with more malevolent values, they could do irreparable damage. For this reason and numerous others, principles are developing. A group of researchers named “AI4People” devised an ethical framework to develop a “‘Good AI Society’” (Floridi et al. 689). On page 691 of their paper, the authors present a graph showing opportunities and risks that could form from increased AI development, along with a section



stating that AI can be underutilized. Understanding what machine intelligence could become and the balance of its development is essential to designing ethics and rules. The authors use the information gathered within Figure 1 to form structured guidelines. They split the guidelines into

Fig. 1. A graph showing some opportunities for AI in the middle and using arrows to the left and right to highlight the underuse and overuse of AI (Floridi et al. 691).

five main sections: “Beneficence,” the concept that AI should be developed for the ultimate purpose of benefiting humankind and the planet; “Non-maleficence,” the idea that the technology should not infringe on human rights, privacy, security, or general protection; “Autonomy,” referring to the point that AI can make its own “decisions,” but the decisions should continue to

receive monitoring; “Justice,” a principle stating that AI should be fair, unbiased, and impartial to all individuals; “Explicability,” which mentions that AI developers should keep in mind the purpose of development, the reasoning behind changes in development, why it works, and what audience it is intended for, and be able to hold transparency with possible clients (Floridi et al. 695-700). The principles mentioned articulate the specifics that developers, clients, and beneficiaries of the product should follow or keep in mind when creating intelligent programs. Furthermore, not only do the authors suggest possible guidelines, but they give insight into what their “Good AI Society” would do to control the use and development of AI. AI4People is not the only reasonable source for AI guidelines; there are multiple options available (695-6), and as technology evolves, the rules will be forced to follow. If the guidelines created are considered now, the future could provide a safe place where AI is used for the ultimate purpose of helping humanity.

Some organizations, such as Stop Killer Robots, are looking into the future and trying to prevent AI from making humanity’s decisions, especially regarding life or death. They seek to influence governmental systems to limit the development of weapons systems capable of harnessing AI's power (Stop Killer Robots Campaign). Organizations like this are needed in this world. They are part of the change that leads to a more safety-minded development society of AI. Nevertheless, though their claims and ideals favor the “greater good,” the development of weaponized or lethal AI will not go away, even if it may temporarily halt. In Kavanagh’s list of possible wildcards in the development of AI, she mentions a situation that she calls a “Sputnik Event” (53): a concept referring to the conflict of the “space race” and the Cold War with Russia. Both countries, feeling threatened or motivated by the other’s development, push or “race” each other to expand their technical abilities to achieve a goal of dominance over the other. Instead of

Russia, however, China is noted as the other possible leader in the conflict. According to Maria Stefania Cataleta, a researcher of international law, "China [...] has an ambitious AI strategy to become the global leader by 2030, while already being on target to outperform the US in academic research in the field by this year" (22). The Chinese government, run on communism, holds different ideals than the United States. If laws limiting AI development were designed in the United States for the rest of the world, China would most likely not even consider holding those rules, especially if it goes against their technological goals (26). In other words, AI should be controlled, but the United States could fall behind if rules become too strict. Now, AI should not be allowed to run rampant on human rights – it can still be used ethically – but it should be allowed to continue open development to benefit society and ensure a possible "AI Race" will stay within the United States' grasp to ensure safety and dissuade conflict.

AI has numerous benefits and gives hopeful insight into a future of ease with automation, but some believe AI is heading in the wrong direction and will cause more harm than good. An article published on the website *The Conversation* called "Artificial Intelligence Can Discriminate on the Basis of Race and Gender, and Also Age," by a group of professors and researchers, says that AI, though it provides multiple benefits, can be impartial toward certain cliques or groups of people (Chu et al.). The bias shown in machines is an enduring reason for some to doubt the prospect of "fair AI." Based upon particular trends in the data an intelligent program receives and organizes, it can, in essence, become "biased" if it has to make decisions based on the data. The program may seem to form stereotypes. Because of this, programmers cannot simply "remove" the artificial stereotypes from the code. The original format of a program wasn't necessarily made to harbor biases, so the code that supposedly would hold the bias would not exist in a concrete state. Two, bias can come from indirect forms of data.

Following the fact that possibly sensitive code cannot simply be removed, scholars Feuerriegel, Dolata, and Schwabe discuss deeper bias in their research paper:

> One might be inclined to think that simply omitting sensitive attributes from a decision support system will also solve fairness issues. However, this is a common misunderstanding: several non-sensitive attributes act as proxies (e.g., salary is a proxy of gender, ZIP code is a proxy for ethnicity, family structure is proxy of race or religion) and, hence, even decision support systems without knowledge of sensitive attributes are deemed unfair. (379)

When AI collects data, it can analyze trends, interpret patterns, and find outliers to create an output. Distinguishing between the ethical side of the "proxies" mentioned above is ability programs do not currently possess, even though many were modeled to replicate the human mind. One might ask, "Can the program be biased if it cannot know the difference between right and wrong?" The answer to the question is no. A program cannot hold its own bias if it does not have the capability to produce natural thought; however, the purpose of the program, how the program is used, and the data that is entered into the program can be. AI is a double-edged sword; its purpose is crafted in the hands of the creator, and its destiny is gifted to the trends in data collected.

AI will continue to change the world with every step it climbs toward its unknown limits. The technology has taken exponential leaps in progress as time progresses. Its capabilities, ranging from solving protein folding to analyzing the best way to sell a product, multiply as more fields seek the numerous benefits it provides. With it, the human race will reach heights previously thought incomprehensible. AI will apply a sense of ease, convenience, and efficiency within society that will catalyze technological development. Nevertheless, like most technology,

it does not come without downsides, but the nature of humankind's curiosity will continue its development. In the words of the Australian roboticist Rodney Brooks, "Artificial Intelligence is a tool, not a threat." AI will become an essential resource in the future as humanity further relies upon its benefits and utilizes its ever-increasing capabilities.

Works Cited

Abebe, Wondmagegn Taye, and Demeke Endalie. "Artificial Intelligence Models for Prediction of Monthly Rainfall Without Climatic Data for Meteorological Stations in Ethiopia." *Journal of Big Data*, vol. 10, no. 1, Dec. 2023, p. 1. *Gale In Context: Science*, link.gale.com/apps/doc/A732220157/GPS?u=wylrc_western&sid=bookmark-GPS&xid=fa960eba. Accessed 10 Mar. 2023.

Brooks, Rodney. "Artificial Intelligence is a Tool, Not a Threat." Quotefancy, 2019, https://quotefancy.com/quote/1599866/Rodney-Brooks-Artificial-intelligence-is-a-tool-not-a-threat.

Campaign to Stop Killer Robots. "Vision and Values." *Campaign to Stop Killer Robots*, https://www.stopkillerrobots.org/vision-and-values/. Accessed 13 Mar. 2023.

Cataleta, Maria Stefania. *Humane Artificial Intelligence: The Fragility of Human Rights Facing AI*. East-West Center, 2020. *JSTOR*, www.jstor.org/stable/resrep25514. Accessed 20 Feb. 2023.

Chu, Charlene, et al. "Artificial Intelligence can Discriminate on the Basis of Race and Gender, and also Age." *The Conversation*, 18 Jan. 2022, theconversation.com/artificial-intelligence-can-discriminate-on-the-basis-of-race-and-gender-and-also-age-173617. Accessed 21 Feb. 2023.

De Spiegeleire, Stephan, et al. "AI – Today and Tomorrow." *Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers*, Hague Centre for Strategic Studies, 2017, pp. 43–59. *JSTOR*, www.jstor.org/stable/resrep12564.8. Accessed 20 Feb. 2023.

Feldstein, Steven. “Types of AI Surveillance.” *The Global Expansion of AI Surveillance*, Carnegie Endowment for International Peace, 2019, pp. 16–21. JSTOR, www.jstor.org/stable/resrep20995.8. Accessed 10 Mar. 2023.

Feuerriegel, Stefan, et al. "Fair AI: Challenges and Opportunities." *Business & Information Systems Engineering* 62 (2020): 379-384.

Floridi, Luciano, et al. “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” *Minds and Machines*, vol. 28, no. 4, 2018, pp. 689-707. doi: 10.1007/s11023-018-9482-5.

Floridi, Luciano, et al. “Overview of the Four Core Opportunities Offered by AI, Four Corresponding Risks, and the Opportunity Cost of Underusing AI.” *Minds and Machines*, vol. 28, no. 4, 2018, p. 691. doi: 10.1007/s11023-018-9482-5.

IBM Global AI Adoption Index 2022. IBM Corporation, 2022.

"John McCarthy." *Encyclopedia Britannica*, 20 Oct. 2022, https://www.britannica.com/biography/John-McCarthy. Accessed 7 Mar. 2023.

Kavanagh, Camino. “Artificial Intelligence.” *New Tech, New Threats, and New Governance Challenges: An Opportunity to Craft Smarter Responses?*, Carnegie Endowment for International Peace, 2019, pp. 13–23. *JSTOR*, www.jstor.org/stable/resrep20978.5. Accessed 20 Feb. 2023.

Korteling, J., et al. "Human- Versus Artificial Intelligence." *Frontiers in Artificial Intelligence*, vol. 4, 2021, doi.org/10.3389/frai.2021.622364. Accessed 22 Feb. 2023.

Michele, Farisco, et al. "Towards Establishing Criteria for the Ethical Analysis of Artificial Intelligence." *Science and Engineering Ethics*, vol. 26, no. 5, 2020, pp. 2413-25. *ProQuest*, doi.org/10.1007/s11948-020-00238-w.

Newton-Rex, Ed. “59 Impressive Things Artificial Intelligence Can Do Today.” *Medium*, 16 Mar. 2023, medium.com/59-impressive-things-artificial-intelligence-can-do-today. Accessed 15 Mar. 2023.

Ouchchy, L. et al. “AI in the Headlines: The Portrayal of the Ethical Issues of Artificial Intelligence in the Media.” *AI & Society*, vol. 35, 2020, pp. 927–36. doi.org/10.1007/s00146-020-00965-5.

Rajakishore, Nath, and Sahu Vineet. "The Problem of Machine Ethics in Artificial Intelligence." *AI & Society*, vol. 35, no. 1, 2020, pp. 103-11. *ProQuest*, www.proquest.com/scholarly-journals/problem-machine-ethics-artificial-intelligence/docview/2356003007/se-2.

Sioli, Lucilla. “A European Strategy for Artificial Intelligence.” *European Commission*, 23 Apr. 2021. AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf.

The AI Act. “The Artificial Intelligence Act.” *Future of Life Institute*, www.artificialintelligenceact.eu. Accessed 24 Mar. 2023.

"Turing Test." *Encyclopedia Britannica*, 7 Feb. 2023, www.britannica.com/technology/Turing-test. Accessed 7 Mar. 2023.

Velykanova, Maryna, et al. "Artificial Intelligence: Legal Status Determination." *Law, State and Telecommunications Review* 14.2 (2022): 68-80.

Whittaker, Meredith, et al. *AI Now Report 2018*. New York: AI Now Institute at New York University, 2018.

Yerushalmy, Jonathan. “‘I Want to Destroy Whatever I Want’: Bing’s AI Chatbot Unsettles US Reporter.” *The Guardian*, 17 Feb. 2023,

www.theguardian.com/technology/2023/feb/17/i-want-to-destroy-whatever-i-want-bings-ai-chatbot-unsettles-us-reporter. Accessed 15 Mar. 2023.