

AIRES 2021 Research Conference: *Data and Privacy Applications*
**Ethical Review in The Age of Artificial
Intelligence**

Jung Hwan Heo¹

¹Viterbi School of Engineering, The University of Southern California, United States of America

DOI: <https://doi.org/10.47289/AIEJ20210716-4>



Abstract

Artificial intelligence (AI), particularly machine learning, has made significant strides in the past decade. Due to the widely applicable nature of this technology, the emergence of increasingly intelligent machines is poised to transform today's society. Recently, the rate of AI development has aroused significant concerns due to the lack of guiding policy and regulation. Thus, it is integral for the public to recognize the technology and make informed choices regarding the future of AI. This paper serves to acquaint the layperson and other stakeholders involved in AI development with the current progress of AI and the ethical concerns that must be addressed before significant advancements. The subject of discussion is narrowed down to three fields of AI's most prominent use: (1) the internet; (2) the automotive industry; and (3) the healthcare industry. For each sector, the foundation of the domain-specific AI technique is introduced, the benefits and ethical ramifications are discussed, and a final cost-benefit analysis is provided.

History

Received 31 January 2021

Accepted 14 July 2021

Published 16 July 2021

Keywords

Recommender Systems, Computer Vision, Natural Language Processing, Algorithmic Bias, Privacy Rights, Trust in AI

Contact

Jung Hwan (John) Heo

The University of Southern California –
Viterbi School of Engineering
2650 McClintock Ave, Los Angeles, CA,
USA 90089
(johnheo@usc.edu)

Acknowledgements

Many thanks to Professor Dr. Elisa Warford for introducing me to the world of AI and ethics, as well as her support and guidance throughout the process. I also want to thank Grace Song for professional feedback during the ideation phase.

Disclosure of Funding

None

AI Ethics Journal

Introduction

Artificial intelligence (AI) is a rapidly advancing form of technology that has the potential to transform the digital era. From virtual assistants, robot doctors, to self-driving cars, the application of AI in society is immense. The superhuman capability of AI has already been demonstrated in a few areas. One is Google's AlphaGo program that defeated the best human player of the board game Go in 2016. Given the complexity of the game, such a feat was anticipated to happen in at least a decade. Yet, the turnout of the game proved that advancements in AI far exceeded previous expectations. The continued growth of AI technologies that better predict and analyze data inspires other grand accomplishments such as virtually eliminating global poverty, massively reducing disease, and providing better education to everyone (Anderson et al., 2018).

Nonetheless, the current state of AI is very limited to performing a single task; thus it is often referred to as "narrow AI". The difficulty of generalizing and understanding contexts remains a short-term challenge. In the long-term, the advent of *singularity* is a topic of concern for many AI researchers, where future intelligent machines will recursively build a more intelligent version of themselves, potentially going beyond human control.

With great powers to benefit humanity at an unprecedented scale, many problems surround the malevolent use of AI and its unintended consequences as well. The ethics of AI has only recently been receiving public attention, particularly in social media. Rising to become a major medium for information exchange, social media platforms are being held accountable for political polarization and echo chambers through multiple studies (Cinelli, 2020). Other ethical concerns such as workforce displacement, data privacy, and human autonomy are projected to only become more challenging with the

accelerating adoption of AI across multiple industries (Pazzanese, 2020).

It is thus critical for every member of today's society to acquaint themselves with the potential impact of AI. Members should contribute their perspectives to develop a universal ethical framework that will guide how AI should affect society. To kickstart this process, this paper will introduce the readers to three different industries that are being most actively used and researched for AI. Beginning with the internet sector where AI has already been widely implemented, its real-world impact will be used as a precedent to guide the ethical analysis of subsequent industries in automobile and healthcare.

1. The Internet

1.1 Rise of a Digital World

With the advent of the internet, digital infrastructures have revolutionized the way people interact. A massive increase in the interaction between the physical and digital world is producing an unprecedented amount of data. The surplus of data allows machine learning (ML) algorithms, especially neural networks, to build precise models that can understand human behavior and even manipulate them. Recommender systems (RS) are one of the defining techniques of neural networks, and they're being widely applied to social media and other digital platforms. This chapter will first introduce the basic mechanisms and applications of RS in the social media and e-commerce industries. Then, its ethical challenges concerning the conservation of democratic values will be discussed to leverage the overall utility of RS.

AI Ethics Journal

1.2 Overview of Recommender Systems

The internet provides people with an overwhelming amount of information, which recommender systems effectively filter through and prioritize according to the user preferences. Recommender systems can be categorized into two main categories: content-based filtering (CBF) and collaborative filtering (CF). Each of these recommendation techniques will now be introduced.

Content-based filtering emphasizes the *features* of the item, where other items that encompass similar characteristics are recommended to the user. This requires a rich description of the item to identify positively labeled features, which makes CBF most effective in web pages, publications, and news (Isinkaye et al., 2015). By focusing on the relationship between the item and the user, the data of the user is localized; however, an absence of descriptive data on the item limits CBF from making useful recommendations.

Collaborative filtering is a prediction model for content that cannot be very descriptive, such as movies and music. It has two large subcategories called memory-based and model-based approaches. Memory-based CF utilizes the user's past interaction with different products. It builds a database of preferences for items by users, forming a clustered "neighbor" of users with similar preferences. Then, it simply recommends the most popular products among the neighboring users. In contrast, a model-based CF incorporates a more sophisticated recommendation by building an underlying generative model (Rocca, 2019). By leveraging much more user data, it can predict the preferences of items that haven't been rated or seen by users. Although CF can make serendipitous recommendations even when the relevance between the user and an item is weak, a lack of information about the user (data sparsity) can greatly damage the effectiveness of CF algorithms.

In the industry, social media companies heavily rely on both methods of recommender systems. The decision to use one model over the other varies with how much data is available and how explainable the decision-making process needs to be. Major technology companies have an abundance of data, so a model-based CF is preferred for its higher accuracy and scalability. For example, Facebook incorporates a machine learning model through a branching decision tree simulation. The model outputs clusters of users that show similar consumption behaviors or interests, allowing for more personalized advertisement (Biddle, 2018). In this case of advertising, the model may not be required to explain why it came to the conclusion it did, as long as it is effective. However, if a model is being used to diagnose cancer per se, then the decision must incorporate explanations to convince the doctors and patients. Challenges of explainability and transparency of AI will be explored in later chapters.

1.3 Effective Personalization

With a foundational understanding of the technology, popular applications and benefits of recommender systems will be discussed. One of the major applications of this technology resides in the internet retail industry, where recommender systems play a vital role in motivating purchase decisions through personalizing items to each user. The efficacy of such personalization is often directly linked to enhancing sales, shown by a wide range of today's internet retailers (Smith, 2017).

One of the prime examples of recommender systems is illustrated by the internet commerce company, Amazon. Experiencing more than a 1000% growth in revenue from 2009 to 2019, Amazon's success in its retail business is largely ascribed to personalized product recommendations (Davis, 2020). Their recommendation system algorithm, item-based collaborative filtering, effectively filters items by user-specific preference (Blake, 2017). Besides the purchase history of the customer,

AI Ethics Journal

Amazon incorporates information from other customers that purchased the same product, giving “frequently bought together” suggestions. From the perspective of a marketplace, recommender systems spur a win-win relationship, providing a better customer experience to the user and higher profitability to the retailer.

Another prevalent usage of the system is in social media. With the mission to build community and bring the world closer together, Facebook helps users connect through a myriad of recommendations. Not only can users more easily find and connect with their family, friends, and colleagues, but they can also discover users that share the same interests. It can be shown through social plugins such as likes and share buttons, where Facebook recommends more of the preferable content on their news feed section. Indeed, recommender systems are extensively used to personalize the information that is presented to the user, which requires little or no effort to find engaging content of their interest.

1.4 Echo Chamber and Democracy

While digital personalization provides convenience for finding relevant and engaging information, the lack of information diversity hinders the democratic system. The political landscape of the social media network has a massive impact on millennials. According to a survey conducted by the Pew Research Center, six out of every 10 millennials get their political news on Facebook alone, compared to local TV or other broadcast sources (Mitchell, 2015). With a significant portion of the young audience exposed to politically incentivized ideas, social media companies hold high stakes in preserving a healthy ecosystem. This includes the responsibility to proactively encourage healthy discourse among users and manage the quality of information by filtering misleading and harmful content.

Nonetheless, personalization algorithms are undermining fundamental pillars of a democratic system required to facilitate any civic communities. Prominent legal scholar Cass Sunstein asserts that citizens of a healthy democratic system should be exposed to materials that are purely encountered by chance. Such unanticipated encounters usher a mix of viewpoints, which ensures against fragmentation, polarization, and extremism (Sunstein, 2018). Yet, the diversity of information is lost during the filtering process of recommender systems. Exposure to content that only reinforces their pre-existing viewpoints makes users more susceptible to echo chambers and confirmation biases, leading to political hyper-partisanship.

The systematic information filtering on social media has dangerous implications due to the nature of human cognition. Studies suggest that when perceived information is misaligned with one’s beliefs, psychological stress is triggered (Knobloch-Westerwick et al., 2020). In turn, people adjust their beliefs accordingly to resolve this contradiction and therefore reduce the stress. Conversely, individuals presented with information aligned with their viewpoint help avoid psychological stress and even enhance affective mood. In this case, confirmation bias can be a form of entertainment for human cognition. The proclivity for humans to seek psychological comfort is precisely amplified by RS because it meticulously presents information that only reinforces one’s ideas; though users may be psychologically comforted, it weakens the ability for individuals to form a balanced viewpoint required to form a civic society. In fact, social media users are systematically labeled by recommender algorithms; for example, the Facebook interest settings allow users to select their political leanings from “very conservative” to “very liberal” (Merrill, 2016). The News Feed—the primary source of information—is accordingly synchronized to display politically aligned posts and advertisements.

AI Ethics Journal

The personalization architecture for social media is particularly susceptible to the spread of misinformation. Due to financial and political motivations, the accuracy of news on social media is compromised by various conspiracy theories, hyper-partisan content, pseudoscience, and fabricated news reports (Ciampaglia, 2018). When the social media platform disengages from regulating false content, the spread of misinformation is accelerated by the combination of recommender systems and cognitive biases. Because human cognition has a finite information processing capacity, brains tend to take “cognitive shortcuts” by responding to strong stimulants of interest (Gigerenzer et al., 1999). Then, news headlines with emotional connotations and hyper-partisan ideas are prioritized over accurate and nuanced ones. Thus, recommender systems constantly feeding bias-prone ideas to users are posing a significant threat to democratic discourse. Indeed, when the largest political information source for millennials promotes siloing, polarization, and confirmation bias, continued exposure to politically selective materials jeopardizes the future of nuanced political discussion.

1.5 Data and Privacy

User privacy is another major ethical challenge that is deeply associated with recommender systems. Most of the commercially successful RS utilize collaborative filtering techniques, building models of users with inference from a similar group of users. Here, privacy risks occur in multiple stages: data collection, data leakage, and model-level inferences (Milano 2020). During data collection, user data could be stored by third-party businesses or agents without the consent of the users. Even if the data has been consensually obtained, stored data could be leaked beyond the control of the user. Yet, the most troubling aspect is that collaborative filtering algorithms may infer extensive information about the user even with minimal knowledge of that user’s data. This is possible since the algorithm can build a model of a user based on

the interactions of similar users (or aforementioned “close neighbors”). The ability to build a fairly accurate profile of the user poses a significant ethical risk, as individuals have little privacy control over revealing personal information.

Once a security breach occurs in any of the three stages, it can lead to malicious consequences during user-software interaction. By capitalizing on personal information, Intelligent Software Agents (ISA) utilize methods such as deception and persuasion to manipulate user behavior. Popular forms of deception are clickbait and phishing content, which employ misleading descriptions to get users to access unwanted links. At a more intricate level, persuasion techniques nudge individuals to desired behaviors by intervening in their decision-making process. Since humans largely rely on intuition over sophisticated rationale to make quick decisions, individuals are prone to making irrational choices depending on what information is present to them (Ariely, 2018). ISA, therefore, can influence the “choice architecture” of individuals by leading them to a predictable course of action. For example, manipulating the order or layout of choices alters an individual’s preference by primacy bias (Burr et al., 2018). That is, the cognitive bias to prefer the first and last item in a series of options can be used for manipulation. In such cases, failure to protect personal information increases the risk of further manipulation, since the ISA has captured the user’s point of susceptibility.

To address these challenges, many privacy-conserving approaches have been introduced, including a user-centered recommendation framework by Paraschakis (2017). Here, users are given explicit privacy controls that allow them to decide the sharing settings themselves. However, this user-centric framework may not be a sustainable solution, since it puts an undue burden on the users to be responsible for all privacy controls. Besides, the user preference itself can be informative metadata, where external agents may infer that a user that has strong privacy settings has “something to hide” (Milano, 2020).

AI Ethics Journal

Other solutions include privacy-enhancing architectures that store personal information on decentralized databases to minimize data breaches.

1.6 Takeaways

Considering both the benefits and ethical challenges of recommender systems, the need for social media platforms to prioritize democracy over capitalistic needs is evident. Nonetheless, this problem is very difficult, since the challenge is inherently rooted in the social media ecosystem. Digital platforms provide free communication services while displaying advertisements as the primary source of revenue. Companies will inevitably maximize the engagement of users by presenting information tailored to their interests. Content outside of their ideological comfort zone would irritate users, leading to less engagement and therefore less advertising revenue. Moreover, companies are motivated to collect personal data to build better personalization algorithms that will influence user behaviors more profitably.

From a utilitarian perspective, RS are undoubtedly seen as effective. Social media can maximize happiness from personalized content, while companies profit from enhanced advertisement revenue. Yet, such a capitalistic approach to influencing human behavior poses a significant threat to the political landscape, compromising the ability of people to engage in nuanced political discourse. A greater societal urge to change social media's hyper-utilitarian architecture is needed. At a governmental level, antitrust suits against major tech companies can regulate capitalistic motives by revealing unconsented use of data brokerage. At an individual level, personal vigilance of understanding how recommendation engines work can greatly reduce the risk of forming confirmation bias. As such, a collective effort from both personal and governmental levels is necessary to retain democratic pillars amid free-market capitalism.

Admittedly, capitalism has allowed humanity to make extensive progress in creating goods for the world. Yet, the democratic pillars must be prioritized over capitalism to provide equitable space for every individual to function as a society. This makes it imperative for social media companies to proactively filter misinformation and disinformation, maintain diversity in RS algorithms, and uphold ethical standards in data usage despite their financial motives.

2. The Automobile Industry

2.1 Introduction to Autonomous Vehicles

Autonomous driving technology has been a decades-long aspiration for many AI researchers. Due to the recent successes in deep learning, massive improvements in computer vision (CV) techniques are allowing vehicles to perceive and navigate the world better than ever before. While the timeline for complete automation of driving is still unclear, it is evident that the full deployment of autonomous vehicles (AVs) will bring significant changes to society.

Along with the convenience for drivers, AVs promise a massive reduction in carbon emissions and road traffic, while increasing the safety of mobility. Yet, workforce displacement and the lack of an ethical framework to guide the development and deployment of AVs remain a significant challenge. In this chapter, the role of AVs powered by CV techniques will be discussed.

2.2 Overview of Computer Vision

CV algorithms utilize images as the input, where each image is decomposed and processed at a pixel level. For colored images, a square image can be represented as three matrices that describe the intensity of the RGB scale at each pixel. In a grayscale image, only one matrix is

AI Ethics Journal

required to quantify each pixel in the image from black to white. To collect such images for processing, automobiles use a combination of different sensors including cameras, RADARs, and LiDARs. The choice to use a particular combination of sensors varies across companies and research groups, while the task of object recognition lies at the center of CV.

Object recognition is most effectively executed through a machine learning algorithm called a convolutional neural network (CNN). CNN is a type of neural network that is inspired by how the human visual cortex processes visual information (Hubel, Wiesel, 1968). Closely resembling the connectivity patterns of neurons, the CNN model consists of various layers of information-processing units, where each layer statistically extracts some useful feature of the image. For example, in a face recognition task, earlier layers would extract basic features, such as lines and edges, while middle layers would extract more advanced features, such as eyes, nose, ears, and lips (Brownlee, 2019).

Similarly, the autonomous car learns object recognition tasks by extracting features of a pedestrian, streetlights, road lanes, other cars, and more.

With the ability to recognize surrounding objects, the vehicle can now reason about the environment and plan its navigation. Here, humans can either build the path planning algorithms themselves or let the system learn from trial and error. The latter method is a supervised learning technique, which is represented in Fig. 1. Supervised learning is a method that maps the input (image of the perceived environment) to output (steering direction, speed, and more) from labeled training examples. With enough examples that can meaningfully show the relationship between the input and output, the

system will learn the ideal driving motion in various environments.

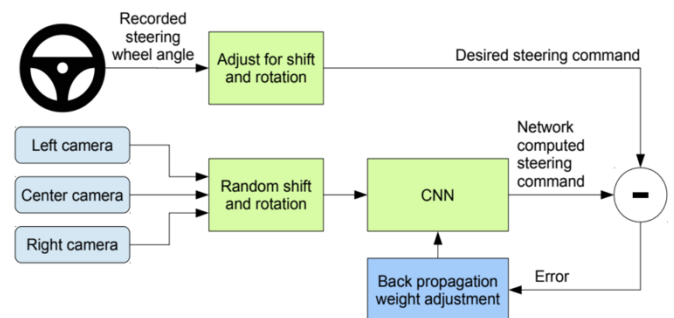


Fig. 1. Bojarski et al. "Training the neural network" arxiv.org, 25 April 2016, <https://arxiv.org/pdf/1604.07316.pdf> Accessed 10 Dec. 2020

2.3 User Convenience

Potential benefits of self-driving cars will now be provided. Driving, for many people, is a laborious process. While 86% of US adults own a car and 64% of them drive every day, only 34% personally enjoy driving (Brenan, 2018)¹. By automating the task of driving, an individual can pursue other activities. In addition to autonomous road navigation, automobile companies are also discovering new ways to automate other processes to maximize convenience for customers. For example, Tesla introduced an auto-park feature, whereby the software would execute the entirety of the parking process without any human intervention. Similarly, the smart summon feature allows the user to simply summon the car out of the parking space to where he or she is located. The addition of such convenient features, on top of self-driving, highlights a paradigm shift in the purpose of driving for people: rather than viewing a car as a property that needs to be operated by specific owners, it can be seen as a shared commodity that provides personal mobility

¹ Telephone interviews were conducted April 23-29, 2018, on the Gallup U.S. Poll with a random sample of 1,503 adults, aged 18 and older, living in all 50 U.S. states. The margin of sampling error is $\pm 3\%$ at the 95% confidence level.

AI Ethics Journal

service to consumers (Chan, 2017). To this end, self-driving cars would benefit individuals who do not wish to partake in the driving experience by allowing them to enjoy a fully automated, personalized transportation service.

2.4 Sustainable Transport

Society's decreasing need for car ownership has given rise to the idea of sharing a vehicle, which can potentially address many urban problems. Mainstream mobility platforms such as Uber and Lyft have already implemented ridesharing services; instead of needing personal vehicles, consumers can simply pay someone else to drive them to their desired destinations. Despite the available use of public transport, many people choose to use personal vehicles as their modes of transport for their relative convenience in reaching the destination directly. This contributes to increased traffic congestion in cities, more spaces taken up by parking lots, and more carbon emissions. However, inspired by autonomous driving technology, ridesharing service is being discussed as a solution to the aforementioned sustainability issues. Both Uber and Lyft are actively developing autonomous driving technology, ultimately envisioning their services to be run without the need for human drivers. By removing the costs of human labor, such services would be cheaper while also ensuring that there are fewer vehicles on the road. This allows urban cities to dedicate less space for parking lots, less traffic congestion, and most importantly, ensures a significant reduction in total carbon emissions.

2.5 Mobility Safety

In addition to convenience and sustainability, the most practical benefit of autonomous driving is mobility safety. Each year, approximately 1.35 million people are killed in roadways around the world. Crash injuries are the eighth leading cause of death globally for all age groups, which is a higher fatality rate than HIV/AIDS (World Health Organization [WHO], 2018). With 94% of motor vehicle

deaths attributed to human error (Brown, 2017), autonomous vehicles may offer a solution for eliminating human error and reducing automobile fatalities.

Autonomous driving software helps augment human's driving capability, greatly reducing the risk of accidents. When drivers fail to identify an obstacle in road conditions, computer vision can be used to continuously monitor the environment of the car, with 360-degree coverage. Decision errors that result from poor judgment can be addressed by precision AV software that ensures that cars are operated with legal and safe practices.

Performance errors that stem from undesirable conditions that compromise driving safety, such as physical fatigue or impaired cognitive abilities, can also be managed by the consistency of software control. This could particularly benefit the elderly since older drivers require about twice as much time to detect unexpected hazards than younger drivers (Wolfe et al, 2020). As a longer response time could lead to fatal injuries, it is indeed a much more daunting task for the elderly to drive by themselves. With the help of AVs, the elders are not only provided a safer mode of transportation, but also a sense of empowerment to travel without relying on other human drivers for support.

2.6 Job Replacement

Despite the benefits of vehicle automation, AVs pose a significant threat to the population most directly impacted by vehicle automation — human drivers. Fully self-driving cars will render chauffeurs, taxi, and ride-hailing drivers obsolete since navigation software will eventually become more reliable than a human. Given that job replacement is a serious threat, the demographic of taxi drivers makes it challenging to address unemployment. The global average age of taxi drivers is in the range from 51 to 60 years old, far exceeding the average age range of 35 to 44 in transportation, warehouse, and express services shown in

AI Ethics Journal

Census 2011². Moreover, their income has not shown any growth during the period 2007-2013, which implies a lack of career prospects (Hung, 2015). With the workforce continually aging and career prospects in compromise, the taxi industry fails to attract younger workers and is thus prone to be rapidly displaced by AVs once the technology is deployed.

A support system to make the transition fairer for people negatively impacted by automation is necessary to encourage societal acceptance. A South Korean ride-hailing company TADA was met by a fierce pushback from the taxi union, where one of the drivers self-immolated himself in an attempt to defend the union from being overtaken (Hosokawa, 2020). To that end, it is imperative to provide support for those whose jobs have been replaced by automation.

Before AI-enabled automation, governments and communities were already making investments in education, training, and social safety nets to mitigate the negative effects of automation (Fitzpayne et al., 2019). However, due to the highly competent and widely applicable nature of AI-enabled technologies, traditional methods to solve automation problems seem inadequate. The recent coronavirus pandemic has incentivized companies to further automate workplaces to discourage the spread of the virus. Studies show that 400,000 jobs have been lost due to automation in the US factories from 1990 to 2007, and it may replace up to 2 million more jobs by 2025 (Semuels, 2020). Thus, greater means of protection are critical to safeguarding the population susceptible to automation.

In the long term, where AI can automate a significant number of human tasks and replace jobs, direct monthly payment such as universal basic income (UBI) is being

discussed. As a growing number of workers lose their jobs to automation, the entities that own AI will procure greater wealth. This additional gain of wealth can be then distributed to those workers through value-added tax (VAT). A 10 percent VAT would raise about \$2.9 trillion over 10 years, even after covering the cost for UBI (Gale, 2020). Nonetheless, zero return for the money distribution has raised concerns for its demotivation to work. To ensure accountability, alternative forms of UBI, such as requiring a certain number of hours of valuable volunteer work, can be explored further.

2.7 Complicated Moral Issues

While automation may potentially be addressed by UBI, the lack of an ethical framework for autonomous driving systems poses a threat to building a robust human-AI system. While AVs could make more quick and accurate decisions to avoid unexpected hazards, moral problems arise in more complicated cases. Such cases can be represented by the Trolley Problem made famous by philosophers Judith Thomson and Philippa Foot. For example, when a car must decide between saving 10 elderly humans or one baby, what should it do? The human driver, in such morally conflicting cases, could decide by using his or her value system. Yet, a machine, an immoral and inanimate system, cannot make judgments based on value systems; it must always follow a command programmed by the human. Researchers at MIT have shown through an experiment of the Moral Machine that moral preferences vary by country and their respective cultural traits (Awad et al., 2018). Indeed, society as a whole struggles to agree upon a prevailing moral framework that guides the behavior of autonomous systems.

But the Trolley Problem is impractical since the possibility

² Surveyed cities are: Hong Kong, New York, London, Tokyo, Seoul, Singapore, Taipei. Interviewed 600 passengers and 700 practitioners in total. Sample size per city is unclear.

AI Ethics Journal

of its theoretical conditions actually applying to everyday life is very low. Also, it can be dangerous to discuss moral issues in such constrained situations, as it can lead to a narrow ethical framing that could lead to overly simplistic metrics in human affairs (European Group on Ethics in Science and New Technologies, 2018). Still, the Trolley Problem and its complementary Moral Machine experiment help us recognize that establishing a general moral framework across countries is imperative for the development of sustainable autonomous systems. This can be done by continued moral reflection and dynamic dialogues, such as IEEE's policy paper on "Ethically Aligned design" or ITU's global summit "AI for good". The people must join this conversation and encourage others as well to truly procure a general ethical framework that would create trustworthy and beneficial AI-robotic systems in the future.

2.8 Takeaways

Compared to recommender systems, self-driving technology offers utility beyond user convenience through its potential for providing sustainable transport and safety improvements. Ride-sharing services can dramatically reduce the number of cars on the road which solves many clustering problems in urban areas. The digitalization of vehicles coupled with electric vehicles promises net-zero carbon transportation. The cutting-edge navigation software, currently still in development, can reduce human-error accidents to save millions of lives. These benefits are essential to the continuation of people's lives, as climate change is a planetary threat and vehicle accidents are still prevalent around the world. Also, surveys show that developing countries have higher rates of death from vehicle accidents (Nantulya et al., 2002), thus AVs can help improve conditions of disparity.

The challenges remain to build a better ethical framework for AVs to safely integrate with the rest of society. This includes developing effective policies to regulate the

operation of AVs so that human drivers are ultimately still in control. The challenge of workforce displacement can be addressed, as there is still time before AVs are widely deployed; active investigations on workforce implications from AVs and automation, in general, are being pursued by researchers such as Frank et al. (2019) and Brynjolfsson (2017).

Indeed, new and innovative systems must be implemented to meet a new paradigm of work in the digital era. Taking account of the immense benefits that AVs bring, and given the extensive effort to address workforce displacement, the development of self-driving cars should be continued with great vigilance to its known challenges.

3. The Healthcare Industry

3.1 Data-driven Healthcare

The healthcare industry is experiencing a grand 36% compound annual growth in healthcare data, due to the increasing digitization across the industry (Reinsel et al., 2018). In addition to traditional medical devices inside hospitals, wearable devices such as Apple Watches can record a range of physiological data from heart rate, sleep cycle patterns, to blood-oxygen levels. Moreover, electronic health records (EHR) that capture clinical notes from physicians and other care providers generate a rich description of patient data, ideal for ML inference.

Yet, AI in healthcare is the least matured compared to the internet or the automobile industry. This means many different AI techniques are being simultaneously researched, including the aforementioned RS and CV techniques. To extend the scope of ML techniques covered in the paper, this chapter will focus on Natural Language Processing (NLP). NLP allows machines to make sense of text and speech data, which enables a variety of applications: text data on social media can be used to predict mental illnesses; understanding human

AI Ethics Journal

language allows a deeper human-robot interaction (HRI), leading to socially assistive robots for eldercare. These use cases represent only a bare minimum of all the possible cases and will be explored further after basic NLP techniques are explained.

Despite the vast array of ML applications in healthcare, grand ethical challenges remain in such an industry with especially high ethical standards. Hence, current challenges in mitigating bias, preserving privacy and security of health data are discussed, as well as the future directions for creating ethically aligned AI in healthcare.

3.2 Overview of Natural Language Processing

Natural language processing uniquely deals with speech and text data, in which the complexity of language is a challenge. Compared to image data that has a granular structure of pixel components, words can be interpreted differently depending on the context. As one of the popular methods to represent language, word embeddings convert words into a vector form. Converted word vectors can be spatially distributed in a vector space, where words with similar contexts occupy close spatial positions (Stewart et al., 2020). Various analysis methods will be discussed below.

Algorithms can effectively analyze the arrangement of words in a sentence such that it makes grammatical sense. Lines of text are first truncated into individual words through word segmentation. Then, similar words that represent the same fundamental meaning are grouped through the process of lemmatization; this is so that they can be analyzed as a single item. For example, the word 'to walk' can appear as 'walk', 'walked', 'walks', or 'walking'. Then, words can be computationally analyzed with the rules of formal grammar (Cattel, 2013).

NLP can also derive meaning from text through semantic analysis. It is one of the difficult aspects that are still in

experimental research. Named entity recognition (NER) categorizes parts of a text, such as names of places or people. Word sense disambiguation computationally generates a meaning of a word based on the context. Natural Language Generation involves using databases of texts to derive intentions and converting them into human language.

3.3 Administrative Efficiency

As one of the applications of NLP in healthcare, optimizing administrative tasks for health care providers could massively increase their workflow efficiency. Besides interacting with patients for care delivery, nurses, doctors, and physicians spend a relatively large amount of time on repetitive tasks during work. The average US nurse spends 25% of the time completing "regulatory requirements, redundant paperwork, and other non-direct care" (Commins, 2010). The repetitive nature of such work can be automated through robotic process automation (RPA), where NLP-powered computer programs can process texts and speech data: examples of administrative tasks include prior authorization, updating patient records, or billing (Davenport et al, 2019).

With automation in place, an increase in administrative efficiency could bring economic and social benefits. Implementing RPA leads to lower transcription costs and delays, improvements in the quality of health records, and a decrease in stress levels of workers (Burroughs, 2020). While healthcare entities can expect financial benefits from higher productivity, primary care workers can have better care delivery experiences that require more human-like traits such as empathy, adaptability, and communication. It is worth recognizing that RPA for health care workers is indeed a positive use of automation since it does not replace the human agents but simply complements them. Similar to agricultural and manufacturing automation that continues to yield higher productivity, NLP-powered automation in healthcare

AI Ethics Journal

promises a higher quality of both patient and health care worker experience.

3.4 Language-based Diagnosis

Not only do NLP techniques lead productivity gains for healthcare systems, but also serve as tools for diagnosis, particularly for mental illnesses. In the chronically underfunded sector of mental health, 1 in 5 of US adults experience mental illness annually (National Alliances on Mental Illness, 2020). Extended social isolation, health risks, and financial burdens during the coronavirus pandemic have exacerbated the occurrence of mental illnesses twofold (Czeisler et al., 2020), highlighting the need for a low-cost, scalable diagnosis tool. Here, NLP can precisely be applied for early diagnostics, which leads to significantly improved treatment costs and outcomes (Costello, 2016).

Utilizing semantic analysis from NLP, researchers could make inferences using both clinical and non-clinical texts. In a clinical setting, EHRs can be used as secondary informatics for making clinical decision support systems. In one use case, NLP models capture key symptoms of severe mental illness (SMI) from EHRs as clinical decision systems. From automatically extracting a broad range of SMI symptoms across a high volume of text summaries, the model extracted symptomology in 87% of patients with SMI (Jackson et al., 2017). In a non-clinical environment, researchers take advantage of the fact that writing is a means to not only convey factual information, but also convey feelings, mental states, hopes, and desires (Calvo et al., 2017). Therefore, popular social media could be used as a database to look for subtle cues of mental health in a day-to-day setting. For instance, De Choudhury et al. (2013) determined Twitter posts with an onset of depression by building a classifier that utilizes the Social Media Depression Index (SMDI) to a group of tweets by the likelihood of having depression.

Yet, applied clinical NLP for diagnosis is still very dispersed due to its nascent nature: it is also met with challenges of high variability across personal background information such as demographics, geographics, and social conditions. Thus, developing a research environment with a combination of computational and clinical expertise is crucial to bring a real service impact (Stewart, Velupillai, 2020).

3.5 Robot Care for the Elderly

NLP-powered assistive robots have the potential to innovate healthcare systems for the aging population. It is by over 50% for the next fifteen years, while home health aides are projected to proportionately grow by 38% (Ortman et al., 2014). Yet, the number of caregivers is expected to drop from 7 caregivers for every person in 2010, to 4 for every person in 2030 (Redfoot, 2013). Here, social robots that can dynamically interact with elders can be used to support physical, emotional, social, and mental health. Conversational robots are especially effective, since nearly half of all people over 75 live alone, while more than a third do not speak to anyone on an average day (Age UK, 2019).

Personalized care robots still lack the capabilities to be truly efficient. Technological challenges remain as NLP techniques still struggle with understanding the context of the conversation. Current robots can ask simple questions such as “How are you?”, but struggle with predicting the next course of sequence to carry out a compelling conversation. Nonetheless, NLP is an active area of research that happens in both industry and academia. With greater language datasets being gathered and ML algorithms advancing, AI-powered chatbots are ultimately expected to emulate a vast majority of human speech interactions.

AI Ethics Journal

3.6 Algorithmic biases

Social determinants of health

As NLP is fundamentally a machine learning technique, it cannot avoid the ethical challenges regarding algorithmic fairness and bias. It can be especially challenging in the healthcare sector, where the stakes to maintain an unbiased system are high. Indeed, AI is an incredibly influential double-edged sword: it could both democratize and discriminate healthcare for vulnerable communities. The problem of bias exists at the core of machine learning, which is the data that is used to train the models from the beginning. If the datasets being used to train a model are not representative of certain minority groups, ML models can create a bias that marginalizes those groups. The challenge of attaining a representative dataset is hard due to social factors that affect the health of an individual, known as the Social Determinants of Health (SDOH). Because “various social and structural factors affect health across multiple intersecting identities” (Mammot, 2005), having a narrow representation of data leads to discriminatory models.

Recent studies have shown through several real-world examples that algorithmic bias can occur regarding ethnic origins, skin color, gender, age, and disabilities. For example, an algorithm that UnitedHealth Group developed to identify vulnerable patients requiring care management was found to have a severe racial bias against Black patients (Williams et al, 2018). Preconceived biases and harmful stereotypes in the medical industry concerning Black individuals were relayed into the algorithm. The manifestation of racial biases as an SDOH contributed to negative health outcomes. The effort to remedy this disparity was found to increase Black patients receiving help from 17.7 to 46.5%, which highlights the need to offset algorithmic biases (Obermeyer et al., 2019).

There have been attempts to engineer fairness into the models themselves. At the beginning of the data collection

stage, choosing a representative dataset that includes a diversity of demographics, age, and gender can potentially prevent the algorithm from creating a bias for a particular trait. During model implementation, developers can carefully choose the architecture of the model, as every model has its characteristic method of processing data. Kilpatrick et al. (2019) found that supervised learning may be more effective but can be prone to human biases; they also observed that unsupervised learning can be trained more quickly, but maybe more prone to errors.

However, some argue that computational methods to account for fairness have their limitations, considering the sheer complexity of SDOH. McCradden et al. (2020) have demonstrated the inconclusiveness in past efforts to understand the mechanisms in which SDOH affects health outcomes. Given the scope of complexity in maintaining a fair data-driven system, perhaps the most important practices are proactive monitoring and testing. Throughout the model training and validation stage, developers can actively monitor and test their models for biases: if any are to be found, they can resolve the issue before the model is deployed to the public. Here, identification and efforts to resolve bias should be actively communicated to the stakeholders to ensure transparency and form a trustworthy relationship with its end-users.

3.7 Transparency and Trust

In addition to maintaining a fair system, ensuring the safety of the models through transparency is critical in healthcare applications. A counterexample of a transparent ethical practice is IBM’s AI doctor Watson, which gives physicians treatment recommendations based on the medical records of patients. Contrary to its supposed functionality, the Watson model actively gave “unsafe and incorrect” recommendations (Brown, 2018). The problem was raised that instead of training on actual patient data, Watson utilized only a few synthetic datasets that were artificially devised by doctors at the Memorial

AI Ethics Journal

Sloan Kettering Cancer Center (Ross et al., 2018). The poor quality of the recommendations could have been easily prevented by unveiling their process of training the AI model. Outside developers, doctors, and other stakeholders would have tested the robustness of the model, if not arouse concerns for using only a few synthetic datasets: this is while knowing that high-performance ML models typically require large, high-quality clinical datasets.

Not adhering to transparent practice could not only lead to poor model performance, but it can compromise its relationships with the users. Continuing what Brown (2018) has found in the previous example, IBM kept Watson's faulty treatment recommendations hidden from the public for over a year. The company also holds the responsibility to communicate any shortcomings to establish a trustworthy relationship with patients and clinicians, which is key to a successful implementation in practice.

3.8 Takeaways

Though NLP is still in its early adoption, its potential in healthcare applications is immense. One estimate from McKinsey predicts that big data-driven strategies could generate up to \$100 billion annually across the US healthcare system alone (Cattell et al., 2013). From better diagnostics to streamlined patient care environments and effective treatments, NLP applications can range the entire healthcare ecosystem. Compared to other industries, AI use cases for health care have their unique challenges and opportunities. While recommender systems and autonomous driving are entirely composed of quantitative tasks, conversational robots require traits such as empathy. To this end, empathetic AI is also a promising area of research that can help bring massive benefits to how humans understand emotion.

For these applications to truly be effective, healthcare technology must be treated with special safety requirements. Extensive validation and verification should be done by not only regulators such as the FDA, but vigilance from the level of care deliverers (doctors) as well as receivers (patients) are needed. AI policy on healthcare is still nascent; primitive guidelines in developing and implementing AI to health care practice continue to be developed. More coverage on the ethical guidelines can be found through Crigger (2019).

Similar to the automotive industry, AI technology has massive potential benefits that put it in favor of the development of AI systems. However, it is required to be as robust and ethically aligned as possible, since it directly deals with patient lives that are fundamental to human survival.

Conclusions

This paper provides the necessary background for any stakeholders to reasonably understand and join the conversation of AI ethics. The basic mechanisms of recommender systems, computer vision, and natural language processing were covered, along with their applications in the internet, automotive, and healthcare fields respectively.

The progression of the techniques introduced follows from the most widely deployed to the least. Compared to the RS on the internet, autonomous vehicles and medical NLP technologies are still in development. This allows the ethical challenges to be thoroughly considered before raising negative social consequences in practice. One should take the issue of RS on the internet as a precedent to proactively address the ethical implications for the other two sectors. Due to the scope of the topic covered, there are many applications and ethical challenges that remain other than those mentioned in this paper. Should one be willing to discover a scoping review focused on each of the three ML techniques, other works attempt to

AI Ethics Journal

analyze the ethics of RS on the internet (Milano et al., 2020), CV for self-driving cars (Lim et al., 2019), and clinical NLP (Gerke et al., 2020). In the time of accelerating technological advancements, the net influence on society is not rooted in the technical solutions themselves, but in thorough planning for ethically aligned AI systems. Every individual should contribute to building a better ethical framework to ensure a future that is positively impacted by AI.

AI Ethics Journal

References

- [1] Anderson, Janna, Lee Rainie, and Alex Luchsinger, 2018, *Artificial Intelligence and the Future of Humans*, Washington, DC: Pew Research Center.
- [2] Cinelli, Matteo, et al. "Echo chambers on social media: A comparative analysis." *arXiv preprint arXiv:2004.09603* (2020).
- [3] Pazzanese, C. (2020, December 04). Ethical concerns mount as AI takes bigger decision-making role. Retrieved February 01, 2021, from <https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>
- [4] Isinkaye, F.O., et al. "Recommendation Systems: Principles, Methods and Evaluation." *Egyptian Informatics Journal*, Nov. 2015, pp. 261–273., doi:<https://doi.org/10.1016/j.eij.2015.06.005>.
- [5] Rocca, Baptiste. "Introduction to Recommender Systems." *Medium*, Towards Data Science, 12 June 2019, towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada.
- [6] Biddle, Sam. "Facebook Uses Artificial Intelligence to Predict Your Future Actions for Advertisers, Says Confidential Document." *The Intercept*, 13 Apr. 2018, theintercept.com/2018/04/13/facebook-advertising-data-artificial-intelligence-ai/.
- [7] Arora, Shabana. "Recommendation Engines: How Amazon and Netflix Are Winning the Personalization Battle." *MarTech Advisor*, 28 June 2016, 05:40 PM, www.martechadvisor.com/articles/customer-experience-2/recommendation-engines-how-amazon-and-netflix-are-winning-the-personalization-battle/.
- [8] Smith, Brent, and Greg Linden. "Two decades of recommender systems at Amazon. com." *Ieee internet computing* 21.3 (2017): 12-18.
- [9] Davis | Oct 29, D. (2020, October 29). Amazon sales, Amazon revenue and Amazon annual profits. Retrieved January 13, 2021, from <https://www.digitalcommerce360.com/article/amazon-sales/>
- [10] Blake, M.. "Two Decades of Recommender Systems at Amazon." (2017).
- [11] Mitchell, Amy, et al. "Facebook Top Source for Political News Among Millennials." *Pew Research Center: Journalism & Media*, 1 June 2015, www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials/.
- [12] Sunstein, Cass R. "Is Social Media Good or Bad for Democracy?" *Facebook Newsroom*, 22 Jan. 2018, about.fb.com/news/2018/01/sunstein-democracy/.
- [13] Knobloch-Westerwick, Silvia, et al. "Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information." *Communication Research*, vol. 47, no. 1, Feb. 2020, pp. 104–124, doi:[10.1177/0093650217719596](https://doi.org/10.1177/0093650217719596).

- [14] Merrill, Jeremy B. "Liberal, Moderate or Conservative? See How Facebook Labels You." *The New York Times*, 23 Aug. 2016, www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html.
- [15] Ciampaglia, Giovanni Luca, and Filippo Menczer. "Biases Make People Vulnerable to Misinformation Spread by Social Media." *Scientific American*, 21 June 2018, www.scientificamerican.com/article/biases-make-people-vulnerable-to-misinformation-spread-by-social-media/.
- [16] Gigerenzer, Gerd, and Peter M. Todd. *Simple heuristics that make us smart*. Oxford University Press, USA, 1999.
- [17] Milano, Silvia, Mariarosaria Taddeo, and Luciano Floridi. "Recommender systems and their ethical challenges." *AI & SOCIETY* (2020): 1-11.
- [18] Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. London: Harper Collins.
- [19] Burr, Christopher, Nello Cristianini, and James Ladyman. "An analysis of the interaction between intelligent software agents and human users." *Minds and machines* 28.4 (2018): 735-774.
- [20] Paraschakis, Dimitris. "Towards an ethical recommendation framework." *2017 11th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2017.
- [21] Hubel, D. H.; Wiesel, T. N. (1968-03-01). "Receptive fields and functional architecture of monkey striate cortex". *The Journal of Physiology*. **195** (1): 215–243.
- [22] Brenan, Megan. "Driverless Cars Are a Tough Sell to Americans." *Gallup News: Economy*, 15 May 2018, news.gallup.com/poll/234416/driverless-cars-tough-sell-americans.aspx.
- [23] Chan, Ching-Yao. "Advancements, prospects, and impacts of automated driving systems." *International journal of transportation science and technology* 6.3 (2017): 208-216.
- [24] World Health Organization (WHO). Global Status Report on Road Safety 2018. December 2018. [cited 2020 October 28].
- [25] Brown, B. (2017, October 06). 2016 NHTSA Fatality Report Adds to Evidence for Self-driving Cars. Retrieved January 15, 2021, from <https://www.digitaltrends.com/cars/2016-nhtsa-fatality-report/>
- [26] Wolfe, Benjamin, et al. "Rapid holistic perception and evasion of road hazards." *Journal of experimental psychology: general* 149.3 (2020): 490.
- [27] Hung, Wing-tat. "Taxi Industry Is Aging?" *Taxi Industry Is Ageing?*, Hong Kong Polytechnic University, Sept. 2015, www.polyu.edu.hk/cpa/excel/en/201509/viewpoint/v1/index.html.
- [28] Hosokawa, Kotaro. "South Korea Bans Ride-Hailer in Win for Taxi Lobby." *Nikkei Asia: Startups*, 11 Mar. 2020, asia.nikkei.com/Business/Startups/South-Korea-bans-ride-hailer-in-win-for-taxi-lobby.
- [29] Fitzpayne, Alastair, et al. "Automation and a Changing Economy: The Case for Action." *The Aspen Institute*, 2 Apr. 2019, www.aspeninstitute.org/publications/automation-and-a-changing-economy-the-case-for-action/.
- [30] Semuels, Alana. "Millions of Americans Have Lost Jobs in the Pandemic—And Robots and AI Are Replacing Them Faster Than Ever." *Time: Business, Covid-19*, Time, 6 Aug. 2020, time.com/5876604/machines-jobs-coronavirus/.

- [31] Gale, William G. "How a VAT Could Tax the Rich and Pay for Universal Basic Income." *Brookings*, Brookings, 30 Jan. 2020, www.brookings.edu/blog/up-front/2020/01/30/how-a-vat-could-tax-the-rich-and-pay-for-universal-basic-income/.
- [32] Awad, Edmond, et al. "The moral machine experiment." *Nature* 563.7729 (2018): 59-64.
- [33] European Group on Ethics in Science and New Technologies. "Statement on artificial intelligence, robotics and 'autonomous' systems." Retrieved September 18 (2018): 2018.
- [34] Nantulya, Vinand M., and Michael R. Reich. "The neglected epidemic: road traffic injuries in developing countries." *Bmj* 324.7346 (2002): 1139-1141.
- [35] Frank, Morgan R., et al. "Toward understanding the impact of artificial intelligence on labor." *Proceedings of the National Academy of Sciences* 116.14 (2019): 6531-6539.
- [36] Brynjolfsson, Erik, and Tom Mitchell. "What can machine learning do? Workforce implications." *Science* 358.6370 (2017): 1530-1534.
- [37] Reinsel, David, John Gantz, and John Rydning. "Data age 2025: the digitization of the world from edge to core." *IDC White Paper Doc# US44413318* (2018): 1-29.
- [38] Commins, John. "Nurses Say Distractions Cut Bedside Time by 25%." *HealthLeaders Media*, 9 Mar. 2010, www.healthleadersmedia.com/welcome-ad?toURL=%2Fnursing%2Fnurses-say-distractions-cut-bedside-time-25.
- [39] Davenport, Thomas, and Ravi Kalakota. "The potential for artificial intelligence in healthcare." *Future healthcare journal* 6.2 (2019): 94.
- [40] Burroughs, Amy. "Language Processing Tools Improve Care Delivery for Providers." *HealthTech: Patient-Centered Care*, HealthTech, 4 May 2020, healthtechmagazine.net/article/2020/05/language-processing-tools-improve-care-delivery-providers.
- [41] "Mental Health By the Numbers." *National Alliance on Mental Illness*, Dec. 2020, [www.nami.org/mhstats#:~:text=aged%2010%2D34-You%20Are%20Not%20Alone,2018%20\(11.4%20million%20people\)](http://www.nami.org/mhstats#:~:text=aged%2010%2D34-You%20Are%20Not%20Alone,2018%20(11.4%20million%20people)).
- [42] Czeisler, Mark É., et al. "Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—United States, June 24–30, 2020." *Morbidity and Mortality Weekly Report* 69.32 (2020): 1049.
- [43] Costello, E. Jane. "Early detection and prevention of mental health problems: developmental epidemiology and systems of support." *Journal of Clinical Child & Adolescent Psychology* 45.6 (2016): 710-717.
- [44] Jackson, Richard G., et al. "Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project." *BMJ open* 7.1 (2017).
- [45] Calvo, Rafael A., et al. "Natural language processing in mental health applications using non-clinical texts." *Natural Language Engineering* 23.5 (2017): 649-685.
- [46] De Choudhury, Munmun, et al. "Predicting depression via social media." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 7. No. 1. 2013.

- [47] Stewart, Robert, and Sumithra Velupillai. "Applied natural language processing in mental health big data." *Neuropsychopharmacology* 46.1 (2020): 252.
- [48] Ortman, Jennifer M., Victoria A. Velkoff, and Howard Hogan. *An aging nation: the older population in the United States*. Washington, DC: United States Census Bureau, Economics and Statistics Administration, US Department of Commerce, 2014.
- [49] Redfoot, Donald, Lynn Feinberg, and Ari N. Houser. *The aging of the baby boom and the growing care gap: A look at future declines in the availability of family caregivers*. Washington, DC: AARP Public Policy Institute, 2013.
- [50] Age UK. (2019) Later Life in the United Kingdom | Age UK. Retrieved January 15, 2021, from https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/late_life_uk_factsheet.pdf
- [51] Marmot, Michael. "Social determinants of health inequalities." *The lancet* 365.9464 (2005): 1099-1104.
- [52] Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. "How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications." *Journal of Information Policy* 8 (2018): 78-115.
- [53] Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366.6464 (2019): 447-453.
- [54] Kilpatrick, Steve. "How to Address and Prevent Machine Bias in AI." *Logikk*, 22 Jan. 2019, www.logikk.com/articles/prevent-machine-bias-in-ai/.
- [55] McCradden, Melissa D., et al. "Ethical limitations of algorithmic fairness solutions in health care machine learning." *The Lancet Digital Health* 2.5 (2020): e221-e223.
- [56] Brown, Jennings. "IBM Watson reportedly recommended cancer treatments that were "unsafe and incorrect"." *Gizmodo*, July 25 (2018).
- [57] Ross, Casey, and Ike Swetlitz. "IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show." *Stat News* <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments> (2018).
- [58] Cattell, Jamie, Sastry Chilukuri, and Michael Levy. "How big data can revolutionize pharmaceutical R&D." *McKinsey & Company* 7 (2013).
- [59] Crigger, Elliott, and Christopher Khoury. "Making policy on augmented intelligence in health care." *AMA journal of ethics* 21.2 (2019): 188-191.
- [60] Lim, Hazel Si Min, and Araz Taeihagh. "Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities." *Sustainability* 11.20 (2019): 5791.
- [61] Gerke, Sara, Timo Minssen, and I. Glenn Cohen. "Ethical and Legal Challenges of Artificial Intelligence-Driven Health Care." *Forthcoming in: Artificial Intelligence in Healthcare, 1st edition, Adam Bohr, Kaveh Memarzadeh (eds.)* (2020).
- [62] Brownlee, Jason. "Overview of Object Recognition Computer Vision Tasks." *Machine Learning Mastery*, 22 May 2019.