Table of Contents

Volume XLVII, Number 1, Spring 2021

The Journal of Technology Studies

Autonomy in AI Systems: Rationalizing the Fears

By Kenneth R. Walsh, Sathiadev Mahesh, and Cherie C. Trumbach

ABSTRACT

The news, popular culture, and legislatures are concerned with the recent reemergence of artificial intelligence (AI) technology. Some of the fear is fueled by the terminology including artificial intelligence, machine learning, deep learning, and superintelligence that have specific meaning within the technology community, but can be misunderstood by the general public or by other fields of inquiry. Because of this, the fears are not well linked to what the technology actually does. The type of AI technology such as neural networks or decision trees does little to clarify the conversation. However, considering where an AI system exhibits autonomy better highlights what the systems capabilities are and what may be rationally feared from such capabilities. This paper develops a typology of autonomous functions within AI systems and why they matter.

Keywords: *artificial intelligence, machine learning, computers in society, artificial intelligence types*

INTRODUCTION

To the general public, the term "artificial intelligence" has conjured up thoughts of machines that can be taught to mimic the human body, primarily the brain. We imagine these machines have human personality, thought processes, and decision-making capabilities, with speed and capacity that may outpace the humans who created it. The terminology of the field has developed accordingly with terms such as machine learning, superintelligence, and artificial neural networks. However, while Artificial Intelligence (AI) terminology has been useful describing a vision for what may be possible with computer technology, the terminology may imply computer capabilities that do not exist while hiding computer mechanisms that are being used in real-world applications. AI may appear magical in its capabilities, replicating barely understood human thought processes, but is often incapable of adapting to changes in the environment, extending its reach outside its narrowly specified domain, and requires considerable human effort to be re-trained (Brooks, 2017). Young and Carpenter (2018) found that people with greater exposure to science fiction literature had a greater fear of AI technology. While, O'Sullivan (2017) argued that fear and misunderstanding of AI could lead to a stifling AI regulatory environment.

AI terminology has developed, historically, as a shorthand for the capabilities of AI, while hiding the complexity of its practical application and real capabilities. For example, on the one hand, artificial intelligence implies machine operations, which can be thought of as both unbiased and logical, but also implies unfeeling and lacking compassion. On the other hand, those machines are also programmed by humans who are training machines to act like humans. In a world where related fields of governance, law, and ethics are increasingly concerned about the capabilities, biases, errors, impact on employment, and even dangers from rogue AI, such language may be misleading and leads to many irrational fears. What is most unusual about AI systems is their ability to make autonomous decisions and display autonomous behavior without real-time human intervention, and it is this autonomous capability that requires discussion among stakeholders from a variety of backgrounds. This paper develops a classification scheme for the autonomous capabilities of AI systems to improve cross-disciplinary communication. It provides a common way to sort out irrational fears from rational fears in regards to AI in order to have productive discussions regarding policy.

BRIEF HISTORY OF AI TERMINOLOGY

In 1948, Turing (1948) proposed the term "Intelligent Machinery" to describe what he said was the possibility of building a machine that could closely simulate the behavior of the human brain. Rather than building an entire robot that mimicked human capability, he proposed developing systems that could act as subsets of intellectual thought as a practical way of exploring human intelligence. He then theorized that a future machine with infinite memory and speed could process all such subsets of thought.

John McCarthy first used the term Artificial Intelligence, professionally as it is applied today, in his 1955 proposal for the conference, Dartmouth Summer Research Project on Artificial Intelligence (McCorduck, 1977). Webster's dictionary defines intelligence as the ability to acquire and apply knowledge and skills. One can argue whether or not the systems that have been created so far are actually

intelligent and whether or not they should be called Artificial Intelligence. "To ascribe certain beliefs, knowledge, free will, intentions, consciousness, abilities or wants to a machine or computer program is legitimate when such an ascription expresses the same information about the machine that it expresses about a person" (McCarthy, 1979). McCarthy, certainly argued well for the usefulness of the terminology. He is not concerned so much with the question of machine intelligence being equivalent to human intelligence but rather whether a sound description of AI systems can be built using such language. In his example, he describes how a thermostat decided to signal the furnace when it thinks the temperature is too cold. Such an example is useful for describing the working of a thermostat but should not be thought of as suggesting that thermostats can think of things other than whether it is too cold or not. In other words, authors may have established a good way of communicating how a system behaves, without literally implying that the word "think" resembles the full range of thinking ascribed to humans' thinking. Since McCarthy probably was considering the AI community as his audience, his arguments are probably useful. However, when the general population needs to understand AI, such descriptions may be misleading.

In 1959, Arthur Samuel coined the term "Machine Learning" to refer to the subset of AI techniques where the problem-solving algorithm was not directly programmed by the analyst but was found from the analysis of data. Samuel (1959) compared the results of two machine learning systems to play checkers, one a general neural network and the other a state network structured by the states in the checkers game.

In 1986, Rina Dechter coined the term "Deep Learning" to refer to the collection and use of multiple conflict sets in solving problems using backtracking such as finding a path through a complex maze. At the present time, deep learning is often used to refer to neural networks with many hidden layers and complex internal interconnections.

A problem with the line of terminology used within the AI community is that the terms have very different meaning in standard English. The AI community probably cannot and should not change terminology, because they have defined the terms well for their purposes and create new terms built upon old terms. For example, deep learning is an advancement in machine learning. However, the terminology in line of enquiry is not likely to come closer to the language used by the general public. While at the same time, the general public and even researchers in other fields are becoming more interested in and more affected by the field of AI.

There are different ways AI technologies are categorized. A common breakdown of AI approaches is defined by Russell and Norvig (2020), Acting Humanly, Thinking Humanly, Thinking Rationally, Acting Rationally. The Turing Test is an example of a computer acting humanly. Today's uses of AI can be categorized narrow AI because of their focus on solving specific problems while researchers continue to strive for systems that might be categorized as Artificial General Intelligence or "strong AI." Artificial General Intelligence replicates human intelligence in a machine (Sullins, 2015). Another way of classifying AI is by its ability to use memory and draw on past experiences as well as its awareness of the outside world and emotional abilities. AI can currently encode and subsequently use human knowledge. It can add data from past experiences and use them in future actions thereby improving its decision-making capability. Though there is research interest, AI cannot understand human emotions and alter its emotional behavior suitably, become self-aware, and make representation of itself.

There are numerous methods, fields, and approaches under the artificial intelligence heading that are categorized in different ways. Some of them are Machine Learning (including Artificial Neural Networks and Evolutionary Algorithms), Expert Systems, Natural Language Processing and Computational Creativity, among others. Machine Learning consists of algorithms that allow the computer to make predictions, learn from mistakes and make adjustments without human intervention. The most common type of Machine Learning, is supervised learning, the most well-known of which is Artificial Neural Networks (though ANNs can also be used for unsupervised learning, also) (Ongsulee, 2017). The algorithms can recombine into variations previously non-existent. Therefore, the AI can change the model itself (Eiben & Smith, 2015). Applications of adversarial learning range from the unsupervised training of ANNs to learn new strategies for playing games to network security by testing new rules against network attack vectors (McDaniel, Papernot, & Celik, 2016).

Expert Systems are developed by obtaining knowledge from human experts and formulating that knowledge into structured rules. The primary source of expert system knowledge

is human experts. It is extracted by structured interviews on the decision-making process used to select actions in response to real-work cases and analysis of case files maintained by human experts. Expert systems have a very narrow domain of expertise and by default do not automatically learn from experience. They, however, do provide detailed documentation of their problems and this data can be analyzed by human experts to alter the knowledge coded in the expert system.

Computational Creativity is the development of AI that can create stories, art, or music without human intervention. Creativity refers to an output that is novel and useful, a previously unknown clarification for an unexplained problem, or an output that changes currently held opinions. Some tasks performed by humans that are commonly termed creative are not truly creative, such as the creation of a narrative from data in sports writing or financial data which has been successfully automated (Conde-Clemente, Trivino, & Alonso, 2017). Creating new art by combining genres or writing new fiction or poetry is creative and a challenge for many humans. Computer co-creation teams AI with a human to enhance the quality of artistic output by altering the writing style (Manjavacas, Karsdorp, Burtenshaw, & Kestermont, 2017) or art. Super Intelligence refers to AI that can exceed the capacity of the best human brains. Technological progress has to continue unimpeded for decades before we will see such AI, and any plateauing of processor technology improvement will further delay this development.

FEAR OF AI

For many, AI creates a deep fear of rouge automatons, which may think for themselves and determine that human life is not valuable. In Karel Capek's play which introduced the term robot and in the sci-fi movie *2001: A Space Odyssey* machines went rogue and attacked humans. The fear of sentient AI, which may make a choice to control and even eliminate humans rather than tamely work as commanded leads to fears of large numbers of self-driving vehicles or smart homes. Essays on the future of robotics often verge into fictional territory and misuse common assumptions such as Moore's Law for computer chip capability. By extending the observation of past growth encapsulated in Moore's law, Kurzweil (1990) projects the capability of computers to exceed that of the human brain in a couple of decades and uses this to predict a future in which humans need to blend with silicon chips to create a singularity, a future in which man and machine need to become one for survival. Will AI have superhuman calculation capabilities and gain the human capabilities of free will, survival instincts, mobility, dexterity, and emotional passions? There is the fear of discovering that humans are not as unique as otherwise thought. The wrestling with the humanity of robots is evident in many science-fiction plots. The fear rests in the idea of autonomy. As AI technologies such as deep learning, affective computing, written and oral natural language interfaces, and swarm with distributed intelligence improve, will it also gain the ability to improve itself beyond human capabilities and break free from human control over its learning and actions? Machine consciousness is considered possible only if we reduce all human experience to computationalism, i.e., the theory that the human mind is merely an information processing system. Koch and Tononi (2017) described an integrated information theory in which consciousness requires a specially configured system with an architecture that gains synergy from its components, rather than merely cumulating the power of its components. Johnson and Verdicchio (2017) attributed such fear to sociotechnical blindness. It is "the failure to recognize that AI is a system… [that] only operates in combination with people and social institutions." While the word "autonomy" is used for AI systems, it does not have the same connotation as for humans. Autonomy may refer to nothing more than the ability to generate pseudo-random numbers for a chess program to play different strategies. They (Johnson & Verdicchio, 2017) argued that such autonomy is merely computational and warned against shifting from a metaphorical description of intelligence to sameness as done by Muller (2015). Muller describes a scenario in which robots develop a drive for self-preservation that leads to resource acquisition behavior. However, the structure of AI systems is such that human designers have programmed goals into such systems and they are not free to make alternative decisions.

Keating and Nourbakhsh, recognizing merging anxiety in society, asked, regarding IBM's Watson, "Does such a machine learn? and concluded there are a number of issues to consider about reasoning, agency, and society to name a few. (2018, p. 30).

Apart from these irrational fears, are the rational fears that are often hidden, overshadowed by the irrational. The problem with present day AI systems arise primarily from programming that does not anticipate unique conditions, errors in the collection and processing of data, hacking by

malicious individuals (Garfinkel, 2017), and even human operator complacency and overcorrection due to the mind-numbing nature of the task of operating systems on auto pilot (Carr, 2014) rather than a robot revolution. Chopra, in 2010, began making the case for considering legal personhood status to intelligent agents to allow for legally binding transactions. Johnson and Verdicchio (2017) concluded that the fear should rest with the decision-making of humans who decide when an AI is ready for deployment, the boundaries placed on these systems, or the legal responsibility or even rights of AI systems. Humans do not always share common goals, interests, or even moral codes. It is in the context of these rational fears and the true meaning of AI autonomy that we can better understand the range of variations possible and where the real risks, real solution, and rational concerns lie by better understanding the points of autonomy within the AI.

CHALLENGES TO POLICY MAKERS AND DEVELOPERS

Emerging technologies that experience rapid growth like AI create a challenge for policy makers and developers in that their development and use outpace the ability for them to respond to unexpected issues that arise (Munoko, Brown-Liburd, & Vasarhelvi, 2020). Regulation and oversight cannot keep pace. At a GAO Forum in 2018, participants identified several key policy considerations: incentivizing data sharing, improving safety and security, updating the regulatory approach, assessing acceptable risks and ethical decision making, establishing regulatory sandboxes, developing high-quality labeled data, understanding AI's effect on employment and reimagining training and education, exploring computational ethics and explainable AI (Persons, 2018). When these policy decisions are made, it has a significant impact on the rate of development, speed of diffusion, and the way in which AI develops. AI impacts policy and policy impacts development. The primary areas of concern related to policy and development include privacy, trade, and liability (Agrawal, Grans, & Goldfarb, 2019). As it relates to the fears that individuals hold, liability is central to the discussion. There are many issues surrounding liability policy. Three assumptions often made are that the systems are always right, will behave within constraints, and divergences will be detected. However, when these assumptions fail there are ethical, legal, and economic implications (Munoko et al., 2020). In the United States, the Algorithmic Accountability Act of 2019 requires companies to assess their AI systems for bias and discrimination and reasonably address those issues. AI systems are designed by humans and may simply encode human biases which would clearly be problematic in hiring, for example, or reinforce gender bias. However, consumers also find issue with being classified by a narrow set of information. Companies have launched efforts to offer choices that are outside of the algorithms (Puntoni, Reczek, Geisler, & Botti, 2021). There are also policy implications related to jobs and education, as certain jobs will be lost by the use of AI and others will be created. There will also likely be a period of mismatch between skills and jobs (Agrawal et al., 2019).

STRUCTURE OF AI SYSTEMS

Russell and Norvig (2020) used the four categories of approaches to artificial intelligence, thinking humanly, thinking rationally, acting humanly, and acting rationally. This classification scheme is useful in both understanding the historical development of AI systems and the differing schools of thought that are advancing the field. However, the school within many cases idealized goals, do not shed light on what types of AI are more or less a real threat.

Russell and Norvig (2020) describe the structure of AI systems as agents stating, "an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" (p. 34). The way an agent acts upon its environment is based upon what is perceived through its sensors and its internal algorithms. Russel and Norvig (2020) noted that many engineering systems could be described as agent systems, but AI operates "where the artifacts have significant computational resources and the task environment requires nontrivial decision making" (p. 36). They explained that "given an agent design, learning mechanisms can be constructed to improve every part of the agent" (p. 55). They described an autonomous agent as one that can learn from its perceptions of its environment beyond what was originally given to it by its designer. Such an agent placed in an environment can begin to affect the environment, and the impact is enhanced when there are multiple agents placed in the environment. Since agents are digital objects, they can be easily replicated, leading to a proliferation of profitable agents. As a result, an agent-rich environment is fundamentally altered, and the agents operate in an

environment that did not even exist when they were designed. Since AI agents are designed for autonomous operation, without human intervention and limited oversight, the threats they pose in an altered environment are a reasonable fear.

It is specifically this conception of learning that leads to autonomy that leads to AI systems being efficient and interesting, two characteristics that will drive their greater use in society and should demand our analysis. Note that efficiency lies at the heart of a designer being able to program the agent to learn rather than programming everything the agent needs to learn. This is one aspect of their growing use in business applications. However, it is precisely the fact that the behavior of the AI agent is not fully documented a priori, that the systems are both interesting and could be dangerous. However, AI agents are not fully autonomous. They are autonomous within certain ranges of autonomous behavior. What they can do autonomously is precisely what should be analysis. The following section defines types of AI autonomy that can be used to identify risks areas of AI agents that have such types of autonomy and what potential mitigating actions might be appropriate when those agents are deployed.

TYPES OF AI OF AUTONOMY

Types of AI autonomy distinguish between what aspects or dimensions of the AI system is programmed or hard wired by the human vs. what aspects or dimension the computer system can choose on its own. Table 1. summarizes the types of AI autonomy.

Data Autonomy

In most data mining applications humans determine the data to be used, and often select data filtering, cleaning, and formatting options before providing it to the ML tools. Consider the case of a system analyzing retail store data from POS systems. In this case the ML does not have data source autonomy. While the ML system continually gets new data from the sensors which collect data, and this new data is used in analysis, and may result in new decisions being made, the ML system does not autonomously collect data from new sources.

Data source autonomy: Does the AI gets its own data from multiple sources or does a human intervene in this process? In an AI with data source autonomy, the AI has the ability to collect and process data from new sources it encounters. In the case of the ML tool, in order to possess

Table 1. Types of AI Autonomy

Type of AI Autonomy	Description
Data Autonomy	
Data Update Autonomy	The degree to which the AI can determine based on its logic, when to update new data and include it in calculations
Data Source Autonomy	The degree to which the AI has the ability to collect and process data from new sources it encounters.
Data regulation (transformation) autonomy	The degree to which the system has the autonomy to transform new data that does not meet the standards clearly defined by the human designer
Model Autonomy	
Model parameter autonomy	The degree to which the AI can automatically update the parameters as required based on output errors
Model type autonomy	The degree to which the AI has the autonomy change the decision model used to make a choice
Decision Autonomy	The AI has the autonomy to implement its own decisions or human intervention is required before a decision can be enacted.
Objective Autonomy	The degree to which the AI can set new objectives that meet the defined goals coded in the system is a threat with complex, autonomous AI. When broad goals are built into the AI rather than specific objectives, the AI could select objectives that meet the goal, but conflict implicit ethical guidelines built into human decisions.

data source autonomy, it must be able to scan and find new sources of data. For example, if the ML system analyzing POS data finds insufficient customer data from its internal sources for decision making, searches a list of other data sources, and contacts credit bureaus to obtain customer data which it adds to its dataset to make better inventory decisions, then the system has data source autonomy.

Data update autonomy: The system described previously has data update autonomy if it does not merely get new data as it is collected, but can determine based on its logic, when to update new data and include it in calculations. However, the sources are not changed, and no new data sources enter into the decision-making process. The risks are small and there is no rational fear of such models.

Data regularization (transformation) autonomy: Often, systems using data analytics to drive decisions may regularize the data for processing. Does the system have the autonomy to transform new data that does not meet the standards clearly defined by the human designer? What happens when a video camera which proved a data stream is repositioned? Does the AI recognize this change and automatically transform the data for vision analysis? Does this require human intervention? This is of particular interest in the Internet of Things (IoT), where a swarm of devices monitor the environment for security and system optimization. The challenge is the meta data provided by the individual components in the IoT. For example, if the devices merely send their sensor data to the AI, repositioning the devices will lead to errors in conclusions derived from analysis of old data. However, if the physical position of the device is also transmitted with the sensor data, physical repositioning will merely require transforming the data to generate parameters consistent with old data. Replacing the camera with a device having a different sensor, or even changing the color temperature of room lighting will create data transformation problems for the AI.

Model Autonomy

Model parameter autonomy: Many ML toolsets allow the analyst to set model parameters such as the number of layers and nodes in a neural network or kernel type, penalty parameter and the gamma of the kernel. In a system with parameter autonomy, the parameters may have initial default values set by a human, but the system can automatically update the parameters as required based on output errors.

Model type autonomy: Does the AI have the ability to select the appropriate analytic model based on its results and vary the selection over time. For example, if the AI is designed to start with an ANN model, uses this to make decisions, and over time compares the performance of these decisions in the real world with that of using Support Vectors, Random Decision Forests, or Logistic Regression and can change models as they prove to be better, the AI will function with different "personalities" based on the outcomes of its decisions.

Decision Autonomy

Does the AI have the ability to implement its own decisions or is there a human intervention step before a decision can be enacted? As AI becomes more intelligent and utilized in society, opportunities for AI to make decisions will increase. A simple example is the evolution of self-driving cars where early generations give warnings to the human driver to take action while more sophisticated implementations allow the computer itself to take the evasive action. If such decision autonomy were expanded, there could come a time where the AI chooses where the vehicle should go.

A claims processing AI may reject a claim and send the rejection notice to the customer. This is similar to the action of a human DM determining claims eligibility. This is not a risk of AI. However, when the customer complains or when there is negative news media coverage of the action, effective human decision makers respond to the outcry. Does the AI with result autonomy have the ability to understand when its actions go awry, and take appropriate remedial action? This requires the AI to (a) monitor environmental data on the consequences of its actions and (b) request outside assistance in crisis situations. The problem faced by AI with results autonomy is often the result of failure to understand the degree of intervention by human employees when taking actions with serious consequences and automating the process with an AI that lacks these features.

An article in *Forbes* magazine (2018) highlights a few of these important decisions. Consider autonomous cars. Autonomous vehicles will have to make decisions in a split second. Should it swerve to avoid hitting a pedestrian? What if swerving threatens the passengers who are in the car? What about military drones? Currently, humans make the final decision as to whether the drone should fire upon a target. How much decision-making autonomy can or should be left in the hands of the drone?

Objective Autonomy

The goal of the AI is set during design typically by the user. For example, in a mapping AI used to guide a self-driving vehicle, the user (passive driver) sets the destination (objective), and tactical choices to reach the destination such as using tollways, local roads, or a scenic route. In some cases, the strategy may be pre-selected by the system, and it may be altered by the user. The AI designer may have additional goals embedded within the system to either focus on fuel efficiency, safety, or learning from user interaction. Does the AI retain goals coded into it by the designer even while being used by different users, and does it get new goals from external agencies or does it even learn new goals?

RATIONAL FEARS

Each of these types of autonomy brings with it fears. It is these rational fears that must become the focus of AI decision makers ranging from system designers and policymakers.

Data Autonomy

If an AI has the ability to seek out new sources, there is the possibility that the AI may collect data from unreliable sources and corrupt its decision-making process. Consider an AI seeking out new sources: will the AI have the ability to determine if the source is satire, "fake news," or includes extreme bias. The problems with source data quality arise from anecdotal data which leads to biased conclusions, incorrect data generated by a badly designed system, and deliberately falsified data created to set the AI awry.

AI analysts are concerned with the quality of information used as input to the AI systems they design because poor quality information input to an AI system can rendered its output useless. This need for quality information in systems has been recognized by both researchers and analysts. For example, Eppler (2006) provides a summary of a number of information quality models and distills them into a list of 16 information quality criteria. However, those quality models are designed to guide humans in selecting and creating high-quality information and are not used when an AI autonomously selects data input. Madnick, Wang, Lee, & Zhu, (2009) also conducted a review of data quality, which they described as including information quality; however, they also described structured organizational contexts with significant human intervention. Knight and Burn (2005) described how web-based data exacerbates the problem because information can be uploaded with no assessment of quality.

In addition, there may be violations of data ownership and privacy regulations when using web sources. The timeliness of the data is also important. Suppose the data is no longer valid after a certain amount of time has passed. There is rational fear that updates may not happen frequently enough or perhaps too frequently, therefore providing too much information to the AI about individual behavior. There may be expenses associated with collecting and using this data.

These are rational fears from data autonomy and need to be addressed when setting AI to collect and use data from new sources. If an AI has the ability to update data, it is important to ensure that the data is updated accurately. The frequency of the updates and the communication of that information is relevant. Furthermore, as data is updated, the AI must determine if the format of the original source has changed. The structure of the source database may change over time resulting in import problems. However, if the AI has the ability to regularize the data on its own, it may do so in an erroneous format and lead to mistakes in operation.

Model Autonomy

Model parameters - While there is some likelihood of overfitting to small perturbations in the data, there are no major consequences that should create a rational fear of parameter autonomy.

Consider a regression model that uses data from multiple sources to generate a causal forecast for a production system. When the data changes, the model's parameters change, leading to a new forecast model. As long as there are limits on behavior imposed using constraints defined by the system, for example, a maximum and minimum for production, the AI will not create a catastrophe.

Model type - Decisions taken by some of these models can be explained and rationalized while others are black boxes with no explanations possible. A shift in the model can lead to the AI making decisions that cannot be explained or rationalized by human owners, leading to legal and regulatory problems. This is a rational fear from model type autonomy.

A recruitment AI has a model base allowing it to use a decision tree based on experience with recruits or an artificial neural network (ANN) to select candidates. While the decision tree can be explained, the ANN is a black box and its decisions cannot be explained when faced with a lawsuit. AI that can select the best model can lead to unfair and unsupportable decisions.

Decision Autonomy

AI used to schedule airline flights is designed to optimize costs while meeting traffic and regulatory requirements. The scheduling AI will assign shifts to employees and aircraft to routes. If the AI has the ability to implement these decisions, that is, to roll out new employee and aircraft schedules autonomously, there is a likelihood that the optimization routine would recommend cancellations of flights based on profitability. It may end up canceling all flights out of an unprofitable airport. While the decision is justified by cost analysis, it may be politically expensive to the airline and lead to either new restrictive legislation or customer dissatisfaction. The AI's model often will not incorporate these subjective factors in its decision making. There is a rational fear the autonomous decision implementation by AI will lead to extreme outcomes that trigger a hostile marketplace reaction.

Flash crashes in the financial market are not new, and the earliest documented occurrence is the 1987 market crash in the United States, blamed in part on "program trading," the use of computer algorithms to enact arbitrage deals based on minute variations of stock prices and underlying indexes (Furbush, 1989). Regulatory changes following this crash forced limits on automated trading when the major indexes vary more than a set percentage, to prevent similar AI enabled market volatility, which can impact investor sentiment, and lead to longer lasting impact on the market. When AI can enact its decisions autonomously, and more critically, when multiple AI systems can interact with one another, the herd behavior of bots can have serious consequences (Ferrara, Verol, Davis, Menczer, & Flammini, 2016). The rational fear is that multiple AI systems can interact with one another, leading to herd behavior on a very rapid scale, and consequent crises.

Objective Autonomy Fears

AI with broad goals, allowing for objectives to be set by the AI is designed to find new threats and opportunities as they occur and protect from attacks or exploit them as needed. The fear is that these goal-based systems would find results that meet the goal but have undesirable consequences.

A tax planning AI with a broad objective of minimizing tax payments may select risky tax shelters or even decide to stop filing tax returns.

An AI to manage an electric grid with a goal of making the grid more energy efficient could analyze energy usage patterns and turn systems on/off to optimize energy usage. However, over time it could find that the best way to optimize peak energy usage on a hot day is to create a rolling blackout of the grid.

A router with intelligence and a broad objective to route data packets efficiently may decide to slow down streaming media traffic to reduce system overloads.

One other reasonable fear is that the AI may have a published objective of optimizing network efficiency, but a hidden agenda to monitor traffic and manipulate information flows, which may not be known to users. This type of AI hijacking, where the goals of the user and designer differ is a real fear from the proliferation of AI.

CONCLUSION

AI systems are developing quickly and are utilized in a variety of settings. They have for decades captured the mind of the general public in science fiction literature and media. It is through this lens of grappling with the humanness and autonomy of AI systems that the general public understands AI. Addressing AI through that lens of public understanding and concerns provides a context for policymakers to who are drafting policy in response to AI development and AI developers who are seeking product acceptance. The public must see AI as a support to humanity and not as a competitor. Policy makers and developers must be sure to address the fears that society holds whether rational or irrational. It is therefore, our premise that the language for the categorization of AI be put in the context of this same filter since it is those fears that will drive policy makers to formulate a response. Policymakers and developers cannot ignore the fears that will prevent the acceptance of a technology, whether they are rational or irrational. The typology presented in this article captures the rational fears actually present at varying types of autonomy and bounds those fears in the context of the ever presence of the human that exists in the system.

Dr. Kenneth R. Walsh *is an Associate Professor in the Information Systems Group of the Management and Marketing Department, College of Business, at the University of New Orleans.*

Dr. Sathiadev Mahesh *is a Professor of Management at the University of New Orleans.*

Dr. Cherie Courseault Trumbachis *an Associate Professor of Management/ Information Systems at the University of New Orleans.*

REFERENCES

Agrawal, A., Gans, J., & Goldfarb, A. (2019) Economic policy for Artificial Intelligence, *NBER Innovation Policy & the Economy*, 19(1). Pp139-159

Brooks, R. (2017). The seven deadly sins of Al predictions: Mistaken extrapolations, limited imagination, and other common mistakes that distract us from thinking more productively about the future, *MIT Technology Review*, (120:6). pp. 79-86.

Carr, N. (2014). *The glass cage: Automation and us*. New York, NY: W.W. Norton & Co.

Chopra, S. (2010). Rights for autonomous artificial agents? *Communications of the ACM*, (53:8), pp. 38-40

Conde-Clemente, P, Trivino, G., & Alonso J. (2017). Generating automatic linguistic descriptions with big data, *Information Sciences*, (380), pp. 12-30

Dechter, R. (1986). Learning while searching in constraint-satisfaction-problems, in *Proceedings of the 5th National Conference on Artificial Intelligence*. Philadelphia, PA, August 11-15, 1986. Volume 1: Science.

Ferrara, E., Varol O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots, *Communications of the ACM*, (59:7), pp. 96-104

Furbush, D. (1989). Program trading and price movement: Evidence from the October 1987 market crash. *Financial Management*, (18:3), pp. 68-83. Retrieved from http://www.jstor.org/stable/3665650

Garfinkel, S. (2017). Hackers are the real obstacle for self-driving vehicles", *MIT Technology Review*, August 22, 2017. (accessed 4/27/2018 https://www.technologyreview.com/s/608618/hackers-are-the-real-obstacle-for-self-driving-vehicles/)

Eiben A. E., & Smith J. E. (2015). What is an evolutionary algorithm? In *Introduction to Evolutionary Computing*. *Natural Computing Series*. Springer, Berlin: Heidelberg

Eppler, M. J. (2006). A framework for information quality management. In *Managing Information Quality*. Springer, Berlin: Heidelberg. https://doi.org/10.1007/3-540-32225-6_3

Johnson, D., & Verdicchio, M. (2017). AI anxiety, *Journal of the Association for Information Science and Technology* (68:9), pp. 2267-2270.

Keating, J., & Nourbakhsh, I. (2018). Teaching artificial intelligence and humanity. *Communications of the ACM*, 61(2), 29–32. https://doi.org/10.1145/3104986

Knight, S., & Burn, J. (2005). Developing a framework for assessing information quality on the World Wide Web, *Informing Science Journal* (8), pp. 159-172.

Koch,C., & Tononi, G. (2017). Can we quantify machine consciousness? *IEEE Spectrum*.

Kurzweil, R. (1990). *The Age of Intelligent Machines*. MIT Press, Cambridge, MA: USA

Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and framework for data

and information quality research, *ACM Journal of Data and Information Quality* (1:1), pp. 2:1-1:2:22.

Marr, B. (2018). The life and death decision AI robots will have to make, *Forbes*, June 29 https://www.forbes.com/sites/bernardmarr/2018/06/29/the-life-and-death-decision-ai-robots-will-have-to-make/#34379ea0480a

Manjavacas, E., Karsdorp, F. B., Burtenshaw, B., & Kestemont, M., (2017). Synthetic literature: Writing science fiction in a co-creative process. *Proceedings of the Workshop on Computational Creativity in Natural Language Generation* (CC-NLG 2017) Santiago de Compostela, Spain: Association for Computational Linguistics (ACL), pp. 29

McCarthy, J. (1979). Ascribing mental qualities to machines, In Martin Ringle (Ed.), *Philosophical Perspectives in Artificial Intelligence*. Humanities Press.

McCorduck, P. (1977). The early history: History of Artificial Intelligence discussion panel. *International Joint Conference on Artificial Intelligence* (IJCAI).

(accessed 4/10/2018 https://www.ijcai.org/ Proceedings/77-2/Papers/083.pdf).

McDaniel, P., Papernot N., & Celik, Z. (2016). Machine learning in adversarial settings. *IEEE Security & Privacy*, (14)3. pp. 68-72, May-June doi: 10.1109/MSP.2016.51

Muller, V. (Ed.) (2015). *Risks of Artificial Intelligence*, Boca Raton, FL: CRC Press.

Munoko, I., Brown-Liburd, H. L & Vasarhelvi, M. (2020) The ethical implications for using artificial intelligence in auditing, *Journal of Business Ethics*, 167(2). pp. 209-234

Ongsulee, P. (2017). Artificial intelligence, machine learning, and deep learning. *IEEE Fifteenth International Conference on ICT and Knowledge Engineering*, Nov 22-24, Bangkok, Thailand 10.1109/ICTKE.2017.8259629

O'Sullivan, A. (2017). Don't let regulators ruin AI, *MIT Technology Review*, 120(6). Pp. 73.

Persons, T. M. (2018) Artificial intelligence: Emerging opportunities, challenges, and implications for policy and research, GAO Reports 6/26/2018, pp. 1-11.

Puntoni, S., Reczek, R.W., Giesler, M., & Botti, S. (2021) Consumers and artificial intelligence: An experiential perspective, *Journal of Marketing*, 85(1). pp. 131-151.

Russell, S. J., & Norvig, P.(2020). *Artificial intelligence: A modern approach*. (4th ed.) Upper Saddle River, NJ: Prentice Hall

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers, *IBM Journal of Research and Development* (3:3).

Sullins, J. P. (2015). Artificial intelligence, *Ethics, science, technology, and engineering: A global resource*, pp. 119-124.

Turing, A. M., (1948). Intelligent machinery, *National Physical Laboratory Report*. (accessed 4/15/2018, http://www.npl.co.uk/about/history/notable-individuals/turing/intelligent-machinery).

Young, K. L., & Carpenter, C. (2018). Does science fiction affect political fact? yes and no: A survey experiment on 'killer robots'," *International Studies Quarterly*, 62, pp. 562-576.