PROGRAM MATERIALS
Program #29127
September 19, 2019

US Artificial Intelligence Regulation Has A Long Way to Go

Celesq® AttorneysEd Center
www.celesq.com

5301 North Federal Highway, Suite 180, Boca Raton, FL 33487

Scoolidge Peters Russotti & Fox LLP

2 Park Avenue, 19th Floor | New York, NY 10016

US Artificial Intelligence Regulation Has A Long Way To Go – PROGRAM MATERIALS

By: Peter Scoolidge, Esq. & Jasmine Weg, Esq.

Part I – Introduction

Artificial Intelligence, often referred to as "AI", is the use of software to perform tasks, which normally require human intelligence such as visual perception, speech recognition and decision-making. At present, we interact with and use AI in our everyday lives quite often. Prime examples are the popular virtual digital assistants: Apple's Siri, Amazon's Alexa, and IBM's Watson.

Humans presently interact with AI to varying degrees throughout their daily lives and given the proliferation of AI with ever-increasing decision-making capabilities, the question naturally arises as to what policies and regulations are necessary in an AI-integrated society?

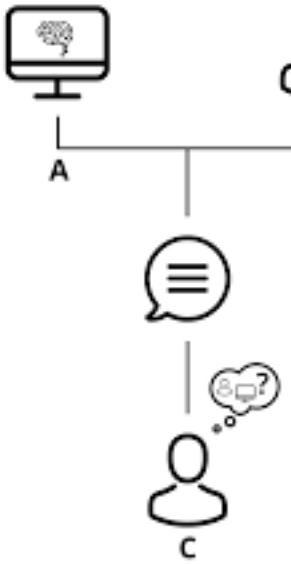
In this program, we will provide insight into what AI is and what it can do while exploring some of the potential nightmare scenarios people have envisioned for AI that has run amok. We will talk about the present efforts by various international state actors at initiating a set

of legal or ethical guidelines governing AI development and use, and talk about the basis in other existing legal and ethical principles underlying the proposed regulations.

Part II – What is AI?

To understand the legal implications created by the widespread use of AI, we must first take a step back and discuss what AI is.

This conversation must logically begin with the **Turing Test**. Coined by computer scientist Alan Turing in 1950; the test was designed to determine whether or not a computer is "intelligent" or can "think". The test, essentially, was an "imitation game" whereby a human must chat with some mysterious speakers most of which are human and one of which is a chatbot. The goal is for the chatbot to trick the judge into believing he is communicating with a human. Turing's prediction at the time was that by the year 2000, a computer "would be able to play the imitation game so well that an average interrogator will not have more than a 70-percent chance of making the right identification (machine or human) after five minutes of questioning." Turing's prediction has yet to come true but AI is getting closer and closer.



AI is constantly evolving so it can be difficult to succinctly define. Indeed, John McCarthy, who coined the term “Artificial Intelligence” in 1956, explained “as soon as it works, no one calls it AI anymore.” In other words, what is classified as AI has changed over the years as computer scientist develop more and more robust technology with greater capabilities.

To better understand AI and its potential usage, consider the two categories it is often broken into: **Narrow AI** and **General AI**. Narrow AI, also commonly referred to as “weak” AI is essentially present day AI – AI developed to perform a single task, which can be anything from analyzing data to checking the weather. Narrow AI performs a task based on a specific data set so it cannot perform a task outside of what it has been designed to perform. A great example of narrow AI is Siri. While the popular phone assistant interacts with us, it is only within the confines of its design. In other words, Siri Can process language and perform google searches or

other tasks it is programmed to perform, but pose an open-ended abstract question and you'll receive a vague, unsatisfying response. General AI, or Strong AI, on the other hand is when a machine exhibits human intelligence and is capable of performing any intellectual task that a human being can. Most notable about General AI is the sophisticated ability for experiential understanding. This is the type of AI that is often envisioned during a futuristic time when robots roam and often depicted in movies and books. Qualities of General AI would be the ability to reason, make judgments under uncertainty and integrate prior experiences into decision-making – as well as be creative. While we are not in the general AI era at the moment, many scientists believe general AI will be achieved over the next few decades. Machine learning, or the ability of AI to find patterns and make decisions without instruction is part of the path to general AI. A subset of Machine learning – called deep learning, is when a machine learns from unsupervised data through neural networks which have brain-like functions.

With that very basic understanding of what AI can do presently and where AI might be in the next twenty, thirty years – you can see how critical it is for governments as well as major tech companies to develop both ethical guidelines and laws and regulations for the development and use of AI. Because although Artificial Intelligence is expanding at a rapid speed, the development of limitations on AI by way of laws and regulations is unfortunately straggling behind. The US has yet to codify guidelines or legislation to regulate the usage of AI, however,

globally, various state players through the implementation of ethical guidelines are addressing this issue.

Part III – EU's "Ethics Guidelines for Trustworthy AI"

On April 8, 2019, the European Union's High-Level Expert Group (HLEG) on Artificial Intelligence published a comprehensive document entitled Ethics Guidelines for Trustworthy AI, laying a foundation of ethical considerations to uphold while employing the use of AI. This framework is a useful tool for the development of meaningful policy and regulation and at the very least is a starting point for opening a dialogue on the matter.

An underlying principle of the guidelines is that "AI systems need to be **human-centric**, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom."

Trustworthy AI – Society's ability to trust AI systems and the humans developing, deploying and using them is critical. Trustworthy AI has three principal components. It should be lawful – that is, of course, that the development and use is compliant with all applicable laws and regulations; it should be ethical, i.e. adhering to ethical principles and values; and finally it should be robust, both technically and socially. The three components are interwoven. Essentially AI must be developed and used within the confines of the all laws regulating it, depending on the specific industry the AI falls into. But ethics fill in where the law lags so AI must always be designed and

used in alignment with ethical norms. Robustness ties the other two components together. Trustworthy AI must be robust in that the systems perform as designed – safely, reliably, securely and are integrated with safeguards to prevent foreseeable issues and adverse impacts.

The guidelines highlighted the following ethical considerations: respect for human dignity; freedom of the individual; respect for democracy, justice and the rule of law; equality, non-discrimination and solidarity and finally, citizens' rights.

Furthermore, the guidelines categorized four ethical imperatives that AI practitioners must always strive to adhere:

(i) Respect for human autonomy; (ii) Prevention of harm; (iii) Fairness and (iv) Explicability.

Respect for human autonomy: AI systems should never unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. On the flipside, AI systems must augment, complement and empower human skills. Thus, AI systems should be designed as human-centric, always mindful to preserve human choice.

Prevention of harm: the systems should be safe and secure and there must be safeguards to prevent AI systems from malicious use. In this imperative, it is also important to consider how AI systems can exacerbate asymmetries of power and ensure that AI is not used to further

disenfranchise the less powerful, such as in the employer-employee relationship or government-citizen relationship.

Fairness: There must be efforts to avoid AI system design that implements unfair biases. The guidelines suggest that AI that is designed free of bias, discrimination and stigmatization could be used to increase societal fairness.

Explicability: Trustworthy AI necessarily depends on transparency. Without transparency as to the capabilities and purpose of the AI systems, a decision as to the application of a system cannot be duly contested.

Human agency: users should be able to make informed autonomous decisions about AI. This goes to the explicability imperative. Humans should be educated on AI as it impacts their lives. This will allow humans to challenge the system and ensure the ethical principles we have been discussing are being adhered to. Where a decision will have a significant effect on an individual, that decision should not be based solely on automated processing.

Human oversight: human oversight is the key to ensure that AI is not adversely affecting society or undermining human autonomy. Human oversight of AI can be broken into three categories:

(i) Human in the loop (HITL) which would be the capability for human intervention in every decision cycle of the system – the guidelines point out that in many cases this would be neither possible nor desirable;

(ii) Human on the loop (HOTL), which is more practicable, allows for human intervention during the design phase of the system and then human monitoring throughout the systems operation; and

(iii) Human in command (HIC), which allows for overall human oversight of an AI system, integrating human discretion during the use of the system and ensuring the ability to override decisions.

Technical robustness and safety: AI systems must be safeguarded against vulnerabilities that could adversely affect humans such as hacking. To that end AI should be developed with a fallback plan in case of problems. One example is when an AI system asks for a human operator before continuing an action. AI systems must be accurate. The guidelines state that an "explicit and well-formed development and evaluation process can support, mitigate and correct unintended risks from inaccurate predictions. When occasional inaccurate predictions cannot be avoided, it is important that the system can indicate how likely these errors are. A high level of accuracy is especially crucial in situations where the AI system directly affects human lives." Furthermore, AI systems must be both reliable and reproducible. In other words, the AI performs the same functions when repeated under the same conditions.

Privacy and data governance: it is imperative that AI systems guarantee privacy and protect data throughout a system's lifecycle. Consider this example: in 2006 Netflix sought to improve their

movie recommendations algorithm and crowd-sourced the problem, offering a $1 million prize. The company release 100 million anonymized movie ratings, assigning subscriber IDs and releasing the movie title. Rating and date on which the subscriber rated the movie. The task was for contestants to develop an algorithm 10% better than Netflix at predicting how a subscriber would rate other movies. 16 days later, University of Texas researchers announced that they were able to take the anonymized data and identify the names of the users. They cross-referenced the dates the Netflix users rated the films with other databases like the movie-rating system IMDB. Netflix tried to launch another such contest in 2009, releasing the gender, zipcode and age of the anonymized rater but this time they were hit with a lawsuit. While the user ID was supposedly anonymized, the ability to reverse-engineer the data meant that the user's identity and entire viewing history could be ascertained. The plaintiff in that lawsuit was a lesbian mother who was not open about her sexual orientation and feared that her personal information would be exposed against her will. The protection of privacy and individual data must be front and center in the development of trustworthy AI. Furthermore, precautions must be taken to ensure the quality and integrity of the data, including protocols governing the access of such data.

Transparency: the data sets and processes that form an AI system's decisions should be documented for transparency. Mistakes should be identified as well to prevent future mistakes. The decisions made by AI should be traceable and explainable. If an AI system's decision is

capable of having a significant impact on a person's life then the guidelines urge that there should be an option to demand an explanation into the AI's decision-making process.

Transparency is also important when AI systems communicate with humans. AI should be identifiable as such when it is interacting with human beings. So while we discussed the Turing Test earlier as the marker of AI intelligence, from an ethical prospective we don't want an AI system interacting with humans without their knowledge. For instance, AI is being developed for use in conducting certain rote phone calls such as for customer service but these ethical considerations would require the AI to identify itself and inform individuals when they are interacting with an AI system.

Diversity, non-discrimination and fairness: to avoid discrimination its important to consider the data sets used by the AI systems in terms of biases. Data sets could easily include inadvertent biases that if used could result in unintended prejudice and discrimination.

Here is a hypothetical the UK's Information Commissioner's Office posed: *A bank has developed a ML system to calculate the credit risk of potential customers. The bank will use the ML system to approve or reject loan applications. To train the system the bank has collected a large set of data containing a range of information about previous borrowers, such their occupation, income, age, and whether or not they repaid their loan. During testing, the bank*

wants to check against any possible gender bias, and finds the ML system is giving women lower credit scores, which would lead to fewer loans being approved.

This might be caused by one of two problems: 1 – imbalanced trade data – for example, if more men have taken loans from the bank in the past than women, the bank would lack data on women which would result on the ML system training on the male data predominantly. It would thus have stronger data on male repayment rates. "Put another way, because they are statistically less important, the model could systematically predict lower loan repayment rates for women, even if females in the training dataset are on average more likely to repay their loans than men." 2 – another reason for this may be because the training data reflects past discrimination - if in the past women's loan applications were rejected more frequently than men's on the basis of gender, then any model based on such training data is likely to reproduce the same pattern of discrimination.

If programmers consider these potential issues that may be latent in a given data sets they can proactively try to remedy the biases before training AI systems so as not to perpetuate further discrimination.

Accessibility and Universal Design: it is important that AI systems which individuals will interact with in their daily lives such as those used in a business to consumer domain be universally accessible. Thus the technology should be accessible across age groups, gender and

accessible to those with disabilities. This is a broad directive, but it's an important consideration for AI developers and policymakers alike.

Societal and Environmental impact: along the lines of the general goal of prevention of harm, AI should be designed in a sustainable manner and as environmentally-friendly as possible.

Implementation: Implementing a trustworthy AI regime requires a system of monitoring, reporting, testing and auditing.

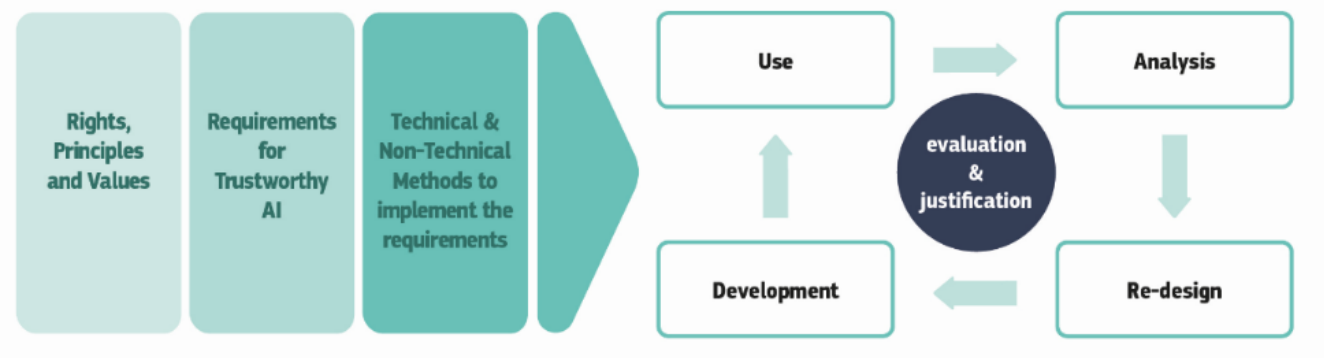
Presenter to read NY Code

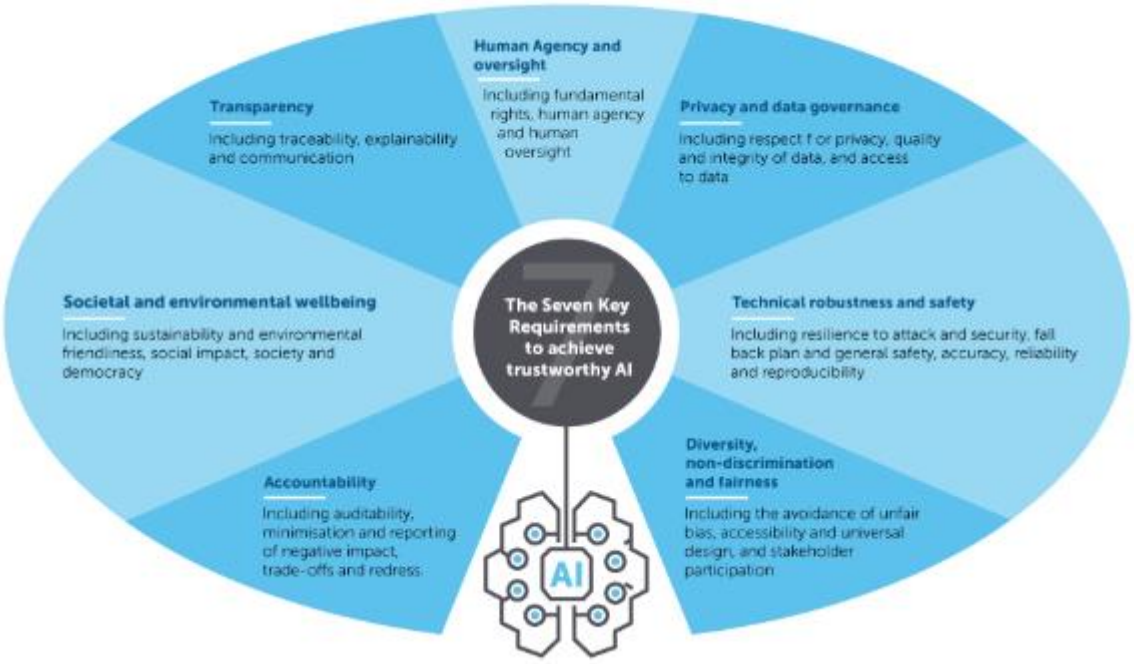
This code is required for all attorneys wishing to receive CLE credit in the state of NY and taking the program 'on-demand' at Celesq AttorneysEd Center either online or via CD

Please notate it carefully

The presenter will only be able to read the code twice and will not be able to repeat it or email it to you.

Thank you!





Rule of law by design: The guidelines encourage AI systems developers to identify both a set of norms that the AI system should adhere to as well the potential negative effects that might occur with deployment of the system *prior to programming*. They then must develop systems with these norms and potential adverse effects in mind. Testing must occur throughout the lifecycle of the system. A useful tool for testing AI systems is the use of "red teams" "deliberately

attempting to “break” the system to find vulnerabilities”, as well as the use of “bug bounties, “which” incentivize outsiders to detect and responsibly report system errors and weaknesses”.

There are other non-technical ways of ensuring trustworthy AI:

- Regulation – obviously through governmental bodies
- Codes of conduct – companies creating the technology or other stakeholders can develop internal standards embracing the concepts we have illustrated
- Standardization – industry standards adopted for specific types of AI systems, possibly through the use of accreditation systems
- Certification – a certification program for AI techniques or training is a useful means of achieving transparency and accountability
- Education – this is an important tool. Widespread education on AI will inherently lead to more trustworthy AI because a large part of the fear and mistrust that surrounds the topic is fueled by the mystery surrounding it
- Diversity – we discussed inclusivity as a goal for trustworthy AI – but part of that begins with diversity in the development space – AI systems developed by a diverse group of people will result in systems that are inclusive of the values we have discussed

Part IV: Global Efforts

China: China has developed a government agency called the Chinese Association for Artificial Intelligence and in January 2019 they established their own AI ethics committee. Two months later in March, various private sector actors such as Baidu and Tencent, called for a government-established set of ethics rules for AI development. China has long-endeavored to be a leader in the field of AI and we predict that they will publish a comprehensive guidelines for AI development in the coming future.

Singapore: Singapore has also attempted to address AI concerns – their own Monetary Authority released a document illustrating aspirational principles for AI applications. They call for AI that is fair, ethical, accountable and transparent. Much like the EU's comprehensive guidelines, the Monetary Authority stresses the need to avoid biases, set up systems for testing and review, and disclose the use of AI to all who engage with it.

Canada and France: Canada and France established an intergovernmental panel to study the effects of AI in December 2018. The main takeaway from that panel was A proposed mandate issued by Canada's executive branch. The mandate listed several areas of interest including:

(1) How data for AI applications is collected and processed, (2) the effect of AI on human rights, (3) whether AI can be "trusted," and (4) military uses of AI.

World Economic Forum: The World Economic Forum has taken up the topic, expressing concerns over the proliferation of AI and its impact on the labor market. With the automation that AI enables, humans are being replaced by machines in many different fields.

This will be definitely be a topic to further explore as labor and employment paradigms shift creating a need for a shift in education and technical training.

Australia: In January 2019 the Australian Human Rights Commission released a discussion paper entitled "Discussion paper to boost conversation about AI ethics in Australia". The Commission espouses the opinion that the key to unlocking the potential of AI is to guarantee that the public has trust in AI-driven solutions. The paper proposes eight principles to guide developers, the industry and the government in regulating AI systems:

1. Generate net benefits – the benefits of employing AI systems must always outweigh the costs
2. Do no harm – as well guidelines on the topic stress, AI systems are meant to improve quality of life by adding the value of efficiency and technology, it is important to regulate AI systems so that they are not designed or used to cause harm or deception and to always be mindful of minimizing negative outcomes.
3. Regulatory and legal compliance – AI systems designed and operated in compliance with all relevant law

4. Privacy protection – private data is protected and breaches are prevented as best possible
5. Fairness – AI should be free from training biases and the systems must not result in unfair discrimination
6. Transparency and explainability – The purpose of AI systems, the manner in which they are to be used and how they will impact people must be fully disclosed. When algorithms are used to make decisions that will affect individuals, they must be informed as to what information was used to form those decisions.
7. Contestability – in the situations where algorithms are used as described above, then a mechanism must be developed by which a person can contest the decision or the use of the algorithm.
8. Accountability – creators and implementers of AI systems should be identifiable and thus accountable for the impacts their technology has. The Paper even urges that there should be accountability for unintended impacts as well.

Part V: The United States

Surprisingly the United States has been very quiet on developing policy and regulation for AI. Department of Defense (DoD) has encouraged a vague directive of usng "appropriate levels of human judgment" in the development and employ of autonomous weapons but this of course gives no meaningful guidance.

Congress has yet to pass legislation or establish a committee to explore the topic.

Google established a committee This past year to guide that company's development of ethical/responsible AI applications but that was quickly dismantled before we could take away any useful guidance. Nonetheless Google has implemented a set of principles internally for the use of AI. The principles Google illustrated as imperative for AI applications and development are:

- AI must be socially beneficial
- Developers must avoid creating or reinforcing unfair bias
- AI must be built and tested for safety
- Accountability to the public
- Privacy incorporated into the design of AI
- High standards of scientific excellence
- AI is created and used in accordance with these principles

Microsoft – In 2018 Microsoft established an AI ethics committee called AI and Ethics Engineering and Research (AETHER) with the goal of crafting internal policy and addressing ethics. Microsoft recently released the White Paper: "Microsoft's Vision for AI in the Enterprise" outlining their approach to the use of AI. The company states that trustworthy AI

requires "solutions that reflect ethical principles that are deeply rooted in important and timeless values." Microsoft outlined these principles as: fairness, reliability and safety, privacy and security, inclusivity, transparency and accountability.

Part VI : Opportunities and Concerns

Lets explore some concrete areas in which trustworthy AI can be designed to enhance society and then discuss some critical concerns that AI systems might pose.

Opportunities:

Sustainability: AI has the potential to aid in tackling the issue of climate change by enabling efficient and effective use of energy and natural resources. AI is being used to optimize energy efficiency as we see in the automobile industry. AI is also useful in analyzing energy data to detect and guide energy needs.

Health: robotics and AI generally are monumental in health and science. AI has changed the way procedures can be done, data analysis allows more tailored and preventive treatment and life-saving tools have emerged through AI.

Education: AI developments have made education more easily accessible and widespread while also providing more innovative useful tools in the classroom.

Concerns:

Identifying/tracking individuals: AI has completely changed the way we identify individuals with face recognition and biometric data (lie detector, voice detector). In many ways these tools are incredibly beneficial to society as it allows for better fraud prevention, the identification of money laundering, terrorist financing, etc. However, there are of course privacy concerns that arise. As governments gain the ability to track the masses, this invites questions as to informed consent and the infringement of freedom.

Covert AI systems: many believe it is inherently unethical for a human to not know they are interacting with AI as opposed to another human being. The development of human-like robots is underway which will exacerbate this concern.

AI enabled citizen scoring: AI is already being used to "score" citizens. At present it is used as a methodology for evaluating students, generating school systems, and other methods. However, this area is fraught with potential misuse. Scoring could theoretically be done to assess individuals moral personalities or ethical integrity by public or private actors resulting in infringing on individual autonomy and furthering discrimination.

Lethal Autonomous Weaopons Systems (LAWS): LAWS are AI-designed weapons that can engage autonomously. The idea of engaging in deadly warfare without a human controlling or overseeing the use of weapons is both dangerous and scary.