Testimony of Professor Jaime Carbonell

before the Committee on Science, Space, and Technology, Subcommittee on Research and Technology and Subcommittee on Energy, of the U.S. House of Representatives on the hearing titled,

"Artificial Intelligence - With Great Power Comes Great Responsibility"

June 26, 2018

Good Morning. My name is Jaime Carbonell. I am a professor of computer science and the director of the Language Technologies Institute at Carnegie Mellon University.

I would like to thank Chairman Comstock of the Research and Technology Subcommittee and Chairman Weber of the Energy Subcommittee for inviting me to testify today. I would also like to thank Ranking Member Lipinski, Ranking Member Veasey, and all of the Members of the Committee for your interest in Artificial Intelligence. It is a field in which I have been working for over 40 years.

Background: Artificial intelligence (AI) is the ability to create machines who perform tasks normally associated with human intelligence, from autonomous driving and spoken language comprehension to medical diagnosis and scientific discovery. Throughout its roughly 60-year history, AI has been incubated, admired, dismissed as infeasible, respected, maligned, feared, and more recently mainstreamed. Progress has been inexorable over the years, based on faster hardware, larger memories, new computing paradigms such as parallelism in graphical processing units (GPUs), major algorithmic innovations such as deep neural networks, and a plethora of successful applications driving the current industrial AI boom.

A Brief History: The term "Artificial Intelligence" was coined in 1956 at a conference in Dartmouth University by the four founders of the field: Herbert Simon, Alan Newell, Marvin Minsky, and John McCarthy, in their shared conviction that human thought could be reduced to symbolic computation on digital computers in the near future. Simon and Newell founded the AI efforts at Carnegie Mellon University, Minsky at MIT, and McCarthy at Stanford. Although there were some early successes, such as propositional logic provers and checker-playing programs, the true magnitude and complexity of human reasoning and the difficulty in replicating aspects of human intelligence in digital computers ensured that AI was a very long-term endeavor, not just a single moonshot or even a Manhattan Project.

The 1960's witnessed the founding of most of the AI subfields: computer vision, natural language processing, speech recognition, robotics, automated reasoning and planning, and so on. In the 1970's symbolic approaches to AI dominated, including reasoning based on first-order logic, and in the 1980's the first substantial commercial benefits of AI were reaped with the advent of rule-based "expert-systems". These systems encoded narrow human expertise in commercially important areas such as automatically configuring digital computers, or deciphering mass-spectrograms. Another example of deep narrow expertise was in chess playing machines. IBM's Deep Blue, based on the earlier Deep Thought chess playing program from Carnegie Mellon University, beat world champion Gary Kasparov in 1989. Additionally, the field of modern Machine Learning was founded in that decade. During the 1990's steady progress was made on virtually all the component AI technologies, including neural networks for learning, combining vision, navigation and manipulation into operational robotics, and combining speech recognition and natural language processing into dialog systems. The 90's also heralded the statistical approaches in AI and machine learning, reducing many cognitive problems to mathematical optimization. Strangely, the 1990's are sometimes called the "AI winter", since much of the progress was not externally visible, and other technologies such as the advent of the internet and early electronic commerce captured the limelight.

After the turn of the century, industry started taking AI more seriously. The search engine companies, for instance worked on semantic associations to find information beyond the exact keywords in a query. For instance, asking for "inexpensive vehicle insurance" yields results about "cheap car insurance." Those same companies accessing huge amounts of electronic text in multiple languages to train cross-language associations were able to build the first widely-used general-purpose machine translation engines. The rule-based expert systems of earlier decades evolved and became embedded into manufacturing systems and workflow-planning, and otherwise entered routine practice, and were no longer the subject of academic inquiry. Instead, academia sought new challenges, such as integrated virtual agents, autonomous robots, multilingual speech-to-speech translation, and so forth. The movement of "Artificial General Intelligence" (AGI) took root where the focus on performing very well at a narrow difficult task was replaced by the desire to exhibit adaptive intelligence to perform reasonably across many tasks, i.e. re-focusing on the initial goal of AI to create human-like intelligence.

In the current decade interest in AI has exploded, in large part due to the resurgence of deep networks in machine learning, at the core of AI, and their ability to perform many tasks much better than before – tasks ranging from self-driving vehicles and robotic assistants to robust speech understanding to semi-intelligent chatbots such as Siri, Cortana and Alexa. All the large tech companies have sprouted AI labs or divisions, as have many in other disciplines including finance and manufacturing.

AI and its sub-disciplines: Borrowing the title of the famous 1955 movie we can say that human intelligence is a many-splendored thing; it is also a many-faceted thing. Vision and physical coordination is not a trait unique to humans, but inherited from our ancestor species. On the other hand, complex language, writing, higher mathematics and abstract reasoning, most scientists would attribute uniquely to our species or at least our genus. AI attempts to understand and replicate all aspects of human intelligence, including those that evolution took many millions of years to perfect. But, to make the ultimate challenge somewhat more tractable, AI is divided into multiple somewhat overlapping disciplines:

- ***Sensing***: Using vision, sound, touch and super-human sensors such as sonar, radar, and lidar to sense and form a mental model of the immediate environment, and update the model dynamically.
- ***Communicating***: Using language, written or verbal, to convey information, questions or commands to other agents in the social group, as well as to record knowledge for future reference.
- ***Acting***: Effecting changes in the physical world by manipulating objects, navigating to other locations, observing the outcomes of the actions, and recording the results
- ***Reasoning***: Envisioning performing physical or communicative actions in the world, inferring their consequences, and planning sequences of actions to achieve desired goals, as well as inferring the likely actions of other agents.
- ***Learning***: Acquiring knowledge and experience to improve other cognitive tasks: sensing, acting, communicating and reasoning, including the ability to solve new problems or address new situations.

Many established AI areas map onto one or a combination of the above disciplines. For instance, computer vision is primarily sensing, but also learning to recognize, store and classify patterns. Natural language processing is primarily communicating, but contains elements of reasoning (e.g. semantic interpretation) and learning. Robotics, while centered on sensing and acting, touches on all the AI disciplines. Sometimes AI disciplines are confused with the overall endeavor, as in "are you doing AI or machine learning?" This is like asking "are you a car repairman or an engine mechanic"? If someone is practicing the latter they are also practicing the former. Similar confusion is evident in "should we invest in robotics or AI?" Or, "Should we build a chatbot rather than doing AI?" Instead the better question is "should we focus in Robotics – or on chatbots – as the most relevant part of AI for our business?" All of the subfields of AI are reinforced by advances in adjacent subfields, and some like machine learning are cross-cutting and ubiquitous.

Applications of AI: In many ways AI is the ultimate meta-technology, the technology that enhances most other technologies whether they be in strategic planning, manufacturing, healthcare, transportation, customer contact, entertainment, forensics, mining, agriculture, or even scientific research. AI extends to defense and national security, ranging from intelligence gathering and interpretation to operational warfighter capabilities. And, AI is playing an increasing role in humanitarian areas such as workplace safety and disaster relief. Let us look at a few illustrative applications of AI, focusing on the areas this writer knows best.

- ***Question Answering***: Wouldn't it be great if we could bring to bear all the world's knowledge – or at least all the publicly available knowledge – to bear in answering automatically every burning question? Starting with DoD funding, followed by IBM's Watson in the Jeopardy Challenge, to today's open-domain systems, Q&A has emerged as a major challenge task for AI. I am proud to say that Carnegie Mellon has been a central player in Q&A from the get-go in government programs, through helping IBM build the original Watson, to heading the leader boards in current Q&A challenges.

- ***Autonomous Driving***: Imagine drivers who do not get distracted, do not get sleepy, always know their routes, and can be counted on to stay always sober. That's the safety promise of autonomous vehicles – not an absolute safety guarantee, but nonetheless a major improvement. From the Navlab project at CMU in the 1980's to "No hands across America" autonomous highway driving in the 1990's to present large-scale commercial endeavors, autonomous vehicles are the future of safe transportation.
- ***Workplace safety***: What if blue-collar workplace accidents could be predicted and thereby reduced? AI, in the form of historical workplace data analysis, correlating accidents with workplace conditions, safety inspections and behavioral indicators, can help do just that, as evidenced by joint work between our university and a company called Industrial Scientific, already helping improve worker safety in many companies across different industries such as construction, mining, manufacturing, etc.
- ***Massive multilinguality***: There are some 6,000 languages in the world, but only the top 2% or so by population and economic significance have translation software. What if there is a natural disaster requiring international assistance, or what if the US military needs to coordinate with local authorities for the other 98%? NSF, ARO and now DARPA are addressing this challenge to build information extraction from low-resource languages and rudimentary translation in a matter of days, rather than years.
- ***Game theory***: AI has proven itself at full-information games such as chess and go, and more recently at partial-information games such as Texas holdem poker, beating the human champions in each of these games. The same technology is used to optimize kidney exchanges, negotiations, and other partial knowledge high-payoff decision tasks.

The above examples represent a very small cross-section of a much larger AI application space. How about agriculture? AI is optimizing planting and fertilizing plans and schedules as well as creating robot-controlled tractors. Law enforcement? Voice and facial identification are helping forensic analysis. Music? AI-based accompaniment and AI-based tutors for different instruments. Healthcare? AI-based DNA analysis for risk factors, robo-surgery for super-human precision, analysis of patient records, and much more, are emerging. Finance? Hedge-fund analytics based on machine learning, investment risk, and so on. Education? AI based tutors in mathematics, language, and computer programming are emerging. There is nary an industry untouched by AI, and its impact will significantly grow and become more manifest in the coming years across the board.

Artificial General Intelligence (AGI): AI is rapidly embedding itself into just about every aspect of our lives; its utility is beyond question. However, how about the original AI dream of creating general human intelligence? Many AI researchers have not given up on that goal, though many feel the path leads through the various components of intelligence before striving for a grand unification. Other AI researchers are content with narrow AI, that is, task-oriented AI and not general intelligence, since that is what drives our economic engine. Yet others, impatient to return to the AI-genesis goal of creating human-level intelligence coined the phrase "Artificial General Intelligence" seeking to leapfrog a component-based approach with a more holistic one.

Some AI researchers, such as this writer, view AGI as an aspirational goal, working to generalize AI methods from the narrow to the broad, from the specific to the general. For instance, consider transfer learning. If a person learns to drive a car, perhaps through instruction and considerable practice, and then she is asked to drive a van or a small truck, chances are she can do so, albeit with a few awkward moments adjusting to the height and size of the new vehicle. However if the software trained for a self-driving car were to be extracted and implanted onto a small truck, disaster may ensue – judging distances to other vehicles would be off due to the higher camera angle, the breaking distance and turning radius would not behave as expected, and even the lane-detection and following might fail. The difference is that humans transfer what they learned previously while compensating for the known differences, such as height, angles, size, vehicle responsiveness, etc. Transfer learning strives to do exactly that in an automated manner, learning what to keep, what to discard, what to modify and how to modify it across related tasks. Currently fielded AI approaches require retraining from scratch, insensitive to having learned a very similar task. Hence transfer learning is a small but important step towards AGI.

Another example is deep neural networks whose topology must be currently hand crafted by researchers for a given task, e.g. for face recognition vs reading MRIs, or for machine translation vs text mining. Each task requires determining how many nodes are required, arranged into how many layers, what type of layer-based combination functions, what kinds of connectivity among the layers, and so on. Researchers, including this writer, are striving to create self-configurable deep networks that change their structure automatically to optimize task performance. That is another small step towards AGI. Many other researchers are also investigating how to make AI more general, one step at a time, in many different ways and in different subfields such as in robotics, natural language processing, automated reasoning or machine learning.

However, a few researchers are less patient, trying to reach AGI by more direct means, and though not yet having achieved the desired breakthroughs, are nevertheless determined to pursue the proverbial gold at the end of the long rainbow.

Some common AI myths: Some popular beliefs or claims about AI deserve the label of "myth," as they are contrary to observations and informed opinion, including the following:

- ***One stop shop for all of AI!*** AI is not binary; it is not something one has or fails to have. The field continues to progress at a good pace; over time more sophisticated and capable

AI methods and systems are created. Any vendor claiming to be the ultimate provider of everything AI is selling a bill of goods or a pig in a poke. As more AI capabilities come online we must be ready to see whether and how to employ them, and we never know for certain where or by whom they will be created.

- ***AI = deep neural nets.*** Over its history many AI paradigms emerged, dominated the field, and then fell out of favor to newer methods: first-order logic, rule-based systems, the first coming on neural nets (1980s), symbolic machine learning, the second coming of neural nets (1990s), statistical machine learning, probabilistic reasoning, and now the third coming of (deep) neural nets. Actually, every AI paradigm leaves its mark, and many powerful systems use hybrid approaches, not just the latest method *du-jour*, even as convolutional and recurrent deep networks are proving to be very powerful.
- ***AI = suite of software tools.*** This is no more true than carpentry being just a box of carpenter's tools; one also needs the master carpenter. Surgery is not just a suite of scalpels, sutures and other surgical implements; one needs the experienced surgeon. AI tool suites are major enablers of novel AI applications, but the skilled AI practitioner is an integral part of the equation.
- ***AI is impossible.*** This used to be a common way for pundits to gain attention, including some philosophers, claiming that AI would never create a champion chess player, and would never understand human speech, and would never drive a vehicle in traffic. Over time all were proven wrong. Now the more common claim is that AGI is impossible. That one will be harder to disprove as AI systems become more general and more powerful, the proverbial goal posts will be moved from passing the Turing Test[1] to ever harder tasks as each is accomplished over time. Time and much research will tell whether true AGI is indeed achievable.
- ***AI will cause massive unemployment.*** A variant of this claim has been made after every major technological advance, but the opposite is typically true – there was more employment after the industrial revolution than before, or after the introduction of information technology and the internet. AI is already deployed in many industries, yet we have the lowest unemployment rate in recent times. Instead, AI will likely change the nature of work, displacing workers in some fields over time and creating other remunerative jobs. Rather than speculation, careful studies by economists, AI scientists and policy makers are required to make better predictions of labor market effects of AI – we simply do not know them in any detail. But it is fairly clear that the countries most advanced in AI technologies will reap its benefits, and others will be left behind, just like in earlier industrial revolutions, as the effects of AI will indeed be global.

[1] The most popular version of the test designed by Alan Turing, the famous code-breaking British mathematician, is whether a person asking questions of two hidden responders, one a human the other a machine, can tell which is which after a brief period of time based on their responses. If the questioner cannot tell them apart, then the machine is said to have passed the Turing Test.

Glimpsing into the future of AI: The most reliable prediction one can make relating to AI is extrapolating current trends into the near future:

- *AI will become increasingly ubiquitous in everyday life*, from general chatbots, to individualized health monitoring, to personal assistants. Turing award winner Professor Raj Reddy from CMU, predicts that personal assistants will evolve into “guardian angels”, always-on AI system helping people prepare for meetings, monitoring news or social media for items of immediate interest, reminding users to exercise or to not touch that temping chocolate dessert. These “angels” or “cogs” as this writer prefers to call them will be driven by observing, learning, and receiving instructions from their users.
- *The power of underlying AI technologies will increase,* as machine learning, language technologies and robotic sensing improve due to more powerful GPUs, improved deep neural network architectures, better training algorithms that combine raw data with domain knowledge, and so on. These will increase the range of practical AI applications across the board.
- *AI will expand into new areas including cyber security*, where research is already establishing that zero-shot malware detection is possible via machine learning methods trained on prior cyberattacks and malware profiles. This is a major improvement to waiting for a new virus to infect many machines before viral signature is disseminated by the cyber protection provider. AI can also help in energy conservation, including smart-grid and smart-home power consumption, as well as optimizing renewable power installations via predictive local-climate models.
- *Artificial general intelligence will witness a resurgence*, not so much from a major paradigm switch in AI, though that is what its proponents desire, but rather from evolutionary forces in bringing to bear transfer learning, self-configurable neural networks, more flexible robotics, more natural and general dialog systems, and other improved AI methods coming together into larger systems of wider capabilities.
- *The demand for AI talent will far outstrip the supply*. Industries engage in bidding wars and employee poaching in attempts to fill open positions in AI, raising compensation levels well beyond any previously seen. Universities strive to increase the supply, but when industry attracts top university faculty – and this is happening with increasing frequency – it is very difficult to maintain, let alone expand, educational offerings.
- *AI in other countries will surge*. This is already happening in Russia, India, and especially China (see the following section), but the trend will expand to many more countries that realize AI is the crucial technology of tomorrow, and quickly ramp up efforts in research, education and especially industry.

- *AI ethics will become a central concern.* As with virtually all technologies, AI can be used for the benefit of human kind or its detriment. For instance, electricity powers our world, but also can electrocute; the internet is a major economic driver and wonderfully empowering by democratizing information but it is also used to spread hate-speech and enable cyber bullying. AI is no different, and we must consider ethical implications, where "we" refers to AI researcher and practitioners, ethicists, economists and policy makers working together. For instance we may want AI to help us establish and maintain healthy lifestyles, but we may not want that same AI to breach our privacy and inform others of our individual health problems.

US Dominance in AI: The United States has been the undisputed world leader in AI from its inception in 1956 through the early days of overoptimistic expectations that AI was just around the corner, to the realism that AI is difficult – much harder than rocket science – and its ultimate success in a wide variety of challenging tasks. In the 1980's Japan's Fifth Generation Project challenged the US AI dominance, and the 1990's Europe's Economic Commission attempted to surpass US AI efforts, but neither succeeded in large part due to the foresight and perseverance of the US public sector. Unlike the above limited-lifetime efforts, the National Science Foundation (NSF), the Defense Advanced Projects Agency (DARPA), the US intelligence community including the Intelligence Advanced Projects Activity (IARPA), and the research divisions of each of the armed services (ONR/NRL, AFRL, ARL/ARO) persevered in supporting AI through all of its growing pains, its successes, its more challenging times, and its arrival in the mainstream to become a top candidate for the most versatile and valuable industry of our century. The foresight to persevere, to strive to win the long game of AI, is a credit to the cited government institutions through multiple administrations and through remarkable bipartisan unity in the national interest. Hence, my top recommendation is kudos to these agencies; they do an excellent job, please extend them as much support as possible.

However, the US dominance in AI is being challenged like never before. Many countries are striving hard to improve their AI know-how, work-force, and industry, including China, Russia, Korea, Japan, Germany, the UK and India. Consider China, for instance, which has made AI a national priority. On May 4, 2018 CNBC reported "China is determined to steal the AI crown from the US and nothing, not even a trade war will stop it. China's 2030 plan envisions building a $1 trillion AI industry." Wired Magazine reports: "China will be the world's dominant player in artificial intelligence by 2030. This isn't a prediction by a researcher or academic, it's government policy from Beijing." Whereas these statements may be on more alarmist than reliable predictions, they clearly indicate Chinese intent. China's national priority is AI pre-eminence. Even General Secretary Xi Jinping is reported to have AI books on his shelf.

It is difficult to estimate the very substantial level of AI funding in China, but there some components include: 1) The city of Tianjin is committing $5 billion to support the new AI industrial park. 2) The Feb 20, 2018 statement in the Financial Times saying "*Last year almost half the global investment into AI startups went to China, up from a mere 11.3 per cent slice in*

2016", 3) On June 22, 2018 the South China Morning Post reported: "*China's Ministry of Science and Technology has funded at least eight AI-related research projects over the past six months to the tune of 2.73 billion yuan (US$430 million) from the central government budget*" and "*The China Academy of Sciences (CAS) which has over 300 labs and four national research centres, received over 2.7 billion yuan for its 11 fundamental science projects last year, although it's unclear how many of these are directly-AI related*." China has already far surpassed the US in terms of patents granted for AI technologies, according to Quartz, May 2, 2018.

In contrast, the United States spends about $2 billion per year on AI research, according to OSTP, not counting substantial DoD spending on procurement that includes applied AI research. The bottom line is that both countries take AI very seriously and that China is striving and investing to achieve AI leadership. It will take a greater effort for our country to maintain international leadership in AI.

Some recommendations: The following suggestions address what this writer believes to be national priorities with respect to AI, whose neglect will be to our collective peril.

- ***Support the US government agencies which helped create and foster AI.*** NSF, DARPA, IARPA, parts of NRL/ONR, ARL/ARO, AFRL, etc. Without DARPA in the early days, AI would not exist in its present form. These agencies do a good job of funding and shepherding AI research. They deserve increased financial support to do their jobs even better, especially in light of the sudden large-scale international competition. No matter what other measures the US government takes with respect to AI, keep and grow the proven successful processes and agencies.
- ***Address the AI personnel shortage.*** We need to produce more AI researchers, especially more US-citizen or permanent-resident AI researchers. One way to do so is to provide scholarships for US students pursuing AI careers. One thought is that these could be funded by asking companies in an expanded merit-based H1B visa program to pay $30K or so per sponsored visa. Then if there are some 100K new H1B visas per year that would create a fund of $3 billion per year, which would provide something like 30 thousand scholarships. Whatever the mechanism or the numbers, we should be training more AI researches and engineers. The flipside is to retain more the foreign AI talent trained in US universities, which is also addressable by increasing H1B visas
- ***Address the exodus of AI faculty from universities.*** Industry is making offers that few AI faculty can refuse, with the inevitable loss of top university AI talent making it very difficult to teach and mentor new generations of students. We must stop eating our seed corn, though it is not clear how to change incentives to attract and preserve more top-talented AI faculty.
- ***Encourage AI at the undergraduate level.*** AI used to be for PhD students, then MS programs in AI or sub-disciplines of AI started including several at CMU. Now is the time to start AI as an undergraduate major, as we are experimentally doing at CMU.

- ***Consider funding a national AI center.*** Other countries have national AI laboratories or are creating same, as a nexus between government, industry and academia. The US should do so as well, both as a vehicle for growing the volume and scope of AI activities but also to enable large-scale projects that require sustained collaboration among dozens of researchers, thereby balancing the smaller efforts supported by the NSF and other funding agencies. Such a center, though, should be stable for a long period of time to be truly productive, not a one-year flash-in-the-proverbial-pan.