New Technologies to Enhance Computer Generated Interactive Virtual Humans

Ke-Thia Yao, Jeremy J. Liu, & Nicholas J. Kaimakis
University of Southern California
12015 E Waterfront Dr,
Los Angeles, CA 90094
310 448-8297, 310 448-9187, 847 487-7291
kyao@isi.edu, {jeremyjl, kaimakis}@usc.edu

Dan M. Davis
HPC-Education - Consulting for
University of Southern California
6275 E. 6th St.
Long Beach CA 90803
310 909-3487
{dmdavis@acm.org}

Abstract: *Conversation-like exchanges of information between users and computer-generated virtual humans can be computationally demanding. Many of the Grand Challenges facing the simulation community will further stress the concerned computer resources. These computer/human interfaces are currently being enhanced and will require both technical advances and societal changes. This paper offers a list of Virtual Human Grand Challenges, solicited from leaders in the community, for purposes of this discussion. Then the authors report on emerging technologies that will further enhance the "human" qualities of the previously mentioned interfaces. Some of these technologies are hardware based while others come from software developments. Further, system engineering advances are presented with the intention of making better use of all the aspects of the system to achieve simulation goals. On the hardware side, there is the emerging technology comprised of several threads of quantum computing, e.g. USC's 2,000 Qubit Quantum Annealer. Software advances of note are discussed in the areas of voice recognition, natural language processing, and deep learning. The use of quantum annealing to enhance deep learning is particularly examined. Systems engineering issues are analyzed, as these emerging technologies do not stand alone and must be integrated into or replace legacy systems, mandating close adherence to a range of standards. In each of these cases, the authors discuss theirs and others' experience with the technologies and explicate both the benefits and the costs of such implementations. The paper is written so as to assist the researcher and the systems implementers in choosing which of the new technologies warrants consideration for inclusion in their own work and which technologies will eventually require adoption in order to stay compatible with other systems. Other technologies may require more effort in implementation than they would provide utility in practice.*

1. Introduction

The defense establishment of the United States is faced with accelerating operations tempos and declining funding, yet they must defend against a proliferation of foes (Figure 1) that are increasingly sophisticated and may be either asymmetric movements or nation-state adversaries. One way to resolve this conundrum is to increase the reliance on computer assistance, simulations, robotics, virtual humans, and artificial intelligence. Although many of the technologies and techniques in this paper would be applicable to all of these disciplines, the paper will focus most closely on interfaces using virtual humans. While the strides in making the virtual humans more "human" have been dramatic, and there is growing evidence of their effectiveness, there are still areas that are ostensibly outside of current capabilities. Several of these shortcomings are vital to conveying the proper feeling of "liveness." These virtual humans have already demonstrated great utili-

Table 1 - Gallup Poll of US Enemies

N. Korea, Russia, Iran, China Rotate Top Spots on "Greatest Enemy" List
What one country anywhere in the world do you consider to be the United States' greatest enemy today? (open-ended)

	2014	2015	2016
	%	%	%
North Korea	16	15	16
Russia	9	18	15
Iran	16	9	14
China	20	12	12
Countries in which ISIS operates	0	4	5
Iraq	7	8	5
Afghanistan	5	3	4
Syria	3	4	4
Other	13	15	11
None	2	1	4
No opinion	9	12	11

GALLUP

ty in the areas of training, education, mentoring, psychotherapy, counseling and interactive public access to limited personnel resources of historical interest, *e.g.* the personal stories of holocaust survivors.

A note of definition: “Virtual Human” is a concept that has slightly differing definitions in the simulation, artificial-intelligence, motion-picture, and other communities. This paper does not attempt to resolve those differences, and it will instead state how the term is to be used below. In this paper, a Virtual Human refers to any series of computer programs that effectively portray human appearance, voice, and actions that can interact with “live” humans without the computer generated half of the dialogue receiving on-line human direction or live intervention. As used here, the term encompasses the most extreme expression of “Virtualness” in which it entails using completely computer generated images, voice, and conversational responses. However, it also considers sets of conversational responses selected by using natural language processing to analyze user input and play appropriate video clips. These clips will have been recorded by a real person at an earlier time, giving their own answers to a comprehensive bank of anticipated question, thus providing the imagery and the voice.

The paper will begin with an introduction and some background of military needs for and uses of Virtual Humans (VH). This will naturally segue into a discussion of the Grand Challenges facing the creators and implementers of VH avatars in operational programs. The next major issue to be covered will be the emerging capabilities of Deep Learning (DL): genesis, uses, need and future. Then, a section addresses the new technology of Quantum Annealing (QA), which is a restricted, but operational, branch of the larger area of Quantum Computing (QC): its history, progress and future. This analysis will cover the facilitation of deep learning. These all lead to the author’s analysis of how and why QA and DL could address the delineated Grand Challenges.

1.1 Virtual Humans in the Defense Environment

A virtual human is a virtual reality creation in which an avatar is created, often based on a real person, and attempts to recreate the appearance, voice, feel, and interaction that a live human would produce. With the advancement of several new technologies, including but not limited to natural language processing, virtual reality (VR), computer generated imagery (CGI), machine learning, and virtual learning, the uses, as well as the limits of virtual humans are becoming evident. As a research institute with a specialization in virtual reality programming, the Institute for Creative Technologies (ICT) is the home of myriad simulation projects, and it considers virtual reality to be its primary focus. Researchers at ICT have generated SimCoach, New Dimensions in Testimony (NDT), PAL3, MentorPAL, and other generalized programs under learning sciences, medical VR, mixed reality, narrative storytelling, social stimulation, virtual humans, and vision and graphics. The breadth of knowledge available is significant and the use of these resources has allowed advances in the implementing of virtual humans. The presentation to the user can take many forms, as shown in Figures 1-3 (All ICT Photos).

Figure 1 – Video of NDT speaker shown in 3-D holographic display

Figure 2 – MentorPAL video clip on 2-D monitor

Figure 3 – Fully Animated CGI in SimCoach

Although a virtual human may seem as simple as remodeling a human using CGI, it turns out that it takes significant study and effort to implement a Virtual Human, and the process can consume considerable computing power to do so effectively. The essential elements that go into the creation of a virtual human with lifelike abilities include natural language processing, machine learning, VR, CGI, and social stimulation. Natural language processing (NLP), will be the main focus of this discussion, though the same argument concerning the limits of virtual humans can be made with several of the other components. Natural language processing composes “an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things” [1]. Using this definition within the context of virtual environments, NLP tools allow computer technology to recognize voice input, analyze voice tone, provide lifelike conversation, retrieve information, and many other applications in combination with machine learning. Recent developments in NLP have made significant advances, like "a single convolutional neural network architecture that, given a sentence, outputs a host of language processing predictions: part-of-speech tags, chunks, named entity tags, semantic roles, semantically similar words and the likelihood that the sentence makes sense (grammatically and semantically) using a language model" [2].

Many applications, through interpretation of language, have been quickly advanced by the researchers focusing their studies on the field. At ICT, progress has been made in training and learning environments [3], multi-party dialogues [4], ethics and cooperation [5], health applications [6], and representation and reasoning [7]. Although automated speech recognition (ASR) is far from perfect, prime software makes Virtual Humans practicable, "Google achieved 73.3% of exact recognized phrases with a 15.8% [Word Error Rate]" [8], and the technology will continue to improve. It is also important to note that many of the errors in ASR are caused by slurred speech, cultural slang, and context, stemming from a "lack of consistent units of speech that are trainable and relatively insensitive to context" [9]. Although this causes problems when comparing these transcriptions with global data, models can be trained locally to a particular person's voice or a low resource language, using software such as CMU Sphinx from Carnegie Mellon University as applied to languages such as Arabic [10]. Although these customizations and build-your-own languages can be more accurate, they take time to implement and often require increased local processing power.

As it stands now, the primary issues with natural language processing include machine translation, precision, data storage, efficiency, and computation power; meeting these are the foci of this paper. All of these will continue to improve with time, as hardware, software, data storage and ease of access are areas within which new research is emerging and seems directly applicable. Specifically, the speed and application of quantum computing will enable significant advances in NLP and its applications: one of which is the critical area of enhancing virtual humans.

Virtual humans constitute-and should continue to constitute-a major role in virtual learning, virtual storytelling, VR/AR, and constructive simulation. Within the military context, virtual learning environments provide useful mechanisms for initial training as well as lifelong training. An example of this in implementation can be found in PAL3; "the PAL3 system was designed to accompany a learner throughout their career and mentor them to build and maintain skills" [11]. Modern learning calls for new methods of information transfer. Online tools like Khan Academy, YouTube, Coursera, Lynda, and other massive open online courses (MOOCs) are growing in popularity, and the trend does not seem to be slowing down. Many of these tools suffer from the lack of the immediacy and motivational qualities of personal interaction with a human.

Figure 4 – Subject receiving direction in light stage used for generating 3-D imagery for holographic display

Storytelling and gaming are also important facets of the virtual field. The New Dimensions in Testimony project "allows people to have an interactive conversation with a human storyteller (a Holocaust survivor) who has recorded a number of dialogue contributions (Figure 4), including many compelling narratives of his experiences and thoughts" [12] . The project's mandated mission of preserving the stories of Holocaust survivors can be applied to other important persons and historical sagas of value and interest. Advances in gaming allow for utilization in learning, entertainment, healthcare, and lifelike training for areas like combat, base security, reaction team strategy, medical emergency responses, interfaces with civilian populations, and critical thinking. The SimCoach system was developed to assist Post Traumatic Stress Disorder (PTSD) subjects and to "motivate users to take the first step – to empower themselves to seek advice and information regarding their healthcare" [13]. These virtual systems have been shown to have more success, generating deeper levels of confidence with patients than even live healthcare interactions [13]. The successes of these operations thus far encourage research into potential future applications.

The benefits of virtual environments include making the most skilled instructors and resources available and scalable, engaging new learning environments, confidentiality, and global accessibility. Some of the negatives of virtual learning are the difficulties of personalizing learning to fit individual students, handling live question, responding to different learning styles, and similar limits to NLP and other processor and data heavy systems. On the positive side, this approach does allow for using only the best instructors and can avoid the classroom stricture against teaching to the lowest common denominator, having to respond to questions in a generic way. It can be trained to detect and cater to every student's learning style, and other constraints faced by live teachers. Future fields of focus will include streamlining the creation of such virtual environments and interaction. For example, it is possible to conceive a system which takes in a live recorded interview with a U.S. President, or other prominent or knowledgeable figure, and generates an interactive environment in which a user can experience the interview by asking the questions themselves. Or a true virtual assistant that obviates the need for an assistant might provide a radical shift in the way the DoD functions. Such tools are already in existence, and they will only become more prevalent and powerful. However, the need for extensive processing power, efficiency, and data storage and transfer remains a limiting factor for many such developments. The aforementioned applications call for considerable further exploration and research. They may well be amenable to a quantum annealing approach to deep learning enhancement.

1.2 The Grand Challenges of Virtual Humans

In establishing a list of grand challenges, the authors first began with conceptualization sessions among them. They then went to local experts in VH implementations, including some of the "Deans" of the discipline. From that input, they consolidated and articulated the resultant list. Taking that list, they shared it for comment with others in this discipline located outside of Southern California. When no new suggestions were forthcoming from other groups, the authors were satisfied they had, at the very least, identified a list of essential challenges in this area. The list is presented below without any assertion that there are no others, or even that some of these challenges may have been resolved without the author's knowledge. Further, they invite the reader to comment on or add to the list by contacting Dan Davis at ddavis@hpc-educ.org.

Table 2 - Grand Challenges in Virtual Humans (VHs)

1	VH's recognizing and responding appropriately to: • Sarcasm • Humor • Irony
2	Distinguishing multiple speech acts in a single sentence: (*e.g.* "That's interesting about the present, but what about your childhood?": *i.e.* recognizing the segue and responding to the second part of the prompt)
3	Haptics (*e.g.* a handshake)
4	Emotional expression(s) reflecting the VH's own feelings
5	Knowing when and how for VH to utilize lifelike body language or hand movements
6	Being able to act in the world (virtual or real) as well as talk. (That is, have the VH be able to plan out and take actions that affect the state of the real or virtual world.)
7	VH's ability to learn from interactions with persons and use what it learns intelligently
8	Merging clips to deliver responses with continuous smooth motion
9	Providing believable empathy/understanding as an AI
10	Recognizing and reacting to gist of user statements, emotions, and body language
11	VH ability to interact with multiple humans or sources of input simultaneously
12	VH interrupting speech or being interrupted

In Table 2 above, no inferences should be drawn from the order or ordinal number attached to any challenge. They are presented in the order in which they were discussed. This paper is not considered to be an in appropriate venue to settle which challenge is the most critical. This will vary in the eyes of the stake-holders in the various disciplines. Each of these challenges is recognized in other fields of inquiry, *e.g.* linguistics, as some of the most difficult to define and to master. Yet they are all part of "being human," and persons lacking abilities in any of these areas are somehow considered deficient.

As these are complex behaviors that are generally learned partially from explicit instruction, they are difficult to learn in a digital computing environment. Some, *e.g.* sarcasm, are often dependent on tone of voice for identification. A way to detect context and voice tone would go a long way in helping resolve these challenges, but there are limits to digital power.

2. New Technologies and Approaches

2.1 Quantum Computing

Nearing the end of the Moore's Law growth of digital computing, many simulation professionals are concerned with solving the grand challenges set forth above. One of the alternatives frequently mentioned is Quantum computing. It has hopefully been considered an extension of computational capability since the Nobel Laureate Richard Feynman presented the seminal paper in 1982. In that paper he held that: "… with a suitable class of quantum machines you could imitate any quantum system, including the physical world." [14]. The authors have assiduously followed the development of such a machine and those devices still are, as near as can be ascertained, almost entirely at the test-bench phase. There seem to be no such "general purpose" quantum computer that is even nearing operation. There is one operational design in the quantum world: while not a general purpose quantum computer, it relies on very cold temperatures (15 milliKelvin) to create a useable quantum effect. [15]

Figure 5 - D-Wave Quantum Annealer at USC

This adiabatic quantum annealing device has been conceived, designed, produced, and delivered to the University of Southern California. It has been in operation since 2012. In its current configuration, D-Wave computers have a design providing approximately 2,000 qubits. The "approximate" figure is required as some delivered machines have some fraction of these qubits turned off and a small number of the qubits (~ 1%) are not stable after the processor reaches the target 15 mK. Figure 5 shows the D-Wave Two, as installed in the USC-Lockheed Martin Quantum Computing Center (QCC) at the Information Sciences Institute (ISI) in Marina del Rey. There is another in the San Francisco Bay area in a joint Google and NASA project, and one has putatively been ordered by the DoD's High Performance Computing Modernization Program. Others are in varying stages of procurement and delivery.

In recent years, other authors have touted quantum computing's ability to produce more power, using terms like "magic" to stir the imagination and whet the appetites of the user community. They point out that the capability of quantum computers arises from the different way they encode information. Digital computers represent information with transistor-based switches having a state of 0 or 1, labeled as a bit. In contrast, the basic unit of quantum computer operation, the quantum bit or qubit, can exist simultaneously as 0 and 1. A quantum bit, called a qubit, might be represented by an atom in one of two different states, which can also be denoted as 0 or 1. Two qubits, like two c1assical bits, can attain four different well-defined states (0 and 0, 0 and 1, 1 and 0, or 1 and 1). But unlike classical bits, qubits can exist simultaneously as 0 and 1, with the probability for each state given by a numerical coefficient. Data in a two-qubit quantum computer thus requires four coefficients. In general, n qubits demand 2^n numbers, which rapidly becomes a sizable set for larger values of n. For example, if n equals 50, about 10^{15} numbers are required to describe all the probabilities for all the possible states inside the quantum machine. That large number exceeds the capacity of the largest digital computer. A quantum computer should demonstrate incredible computational power because it can be in multiple states at once, a condition called "superposition." Also, perhaps more importantly, it can act on all its possible states simultaneously. The quantum computer can evaluate a series of optima that would be beyond the power of the largest digital cluster [16]. Evaluating a large number of parameters to "learn" patterns and behaviors is at the heart of Virtual Human enhancement.

The authors have witnessed and participated in the development of high performance computing for several decades and have developed a significant body of experience with newly introduced technologies. They were engaged in the very early introduction of parallel computing and aware of its rivalry with sequential computing and with vector computing. They heard the detractors of parallel computing argue the limits of parallelism and the proponents who argued that it could be used more universally. While acknowledging there are many problems that have remained outside of the easily parallelized arena, it is evident that the majority of all large-scale computational problems are now run in parallel. This is due to the application of new techniques to decompose both data and computation in effective ways. Such technology has proven very useful to the simulation community, which has many issues identical to the test and evaluation environment. By using super-cold processors, the D-Wave has been able to demonstrate accepted quantum computing, see Figure 6 to the right. Even if the projected speed-ups are not realized on this design, it is a workable and verifiable quantum computing device.

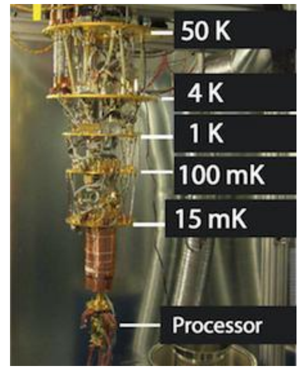

Figure 6 - D-Wave Cooling Mechanism **(D-Wave photo)**

The D-Wave machine in its developmental adolescence; it is no longer helpless, but its potential remains to be seen and characterized. This phenomenon was described by Professor Clayton Christensen in his series of books on disruptive technologies [17], *i.e.* a period in which the new technology is operational, but not yet competitive with technologies that have had decades of development and optimization. The criticisms and skeptical comments are familiar to all who have witnessed the displacement of older technologies. However, this community has also seen the claims of new capabilities that were then quickly set aside when old technologies improved faster than the new technology could prove itself. These issues are all part of this community's history and provide insights for analyzing the value of new technologies.

Several things need to happen before that D-Wave becomes a tool rather than a proof of concept device. The hard-wired connectivity paths need to be increased to represent a viable representation of the experimental environment sought to be simulated. A huge code base, both theoretical underpinnings and practicable software approaches, must be conceived and implemented. Perhaps decades of developmental advances will be needed to bring even this limited capability to its promise of revolutionary change. But this community will also remember the naysayers of the early days of personal computing and the skeptics of the first half of the 90's when the internet was first becoming available to the public at large.

Current plans at D-Wave include: more connectivity pathways, more qubits and, perhaps most critically, improved software in both the applications and the digital-to-quantum interface. The programmers at USC have used all of the D-Wave programs and become familiar with their Application Program Interface (API). D-Wave currently supports C/C++, Python, and MATLAB, three very common research languages. The company provides both on-line and live tutorial sessions to bring new users on-board. Much of the USC work focused on the treatment of large databases, known today as "big data.".

2.2 Deep Learning

Deep learning is another term that is used to describe an emerging technique in machine learning and artificial intelligence. The authors take it to be an extension of earlier work in the areas of neural networks, evolutionary computing, and data mining. But in deep learning, the refinement is done by several layers of convolution processing, in which some evaluation process sends along only such information that the layer values as beneficial. These successions of processing steps are call hidden layers and are the distinguishing feature of deep learning. Deep learning advocates note that this hierarchical approach to sifting large amounts of data is more in keeping with real word issues, albeit its need for more compute power.

Deep learning has achieved state-of-art in performance in a wide range of problems, including hand writing recognition, object recognition, classification in images, speech recognition, understanding natural language text, and adversarial games, [18]. A key characteristic of being able to tackle these problems is the need to find complex structures in high-dimensional data. Deep learning with its multiple layered structure and distributed representation turns out to be well suited for these problems as compared with other machine learning approaches and other manual knowledge engineering approaches.

The current interest in deep learning can be viewed as the third wave of neural network development. In the late 1950's the Perceptron ® captured the public imagination as a new type of electronic brain [19]. But, the interest in Perceptron waned when Minsky and Papert pointed out the limitations of a single Perceptron, which can only capture simple linearly-separable structures. In the 1980's, interest renewed as researchers constructed much more capable neural network models with multiple layers with multiple neurons per layer. But, again interest waned, because multiple layered networks are difficult to train and the computation power was not sufficient at that time. The current interest in neural network started in 2006 as Hinton *et al.* introduced a novel method to pre-train the network evaluation weights for deep belief networks using Restricted Boltzmann Machines (RBM) [20]. Moreover, the advent of graphical processing units (GPUs) with their Single Instruction Multiple Data (SIMD) capabilities reduced the computational bottleneck, often reducing weeks of computation to a few hours. GPUs are increasing in their power, but also face physical limits to their computational growth.

Of the three major deep learning approaches (feedforward neural networks, recurrent neural networks, and Boltzmann machines), the Boltzmann machine approach is the most amenable to quantum annealing. Boltzmann machines are probabilistic neural networks that implicitly define probability distribution over the activation states of the neurons in the deep learning network. Training a Boltzmann machine requires being able to repeatedly sample from the distribution of activation states. However, sampling for Boltzmann network with loops can be computationally expensive.

The reason restricted Boltzmann machines can be trained efficiently is because they avoid edges within a layer, which enables layer-based approximate sampling of the network. With quantum annealing, there is the potential of efficiently sampling networks with loops. This enables the creation of a broader range of networks, with more complex topology. The current generation of quantum annealers does support complete graphs; they also impose limits on the topology of the intra-layer edges. This is the definition of limited Boltzmann machines (LBM), which are strict supersets of RBMs, and demonstrate the effectiveness of the additional edges.

The rise of deep learning is related to the rise of big data. Deep learning models require very large datasets to properly train the neural network weights. The activation of a neuron depends on the activation of its neuron neighbors mediated by the weights of the connection, see Figure 7. The winner of the 2012 ImageNet Large-Scale Visual Recognition Challenge AlexNet [21] has over 60 million weights with 5 convolutional layers, max pool layers, and 3 fully connected layers. The network was trained on over 15 million images with over 22 thousand labeled categories. In 2014, Simonyan and Zisserman developed VGG net, which has 19 convolutional layers, and Microsoft ResNet [22] has 152 layers.

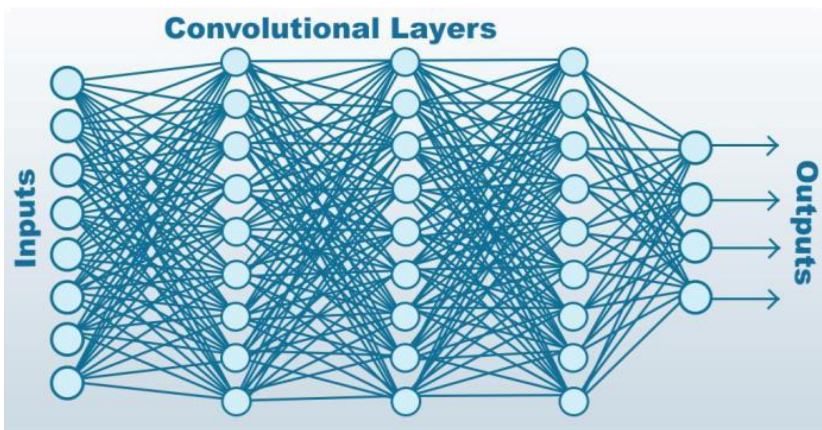

Figure 7 - Diagram Showing Successive Layers of Deep Learning Analysis

Deep learning does have limitations and there are recent research efforts that are addressing these limitations. One limitation is brittleness. Deep learning models have demonstrated remarkable ability to perform well in noisy data, for example the ability to recognize objects in photos with busy backgrounds. But, paradoxically it is often quite fragile. Given new classes of objects to recognize, the model must be extensively retained with many instances of the new objects as well as the previous know objects. In the games area, a deep learning-based Alpha Go program can beat top ranked human players, but the number of handicap stones is hardwired. Changing from the default handicap of 7.5 stones requires extensive retraining. An approach to address this limitation is to develop less monolithic deep networks, such as developing visual attention models to incrementally decode images and combining deep networks with artificial reasoning capabilities. Another limitation is the need for large labeled training dataset. Large datasets are often easy to find, but labeling is often a labor intensive process, *cf*. discussion of "sarcasm" recognition below. Researchers are developing unsupervised learning techniques, such as using Boltzmann machine-based auto-encoders; and incorporating artificial intelligence techniques (like reinforcement learning in the AlphGo) to self-generate labeled data.

3. Quantum Annealing

3.1 Simulated Annealing

Finding the extrema of some function is an essential tool in many learning algorithms. Most problems of interest can be described in terms of some loss function to minimize given training data and model parameters. Finding parameters to minimize this function is the heart of machine learning, and it is an arduous task. Non-trivial problems have dizzyingly complex feature spaces that are nigh-impossible to navigate. The number of possible solutions grows exponentially with the number of parameters one uses, and one cannot be sure any proposed solution is optimal. Exhaustive search is not viable in the least, so there exists a plethora of methods to estimate good parameters, among which include simulated annealing.

Simulated annealing is a random statistical process that allows us to move around a model's parameter space in search of a global optimum. Broadly speaking, the process involves starting with some set of parameters, making semi-random alterations to those parameters, evaluating the new parameters, choosing whether to stay with the old or new parameters, and then repeating all previous steps. The "annealing" part of the method comes from our probability of accepting the new parameters: the higher the "heat" of the process, the higher chances are of accepting "bad" parameter alterations in the hopes that there is a move out of local minima/maxima. The start of annealing has high heat so it explores larger portions of the feature space whereas the end of annealing has low heat, giving better precision and local search. Although simulated annealing has proven useful for optimization problems, there always remains the problem of getting trapped in local optima. Here, quantum devices may prove useful.

3.2 Quantum Annealing

Quantum Computing in a general, Feynmanian sense is still a ways off. D-Wave has offered a Quantum Annealer that depends of very cold temperatures that can be a first step in understanding the quantum computing processes. Due to design limitations, it can only do annealing at this time, but annealing alone, as discussed previously, is a critical capability and a is central to many of the grand challenges potential approaches. The annealer has shown quantum computing abilities, but has not as yet shown the kinds of quantum powers that its proponents envisioned.

Quantum tunneling is a phenomenon where an annealing process on a quantum device can escape local optima [22] whereas a classical annealing process may remain stuck. Better parameters may be obtained using quantum annealing, or at least, a different set of optimal parameters may be generated. Devices like the D-Wave promise such quantum effects and it is in the nation's interest to investigate how one might exploit these purported benefits for use in deep learning algorithms.

3.3 Quantum Annealing and Deep Learning

The following attempts to draw a clearer connection between deep learning algorithms and quantum annealing on the D-Wave device. A deep learning architecture is most simply described as a neural network with many stacked layers, and each layer's neural units are connected to the neural units of the layer above. The deep learning problem is to find a set of connection weights, J ,and unit biases, h, that will produce the correct output when given a particular input, where each neural unit $\sigma_i \in \{0, 1\}$ is either "on" or "off". It so happens that such a connection weight and bias description is easily converted into an Ising model with the following energy expression:

Equation 1

$$H = -\sum_{i<j} J_{ij}\,\sigma_i\,\sigma_j - \sum_i h_i \sigma_i$$

Happily, this is also exactly the type of optimization problem D-Wave is set up to solve. One can translate the deep learning architecture into a form D-Wave solves and vice versa, *e.g.* Figure 8.

What may not be obvious at first glance is why this could be a potential benefit to deep learning aside from the previously mentioned quantum tunneling effect, which might give different or better answers to the same optimization problems. Normally, deep learning architectures only allow connections between units of adjacent layers; units within the same layer or units in distant layers do not interact. This is done to

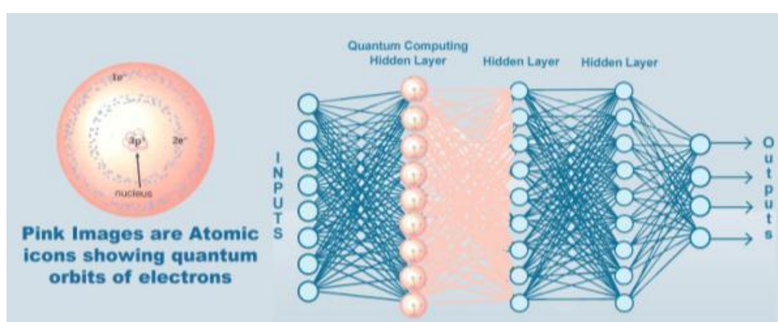

Figure 8 - Notional Diagram of Deep Learning using a Quantum Computing Layer

make computation feasible; otherwise the Ising model is an NP-complete problem [24]. However, what D-Wave offers is the ability to help optimize deep learning architectures that allow connections between neural units in the same layer, as in Figure 8. While Figure 8 shows only one layer as being enhanced by quantum annealing, it is possible that all of the layers would be quantum enabled. The proper mix of quantum and digital computing is one of the myriad research issues already apparent. This expands our possible network architecture choices and configurations which may not be feasibly computed on classical machines. Some preliminary work suggests that inclusion of extra connections between intra-layer units may produce better results for some learning problems.

One potential improvement is in Boltzmann Machines (BMs), where a small experiment indicated the inclusion of intralayer connections helped speed up the training process [25]. Like Ising models in equation 1, BMs are energy models composed of connected units, and it is straightforward to see one can try using D-Wave's device to represent them. Being an Ising-like model means a user would also have to worry about computational feasibility - full connectivity cannot be allowed on a classical machine. But with a quantum device, perhaps limited connectivity can be allowed. BMs are divided into visible and hidden units. RBMs add a condition that within each group there are no connections. The RBM can thus be thought of in terms of a bipartite graph, and it is the RBM that typically applied to problems because it is easy to compute [21]. The D-Wave device offers the opportunity to relax the RBM condition that no intragroup connections exist (see Figure 9) and explore if additional connections (called "limited" BMs) can benefit BMs in a practical way.

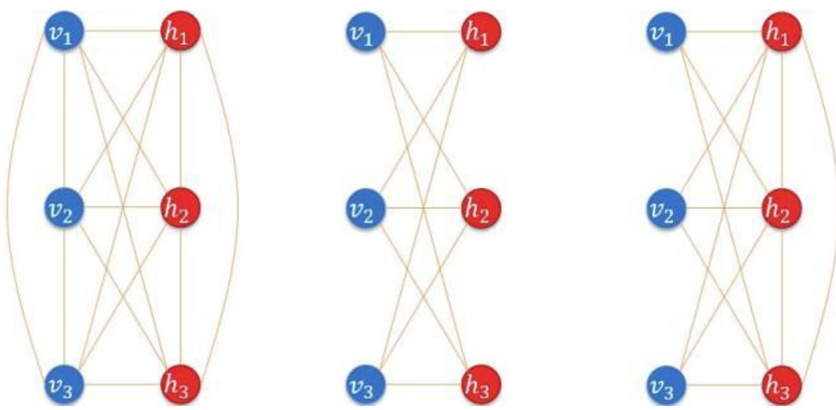

Figure 9 - Boltzmann machine configurations. Left: full connectivity.
Center: "restricted" BM with no intraroup connections.
Right: "limited" BM with intrahidden connections.

Currently limited BMs are being applied to neutrino experimental data to classify impact sites. That is, given some trajectory information captured from different angles, there is an attempt to predict which section of the experiment chamber a neutrino will hit. The effort seeks to determine if limited BMs can give some performance benefit over restricted BMs in this same task, and it is also hoped to make a similar comparison between limited BMs and other standard classification methods.

4. Discussion

The above much abbreviated reviews of the emerging capabilities of deep learning and quantum computing, as well as a synergistic combination of the two, lend credence to the pre-nascent breakthroughs in some of the grand challenges enumerated. The space limits on this paper do not permit going through each of the grand challenges, but an example may help illuminate the process the authors' envision. The first example is the first one on the list: "How does a machine learn to recognize sarcasm?" These are difficult linguistic and lexicographic questions for humans. How does one recognize sarcasm? Tone of voice? Context? Known intent and bias of the speaker? So hurdle number one may simply be coming to a reproducible understanding and identification of the meaning, use and indicia of these concepts. A preliminary step might be taking a large corpus of literature and engaging on a herculean effort to flag sarcastic remarks, plus a similar effort in video clips from motion pictures or from candid clips from real life. Then the deep learning process could be loosed on the identified segments, as well as large bodies of non-sarcastic language usage to give the system a chance to extract markers of sarcasm, the same way most humans do as they mature. The results from this iteration could then be tested against human recognition of a new set of data, with further learning to follow on to the heels of the new revelations and adjustments to the learning algorithms. As humans also often make mistakes in recognizing sarcasm, *cf*. the character Sheldon in the T.V. show “The Big Bang Theory,” the ability of the system to recognize it may be over-harshly challenged.

Of the other Grand Challenges, three seem least likely to be uniquely resolvable via Quantum Computing technologies. Interaction with real or virtual world objects, smoothly merging video clips, and responding to multiple persons, both real and virtual, all seem to be issues for which the optimization functions of Quantum Annealing are not required. Not having experience with a general purpose Quantum Computer, the assessment of the applicability and benefits of that technology is more problematic, but the authors cannot envision it at this time. The other Grand Challenges do seem to be amenable to break-through advances via optimization analyses on a Quantum Annealer.

Quantum Computing’s promise of being able to better identify patterns of context, image content, and audio sonority may be a necessary element in resolving many of these challenges. The putative ability of the envisioned machines to analyzed vast quantities of data in the learning process does hold out hope that such machine discriminations can be made, at least to the level of not being disruptive of the creation of a life-like conversation.

Carrying this suggested approach over to the other nine challenges may seem this is a daunting task, for even alone the above seems daunting. Again, how daunting would it have been 25 years ago to view the effort that has gone into the implementation of the Internet? Few would have predicted the explosion of the code base, the infrastructure implementation, and the emergence of titans like Amazon, Google, and Facebook. If the benefits of quantum annealing, even if restricted to simulated annealing, become more apparent, the code base will be developed, the infrastructure will be installed. If, on the other hand, the technology does not bear fruit, the initiative will recede into the shadows with so many that have faded before it. The issues remain in flux and the questions will be in the technical discourse for some time.

5. Conclusions

There are uses for Virtual Humans and these uses will increase. Virtual Humans will become increasingly “human.” There are certain aspects of this humanness that are desirable in general and necessary in certain circumstances and for certain uses, which are very challenging and meet the definition of Grand Challenges. The emergence of theoretical and software approaches like deep learning, may be enabled by system and hardware advances like quantum computing. The rate of acceptance and the speed of efficacious implementation is still in question. The simulation community would be well advised to keep careful watches on both technologies. The potential uses can be envisioned long before they are enabled. When the time is ripe, a careful analysis must be made of the costs of implementing any new technologies and a critical thought process needs to be engaged to ensure the timely, but not premature use of these new tools. Further, system engineering advances in sustainability and the proper use of standards will be required to fully exploit these new technologies. The past has shown that a thoughtful and early application in these disciplines will pay significant dividends in the future.

6. Acknowledgements

The authors would like to thank all of their simulation colleagues who ventured their opinions on the Grand Challenges of Virtual Humans, but they want most especially thank Dr. William Swartout, of the Institute for Creative Technologies at the University of Southern California for his counsel, insights and support. Also, Dr. Robert Lucas, the Operational Director of the USC-Lockheed Martine Quantum Computing Center was very helpful and supportive, recommending some important conceptual issues relating to Quantum Computing, which were considered above.

7. References

[1] Chowdhury, G. G. (2003). “Natural language processing.” Annual review of information science and technology, 37(1), 51-89.

[2] Collobert, R., & Weston, J. (2008, July). “A unified architecture for natural language processing: Deep neural networks with multitask learning.” In Proceedings of the 25th international conference on Machine learning (pp. 160-167). ACM.

[3] Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., & Piepol, D. (2007, November). “Building Interactive Virtual Humans for Training Environments.” In Proceedings of the Interservice/Industry Simulation, Training and Education Conference, Orlando, Florida, 2007 (Vol. 174).

[4] Traum, D., & Rickel, J. (2002, July). “Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds.” In Proceedings of the first International Joint Conference on Autonomous Agents and Multi-agent Systems: part 2 (pp. 766-773). ACM.

[5] Allwood, J., Traum, D., & Jokinen, K. (2000). “Cooperation, Dialogue and Ethics.” International Journal of Human-Computer Studies, 53(6), 871-914.

[6] Rizzo, A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., ... & Parsons, T. (2011). “SimCoach: an Intelligent Virtual Human System for Providing Healthcare Information and Support.” International Journal on Disability and Human Development, 10(4), 277-281.

[7] Swartout, W. R., Gratch, J., Hill Jr., R W., Hovy, E., Marsella, S., Rickel, J., & Traum, D. (2006). “Toward Virtual Humans.” AI Magazine, 27(2), 96.

[8] Kudryavtsev, Andrey. "Automatic Speech Recognition Services Comparison." Grid Designer's Blog. N.p., 11 Jan. 2016. Web. 20 Apr. 2017. Available: <http://blog-archive.griddynamics.com/2016/01/automatic-speech-recognition-services.html>.

[9] Lee, K. F. (1988). “Automatic speech recognition: the development of the SPHINX system” (Vol. 62). Springer Science & Business Media.

[10] Satori, H, Harti, M., & Chenfour, N. (2007). “Introduction to Arabic Speech Recognition using CMU Sphinx System.” arXiv preprint arXiv:0704.2083.

[11] Swartout, W. R., Nye, B. D., Hartholt, A., Reilly, A., Graesser, A. C., VanLehn, K., ... & Wang, L. (2016, March). “Designing a Personal Assistant for Life-Long Learning (PAL3).” In The Twenty-Ninth International Flairs Conference.

[12] Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., Jungblut, K., Leuski, A., Smith, S., & Jungblut, K. (2015, November). “New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor’s Interactive Storytelling.” In International Conference on Interactive Digital Storytelling (pp. 269-281). “Springer International Publishing.

[13] Rizzo, A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., ... & Parsons, T. (2011). “SimCoach: an Intelligent Virtual Human System for Providing Healthcare Information and Support.” International Journal on Disability and Human Development, 10(4), 277-281.

[14] R. Feynman, , (1982), "Simulating Physics with Computers", International Journal of Theoretical Physics 21 (6–7): 467–488.

[15] Lucas, Robert F., Gene Wagenbreth, John J. Tran, David R. Pratt, and Dan M. Davis. (2013) "Practical Adiabatic Quantum Computing: Implications for the Simulation Community." in the Proceedings of the Interservice/Industry Simulation, Training and Education Conference, Orlando, Florida, 2013

[16] Gershenfeld, Neil, and Isaac L. Chuang. (1998) "Quantum computing with molecules." Scientific American 278.6 (1998): 66-71

[17] Christensen, Clayton. (2013). “The Innovator's Dilemma: when new Technologies Cause Great Firms to Fail”. Harvard Business Review Press, 2013.

[18] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep learning." Nature 521.7553: 436-444.

[19] . Bernstein, Jeremy, (1981), “A. I.”. New Yorker Magazine, New York, New York,

[20] Hinton, Geoffrey E. "Training Products of Experts by Minimizing Contrastive Divergence." Neural computation 14.8 (2002): 1771-1800.

[21] Krizhevsky, Alex, and Geoffrey E. Hinton. "Factored 3-way Restricted Boltzmann Machines for Modeling Natural Images." International conference on artificial intelligence and statistics. 2010.

[22] Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in Neural Information Processing Systems. 2015.

[23] Boixo, Sergio, T. F., Isakov, S. V., Wang, Z., Wecker, D., Lidar, D. A., ... & Troyer, M. “Evidence for Quantum Annealing with More than One Hundred Qubits.” Nature Physics 10.3 (2014): 218-224.

[24] Cipra, Barry A. ”The Ising model is NP-complete.” SIAM News 33.6 (2000): 1-3.

[25] Potok, Thomas E., Catherine D. Schuman, Steven R. Young, Robert M. Patton, Federico Spedalieri, Jeremy Liu, Ke-Thia Yao, Garrett Rose, and Gangotree Chakma "A study of Complex Deep Learning Networks on high Performance, Neuromorphic, and Quantum Computers." Proceedings of the Workshop on Machine Learning in High Performance Computing Environments. IEEE Press, 2016.

Authors’ Biographies

Ke-Thia Yao is a Project Leader and Research Scientist in the Computational Systems and Technology Division of the University of Southern California (USC) Information Sciences Institute (ISI). His primary research interest is in understanding large complex systems and data sets. He teaches data management classes at USC’s Viterbi School of Engineering. He is part of the team investigating uses for USC’s 2,000 Qubit Quantum Annealer. Dr, Yao was one of the leads on the JESPP project where he analyzed the computational and behavioral properties of large-scale simulations. That project had the goal of supporting very large-scale distributed military simulation involving millions of virtual and constructive entities. He received his B.S. degree in EECS from UC Berkeley, and his M.S. and Ph.D. degrees in Computer Science from Rutgers University.

Dan M. Davis is a consultant for the Institute for Creative Technologies at USC, focusing on large-scale distributed DoD simulations and Virtual Humans. He was the Director of the JESPP project at JFCOM for a decade. As the Assistant Director of the Center for Advanced Computing Research at Caltech, he managed Synthetic Forces Express, bringing HPC to DoD simulations. Prior experience includes serving as a Director at the Maui High Performance Computing Center, at JPL and in industry. As early as 1971, Dan was writing programs in FORTRAN on one of Seymour Cray’s CDC 6500’s. He saw duty in Vietnam as a USMC Cryptologist and retired as a Commander, U.S.N.R. He received B.A. and J.D. degrees from the University of Colorado in Boulder.

Jeremy J. Liu is a fourth year PhD student at the Viterbi School of Engineering at the University of Southern California. His research interests are quantum computing and deep learning, especially using the former to enhance the latter. He has conducted research into Oil Field Management and has authored papers on that topic. He received a B.S degree in Information and Computer Science from the University of Pennsylvania.

Nicholas Kaimakis is active in research at the Institute for Creative Technologies of the University of Southern California. His current research thrusts are in the use of computer generated avatars or video clips, animated and directed by natural language-optimized Artificial Intelligence (A/I) programs that present a life-like dialogue capability to interact with remote users via the internet. His current project is funded by the Navy and is designed to help improve knowledge of STEM fields. He is an undergraduate student studying Computer Science within the Viterbi School of Engineering at the University of Southern California.