Raising Artificial Intelligence

By: Mark Cummings, Ph.D., Bill Yeack, Joel Holland

Artificial Intelligence (AI) is currently experiencing a growth spurt. Just as when children go through growth spurts, it is helpful to be able to understand what is happening in the context of the overall development process. And, just like in a teenager, the growth spurt brings significant benefits but also limitations and dangers. Understanding this is key to being able to make the right decisions. So, if someone tells you that they have the ultimate AI that will solve all your problems, ask them the following questions: How do you fit in AI's evolutionary path? What is the cycle time or bias for training relative to the cycle time of the underlying problem? How do you handle the data scale problem?

AI's Evolutionary Path

One senior person in the field says that there is a new academic paper on AI published every minute. Considering that pace, there are many ways to look at AI's evolution. In our research, we have found that this paper by Steve Sovolos is helpful in explaining the progression. As shown in Figure 1, there are three bands of AI evolution. The first band is Artificial Narrow Intelligence (ANI), or Band 1. It is the kind of artificial intelligence that is good at performing single tasks, such as playing chess or making predictions and suggestions. In the telco world, this kind of AI has been used to predict when certain types of equipment will fail. That allows operators to proactively replace equipment and avoid network interruptions. ANI is the only level of AI achieved so far. And, as we will see, it's a level not yet fully attained, which shows how far we have yet to go.

Artificial General Intelligence (AGI) is the kind of artificial intelligence that can perform human-level intelligence functions and is Band 2. Artificial Super Intelligence (ASI) is the type that is smarter than the collective intellect of the smartest humans in every field. Smart as it is, however, it still needs human input of some kind in order to be deployed—and to work. ASI is Band 3. Anything beyond ASI is all speculative.

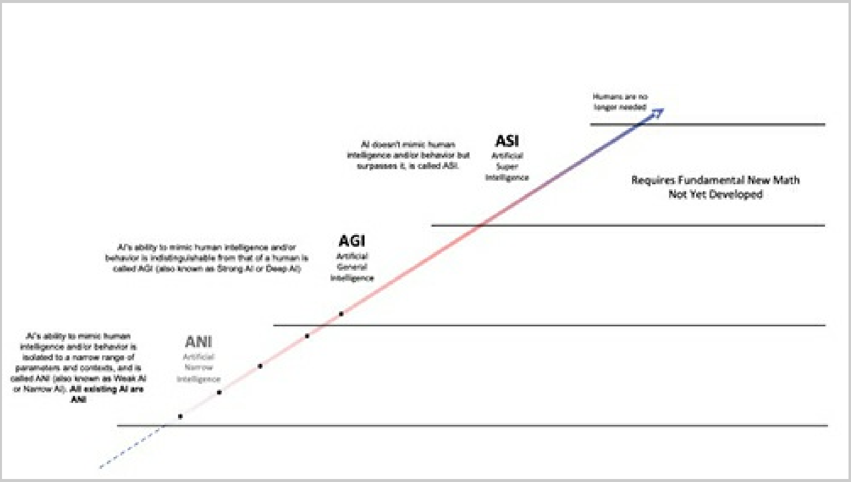
Humans are no longer needed
AI doesn't mimic human intelligence and/or behavior but surpasses it, is called ASI.
ASI
Artificial Super Intelligence
Requires Fundamental New Math Not Yet Developed
AI's ability to mimic human intelligence and/or behavior is indistinguishable from that of a human is called AGI (also known as Strong AI or Deep AI)
AGI
Artificial General Intelligence
AI's ability to mimic human intelligence and/or behavior is isolated to a narrow range of parameters and contexts, and is called ANI (also known as Weak AI or Narrow AI). All existing AI are ANI
ANI
Artificial Narrow Intelligence

Figure 1 - AI Evolution Band

In addition, there was a Band 0, which was the precursor to all of these bands. Band 0 represents the Marvin Minsky–David Horowitz era, in which most of the foundation issues were discovered. This was the era of LISP and other approaches, which were revolutionary but by today's standards seem elementary. The big growth spurt is what is getting us from Band 0 into the beginning of Band 1 (ANI). To understand this, as well as its promise and limitations, we have to break Band 1 down.

Figure 2. shows such a breakdown of Band 0 into four generations. The generations are distinguished by:

- Purpose, such as whether the question the system is seeking is known at the beginning of the process;
- Characteristics of the data the system is using, such as whether the data source is known.

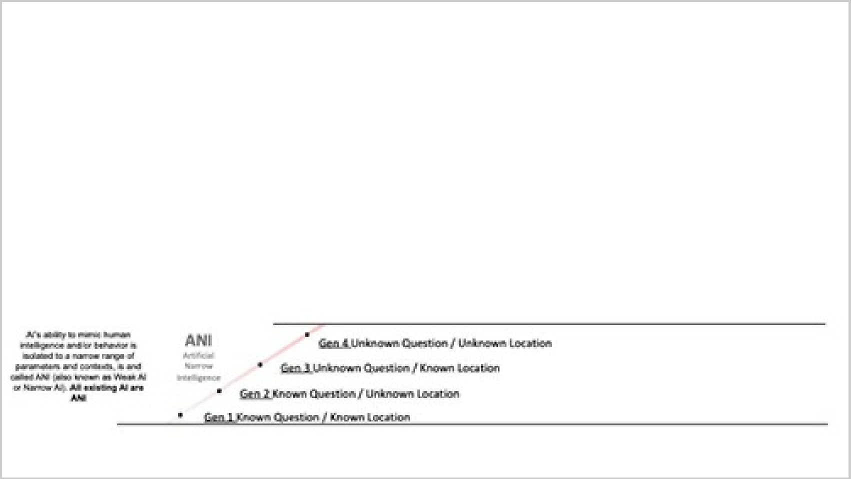
AI's ability to mimic human intelligence and/or behavior is isolated to a narrow range of parameters and contexts, is and called ANI (also known as Weak AI or Narrow AI). All existing AI are ANI
ANI
Artificial Narrow Intelligence
Gen 4 Unknown Question / Unknown Location
Gen 3 Unknown Question / Known Location
Gen 2 Known Question / Unknown Location
Gen 1 Known Question / Known Location

Figure 2 - Breakdown of Band 1

Each of these two elements can be known or unknown. In Generation 1, the system knows both the question and the source of the data (typically a specific data lake). This combination is then an ANI Gen1 solution set. An example of such a system is one that seeks to answer the question, "Is someone trying to attack my SQL Database?" SQL D/B attacks show a specific pattern in SQL D/B log files. In this example, the AI system knows what it is looking for (the pattern) and knows where to look (the SQL D/B data lake that stores all of the log files). The SQL D/B may support Communications Service Providers' (CSP's) infrastructures. For example, the OSS (Operations Support Systems) and BSS (Business Support Systems) that were recently attacked in European CSPs. Here again, today's systems are primarily Gen 1. So, we have three more generations in ANI before we even get to the beginning of the Band 2.

The generations proceed. In Generation 2, the initial system knows the question but not the location of the data to analyze to get to an answer. Extending the example above, the initial system knows the pattern it is looking for but does not know where the data exists to allow it to find that pattern. To accomplish its purpose, the system must "reason" and find the location of the data it needs.

In Generation 3, the initial system does not know the question but does know the location of the data to analyze to get to an answer. Again, extending the above example, the initial system may have a general purpose, such as to maintain the integrity of another system, but it doesn't know anything about potential attacks, failures, and so on; or their patterns. It does know that it has access to a data lake containing the target system's log files.

In Generation 4, the initial system knows neither its question nor the location of the data. This begins to blur into the beginning of the AGI Band. The best example of such a system might be the very early research work being done on LEELA. In this example, the initial system has been developed based on the human development theories of Piaget to have some basic cognitive abilities independent of questions (purpose) and data source. Work in this space is very preliminary, so it is difficult to say anything with certainty. But it appears that although Generation 4 could be very powerful, it probably requires massive amounts of data and long training cycles.

Training Bias?

We have talked about the data as a single thing. But there are (at least in Generation 1 as it exists today) two types of data: the data the system is “trained” on and the data the system uses in production. This concept of training is where the term Machine Learning (ML) gains importance. There are other valuable and important kinds of ML, but we will only talk about the one as it is used in Deep Neural Networks (DNN) here. One of the leading AI experts and the architect of a very significant AI microprocessor (commonly referred to as a TPU) said in a private workshop that we actually don’t know exactly how or why this kind of machine learning works, but it does. In these systems, a Deep Neural Network (DNN) is fed a “training data set.” This training data set is a sample of the real world and is subject to all the statistical problems with samples. Based on this data set, the DNN is trained to find certain patterns (the question). The resulting system is then pointed at “real world” data sets.

The statistical sampling and bias problems we skipped over lightly above can be serious. Let’s consider this example. A facial recognition system designed to recognize felons was applied to the members of the US House of Representatives. This system identified the majority of the House’s dark-skinned members as felons. Alternately, consider the autonomous vehicle system in Phoenix that didn’t recognize a homeless person as a person, but rather considered her as insignificant debris and killed her. If such a facial recognition system is employed by a CSP to control access to sensitive facilities, these kinds of problems can be very significant.

Supervised training can overcome some of these problems. But supervised training is expensive, takes a long time, and there are not enough people who can do it effectively to support all the applications desired. These facts bring us to unsupervised training—the machine learning discussed above.

Cycle Time?

In addition to the statistical sampling and bias problems, and even with an advanced TPU multi-processor system and very experienced engineers, it can take 15 to 20 days to train the solution with the training data set. In many application areas, the underlying problems (questions) are changing faster than the training period. For example, in dynamic industries like security threat hunting, the threats are evolving at an alarming rate. So, with a 15-day training period, the basis of the trained solution set is invalidated even before it goes live.

It is important to point out that the actual run time of training is not the only factor to be considered in cycle time. Significant time is involved in deciding and planning. Then there is more time involved in developing and acquiring the training data set. Finally, staff and machine resources are not infinite, so specific training has to be scheduled and may have to wait its turn. If there is not a problem acquiring a satisfactory training data set, the planning and development can double or triple the cycle time. If there is a problem obtaining an unbiased training data set, or there is a staff or machine bottleneck, adding those delays to the cycle time can also produce a dramatic increase.

Data Scale?

Data scale is important because when the data sets grow beyond critical points, processing time grows dramatically. For example, a typical security service within a CSP can collect 20+ BEPD (billions of events per day). Looking through a month’s worth of actual data will equal over 600 billion events. One query with a conventional database this size could take over 10 days. Thus, the data scale problem impacts both training and run time.

Statistical theory tells us that as we increase the sample size, we decrease the probability of bias. Therefore, one way to address the training data set bias problem is to use a larger training data set. Another approach is to use different training data sets for different contexts. For example, using one training data set for a midwestern US population, a different one for a Chinese population and a third for a Nigerian population offers the advantage of being able to match the training sets to the different population characteristics. Each of the training sets can be refreshed based on the different problem change rates in the different populations. Both approaches add to the data scale problem. Similarly, having larger problem data sets to work with increases the probability of achieving correct results. Again, this increases the time it takes to achieve a result.

Work is ongoing to solve the data scale problem that will be discussed in a future article, but—going back to the beginning of this article—for now, just remember to ask the questions.

Conclusion

AI is going through a growth spurt. This spurt is having a very significant positive impact in many areas, but it also has limitations and dangers. Understanding both the promise and the cautions is key to making good decisions around AI. So, if someone tells you that they have the ultimate AI that will solve all your problems, remember to ask them the three key questions we have discussed: Where are you in AI’s evolutionary path? What is the cycle time and bias for training relative to the cycle time of the underlying problems? How do you handle the data scale problem?