Artificial Intelligence
Who is responsible? And who cares?
Compsim White Paper

In the past, when someone built machines for the market, they first learned what the market wanted and then they created a solution. The manufacturer usually had a quality control function to validate the solution. The manufacturer was responsible for delivering a quality solution. Sometimes the manufacturer would go through a third-party testing organization for an external validation. If the machine was computer-controlled, the machine software would likely go through code certification in addition to the application certification process. Again, the manufacturer was responsible for the performance and safety of the system.

With Machine Learning the behavior of the solution is at least somewhat being delegated to the machine. The assumption is that the machine can learn what the manufacturers did not want to extend the effort to learn by themselves. And, with current laws, it appears that manufacturers may be absolved of any responsibility by adding small print to the contract that states the user is assuming some responsibility. There are now cases when the capabilities of systems can be misused, such as creating deep fakes[1] and stealing proprietary information[2]. If there are some flaws in the solution, who cares? Adults (and children) are learning how to use Generative AI to write their papers. If they get caught, who cares? Public speakers can use Generative AI to prepare their presentations now. It is easier than doing the work themselves. Who cares if there is stolen information, or sometimes incorrect information? Humanity is operating in information overload, and it now appears to accept that some people lie all the time. Misinformation, disinformation, and lies drive daily life, so why not machines?

The case that misinformation and lies can be mass-produced at automated speeds may not be a concern that is widely expressed. Nor is the case that the machines may be used in safety-critical applications where the cost of error is high.

We are entering a time when fully autonomous machines will be fighting our wars: machines with weapons. These machines will be (are) making decisions about who lives and who dies. Self-driving cars are also making these life-and-death decisions, and are being mass-produced. Will these machines "learn" who to kill? Will they be able to explain their behavior? Will their human owners accept a response from the machine that "it seemed like the right thing to do at the time?" Or will humans just stop asking or searching for an alternative solution?

---

[1] Deep Fakes:
https://search.yahoo.com/search?fr=mcafee&type=E210US978G91801&p=describe+%22deep+fakes%22
[2] Generative AI stealing proprietary information:
https://search.yahoo.com/search?fr=mcafee&type=E210US978G91801&p=Generative+AI+stealing+proprietary+information

We are living in a "data-driven economy"[3]. Robotic Process Automation (RPA)[4] is marketing as a new area for enterprise automation and investment. This might be restated as there is too much data to comprehend by humans, so it is best to turn responsibilities over to machines that can interpret the data and execute the day-to-day business functions. It is likely that humans are accepting that these systems are performing as desired because they trust their tool suppliers and they trust their data sources without fully understanding everything inside the process. Again, the systems are too complex for humans to comprehend.

Beyond conspiracy theories that are based purely on speculation, since the data-driven economy is based on data, there likely are (or will be) those attempting to corrupt the data for their own purposes; or just because it is an interesting challenge. The result will be that inappropriate decisions or actions will be taken by the enterprise and the battlespace.

When machines (or humans) "learn", they are being taught how to interpret information (data). For humans, we may hope that they develop a "value system" throughout their lives that helps them interpret the information. We hope they do not lie or attempt to deceive others. But we can observe that the value system developed by humans evolves differently for each human. They do not all behave the same. Some do lie. Every human develops their own value system that ends up providing bias in every decision and action they make.

It is relatively easy to see that there are two parts to the decision-making process. At the broadest level, there is the data, and there is the processing of the data. The data provides the influencing factors that drive the decisions and actions. The processing of the data defines how the decisions and actions are derived from the data.

When adult humans learn (or are taught) how to perform a task they never start at ground zero. They have been educated, and they developed some basic skills. They play and learn how to interact with others throughout their life. Perhaps they have some religious training. Parents will have provided some guidance. They will have developed some language skills. They may have gained life lessons about how to work with others in a shared civilization. They may have been punished when they do inappropriate things, and applauded when they perform as desired. One might suggest that as they progress through life, humans develop their value system. To accomplish this, focused training attempts to add to that accumulated body of knowledge with training on how to perform specific tasks (what to do, how, when, how much, and where).

---

[3] Data Driven Economy: https://search.yahoo.com/search?fr=mcafee&type=E210US978G91801&p=Define+a+Data+Driven+economy%3F

[4] Robotic Process Automation aka Software Robotics: https://search.yahoo.com/search?fr=mcafee&type=E210US978G91801&p=Define+Robotic+Process+Automation

Compare this to Machine Learning where the machine is exposed to numerous "patterns" on which to learn how to address the problem areas. There is no separate value system. There is just a collection of patterns, where the machine will interpolate between those taught patterns for the closest match at any instant.

Now back to the machine-building mentality. A machine builder, in the Industrial Revolution timeframe, built machines to perform desired functions. Because the machine builder was in total control of the design and functionality, the machine builder had full control of how the inputs controlled the outputs. If the machine broke or performed unexpectedly, the machine builder could trace through the design, identify the problem, and address it. As machines grew more complex, the machine builder built mechanisms into the machines that would assist in the process. But since most machines were constructed with the simple process where inputs drove the outputs, all you had to do was trace through the design from inputs to outputs to fix any issues. Some processes for the machines allowed for tuning of the processes. This would allow users of the machines to adjust parameters through some type of operator interface. This could be stated as allowing the user to adjust the "value system" of the machine in the performance of its function.

Adding Intelligence to Machines

For many years almost every industry has been talking about intelligent machines and smart "everything". Smart machines, smart sensors, smart vehicles, smart buildings, smart cities… Mainly this has meant that microcontrollers were being integrated into devices and using computers made them smart.

With the topic of Artificial Intelligence, we are extending the machines from following explicit rules into areas where humans apply judgment and reasoning skills to address more complex problems. Dr. Horst Rittel (UC Berkley) coined the phrase "Wicked Problems" in the 1970s[5]. He was a city planner and stated that for "Tame Problems" you could write a formula (a set of rules) to get a "correct answer". But for Wicked Problems, if you collected all the information about your problem space, all you could hope for was a "best solution". The suggestion was that it would not be economically feasible to attempt to write a formula to run a city and get a correct answer to handle all situations affecting the city. In his turf, this meant if a city planner knew everything about a city, and how it functioned, a skilled city planner could run a city the best way using means available to that city planner at the time. The skilled city planner would utilize his/her "skilled" value system in the process.

The Value System is Key

We are highlighting that the value system is a key component of intelligent behavior. There are pieces of information AND values or importance levels of each piece of

[5] Horst Rittel / Wicked Problems: https://en.wikipedia.org/wiki/Horst_Rittel

information when considering every scenario. We are also suggesting that the importance of information can change. The easiest way to understand this is by describing a situation using temporal factors (time and space). A threat that is farther away in distance or time may be less important than the same threat that is close by in space or time. A big threat may be more important than a smaller threat. But at the same time, depending on the scenario, a big threat that is further away may or may not be less important than a smaller threat that is closer. It is all about the value system. Adding to the complexity is the accumulated stress of different parts of the problem set. In many complex problems where humans apply their judgment and reasoning skills, there are competing problems that must be addressed together. Stated differently, it will likely not be possible to satisfy all needs as there is never enough time or resources to do everything the best way. A compromise is required and procrastination is not acceptable.

We would suggest that it would be hypothetically impossible to train a machine by showing it patterns of every situation where every influencing factor is changing in importance continuously. And consider that in temporal areas change is continuous. It is not going from state to state. Stress based on temporal factors is continuous.

Building an Intelligent Machine

Rather than turning responsibility over to a machine that learns on its own how to behave, we suggest the machine builder should retain that responsibility.

To accomplish this task, we suggest the machine builder instructs the machine on how to think. By this, we mean that the machine builder gives the machine a value system as part of telling the machine how to think. To accomplish this task, we are suggesting the use of a new "technology": a new way to deliver Artificial Intelligence.

Knowledge Enhanced Electronic Logic (KEEL®) Technology was developed to satisfy this need[6]. KEEL Technology allows a machine builder (domain expert) to tell the machine how to think (integrate valued influencing factors that are used to address one or more competing valued problems) and behave. And, by "behave" we are highlighting that behavior is both point decision-making AND adaptive operational control. Driving a vehicle is an example of operational control where the driver is continuously correcting for temporal factors, like curves in the road ahead, signage, changes in surface conditions, changes in visibility, and moving obstacles at different distances and headings.

There are two parts to KEEL Technology:

1. The KEEL Dynamic Graphical Language (DGL) makes it relatively easy to define complex judgment and reasoning skills that can be executed by a machine. In the process, the operating model is given a "visual value system".
2. The auto-coded KEEL Cognitive Engine that processes information as if it was

[6] About KEEL Technology: https://www.compsim.com/About_KEEL.htm

processed on an analog computer, yet runs on today's off-the-shelf microcontrollers.

The DGL with a methodology called "Language Animation" that allows one to watch the system "think" in almost real-time. This methodology supports the need for 100% explainable and auditable AI.

For the machine/system builder, this allows the supplier of solutions based on KEEL Technology to create safe systems and be responsible for those designs just like the machine builders of old that created machines to do the jobs they were designed to perform.

The designer can add as much flexibility into the designs as demanded by the customer. Often this means that the customer can adjust the "value system" to control the importance of different influencing factors.

By "telling the machine how to think" rather than how to address specific problems, the systems can address problems they have never encountered before. The machines will be able to handle multiple problems collectively, just like humans. Within the KEEL Cognitive Engine, the machine can balance interrelated problems, even when they have conflicting attributes, like demands for the same resources in order to achieve the best overall set of solutions. KEEL Technology along with other supporting tools allows a detailed examination of the decision-making process. It may be obvious to some that when complex problems are delegated to machines there will still be a need for continuous audit and refinement. This is especially true as we enter the domain of adversarial computing[7], where adversaries will constantly be incorporating new data sources, new tactics, and new weapons and will be probing for weaknesses in existing systems.

Who cares? Recipients of AI behavior care:

The recipients of inappropriate results will care. If a self-driving car decides that you die, you may care. If your weapon system decides that you die, you may care. If your weapon system decides to kill a large number of non-combatants, you may care. If you cannot explain the behavior of your machines beyond "it seemed to the machine that it was the right thing to do at the time, because that is what it did", then lots of people may care. If you are promoting that your systems are safe, ethical, unbiased, and trustworthy, your followers may care. And these are just life and death decisions. When consumers of your expert decisions and actions find out that your solutions cannot be explained and therefore may be flawed, they will find other suppliers.

Concerns have been expressed regarding Machine Learning systems that have potential

[7] Adversarial Computing: https://www.compsim.com/PublicPapers/Adversarial%20Systems-AdversarialComputing.pdf

hidden biases built into their solutions. The same concerns have been stated relative to the ethical behavior of AI systems. These concerns can be summarized as the need for Explainable AI. Some quotes from the past may summarize the concerns:

> "If you can't explain it simply, you don't understand it well enough." Albert Einstein.
> "If you cannot measure it, you cannot improve it." Lord Kelvin.

Note that these quotes did not even consider that decisions and actions made by autonomous systems will be mass-producing those decisions and actions.

Without a clear and detailed explanation of AI behaviors, it will be impossible to understand the consequences of mass-producing systems that may sometimes exhibit unexpected, undesirable behavior that may have significant global consequences.

Here is another important question: Is it ethical to deliver any form of artificial intelligence that is not easily 100% explainable and auditable?

Who cares? Suppliers care:

On the delivery side, suppliers of AI-based solutions that want to be recognized as purveyors of quality solutions also care about controlling costs and emphasizing development efficiencies. They may have the opportunity to become a recognized center of excellence for the development of cognitive technology in their domain because they retain that skill set. They become the expert in the field, rather than only using a set of training patterns. They can describe their expert solutions better than any human can because they can provide a mathematically explicit explanation (without focusing on the math). This means they recognize the value of KEEL Technology that:

1. Is easy to learn (novice to productive in weeks),
2. Can be audited and explained (provided at no additional cost),
3. Is easily fixable (hours from detection to auto-coded production solution),
4. Is application-platform-architecture independent (auto-coded solutions in many languages for different development environments),
5. Is a component technology with a simple API (easily insertable into new or existing applications),
6. Has an extremely small memory footprint (suitable for small devices and sub-assemblies),
7. Can be "designed-once-deployed-many" where a single design can be deployed in different programming languages with the click of a mouse (M&S, emulation, different production platforms),
8. Has no dependencies on hardware or external software libraries (saving costs, complexity, and memory).

Summary:

Both consumers and suppliers care about the quality of the AI-based delivered behavior. When users find they have a choice between fully explainable solutions and those that are not, especially when the explainable solutions might even be less expensive to support, some might think the preferred approach is clear. At the same time, suppliers that have invested heavily in unexplainable AI approaches will resist as long as they can.

Tom Keeley
President Compsim
https://www.linkedin.com/in/tomkeeley/