Compsim™ Expert Decisions by Machines

KEEL®
Knowledge
Enhanced
Electronic
Logic

What should you expect/demand

when you ask a machine for expert advice?

What should you expect/demand

when you are trusting the machine with your life?

We (humans) are exposed to more information every day. It is impossible to keep up. There are always other humans wanting to take up the role of thinking for you, but we know that humans have their own bias, their own ethics, their own understanding of right and wrong. And organizations are always looking for ways to reduce cost and AI Agents may provide a path to the future. And with many organizations jumping on the AI train they must know what they are doing. Right?

Historically humans have built machines to operate under their control to increase speed, increase strength, increase precision, be always on and ready, or do more with less. Mass-produced commoditized machines are part of everyday life. Yes, sometimes they break, but for the most part, humans remained in control of the machines by giving them a strict set of rules. Humans have been responsible for building good machines. Machine producers have tested their machines before releasing them to production. Consumers have had confidence that the machines would perform as advertised.

Now, however, the rules are gone, and the AI machines are being trained on patterns of data. The machines are expected to behave in the most appropriate way based on the patterns on which they have been trained. The organizations providing services based on the trained datasets are “hoping” that the output of the machines will satisfy their clients. Humans receiving the expertise from these systems are “hoping” they are receiving the best answers to their needs. Some humans are “expecting” that no organization would release a solution based on “hope”, so surely the machines can be trusted. There is too much going on for some, that just accept that it is ok for someone else to do all the thinking for them. In the AI space, there is concern that unacceptable bias may be applied to decisions and actions. Concerns about Ethical AI, Safe AI, and Trustable AI in the ML/LLM/GenAI space are being raised. One might suggest that all the issues can be addressed with 100% Explainable and Auditable AI. If the AI can “easily” be reviewed and explained, then it can be trusted. And if it is easy, there is a good chance it will be done.

Reasonable Expectations:

It may seem reasonable that any machine delivering expert decisions and actions should be able to:

1. **Simply "List" ALL the decisions and actions that could be made by the system.**
    a. Do something / Don't do something
    b. Choose from one or more mutually-exclusive options
    c. Allocate resources
    d. Prioritize actions

   By explaining what the machine "could do" one might infer what it "could not do". If your life depended on decisions and actions of the machine, one might want to know what it could not/would not do.

2. **Explain any decision or action by listing all the influencing factors contributing to decisions or actions.**

   Some decisions and actions require the sharing of resources, so it is important to know which decisions and actions are sharing resources.

   By listing all the influencing factors one can get confidence that the system has considered the right factors. It also leaves open the opportunity that the system could be reviewed by outside groups that can provide an additional level of confidence in the system.

3. **Explain the values assigned to all factors used in the decisions or actions.**

   It is the perceived importance or value of influencing factors as well as importance assigned to possible decisions and actions that allows one to apply judgment and reasoning. We expect an expert to understand the importance of influencing factors. If decisions and actions are allocated to machines, we should expect that the machines can explain the values of influencing factors that lead to decisions and actions. Machines work on explicit values, so it should be easy for machines to expose how information is valued.

4. **Explain how all Influencing factors are combined to make decisions or control actions.**

   We trust that human experts combine the reasons for decisions and actions appropriately. With machines, however, we might demand their decisions and actions are explained by exposing exactly how the influencing factors are combined to deliver the decisions and actions.

By explaining how information is integrated, one could review the reasoning model to understand how thoroughly the machine was considering the situation and open the door for external parties to review the process.

5. **Explain why other decisions and actions were not taken.**

   There may be many cases where one is not only looking at why certain decisions and actions were taken, but also investigating why other decisions and actions were not taken. The concept of a "second opinion" is often used by humans when there is concern that the first opinion may have been erroneous.

   With a machine, it should be just as easy to review why a decision or action was not made by the machine, as it was to explain the selected decision or action.

6. **When the system is considering multiple problems together, where resources are spread across multiple problems, it will be important to explain how resources are being allocated and problems prioritized.**

   While seldom considered in today's conventional AI discussions, more intelligent systems will be considering multiple problems together: pursuing goals, responding to new threats and new opportunities, responding to depleted resources, etc.

   It will be very important to understand why autonomous systems behave the way they do. Policies will need to be changed rapidly to keep up with change. If policies cannot easily be explained, they cannot easily be extended and fixed. It this is not provided, then mass-produced chaos will occur.

Decision-making by Machines

Much of the AI today is focusing on human-related communications, where Machine Learning/Large Language Model/Generative AI are based primarily on the human language that has been captured and used for pattern training purposes. But if one looks at the life of digital computers and microcontrollers, many are used for controlling the behavior of machines rather than generating and manipulating text. They are used in controlling games, toys, modeling and simulation components, micro-controlled edge devices, subassemblies of vehicles, power distribution system, etc.

According to a Gemini query, "it is safe to conclude that well over 50% of the computer and microcontroller market is focused on applications that do not primarily involve processing human-generated text and voice."[1]

It is easy to suggest that these computerized devices and software applications will benefit from human-like judgment and reasoning as if a human expert was embedded in every component.

It is also easy to suggest that many of these mass-produced devices will be performing safety-critical functions. Delivering systems that can hallucinate at times and perform randomly at times should be discouraged.

If humans are to remain in charge of their own future, at least those that have not already given up, they will demand systems that can respond to the **Reasonable Expectations** above.

This is especially true, because all expectations can be met with technology available today[2].

[1] Gemini query: "What percentage of the computer/microcontroller market is not focused on manipulating human generated information (text, voice)?"

[2] Compsim's Knowledge Enhanced Electronic Logic (KEEL®) "Technology" https://www.compsim.com/papers/What%20can%20one%20do%20with%20KEEL%20Technology.pdf