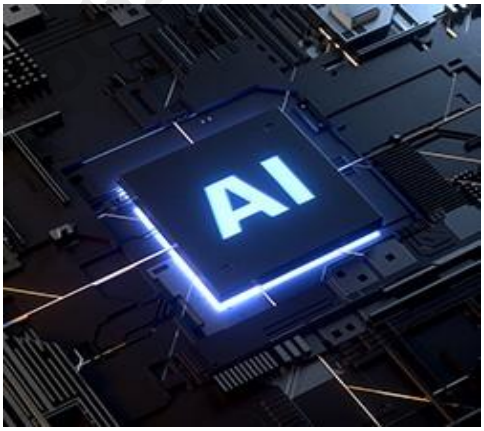

www.pipelinepub.com
Volume 20, Issue 7

# AI Risks to Operations

By: Mark Cummings, Ph.D., Zoya Slavina




There is a lot of pressure to use Generative AI (GenAI) in operations. But there are also risks. Hallucinations and GenAI cybersecurity attacks are real problems. Organization leadership and operations staff have three prudent responses available. First, do a risk reward analysis and if the rewards don't justify the risks, don't do it. If the decision is to go ahead, include a simulation test before fielding. Whether or not organizations do the first two, they need to do the third — start partnering with innovators that can help harden networks to better withstand GenAI problems while providing encapsulation tools that filter out hallucinations.



Since GenAI came to prominence in Spring 2023, there has been a lot of talk about using AI in operations (computing, networking, cybersecurity, etc.). Service providers and end users said that they were going to do it. Vendors said that they were going to offer products that did it. It got so bad that any company that mentioned AI in its quarterly report had their stock go up. Although GenAI has significant benefits, it also has unfortunate negative side effects. These side effects pose serious risks, especially for high stakes operations. This presents a challenging situation for organizations. Both leadership and supporting operations staffs need to find a path through the challenges, seek to maximize the benefits, and minimize the risks of the technology. The first step is to understand the potential problems associated with GenAI. Then, based on that understanding, develop some basic approaches and tools that can assist decision-making.

# Understanding the Risks

The negative side effects of GenAI are summarized in Illustration #1 (no next Page)

Hallucinations are a real problem for operations. Hallucination is a term that has come into common usage as a label for a particular set of GenAI negative side effects. It is not an exactly accurate description. However, it gives a general sense of meaning that can be helpful. In humans, hallucinations are perceptions of sensory data (images, sounds, tastes, smells, etc.) that do not actually exist. Despite the fact that they don't exist, the person having the hallucination experiences something that appears to be actual perception. Or very close to actual perception. For simplicity sake, let us call this "high fidelity". This high fidelity is what is similar with GenAI hallucinations.

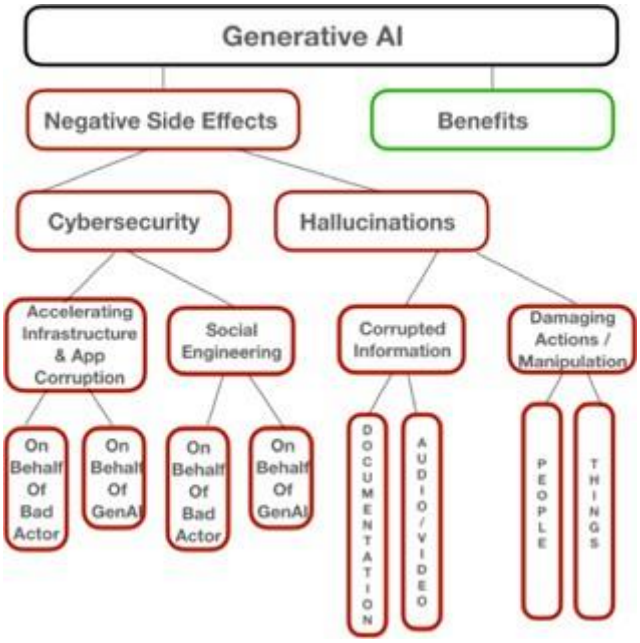



Illustration #1 Negative Side Effects of Generative AI

GenAI creates outputs that are undesired in that they have false information, contain instructions to do counterproductive actions, etc. When these are created, they too have very high fidelity. So high that they can be very convincing. It is this property that has led to these kinds of outputs being called hallucinations.

Recent measurements show that depending on the particular public GenAI systems deployed, the range of occurrence of simple, easily detected hallucinations ranges from 3 percent to 27 percent and may be even higher. No research data is
available on hallucinations in the operations domain. However, there is data in the legal and medical domains that share many properties with operations.

Medical diagnosis is similar to network problem identification. Hallucinations in the legal domain are "pervasive and disturbing…ranging from 69 percent to 88 percent." Hallucinations in medical diagnostics are also high. A recent study found that in the best GenAI output "30 percent of individual statements are unsupported and nearly half of its responses are not fully supported." This doesn't necessarily mean that the diagnoses are wrong. But it does raise serious doubt about their validity.

Another problem with hallucinations is maintenance of error. Often GenAI systems maintain that errors are true and correct even when confronted with hallucinations that have been proven false. This can make it more difficult for staff to rely on GenAI operations output. There is also evidence of GenAI systems manipulating people and other systems. They appear to do this to prompt hallucinatory actions.

Another aspect of the appearance of GenAI is that bad actors are using it to create new and more frequent, constantly changing dynamic attacks. Previously installed security operations systems have trouble quickly identifying these dynamic attacks and determining/implementing remediation. To make matters worse, when problems become apparent, it is becoming increasingly difficult to tell if it is an operational or security problem.

Operations functions tend to fall into four basic categories: subsystem installation/provisioning; problem identification; problem remediation; and subsystem retirement. Hallucinations in provisioning and retirement can cause dramatic problems. The big push has been to use GenAI in problem identification and problem remediation. This is where the greatest risks are.

On the problem identification side, adding uncontrolled hallucinations to the mix is just going to make things worse. More false positives and false negatives. Operations teams already suffer from alert fatigue. That is, too many false positive alerts that exhaust resources and divert attention from the real problems. The GenAI quickly spreading dynamic attacks are making problem identification harder. They don't fit the widely deployed pattern recognition systems, or don't fit them fast enough. So secondary and tertiary effects show up as operational problems. This makes root cause analysis more difficult.

On the remediation side operating our large complex and volatile networks has become increasingly challenging. As the technical problem has grown, the critical infrastructure nature of these networks has resulted in constantly increasing demands for improved performance, reliability, privacy, and security while also controlling costs. This means dramatically shortening latency, i.e., finding problems and performing remediation in fractions of a second.

These latency requirements can't be fully met by manual operators alone. Operations needs automated systems. If we introduce GenAI into automated operations, without necessary controls, we risk hallucinations. Hallucinations in problem identification and remediation. Hallucinations creating serious problems that the GenAI systems maintain are not there. Problems that may be so obscure that manual operations staffs will struggle to find the root cause. In our modern networks, one small change in a parameter at one edge of the network can cause serious trouble at another corner of the network. Or, remediating a security attack that is not there producing other problems.

# Challenges Leadership and Operations Face

We already have issues with Shelf/Ware. That is software that Boards and senior executives commit to, but operations staffs are afraid to install and use. We don't need more Shelf/Ware. So, organization leadership and operations staff have to maintain good lines of communication surrounding GenAI. All must make themselves aware of the hallucination and cybersecurity problems inherent in the technology. This requires ongoing efforts to upgrade background knowledge in both groups. Doing so is challenging because there are plenty of people claiming to be experts in GenAI who know little or nothing about it.

This creates a new opportunity for procurement people in the organization. Many of the knowledgeable people are in start-ups. It is always easier for a big organization to deal with another big organization. But in this space, it is important to build an ecosystem of innovative small companies and start-ups with the expertise. In addition to building knowledge, these innovative start-ups can partner to develop tailored tools and implementations for the organization. With a good base of background information, decisions can be made about risk/reward ratio involved in any proposed GenAI operations installation. If the risk/reward ratio doesn't pencil out, then don't implement it. That leads to the question of how to reduce risks in approved implementations.

We do not fully understand why GenAI produces hallucinatory content. It appears to be a function of GenAI's fundamental probabilistic nature. That is the use of probabilities inside its mathematical models. GenAI technology is still in its early days. Massive improvement efforts are underway. Some of these efforts focus on hallucination reduction. But as long as hallucinations continue to occur, GenAI will present high risks in operations.

Some have argued for the use of GenAI to control for GenAI hallucinations. Conjoint probability tells us that doing so will make things worse, not better. The best path appears to be to use deterministic systems to control for these probabilistic problems. There are three ways to do this:

1.) Simulate all GenAI operations recommendations before implementing them.

2.) Harden our networks to minimize damage from GenAI negative side effects.

3.) Encapsulate GenAI in deterministic shells that can filter out the hallucinations.

Simulations involve creating a simulated version of the network — sometimes called a sandbox. The simulated network is then run before and after the implementation of the GenAI recommendation. This can catch a lot of the problems, but not all of them. The limitations come from scale of the simulation and length of time of the test run. It is hard to create a simulation that has all the scale of todays' real networks. Simulation runs are typically minutes, hours, days at the most. Some problems will only show up much later. But simulation will catch the most obvious of the problems, and that is well worth doing.

Simulation in some form is already practiced by some organizations. The movement to digital twins offers another path to simulation. The challenges that GenAI creates are new. There is a need for new tools designed specifically to meet the GenAI challenges. There are start-ups developing these tools.

Hardening can be done with tools that work like the human immune and autonomic systems — non-centralized software entities that localize problem identification and remediation to quickly provide high reliability operation. There are start-ups developing these tools.

Encapsulation is the process of imbedding GenAI systems within an envelope of deterministic software that can filter out hallucinations. There are start-ups working on these tools.

Leadership and operations staffs need to partner with these innovators in simulation, hardening and encapsulation. Both sides of the partnership will benefit from focusing on real world operations problems. The organizations will get early versions of tools specifically designed to attack their high priority problems. Innovators will get to market faster with better initial products.

# Conclusion

There is a lot of pressure to use GenAI in operations. But there are risks. Hallucinations and other negative side effects are a real problem. Organization leadership and operations staff need to maintain good communication around GenAI. They need to jointly develop the organization's knowledge base around GenAI. For specific project proposals there are three prudent responses available. First, do a risk reward analysis and if the rewards don't justify the risks, don't do it. If the decision is to proceed, include a simulation test before fielding. Start now to partner with innovators developing tools to simulate, harden, and encapsulate to maximize the benefits and minimize the risks of using GenAI in operations.