On the Loop

Supervising Automation in Manufacturing

Aaron Fagan, Senior Editor

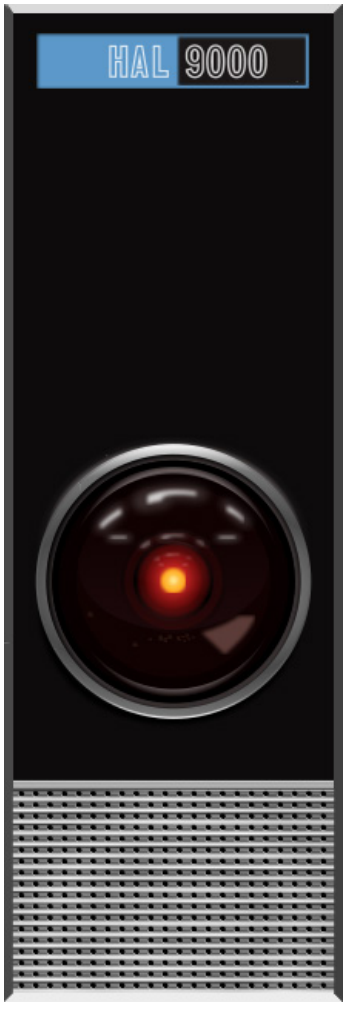
The mere mention of artificial intelligence (AI) often conjures one dystopian vision or another—perhaps the prime example of all is the HAL 9000 going spectacularly awry in the film *2001: A Space Odyssey.* The prospect of the widespread adoption of AI is understandably alarming to people in a host of ways, but be that as it may, various forms of it are already a central part of how things are done—from finance to health care, from heavy machinery to retail—and the reason for this is simple: AI allows computers to do things people can't unassisted, and by pairing algorithmic accuracy with automation, this helps save valuable time and resources. However, as AI and other control mechanisms affecting systems grow increasingly sophisticated, the human link to these processes becomes critical.

We are familiar with closed-loop manufacturing, but as AI systems are being leveraged and scaled, what will meaningful human interaction "on the loop" look like. Originally an engineering term, requiring a human "in the loop" who could halt systems, the term "on the loop" aims to distance humans further from systems.

Having an operator in the loop means a person still has complete control over starting or stopping any action performed by an intelligent system after receiving a cue. Moving toward an on-the-loop standard pushes human control farther from the center of automated decision-making. It would still give humans oversight of an automated system, but the AI would function without the need for human pre-approval as it would with an in-the-loop design.

You cannot have "meaningful human interaction" with data or sensors, or actuators at the time of data collection and operation. "The loop" is the whole system—the sensors, the actuators, the data (mostly historical, often poor quality, almost always difficult to interrogate), the machine learning or AI, the pieces separately, and the interoperable whole. No single human has the capacity to understand and oversee all these parts, let alone to meaningfully intervene.

An operator cannot meaningfully interact with active AI code. AI is a different coding discipline from machining that requires an entirely different competency. In the case of the code being embedded in other systems, few people, if any, can parse these elements in real-time. It also bears mentioning people get quantifiably bored when working with autonomous systems. The situations where machines can be autonomous but require human supervision are often the most problematic. Humans tune out or get distracted—with costly effects. Research data shows that humans cannot actively supervise machines for long periods of time without risk increasing, particularly where the systems are largely autonomous. The reason is related to the concept of "magical thinking," meaning humans are prone to assume systems cannot fail, and yet they do.



The HAL 9000 from Stanley Kubrick's *2001: A Space Odyssey* was presumed to be "incapable of error."

The complexity, speed, and scale of many autonomous, and even automatic, systems do not allow for enough time to challenge them. The speed at which information is provided, and the time-sensitive decisions that need to be made, will often render potentially appropriate human intervention impossible. If meaningful human interaction on the loop is remote, there are even greater risks. Network delays—due to issues of bandwidth, lag time, human cognitive delays, and data poverty (i.e., not having all the information you need, some of which cannot be captured by automated or autonomous projects)—amplify existing risks. Lack of transparency prevents predictability. Even where low risks exist, the speed and scale of autonomy may expedite or expand the potential of serious errors.

Meaningful human interaction on the loop will mean something different in the context of each specific system. Having experts with diverse and interdisciplinary skills involved throughout the development and lifecycle of a system directed at solving a challenge will have a far greater impact than any human on the loop. After the system is active, just being on the loop will almost certainly not provide the capabilities for any human to pause, reflect, question, and stop the trajectory of the machine.

A strong link exists between the amount of information to be processed, the tempo, and the position of humans in the decision process. The faster automation goes, the more humans will be on the loop.