Designing AI Systems With Human-Machine Teams

The greatest potential from artificial intelligence will come from tapping into the opportunities for mutual learning between people and machines.

Maria Jesus Saenz
Elena Revilla
Cristina Simón

Designing AI Systems With Human-Machine Teams

Maria Jesus Saenz, Elena Revilla, and Cristina Simón

The greatest potential from artificial intelligence will come from tapping into the opportunities for mutual learning between people and machines.

Artificial intelligence promises to augment human capabilities and reshape companies, yet many organizations are finding that the results are falling far short of their expectations. This is frustrating but not surprising. Too often, companies try to implement AI without having a clear understanding of how the technology will interface with people. [1]

Over the past decade, we have done a number of studies to examine how companies use digital capabilities to become more competitive, including a recent study on human-machine collaboration in a cross-industry setting, where we sought to better understand the contexts in which organizations use particular digital systems. [2] In this research, which included more than 20 case studies, we found that many organizations underestimated the value of teaming the predictive capabilities of algorithms with the expertise and intuitions of humans, especially in decision-

framing. This led to unsuccessful applications or missed opportunities to learn.

Finding the right balance between machines and people can be extremely challenging even for the most tech-savvy companies. Intel, for example, is seeking to use AI to assess and manage its relationships with more than 19,000 suppliers. [3] Intel's objective is to leverage massive amounts of data in order to select and monitor suppliers to build a more efficient and responsive supply chain. The specific applications it is trying to include in the program are varied. Some tools can help find potential suppliers, while others monitor supply disruptions and reputational risks.

Given its huge data-processing capacity, Intel expects to be able to handle most of the analysis with ease. However, some of the supplier assessments and sourcing decisions will be more complicated, requiring more than the *formal* rationality that underpins AI algorithms; these will necessitate using the *substantive* rationality of humans. [4] The human input will be particularly critical in situations where it's necessary to make complex decisions quickly or the signals are weak (for example, when assessing a supplier's trustworthiness or determining what to expect from a strategic agreement). For these reasons, the system relies on both human and AI-driven elements.

Assessing the Context of AI Application

On the basis of our experience studying the implementation of AI projects in a number of settings, we have found that bringing together the formal rationality of AI and the substantive rationality of humans can help companies meet their project goals and optimize the chances of success. However, before setting out to design an AI system, managers need to assess the decision-making context on two dimensions: (1) the openness of the decision-making process and (2) the level of risk. These assessments will help managers figure out the teaming options for implementing their AI systems and maximizing further learning.

Openness of the decision-making process. A closed decision-making process implies that all the relevant variables have been considered and that there are predefined rules for framing decisions. An open process, in contrast, anticipates that there will be problems that aren't well defined and that some variables (for example, supplier contract terms or behaviors) may not be known in advance.

Closed and open decision-making require different approaches with regard to AI. Closed applications have well-established, structured performance indicators and work with a set of fixed variables. Open system decisions require additional information, often from multiple sources.

Assessments as to whether the process should be open or closed may vary. Consider the challenges involved with language translation. Automatic language translators are based on preset rules of grammar and meaning, and are therefore closed; AI systems learn by accumulating terms and colloquialisms. In undefined situations, however, the process might be assessed as open; here, organizations can draw on AI systems such as natural language processing to access contextual information and learn how certain experts handle specific situations.

Level of risk. There are many different types of risk, including poor decisions that lead to physical damage, reputational damage, and financial loss. The severity of a risk depends on the specific elements. An acute risk might be tolerated if the chance of the event occurring is small. Conversely, if the chance is high, the risk may be unacceptable — even if the specific danger is small.
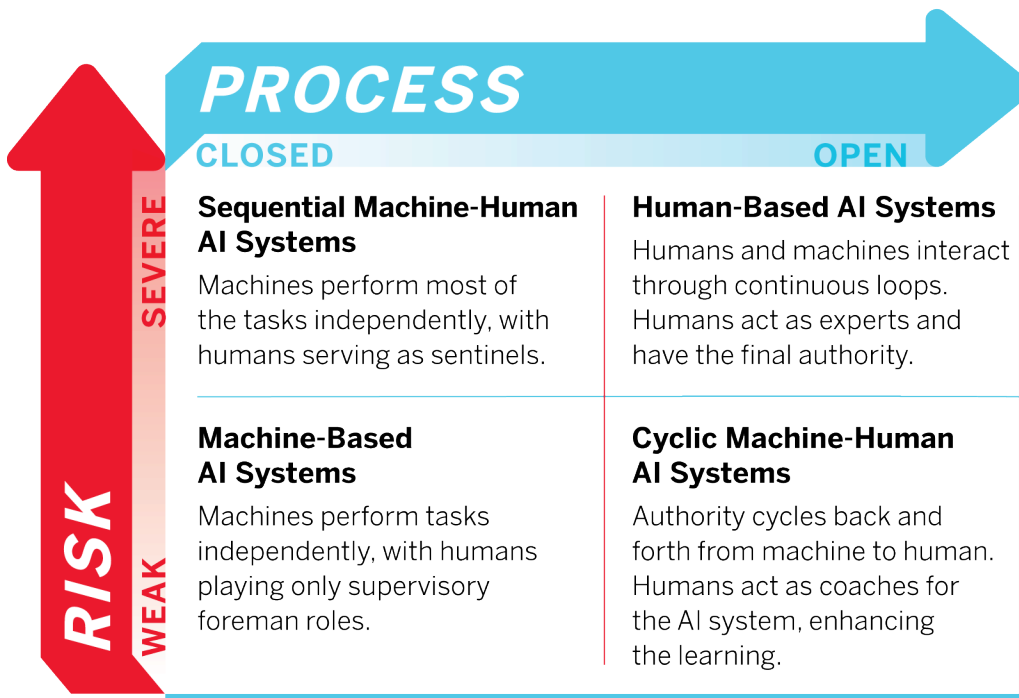
Knowing the risk level can help you decide whether you'll be comfortable making decisions entirely based on algorithms or whether you'll want additional resources like human experts on hand to help you handle unexpected situations. In touchy situations — for example, a delicate negotiation between two governments — an automatic language processor may not be able to provide the level of clarity needed to protect you from serious misunderstandings. In such cases, you may want to have human translators who understand the subtleties of language. In addition, human experts might be able to enrich the AI system, allowing for further learning.

What Role Should People Play?

An important consideration in designing the right human-machine teams is situational awareness. Combinations of human awareness and AI system design can take different forms, making different configurations possible. (See "Human-Machine Teaming Capabilities.") When the contextual factors are well defined, algorithms can "learn" by interacting with the environment through supervised machine learning. In these instances, the need for human involvement is low. Humans, in effect, act not as active decision makers but as foremen.

Human-Machine Teaming Capabilities

Depending on the circumstances, humans and machines can work together in four different ways. Humans play a distinct role in each of the quadrants.

	PROCESS	
	CLOSED	**OPEN**
SEVERE	**Sequential Machine-Human AI Systems** — Machines perform most of the tasks independently, with humans serving as sentinels.	**Human-Based AI Systems** — Humans and machines interact through continuous loops. Humans act as experts and have the final authority.
WEAK	**Machine-Based AI Systems** — Machines perform tasks independently, with humans playing only supervisory foreman roles.	**Cyclic Machine-Human AI Systems** — Authority cycles back and forth from machine to human. Humans act as coaches for the AI system, enhancing the learning.

RISK (vertical axis, from WEAK to SEVERE)

But when the potential consequences of bad decisions are serious, greater situational awareness is required. In such cases, humans can act as sentinels, relying on their experience to manage risky situations. Whereas algorithms may be good at identifying ill-defined processes, an experienced person might also be needed to train the AI system, taking on the role of coach. In situations where the levels of complexity and risk are great, the need for human-machine interaction will peak to the point where it becomes a reciprocal learning relationship. In this context, the human expert is a companion in a long-term, peer-to-peer relationship.

Human-Machine Teaming Capabilities

By combining humans and machines in AI systems, organizations can draw on four main teaming capabilities:

Interoperability. The interaction between people and machines needs to be facilitated, depending on the context and the desired outcomes. To be effective, systems should be able to share the right piece of information and analysis whenever it's required, as well as leverage the strengths and complementarities of the different agents. An AI system should also be able to specify the precise role that a human needs to play in the interaction.

Authority balance. In examining dealings between people and machines, it's essential to know which one has the final control and when. In low-risk situations, the ability to control for the outcome might be enough (even if that means that some operations are not performed properly and fixed later). But in high-risk situations, the process might require a more immediate response. The system could also decide to revise how authority is assigned in order to prevent actions that could endanger people or assets.

Transparency. Given the need for reinforcement loops

between algorithms and humans, transparent decision-making processes are key to building trust. The human needs to know which variables, rules, and performance parameters the algorithm uses. At the same time, the machine should know which decisions the human is authorized to make in order to integrate them into the learning loops.

Mutual learning. Machines learn from various sources, including the external environment, repetitive patterns, and the expected versus actual outcomes of decisions. However, they can also develop insights from human experience and intuition. This learning takes two forms: when humans make decisions that the machine analyzes and when human experts train the machines with their intuition. In other words, learning goes both ways. Just as machines learn from humans, humans can acquire insights from algorithms. These two-way learning loops increase the overall scope and performance of the AI system.

Configurations of Teaming Capabilities

Different decision-making scenarios call for different teaming capabilities. We see four different ways humans and machines can work together to make decisions.

Machine-based AI systems. In settings where machine-based designs are central and no surprises are expected, machines can perform tasks independently, with humans playing only supervisory roles and making changes only when necessary. Since potential mistakes are visible and do not pose major risks, the interoperability is for audit purposes only, and transparency is not required.

A good example of how this works can be seen at GreyOrange, a company that designs and manufactures AI-powered autonomous guided vehicles (AGVs) for warehouses, distribution centers, and fulfillment facilities used by online retailers and other companies. [5] The variables that define GreyOrange's operation scenarios in the warehouses are clear — they typically include location, speed, and type of product — as are their expected relationships. The systems are managed by machines that

adhere to a precise set of rules, goals, and key performance indicators. The warehouse operator plays the role of foreman, with little need for situational awareness; the authority resides in the AGV. Since the operating parameters are fixed, only occasional interaction with the system is needed to fine-tune performance, with the goal of progressively improving operational efficiency.

Sequential machine-human AI systems. In other settings, machines are capable of performing many of their required tasks independently. But humans need to do more than monitor the outcomes — they need to be prepared to step in to deal with unplanned contingencies. This requires humans to have situational awareness and to be ready to identify events that extend beyond the capacity of the machine and intervene. To know when such interventions are required, the AI system needs to have a level of transparency.

Amazon, for example, is in the midst of trying to figure out how to execute the handoff between machines and humans for the last mile of delivery with the help of drones. [6] In scarcely populated areas, the company may permit drones to operate largely on their own. However, in densely populated areas, Amazon foresees the need for greater interaction between the machines and humans. Whenever there's a hint of danger, the company anticipates a need for additional human support. The AI system will have the ability to register the human interventions to feed the learning process. [7]

Cyclic machine-human AI systems. In settings where the processes are open and low-risk, organizations have wide latitude for shifting decision-making authority from machine to human and vice versa. Even though a high degree of transparency may be needed, as long as the AI system is operating smoothly, the human agents' task is to monitor the outcomes without intervening in the activity. Their role is that of a coach: to train the AI system by providing new parameters and generally improving the performance.

Qoints, a digital marketing agency that uses AI-powered solutions in brand-marketing campaigns, offers a useful example. [8] The Toronto-based company uses algorithms and machine learning to identify the most effective social

media influencers for a product. The goal is to teach the AI system to make better predictions of the level of engagement influencers will drive.

Human-based AI systems. Decision processes that are both open and high-risk call for human-based AI systems, with the final authority in the hands of humans. Although the AI systems may have enough stored and processed data to make educated guesses, the risk of something bad happening can't be overlooked. Therefore, experts must maintain high situational awareness. It's critical, moreover, that the various decision rationales be sufficiently clear and transparent to advance the learning of both humans and machines.

In health care, for instance, technicians are increasingly using AI to identify problems early. [9] Radiology centers can study data patterns and 3D images for irregularities and then use human-machine teams for deeper analysis. Even though this can be a highly effective application of AI-human teams, many factors can affect a person's health that can't be integrated and coded into an algorithm. Humans and machines need to collaborate to take advantage of the different sources of mutual learning, which will improve final outcomes and create more-effective AI systems over time.

Humans and machines can work together productively in AI applications to maximize project achievements. Successful AI implementations shouldn't be tied to a single solution. Rather, they should draw on a variety of configurations that can be adapted to the scenario at hand. This will enable AI systems to shift from one configuration to another depending on the environment and human factors.

About The Authors

Maria Jesus Saenz is a research scientist at MIT and director of its Digital Supply Chain Transformation research initiative. Elena Revilla is a professor of operations and technology management at IE Business School in Madrid, where Cristina Simón is the Inditex Professor of People Management.

References

1. D. Lindebaum, M. Vesa, and F. den Hond, "Insights From 'The Machine Stops' to Better Understand Rational Assumptions in Algorithmic Decision-Making and Its Implications for Organizations," Academy of Management Review 45, no. 1 (January 2020): 247-263.

2. M.J. Saenz, J. Rice, and K. Cottrill, "Do You Have the DT's? Why It's Time to Bring Digital Transformation Down to Earth," Supply Chain Management Review, April 1, 2019, www.scmr.com.

3. "Intel Embarks on a Digital Supply Chain Journey," SupplyChainBrain, Dec. 19, 2018, www.supplychainbrain.com.

4. S. Barro and T.H. Davenport, "People and Machines: Partners in Innovation," MIT Sloan Management Review 60, no. 4 (summer 2019): 22-28.

5. "GreyOrange Expands Its Global Footprint, Opens U.S. Headquarters in Atlanta," Business Wire, Aug. 28, 2018, www.businesswire.com.

6. D. Lee, "Amazon to Deliver by Drone 'Within Months,'" BBC News Service, June 5, 2019, www.bbc.com.

7. Lee, "Amazon."

8. D. Takahashi, "Qoints Uses IBM's Watson AI to Find Your Influencers," Venture Beat, May 3, 2018, https://venturebeat.com.

9. K. Loria, "Putting the AI in Radiology," Radiology Today 19, no. 1 (January 2018): 10.

MITSloan
Management Review

PDFs ■ Reprints ■ Permission to Copy ■ Back Issues

Articles published in *MIT Sloan Management Review* are copyrighted by the Massachusetts Institute of Technology unless otherwise specified at the end of an article.

MIT Sloan Management Review articles, permissions, and back issues can be purchased on our website: **shop.sloanreview.mit.edu**, or you may order through our Business Service Center (9 a.m.-5 p.m. ET) at the phone number listed below.

To reproduce or transmit one or more *MIT Sloan Management Review* articles **requires written permission.**

To request permission, use our website
shop.sloanreview.mit.edu/store/faq,
email **smr-help@mit.edu** or call 617-253-7170.