INTELLIGENCE · ARTIFICIAL AND COLLECTIVE INTELLIGENCE

Human-Machine Teaming in Analysis

The best results in analysis often come neither from people alone nor from machines alone but from the two working together. This article explains the research on human-machine teaming, when combinations outperform either partner, why they often do not, and how to design analytic work so that each contributes its strengths.

ModQu research article · 1,784 words · 19 references (APA 7th edition)

In 1960, the psychologist J. C. R. Licklider proposed that the future of computing lay not in machines that would replace people but in **man-computer symbiosis**: a close partnership in which people set goals, form hypotheses and evaluate results, while computers do the routine work of calculation and data handling that prepares the way for insight (Licklider, 1960). More than sixty years later, with machines able to read, classify and predict at enormous scale, Licklider's question has become practical for every analytic team. How should the work be divided, and does the combination actually perform better than either partner alone?

The promise of teaming

The most cited example comes from chess. After losing to IBM's Deep Blue computer in 1997, the world champion Garry Kasparov helped promote a form of the game in which players could consult computers during play. Kasparov (2017) describes a 2005 online tournament in which the winners were not grandmasters with powerful machines but two amateur players using ordinary computers, whose skill lay in managing the process: knowing when to consult the machines, how to compare their suggestions, and when to trust their own judgement. Kasparov's lesson was that a good process for combining people and machines could matter more than the strength of either.

Management writers have developed the idea for organisations. Daugherty and Wilson (2018) argue that the largest gains from AI come from redesigning work so that people and machines complement each other, rather than simply automating existing tasks. Jarrahi (2018) suggests that machines are strongest at handling complexity, the processing of large volumes of data, while people are strongest at handling uncertainty and ambiguity, where intuition, context and judgement are needed. Malone (2018) describes "superminds" in which groups of people and computers together act more intelligently than either could alone.

The evidence is mixed

The research, however, shows that teaming often fails to deliver. Vaccaro, Almaatouq and Malone (2024) carried out a meta-analysis of more than a hundred experiments comparing humans alone, AI alone and human-AI combinations. On average, the combinations performed worse than the better of the two working alone. The results varied with the task.

Combinations tended to lose ground in decision tasks, such as classifying cases or detecting fakes, and to gain in creative tasks, such as generating content. They also tended to help when the human alone outperformed the AI alone, and to hurt when the AI was stronger.

Bansal et al. (2021) examined whether explanations from AI systems helped. In their experiments, explanations increased people's tendency to accept the AI's recommendation whether it was right or wrong. They did not help people distinguish good recommendations from bad ones, so the combination did not reliably beat the AI alone.

Kleinberg et al. (2018) found a similar pattern in a real setting. Their algorithm for predicting whether defendants would fail to appear in court made better predictions than the judges. Judges sometimes overrode the algorithm's implicit advice in ways that worsened outcomes, because they gave weight to signals that did not actually predict behaviour. The combination was only as good as the human's ability to tell when to override.

Figure 1 summarises these findings in a simplified form. Teaming helps most when people can judge when the machine is likely to be right, and hurts most when the machine is the stronger partner but people override it at the wrong moments.

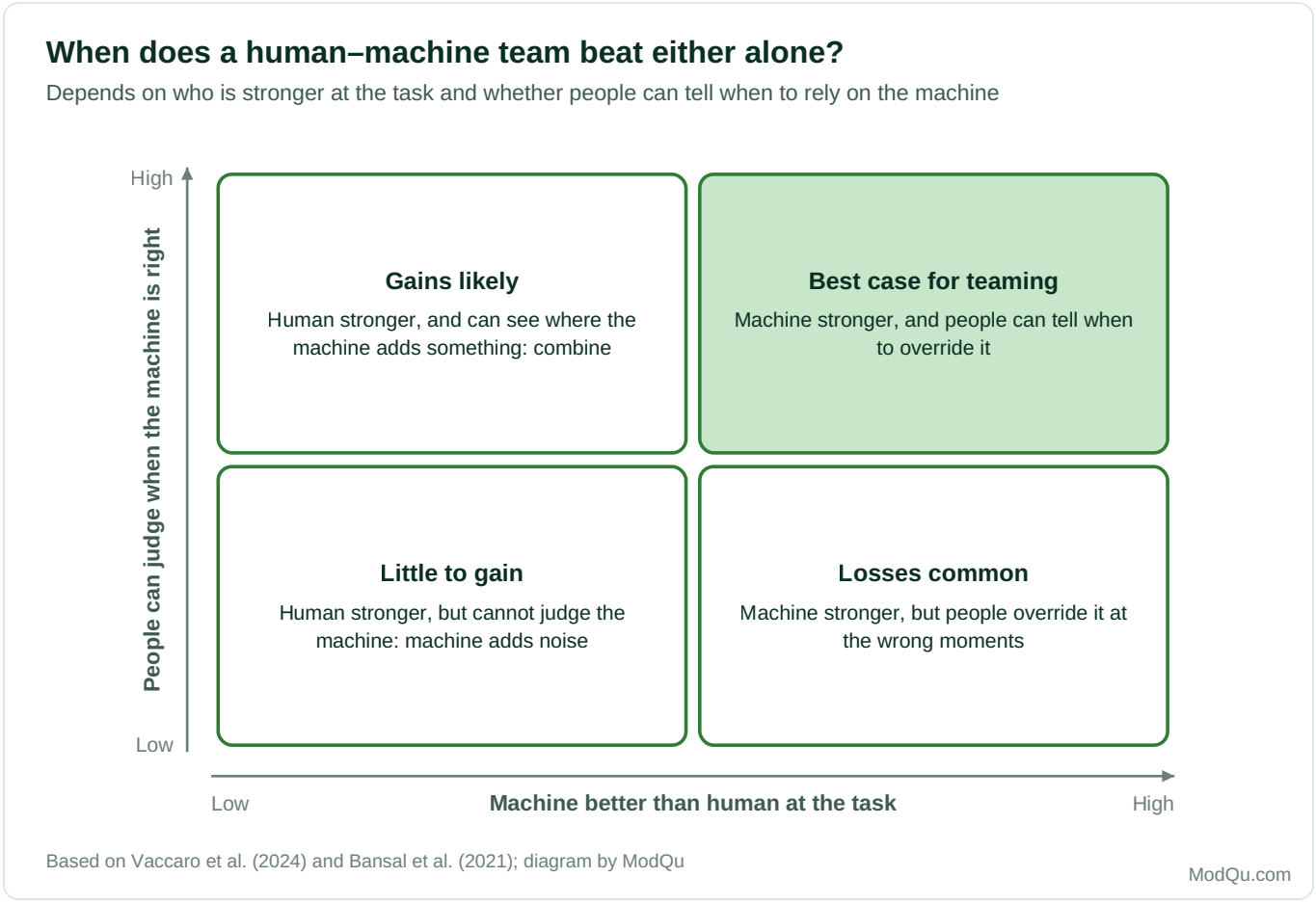


Figure 1. When does a human–machine team beat either alone? A simplified summary of research findings. The key is not only who is stronger but whether people can tell when to rely on the machine. Based on Vaccaro et al. (2024) and Bansal et al. (2021); diagram by ModQu.

Trust: too much and too little

The central problem is calibrating trust. Two opposite errors are well documented.

Too much trust. Parasuraman and Riley (1997) described **misuse** of automation: relying on it when it is unreliable. Skitka, Mosier and Burdick (1999) found that people working with automated aids made errors of omission, missing problems the aid did not flag, and errors of commission, following the aid's incorrect advice even when other evidence contradicted it. Logg, Minson and Moore (2019) found that, in many judgement tasks, people actually prefer advice they believe comes from an algorithm over the same advice from a person.

Too little trust. Dietvorst, Simmons and Massey (2015) documented **algorithm aversion**: after seeing an algorithm make a mistake, people became less willing to use it, even when it remained more accurate than they were. They held the algorithm to a higher standard than themselves. In a later study, Dietvorst, Simmons and Massey (2018) found that people were much more willing to use an imperfect algorithm if they could adjust its output, even slightly. Giving people a small measure of control increased use of the better tool.

Lee and See (2004) argue that the aim is **appropriate reliance**: trust matched to the actual capability of the system in the specific conditions of use. This requires people to understand what the system does well, where it fails and how to tell the difference.

Dividing the work

Parasuraman, Sheridan and Wickens (2000) proposed a model for deciding how much to automate. They distinguish four stages of any information-processing task: acquiring information, analysing it, selecting a decision and implementing an action. Each stage can be automated to a different degree, from none to full. Their advice is to choose the level for each stage according to the consequences of errors and the reliability of the automation, rather than automating everything that technology allows. Figure 2 applies their model to an intelligence team.

Dividing analytic work between people and machines

Four stages of automation, applied to an intelligence team

Stage	Machine role	Human role	Suggested level
Acquire information	Monitor sources, filter, translate	Set sources and filters; spot-check	High
Analyse information	Cluster, flag anomalies, score	Test patterns; seek other explanations	Medium
Select a judgement	Offer options with evidence	Decide, state confidence, sign off	Low
Act on it	Route products, set alerts	Brief decision-makers; take responsibility	Low to medium

Stages from Parasuraman, Sheridan and Wickens (2000); examples and levels by ModQu

ModQu.com

Figure 2. Dividing analytic work between people and machines. Higher automation suits early stages where errors are visible and recoverable; judgement and responsibility stay with people. Stages from Parasuraman, Sheridan and Wickens (2000); examples and levels by ModQu.

In analysis, the early stages of finding, filtering and translating information are well suited to high levels of automation, because errors are usually visible and the volume is beyond human capacity. Analysis can be shared: machines flag patterns and anomalies, people test whether they are real and what they mean. The selection of a judgement, and the confidence attached to it, should remain with people who can be held accountable.

Keeping people in the loop

Automation changes the human role, and the change can weaken the very abilities teaming relies on. Bainbridge (1983) identified the "ironies of automation": the more a system does, the less practice operators have, and the less able they are to take over when it fails, which is precisely when they are needed. Endsley (1995) described **situation awareness**, the operator's perception and understanding of what is happening and what will happen next. Endsley (2017) reviewed decades of research and concluded that as automation becomes more capable, people tend to lose situation awareness, a problem Endsley calls the "automation conundrum". The more reliable the system, the less attention people pay, and the harder it is for them to catch its rare failures.

For analytic teams, this means designing work so that analysts remain engaged with the material, not merely reviewing machine conclusions. Practical methods include requiring analysts to form their own initial view before seeing the machine's output, rotating tasks so that skills are maintained, and regularly reviewing cases where people and machines disagreed.

Augmentation, not only automation

Raisch and Krakowski (2021) describe an **automation–augmentation paradox** in management. Organisations can use AI to automate tasks, removing people, or to augment them, working alongside people. The two are often presented as alternatives, but in practice they are linked: automating routine tasks frees people for augmented work, and augmented work generates the knowledge needed to automate further. Focusing only on automation risks losing the human judgement that gives context to machine output; focusing only on augmentation can leave obvious efficiencies untaken.

Designing an effective team

Drawing on this research, several principles help analytic teams get the benefits of teaming:

1. **Measure both partners separately.** Know how accurate the machine is alone and how accurate analysts are alone before judging the combination.
2. **Give people a basis for judging the machine.** Show confidence levels, past accuracy on similar cases, and the evidence behind each output.
3. **Ask for the human view first.** Analysts who record their own judgement before seeing the machine's are less likely to anchor on it.
4. **Allow adjustment.** Let analysts modify machine output rather than only accept or reject it, and record their changes.
5. **Review disagreements.** Cases where people and machines differ are the richest source of learning about both.
6. **Maintain skills.** Ensure analysts continue to perform core tasks without the machine often enough to keep their abilities.
7. **Assign accountability.** Every product needs a named person responsible for its judgements.

Common mistakes

- **Assuming the combination is automatically better.** It must be measured against each partner alone.
- **Adding explanations and expecting better judgement.** Explanations can raise acceptance of wrong advice as well as right advice.
- **Treating every override as a failure, or every acceptance as success.** Both need to be checked against outcomes.
- **Letting skills fade.** Analysts who never work without the tool cannot catch its failures.
- **Placing the machine's answer first on the screen.** What people see first anchors what they conclude.

A business illustration

A credit insurer uses a model to score the risk that small companies will default on payments to their suppliers. Underwriters review each score before setting cover. An internal review finds that underwriters override the model in about one case in five, and that their overrides are right only slightly more often than chance. However, in one category, firms in sectors affected by recent regulatory changes, their overrides are right far more often, because they follow industry news the model does not yet reflect.

The insurer redesigns the process. Underwriters now record their own risk view before seeing the score for a sample of cases, so that both can be compared with outcomes. The model output shows its historical accuracy for similar firms. Underwriters are encouraged to override in sectors with recent structural change and asked to give written reasons elsewhere. Within a year, the combined decisions are more accurate than either the model or the underwriters alone.

Conclusion

Human-machine teaming can produce better analysis than either people or machines alone, but it does not happen automatically. The research shows that combinations often underperform because people trust machines too much or too little and cannot tell when to rely on them. Effective teams divide work deliberately, give people the information they need to judge machine outputs, keep analysts engaged and skilled, and learn systematically from disagreements. Licklider's vision of symbiosis remains the goal; achieving it is a matter of design, not technology alone.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 81). ACM.

Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*. Harvard Business Review Press.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Kasparov, G. (2017). *Deep thinking: Where machine intelligence ends and human creativity begins*. PublicAffairs.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics*, *133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, *HFE-1*(1), 4–11. https://doi.org/10.1109/THFE2.1960.4503259

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together*. Little, Brown.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1