

RESEARCH SUMMARY

Legal, Ethical and Moral Implications of Machine Learning

Machine learning algorithms can distill vast quantities of information into simpler, more actionable information. But how can decision makers trust that the distilled information is valid and suits their needs? This summary explores research into legal, ethical and moral risks associated with machine learning in the context of personnel management processes with applicability to the U.S. military.

Can machine learning be incorporated into personnel management processes — for recruiting, compensation, promotion, retention and the like — while also maintaining or exceeding the ethical status quo of those processes? Can it move personnel policy closer to an ethical ideal? How might machine learning harm the status quo if it is misapplied?

These questions have complex answers, requiring a closer look at what is gained and lost with machine learning tools. Using a handful of fictional but plausible scenarios, a team of Institute for Defense Analyses (IDA) researchers, including Alan Gelder, Cullen Roberts and Ashlie Williams, illustrated on a

case-by-case basis how the Defense Department’s ethical principles on artificial intelligence might be employed in military personnel management. Although the work focused on a military setting, the underlying lessons apply broadly in various contexts.

The military’s foundational personnel management objective is to acquire, develop, and retain personnel with vital skills, experience and capabilities. Policy supports this objective by shaping attrition risks, personnel quality, recruiting effectiveness, and unit cohesion.

IDA

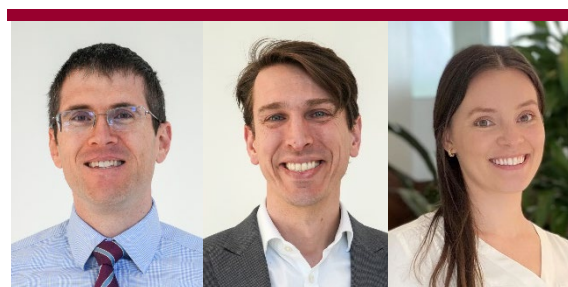
April 2024 | Product 3001921

Plan	<p>Is machine learning the right tool for the job?</p> <p>What are the goals and will machine learning help achieve them?</p> <p>Would a different approach be more effective?</p> <p>How do the risks and benefits of using machine learning differ from the status quo?</p>
Select Data	<p>Is the data appropriate for the job?</p> <p>What is the source of the data? How reliable is the data and what are its limitations?</p> <p>How is the data prepared for analysis?</p> <p>How is the data's accuracy maintained?</p> <p>Is the data periodically checked for unintended bias?</p> <p>How is information that should be excluded from particular decisions dealt with?</p> <p>What privacy and protection safeguards are in place?</p>
Design	<p>What should developers be aware of in designing the machine learning model?</p> <p>Have developers consulted with stakeholders about potential legal or ethical issues?</p> <p>What dimensions of diversity should developers consider in designing the model?</p> <p>Does the machine learning project have end-to-end transparency?</p> <p>How will the system be tested and monitored?</p> <p>Is the code review process robust and the pipeline reproducible?</p> <p>What tests monitor the adequacy of results?</p> <p>Is the pipeline modular so that problematic components can be readily swapped out?</p> <p>What corrective actions can minimize differential outcomes across given populations?</p> <p>How are these actions evaluated and tested?</p>
Implement	<p>Are there processes to enable responsible use of the model?</p> <p>Do stakeholders and users understand the appropriate uses and limitations of the model?</p> <p>Is documentation and training effective?</p> <p>What is the plan for monitoring and evaluating model use?</p> <p>Are safeguards in place to identify unintended consequences and intervene?</p>

Machine learning techniques can help to inform policy by synthesizing information within vast Defense Department personnel databases.

Unfortunately, synthesis can also lead to misuse of information, by intent or accident. The team proposed a generic framework (summarized in the table above) for considering issues like these throughout the life cycle of a machine learning application, whether in the context of military personnel or a broader personnel setting.

For more information, see the report that is the basis for this summary, [IDA Paper P-33087](#).



Alan B. Gelder (agelder@ida.org), Cullen A. Roberts (croberts@ida.org) and Ashlie M. Williams (awilliam@ida.org) are researchers in the Strategy, Forces and Resources Division of the Systems and Analyses Center, an IDA-managed federally funded research and development center. The team's expertise includes experimental, behavioral, health and labor economics; applied econometrics; social and behavioral science research; program evaluation; data science; and machine learning.