# Evaluating Managerial Implications in Research Papers with GPT and Expert Review

Naval Postgraduate School


## Abstract

This capstone examines how academic research can better serve managerial practice by evaluating whether artificial intelligence can reliably assess managerial implications in management and marketing scholarship. The study develops a structured, scalable methodology for scoring managerial implications and tests a refined AI model against expert reviewers. The results point toward a hybrid approach in which AI accelerates preliminary screening while human experts provide substantive judgment. The project highlights both the promise and limits of AI in narrowing the research–practice gap and identifies future opportunities to test alternative rubrics, diversify data sources and explore the role of artificial randomness in evaluation systems.


## Methods

- Exploratory nature
- Rubric framework adapted from Hawkins and Muir (2024)
- Human expert evaluation
- Sample of 10 research articles from JBL, JPSM, and JOM, published between 2015 and 2025
- Multiple AI systems (ChatGPT GPT-4, Claude Sonnet 4.5, and Grok)
- Inter-rater reliability using intraclass correlation coefficients (ICCs)

## Results & Impact

- GenAI scoring of managerial implications is inconsistent across evaluation levels
- The experiments indicate that using GenAI to mirror human scoring is not effective
- Evidence suggests the current rubric may be overly positive
- Future refinement may require consulting experts who use rubrics within GenAI systems to identify potential improvements or overlooked methods

## Future Research

- Test alternative rubric designs
- Larger samples to better understand variance
- Managerial relevance for defense related journal articles
- Explore AI temperature adjustment


Louis Gianneschi, Capt, USAF

Advisor: Lt Col Jamie Porchia, USAF