Uncertainty Quantification

Artificial Intelligence and Machine Learning in Military Systems

Ayla R. Reed

Instituting a military standard for quantified uncertainty metadata represents a solution to the problems inherent in using artificial intelligence/machine learning (AI/ML) for military advantage. By provisioning for metadata now, the Department of Defense can continue to determine the best policy for using AI/ML in parallel with capability development. This coordination will prevent delays in solving difficult technical problems associated with implementing AI/ML in warfighting systems. Uncertainty quantification can enable a practical digital implementation of the observe, orient, decide, and act loop, addressing ethical issues with employing AI/ML in war and optimizing investment in research and development.

Foundationally, the US military does not need artificial intelligence/machine learning (AI/ML). Yet the military needs to be able to observe, orient, decide, and act (OODA) faster—and better—than an adversary to achieve military advantage.[1] Machines have the capacity to observe, orient, decide, and act at a faster pace than humans and thus enable this advantage. The debate remains open, however, on the appropriateness of allowing AI or ML models to "decide" on the best course of military action, when that decision may result in destruction and death.

The potential pitfalls of utilizing AI/ML for military advantage have been propounded ad nauseam.[2] Three issues remain the most concerning: (1) addressing the moral and ethical considerations for giving an AI the authority to destroy things and people; (2) balancing the cost versus military utility of developing AI/ML capability; and (3) ensuring

Major Ayla R. Reed, USAF, program manager at the Maui High Performance Computing Center, Hawaii, holds a master of science in engineering from the University of Arkansas.

1. John R. Boyd, "Patterns of Conflict," in *A Discourse on Winning and Losing*, ed. Grant T. Hammond (Maxwell AFB, AL: Air University Press, March 2018).

2. Arif Ali Khan et al., "Ethics of AI: A Systematic Literature Review of Principles and Challenges," in *Proceedings of the International Conference on Evaluation and Assessment in Software Engineering 2022* (New York: Association for Computing Machinery, June 2022), https://doi.org/; Avi Goldfarb and Jon R. Lindsay, "Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War," *International Security* 46, no. 3 (Winter 2021–22): 7, https://direct.mit.edu/; Nick Starck, David Bierbrauer, and Paul Maxwell, "Artificial Intelligence, Real Risks: Understanding—and Mitigating—Vulnerabilities in the Military Use of AI," in *Compete and Win: Envisioning a Competitive Strategy for the Twenty-First Century,* Competition in Cyberspace Project, Army Cyber Institute and the Modern War Institute, January 18, 2022, https://mwi.usma.edu/; and Emre Kazim and Adriano Soares Koshiyama, "A High-Level Overview of AI Ethics," *Patterns* 2, no. 9 (September 2021), https://doi.org/.

an appropriate level of trust in a machine to make optimal use of the investment into the AI/ML components of capability development. Nevertheless, uncertainty quantification (UQ) included as metadata to military information can address these three pitfalls while adhering to DoD ethical principles for artificial intelligence.

The DoD artificial intelligence strategy prioritizes and incentivizes the maturation of AI/ML technology.[3] The result has been a flurry of activity attempting to expeditiously implement capability, accompanied by minimal planning for sustainability of capability growth or the higher-order implications for use of AI/ML. As one defense researcher has observed, "When technological change is driven more by hubris and ideology than by scientific understanding, the institutions that traditionally moderate these forces, such as democratic oversight and the rule of law, can be eroded in pursuit of the next false dawn."[4]

The Defense Advanced Research Projects Agency argues that current AI/ML systems "lack the necessary mathematical framework" to provide assurance in use, which impedes their "broad deployment and adoption for critical defense situations or capabilities."[5] Assurance requires confidence, and confidence requires minimal uncertainty. Such assurance in systems using AI/ML can help address ethical considerations, provide insight into the cost of development versus utility, and allow the locus of responsibility for its use in war to remain with commanders and operators at the lowest possible echelon.

By implementing a military standard for uncertainty quantification in AI/ML systems, the Defense Department can secure the much-needed trust in those systems. Further, there are feasible ways to apply existing mathematical approaches for uncertainty determination and propagation if the Department makes UQ a requirement for developers. Yet as the military applies this standard to information, it must bear in mind the higher-order effects and challenges of uncertainty quantification.

Uncertainty Quantification for AI/ML

To address the three pitfalls mentioned above, uncertainty quantification should be required within and by any military digital system. Uncertainty quantification, which is the process of assigning some number(s) to the imperfect or unknown information in a system, will allow a machine to express in real time how unsure it is, adding critical transparency for building trust in its use. The Department of Defense should implement a military standard that specifies the quantification of uncertainty tagged as metadata to each data or piece of information available in digital systems. Once available, these

3. Department of Defense (DoD), *Summary of the Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity* (Washington, DC: DoD, February 2019), https://media.defense.gov/.

4. Zac Rogers, "Have Strategists Drunk the 'AI Race' Kool-Aid?," War on the Rocks, June 4, 2019, https://warontherocks.com/.

5. Defense Advanced Research Projects Agency (DARPA) Public Affairs, "Progressing towards Assuredly Safer Autonomous Systems," DARPA, January 29, 2020, https://www.darpa.mil/.

metadata can be propagated to higher levels of information usage through functional relationships, providing an AI or ML model the information needed to always express how confident it is in its output.

Understanding UQ as metadata requires understanding foundational concepts in metrology—the science of weights and measures—related to measurement uncertainty. That is, a measurement has two components: 1) a numerical value which is the best estimate of the quantity being measured, and 2) a measure of the uncertainty associated with this estimated value.

Of note, the 2008 International Organization for Standardization (ISO) *Guide to the Expression of Uncertainty in Measurements* defines the difference between measurement uncertainty and measurement error. These terms are not synonymous: "The ± (plus or minus) symbol that often follows the reported value of a measurand [the quantity being measured] and the numerical quantity that follows this symbol, indicate the uncertainty associated with the particular measurand and not the error. An error is the discrepancy between a measured value and the actual or true value. Uncertainty is the effect of many errors."[6]

In military parlance, a "measurement" is any information collected and used during an OODA loop. Each piece of information has been measured by a sensor of some sort and will have some uncertainty associated with it. Uncertainty quantification as metadata will take at least two forms: empirically generated measurement uncertainty (based on the metrology standards outlined above) and statistically postulated uncertainty (determined by some means, of which there are many).[7]

An operator can use the system-reported uncertainty to inform their tactical decision when using a UQ-capable system. Commanders can set predefined levels of trust needed for various categories of military action at the operational or even strategic level using such systems, which can help operators understand what their authorities are when using an AI or ML model. This would also help acquisition professionals make appropriate investment decisions for AI/ML capability development because it would quantify aspects of utility. Moreover, providing quantified minimum levels of certainty required in systems using AI/ML addresses the three pitfalls discussed above.

In terms of the moral and ethical concerns of using AI, there is no single right answer to the question "Is it moral or ethical to allow an AI or ML model to decide on a military course of action that will result in destruction and death?" As with all moral and ethical debates, dealing in absolutes is impossible.

---

6. Ian Farrance and Robert Frenkel, "Uncertainty of Measurement: A Review of the Rules for Calculating Uncertainty Components through Functional Relationships," *Clinical Biochemist Reviews* 33, no. 2 (2012): 50–51.

7. Moloud Abdar et al., "A Review of Uncertainty Quantification (UQ) in Deep Learning: Techniques, Applications, and Challenges," *Information Fusion* 76 (December 2021), https://doi.org/; and Apostolos Psaros et al., "Uncertainty Quantification in Scientific Machine Learning: Methods, Metrics, and Comparisons," *Journal of Computational Physics* 477 (March 15, 2023), https://arxiv.org/.

Consequently, the Department of Defense should categorize military actions into one of the three well-known relative degrees of machine autonomy: things a machine can never do by itself, things a machine can sometimes or partially do by itself, or things a machine can always do by itself. The Department of Defense then can define a minimum level of certainty as a boundary condition for each of these categories and/or can define minimum levels of certainty needed for specific actions. The criticality of the decision or action will drive the determination of a UQ boundary. Using uncertainty quantification embraces the nuance and ambiguity in addressing ethical considerations for systems using AI/ML.

When it comes to balancing the cost of artificial intelligence/machine learning with its use, the Department of Defense's fiduciary responsibility is to ensure the investment in AI/ML development is proportional to its military utility. There is no purpose in developing and procuring a battalion of fully autonomous killer droids if AI/ML policy prohibits the US military from allowing an AI to decide to destroy something or kill someone. Therefore, predefined minimum uncertainty boundaries will allow acquisition professionals to determine how best to spend limited resources for the greatest return on investment.

Optimizing trust in AI/ML during capability development will require safeguards against widespread inexperience in AI/ML acquisition and the relative juvenility of the science of uncertainty quantification in machine learning. "Uncertainty is fundamental to the field of machine learning, yet it is one of the aspects that causes the most difficulty for beginners, especially those coming from a developer background."[8] All aspects of system development should include metadata tags for uncertainty quantification, whether the system is intended to be used autonomously or not.

These outputs might be rolled up into a higher-level digital capability that will then require the UQ data to calculate uncertainty propagation. For example, an F-16 maintainer's fault code reader should have uncertainty quantification metadata tagged to each fault reading, providing this quantification at the source. The reader itself is not intended to incorporate AI or a machine-learning model, and that data may not be used immediately in an AI/ML application, but the fault data might be compiled with fleet-wide fault data and submitted to an external ML model that forecasts depot-level maintenance trends. The metadata would follow that set of digital information through any level of compilation or higher-order use.

Requiring uncertainty quantification metadata as a military standard achieves the intent of the Secretary of Defense's ethical principles for artificial intelligence that encompass five major areas:[9]

- Responsible: UQ informs judgment and provides the empirical basis for developing, deploying, and using AI capabilities.

---

8. Jason Brownlee, "A Gentle Introduction to Uncertainty in Machine Learning," Machine Learning Mastery, last updated September 25, 2019, https://machinelearningmastery.com/.

9. DoD, "DOD Adopts Ethical Principles for Artificial Intelligence," press release, February 24, 2020, https://www.defense.gov/.

- Equitable: Bias in AI can be measured in the same way that uncertainty is and is based on many of the same statistical principles.[10] Bias can then be addressed and improved.
- Traceable: Requiring uncertainty metadata at every level enables traceability in assurance. Performance issues in machines can be traced back to the culpable component.
- Reliable: UQ allows inspection by developers and allows targeted improvement of the most egregious input factors.
- Governable: UQ as boundary conditions for autonomy trust levels can be used to define guidelines for fulfilling intended functions and avoiding unintended consequences.

These ethical principles were adopted to ensure the Department of Defense continues to uphold the highest ethical standards while embracing the integration of artificial intelligence as a disruptive technology. Uncertainty quantification is a practical way to achieve that goal.

Building Trust in AI/ML

A study by RAND found trust is the root cause of most concerns related to the military use of AI/ML.[11] Department of Defense researchers note that "when it comes to forming effective teams of humans and autonomous systems, humans need timely and accurate insights about their machine partners' skills, experience, and reliability to trust them in dynamic environments."[12] For many autonomous systems, their "lack of awareness of their own competence and their inability to communicate it to their human partners reduce trust and undermine team effectiveness."[13]

Trust in the AI/ML model is fundamentally based on the certainty humans have in the information, whether it be a simple sensor output or the overall competency of an autonomous weapon system. This is supported by MITRE Corporation studies:

> AI adopters often ask about ways to increase trust in the AI. The solution is not for us to build systems that people trust completely, or for users only to accept systems that never err. Instead, lessons point to the importance of forming good partnerships based on evidence and perception. Good partnerships help humans understand the AI's abilities and intents, believe that the AI will work as antici-

---

10. V. Ashley Villar and Michael Little, "Technical Memorandum: Focus Area 3—Uncertainty and Bias," in *NASA SMD AI Workshop Report*, ed. Manil Maskey (Washington, DC: National Aeronautics and Space Administration, September 2021).

11. Forrest E. Morgan et al., *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World* (Santa Monica, CA: RAND Corporation, 2020), https://www.rand.org/.

12. DARPA Public Affairs, "Building Trusted Human-Machine Partnerships," DARPA, January 31, 2019, https://www.darpa.mil/.

13. DARPA Public Affairs, "Human-Machine Partnerships."

pated, and rely on the AI to the appropriate degree. Then stakeholders can calibrate their trust and weigh the potential consequences of the AI's decisions before granting appropriate authorities to the AI.[14]

By thinking of machines—digital or physical—as partners, the military can make analogies to confidence-building techniques with human partners. Sound partnership requires effective two-way communication and a system to reinforce collaboration.[15] Ipso facto, a measure of uncertainty in the digital system output is not useful unless that uncertainty can be conveyed to the human partner. Once machines can quantify uncertainty and can communicate that quantification, they also enable the evaluation of the output and improvement of the system.

Real-time feedback of a machine's awareness of its own competence will increase transparency into the machine's observe, orient, and decide functions by providing quantification of the uncertainty in each of those loops. This feedback improves trust in that specific system and enables quantification of trust in systems-of-systems via uncertainty propagation. For example, consider remotely piloted aircraft (RPA) video surveillance of a potential target. How certain is it that an RPA sensor is accurate and calibrated, that the video stream has not been compromised, and/or that the operator has been given sound baseline intelligence on where to point the sensor in the first place?

Each of these components of the OODA loop has some associated uncertainty that can and should be quantified so that it can be mathematically propagated to the level of decision-making. In this scenario, it would result in a propagated certainty of *x* percent that the target is correct, giving the mission commander confidence in their situational awareness (observation), and allowing them to orient better and decide faster on whether to engage or not.

By quantifying uncertainty and using it in tandem with predefined levels of confidence needed for various categories of action, decisionmakers can create boundary conditions around those military actions that have little to no moral implications as well as those that have serious moral implications. Defense senior leaders can also set thresholds for proportional investment in developing and applying AI/ML capability and can ensure that investment will be used to achieve optimal military advantage. This would provide assurance in a system using AI/ML through a "quantify–evaluate–improve–communicate" cycle.[16]

Uncertainty quantification allows setting if-then relationships for bounding the allowable space of actions for a machine. In another abbreviated example, a space domain

---

14. Jonathan Rotner, Ron Hodge, and Lura Danley, *AI Fails and How We Can Learn from Them* (McLean, VA: MITRE Corporation, July 2020), 43, https://sites.mitre.org/; and see also Andrew Lacher, Robert Grabowsky, and Steve Cook, "A Framework for Discussing Trust in Increasingly Autonomous Systems," MITRE Corporation, updated June 2017, https://www.mitre.org/.

15. Rotner, Hodge, and Danley, *AI Fails*, 43.

16. Soumya Ghosh et al., "Uncertainty Quantification 360: A Holistic Toolkit for Quantifying and Communicating the Uncertainty of AI," arXiv, June 2021, https://arxiv.org/.

awareness mission may use infrared sensor data to identify space vehicles. The if-then relationship may look like this: If a sensor data-to-target correlation model has a certainty greater than 95 percent, then that target identification information can be automatically updated in the National Space Defense Center catalog. If a sensor data-to-target correlation model has a certainty greater than 75 percent but less than 95 percent, then the machine can attempt a match to signals intelligence (SIGINT) with a certainty greater than 75 percent, or it can send the information to a human to verify.

Using quantified uncertainty thus allows commanders to root decision trees in parameters usable by AI/ML models and to guide how those AI/ML models may be used. In considering the three relative degrees of machine autonomy, commanders can predefine levels of uncertainty for the inputs to each of these categories of action as guidelines for when and under what circumstances it makes sense to let a machine decide, clearly defining the rules of engagement for using an AI or ML model.

All weapon systems, whether intended to incorporate autonomy or not, should provide uncertainty metadata within their planned user interface. Knowing the uncertainty of all inputs benefits conventional weapon systems users as much as applications of AI/ML. By provisioning for metadata now, DoD senior leaders can continue determining the best governance and policy for using AI/ML without slowing down technical and engineering development. Any such governance can be implemented in the future by referencing the quantified uncertainty within a system at the component level or at the output level.

Mathematical Implementation

Applying uncertainty quantification and propagation to tightening the OODA loop assumes functional relationships can be used to define military situations. Functional relationships are the best mathematical approach for this application because it can generally be shown that a cause-effect relationship exists between the value of the function and the input variables, without specifically identifying the exact mathematical form of the relationship. By assuming these functional relationships exist, a general equation which describes the propagation of uncertainty can be used.[17]

A generic functional relationship with uncertainty terms looks like:

$$y \pm u(y) = f(x_1 \pm u_1, x_2 \pm u_2, x_3 \pm u_3, \ldots, x_n \pm u_n)$$

where y is the output, $u(y)$ is the uncertainty of that output, and there are n input variables with associated uncertainties that affect that output. This shows that y depends on n input variables, and in the style of "imprecise probabilists," that the exact value of y is within the interval $y + u(y)$ to $y - u(y)$.[18]

17. Farrance and Frenkel, "Uncertainty of Measurement."

18. Barry N. Taylor and Chris E. Kuyatt, National Institute of Standards and Technology (NIST) Technical Note 1297, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, 1994 ed. (Gaithersburg, MD: NIST, September 1994); and T. J. Sullivan, *Introduction to Uncertainty Quantification* (Cham, Switzerland: Springer Cham, 2015), 31.

This direct application of ideas intended to improve medical laboratory research pertains to military decision-making as well. "Uncertainty associated with any measurement and its propagation through a defined functional relationship can be evaluated by differentiation (partial differentiation) and the application of the general equation for the propagation of uncertainty."[19] These mathematical approaches would capture the change in uncertainty as many measurands change in a very complex system. This uncertainty propagation equation can be derived using standard statistical procedures, and most importantly, it is independent of the exact form of the functional relationship.[20]

Those more versed in statistics are invited to submit this approach to further case study and determine the feasibility of calculating propagated uncertainty at very large system-of-systems levels when many input variables need to be included. It has already been shown that "the more complex the problem, the more costly it is to obtain calibrated uncertainty estimates."[21] This approach is probably feasible through operational level AI/ML models (i.e., engagements involving a wing or battalion), but a higher-level strategic propagation of uncertainty (i.e., campaign-level models including political-economic or nuclear factors) may require an infeasible amount of computing power to calculate in real time.

Propagation of measurement uncertainty through a machine learning model as part of the input data set is less common than using statistical methods to estimate uncertainty within the model. Data scientists and AI researchers will be familiar with the mass of studies focused on postulating uncertainty within machine learning models, but much of the historical work does not take an approach of adjusting epistemic uncertainty—an insufficient amount of training data for an ML model—with measurement uncertainty in the training data set.[22]

Uncertainty of measurement can be thought of as noise in data and/or variability in the observation. Other aspects of uncertainty need to be quantified when implementing uncertainty quantification in digital systems, such as the completeness of the coverage of the domain, which is the representativeness of the input data set, and the imperfect modeling of the military problem, which is the result of incorrect baseline assumptions during model development and is ultimately rooted in imperfections in human judgment.[23]

A more modern approach to propagation that may be less computationally intensive may be to use machine learning to postulate uncertainty. Evidence from other disciplines using neural networks shows the inclusion of known input data uncertainty "is advanta-

19. Farrance and Frenkel, "Uncertainty of Measurement," 61.

20. Farrance and Frenkel.

21. Umang Bhatt et al., "Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty," in *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (New York: Association for Computing Machinery, July 30, 2021), 2.4 "Uncertainty Evaluation," https://arxiv.org/.

22. Abdar et al., "UQ in Deep Learning"; and Psaros et al., "UQ in ML."

23. Brownlee, "Gentle Introduction."

geous for making better predictions compared to the case of not using them."[24] These researchers also suggest further investigation into using known input data uncertainty "as the initial values of the uncertainties to be derived" in a Bayesian deep-learning framework, which would be a way to propagate empirical uncertainty in concert with statistically derived uncertainty.[25]

Using the mathematical approach to uncertainty propagation will incorporate and account for the effects of aleatoric uncertainty—the inherent randomness of data that cannot be explained—and epistemic uncertainty. The proposed military standard should enfold the requirement for measurement uncertainty with the requirement of its propagation into higher-order uses, such as machine learning or more abstract modeling and simulation. In military parlance, standardizing UQ by this approach will account for not just the baseline observational data uncertainty, but also data uncertainty related to orientation and action.

Math for Military Utility

To continue the analogy to military strategy, a functional relationship describes how military advantage is gained in the OODA loop, and how uncertainty propagates in that process.

Desired Military Effect ± Uncertainty Success = *f [observation (many variables ± u),*
orientation (some variables ± u),
speed of decision ± u
speed of action ± u]

In this purposely emblematic equation, observation and orientation are constant activities, while decisions and actions are discrete events in time. Probability of success of the desired military effect is based on the propagation of uncertainty of each of the input variables in the loop: how certain is the operator that (a) their observations capture reality, (b) they are orienting in the manner intended, (c) their decision was executed the way intended, and (d) their action has not been disrupted.

The barrier to this approach is that it requires prior knowledge of uncertainties, which is metadata that is currently not available because it is generally costly to determine in the empirical case, and because there are many acceptable methods for its generation in the

24. M. Kiani Shahvandi and Benedikt Soja, "Inclusion of Data Uncertainty in Machine Learning and Its Application in Geodetic Data Science, with Case Studies for the Prediction of Earth Orientation Parameters and GNSS Station Coordinate Time Series," *Advances in Space Research* 70, no. 3 (August 2022): 573, https://doi.org/; and see also Wojciech M. Czarnecki and Igor T. Podolak, "Machine Learning with Known Input Data Uncertainty Measure," in *Computer Information Systems and Industrial Management: 12th IFIP TC8 International Conference, CISIM 2013, Krakow, Poland, September 25–27, 2013, Proceedings*, ed. Khalid Saeed et al. (Heidelberg, Germany: Springer Berlin, 2013), 379.

25. Shahvandi and Soja.

statistical case. This circles back to the recommended solution of levying the requirement and a standard to provide the uncertainties related to each of the input variables as metadata. Once provided, AI/ML systems that compile observational and orientation data can use the metadata for propagation and provide an operator or commander with the overarching quantified uncertainty in the situational picture. When used in real time, this approach intrinsically captures facets of the decision and action steps of the OODA loop.

Higher-Order Effects and Challenges

Fairness in modeling is a well-known issue in the realm of AI/ML capability development, and a large body of work is aimed at ensuring this. Realistically, machines can assist in determining the bias in models by using quantified uncertainty, but a model is only as good as its inputs, and a human will be responsible for determining what those inputs are.[26] Models are "only proxies for the real world and their learning and inference algorithms rely on various simplifying assumptions and thus introduce modeling and inferential uncertainties."[27] Simplistically, the root cause of the uncertainty related to the truthfulness and fairness of a model is based on human psychology. This is problematic for many reasons, but these reasons already exist within executing an OODA loop for military advantage and are not exclusive to using UQ or digital information.

Computers are deterministic in that a developer writes a program and "the computer does what [they] say."[28] If a program is based on bad assumptions, a bad result is not the computer's fault. Trying to quantify how good or bad baseline model assumptions are would still be a problem within this larger UQ framework. This component of uncertainty could be based on any combination of judgment factors during development, such as the choice and preparation of data, choice of training hyperparameters, and the choice of omission. Quantifying uncertainty will only help with fairness in AI/ML models by allowing inspection; it does not necessarily make an AI or ML model fairer.

There are statistical approaches to creating fairness metrics using UQ that can be used to improve models, but the approaches still require human assumptions and decisions in development. Providing uncertainty quantification would allow inspection, and that is the first step needed for improving input assumptions, bias, and output.

Choosing the appropriate mathematical formulae for calculating propagation of uncertainty in a functional relationship requires some baseline assumptions to build the best representation of the partial differential terms. The functional relationship and resulting mapping function may be ambiguous as a result of epistemic uncertainty.[29] Determining

26. Bhatt et al., "Form of Transparency"; and Tongfei Chen et al., "Confidence Scoring Using Whitebox Meta-Models with Linear Classifier Probes," *Proceedings of Machine Learning Research* (*PMLR*) 89 (2019).

27. Ghosh et al., "UQ 360," 1–2.

28. Brownlee, "Gentle Introduction."

29. Bhatt et al., "Form of Transparency."

the correct formulae for uncertainty propagation requires further study, but this challenge does not diminish the value in implementing a UQ military standard.

One analysis has shown that communicating and visualizing uncertainty information to operators of unmanned vehicles helped improve human-AI team performance.[30] But other AI researchers have also shown that "more research is needed into how to best capture and present the developer's [uncertainty quantification] in such a way that is meaningful for the user." They further state, "Giving users seeming control over aspects they don't understand has the potential to give the illusions of clarity and informed control, cause additional automation bias, or simply allow the user to select an option that gives them the answer they want."[31] This finding moves solidly into the body of work on decision theory and psychology. There are statistical approaches that attempt to algorithmically define judgment and decision-making, and there are risks to using those approaches.[32]

A separate analysis provides relevant conclusions from the judgment and decision-making literature that pertain to using uncertainty estimates in decision-making. The study concludes that delivering uncertainty estimates to stakeholders can enhance transparency by ensuring trust formation.[33] A key consideration that the authors cover is the way in which UQ is communicated to stakeholders: "Even well-calibrated uncertainty estimates could be perceived inaccurately by people because (a) they have varying levels of understanding about probability and statistics, and (b) human perception of uncertainty quantities is often biased by decision-making heuristics."[34]

The authors further add that "both lay people and experts rely on mental shortcuts, or heuristics, to interpret uncertainty" and that this "could lead to biased appraisals of uncertainty even if model outputs are well-calibrated."[35] Unsurprisingly, key takeaways on this subject are that chosen methods of UQ communication should be tested first with stakeholders, and that developers should cater their UQ display and user interfaces to different end-user types.[36] For example, the presentation of uncertainty quantification to a data scientist should be different than the presentation of UQ to an operator for wartime decision-making. The Intelligence Community has a long history of determining the optimal method of communicating uncertainty related to military information, so its conventions for "words of estimative probability" may be an appropriate point of departure for the latter type of end user.

---

30. Kimberly Stowers et al., "Insights into Human-Agent Teaming: Intelligent Agent Transparency and Uncertainty," in *Advances in Human Robots and Unmanned Systems: Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems, July 27–31, 2016, Walt Disney World, Florida*, ed. Pamela Savage-Knepshield and Jessie Chen (Cham, Switzerland: Springer Cham, 2017).

31. Rotner, Hodge, and Danley, *AI Fails*, 44.

32. Bhatt et al., "Form of Transparency."

33. Bhatt et al., under "4 Communicating Uncertainty."

34. Bhatt et al.

35. Bhatt et al., under "3.2 Uncertainty and Decision-making"; and see also Amos Tversky and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185, no. 4157 (1974).

36. Bhatt et al.

When thinking of using propagated uncertainty at operational and strategic decision-making levels, there is a chance that using propagation calculations may make UQ numbers irrelevant and unusable because uncertainty approaches 100 percent of the desired output in very complex systems. Incidentally, this is an interesting conclusion that may point to a mathematical proof of the "fog of war." Further investigation into calculating propagated uncertainty at very large system-of-systems levels may better illuminate this conclusion.

Yet this potential shortfall of the benefits of highly propagated UQ is not a strong enough refutation of implementing a UQ military standard. Including the metadata tags at each level allows operators to inspect what factors are contributing the most uncertainty and what factors a commander can have high confidence in, which is still very useful information. When operator bandwidth is available outside of high-stress engagements, these metadata tags allow operators to examine covariance and correlation between input variables in the functional relationship. These metadata can also be used by acquisition professionals for the evaluate-and-improve tasks, by identifying systemic error and eliminating it and identifying the worst offenders contributing to random error.

The potential for highly propagated UQ to be irrelevant also emphasizes the perpetual importance of developing sound military judgment. As in any military situation where uncertainty is very high, operators and commanders with acumen will be required for achieving military advantage. Using AI/ML to observe, orient, decide, and act faster than an adversary will only lead to victory if the actions are superior. This facet of the theory of victory is distinct from the argument for requiring, propagating, and communicating UQ in a standardized way.

Lastly, AI/ML requires input data that is a "suitably representative random sample of observations" of the domain of interest. Importantly, "in all cases, we will never have all of the observations," and "there will always be some unobserved cases" within the domain of interest.[37] Although it is more common that an AI or ML algorithm has been trained on an insufficient data set, attempting to achieve total observational coverage of the domain in a data sampling is not ideal either.[38]

When applying AI/ML to the OODA loop at a higher ops tempo, improving coverage of the domain does not necessitate more sampling, but should be achieved by more randomization in the sampling with focus on determining accurate measurement uncertainty. The study on known input data mentioned above proved theoretically and empirically that incorporating data uncertainty into the learning process for a range of machine learning models made the models much more immune to the problem of overfitting—an unacceptable ML behavior that occurs when the model fits too closely to a training data set, resulting in inaccurate predictions when tasked to evaluate unknown data.[39]

37. Brownlee, "Gentle Introduction."
38. John R. Boyd, "Destruction and Creation," September 3, 1976, https://www.coljohnboyd.com/.
39. Czarnecki and Podolak, "Known Input Data."

The problem of overfitting is not unique to machine learning and is foundationally caused by a deficient input data set. "Simply stated, uncertainty and related disorder can be diminished by the direct artifice of creating a higher and broader more general concept to represent reality."[40] This results in maximum statistical coverage of the domain with minimal intrusion on the system being observed. It also minimizes the size of the data and metadata set, which increases the computational efficiency of the UQ propagation equation in higher-order usage.

Conclusion

> *"We never have complete and perfect information. . . . The best way to succeed in [this ambiguous environment] is to revel in ambiguity."*
>
> Grant T. Hammond [41]

Implementing a military standard for quantified uncertainty metadata and developing the capability to propagate, evaluate, improve, and communicate that information will provide the most flexibility for continuing to pursue AI/ML capability for military use. Using uncertainty quantification with AI/ML systems enables mutual trust and unity within human-machine teams by developing that trust through communication, transparency, and participation in common experiences. Assurance in using AI/ML systems to achieve military objectives requires quantified uncertainty.

Tying back into concepts of military strategy, this entire framework of uncertainty quantification contributes to a winning organization. By provisioning for UQ metadata now, DoD senior leaders can continue determining best governance and policy for using AI/ML without delaying technical and engineering development. As warfighters use UQ to develop trust in AI/ML partners, the military's ability to observe, orient, decide, and act faster than an adversary will increase and ensure military advantage.

40. Boyd, "Destruction and Creation," 7.

41. Grant T. Hammond, "The Essential Boyd," American War: Rediscovering the American School of War (website), n. d., accessed March 6, 2023, https://americawar.wordpress.com/.