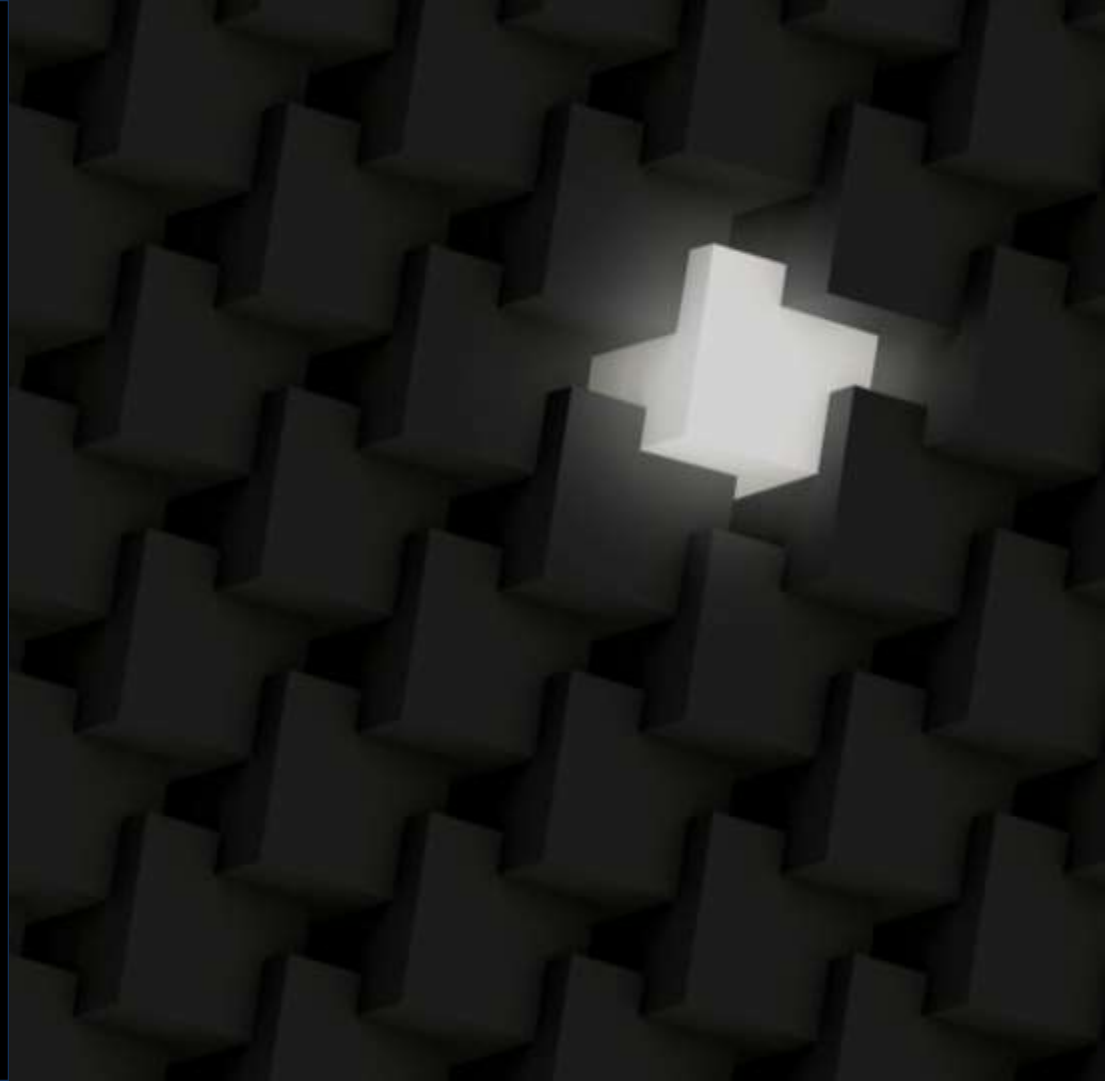
Carnegie Mellon University
Software Engineering Institute

RESEARCH REVIEW 2020

Knowing When You Don’t Know: Engineering AI Systems in an Uncertain World

Eric Heim



This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM20-0913

Why Uncertainty Matters

0.9773 Confident

0.9834 Confident

Images from the Cars Overhead with Context Data Set (https://gdo152.llnl.gov/cowc/), Lawrence Livermore National Laboratory

Why Uncertainty Matters

Guo et al.; ICML 2017

Modern Neural Networks are *drastically overconfident*

(i.e. they often predict with high confidence regardless of their accuracy)

More useful: *Calibrated* confidence (uncertainty) measures – ones that indicate how likely the model is going to produce correct inferences.

Images from the Cars Overhead with Context Data Set (https://gdo152.llnl.gov/cowc/), Lawrence Livermore National Laboratory

Why Uncertainty Matters

0.2463 Confident

0.9834 Confident

Calibrated uncertainty allows humans to compare inferences

Images from the Cars Overhead with Context Data Set (https://gdo152.llnl.gov/cowc/), Lawrence Livermore National Laboratory

In order for the DoD to leverage recent advances in AI, modern Machine Learning techniques need to be able to quantify, reason about, and rectify uncertainty in their predictions. In this work, we will benchmark modern techniques that quantify uncertainty, and develop techniques to identify causes of uncertainty and efficiently update ML models to reduce uncertainty in their predictions.

Quantify, Detect cause of, and Rectify Uncertainty in ML Models 1

Quantify – How do techniques for quantifying uncertainty in predictions practically perform?

Multiple, different approaches to quantifying uncertainty:

- *Post-training calibration* (Nieini, et al; AAAI 2015) (Guo et al; ICML 2017) (Hein et al; CVPR 2019)
- *Bayesian Neural Networks* (Blundell et al; ICML 2016)(Gal and Ghahramani; ICML 2016)
- *Deep Ensembles* (Lakshminarayanan et al; NeurIPS 2017)(Andrey et al; NeurIPS 2018)

We will compare these methods in terms of their computational run time, data efficiency, and ability to accurately quantify uncertainty.

Detect – How do we detect why a deployed model became uncertain in its predictions?

Two (of potentially many) causes for model performance degradation:

1. *Data set shift* - The distribution of data changes from that which the model was trained on.
   Our approach: Explicit detection of data set shift (Rabanaser, Gunnemann, and Lipton; NeurIPS 2019)
2. *Emergence of novel classes* – Never before seen categories of observations emerge in the deployment environment
   Our approach: Open-World Models (Cortes, et al; COLT 2016)(Rudd et al; TPMI 2017)(Oza et al; CVPR 2019)

We will develop techniques to identify the cause a model to degrade in it's certainty.

Quantify, Detect cause of, and Rectify Uncertainty in ML Models 2

Rectify – Once uncertainty is detected and quantified, how do we make models more confident in uncertain cases in the future?

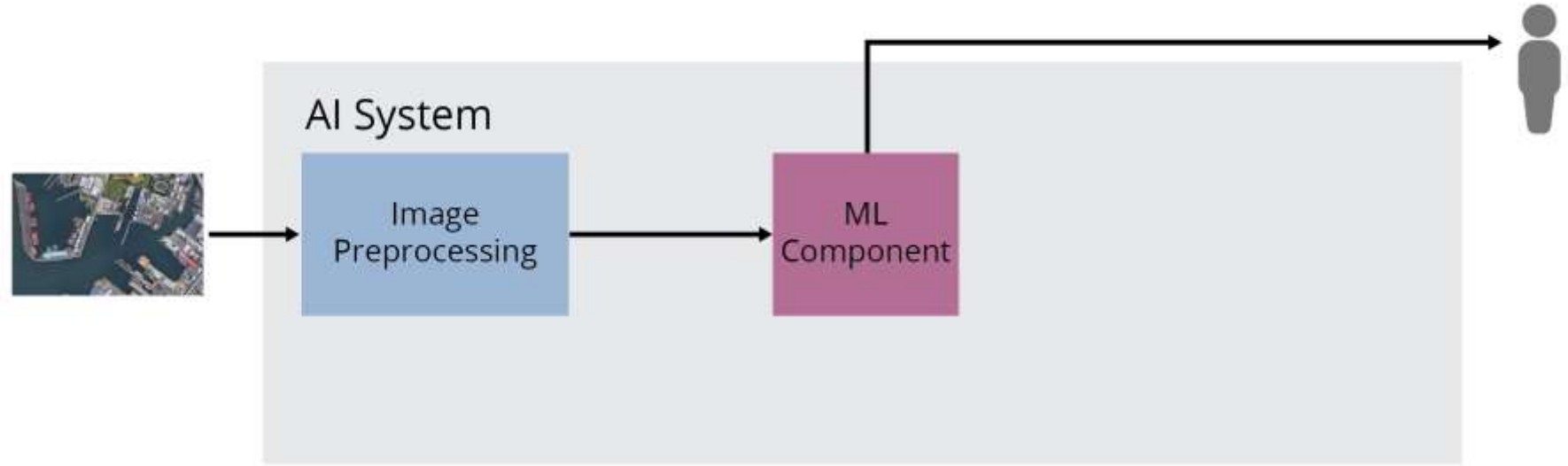
One option is to label the offending instances, and then retrain the model.

Challenges:

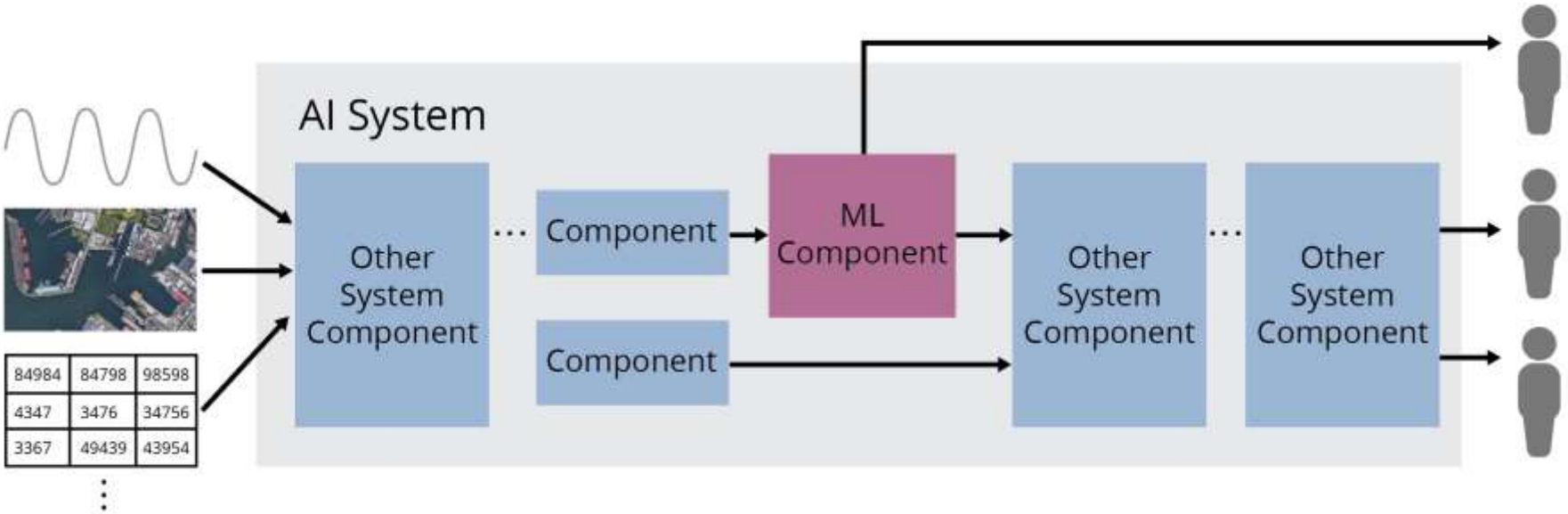
1. Labeling potentially all offending instances can be human-intensive
   Our Approach: Active Learning (Settles; 1995)
2. Retraining, validating and redeploying a model can be time-intensive
   Our Approach: Develop best practices for V&V on ML models; Online Learning where possible (Bottou; 1998)

We will develop techniques to efficiently update models to be more certain in their predictions once uncertainty is quantified and the source is identified.

Uncertainty in AI Systems: An AI Engineering Perspective

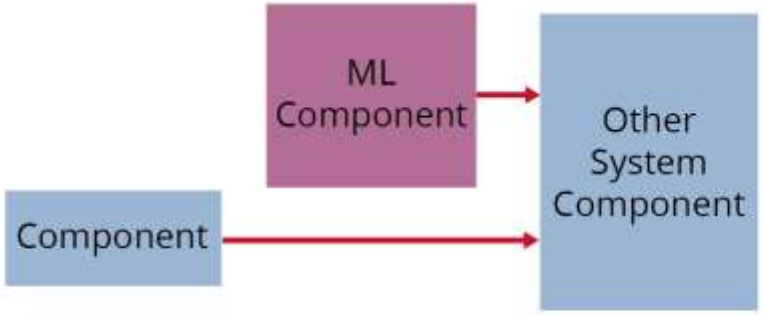


Uncertainty in AI Systems: An AI Engineering Perspective



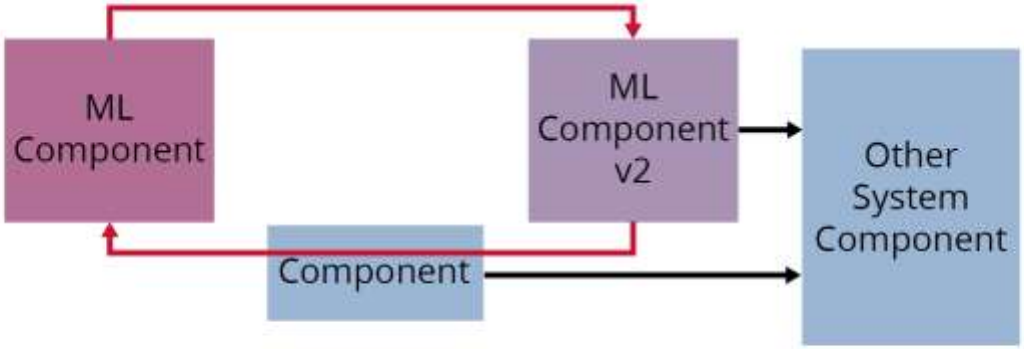
Uncertainty in AI Systems: An AI Engineering Perspective

Calibrated uncertainty from ML components can inform downstream AI system components that their inferences may not be correct, and should use contingencies. i.e. **Robustness**



Uncertainty in AI Systems: An AI Engineering Perspective

Detecting and Rectifying uncertainty enables **best practices** for iterating on ML models to be developed, providing rigor to the process of maintaining ML models.



Mission Statement and Team

In order for the DoD to leverage recent advances in AI, modern Machine Learning techniques need to be able to quantify, reason about, and rectify uncertainty in their predictions. In this work, we will benchmark modern techniques that **quantify uncertainty**, and develop techniques to **identify causes of uncertainty** and efficiently **update ML models to reduce uncertainty** in their predictions.

Through this, the DoD will be able to engineer AI systems that are more robust, and can be more reliably developed and maintained.

Eric Heim
SEI

Jay Palat
SEI

Carol Smith
SEI

Jon Helland
SEI

Zack Lipton
Tepper/MLD

Aarti Singh
MLD

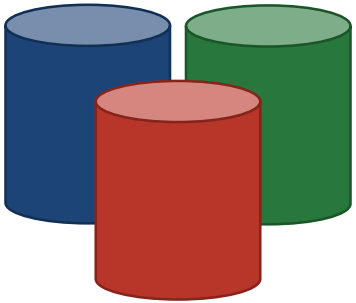
How can we work together?

An important part of this work is making sure we develop techniques to manage uncertainty in a manner that maps to the **needs of real-world DoD missions**.

We want to partner with DoD collaborators to ensure we are doing so.

If you have…

Domain Expertise



Mission-Relevant Data

Real-World Problem

…we would love to work with you! (my email: etheim@sei.cmu.edu)