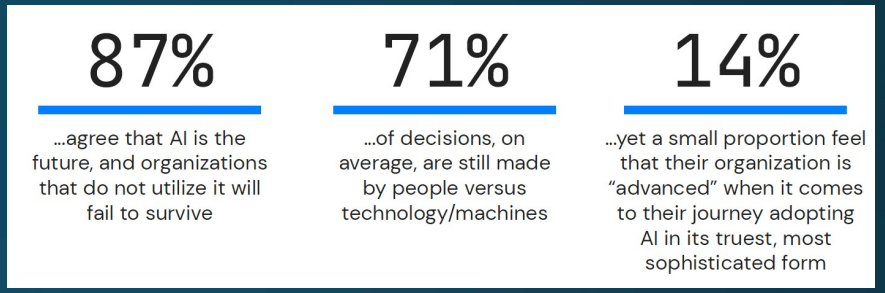
Challenges in Automating Mission-Critical Decision Making Systems: A Practitioner's Perspective

Ozgur Erdinc
Raytheon Technologies Research Center

Lack of Trust in AI models

September 2022 Survey* finds 85% of companies do not trust AI in decision making



Many surveys conclude similarly:

VentureBeat AI reports **87% of data science projects never make it into production**

NewVantage survey reports 77% of businesses report that "business adoption" of big data and AI initiatives continues to represent a big challenge for business.

Gartner says 80% of analytics insights will not deliver business outcomes through 2022 and 80% of AI projects will "remain alchemy, run by wizards" through 2020.

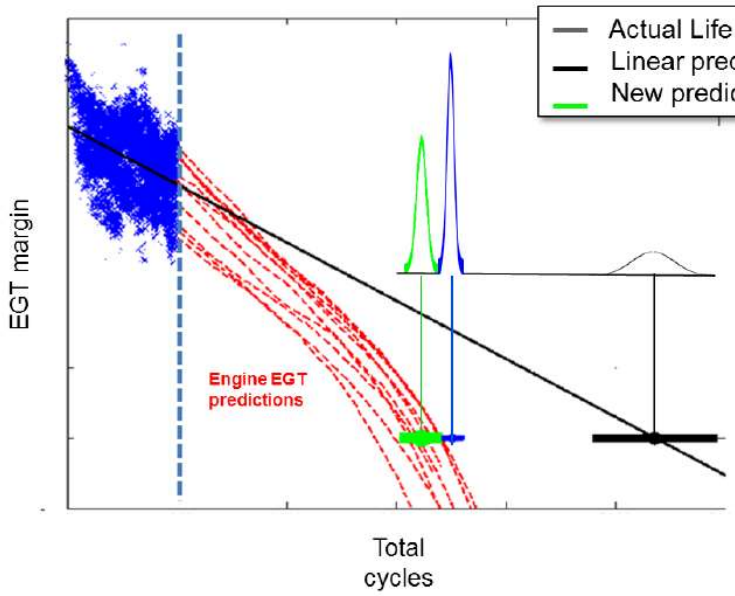
* https://get.fivetran.com/rs/353-UTB-444/images/achieving-ai-a-study-of-ai-opportunities-and-obstacles.pdf
^ https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021

Trust: "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"

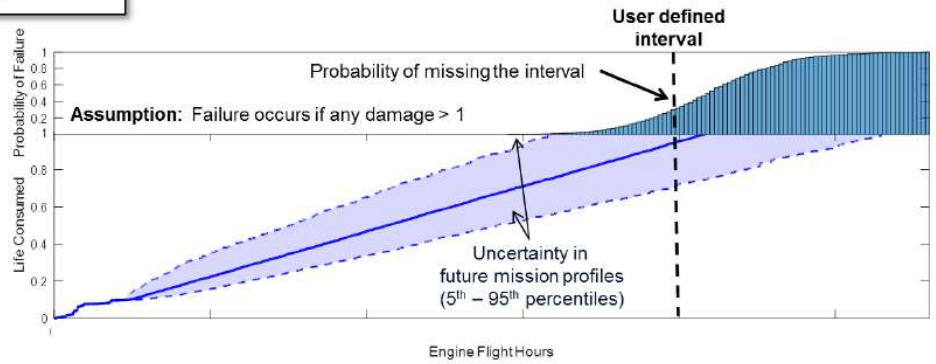
Lee, J.D., and See, K.A. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80.

LIFECYCLE MANAGAMENT OF MISSION-CRITICAL SYSTEMS WITH ADVANCED ANALYTICS

PREDICTING THE FUTURE !



Engine-Level life predictions

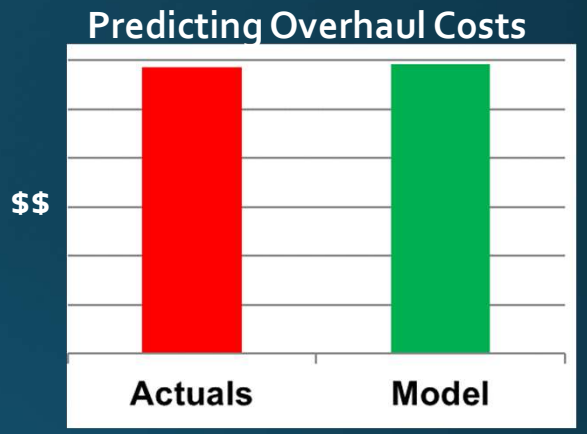


Component-level life predictions
Subsystem-level dependencies

"Prediction is very difficult, especially if it's about the future! – Niels Bohr"

Breaking barriers preventing adoption: Interpretability

Question: block-box model or interpretable model?

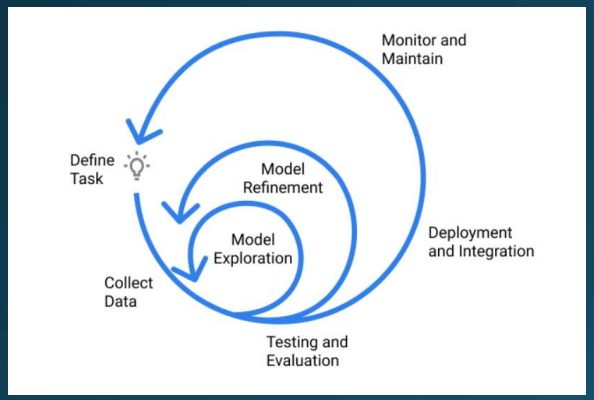


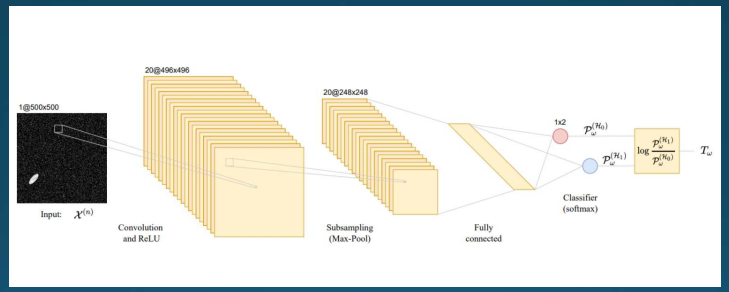
As the stakes get higher, decision makers tend to favor interpretable models.
Regulatory bodies might require interpretable models.
When physics / chemistry / biological models available, ML models should not contradict!

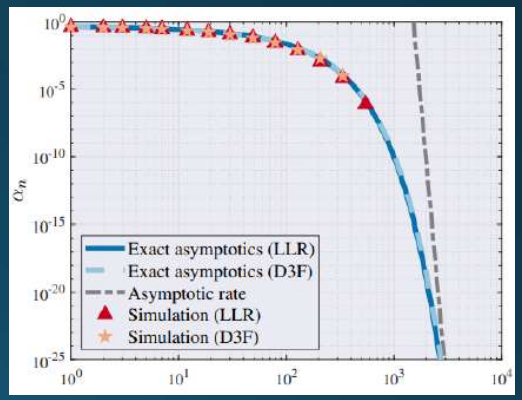
Report on Basic Research Needs for Scientific Machine Learning: https://www.osti.gov/biblio/1478744
Interpretable ML Lab: https://users.cs.duke.edu/~cynthia/papers.html
SciML: https://sciml.ai/

Breaking barriers preventing adoption: Blind Trial & Error

What is the available (Fisher) information in the dataset for a given problem?
what is the theoretical upper bound on performance ?
How many labeled data points needed to achieve a desired performance ?







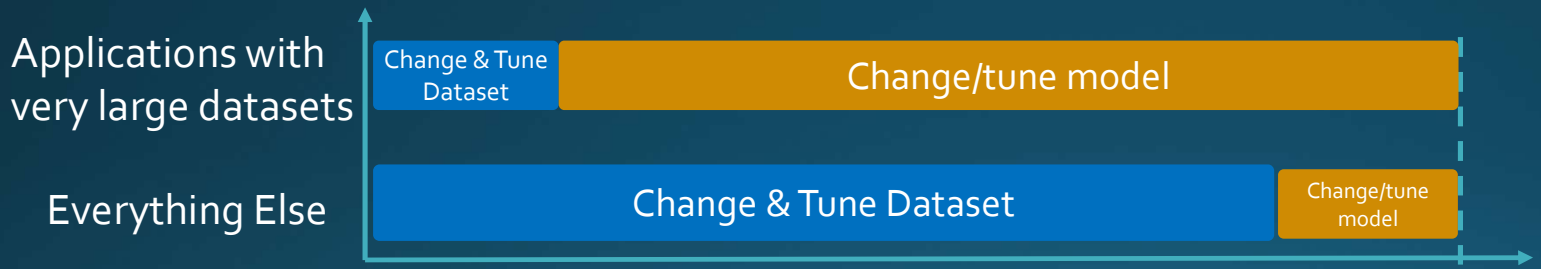
Large Deviations Analysis to predict any machine learning model's performance given a dataset.

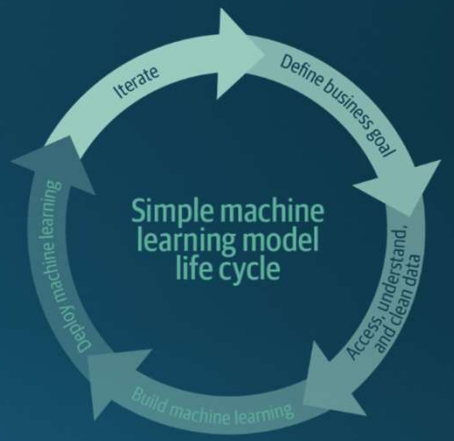
Paolo Braca, et. al. Statistical Hypothesis Testing Based on Machine Learning: Large Deviations Analysis https://arxiv.org/pdf/2207.10939.pdf

Breaking barriers of adoption: Blind Trial & Error

Focusing too much on the techniques/models, too little on the data

DATASET + ML MODEL = AI





Why does this matter? :

1) this is very pervasive mistake amongst practitioners
2) need more principled ways to determine / reduce "uncertainty" in datasets

NeurIPS Workshop on Data-Centric AI: https://neurips.cc/virtual/2021/workshop/21860
Data-Centric AI Competition: https://https-deeplearning-ai.github.io/data-centric-comp/

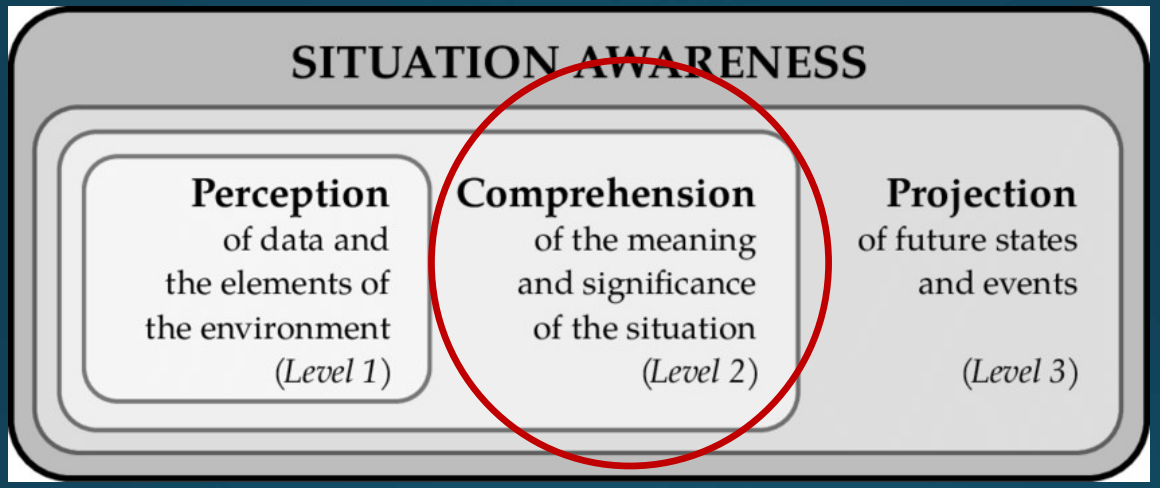
How to achieve Human-AI teaming?

Mission critical decision making support implies a joint-responsibility!

Source: https://www.netapp.com/blog/how-to-build-a-data-pipeline-for-autonomous-driving/

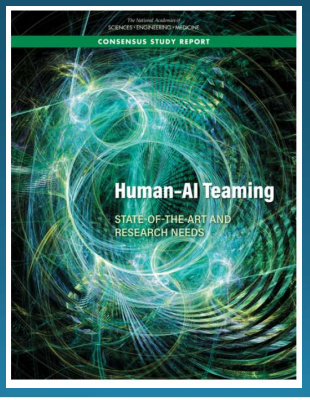
Breaking barriers of adoption: Trusting Model's behavior

Uncertainty Quantification and communication is key in Level 2 - Comprehension state of the AI

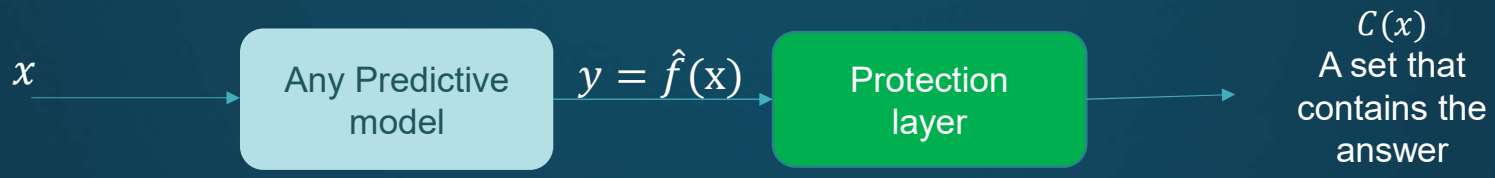


Human-AI Teaming: State-of-the-art and Research Needs:

http://nap.edu/26355



UQ for explainability / trust: Conformal Prediction Framework

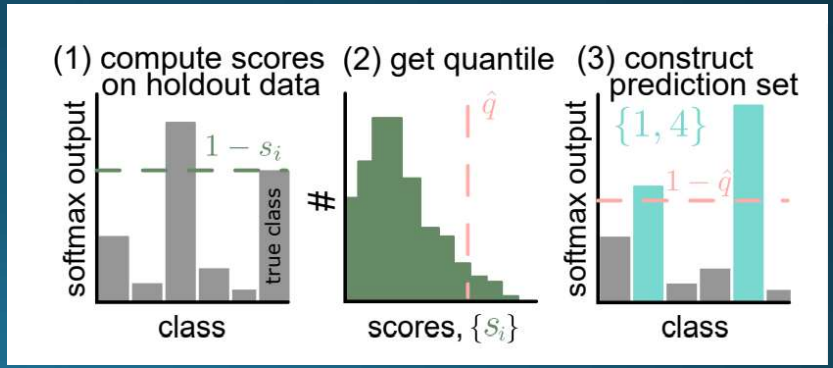


Conformal prediction (CP)*: technology to construct prediction sets $C(x)$ from training/calibration data so that on test data :

Coverage guarantees are met: $P\left(y_{true} \in C(x)\right) \geq 1 - \alpha$

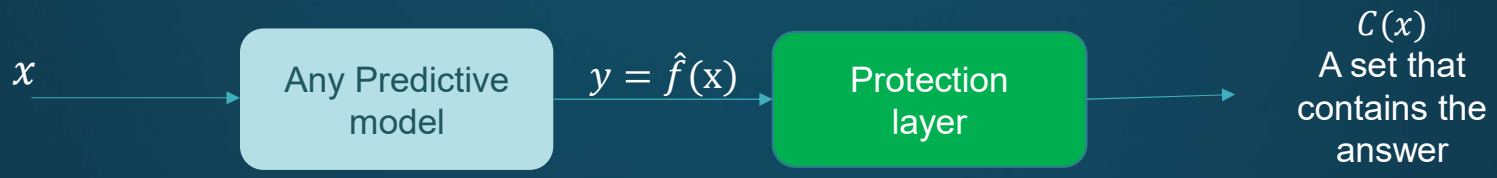
A METHOD TO CONVERT HEURISTIC UNCERTAINTY INTO RIGOROUS UNCERTAINTY

- Identify a **heuristic notion of uncertainty** for the trained prediction model: **softmax output**
- Define the **score function**
- Given **significance level** $\boldsymbol{\alpha}$ and compute class dependent $\widehat{\boldsymbol{q_k}}$ as the $1-\alpha$ quantile of the calibration scores $\boldsymbol{s(x_1, y_1), \cdots, s(x_n, y_n)}$ computed on **calibration (holdout) dataset** per class $k = 1, \cdots, K$
- At inference stage, report all classes that $C(X_{test}) = \{ y: s(X_{test}, y) \leq \widehat{q_k} \}$



Vovk, Gammerman, Shafer, "Algorithmic Learning in Random World", 2005.
Intro to CP: https://arxiv.org/pdf/2107.07511.pdf
COPA: https://cml.rhul.ac.uk/copa2021/

UQ for explainability / trust: Venn Predictors Framework



ANOTHER METHOD TO CONVERT HEURISTIC UNCERTAINTY INTO RIGOROUS UNCERTAINTY

- Define a taxonomy to divide the calibration dataset into multiple categories
- Calculate relative frequencies of labels in each category, including the test sample with all possible labels.
- Use the relative frequencies as class probabilities ranges, assuming the test sample belongs to that class (upper bound), and not (lower bound)
- Predict the class with highest lower bound

Vovk,Gammerman,Shafer, "Algorithmic Learning in Random World", 2005.
Comparison CP vs VA: https://www.sciencedirect.com/science/article/pii/S15320464193
Tutorials: https://cml.rhul.ac.uk/copa2021/

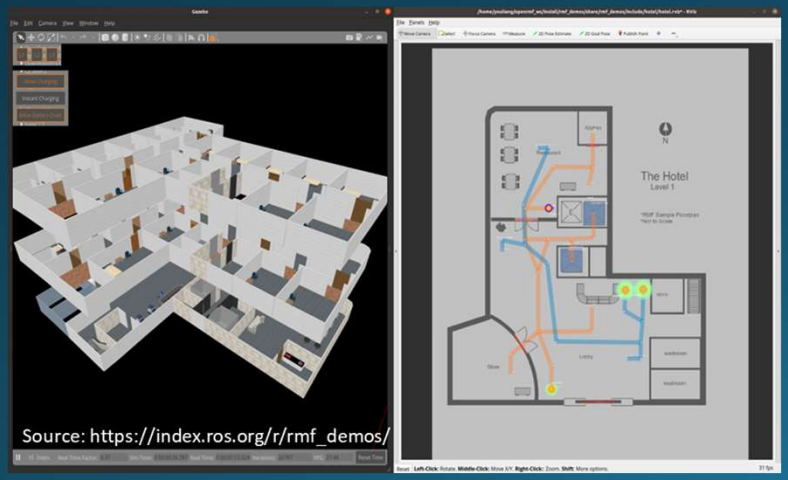
UQ for explainability / trust: other frameworks?

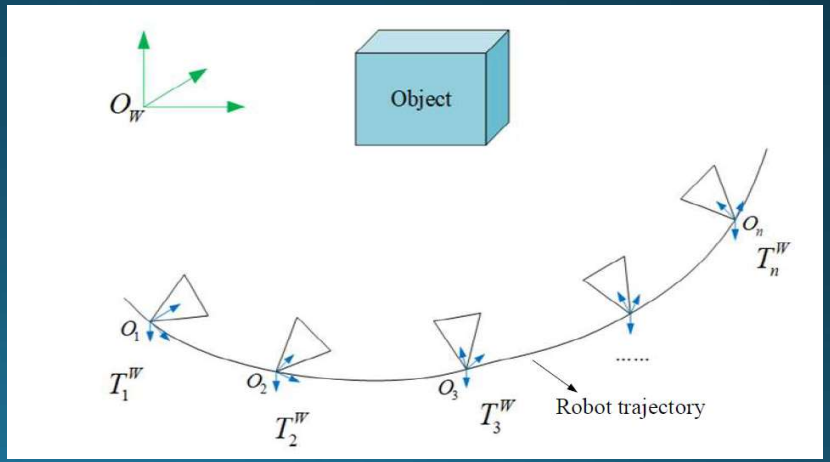
Breaking barriers of adoption: Communicating AI's Comprehension

How to communicate algorithm's findings to a human-decision maker?

Scenario:
Autonomous agents collecting video - but the decision maker need a single-page report. What information should be contained in it?

Objectives:
1. Observe multiple views of each object,
2. Fusion to improve detectability, classification/localization accuracy
3. Prevent information overload.





Beyond Interpretability – Certification / Assurance / Validation

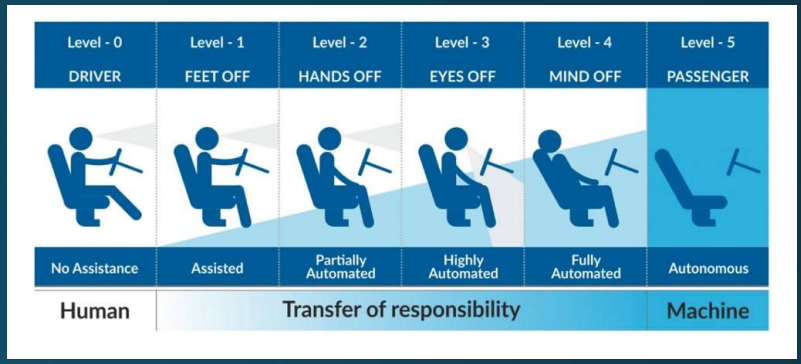
How to ensure the AI model behaves as expected at all times?

Human-AI teaming applications, one way is to revert to Intelligence Augmentation based certification

Safety-critical systems require assurance; guarantees that the model's behavior is acceptable for *all* possible input data.

Scientific ML / Physics-Informed ML can help in achieving guarantees! (self-certification)

Generalization of AI performance via risk bounds (self-certification)



Intelligence Augmentation in Non-Destructive Evaluation: https://aip.scitation.org/doi/pdf/10.1063/1.5099732
Progress in Self-Certified Neural Networks: https://arxiv.org/abs/2111.07737
AI^2: Safety and Robustness Certification of Neural Networks : https://ieeexplore.ieee.org/document/8418593
Physics-Informed ML:
https://www.turing.ac.uk/research/theory-and-method-challenge-fortnights/physics-informed-machine-learning
https://www.pnnl.gov/explainer-articles/physics-informed-machine-learning

Barriers of adoption

Human-AI Teaming Fusion Physics-based ML Validation/Verification
Data-Centric AI Assurance/Generalization UQ for Explainability
Interpretable ML Communicating AI's Comprehension

Challenges after adoption:

Life-long Learning Model-drift Data-Drift
Certification with life-long learning

Ozgur.Erdinc@rtx.com