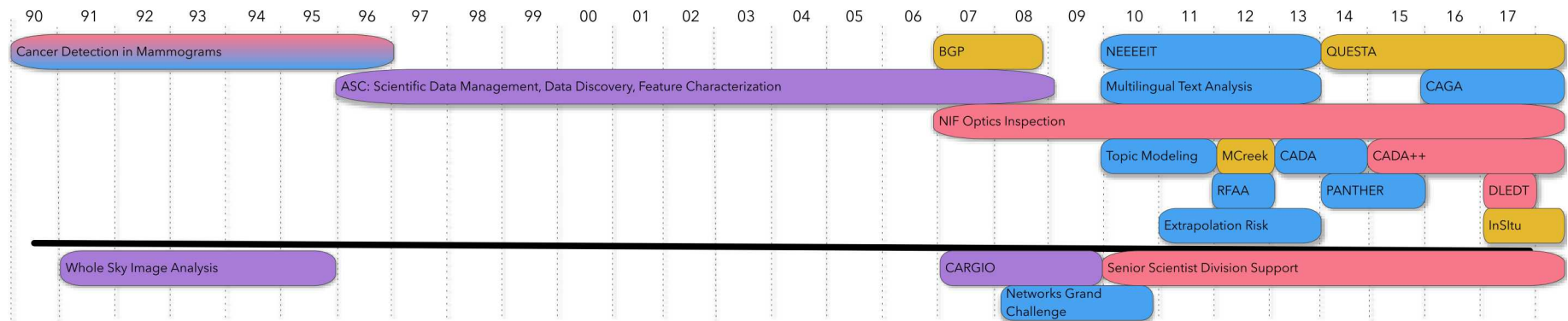


Machine Learning at Sandia: A Personal Retrospective



Philip Kegelmeyer, Sandia/CA, 8700



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.



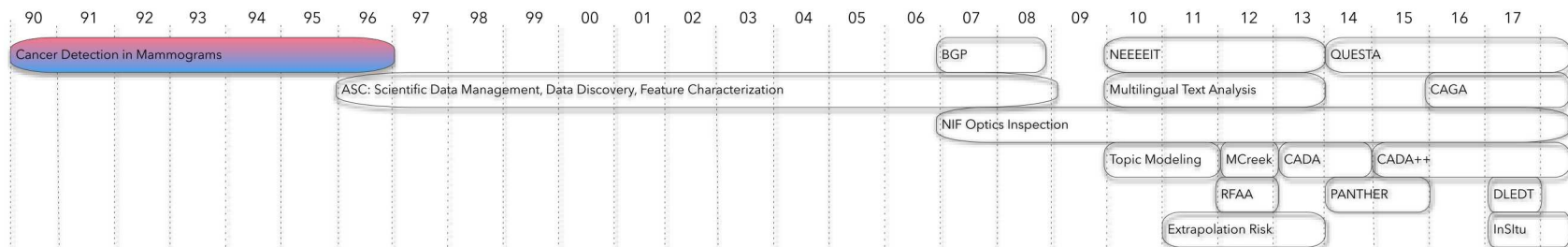
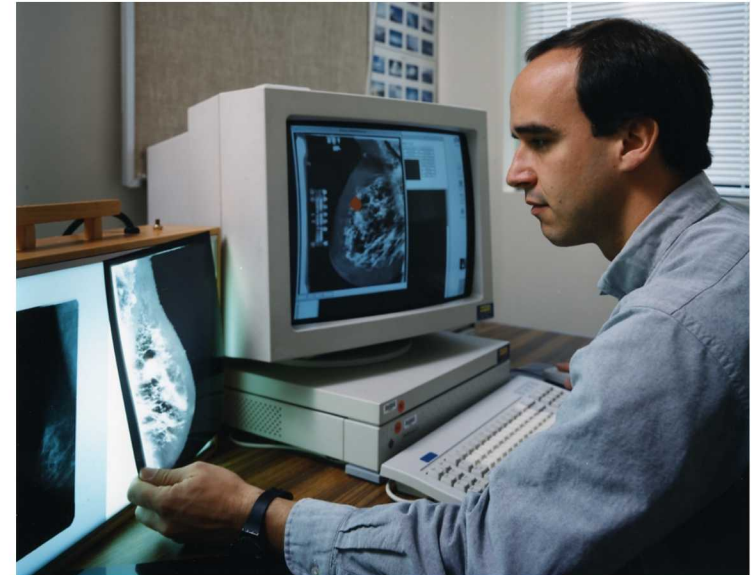
August 8, 2017

Cancer Detection

Cancer Detection in Mammograms

“Tech Maturation Funds” and LDRD:

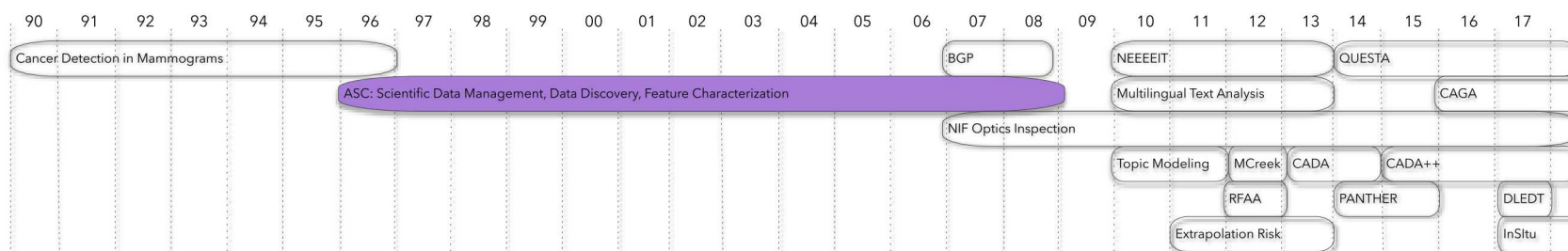
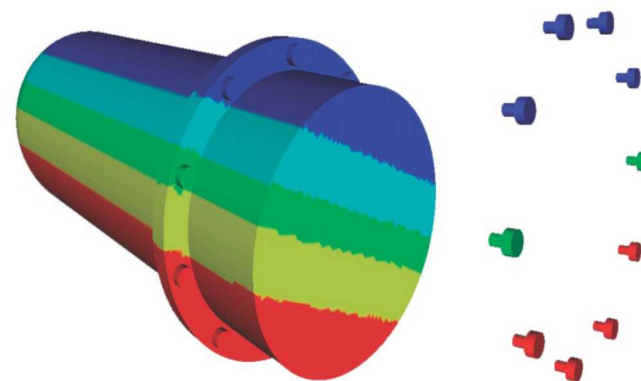
Image analysis and decision trees to detect spiculated lesions.



ASC and NW Simulation Support

Scientific Data Management, Data
Discovery, Feature Characterization

DOE ASC: Scalable machine learning
methods for analysis of physical simulation
data.

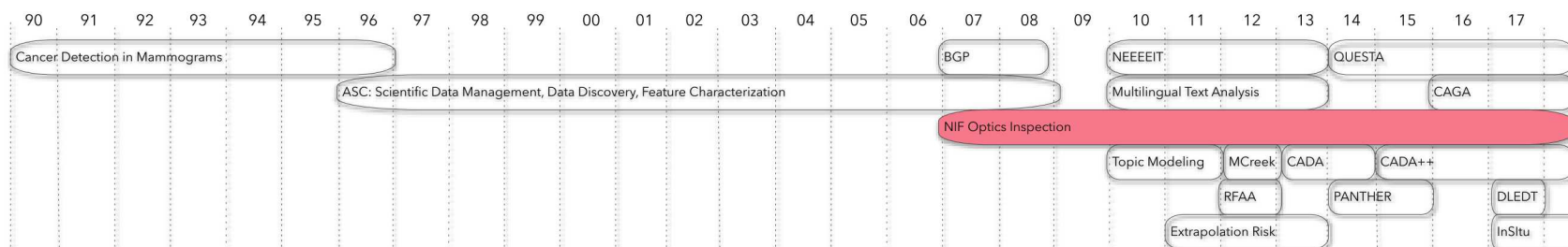
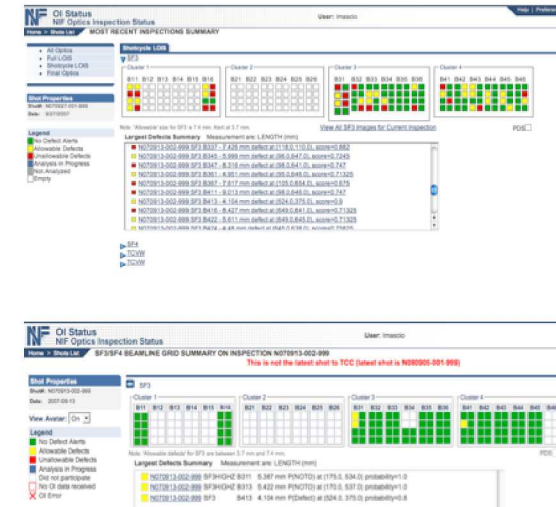


Production Deployment for NIF Optics Inspection

NIF Optics Inspection

Informal Consultation: Support AVATAR machine learning for production use at LLNL's National Ignition Facility

Before
↓
After



Text Analysis

Classifier	Train	OOB	ivoting Score	Easy/Hard	Tokens
115		✓	0.67	Hard	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
116	✓		0.67	Hard	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
117		✓	0.67	Hard	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
118	✓		0.67	Hard	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
119	✓		0.67	Hard	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
120		✓	0.67	Hard	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
121		✓	0.77	Easy	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
122	✓		0.77	Easy	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
123		✓	0.77	Easy	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .
124	✓		0.77	Easy	Russia, China, Britain, Australia, Canada, Chile and Indonesia have sent observers to this week 's exercise 230 miles northeast of Singapore .

Network and Ensemble Enabled Entity Extraction in Informal Text (NEEEEEIT)

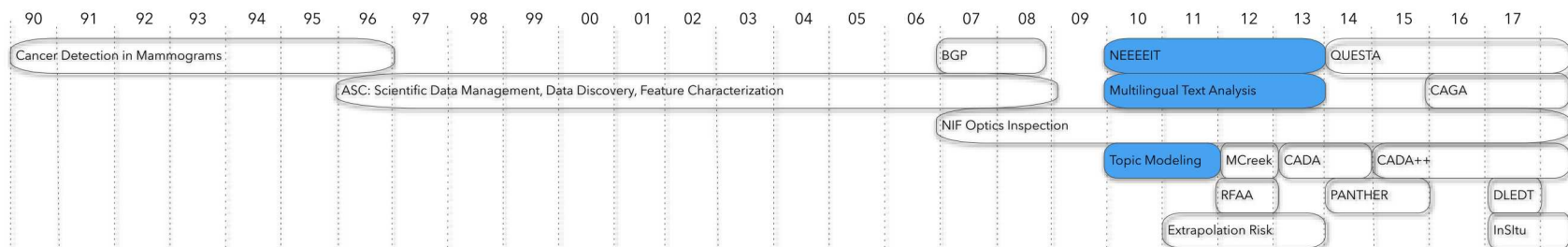
LLRD: Combine conditional random fields (CRFs) and ensemble methods.

Topic Modeling

LLRD and collaboration with LLNL: Study selection of the number of topics in LSA, LDA, and NMF.

Labeling of multilingual text

LLRD: Detect sentiment or ideology in multilingual text without translation.



Machine Learning Methods Research

Data Discovery

ASC: SMOTE, Bozos/Bites, Hellinger trees, ensemble auto-sizing.

Random Forest Accuracy Analysis

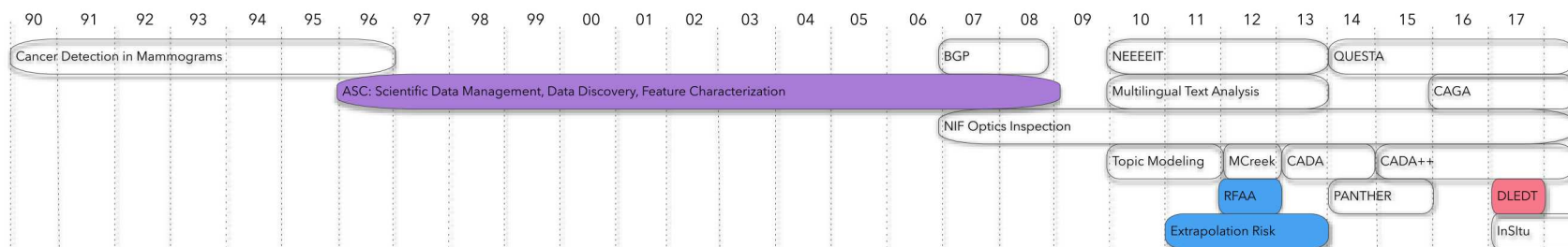
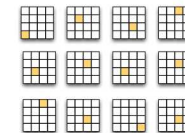
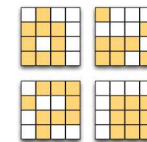
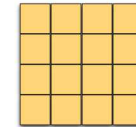
LDRD: Predictively analyze RF accuracy as a function of ensemble size.

Extrapolation Risk

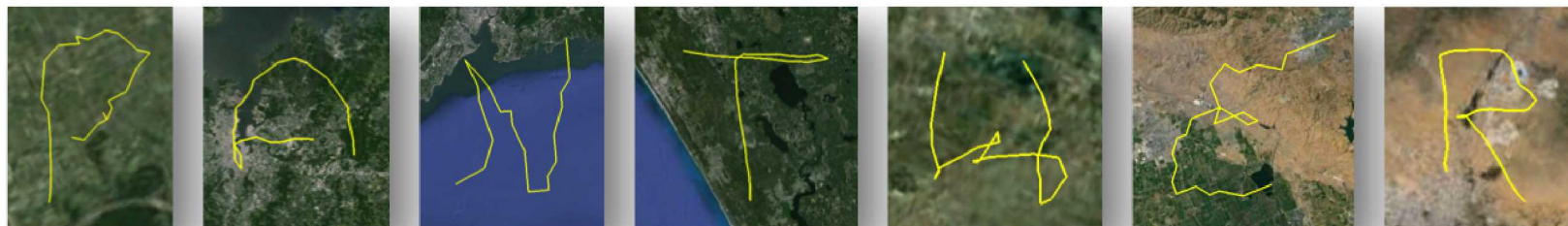
LDRD: Help ML systems understand when they are not competent to have an opinion.

DLEDT

Informal Consultation: Investigate a hybrid of deep learning and ensembles of decision trees.



Applications



Border Gateway Protocol Analysis

SPP: Use BGP traffic for early detection of network disruptions.

Mountain Creek

SPP: Analyze e-commerce transaction data for early warning of bad actors.

PANTHER

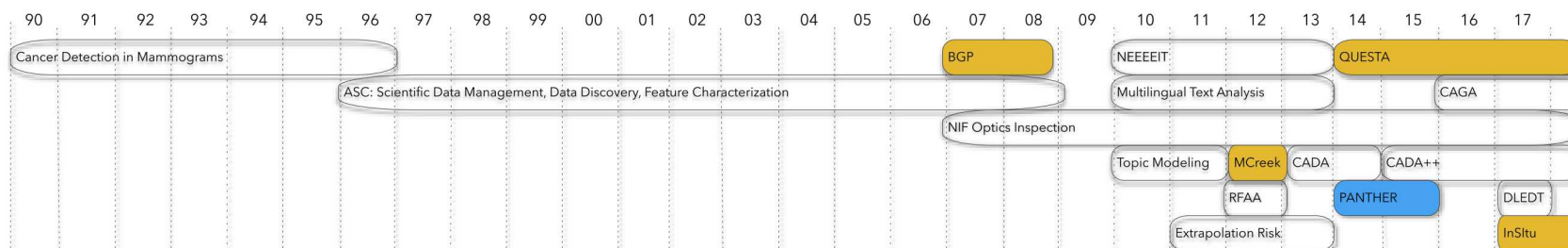
LDRD: Predict properties of aircraft trajectories.

Kachina/Questa

SPP: Build machine learning methods robust to terrible input data.

In Situ Machine Learning

DOE ASCR: Use ML to analyze simulation data while the simulations are in process.



Counter Adversarial Data Analysis

Counter Adversarial Machine Learning (CADA)

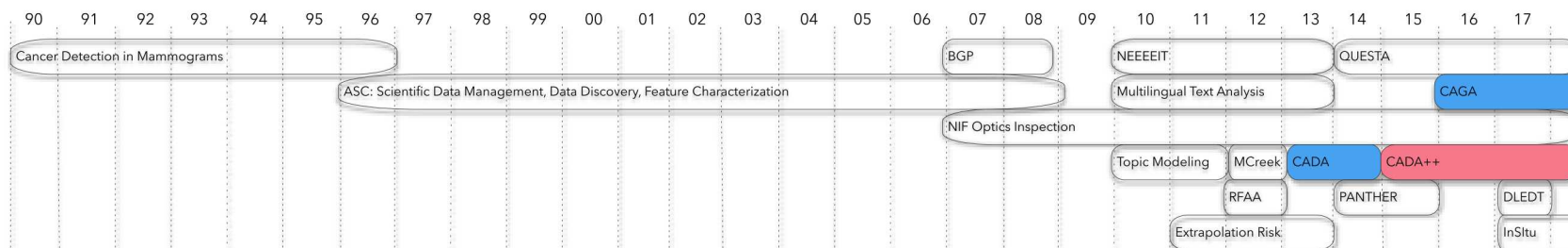
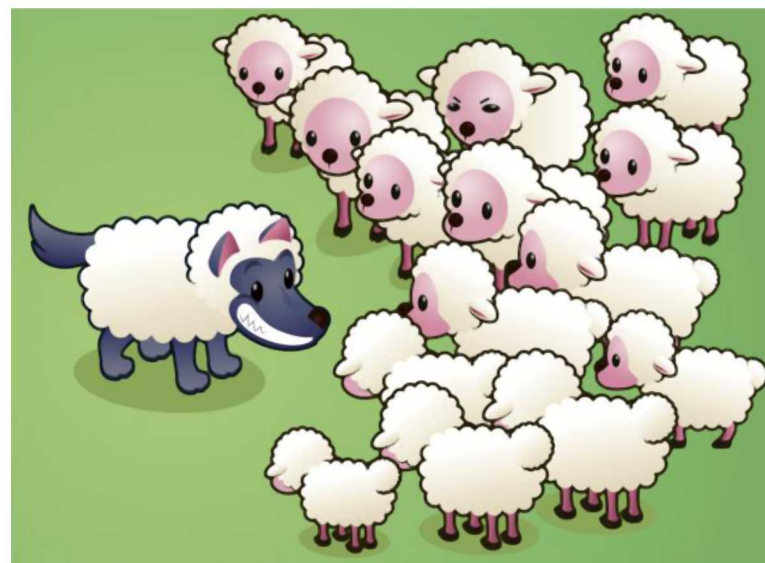
LDRD: Attack and defend machine learning

Counter Adversarial Graph Analytics

LDRD: Use ML to detect attacks around community detection or node labeling

CADA++

Program Development: Internal and external program development



Not Machine Learning

ARM/WSI

DOE: Whole Sky Imagery image analysis in support of atmospheric radiation measurement

CARGIO

ASC: “Clustering of Attributed Relational Graphs for Information Organization”

Networks Grand Challenge

LDRD: Graph analytics for national security problems

Senior Scientist

Division Overhead: Internal and external data science advocacy and strategic planning

