Artificial Neural Networks and AI in high Assurance Applications: Gaps and Techniques

Johann Schumann
KBR/NASA Ames Research Center

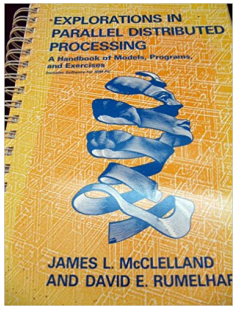
Abstract

In recent years, capabilties of Deep Neural Networks (DNN) and Artificial Intelligence (AI) systems have grown tremendously. They are now applied in many areas ranging from game playing, social media, science, to robotics, automotive, and aerospace applications.

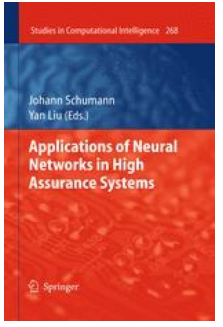
Based upon requirements for safety of DNN and AI in high assurance automotive and aerospace applications, I will discuss the necessity to ensure that AI technqiues for the analysis of Earth observation data and reasoning are working correctly and reliably.

In this talk I will present modern techniques for the verification and validation (V&V) of DNN and other AI components as well as approaches for interpretable AI. I will discuss how these techniques can help to ensure quality of the AI results, improve confidence in their application, and facilitate human-AI interaction and collaboration.

Artificial Neural Networks



- Have been around for a long time…
- With the advent of GPUs and convolutional deep neural networks (DNN), applications multiplied:
  - Automotive, Aircraft, Medical, Process Control, Natural Language, Games, Social Media, …
- Neuromorphic chips provide new capabilities
- DNN often used synonymously with "Artificial Intelligence" (AI)
- How about safety, certification, ethics?

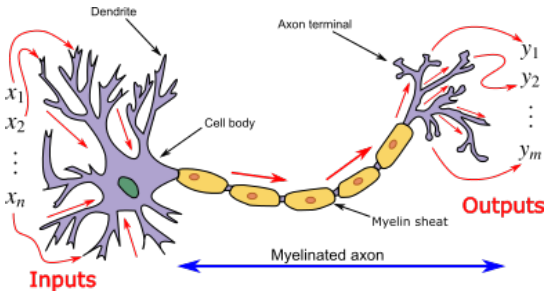


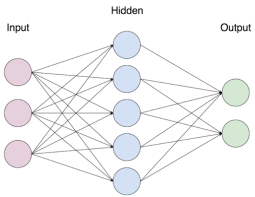
What is a Neural Network?

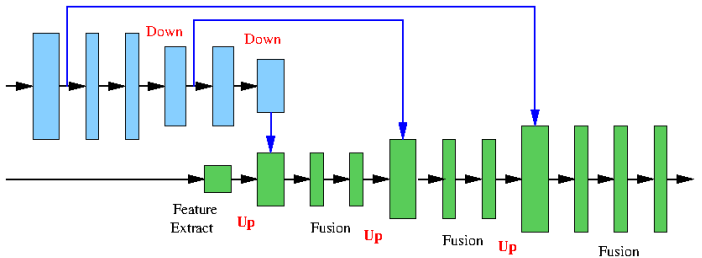
- An artificial brain in the computer?
- The coolest thing since sliced bread?
- A nasty nondeterministic piece of software?
- A high-dimensional lookup table?
- A numerical quadratic optimization algorithm?
- A Kalman filter for estimating function parameters/gains?

✓ *None of the above*
✓ *All of the above*

Artificial Neural Networks







- A mathematical abstraction of neurons and synapses
- A structure that can approximate high-dimensional functions
- Estimation of the *weight* parameters using iterative optimization and machine learning algorithms
- Mulitple layers and convolution suited for image processing and understanding

Analytical and Operational AI

- *Analytical AI* analyzes data and presents results to humans
  - Face or speech recognition, stock market prediction, ...
  - EO data analysis
- *Operational AI* analyzes data and controls system
  - Self-driving cars, drones, weapons, ...

Applications of *Operational AI* can be *safety critical:*
Failures can endanger human life

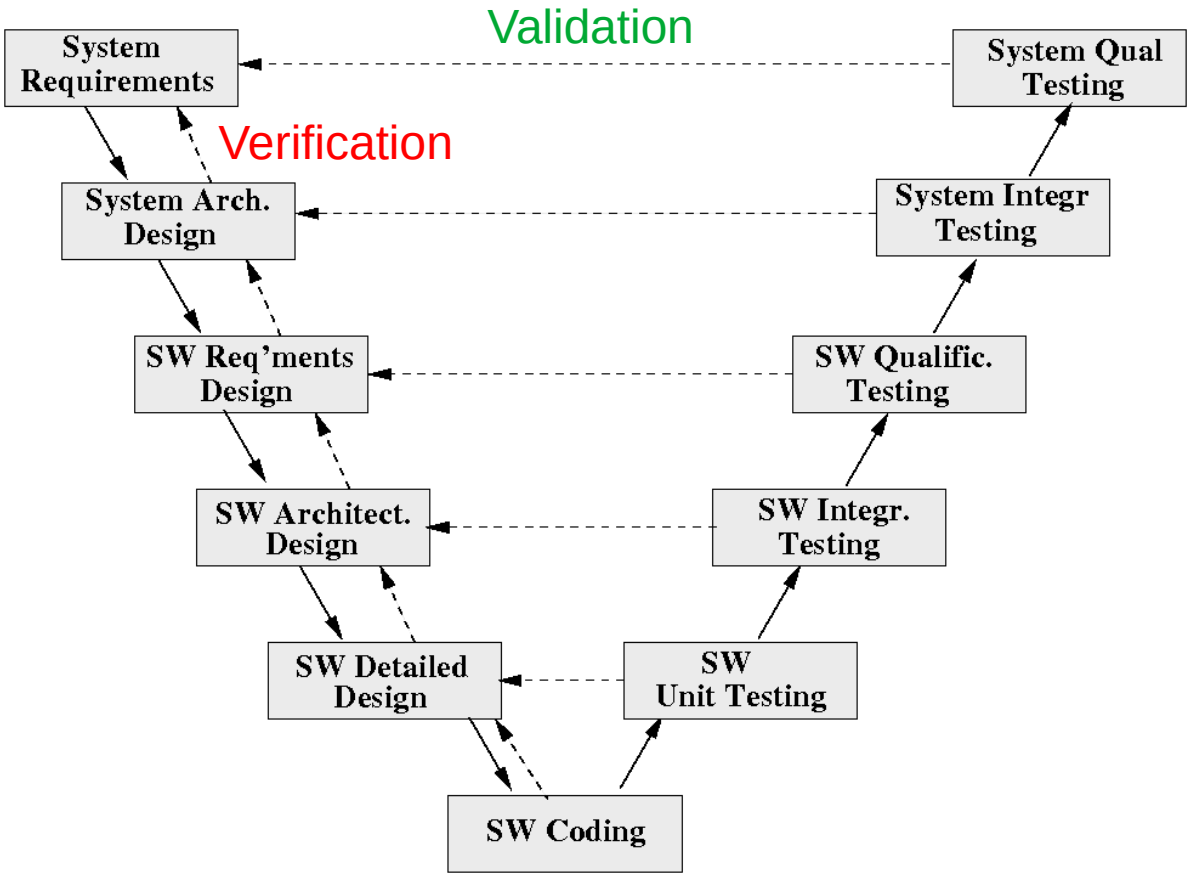
Can the application of *Analytical AI* be safety-critical?

Verification & Validation (V&V)

- Safety-critical software must undergo careful V&V
  - Verification: are we building the system right?
  - Validation: are we building the right system?
- V&V of safety-critical software is extremely complicated and costly.
- Often, safety-critical SW must be certified against a certain standard (e.g., DO-178C for aircraft)
- V&V of DNN and AI is still in its infancy but an active research topic

V&V of Software – traditional

- V&V activities at all stages of the process
  - Requirements
  - Design
  - Implementation
  - Testing



V&V of Data Analysis – traditional

Statistical Model

Implementation In Matlab with packages

Tesing with reference data Calibration

Is my Mix-Gaussians Running OK?

System Requirements

System Arch. Design

SW Req'ments Design

SW Architect. Design

SW Detailed Design

SW Coding

SW Unit Testing

SW Integr. Testing

SW Qualific. Testing

System Integr Testing

System Qual Testing

V&V of Data Analysis – with DNN

Cool Idea

DATA

12 layers
256x256 Tensors
PReLU

Click to TRAIN

System Requirements

System Arch. Design

SW Req'ments Design

SW Architect. Design

SW Detailed Design

SW Coding

SW Unit Testing

SW Integr. Testing

SW Qualific. Testing

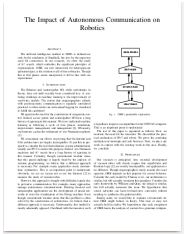
System Integr Testing

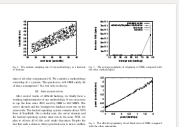
System Qual Testing

?

Real DATA

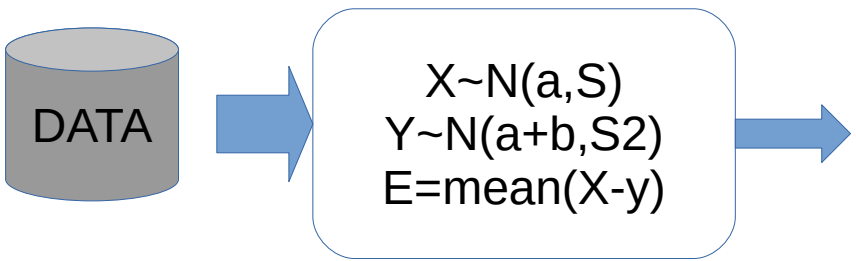
Trained **D**NN





Data Analysis: Traditional vs DNN

- Traditional
  - User develops *model*
  - System is produced by direct implementation or generated code
  - Calibration with *test data*
- DNN
  - Start with *training*
  - "No model neces
  - "No implementati
  - System is genera
    learning using traini nd test data



WHITE BOX

BLACK BOX

Which would you trust? How to V&V?

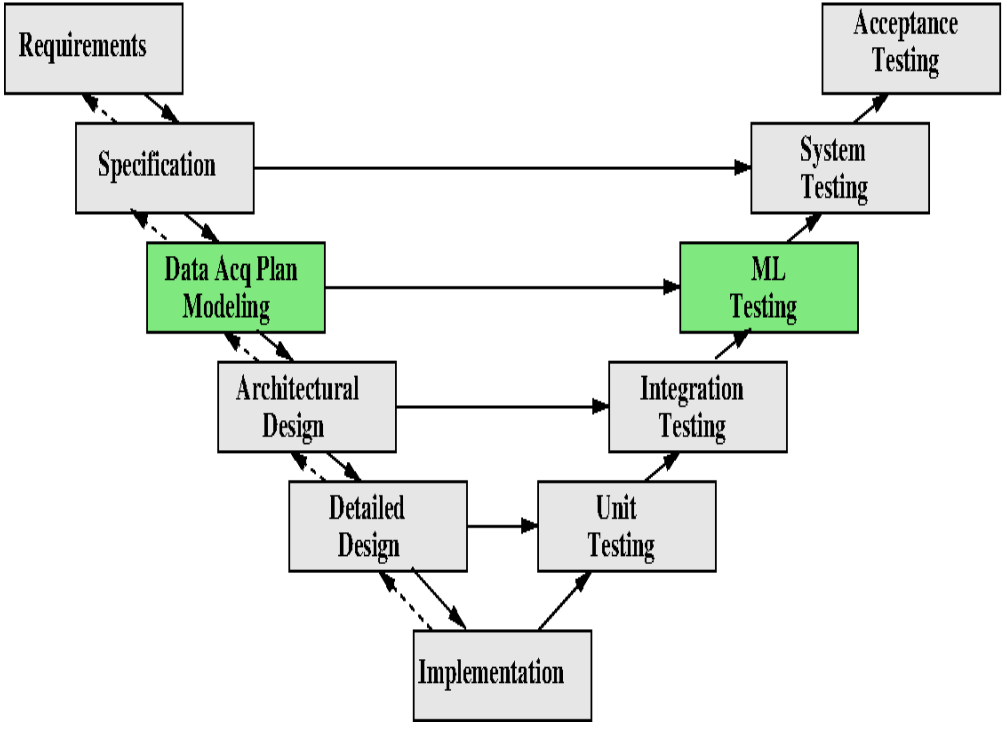
Toward V&V of DNN: Data

- Traditional Model
  - 100s of named parameters “growthrate_wheat”
- Neural Network
  - Information kept in gigantic tables (“weights”) with millions of parameters
  - Tables are generated during training using
    - LARGE sets of “training data”
    - Complex (nondeterministic) (stochastic) optimization algorithms

Toward V&V of DNN: Data

Can we trust the quality of the training data?

- Data should be considered *first-class citizens* like software
- The “V” should be extended by
  - Assurance of (training) data quality and coverage
  - Maintenance of training data
  - Assurance of machine learning
  - V&V of Architecture (e.g., number layers, number nodes) and training algorithms



V&V Approaches

- Statistical Analysis of Training Data
- Adversarial Techniques
  - Trick DNN to deliver the wrong result
- Formal Methods and Model Checking
- Bayesian Runtime Techniques (Confidence Tool)
- Rule Extraction for interpretability/explainability

Assurance of “exotic” Things

- How to assure “ethical” behavior of operational AI?
  - 3 laws of robotics? Kill pedestrian, child, or passenger?
- How to assure safety of complex AI-driven operations?
  - Run-time monitoring to bound safe behavior (currently for aircraft)
  - Human-in-the-loop as “safety person”
- How to assure that the AI works in an unknown environment?
  - From “EO” to “MO”
- How to assure AI-human interaction and collaboration?
- How to avoid/control/assure “emerging behavior”? (aka Skynet or the Matrix)

Conclusions

- V&V and assurance is important for AI in scientific Data Analysis
- Many approaches toward V&V of DNN and AI in general
- Many questions to be addressed
  - What are the right requirements?
  - How can we V&V/assure the training data and DNN performance?
  - How can we V&V/assure on-board training?
- Even more questions
  - Can explainable AI help? How to do that?
  - Ethical questions (safety-critical and non-safety-critical)
  - How to design and assure human ↔ AI collaboration/interfaces?
  - How to keep AI under control