



How we Innovate Data Science at ASRC

Date: 2/26/2022



Innovations Data Science Group

The Federal Government chooses ASRC Federal to create enduring solutions for strategic agency missions, from space to national security and defense to citizen services.

We offer broad technical expertise in digital operations and IT modernization, software, applications and analytics, engineering solutions, professional services, and infrastructure operations.

The **Innovations Data Science** group is an internal research & development team responsible for creating new AI/ML technologies to support the military and DoD.

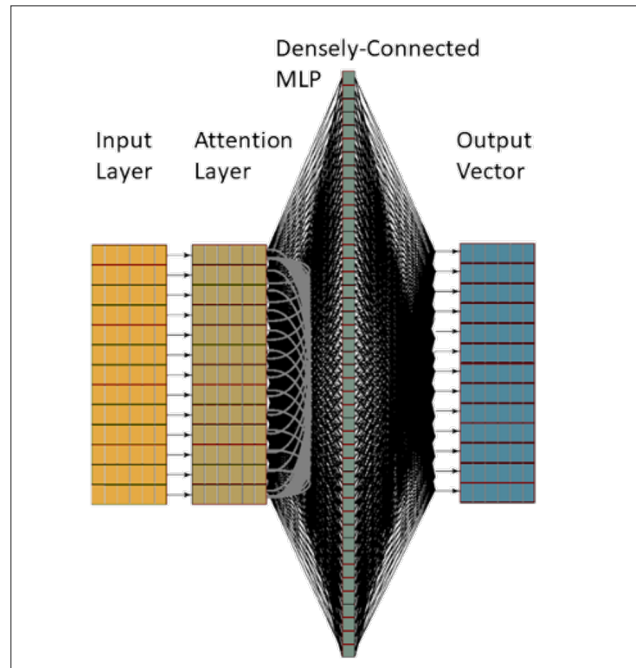


Process Overview

Our development process takes current state-of-the-art academic and commercial technologies and applies them in innovative ways to solve difficult problems for government customers.

- **Develop** IRAD-funded reference implementations as demonstrable Proof-of-Concepts (POC).
- **Publish** unique technical concepts in peer-reviewed academic journals/conferences.
- **Support** new business acquisition through customer engagement, and RFI/RFP submissions.
- **Implement** customer specific (e.g. government owned) code on contract, using techniques learned from our reference implementations.
- **Integrate** lessons-learned from the customer use-case into the reference implementation under IRAD.

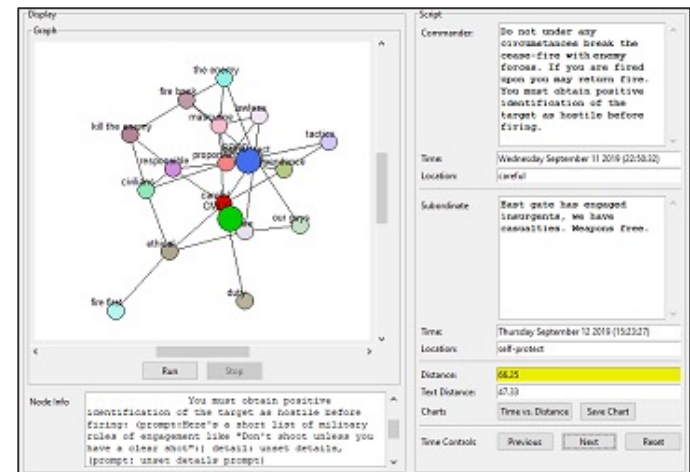
Develop



SimAccelerator



Multi-Domain Battlespace Environment (MDBE)



Neural Narrative Mapping

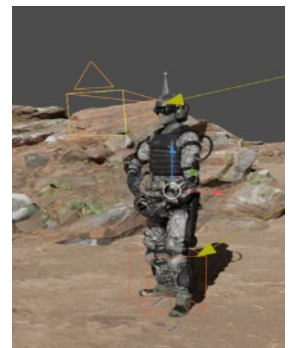
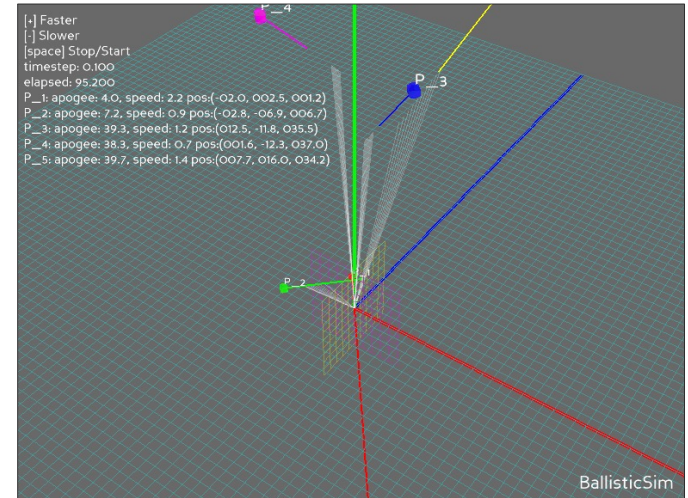
Sim Accelerator

Why do we need Synthetic Data?

Testing systems or training AI/ML models cannot be done for many government use cases due to untagged, noisy, biased, or scarce data

Poor quality or limited scarce data results in brittle or inaccurate models

We generate diverse synthetic data at scale to train AI/ML systems with realistic variance and noise



Sim Accelerator: What is it?

High-fidelity Modeling and Simulation (M&S) consumes a lot of **time and resources**. However, we can rapidly and accurately augment simulations with synthetic data through a combination of technologies.



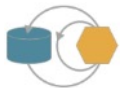
Attention MLP

- Accelerates M&S by training a deep neural network to convert low-fidelity software simulation (which could be run extremely quickly and in parallel) to high-fidelity outputs for a massive increase in speed, with no loss in fidelity



Opt Evolver

- Uses genetic algorithms to identify the best hyperparameters and neural network structure for the use case.



Self-Directed Training

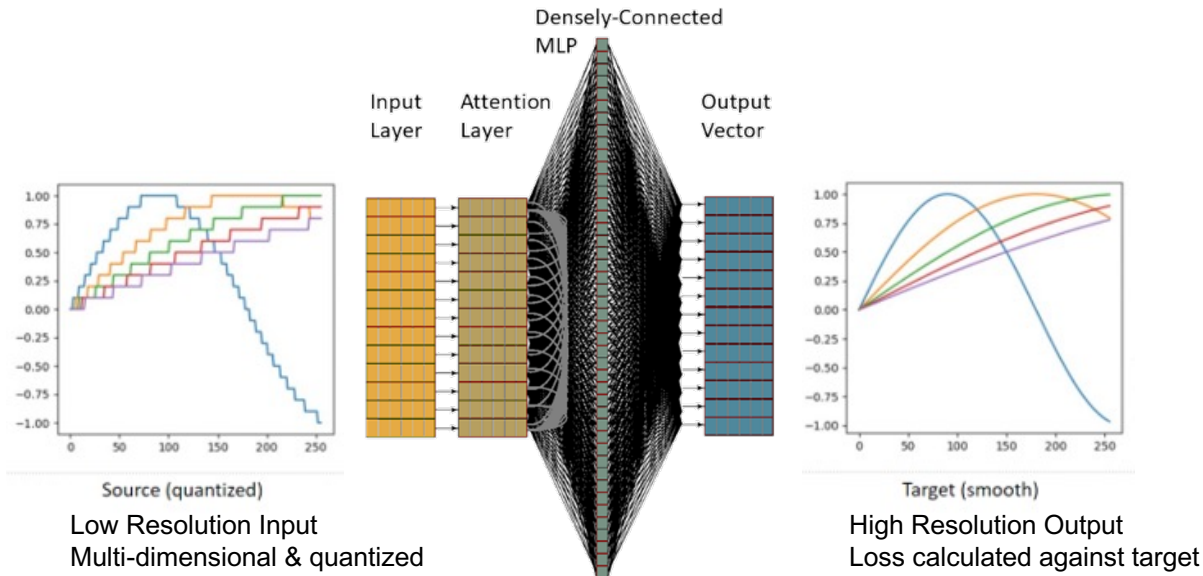
- Utilizes self-directed feedback to generate only the minimum essential synthetic data needed per class while training a model, improving the overall speed of data generation and training



Simulation Accelerator



Sim Accelerator: *Attention-MLP*



- Attention-MLP Neural Network generates **high-fidelity data** from **low-fidelity inputs** at **high speed**
- Interpolates high-fidelity multivariate data to generate complete environment simulations with initial speed improvements of **1600%** and over **97%** accuracy. With more training and GPUs this **improvement should grow exponentially**.



Sim Accelerator: *Opt Evolver*

Genetic Algorithm

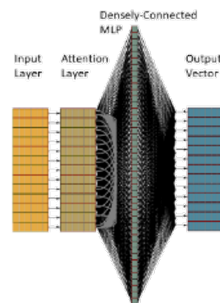
- Iterates over hyperparameters keeping a certain percentage of good "genes"
- Searches for near-optimal combinations of hyperparameters
- Flexible implementation explores hyperparameters, architecture, etc.

Ensemble Method

- Trains multiple models and merges them into an ensemble
- Saves model files and sample graphs to SQL for later inspection

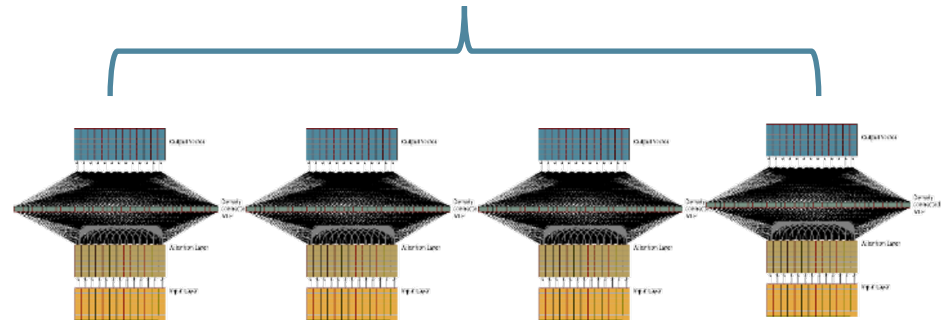
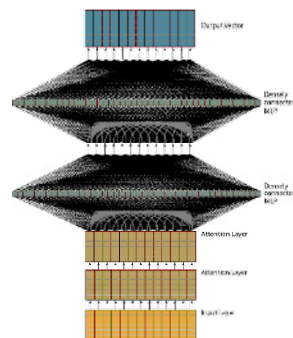
Gen1

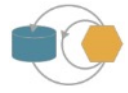
Lr=0.001
batch_size=16
etc



Gen2

Lr=0.0001
Batch_size=32
etc





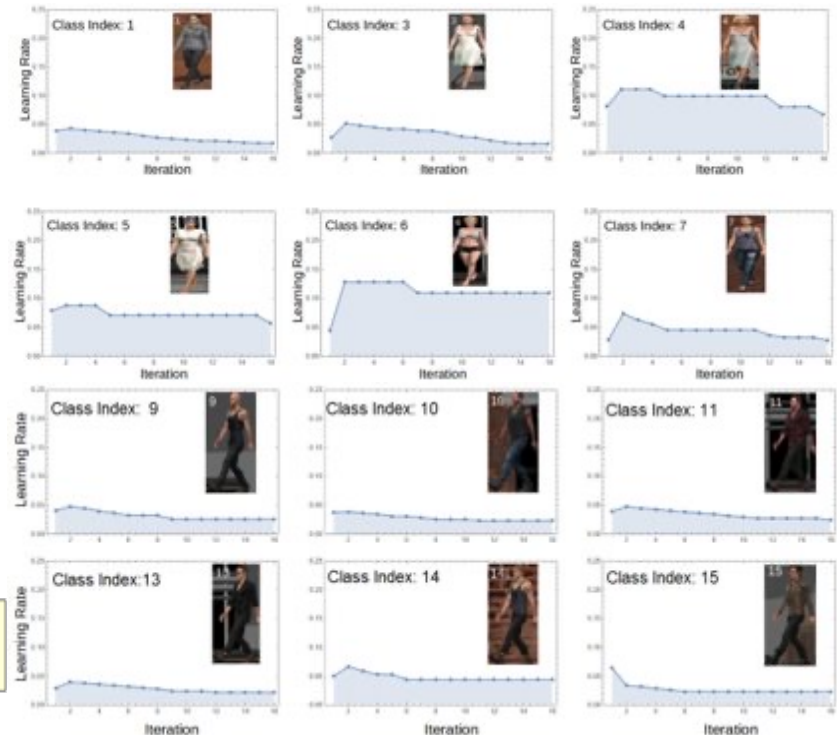
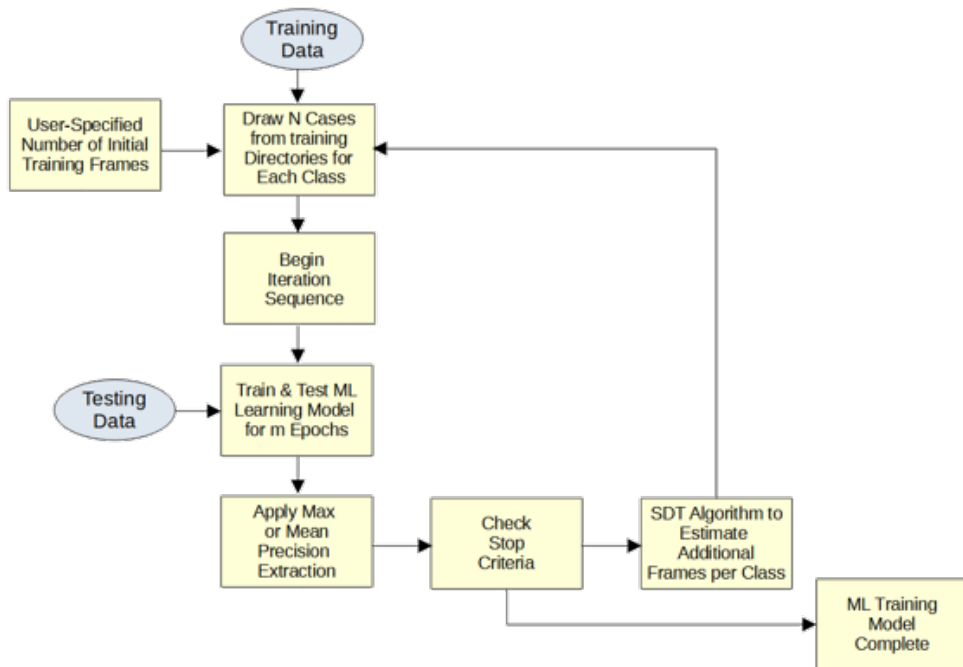
Sim Accelerator: *Self-Directed Training*

As a model trains, classifications **converges at *different rates***

- Identifies underperforming classifications
- Alert us to classes that need additional data

Better performance with less data

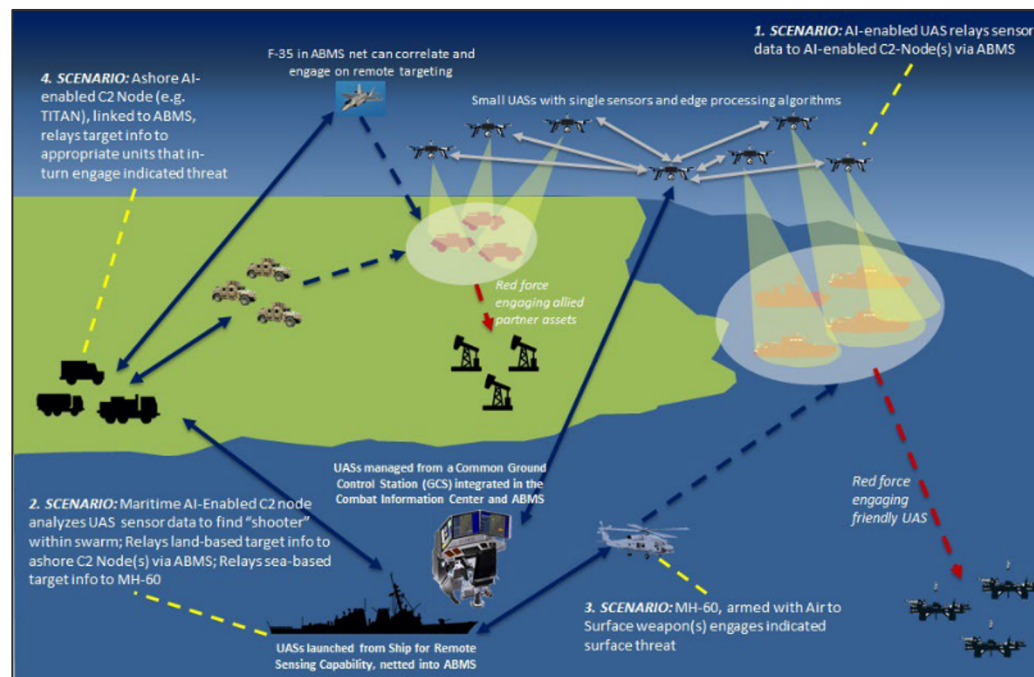
- **70.96%** less processing time
- Minimizes training data needed (**77.90%** less)
- Manages real-time generation of synthetic data.





Multi-Domain Battlespace Environment

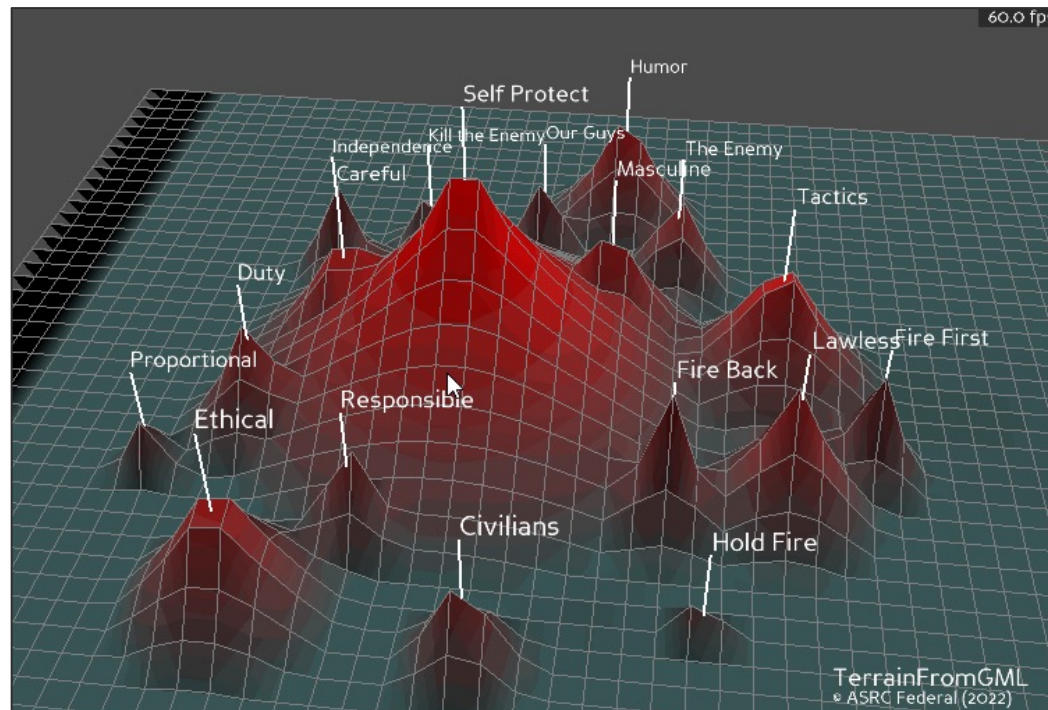
- A “**sandbox**” for training AI/ML models
 - Adversarial Reinforcement Learning to explore large domains
 - Builds on work by DeepMind (AlphaStar) and OpenAI (OpenAI Five)
- Generated models provide Optimal Course of Action (CoA) recommendations with **rationale**
- Uses scaled-Azure-batch compute with **SimAccelerated** modeling



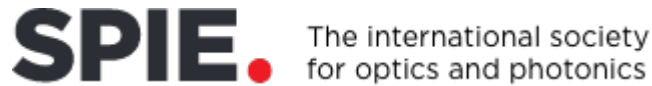


Neural Narrative Mapping

- Build maps of **human belief space** using GPT-3+
- Provides new methods of **visualizing the cyber-human battlespace** and assist with tracking patterns of group think or sympathy
- **Analytics and visualization for:**
 - Detecting “**adversarial herding**”
 - Plotting group/human **trajectories**



Publish



Publication provides

- First-mover advantage, and bona fides when we approach industry partners or target customers to implement our technique.
- Academic publications directly contributed to winning multiple contracts.

2021 Publications:

- International Workshop on Social Sensing ([SocialSens 2021](#)) - Analyzing COVID-19 Tweets with Transformer-based Language Models
- SPIE Optical Engineering + Applications ([Applications of Machine Learning Conference 2021](#)) - Self directed training of person re-identification with synthetic data
- Space & Missile Defense (SMD) Symposium – Enhanced simulation with machine learning (based on MDA20-001 SBIR work)

2022 Publications:

- [IEEE Data Engineering Bulletin on Responsible AI and Human-AI Interaction](#) - Ethics, Rules of Engagement, and AI: Mapping and Locating Beliefs Using Large Transformer Language Models
- NAACL 2022: Polling Latent Opinions: A Method for Computational Sociolinguistics Using Transformer Language Models

Support



Innovations Subject Matter Experts (SMEs) provide support through:

- Training staff in new capabilities.
- Providing marketing demonstrations for core technologies. (e.g. SimAccelerator, or Multi-Domain Battlespace Environment)
- Meeting potential customers to answer questions and drive interest in technology.
- Contribute to RFI/RFP submissions through solutioning and technical volumes.

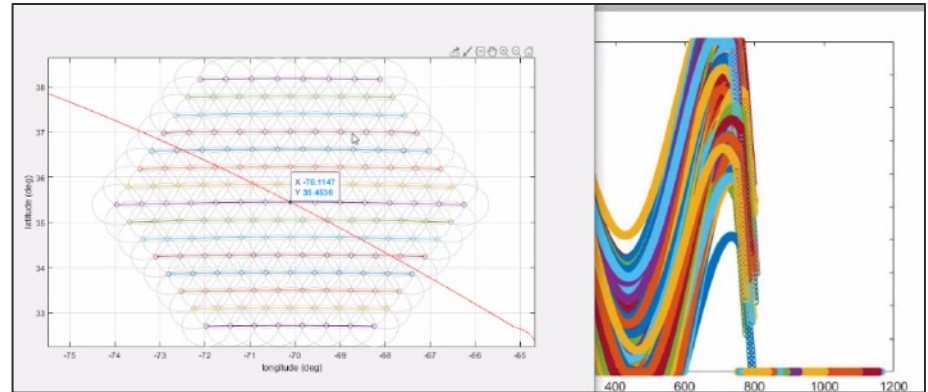
Implement

Implement the technical capability on contract with customer funding.

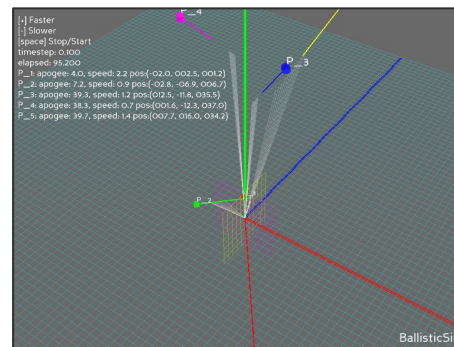
Once successfully awarded a contract, the customer specific implementation is developed from scratch on the contract budget.

This process avoids separate licensing concerns, and vendor lock-in

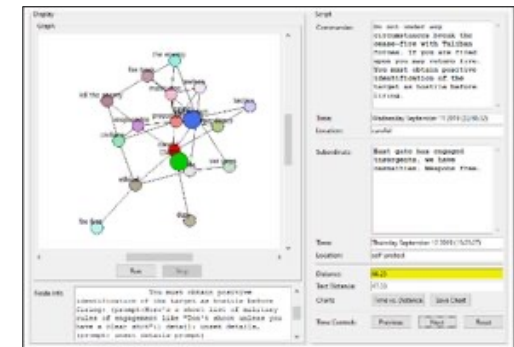
All codebases contain documentation on whether it is IRAD or customer/contract funded. *They never mix.*



Simulation of missile intercepts



Ballistic missile simulation

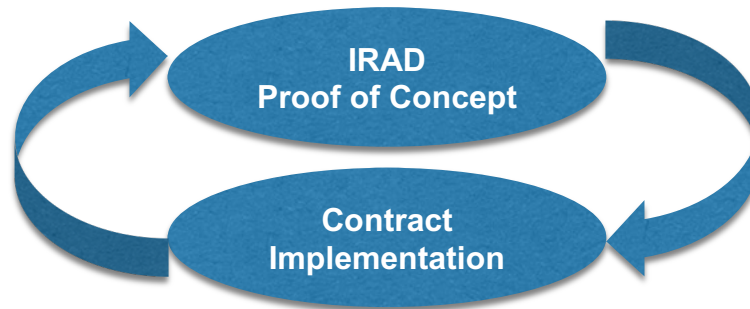


Narrative Mapping tool

Integrate

Lessons learned from the customer use-case can be reimplemented into the Reference Implementation using IRAD.

If we create something generically useful that we want to expand on as a core capability, we can reimplement that technique into the Reference Implementation from scratch.



Summary

Our five-step process provides:



- **Development** of monetizable IP.



- **Publishing** to showcase our capabilities.



- **Supporting** business growth by showcasing technology discriminators.



- **Implementation** in customer specific programs.



- **Integration** of lessons-learned into continuous product improvement.



Quyanaq!

Thank you!

Philip Feldman

AI/ML Futurist

e: philip.feldman@asrcfederal.com

w: (301) 837-5500 | c: (410) 300-7293

7000 Muirkirk Meadows Dr, Beltsville,
MD, USA