COGSEC-DIF-10.19 / Task 002

DARPA Initiatives in Focus

Heuristic Data and Third Wave AI

PREPARED BY RICHARD J. CORDES
RichardJ.Cordes@gmail.com
April 2019
Published through COGSEC

Contents

Executive Summary

This set of DARPA initiatives of interest will be presented individually, but in a format meant to facilitate the construction of a meaningful context with an underlying narrative.

DARPA Initiatives: Heuristic Data

Strategic Initiative 1: SMISC [2011-HR001118S0047]

In 2011, DARPA releases DARPA-BAA-11-64, later tagged as 2011-HR001118S0047, under the name SMISC or Social Media in Strategic Communication, with Dr. Rand Waltzman as a Program manager. Dr. Waltzman was a chief scientist at Lockheed Martin's Distributed Systems Lab. Along with SMISC, he also headed projects to create synthetic Social Media Data and to Detect patterns on associated platforms.

Objectives

DARPA-BAA-11-64 notes that Social Media has the potential to help the military and intelligence community to better understand the environments in which they operate and allow for agile acquisition of actionable intelligence in support of operations. The goal of SMISC is to develop automated and semi-automated support tools and techniques for systematic and methodical use of social media at scale to tackle the following objectives:

1. Detection, classification, measurement and tracking of the formation and development of ideas, perspectives, and concepts. These objectives would contribute to the detection, classification, measurement, tracking and countering of purposeful and deceptive misinformation
2. Detection, classification, measurement and tracking of persuasion campaigns and coordinated influence operations across social media and online communities
3. Identification of participants and intent
4. Development of means to counter stated persuasion campaigns

To clarify, the overarching goal of this initiative was to **Detect False Information and Counter it**.

Challenges

DARPA-BAA-11-64 requests that the proposal writers codify challenges they encounter dependent on the avenue of approach.

Avenues for Solutions

Potential avenues for finding solutions are noted:

1. Patterns of Information flow
2. Sentiment Tracking and Opinion Mining
3. Meme tracking across communities
4. Detecting attachment to cultural narrative or perspective
5. Detecting emergent communities and contagion potential
6. Trust analytics
7. Crowd-sourcing
8. Automated Content Generation
9. Social Media Bots

Technical Areas of Interest

The following Technical Areas are of interest:

[TA 1] Algorithm/Software development
[TA 2] Data Provision
[TA 3] Data/Algorithm Integration

Underlying Problems

While DARPA received valuable research, future BAAs reveal that there were a few underlying problems with implementing highly effective solutions.

1. Data Cataloging - Heuristically Relational Schemas
    a. Data provision was as difficult as algorithm development
    b. In this BAA, it asks algorithm designers to design and then request Data, then Data Provisioners collect and organise the data wanted. In future BAAs, this is **often reversed to imply a "work with what we can get you" attitude towards algorithm designers**
2. Centralised AI Failing at Human Heuristics
    a. Attempts to get AI to discover and catalog data based on constantly changing human heuristics is incredibly difficult and requires enormous amounts of data just to begin the approach of matching the accuracy of a human in handling said heuristics
3. Detection of False Information and Bias
    a. Detecting false information proved to be incredibly difficult, especially when the problem is that the crowd subjected to it isn't detecting it

Strategic Initiative 2: DEFT [2012-DARPA-BAA-12-47]

In 2012, DARPA releases DARPA-BAA-12-47, under the name DEFT or Deep Exploration and Filtering of Text, with Bonnie Dorr as a program manager. Bonnie Dorr holds a Ph.D. in Computer Science from MIT, has held numerous teaching positions and is a leading researcher in the realm of Machine Cognition and

Natural Language Processing. Along with DEFT, she also headed the projects BOLT (Broad Operational Language Translation), MADCAT (Multilingual Automatic Document Classification, Analysis, and Translation), and RATS (Robust Automatic Transcription of Speech).

Objectives

DARPA-BAA-12-47 notes that the ability to see through language to meaning in text, and the cataloging of information contained within, would enable revolutionary advances in science and systems.

1. Develop a range of research foci in order to develop a variety of implementable **algorithms which are capable of robustly handling real world input**
2. Development of HLT (Human Language Technology) for said algorithms to be integrated with

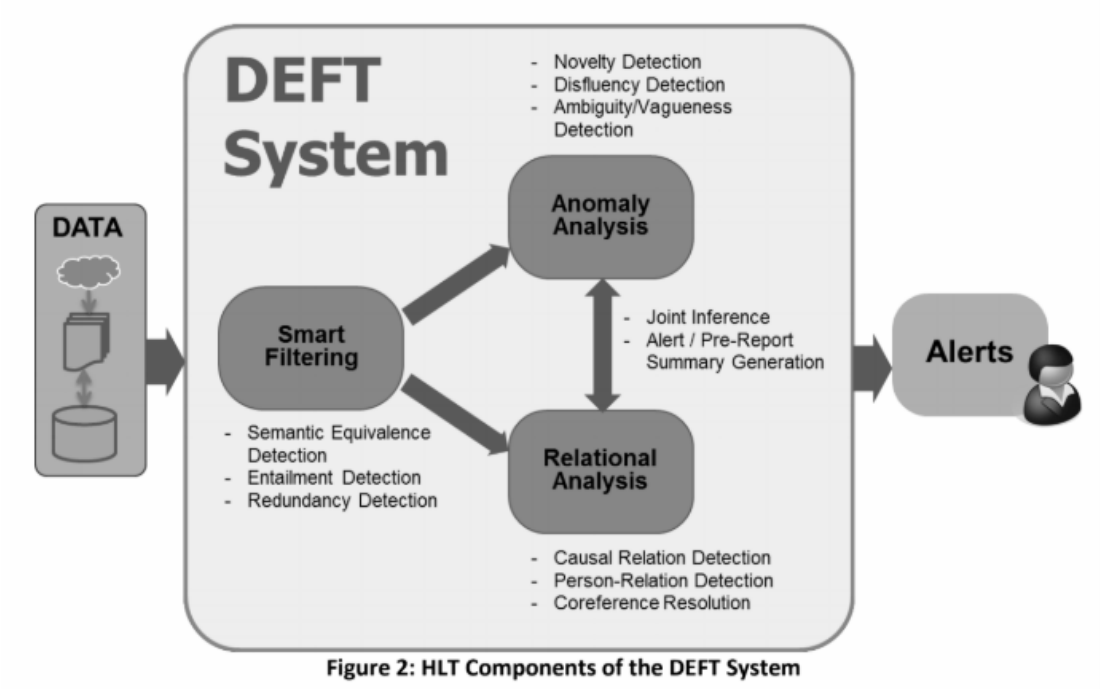

Figure 1 - DARPA-BAA-12-47 HLT Components

Challenges

DARPA-BAA-12-47 addresses the following challenges to discovering meaning in text documents:

1. Identifying people and groups
2. Identifying noted events
3. Discovering geo-spatial and temporal information
4. Identifying causation
5. Identifying modality and belief (perspective)
6. Identifying inter-relationships, coreference of entities and events.
7. Identifying disfluency/disjointedness
8. Identifying information dynamicity or perishability
9. Identifying ambiguity, vagueness, and temporary symbols
10. Identifying inconsistency
11. Identifying implied information

Avenues for Solutions

The following avenues for solutions were given, of particular interest is the development of a means for computers to analyze data ontologically, mimicking how the human brain builds an understanding of its environment and the objects and information within it. Bonnie Dorr, with her background in machine cognition, seemed to be an excellent choice.

1. Automatic Ontological Evolution
2. Axiom Learning
3. Anomaly Signaling

Technical Areas of Interest

The following Technical Areas are of interest:

[TA 1] Research and Algorithm Development
[TA 2] Research Integration

[TA 3] Data Creation and Annotation

Problems

While DARPA received valuable research, future BAAs reveal that there were a few underlying problems with implementing highly effective solutions.

1. Data Cataloging - Heuristically Relational Schemas
    a. Training AI to handle the kinds of tasks within the BAA were extraordinarily ambitious. They require enormous amounts of highly standardised training data
    b. Pedagogist Jean Piaget's conception of how to schematize events and objects is something that even humans have trouble doing despite the fact that human hardware is optimised for this
    c. Trying to coreference and form contextual links between objects was as difficult as identifying said objects
2. Centralised AI Failing at Human Heuristics
    a. Attempts to get AI to discover and catalog data based on constantly changing human heuristics is incredibly difficult and requires enormous amounts of data just to begin the approach of matching the accuracy of a human in handling said heuristics

Strategic Initiative 3: BIG MECHANISM [2014-DARPA-BAA-14-14]

In 2014, DARPA releases 2014-DARPA-BAA-14-14, under the name BIG MECHANISM, originally launched by Dr. Paul Cohen as program manager before handing it over by Dr. Joshua Elliott.

Cohen was the founding Dean of the University of Pittsburgh's School of Computing and Information. He's authored 200 peer-reviewed articles and 5 books. Dr. Cohen holds a Doctor of Philosophy Degree in Computer Science and Psychology from

Stanford University, and a Master of Science Degree in Psychology from the University of California.

Elliott holds a Doctorate in Physics from McGill and spent most of his time at DARPA dealing with socio-technical change, optimal decision and policy making under uncertainty, and environmental variability and its impact on food security. Elliott has also held positions with the Argonne National Lab, the London School of Economics and Political Science, and Columbia University.

2014-DARPA-BAA-14-14 begins with a more humble statement than previous BAAs: "Some of the systems that matter most to the [Department of Defense] are very complicated." They define "Big Mechanisms" as causal, explanatory models of complicated or complex systems in which interactions have important causal effects. Big Data can be discovered and catalogued by machines, but Big Mechanisms cannot. The systems in which they occur: ecosystems, economies, and social groups have literatures and data which is fragmented and poorly distributed. So long as phenomena are studied piecewise, interacting factors are often poorly understood.

DEFT was ambitious in trying to build ontological evolution across internet-wide text data. It would seem that the idea within this BAA was to narrow the scope of such an interest by focusing on one area of text with fairly consistent jargon, and more consistent templates and signals for particular granules of data. The domain chosen was Cancer Biology and the subdomain was signaling pathways.

Objectives

DARPA-BAA-14-14 intended to:

1. Develop technology to read research abstracts and papers and extract fragments of causal mechanisms
2. Assemble fragments of causal mechanisms to produce big mechanisms

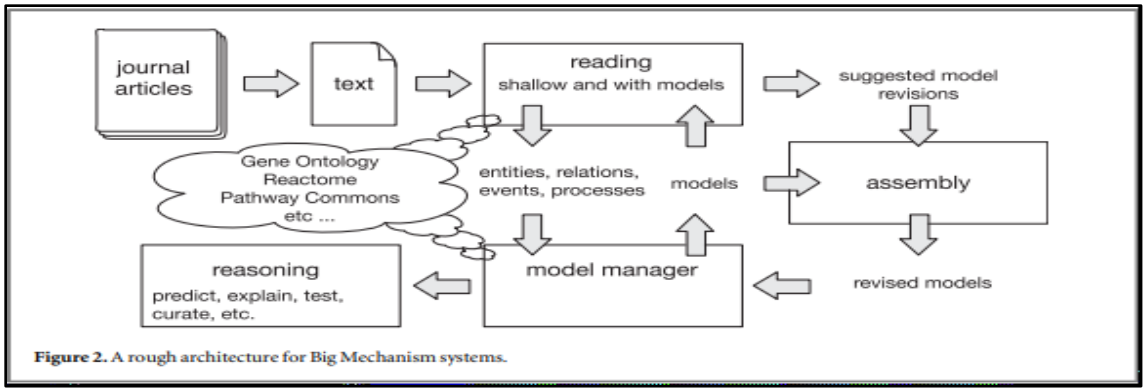

Figure 2 - DARPA-BAA-14-14 Rough Architecture for Big Mechanism Systems (1)

Challenges

2014-DARPA-BAA-14-14 implies the following challenges to extracting meaningful fragments and assembling meaning from them.

1. Identifying people and groups
2. Identifying noted events
3. Identifying causation
4. Identifying modality and belief (perspective)
5. Identifying inter-relationships, coreference of entities and events.
6. Identifying disfluency/disjointedness
7. Identifying information dynamicity or perishability
8. Identifying ambiguity, vagueness, and temporary symbols
9. Identifying inconsistency
10. Identifying implied information

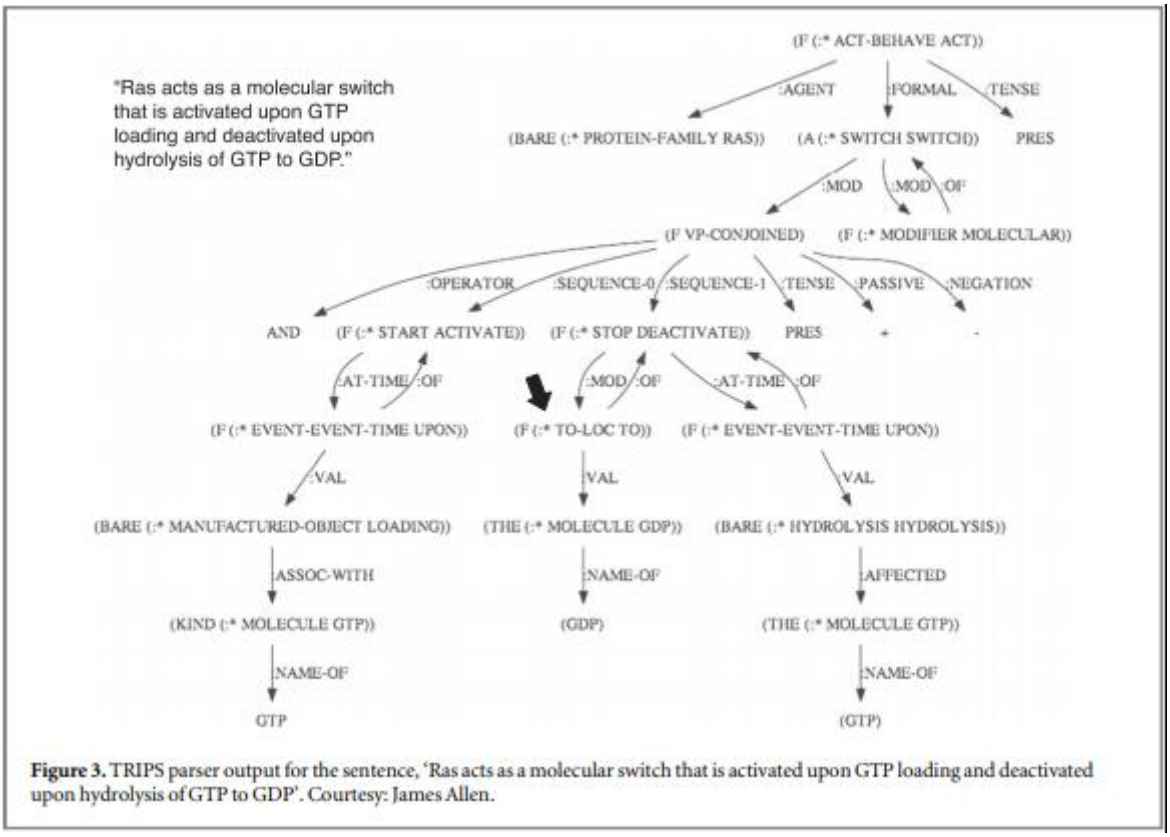

Figure 3. TRIPS parser output for the sentence, 'Ras acts as a molecular switch that is activated upon GTP loading and deactivated upon hydrolysis of GTP to GDP'. Courtesy: James Allen.

Figure 3 - DARPA-BAA-14-14 Rough Architecture for Big Mechanism Systems (2)

Avenues for Solutions

DARPA-BAA-14-14 implies avenues of solutions which are identical to that of DEFT.

1. Automatic Ontological Evolution
2. Axiom Learning
3. Anomaly Signaling

Technical Areas of Interest

The following Technical Areas are of interest:

[TA 1] Natural Language Processing
[TA 2] Assembly of Conceptual Fragments
[TA 3] Explanation of Assembled Big Mechanisms

Problems

DARPA received valuable research from this endeavor. One paper of particular interest, written by Marjorie Green spoke of ontology and the nature of big data itself. Her paper is based on a book by MIT mathematician and founder of the field of *Cybernetics*, Norbert Wiener, called **Human Use of Humans.** The paper, called "A self-organizing system for automatic information classification and retrieval", discussed the book's suggestion, 50 years ahead of its time, that data *analyzed* by machines might fail to produce the insight that humans are capable of and how emergent systems might be required to handle large-scale data analysis. Despite valuable research from a variety of sources from multiple TAs, future BAAs reveal that there were a few underlying problems with implementing highly effective solutions.

1. Data Cataloging - Heuristically Relational Schemas
    a. Identifying relationships between concepts in the research seemed to present massive challenges
2. Centralised AI Failing at Human Heuristics
    a. Developing big mechanisms seems to be outside the scope of current capabilities in Machine Cognition

When Joshua Elliott comes to DARPA in 2017, he is given authority over this project and, soon after, shutters it and begins a new project: World Modelers, which disrupts the exclusive focus on machines and transforms it to **machine assistance of humans.**

Strategic Initiative 4: World Modelers [HR001117S0017]

In 2017, DARPA releases HR001117S0017, under the name World Modelers, with **Dr. Joshua Elliott** as project manager.

HR001117S0017 begins by discussing concern over food insecurity and the ramifications of scarcity in third world economies, with a specific emphasis on Africa.

Objectives

The objectives of this BAA differ from those which preceded it because they, as previously stated, shift focus away from machine-dependent solutions to machine-assisted solutions.

1. Develop technology that will enable **human analysts** to rapidly build models to analyze questions relevant to national and global security
2. Develop a means to rapidly develop comprehensive analyses of complex systems like economies and ecosystems. Food Security was chosen for this reason as it is a crossroads for many different systems: *weather, water supply, safety from conflict, economic stability, credit availability etc.*
    a. Important to this goal is finding a means of constructing causal relationships between stated factors. In other words, finding a means to develop **big mechanisms**

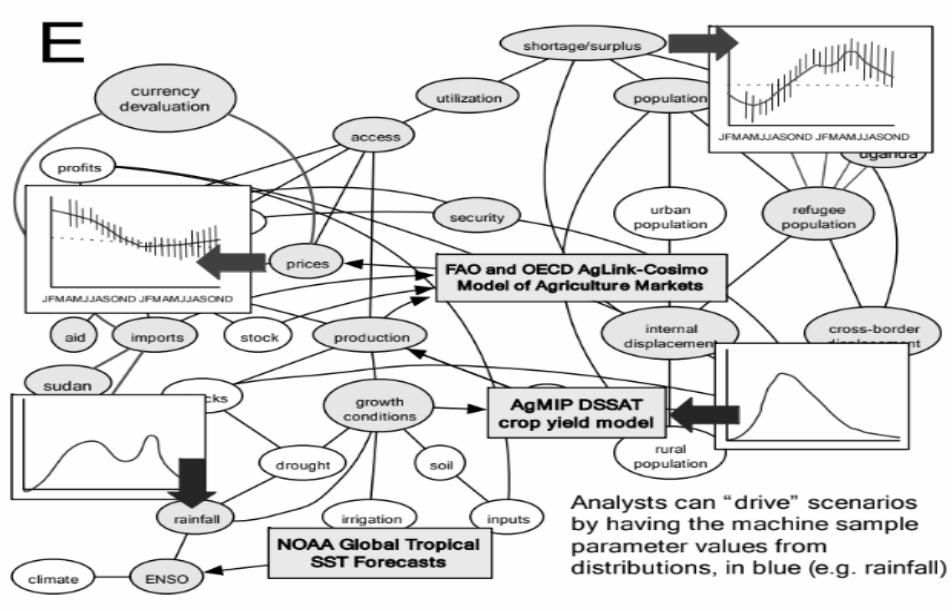

Figure 4 - HR001117S0017 Big Mechanism Food Security Graph

Challenges

The implied challenges within this BAA are similar to that of Big Mechanism and DEFT.

1. Identifying people and groups
2. Identifying noted events
3. Discovering geo-spatial and temporal information
4. Identifying causation
5. Identifying inter-relationships, coreference of entities and events.
6. Identifying disfluency/disjointedness
7. Identifying information dynamicity or perishability
8. Identifying ambiguity, vagueness, temporary symbols
9. Identifying inconsistency
10. Identifying implied information

Avenues for Solutions

HR001117S0017 implies or states the following avenues for solutions.

1. Causal analysis graphs
2. Parameterization of the results of targeted analyses.
3. Machine Reading Platforms as Collaborators to Humans
4. Using Machines to develop models.

Technical Areas of Interest

The following Technical Areas are of interest:

[TA 1] Machine-Assisted Construction of Qualitative Models
[TA 2] Machine-Assisted Marriage of Qualitative and Quantitative Analyses
[TA 3] Machine-Assisted Data Foraging
[TA 4] Interface Design for Users to Handle Scenario Specification and Intervention
[TA 5] Uncertainty Analysis

Problems

1. Set-Theoretic Granularity in Information
    a. Overlapping and redundant data was a problem
    b. Keyword systems are archaic
    c. The ability to compartmentalise and combine data presents large problems
    d. Lack of essential meta-data in order to create unions between identical granules of data
2. Data Cataloging - Heuristically Relational Schemas
    a. Identifying heuristic relationships between objects in the research seemed to present massive challenges
    b. The lack of essential meta-data
3. Centralised AI Failing at Human Heuristics
    a. Developing big mechanisms seems to be outside the scope of current capabilities in Machine Cognition

Strategic Initiative 5: SHARP [IARPA-BAA-13-06]

IARPA is, as mentioned in a previous section, the "Intelligence Advanced Research Projects Activity", which is a sister organisation of DARPA, formed in 2006, specifically targeting the Intelligence Community.

In 2014, IARPA releases IARPA-BAA-13-06, under the name SHARP or Program to Significantly Improve Adaptive Reasoning and Problem-Solving with Dr. Adam Russell as the Project Manager. From DARPA's staff page:

> "Dr. Adam H. Russell joined DARPA as a program manager in July 2015. He is interested in new experimental platforms and tools to facilitate discovery, quantification, and "big validation" of fundamental measures in social science, behavioral science, and *human performance."*

Dr. Russell has broad technical and management experience across a number of disciplines, ranging from **cognitive neuroscience** and physiology to **cultural psychology** and social anthropology. Before joining DARPA, he was a program manager at the Intelligence Advanced Research Projects Activity, where he developed and managed a number of high-risk, high-payoff research projects for the Office of the Director of National Intelligence. Prior to IARPA, Dr. Russell was in industry, where he was a senior scientist and **principal investigator on a wide range of human performance and social science research projects** and strategic assessments for a number of different government organizations.

Dr. Russell holds a Bachelor of Arts in Cultural Anthropology from Duke University, and both an M.Phil. and D.Phil. in Social Anthropology from Oxford University, where he was a Rhodes Scholar.

Objectives

IARPA intended to find ways to rapidly develop the adaptive reasoning and on-site problem solving of government employees and military personnel. However, given the time constraints on their high-performing personnel, who have full-time positions, researchers would have to find ways to do so in very short periods of time.

1. Aggregate Existing Research on Human Learning
2. Develop a suite of reliable psychometrics for competence
3. Test scientifically supported “interventions” which act on high-performing adults to strengthen ARP (Adaptive Reasoning and Problem Solving)
4. Ensure that these interventions require less than 210 minutes of active participation in any given week, so that the results may be practically applied to high performing individuals who are already working in full time positions

Challenges

The following challenges were noted or implied:

1. Aggregation of research within the field
2. Providing an auditable approach. No Black Box approaches allowed
    a. **Explicit refusal to use machine learning**
3. Managing to significantly develop talent in less than 210 minutes of intervention a week

Avenues for Solutions

No avenues for solution were given other than a general suggestion to look to the last two decades of related literature.

Technical Areas of Interest

No Technical Areas of Interest were explicitly given.

Problems

IARPA-BAA-13-06 provided very valuable research to IARPA and partner agencies but seemed to have a problem in common with previous initiatives, but for different reasons.

1. Data Cataloging - Heuristically Relational Schemas
    a. Sorting through the enormous amounts of research proved difficult.
    b. Finding contradictions between conclusions and finding trusting reproducibility of results proved difficult
        i. This was especially true for the social sciences
2. Inefficient Culture Market
    a. As eluded to in the previous section. Social and Behavioural research was found to be very difficult to trust. Many of the papers seemed to present themselves as confirmatory when they were merely

> exploratory, and the nature of their reproducibility was regularly in question.

It should be noted that Robert Hoffman, Senior Research Scientist at the Institute for Human and Machine Cognition, received $400,000 in grants through IARPA-BAA-13-06. The contributed research is not available and all attempts to access it have been unsuccessful. However, by reviewing his later work and the title of the submitted work: "*Microgame-based Interventions and Robust Analyses of Cognitive and Logical Skill Enhancers*", it seems apparent that **he was using gamification and game theory to structure a model for rapid/radical behavioural change, hijacking reward circuitry in the brain to improve the cognitive ability of subjects**. This is less important for this section, but important to the next set of DARPA and Government Initiatives.

Strategic Initiative 6: NGS2 [2016-DARPA-BAA-16-32]

In 2016, DARPA releases DARPA-BAA-16-32, under the name NGS2 or Next Generation Social Science, with **Adam Russell** as project manager.

NGS2 intended to build methods, models, tools, and a **community of human researchers** to perform rigorous, reproducible experimental research at scales necessary to study complex social systems and their **big mechanisms**. They anticipated that they'd require a *reimagining* of the entire social science research cycle. They believed participation from a wide and diverse combination of disciplines and skill sets would be required, not just individuals in the fields of social science, but also *physics, computer science, **game design**, and mathematics*.

Objectives

Objectives were separated by "Performer Categories"

1. End to End

   These teams are expected to engage in multiple cycles of modeling, hypothesis generation, prediction, experimentation, analysis, and reevaluation.

2. Enablers

   Smaller, targeted efforts to develop and test new or early stage high-risk, high-payoff "enabling" technologies in one or more of the TAs. This second category of performer should enable some new capability (e.g., greater speed, more rigor, higher transparency, more dynamic informed consent, larger numbers of participants) and be poised for integration into a complete end-to-end approach. Proposals in this category must detail the unique enabling technical or methodical approach to be developed in regard to (1) the next generation research capability made possible by the proposed approach, (2) the feasibility of achieving this goal, and (3) the effort required for maturation.

3. Testing and Evaluation

   This team would be responsible for assisting the Government with testing and evaluation of the other performers' programmatic and technical progress towards NGS2 goals. Proposers in this category must have demonstrated experience in scientific experimentation, diverse collaboration and evaluation of the accuracy of experimental predictions. They must also have a history of effectively interacting with multiple research communities, including the social sciences, to help advance multidisciplinary, reproducible research in line with NGS2 goals. Proposers should discuss their capabilities in managing experiment registration, innovative forecasting and prediction platforms, engagement with appropriate scientific publications

for registered reports of performer protocols as a form of early peer review, and providing consultation, infrastructure support, and collaboration enhancement across scientific teams and geographies.

Challenges

The following challenges were apparent:

1. Finding means to develop **trust networks** on which to scrutinise or accept research
2. **Building a decentralised community** to work together
3. Accounting for bias in research
4. Determining evidentiary support.
5. **Enabling Reproducibility by having Data Cataloging Standards**
6. Extracting fragments of causal mechanisms from research
7. Assembling fragments of causal mechanisms to produce **big mechanisms**

Avenues for Solutions

Avenues for Solutions were not listed explicitly.

Technical Areas of Interest

The following Technical Areas are of interest:

[TA 1] Predictive Modeling and Hypothesis Generation
[TA 2] Experimental Methods and Platforms
[TA 3] Interpretation and Reproducibility

Problems

Problems found in attempts to respond to this BAA are aligned with the pattern of problems found in previous BAAs.

1. Data Cataloging - Heuristically Relational Schemas
2. Detection of False Information and Bias

Strategic Initiative 7: ASKE [2018-DARPA-PA-18-02-01]

In 2018, DARPA releases DARPA-PA-18-02-01, under the name ASKE or Automating Scientific Knowledge Extraction, with **Dr. Joshua Elliott** as project manager. They were offering up to a million dollars to produce a prototype that could generate new, human-readable hypotheses from scientific research automatically.

Objectives/Challenges

Objectives were presented as challenges.

1. Translate queries into models
2. Demonstrate an approach for machine evaluation (and generation) of hypotheses
3. Demonstrate capacity to elaborate human-readable explanations of results from machine-assisted inference
4. Develop technology to read research abstracts and papers and extract fragments of causal mechanisms and correlative factors.
5. Identify new information and data automatically.

Technical Areas of Interest

[TA 1] The following Technical Areas are of interest:
[TA 2] Machine-Assisted Curation
[TA 3] Machine-Assisted Inference

Problems

1. Data Cataloging - Heuristically Relational Schemas
2. Centralised AI Failing at Human Heuristics
3. Set-Theoretic Granularity in Information

4. Detection of False/Bad Information

Strategic Initiative 8: SCORE [2018-HR001118S0047]

In 2018, DARPA releases HR001118S0047, under the name SCORE or Systematizing Confidence in Open Research and Evidence, with **Dr. Adam Russell** as project manager. Given the accelerating socio-technical complexity of today's world, DARPA's Defense Sciences Office came to believe that new, critical security challenges were emerging and that they would need to leverage Social and Behavioural science at scale to address it. Unfortunately, the inability to aggregate and ascertain the quality, consistency and reproducibility of research was hampering this effort. The BAA addresses the fact that a number of recent empirical studies, meta-analyses, and systematic literature reviews have revealed that Social and Behavioural research artifacts vary dramatically in their ability to be independently reproduced or replicated.

The Department of Defense has since noted that this problem is deeply concerning.

Image 1 - DARPA SCORE Logo

Objectives

HR001118S0047 has the following objectives:

1. Develop parameterised methods for developing "confidence" scores, or an equivalent, in research

2. **Develop platform for crowdsourcing experts** to assign confidence scores in order to develop training data for machines
3. Develop machine learning algorithm to learn from expert assignment

Challenges

The implied challenges are as follows:

1. Sheer number of journals, articles and preprints
2. The difficulty in finding standard measures for confidence given the plethora of problems in the literature.
3. Load balancing and efficiency. The system should be able to handle 3,000 articles and claims per year

Avenues for Solutions

HR001118S0047 explicitly offers a number of avenues for solutions, some of which are essential to understanding the arc of the narrative set forth in this segment of the document:

1. Rating derived from Retractions
2. Evidence of publication bias
3. Undeclared conflicts of interest
4. Unsuccessful attempts at reproduction of results
5. **Crowd-sourcing post publication review**
6. **Social Networks of reviewers**
7. **Incentive structures**

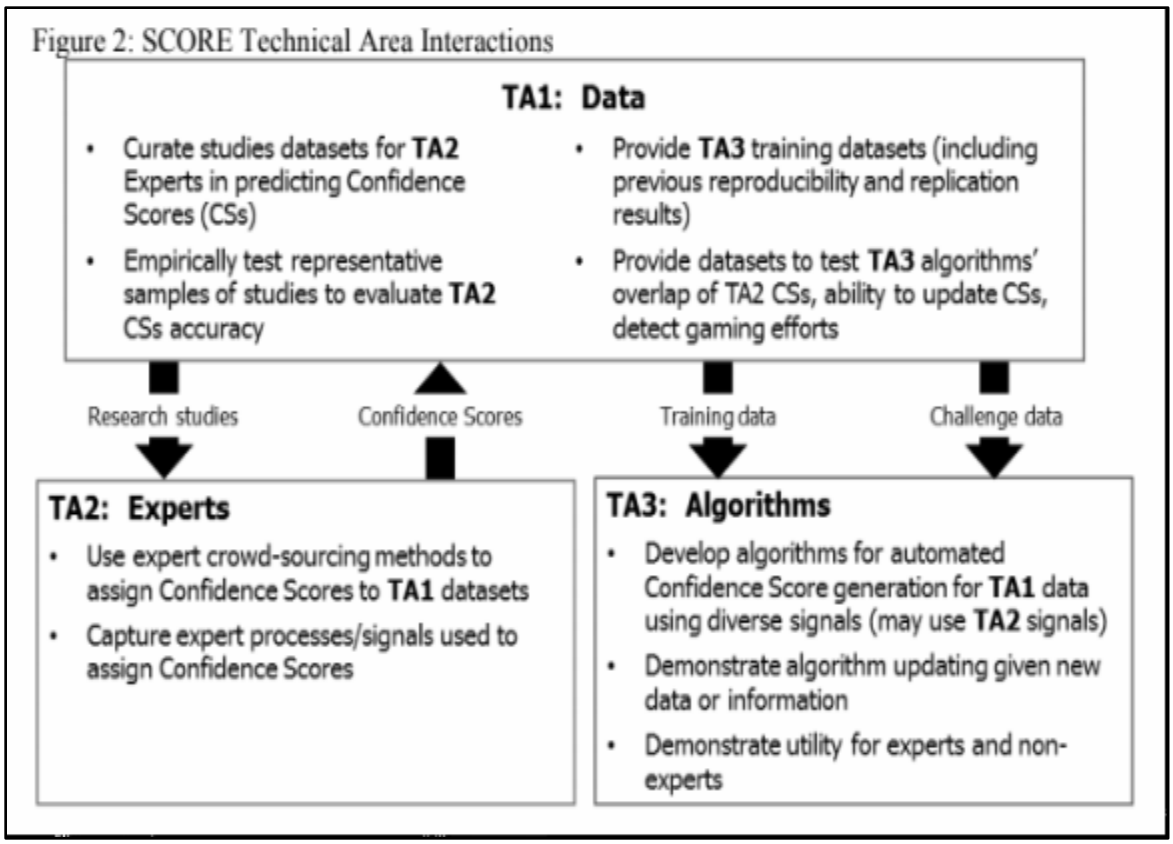

Figure 5 - DARPA SCORE TA Architecture

Technical Areas of Interest

Technical areas of interest are listed below:

[TA 1] Data procurement (research papers and articles)
[TA 2] Teams to develop the scalable methods and platforms for eliciting expert scores.
[TA 3] Machine Learning, using training data from TA 2.

Problems

SCORE is too recent to find results.

Strategic Initiative 9: KAIROS [2019-DARPA-SN-19-19]

In 2018, DARPA also releases DARPA-SN-19-19 under the name KAIROS. No project manager or technical point of contact is listed on the document as this is not the BAA, it is simply a notice for a proposer's day wherein they release and request information from potential participants. KAIROS is a clever acronym, standing for Knowledge-directed Artificial Intelligence Reasoning Over Schemas, Kairos also meaning a propitious, or favourable, opportunity or time to act. The summary they released is short and direct, and the narrative of this segment would be better served by using their own summary in lieu of the previously used format:

> *Rapid comprehension of world events is essential for informing U.S. policy, diplomacy, and national security, a task that becomes more difficult as the amount of unstructured, multimedia information grows exponentially. Vital early indicators are often overlooked in the sheer amount of information available.* ***Many important events are not simple occurrences, but complex phenomena that are composed of numerous subsidiary elements****, some of which happen simultaneously, while others are* ***sequential and dependent on each other****. The KAIROS program will develop mixed-initiative systems that can identify complex events and bring them to the attention of users. KAIROS will explore how to understand complex events described in multimedia input by developing a semi-automated system that identifies, links, and temporally sequences their subsidiary elements, identifying the participants of the complex events and the subsidiary elements, and identifying the complex event type. An event is a recognizable and significant change in either the natural world or human society.* ***Events of interest either create changes that have significant impact on national***

security or participate in causal chains that produce such impacts.

Humans make sense of events by organizing them into narrative structures that occur frequently. ***These structures are abstracted into schemas, which are organized units of knowledge that represent a pattern of memory used in human cognition.*** *The schemas now used in artificial intelligence (AI) research are based on schemas defined by the cognitive scientist* ***Jean Piaget*** *in 1923 as a way in which people process and store memories about event sequences. First-wave (i.e., rule-based, symbolic reasoning) AI systems that incorporated hand-crafted schemas were unable to scale when matching those schemas to real-world data.* ***Second-wave (i.e., machine learning) AI systems require far too many manually-produced, annotated examples as training data for supervised machine learning methods to be practical.***

DARPA is seeking revolutionary ideas that use schema-based AI to comprehend events, their components, and the participants involved. The KAIROS program will seek to overcome the scaling limitations of prior approaches in two stages. The first stage entails the development of automated approaches for learning schemas from big data, and the second stage focuses on the development of automated technologies that apply these schemas to multimedia/multilingual information to discover and extract complex events of interest to KAIROS users.

They note first and second wave AI, what they do not note, is DARPA's initiative regarding the ***third wave*** of artificial intelligence, which will be noted later in the document.

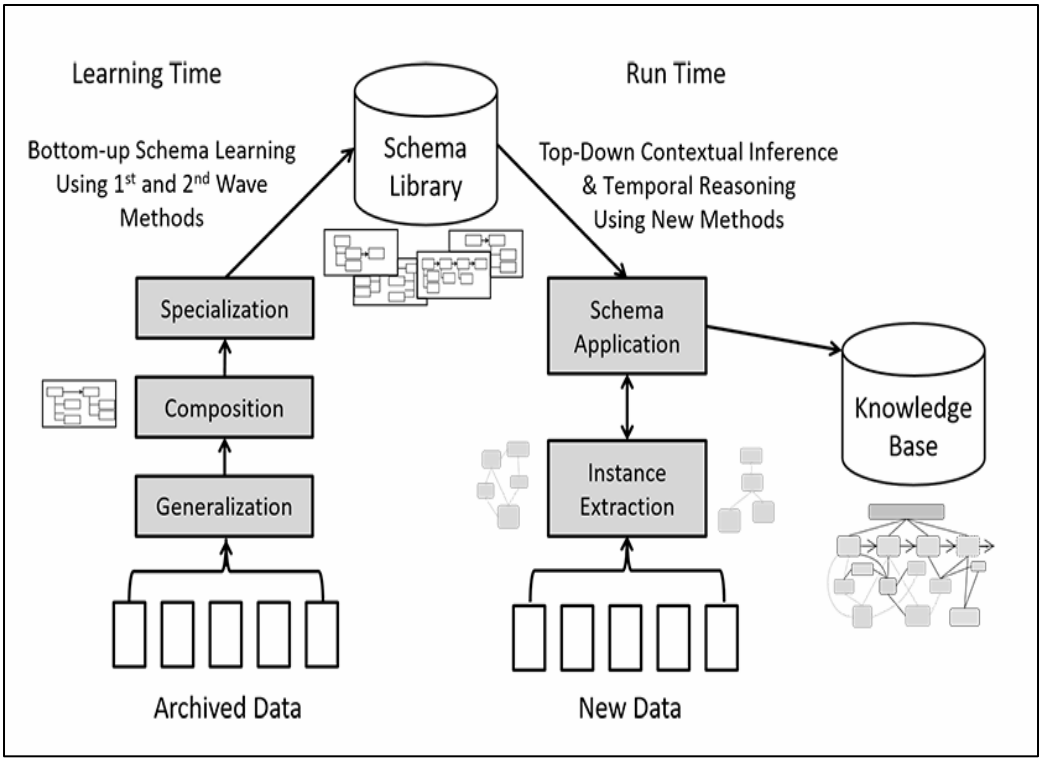

Figure 6 - KAIROS Architecture

Conclusions

There are many individual initiatives which were omitted, although many are worthy of mention. These nine initiatives and the story they tell seemed to be sufficient to develop an underlying narrative that revolves around government and academe failing to develop highly effective, implementable solutions to the following problems:

1. Set-Theoretic Granularity in Information
2. Data Cataloging and Developing Heuristically Relational Schemas
3. Detecting False Information and Bias
4. AI Failing to Handle Human Heuristics

Further Discussion

Third-Wave AI

As hyper-similar initiatives appear within its own programs and within the programs of self-similar agencies, bureaucratic bloat, a tell-tale sign of a rut, is beginning to appear around these four problems. As noted in the KAIROS summary, DARPA has acknowledged first and second wave AI as obsolete for its current goals.

First wave AI were hard-coded systems, the AI isn't learning or thinking, it's just built of sophisticated algorithms which tackle problems in strictly defined environments; Dijkstra path finding is a good example. Second Wave AI uses large sets of training data and sophisticated statistical methods (such as signal desensitization and back-propagated gradient-descent) to develop and refine networks of perceptrons that are interpreted and acted on as tensors in order to feed new parameters and receive predictive value in multi-parameter environments that are not strictly defined.

It has now made announcements regarding its push towards a *third wave* AI. They will be throwing 2 billion dollars towards this single endeavor, which they have named **AI Next**. The investment and interest is not surprising given that the Department of Defense, Pentagon, CIA, and DARPA outlooks on the problems it will solve.

Weaponization of Information: The Need for Cognitive Security

Dr Rand Waltzman, former program manager of **SMISC,** now working for the RAND corporation, submitted a document as testimony to the Committee on Armed Services of the United

States Senate in 2017. The document was called **The Weaponization of Information: The Need for Cognitive Security**. The name would likely only be considered hyperbolic by those who were uninformed as to the nature of its contents. In eight pages, it gives an extraordinarily clear narrative of the dangerous, critical problems SMISC failed to find solutions to.

A couple of scenarios worth noting:

Twitter Hack Causes 136b Loss in Equity Value in 5 minutes

> The information environment can be broadly characterized along both technical and psychosocial dimensions. Information environment security today (often referred to as cybersecurity) is primarily concerned with purely technical features—defenses against denial-of-service attacks, botnets, mass intellectual property theft, and other attacks that typically take advantage of security vulnerabilities. This view is too narrow, however. For example, little attention has been paid to learning how to counter or defend against incidents like the April 2013 Associated Press Twitter hack in which a group hijacked the news agency's account to put out a tweet reading "Two explosions in the White House and Barack Obama is injured." This message, with the weight of the Associated Press behind it, caused a drop and recovery of roughly $136 billion in equity market value over a period of about five minutes. This attack exploited both technical (hijacking the account) and psychosocial (understanding market reaction) features of the information environment.

Clever Placement of Captioned Video Results in Deployment of 13,000 Indian Troops

> Another attack, exploiting purely psychosocial features, took place in India in September 2013. The incident began when a young Hindu girl complained to her family that she had been verbally abused by a Muslim boy. Her brother and

cousin reportedly went to pay the boy a visit and killed him. This spurred clashes between Hindu and Muslim communities. In an action designed to fan the flames of violence, somebody posted a gruesome video of two men being beaten to death, accompanied by a caption that identified the two men as Hindu and the mob as Muslim. Rumors spread over telephone and social media that the mob had murdered the girl's brother and cousin in retaliation. **It took 13,000 Indian troops to put down the resulting violence.** It turned out that while the video did show two men being beaten to death, **it was not the men claimed in the caption; in fact, the incident had not even taken place in India**. **This attack required no technical skill whatsoever; it simply required a psychosocial understanding of the place and time to post to achieve the desired effect.**

Post-Combat Photo-Op Results in a Battalion of U.S. Personnel Out of Commission

Another particularly revealing incident took place during Operation Valhalla in Iraq in March 2006. A battalion of U.S. Special Forces Soldiers engaged a Jaish al-Mahdi death squad, killing 16 or 17, capturing 17, destroying a weapons cache, and rescuing a badly beaten hostage. In the time it took for the soldiers to get back to their base—less than one hour—Jaish al-Mahdi **soldiers had returned to the scene and rearranged the bodies of their fallen comrades to make it look as if they had been murdered while in the middle of prayer. They then put out pictures and press releases in Arabic and English** showing the alleged atrocity. **The U.S. unit had filmed its entire action and could prove this is not what happened.** And yet it took almost three days before the U.S. military attempted to tell its side of the story in the media. **The Army was forced to launch an investigation that lasted 30 days, during which time an entire battalion was out of commission.**

Weaponization of Information: The Need for Cognitive Security cont.

Rand Waltzman goes on to detail Russian IO (Information Operations) strategies. The Russian government has openly admitted that it considers IO an ongoing effort. Chief of Russian General Staff, General Valery Gerasimov stated that war is now conducted by a 4:1 ratio of nonmilitary to military measures. In the Russian view, nonmilitary warfare includes economic sanctions, political and diplomatic disruption, and IO warfare. **The United States, China, and Russia are currently in an open war by their own definitions**. The Russian Military Academy of General Staff released a document of key information regarding information security, it contrasts fundamental Russian and Western concepts of IO by specifically saying that, for Moscow, **IO is seen as a continuous activity whereas the west sees IO as a tactical/intervention activity.**

Waltzman declares to the Senate, probably as their most credentialed expert, that **the only way to counter this is to create a defense that is "whole-of-nation" in character.** He suggests a coordinated effort between government, industry, the intelligence community, media, research, **academia and citizen organised groups**. He suggests there is a need to develop **infrastructure for trust networks** to prevent cognitive hacking of American citizens.

> *"We are in an existential battle for truth in the digital domain… That's, again, where the help of the private sector is important and these data providers. Because that's frankly the digital conflict we're in, in that battle space…This is one of my highest priorities… Data is inherently dual use and like many other benefits of mankind, it is also being weaponized."*
>
> *Andrew Hallman, CIA Digital Directorate*

There is a need to rapidly develop a means to detect false information and prevent coordination in its spread, develop AI to detect deliberate misinformation, and we need set-theoretic granularity in data to rapidly develop models of complex systems. There may be few paths towards solving these problems.

A Crossroads: Wise Crowds

Dr Steven Walker, the director of DARPA, stated in his announcement of **AI Next**, up to this point, machines lack contextual reasoning and their training has to cover every eventuality in order to be accurate, which he notes is "ultimately impossible". This not only reveals the lack of success of many of the BAAs mentioned, but the over-ambitiousness of their goals. He wants the agency to focus its attention machine cognition (once again), on exploring how machines might acquire human reasoning. This was announced in 2018 prior to SCORE and KAIROS. His exact quotes from this release appear in KAIROS.

It would appear that DARPA is at a crossroads on how to approach this problem. Looking over the dual BAAs released in 2018, it would seem that, consistent with DARPA's strategy, they are avoiding a rut by going down **both** roads:

Using pure machine learning or **using crowdsourcing and letting machines learn from wise crowds.** The problem seems to be that **building a platform large enough to house and incentivise wise crowds is outside of DARPA grant responders' capabilities and this problem may have led their project managers to ignore it as an avenue for solutions for some time**, but this is changing. While KAIROS focuses on a pure machine learning approach, SCORE does not. SCORE is explicitly interested in crowd sourcing experts. Also released in 2018, very much in line with Director Walker's stated interests, was a project called Machine Common Sense. MCS explicitly states the intent to use both approaches: pure machine learning **and** crowdsourcing experts

to assist in machine learning. The latter is not a new concept, it had been investigated by some of DARPA's more interesting project managers and produced results that are material to this document.

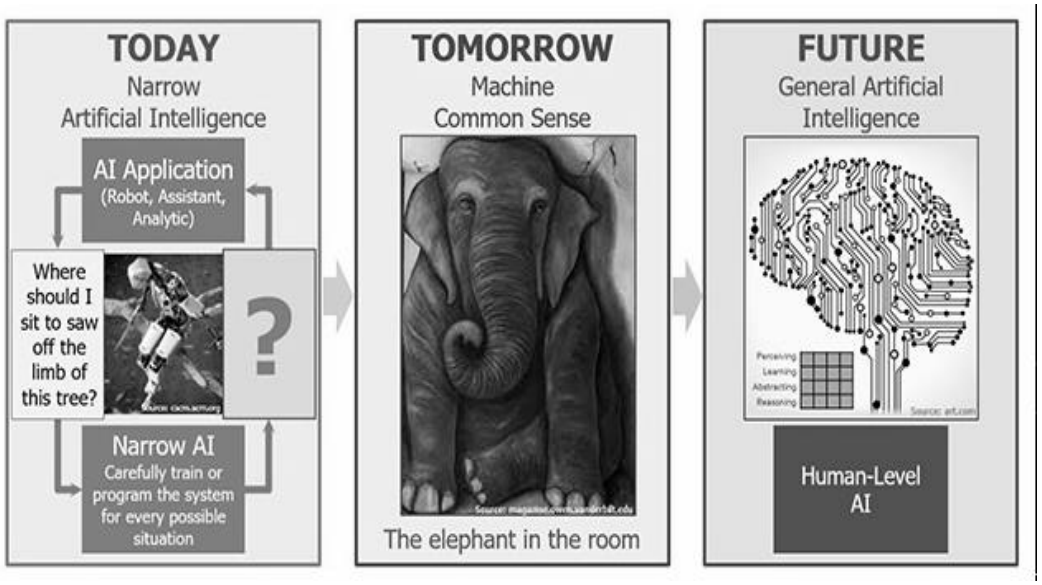

Figure 7 - DARPA Machine Common Sense Diagram

Ralph Chatham

DARPA, as previously stated, is an important object in the stitching together of industry, government, and military—and it depends heavily on revolutionary thinking. This means that occasionally, there are some surprisingly eccentric people who end up in surprisingly serious positions. Gilman Louie at IN-Q-TEL got his start in the Gaming Industry, Stewart Alsop of Alsop Louie was originally a bartender and then a business commentator. DARPA is hyper-aware that the brilliance they seek is often accompanied by eccentricity, and they do their best to, sometimes begrudgingly, avoid looking the proverbial gift horse in the mouth, even when they choose to do things like this in an official report released to several government agencies (see Ralph Chatham):

Rob Oberbreckling is a founder of Perceptive Research Inc. His interests include applying software systems to problems in cognitive science, natural language processing, machine learning, audio signal processing, and automated human performance measurement, modeling, and assessment. Robert previously was a Senior Member of Technical Staff at Pearson Knowledge Technologies where he led team communication data collection efforts in the field as well as created predictive systems for individual and team performance for commercial and military applications. Contact information: Perceptive Research Inc., 3050 24th St., Boulder, CO 80304, rob.oberbreckling@gmail.com

Ralph Chatham is a physicist, storyteller, all-purpose curmudgeon, and lately program manager for the Defense Advanced Research Projects Agency. He is currently a private insultant, delivering advice on technology development, and training in the Defense Department. He has been a submarine officer, laser builder and chairman of two task forces of the Defense Science Board, herding DoD elephants to explore the issues of training superiority and training surprise. He has managed, either inside or outside the government, contract research on: putting lasers in space to talk to submarines patrolling under water and clouds; synthetic aperture sonar; real science applied to detecting deception; and digital tools, games and simulations for training such things as language and information technology troubleshooting. He created and managed from afar the research program discussed in this paper. In addition to the Defense Superior Service Medal, the Secretary of Defense Medal for Exceptional Public Service and other DoD award, Ralph and his wife jointly received a 2003 National Storytelling Network Oracle Award. Contact Coordinates: 2631 Kirklyn Street, Falls Church, VA 22043; 703 698 5456; ralph.chatham@verizon.net.

Joseph Psotka is a Program Manager for basic and applied research in behavioral and social sciences at the Army Research Institute. He earned a Ph.D. degree in cognitive psychology from Yale University in 1975. He taught at several colleges and universities, including Southern Connecticut State College and the University of Waterloo, before becoming Director of Research at NPSRI in Alexandria, Va. in 1978. He was made a Resident Scholar of the National Institute of Education (NIE) in 1981. Dr. Psotka joined the Army Research Institute in 1982 as a team chief within the Training Laboratory, where he has remained. In 1988 his edited volume on Intelligent Tutoring Systems: Lessons Learned was published. His research now focuses on social network analysis, LSA and automated text understanding, leadership, communities of practice, unobtrusive measurement technologies, automated tutoring by intelligent agents, simulation technologies, and higher order thinking. Contact information: Joseph Psotka, US ARI, 2511 Jefferson Davis Highway, Arlington, VA 22202-3926. joseph.psotka@hqda.army.mil

Image 2 - Ralph Chatham Self-Description

Ralph Chatham lived a very full life and, by many accounts, was an exceptionally interesting and brilliant person. He loved Monty Python and Hitchhikers Guide to the Galaxy and was obsessed with folklore, dinosaurs, and the future. He worked at DARPA for 4 years as a program manager, returning later only to become frustrated with bureaucracy. He left to become a private consultant and unfortunately, soon after, developed a brain tumor and died.

Despite his awards, achievements, and predictions, his time at DARPA is not publicly recognised or noted:

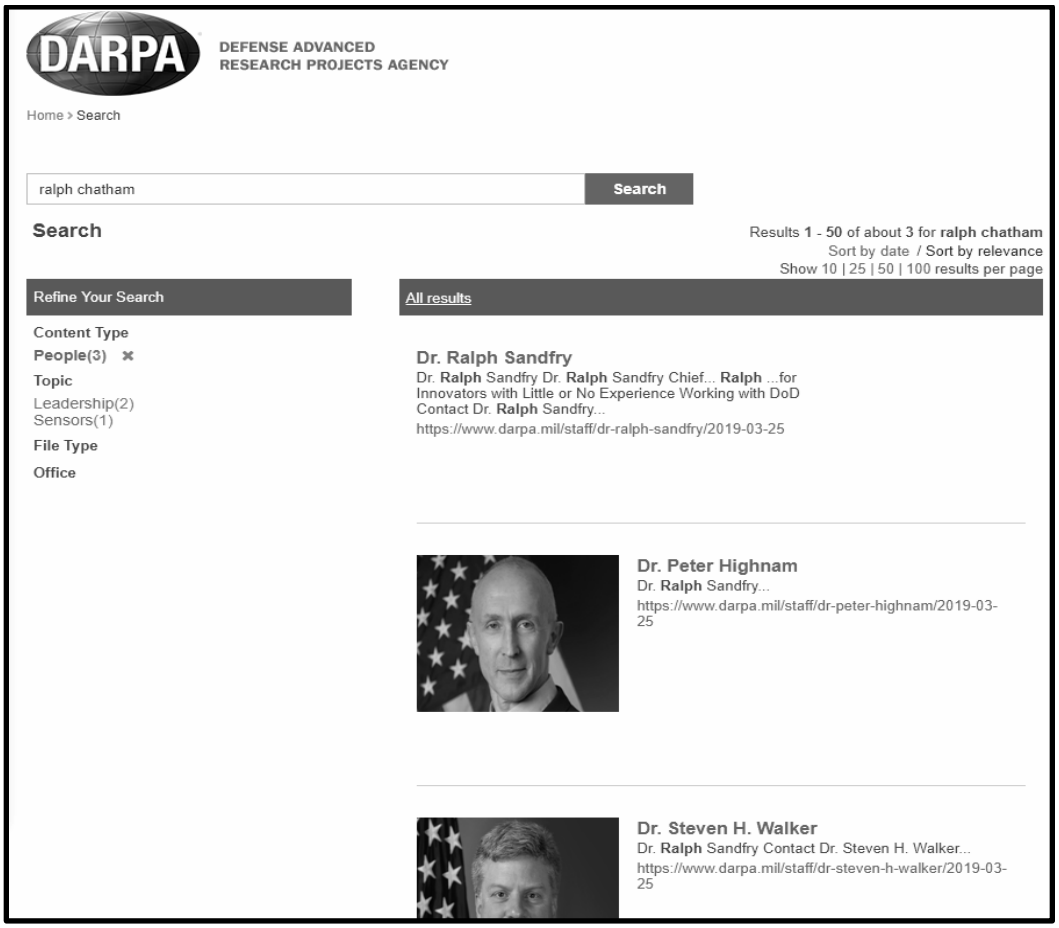

Image 3 - Ralph Chatham Staff Page Not Found

Despite this, it seems he affected a great number of important people within his field—many of whom privately recognise him. Interesting enough, the dedication of the best-selling Tom Clancy novel, The Hunt for Red October, which has since been turned into two feature films, reads:

For Ralph Chatham,
a sub driver who spoke the truth,
and for all the men who wear dolphins.

DARPA has a habit of employing eccentrics, giving them a budget, allowing them to publish results, and then not immediately listening to them. Which is not a slight to DARPA, often the results of these employees are years ahead of their time

and not yet implementable or well understood. Trying to find what Ralph Chatham actually worked on while at DARPA was extraordinarily difficult as there seemed to have been a scorched earth policy on public information regarding any work that he did. There were a lot of dead links and no references to him or his projects on any official pages.

However, after circumventing pay/credential walls, searching through transcribed speeches, dozens of obscure documents, and even the eulogy from his funeral, information was found that helped to develop a path to discovering his work. It was found that the last note of his work anywhere was in the credits area of work by Dr Benjamin Bell (a scientist working on Government E-Learning research, who will be noted in the next segment), and leading up to that, he had been working with the following projects:

1. Robust Self-Forming Human Networks
2. Collaborative Assistance and Rapid Team Optimisation
3. Spontaneous Collaboration and Linking
4. Rapid Formation of Virtual Organisations

More importantly, he had been doing so on the basis that these were potential ways to solve problems in Data Aggregation **and Education,** which will be detailed in a follow-up report.