Office of the Director of National Intelligence
Intelligence Advanced Research Projects Activity
IARPA
Creating Advantage through Research and Technology

BENGAL

BIAS EFFECTS AND NOTABLE GENERATIVE AI LIMITATIONS

CogentGPT

Gheorghe Tecuci

Professor of Computer Science and Director of the Learning Agents Center, School of Computing, George Mason University

Former Chair of Artificial Intelligence and Visiting Professor, Center for Strategic Service and Leadership, U.S. Army War College

Member of the Romanian Academy and Honorary Director of the „Mihai Drăgănescu" Institute for Artificial Intelligence

tecuci@gmu.edu, http://lac.gmu.edu

PROPOSERS' DAY DATE: October 24, 2023

Overview

Instructable Cognitive Agents and Critical Thinking

Cogent

Instrctable Cogent

ChatGPT

CogentGPT

Amazing Applications of CogentGPT

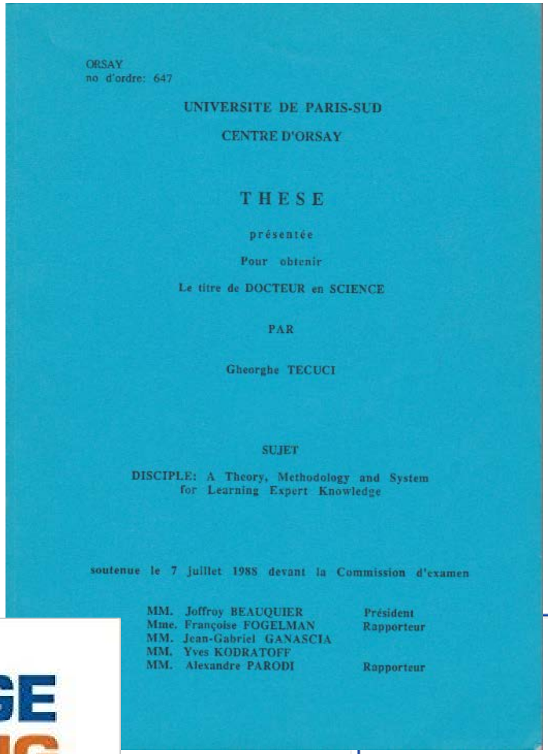
References

Instructable Cognitive Agents

The prevailing approach to the development of knowledge-based agents is through knowledge acquisition from a subject matter expert and representing this knowledge into the agent's knowledge base, which is a form of programming.

This is a *long*, *difficult*, and *error-prone process*.

Agent Instruction researches the development of agents through teaching them as we teach students, rather than programming them.


ORSAY
no d'ordre: 647

UNIVERSITE DE PARIS-SUD

CENTRE D'ORSAY

THESE

présentée

Pour obtenir

Le titre de DOCTEUR en SCIENCE

PAR

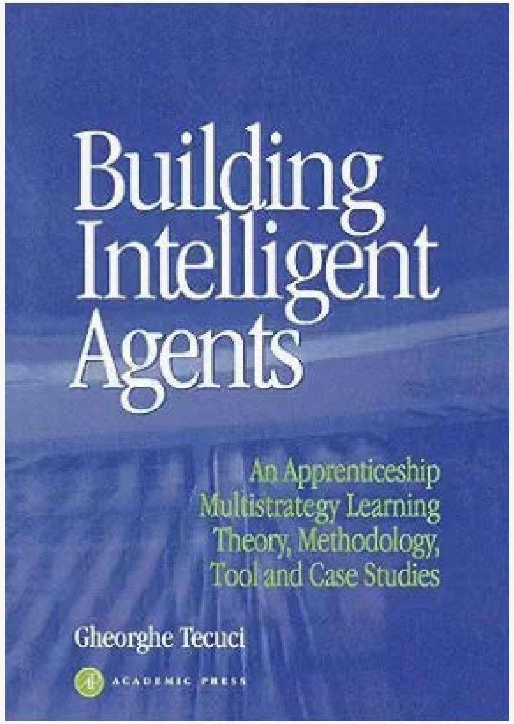
Gheorghe TECUCI

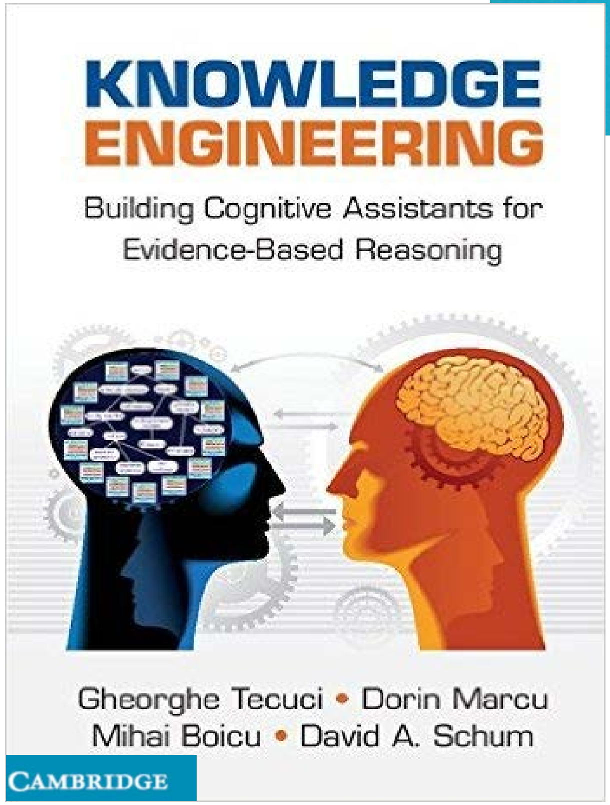
SUJET

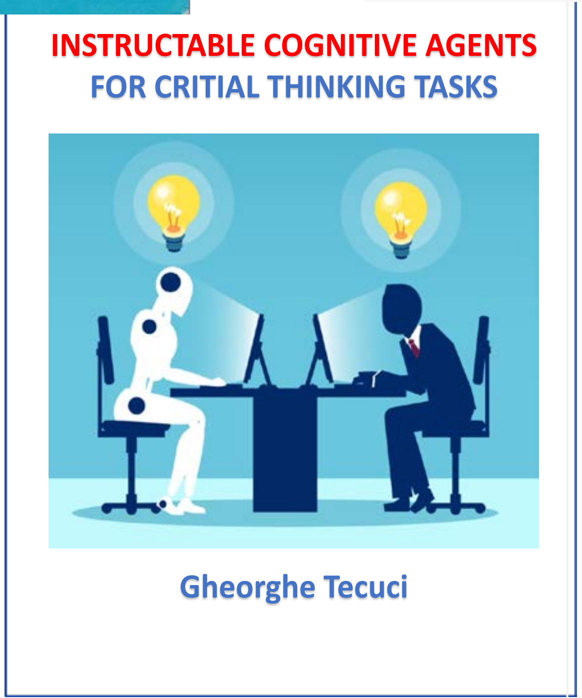
DISCIPLE: A Theory, Methodology and System for Learning Expert Knowledge

soutenue le 7 juillet 1988 devant la Commission d'examen

MM.	Joffroy BEAUQUIER	Président
Mme.	Françoise FOGELMAN	Rapporteur
MM.	Jean-Gabriel GANASCIA	
MM.	Yves KODRATOFF	
MM.	Alexandre PARODI	Rapporteur



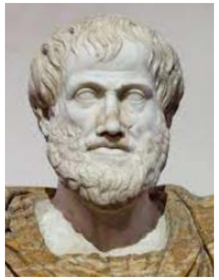




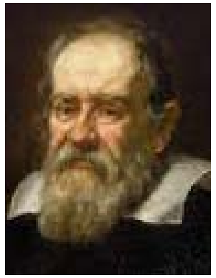
Critical Thinking

Critical thinking is a complex concept that was developed over the past 2500 years through the work of some of the greatest minds, including Aristotle, Galileo Galilei, John Locke, Isaac Newton, William Whewell, Charles Peirce, John Wigmore, and David Schum, who have tried to understand the world through a process of discovery and testing of hypotheses based on evidence.

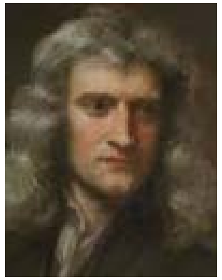
In essence, critical thinking refers to the ability to analyze information objectively and make a reasoned judgment.



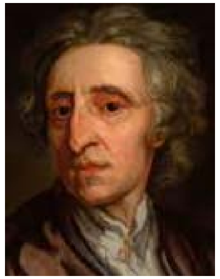
Aristotle [384–322BC]



Galileo Galilei [1564–1642]

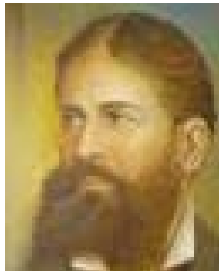


Isaac Newton [1632–1704]

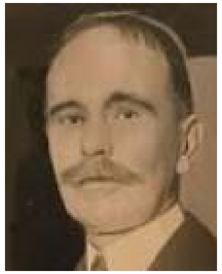


William Whewell [1642–1727]

John Locke [1794–1866]



Charles Peirce [1839–1914]



John Wigmore [1863–1943]

David Schum [1932–2018]

Scientific thinking, mathematical thinking, historical thinking, anthropological thinking, economic thinking, moral thinking, and philosophical thinking, each incorporates critical thinking which is at the core of problem-solving and decision-making in many disciplines, including military science and intelligence, computing, natural and social sciences, education, agriculture, and medicine.

Cogent

For two decades, we have worked on a *computational theory of intelligence analysis* (Tecuci et al., 2011). On this basis, we have developed a sequence of increasingly more practical cognitive assistants for the intelligence analysis education and practice. The first of these systems, **Disciple-LTA** (Tecuci et al., 2005; 2008), is a unique and complex cognitive assistant that integrates powerful capabilities *for Learning, Analysis, and Tutoring*, and is at the basis of other developed systems.

TIACRITIS(*Teaching Intelligence Analysts Critical Thinking Skills*) was developed for teaching intelligence analysis and was experimentally used in many IC and DOD organizations (Tecuci et al., 2011). While praising its solid theoretical framework and deep evidentiary knowledge, the analysts desired a simplified interface and interaction.

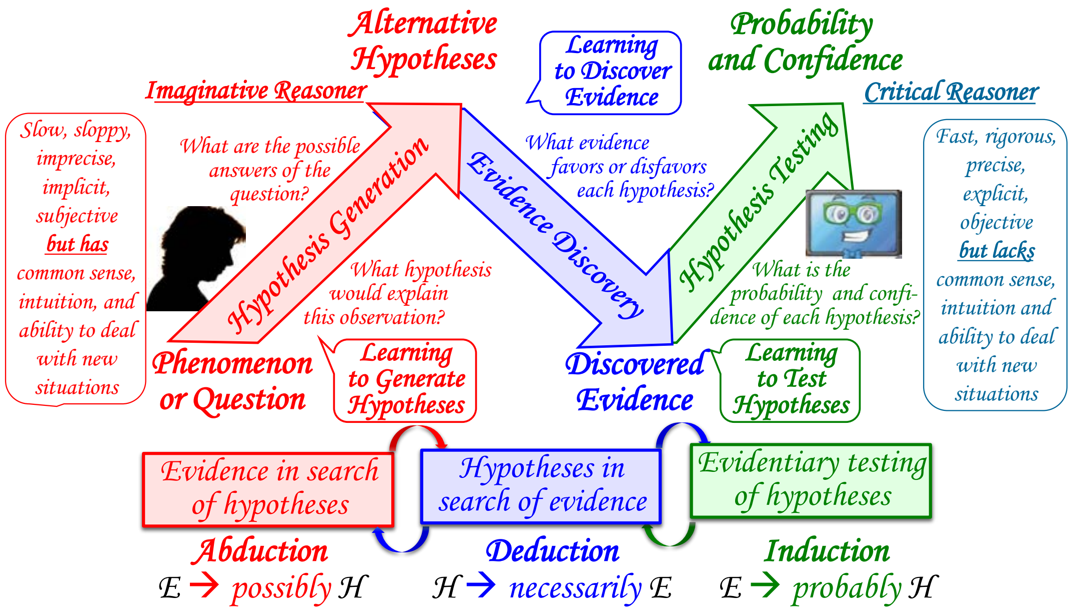
The next system, **Disciple-EBR** (*Disciple Cognitive Assistant for Evidence-based Reasoning*) is a general *learning agent shell for the development of agents for evidence-based reasoning tasks* (Tecuci et al, 2014; Tecuci et al., 2016b). One such agent is **Disciple-CD** (*Disciple Cognitive Assistant for Connecting the Dots*), described in (Tecuci et al, 2016a). Disciple-EBR and Disciple-CD significantly improved TIACRITIS along several dimensions, such as, the use of the *Baconian* and *Fuzzy* probability systems, easier argument development, more flexible management of knowledge bases, improved usability and scalability.

Next, with *significant feedback from intelligence analysts*, we have developed **Cogent** the *Cognitive Agent for Intelligence Analysis* (Tecuci et al., 2015; Tecuci et al., 2018; Tecuci and Schum, 2023; Tecuci, 2023a) that significantly improves the user experience while preserving the Disciple-EBR's sound foundations in the computational theory of intelligence analysis. A short (3 min) video on Cogent is at http://lac.gmu.edu/Cogent/index.html

Instructable Cogent

We propose to develop **Instructible Cogent** by:

- Evolving the mixed-initiative analysis methods of Cogent into automatic analysis methods(Tecuci et al., 2007), , as demonstrated by the **MASH** system (Tecuci et al., 2021), and by developing and integrating automatic capabilities for multi-step abduction, for assessing the credibility and relevance of evidence items, and for assessing the confidence in probabilistic assessments.



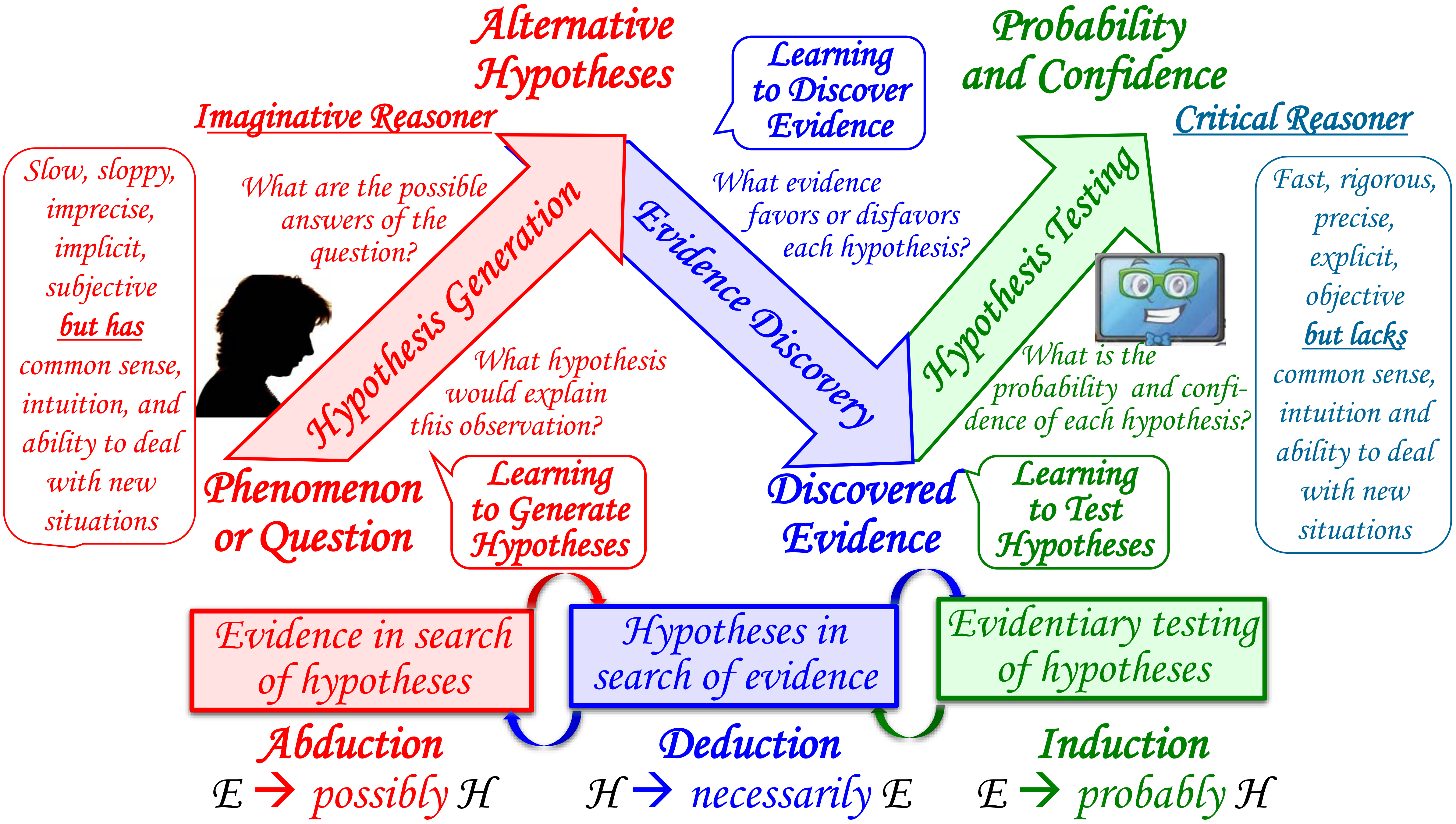
- Integrating and further developing the learning capabilities of **Disciple-EBR** (Tecuci, 1988; 1998; Tecuci al., 2016b; Tecuci, 2023a; 2023b), and by developing and integrating capabilities for automatic rule learning, ontology learning, scenario generation, and rule refinement.

First the user and Cogent will use *abductive (imaginative) reasoning* (that shows that something is *possibly* true) to generate hypotheses hat would explain the observed phenomenon or are possible answers to the question.

Each hypothesis will be used to guide the discovery of relevant evidence, by employing *deductive reasoning* (that shows that something is necessarily true). The user and Cogent will develop arguments that decompose the hypothesis into simpler and simpler hypotheses, until the simplest ones point directly to this evidence.

Finally, they employ inductive reasoning (that shows that something is probably true) to test the hypothesis.

From this analysis, Cogent learns general rules to generate hypotheses, to discover evidence, and to test hypotheses.

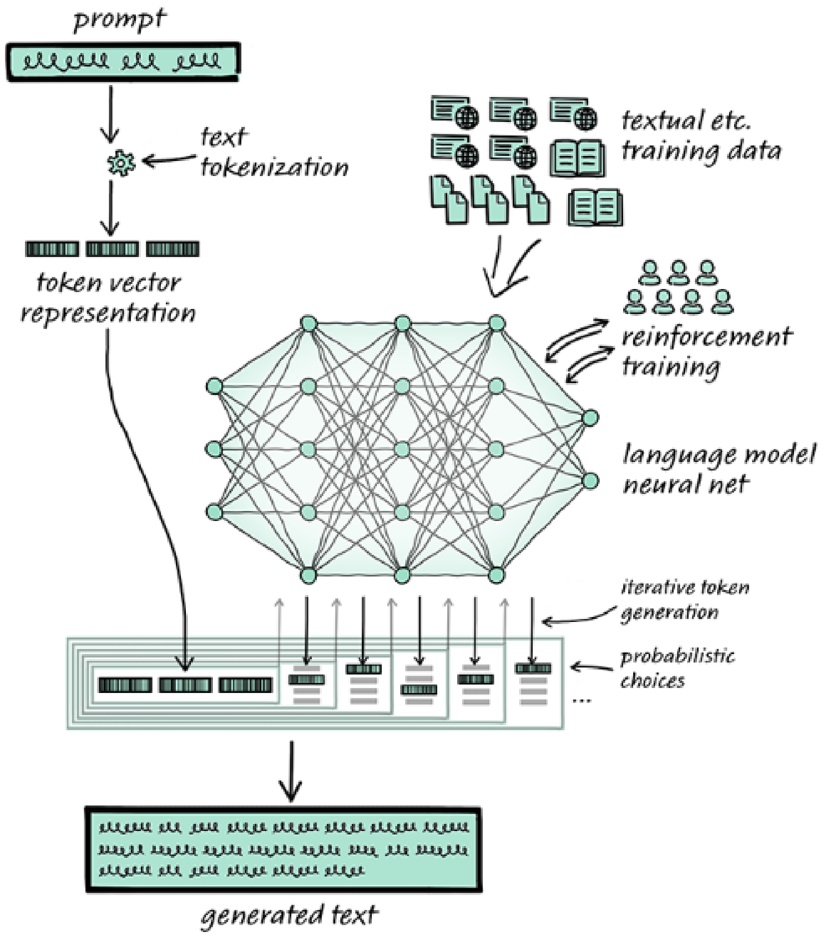

Alternative Hypotheses
Probability and Confidence
Imaginative Reasoner
Learning to Discover Evidence
Critical Reasoner
Slow, sloppy, imprecise, implicit, subjective but has common sense, intuition, and ability to deal with new situations
What are the possible answers of the question?
Hypothesis Generation
Evidence Discovery
What evidence favors or disfavors each hypothesis?
Hypothesis Testing
Fast, rigorous, precise, explicit, objective but lacks common sense, intuition and ability to deal with new situations
What hypothesis would explain this observation?
What is the probability and confidence of each hypothesis?
Phenomenon or Question
Learning to Generate Hypotheses
Discovered Evidence
Learning to Test Hypotheses
Evidence in search of hypotheses
Hypotheses in search of evidence
Evidentiary testing of hypotheses
Abduction
E → possibly H
Deduction
H → necessarily E
Induction
E → probably H

ChatGPT

ChatGPT ingested (i.e., represented and integrated internally) what was posted on the Internet, and combines the information related to the asked question to generate a well-documented answer. As a result, its answer is a kind of average of the information posted on the Internet.

But if much of the information on a topic is wrong, its answer will also be wrong. The answer is also somewhat dated, because it takes time to represent the massive and continuously growing amounts of information available on the Internet.

ChatGPT uses a very sophisticated algorithm and a deep neural network to learn and generate answers. To put it very simply, the answer is generated by using a highly intricate formula that operates on numerical values corresponding to the input, resulting in a set of numerical values that represent the output. ChatGPT has no "understanding" of *why* this is the answer, and therefore cannot explain it. This is the main drawback of neural networks, in general. Additionally, ChatGPT is not (yet) a sophisticated problem solver, and cannot correctly answer questions that require complex (multi-step) reasoning, such as design or planning.



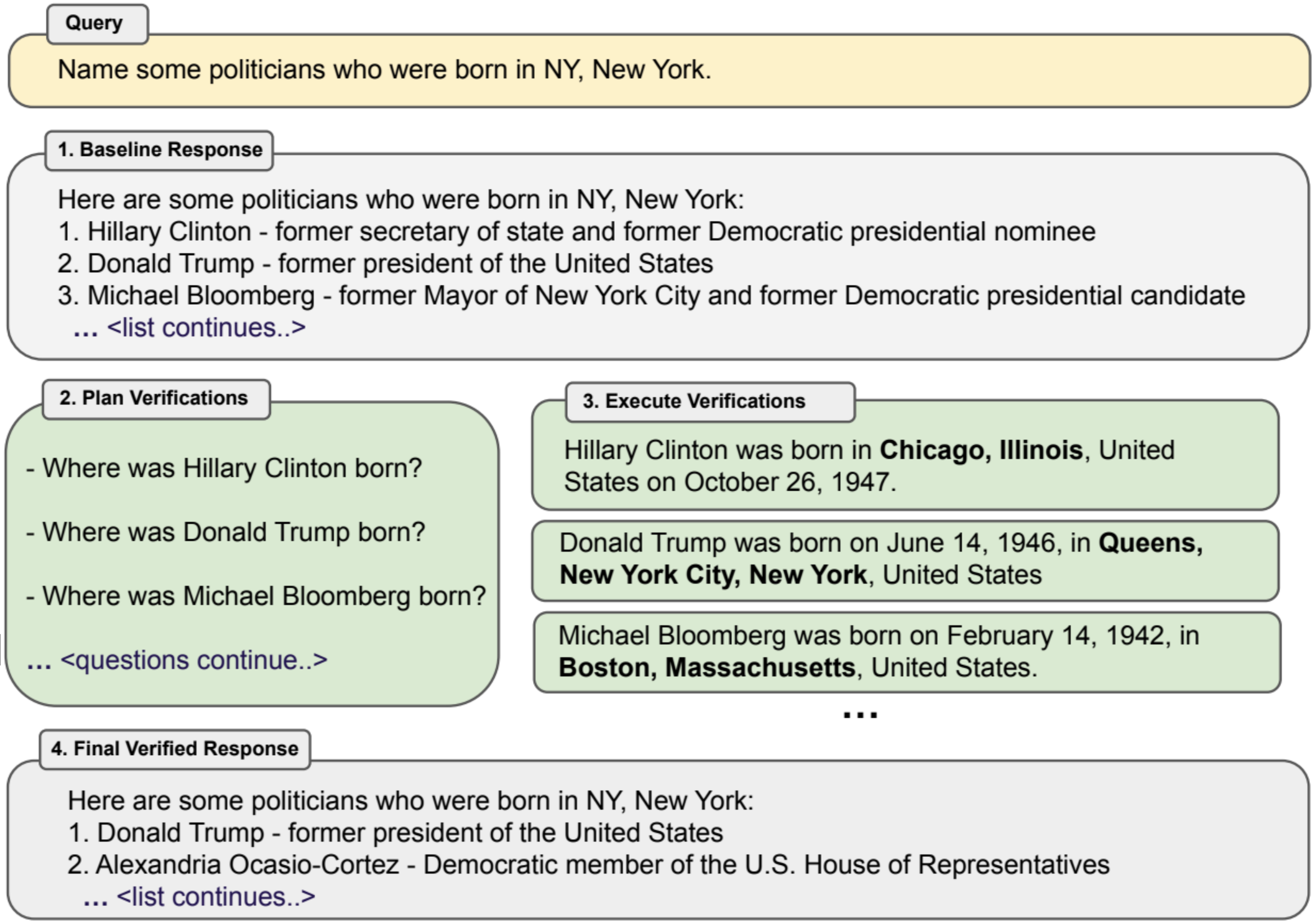
Overall architecture of ChatGPT (from https://writings.stephenwolfram.com/2023/01/wolframalpha-as-the-way-to-bring-computational-knowledge-superpowers-to-chatgpt/)

If people use ChatGPT to generate answers, then they should exercise their critical reasoning to check the generated answers. But it is much simpler to check whether an answer is correct or not, than to find it in the first place, and that is really the power of a tool like ChatGPT.

Chain-of-Verification Method (Dhuliawala et al., 2023)

https://arxiv.org/pdf/2309.11495.pdf#page14

Given a user query, a large language model generates a baseline response that may contain inaccuracies, e.g. factual hallucinations. We show a query here which failed for ChatGPT (see section 9 for more details). To improve this, CoVe first generates a plan of a set of verification questions to ask, and then executes that plan by answering them and hence checking for agreement. We find that individual verification questions are typically answered with higher accuracy than the original accuracy of the facts in the original longform generation. Finally, the revised response takes into account the verifications. The factored version of CoVe answers verification questions such that they cannot condition on the original response, avoiding repetition and improving performance.



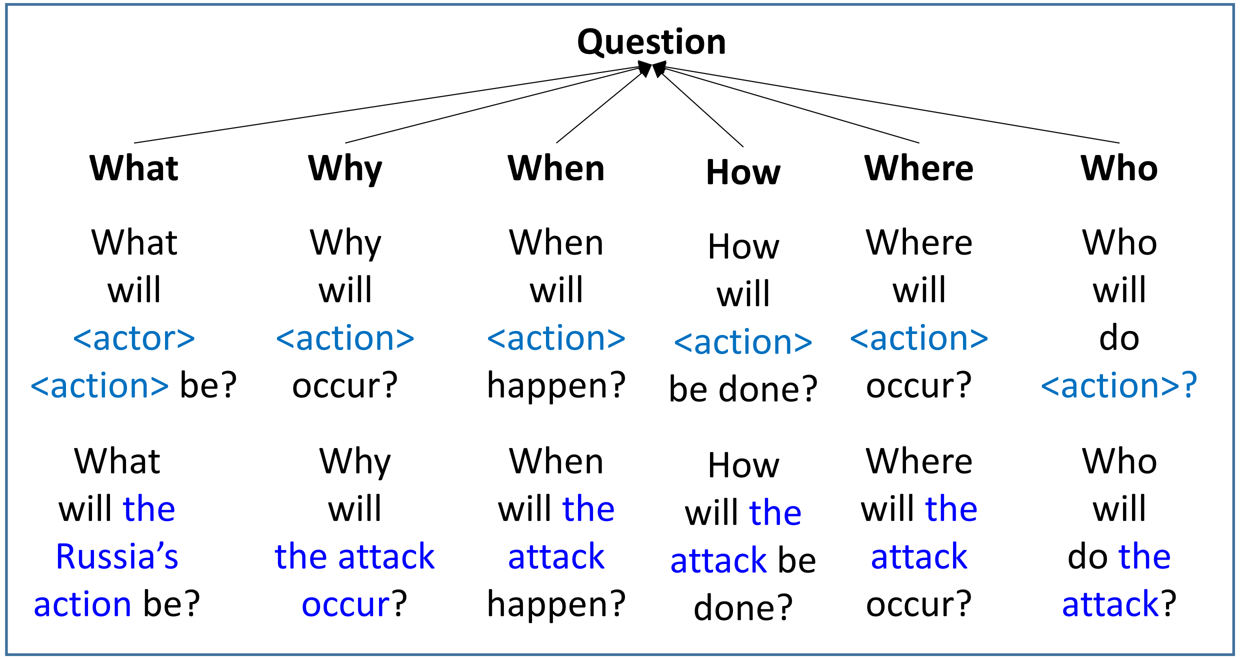
CogentGPT = Instructable Cogent + ChatGPT

We propose to develop **CogentGPT** by integrating the representation, learning and reasoning capabilities of **Instructable Cogent** with the natural language processing capabilities of **ChatGPT**.

CogentGPT will be taught by an expert to answer questions from a wide variety of topic areas, including political, military, social, economic, environmental, and diplomatic topics. These questions will be represented in a “substance-blind” ontology with Kipling’s What, Why, When, How, Where, and Who questions on the top of the ontology.

From the "The Elephant's Child” poem
by Rudyard Kipling [1865 - 1936]:

I kept six honest serving men,
(They taught me all I knew),
Their names are What and Why and When
and How and Where and Who.



When the question to be answered does not match any question from this ontology, an expert will rapidly instruct CogentGPT to answer it.

Because argument construction involves the interplay of *imaginative and critical reasoning*, someone may always find a different route from evidence to the hypothesis. Therefore, *there is no such thing as uniquely correct argument from some collection of evidence to the hypotheses being entertained*.

We would therefore consider that the argumentation developed by CogentGPT is *correct but potentially incomplete*.

CogentGPT can test the answers generated by ChatGPT by comparing them with its own answers,
but it can do much more than that, as illustrated next.

Amazing Applications of CogentGPT

Rieber's REASON Challenge

The REASON (Rapid Explanation, Analysis and Sourcing Online) challenge consists of developing the technology to automatically producing comments (feedback and recommendations) on a draft analytic report, highlighting additional relevant evidence, and identifying strengths and weaknesses in the draft's reasoning. Analysts can use the comments to improve their reports.

As contrasted with current applications of structured analytic techniques, the REASON technology will automatically produce comments with no additional effort from analysts, who can use any comments they find valuable. These comments will be based on the automated application of effective structured analytic techniques.

By making specific comments on draft analytic reports, REASON technology will fit into the existing intelligence analysts' workflow. The comments will be analogous to those made by automated spelling and grammar checks, except that REASON's comments will focus on improving argumentation instead of writing.

REASON

RAPID EXPLANATION, ANALYSIS AND SOURCING ONLINE

INTELLIGENCE VALUE

REASON aims to develop novel technologies that will enable intelligence analysts to substantially improve the evidence and reasoning in draft analytic reports. Intelligence analysts sort through huge amounts of often uncertain and conflicting information as they strive to answer intelligence questions. REASON will assist and enhance analysts' work by pointing them to key pieces of evidence beyond what they have already considered and by helping them determine which alternative explanations have the strongest support. It will do this automatically and on demand by providing evidence and reasoning suggestions as the analyst works on a report. The program will exploit recent advances in artificial intelligence, not to perform the analysis or write the report, but to help analysts do it even better. As a result, decision-makers will receive analytic reports with the highest accuracy, clarity and timeliness.

REASON

CONTACT INFORMATION

PROGRAM MANAGER
Dr. Steven Rieber
steven.rieber@iarpa.gov
301-243-2087

RESEARCH AREA(S)

Analytic Reasoning, Argumentation, Artificial Intelligence, Human Computer Interaction, Human Language Technology, Information Retrieval

BROAD AGENCY ANNOUNCEMENT (BAA)

LINK(S) TO BAA

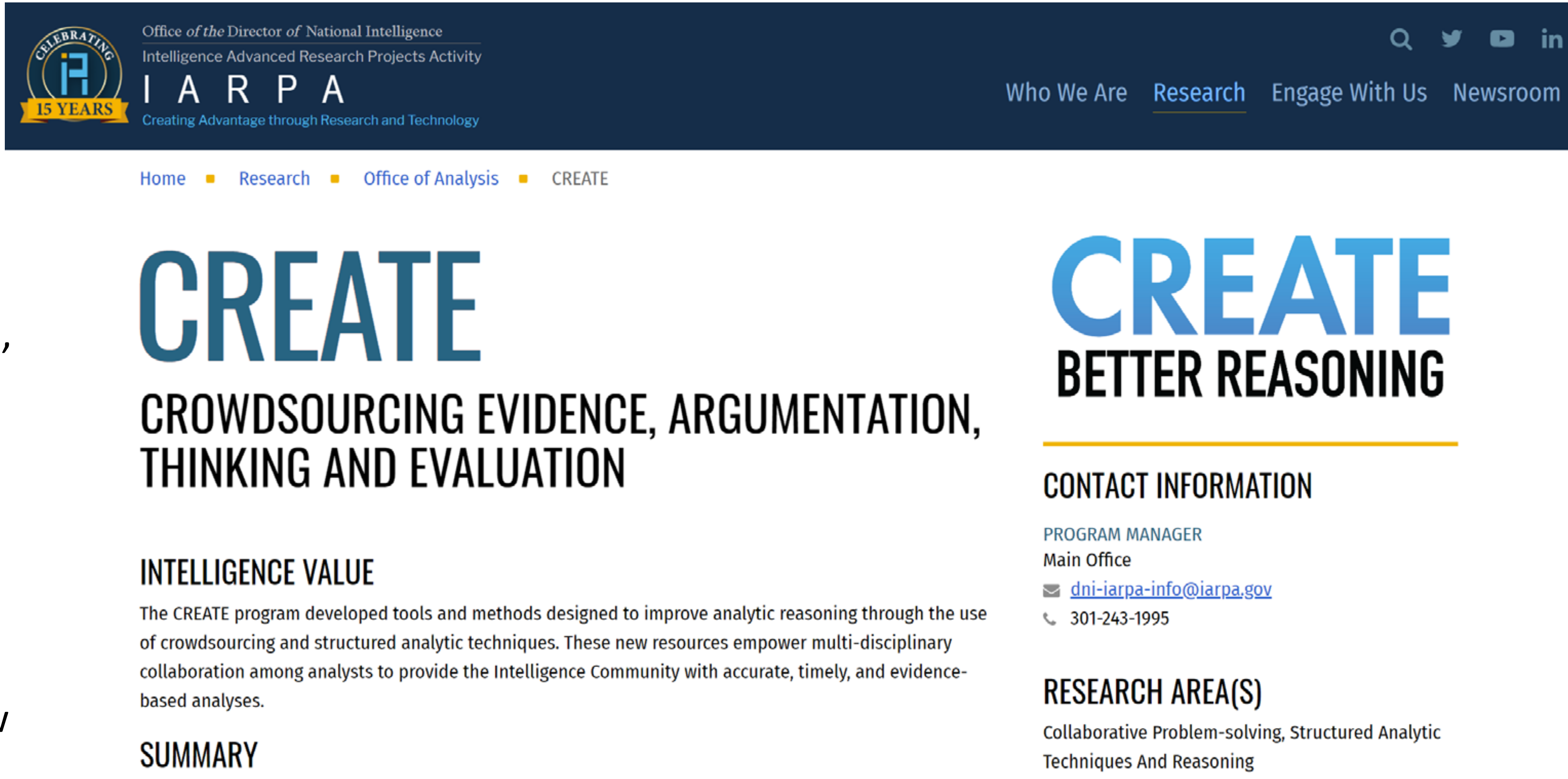
Rieber's CREATE Challenge

To develop tools and methods designed to improve analytic reasoning through the use of crowdsourcing and structured analytic techniques.

CREATE resulted in the improvement of the Cogent system.

The solutions developed using Cogent by students from 4 universities (GMU, Nebraska-Omaha, Nebraska-Lincoln, Mary Washington) were manually evaluated by their instructors.

This very laborious experimentation can now be automated with Cogent-GPT.

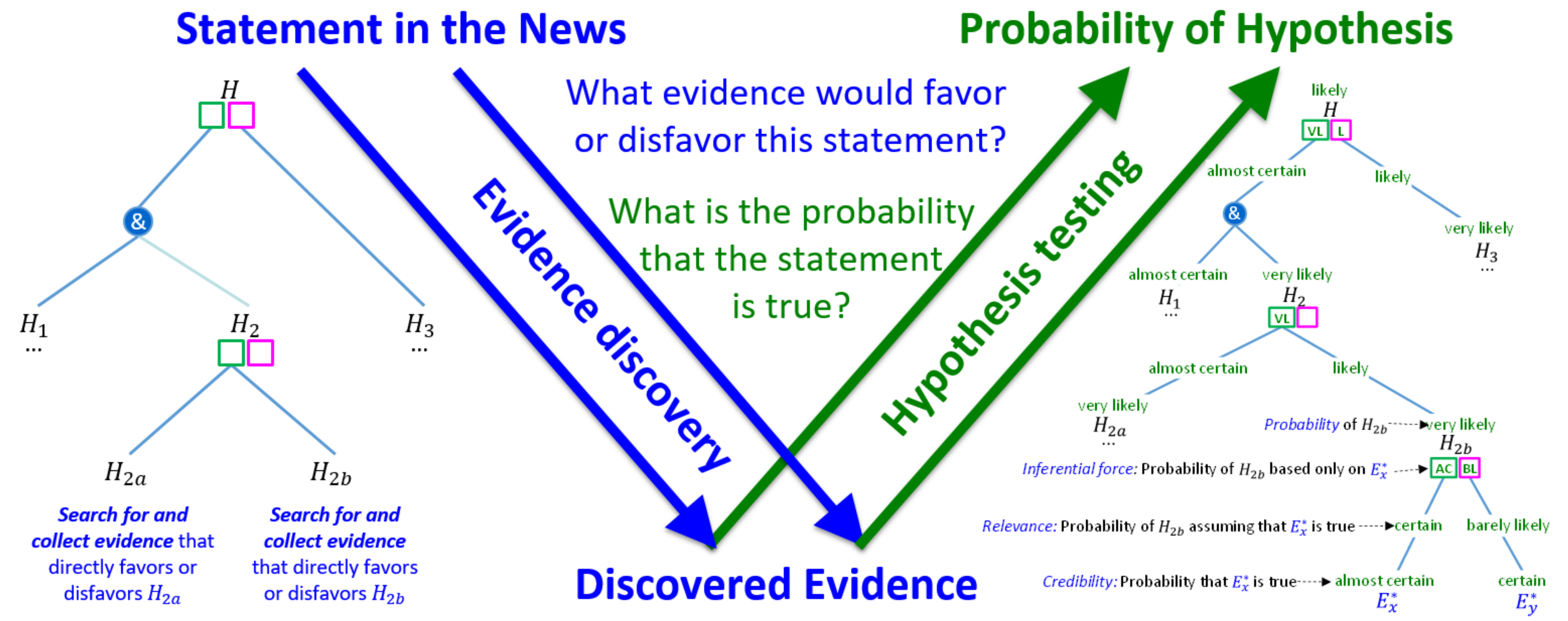


Fake News Detection

Common approaches recommended to “spot” fake new, such as to consider the source, check the URL, look for visual clues, get a second opinion, put your browser to work (https://guides.library.harvard.edu/fake).

Clearly such approaches do not work for the well-elaborated fake news by state actors, such as those that are part of the current propaganda war between Russia and Ukraine. Consider, for example, the “Ghost of Kyiv’ fake news where one of the source was former President Petro Poroshenko who had shared videos and photos purporting to show the mysterious pilot (https://www.dw.com/en/fact-check-ukraines-ghost-of-kyiv-fighter-pilot/a-60951825).

We will treat fake news detection as an intelligence analysis problem, where the claims made in the news are hypotheses to be tested.



References

Tecuci, G. (1988). *Disciple: A Theory, Methodology and System for Learning Expert Knowledge*, Thèse de Docteur en Science, University of Paris-Sud. http://lac.gmu.edu/publications/1988/TecuciG_PhD_Thesis.pdf

Tecuci, G. (1998). *Building Intelligent Agents: An Apprenticeship Multistrategy Learning Theory, Methodology, Tool and Case Studies*. London, England: Academic Press. http://lac.gmu.edu/publications/1998/TecuciG_Building_Intelligent_Agents/default.htm

Tecuci, G., Boicu, M., Ayers, C., Cammons, D. (2005). Personal Cognitive Assistants for Military Intelligence Analysis: Mixed-Initiative Learning, Tutoring, and Problem Solving, *Proceedings of the First International Conference on Intelligence Analysis*, McLean, VA, May 2-6.

Tecuci, G., Boicu, M., Cox M.T. (2007). Seven Aspects of Mixed-Initiative Reasoning: An Introduction to the Special Issue on Mixed-Initiative Assistants, *AI Magazine*, 28(2), pp.11-18, Summer.

Tecuci, G., Boicu, M., Marcu, D., Boicu, C., Barbulescu, M. (2008). Disciple-LTA: Learning, Tutoring and Analytic Assistance, *Journal of Intelligence Community Research and Development (JICRD)*, July.

Tecuci, G., Schum, D.A., Marcu, D., Boicu, M. (2014). Computational Approach and Cognitive Assistant for Evidence-Based Reasoning in Intelligence Analysis, *International Journal of Intelligent Defence Support Systems*, 5(2), pp.146-172

Tecuci, G., Marcu, D., Boicu, M., Schum, D.A. (2015). COGENT: Cognitive Agent for Cogent Analysis, *Proceedings of the 2015 AAAI Fall Symposium "Cognitive Assistance in Government and Public Sector Applications"*, Arlington, VA, Nov. 12-14.

Tecuci, G., Schum, D.A., Marcu, D., Boicu, M. (2016a). Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments: Connecting the Dots, Cambridge University Press.

Tecuci, G., Marcu, D., Boicu, M., Schum, D.A. (2016b). Knowledge Engineering: Building Cognitive Assistants for Evidence-based Reasoning, Cambridge University Press.

Tecuci G., Kaiser L., Marcu D., Uttamsingh C., Boicu M. (2018). Evidence-based Reasoning in Intelligence Analysis: Structured Methodology and System, Special Issue on Evidence-based Reasoning and Applications, *Computing in Science and Engineering*, 20:6, pp. 9-21, Nov/Dec. http://lac.gmu.edu/publications/2018/Cogent-in-CiSE-2018.pdf

Tecuci, G., (2023b) How Intelligent is Artificial Intelligence (http://lac.gmu.edu/publications/2023/How%20Intelligent%20is%20AI.pdf). *Evidence Dialogues* (https://evidencedialogues.wordpress.com/).

Tecuci G., Schum D.A., *Science of Evidence*, 2023. http://lac.gmu.edu/publications/private/SoE%20Book.pdf

Tecuci G., *Practicum in Evidence Marshaling and Argument Construction: Connecting the Dots*, 2023, http://lac.gmu.edu/publications/private/Argument%20Book.pdf

Tecuci G., Critical Thinking for Intelligence Analysis: A Gentle Introduction,2023. http://lac.gmu.edu/publications/private/CTIA%20Gentle%20Introduction.pdf

Tecuci G., Schum D.A., Critical Thinking for Intelligence Analysis: Connecting the Dots, 2023 http://lac.gmu.edu/publications/private/CTIA%20Connecting%20the%20Dots.pdf

Schum D.A., Tecuci G., The Art and Science of the Judicial Proof, 2023. http://lac.gmu.edu/publications/private/Law%20Book.pdf