

Compositional Learning-And-Reasoning for AI Complex Systems Engineering (CLARA)

Benjamin Grosf, PhD
Program Manager, Defense Sciences Office (DSO)
<https://www.darpa.mil/staff/dr-benjamin-grosf>
<https://www.darpa.mil/about-us/offices/dso>

Exploratory: Create new foundations (6.1) for learning-and-reasoning that can provide high assurance

Information Session

February 19, 2026



Program Solicitation has the Final Say

If there is any discrepancy between what is presented today and the solicitation, the solicitation takes precedence





Agenda



- Welcome Dr. Benjamin Grosf 1:00 – 1:05
- Defense Sciences Office (DSO) Overview 1:05 – 1:30
- CLARA Overview 1:30 – 2:30
 - Motivation and Background
 - Technical Dive
 - Proposal Process
 - Optional Slides
- Contracts Management Office Presentation 2:30 – 2:45
- Break + Initial Q&A submission 2:45 – 3:00
- Live Q&A 3:00 – 4:00

(Times above are approximate.)

Motivation and Background





1. CLARA aims to create high assurance, broadly applicable AI systems of systems. This will be accomplished by creating a scientific, theory-driven architectural foundation for the hierarchical composition of both Machine Learning (ML) and Automated Reasoning (AR, also known as Knowledge Representation and Reasoning) subsystems.
2. Such broad assurance is important when composing many distinct ML and/or AR systems into a broader composite system. These systems-of-systems occur in many commercial and DoW domains such as supply chain & logistics, wargaming and course-of-actions, medical guidance, and kill-web.



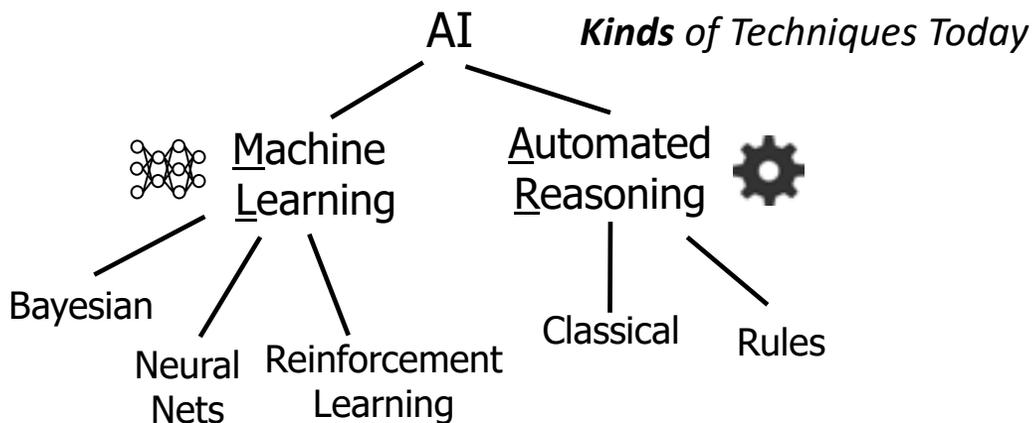
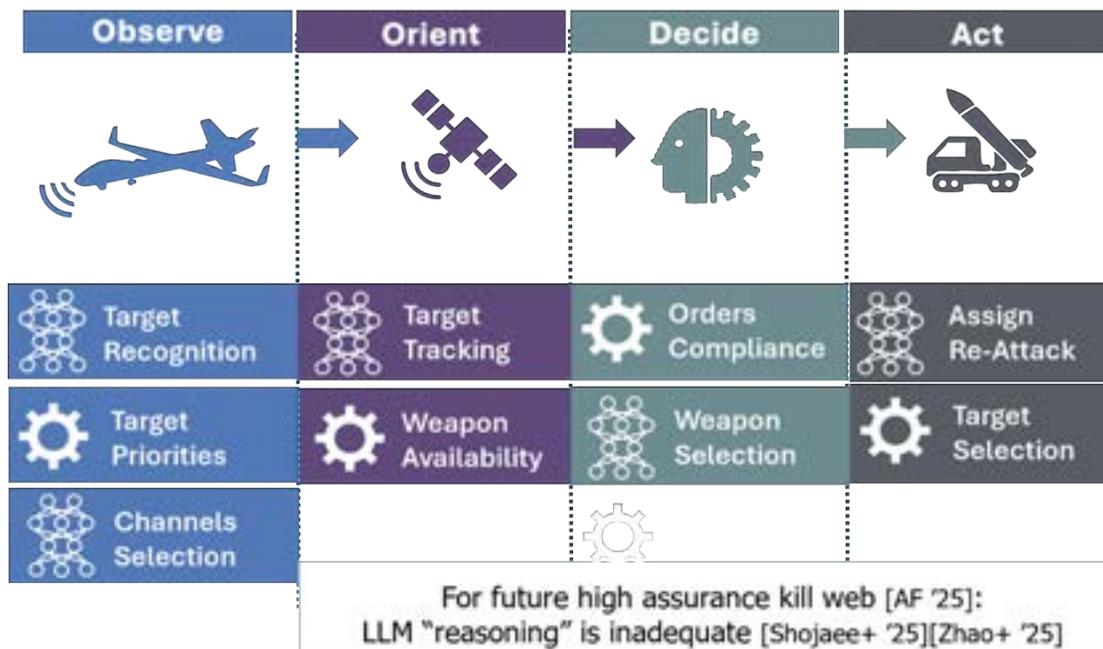
CLARA is an approach to enable Assurance that can be **Built Into** next-generation AI systems of systems for defense



Goal: Create theory-driven foundation that can provide high **assurance**, while enabling broader **applicability**, in AI systems that hierarchically compose a variety of machine learning and automated reasoning subsystems

- **Assurance** = verifiability with strong explainability
 - Via automated proofs in logic that provides shared meaning

Kill Web OODA Loop Must Compose Variety of ML and AR Example is for Illustration Purposes Only



- CLARA will: Put ML on a logic-based (AR) footing
- Potential future benefits: Robustly follows intent, lower cost to test, agile updating of knowledge, rapid adaptability, generalize better, more reusability

[Shojaee+ '25]: The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. *arxiv*.
 [Zhao+ '25]: Is Chain-of-Thought Reasoning of LLMs a Mirage?: A Data Distribution Lens. *arxiv*.

AR: Automated Reasoning, i.e., logical AI
 ML: Machine Learning
 Neural Net: a kind of ML; e.g., LLM
 Reinforcement Learning: a kind of ML
 LLM: Large Language Model, e.g., ChatGPT
 Rules: i.e., Logic Programs; a kind of AR
 Verifiability: via proof, e.g., automatic
 Explainability: for humans, ideally SMEs; incl. editability
 Tractability: computational scalability
 RoE: Rules of Engagement

[AF '25]: <https://www.af.mil/News/Article-Display/Article/4241485/air-force-battle-lab-advances-the-kill-chain-with-ai-c2-innovation>

Image Icon Licenses: AdobeStock_1615277494, AdobeStock_227445055, AdobeStock_281720863

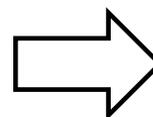


Today's ML and AR both have limitations



	Machine Learning	Automated Reasoning
Advantages	<ul style="list-style-type: none"> • Broad Applicability <ul style="list-style-type: none"> • Strong NL Dialog 	<ul style="list-style-type: none"> • High Assurance <ul style="list-style-type: none"> • Verifiable • Strong Explainability
Disadvantages	<ul style="list-style-type: none"> • Weak Assurance <ul style="list-style-type: none"> • Weak Explainability • Error Prone 	<ul style="list-style-type: none"> • Narrow Applicability <ul style="list-style-type: none"> • Weak Tractability when High Expressiveness

- Industry's dominant approach to composed systems is ML-centric, has weak **assurance**
 - Tacking on specialized AR pieces is insufficient
- Rules have fine-grain composability with high assurance, pioneered by DARPA via a theory-driven foundation
 - But can not yet be applied widely to ML



Technical Challenges for CLARA:

1. "Applicability": Extend Rules assurance widely to ML/AR techniques and composing them
2. "Tractability": Break the tension between expressiveness and tractability

Explainability: For humans. When strong, includes fine grained editability.

Verifiability: Automatically check correctness of system outputs, via automating their proof.

Expressiveness: Flexibility of knowledge/questions that can specified, and of patterns of reasoning

Tractability: Computational scalability, cf. computer science: (worse-case or average) compute-time grows as polynomial (vs. exponential or infinitary) function of input size.

SME: human Subject Matter Expert, not expert in logic/AR nor ML

AR: Automated Reasoning

ML: Machine Learning

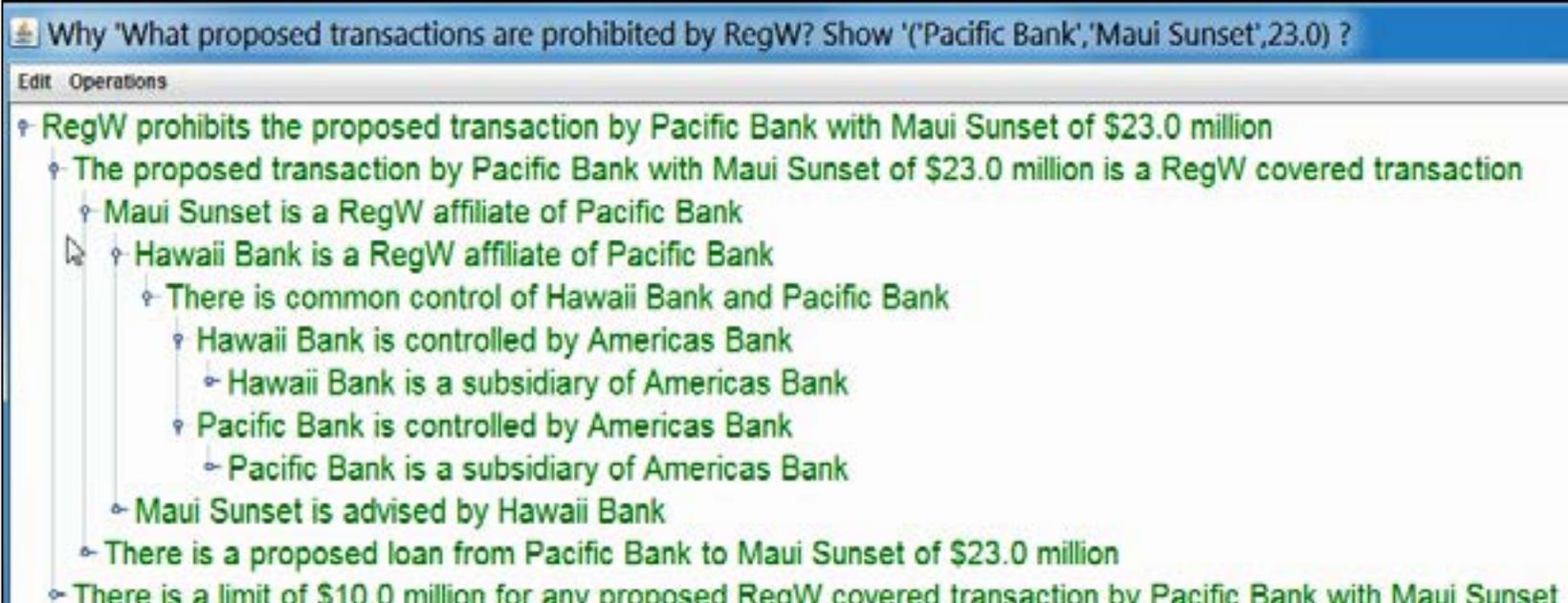
Rules: i.e., Logic Programs; a kind of AR

NL: Natural Language, e.g., English

- Automatic proof of each answer/conclusion
- Comprehensible by human engineers and users – even non-logic-experts (e.g., compliance officer)
- Line-level debugging; Hierarchical, fine-grain on up

Example: Rules for Federal Reserve Regulation W Compliance Pilot [Grosf+ '15]*

Query →



Explanation / Proof

- *Comparison to app-task SOA '15 (which was essentially manual)
- Provides assurance & audit trail
- ~1,000x faster,
- ~1,000,000x cheaper (marginal cost),
- ~100x lower false alarm rate

Rules Explainability can benefit Assurance

Rules: Logic Programs, a kind of AR

AR: Automated Reasoning

~: order of magnitude

[Grosf+ '15] Grosf, B., et al,

Automated Decision Support for Financial Regulatory/Policy

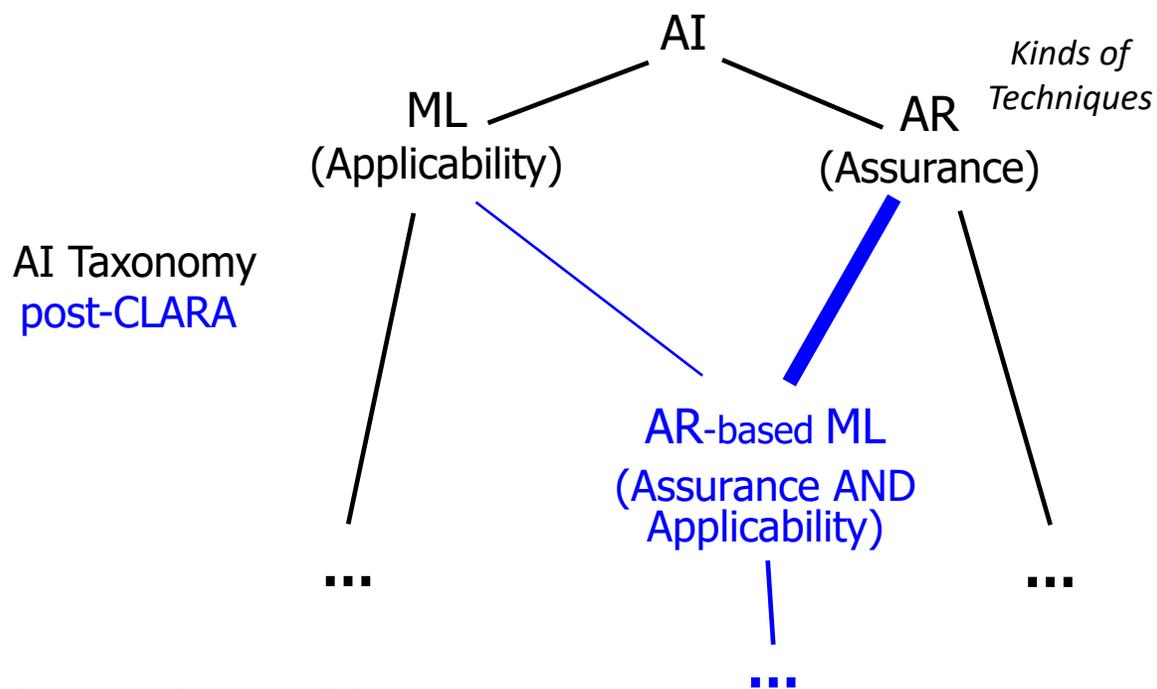
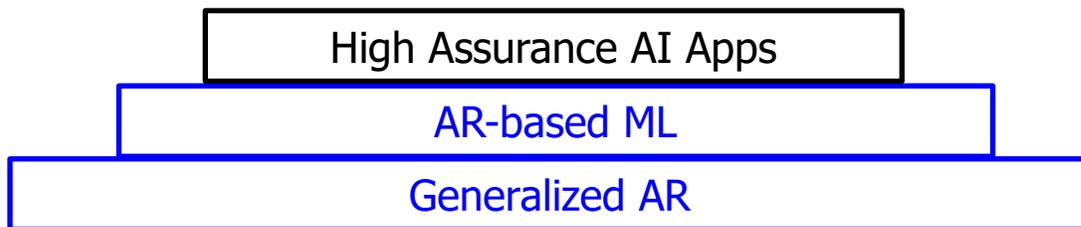
Compliance, using Textual Rulelog, 9th Intl. Web Rule Symposium

(RuleML-2015) industry track demonstration and paper.

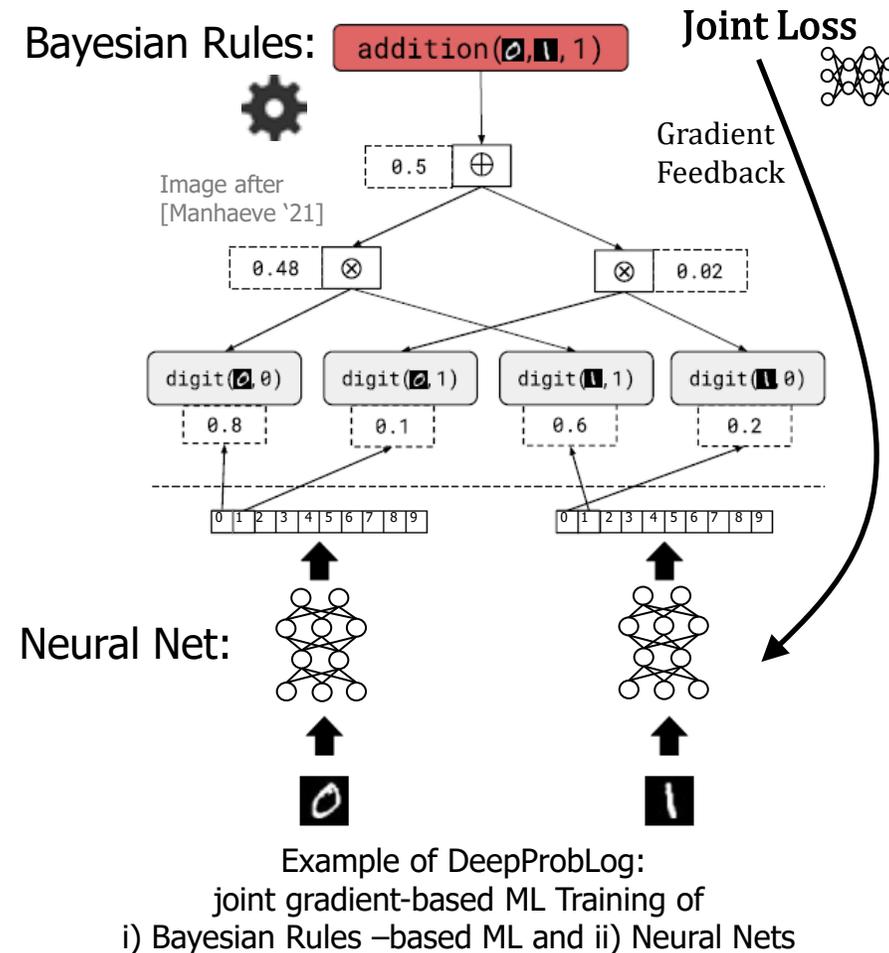
Technical Dive



Architecture of AI Systems of Systems *post-CLARA*



Composing AR + ML is possible



Bayesian: the usual kind of probabilistic Rules: Logic Programs, a kind of AR
AR: Automated Reasoning
ML: Machine Learning
Neural Net: a kind of ML

[Manhaeve+ '21]: Neural probabilistic logic programming in DeepProbLog. *Artificial Intelligence* journal.

[DeepProbLog '24]: Manhaeve+. <https://github.com/ML-KULeuven/deepproblog>

Technical Challenge #1: "Applicability": Extend Rules assurance widely to ML/AR techniques and composing them

- Can leverage and combine two recent expressiveness advances:
 - Bayesian Probabilistic Rules, notably for AR-based ML Inferencing
 - Higher-Order Rules, notably for AR-based ML Training
- Other potential opportunities to create hierarchical, fine-grain compositions of AR + ML:
 - Rules as tensor operations [Inoue+ '24a][Inoue '24b]

[Manhaeve+ '21][Winters+ '22] [DeepProbLog '24][ProbLog2 '25]
[cplint/PITA '25][Riguzzi+ '23]

[ErgoAI '25][Grosf+ '23][Swift+ '24][Kifer+ '24]

Higher-order is a type of meta: knowledge about knowledge; reasoning about reasoning. It's the general case of logic that permits logical relations and functions to be general logical terms, rather than only logical constants; here, also includes defeasibility, reification, and annotation.

Bayesian: the usual kind of probabilistic

[Manhaeve+ '21]: Neural probabilistic logic programming in DeepProbLog. *Artificial Intelligence journal*.

[Winters+ '22]: DeepStochLog: Neural Stochastic Logic Programming. *Proc. AAAI conference*.

[DeepProbLog '24]: Manhaeve+. <https://github.com/ML-KULeuven/deepproblog>

[ProbLog2 '25]: De Raedt+. <https://github.com/ML-KULeuven/problog>

[Riguzzi+ '23]: *Foundations of Probabilistic Logic Programming: Languages, Semantics, Inference and Learning*, 2nd ed.

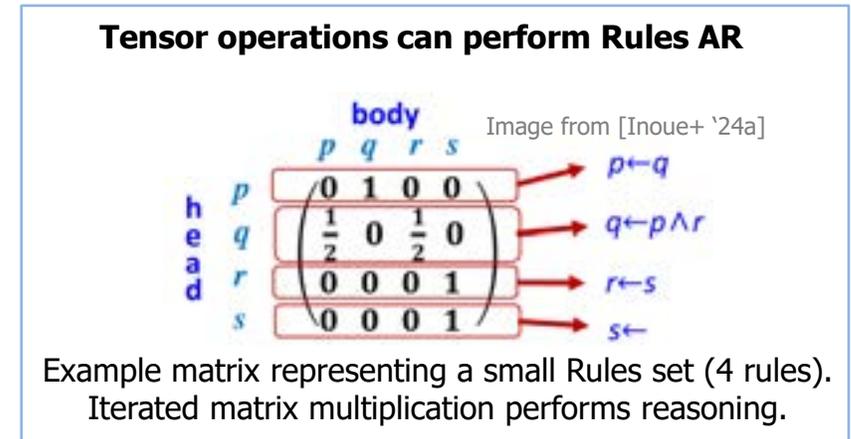
[cplint/PITA '25]: Riguzzi+, cplint (incl. PITA). https://friguzzi.github.io/cplint/_build/latex/cplint.pdf

[ErgoAI '25]: Kifer+, ErgoAI software, manuals & tutorials. <https://github.com/ErgoAI>

[Grosf+ '23]: Ergo: A Quest for Declarativity in Logic Programming. *Prolog: The Next 50 Years*.

[Swift+ '24]: Multi-paradigm Logic Programming in the ErgoAI System. *Proc. LPNMR*.

[Kifer+ '24]: Multi-paradigm Logic Programming in the ErgoAI System. Submitted to *TPLP journal*.



- Neural Nets as fuzzy probabilistic AR
- Concept bottlenecks, to scaffold [Koh+ '20][Hu+ '25]

[Inoue+ '24a]: Linear Algebraic Approaches to Logic Programming. ICLP conf. invited talk.

[Inoue '24b]: Algebraic Connection Between Logic Programming and Machine Learning. *Proc. FLOPS*.

[Riegel+ '20]: Logical Neural Networks. *arxiv*.

[Fagin+ '24]: Foundations of reasoning with uncertainty via real-valued logics. *PNAS journal*.

[Koh+ '20]: Concept Bottleneck Models. *Proc. Intl. ML Conf.*

[Hu+ '25]: Semi-supervised Concept Bottleneck Models. *arxiv*.

fuzzy: non-Bayesian probabilistic, cf. triangular norms and similar

Rules: Logic Programs; a kind of AR
ML: Machine Learning
AR: Automated Reasoning
Neural Nets: a kind of ML



Technical Challenge #2: "Tractability": Break the tension between expressiveness and tractability

- Can leverage opportunity to create **tractable** generalized AR via bounded rationality techniques, notably Restraint & HiLog
 - Tractable Higher-Order
[ErgoAI '25][Grosf+ '23][Swift+ '24][Kifer+ '24]
 - Tractable Probabilistic, via: Higher-Order about "Structure" of Probabilistic

"Structure" of Probabilistic Enables Tractability

Structure (ex. cluster size = 10)	Probability Distribution	Size of Distribution
No Structure (General case)	$p(X_1, \dots, X_{600})$	2^{600} ($> 10^{180}$)
Multiplicative Structure (Conditional Independence)	$p(X_1, \dots, X_{10}) * p(X_{11}, \dots, X_{20})$ $* \dots * p(X_{591}, \dots, X_{600})$	$60 * 2^{10}$ ($< 10^5$)
Additive Structure (note: normalized)	$f(p(X_1, \dots, X_{10})) + f(p(X_{11}, \dots, X_{20}))$ $+ \dots + f(p(X_{591}, \dots, X_{600}))$	$60 * 2^{10}$ ($< 10^5$)

Notation: X_j 's are logical formulas. $p(\cdot)$ is probability. $p(X_i, \dots, X_k)$ is a joint distribution. $f(\cdot)$ is a function. Distribution size: # parameters in probability distribution

Simple Examples of: "Structure" of Probabilistic

bounded rationality: a type of selective don't-care, in knowledge and reasoning
 Restraint: a sub-type of bounded rationality, that employs the truth value "undefined" (i.e., don't-care)
 HiLog: a sub-type of bounded rationality, that syntactically translates higher-order into first-order
 higher-order is a type of meta: knowledge about knowledge; reasoning about reasoning.
 It's the general case of logic that permits logical relations and functions to be general logical terms, rather than only logical constants; here, also includes defeasibility, reification, and annotation.

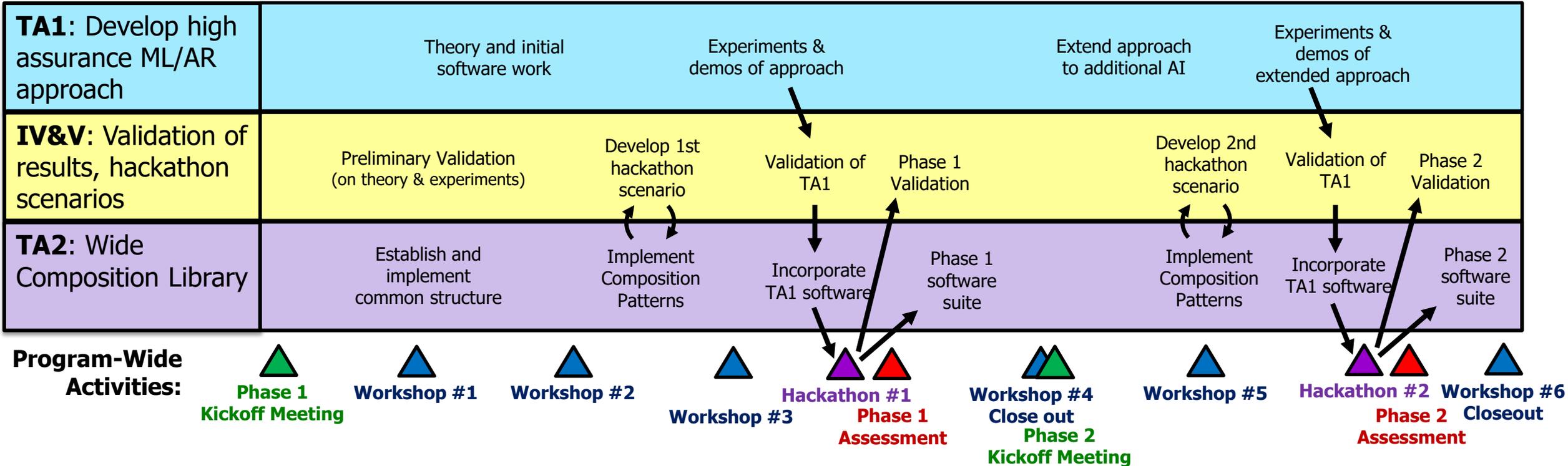
[ErgoAI '25]: Kifer+, ErgoAI software, manuals & tutorials. <https://github.com/ErgoAI>
 [Grosf+ '23]: Ergo: A Quest for Declarativity in Logic Programming. *Prolog: The Next 50 Years*.
 [Swift+ '24]: Multi-paradigm Logic Programming in the ErgoAI System. *Proc. LPNMR*.
 [Kifer+ '24]: Multi-paradigm Logic Programming in the ErgoAI System. Submitted to *TPLP* journal.



Program Schedule has quarterly Milestones & TA interactions



Program Phase	1					2		
3-Month Increment	1	2	3	4	5	1	2	3



- **Workshops:** Intense discussion
 - Strengths and weaknesses of approaches
 - How best to compose. Identify commonalities – for library.
- **Hackathons:** Integration exercises
 - TA1 performers each solve wide-scope scenarios from IV&V
 - Each uses potentially all tools in library

IV&V may begin before competitive performers; working on developing a generalized AR system combining probabilistic with higher-order and restraint, that competitive performers could build upon.



Metrics & Evaluation

address Assurance, Applicability, and Tractability



TA	Category	Metric for performer approach	Phase 1	Phase 2
TA1	Assurance	Verifiability without loss of performance <ul style="list-style-type: none"> Based on automatic proofs (sound, complete, approximation) Includes logical explainability that's hierarchical, fine-grain, natural deduction style, with low (≤ 10) unfolding expansion Performance error rate SOA: Identified by performer + IV&V 	Fully Verifiable; Performance Error rate \leq SOA	Fully Verifiable; Performance Error rate \leq SOA
TA1	Applicability	Multiplicity of Kinds in Composition <ul style="list-style-type: none"> Kinds* are specified programmatically (incl. in solicitation) Intra-approach: # of ML/AR kinds tightly composed Inter-approach: Patterns** specified by IV&V incl. hackathon 	Intra: ≥ 2 kinds (≥ 1 ML & ≥ 1 AR) Inter: Hackathon composition patterns	Intra: ≥ 3 kinds (≥ 2 ML & ≥ 1 AR) Inter: Hackathon composition patterns
TA1	Tractability	Compute Time computational time complexity	For Inferencing: Polynomial	For Inferencing & Training: Polynomial
TA2	Applicability	Wide Integration Competence in Composition <ul style="list-style-type: none"> Integrates functionally the set of TA1 tools, in specific composition patterns***, during programmatic tests incl. hackathons 	IV&V defines and selects patterns, creates rubric Scope: All TA1 tools	IV&V defines and selects patterns, creates rubric Scope: All TA1 tools

* Example kinds: Convolutional Neural Net, Bayesian Rules, Generalized Additive Models

** Similar to TA2 Applicability metric, then particularized for the various TA1 performers

*** IV&V will be selecting specific relevant composition patterns (working ahead of time with the performers to define those most sensibly and flexibly) and then observing and evaluating how well the tools and library work functionally (largely pass/fail) in those composition patterns during use case tests – incl. hackathons – to solve the test problems.

ML: Machine Learning
 AR: Automated Reasoning
 Rules: Logic Programs, a kind of AR

polynomial: as function of knowledge input size (contrast with exponential or infinitary). Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)



Metrics & Evaluation (Continued)



TA	Category	Metric for performer approach	Phase 1	Phase 2
TA1	Assurance • Reliability	Composed Task Reliability of ML+AR system • Correct* decisions, including on edge cases • CLARA built versus traditionally built (SOA)	Composed Task Reliability > SOA	Composed Task Reliability > SOA
TA1	Assurance • Ease of Training	Sample Complexity** in Training • Number/size of ML training data needed to reach a given level of task competence, when training for new task/environment	NA***	Sample Complexity < SOA

* Correctness is expected to be in terms of area under ROC curve (AUROC) or similar

** CLARA will allow users to enter/modify model knowledge in forms (e.g., rules) other than additional training examples

*** In the program overall, focus on training efficiency will be in Phase 2 rather than Phase 1 (see also the Tractability metric)



CLARA will create a scalable, highly reusable, algorithmic foundation for high assurance in AI systems of systems

- **Long-term Vision:** Provide ubiquitous basis for engineering of AI systems of systems that can reach high assurance and spur creation of AI techniques that more deeply combine ML and AR in crucial realms
→ *kill web, supply chain & logistics, wargaming, autonomous systems; financial, medical, science*
- **Potential Transition/Next Steps:** Follow-on efforts with research, testing, training org's to
 - Explore applications in defense; build commercial community
 - Increase TRL, mature tools; mindshare, standardize, e.g., requirements, web



Turn the industry paradigm inside-out:
Integrate AR throughout the guts of ML



ML: Machine Learning
AR: Automated Reasoning
TRL: Technology Readiness Level



Several arenas for applications of Learning-and-Reasoning with high Assurance



RoE: Rules of Engagement
LR: Learning-and-Reasoning

Potential Impact Areas

Kill Web

- Ensure guardrails are built *into* models, not tacked on outside
 - Ex.: don't target if too close to school bus
- Line-level logical understanding

Compose:

Sensor detection systems + RoE + target effects + decision support systems, etc.



Supply Chain & Logistics

- Amalgamate inferences & data from independent sub-systems
 - E.g., low-level tactical, mid-level coordination
- Modular & hierarchical

Compose:

Production models + resource allocation systems + time to arrival models, etc.



Wargaming

- Explainable and decomposable decision paths

Compose:

Sub-game systems: Supply chain models + unit interaction models + game scenario systems, etc.



CLARA benefits *any* system-of-AI-systems and creates a foundation for models with assurance *built in*

Image credits: <https://nara.getarchive.net/media/medium-close-shot-lto-r-flight-crew-cpt-joseph-speight-nav-1lt-dave-jesurun-0cebc7>
https://commons.wikimedia.org/wiki/File:Main_maritime_shipping_routes.png <https://www.pexels.com/photo/birds-eye-view-photo-of-freight-containers-2226458/>

Further Image Credits: <https://www.flickr.com/photos/joncutrer/43252568250>
<https://nara.getarchive.net/media/three-unmanned-aerial-systems-uas-soar-in-the-sky-99dd32>

Proposal Process





Solicitation Dates

- Information Session: February 19, 2026
- **Solicitation** Posted: Day 0 ~February 9, 2026
- Proposer Profile for **Teaming**: February 23, 2026
- **Abstracts** Due: Day 20 ~March 2, 2026
 - Abstracts are recommended but not required
 - Feedback will be provided in a timely manner
- **Full Proposals** Due: Day 60 ~April 10, 2026

Procedure for Future Questions & Answers

- **Questions** can be submitted to CLARA@darpa.mil, include "[FAQ]" in the subject line
- Questions and answers are posted publicly; do not include proprietary information
 - This will include questions and answers from the Q&A live period at 3:00 today
- FAQ Submission Deadline: Day 50, ~April 1, 2026 4:00 p.m. EDT.

Website

- Solicitation link and Information Session recording will be posted to:
<https://www.darpa.mil/work-with-us/opportunities> and later to <https://www.darpa.mil/research/programs/clara>

Read the Solicitation!



TA interaction:

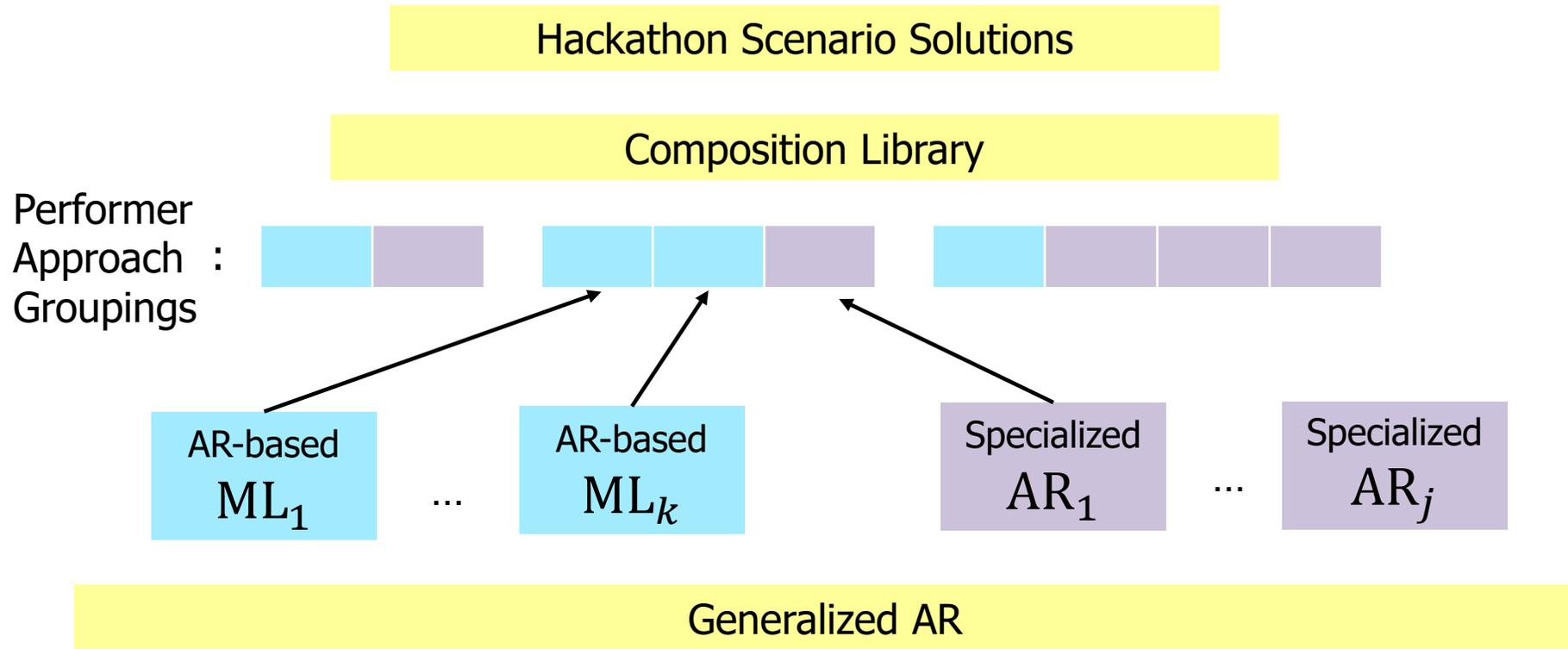
- Proposers may choose to submit separate proposals for both TA1 and TA2 efforts
- HOWEVER, No organization, whether a prime, subcontractor, individual, or any other capacity, will be awarded work on both TA1 and TA2
- Associate Contractor Agreements (ACA's) will be expected between each TA1 performer and each TA2 performer.

Proposal Guidance:

- Abstracts should emphasize technical approach and basis for confidence in success
- Make sure to relate your proposal to the relevant TA's goals, metrics, and the proposal criteria laid out in the relevant sections of the solicitation
- Clearly describe the role and expertise of each key member of your team, and show how each effort contributes to your envisioned CLARA system
- Assume a start date of June 2026

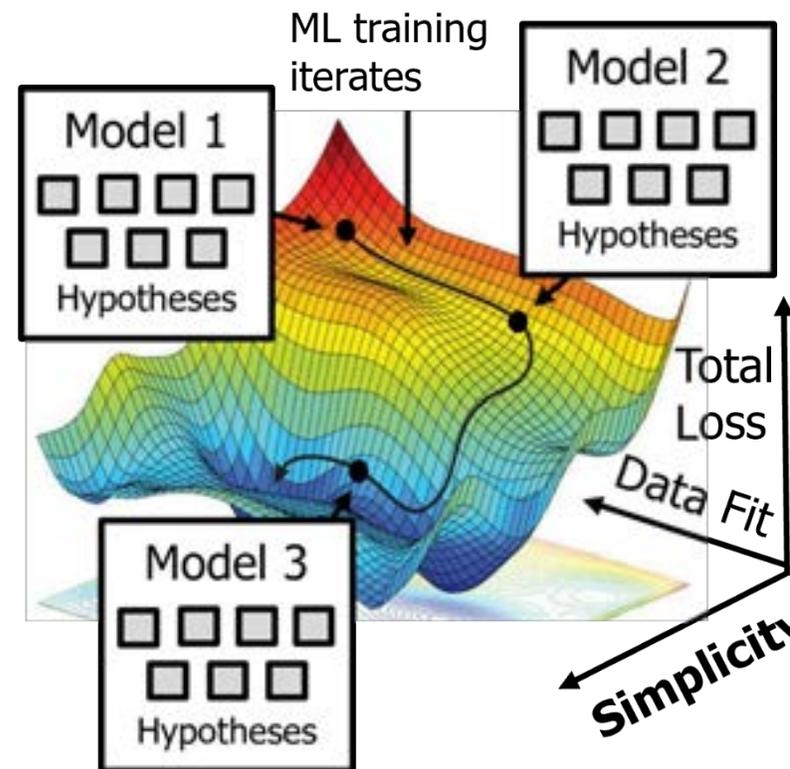
Optional Slides





- Summary: Can leverage higher-order expressiveness for: ML training

- AR-based ML training is searching through a space of candidate ML models, thus involves:
 - space and structures of ML models
 - preferences and changes between models



Higher-order is used in **ML training**, including to specify preference for models that are simpler and have more tractable structures

higher-order: is a type of meta: knowledge about knowledge and reasoning about reasoning.
 It is the general case of logic that permits logical relations and functions to be general terms rather than only logical constants; here, also includes reification, annotation, and defeasible
 defeasible: a.k.a. argumentation: able to flexibly express change (in world and beliefs), exceptions, and debate

AR: Automated Reasoning
 ML: Machine Learning
 Image background credit: Zhu, W., Qiu, P., Chen, X., Yang, Z., Sotiras, A., Razi, A., & Wang, Y. (2025). How effective can dropout be in multiple Instance Learning? Retrieved from <http://arxiv.org/abs/2504.14783>

A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J and M stand for Burglary, Earthquake, Alarm, John Calls, and Mary Calls, Respectively.

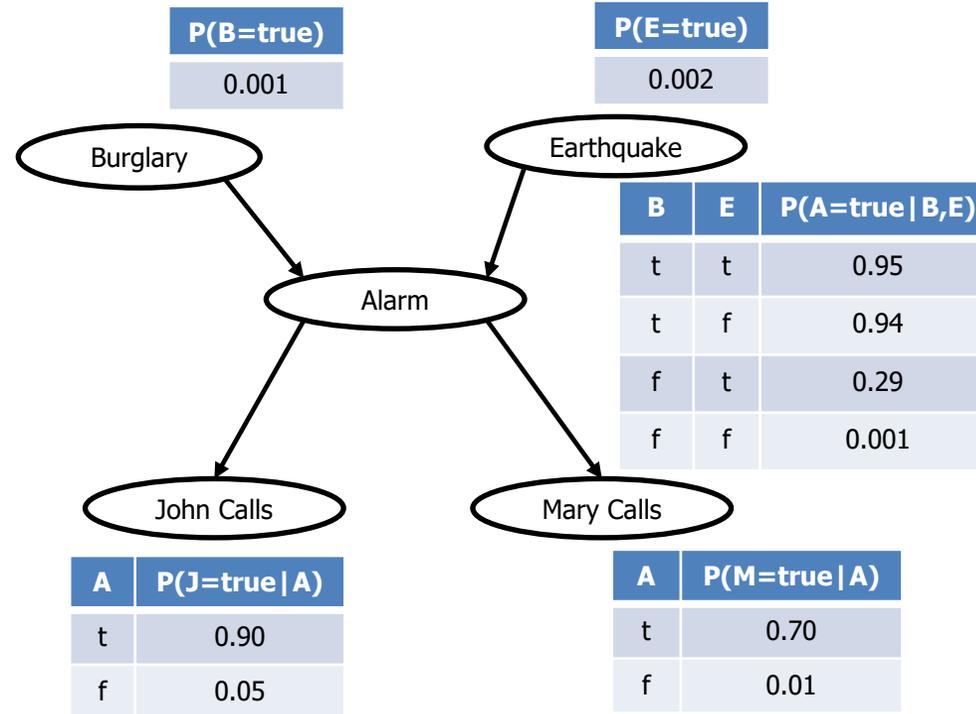
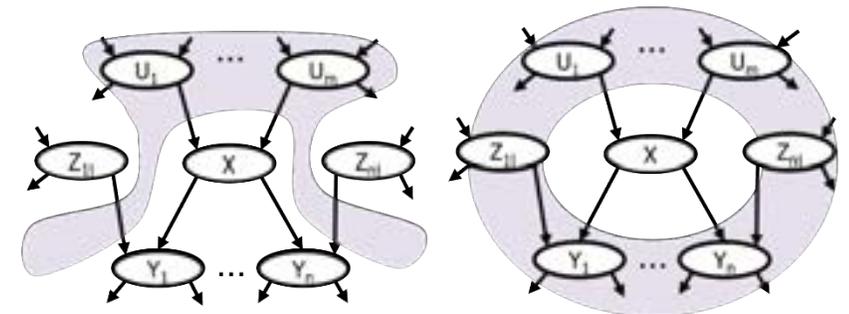


Image and Caption from: [Russell+ '19]

Bayes Nets are useful in both AR and ML



Structure of probabilistic in Bayesian Networks AR -- examples: a node X is conditionally independent of (a) its non-descendants given its parents; and (b) all other nodes given its Markov blanket [Russell+ '19]

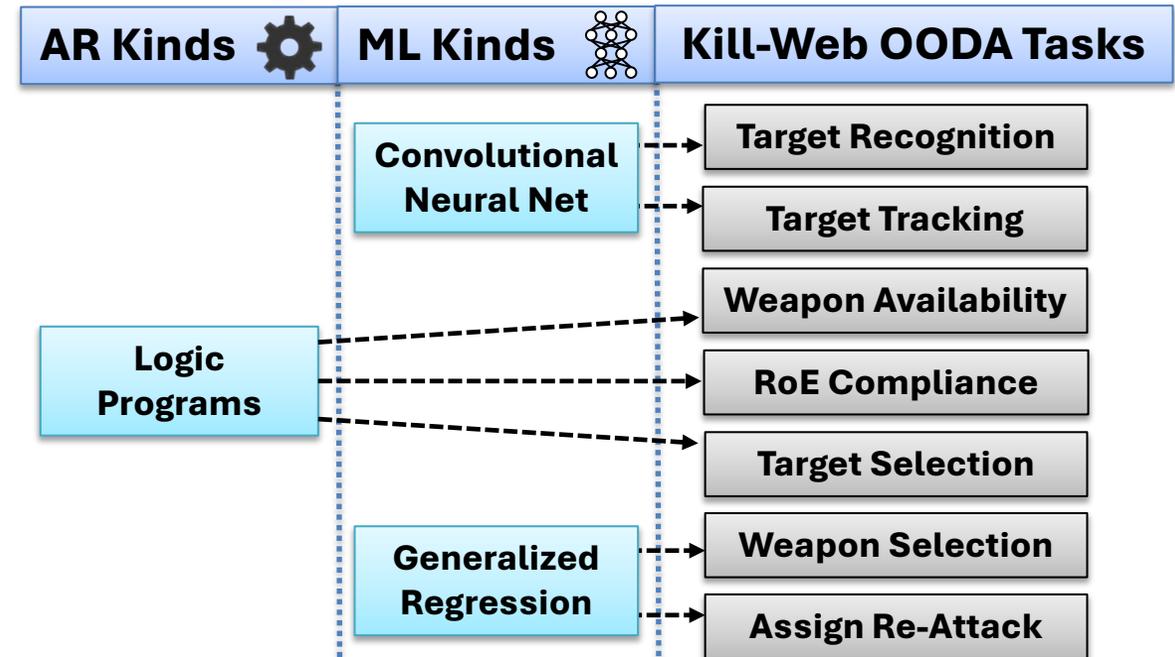


Backup: Hackathons push individual performers towards wide-scope compositional tools & applications



- Hackathons:
 - Each TA1 performer uses TA2 library to solve scenario – by widely composing
 - Requires wide-composition awareness by individual performers
 - Incentivizes clarity, generality, explainability, usability of individual performer approaches

Example Hackathon Scenario (Sketch)



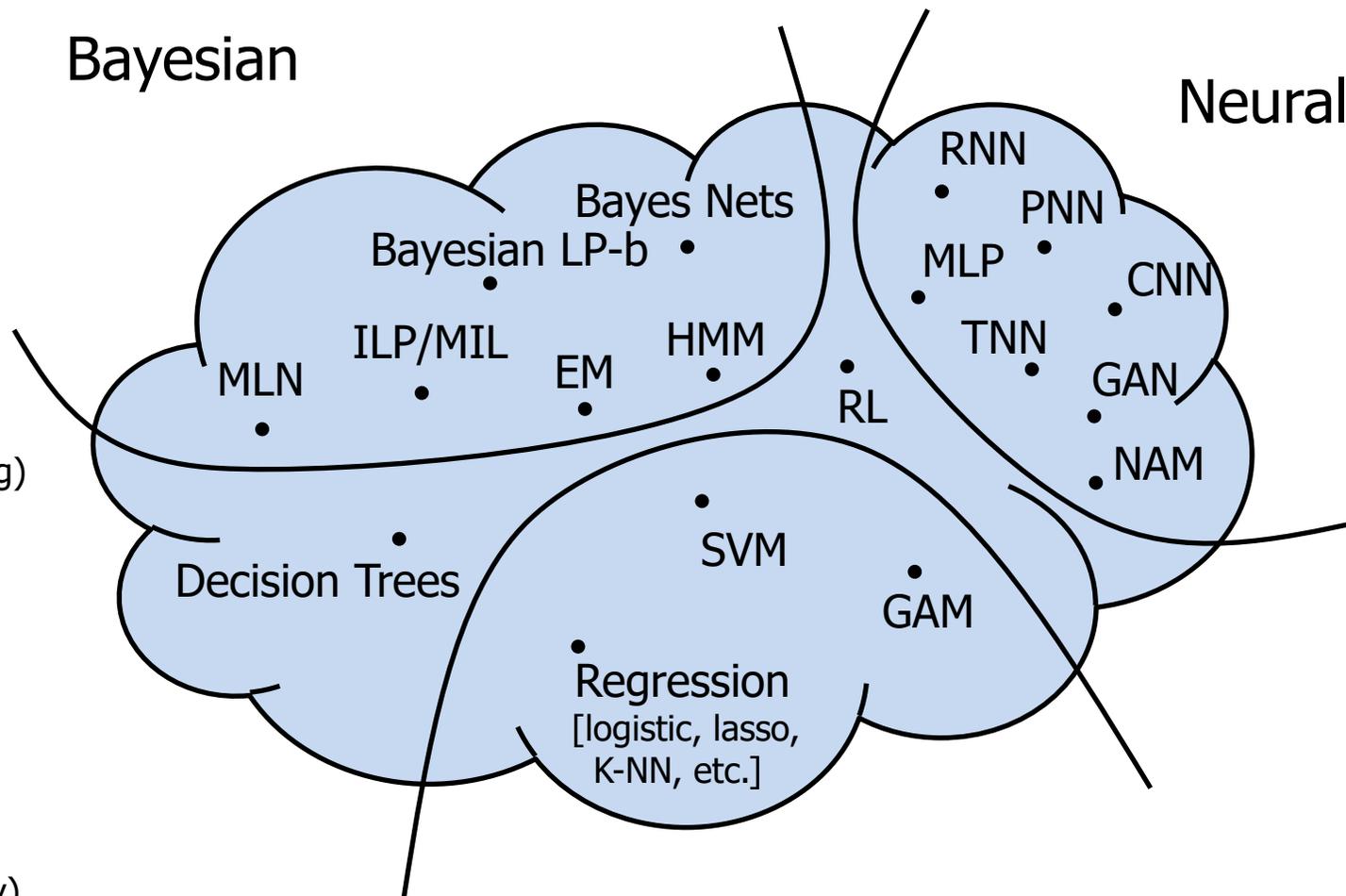
Key:
---> Potential kind for task

Drives wide composition.
Initial integration for application scenarios.

- CNN: Convolutional Neural Network
(incl. GNN: Graph Neural Network)
- EM: Expectation Maximization
- GAM: Generative Additive Model
- GAN: Generative Adversarial Network
- HMM: Hidden Markov Model
- ILP: Inductive Logic Programming
- K-NN: K- Nearest Neighbors (clustering)
- LP-b: Logic Programs -based
- MIL: Meta Interpretive Learning
- MLN: Markov Logic Network
- MLP: Multi-Layer Perceptron NN
- NAM: Neural Additive Model
- NN: Neural Network (NN)
- PNN: plain NN
- RL: Reinforcement Learning
- RNN: Recurrent Neural Network
(incl. LSTM: Long Short Term Memory)
- SVM: Support Vector Machine
- TNN: Transformer NN

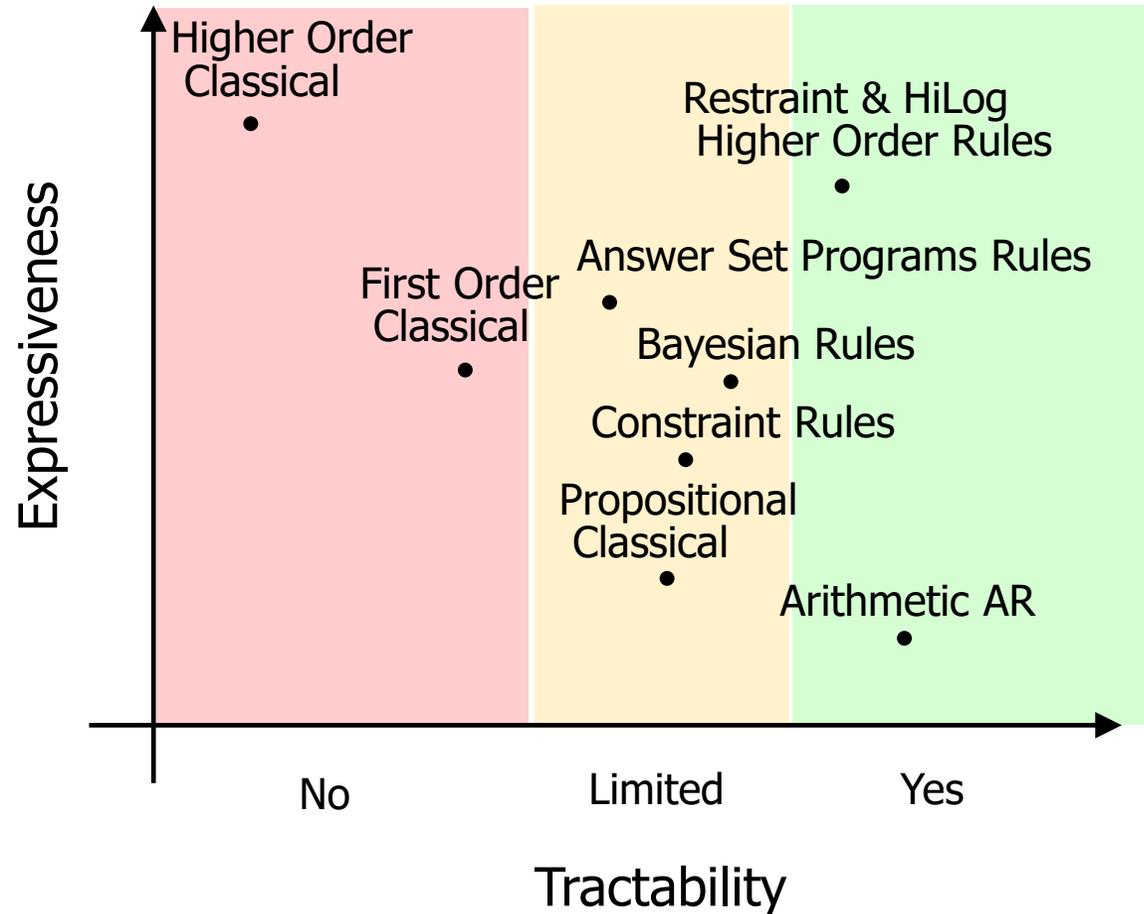
Bayesian

Neural Networks



Arithmetic

Space of AR





www.darpa.mil