PROJECT UPDATE

Knowledge-Driven Machine Learning

Jay Thiagarajan
Machine Intelligence Group
Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

Our Team

Jay Thiagarajan

Rushil Anirudh

Mark Heimann

T.S. Jayram

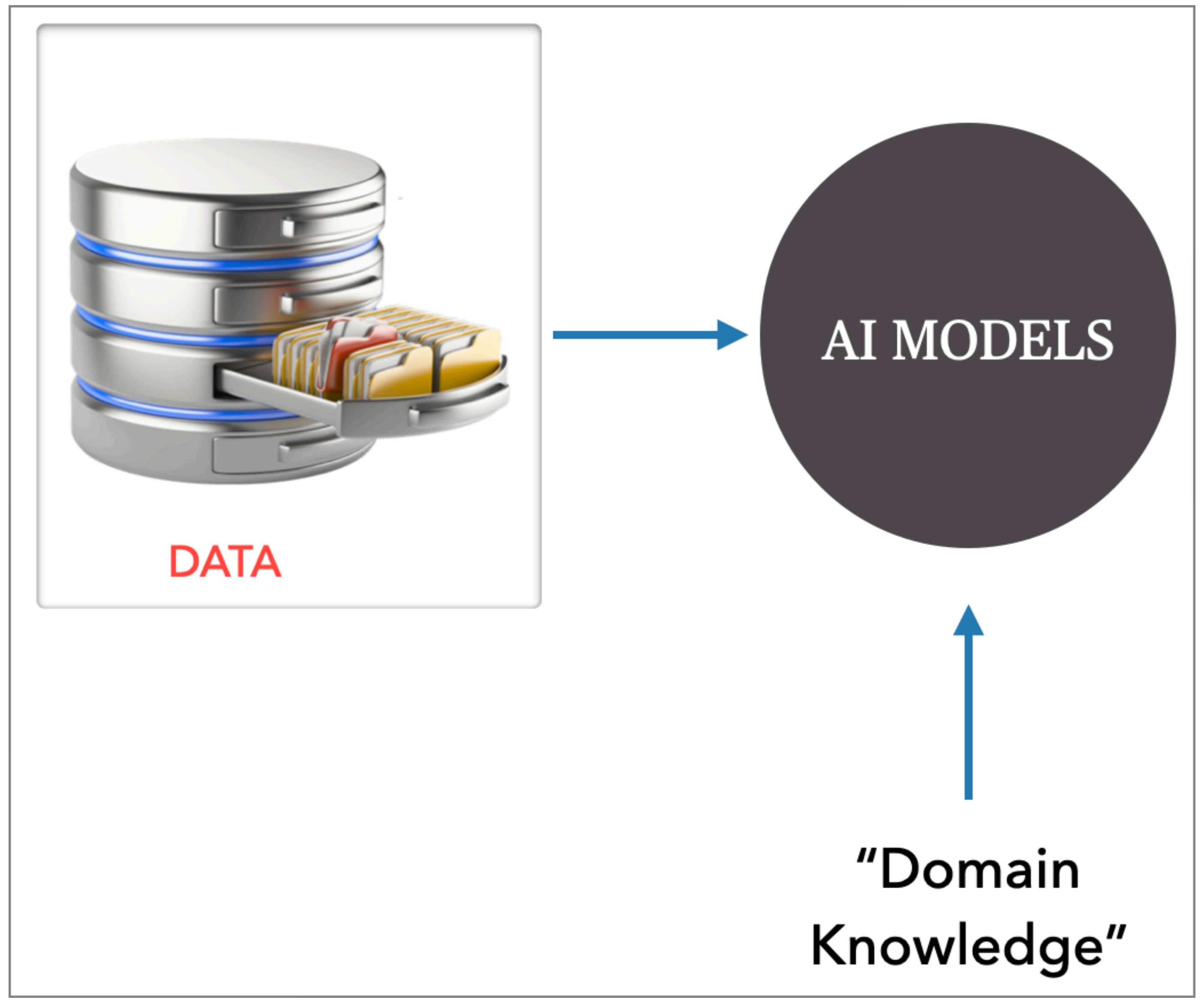
Bhavya Kailkhura

Students

Puja Trivedi
(UMich)

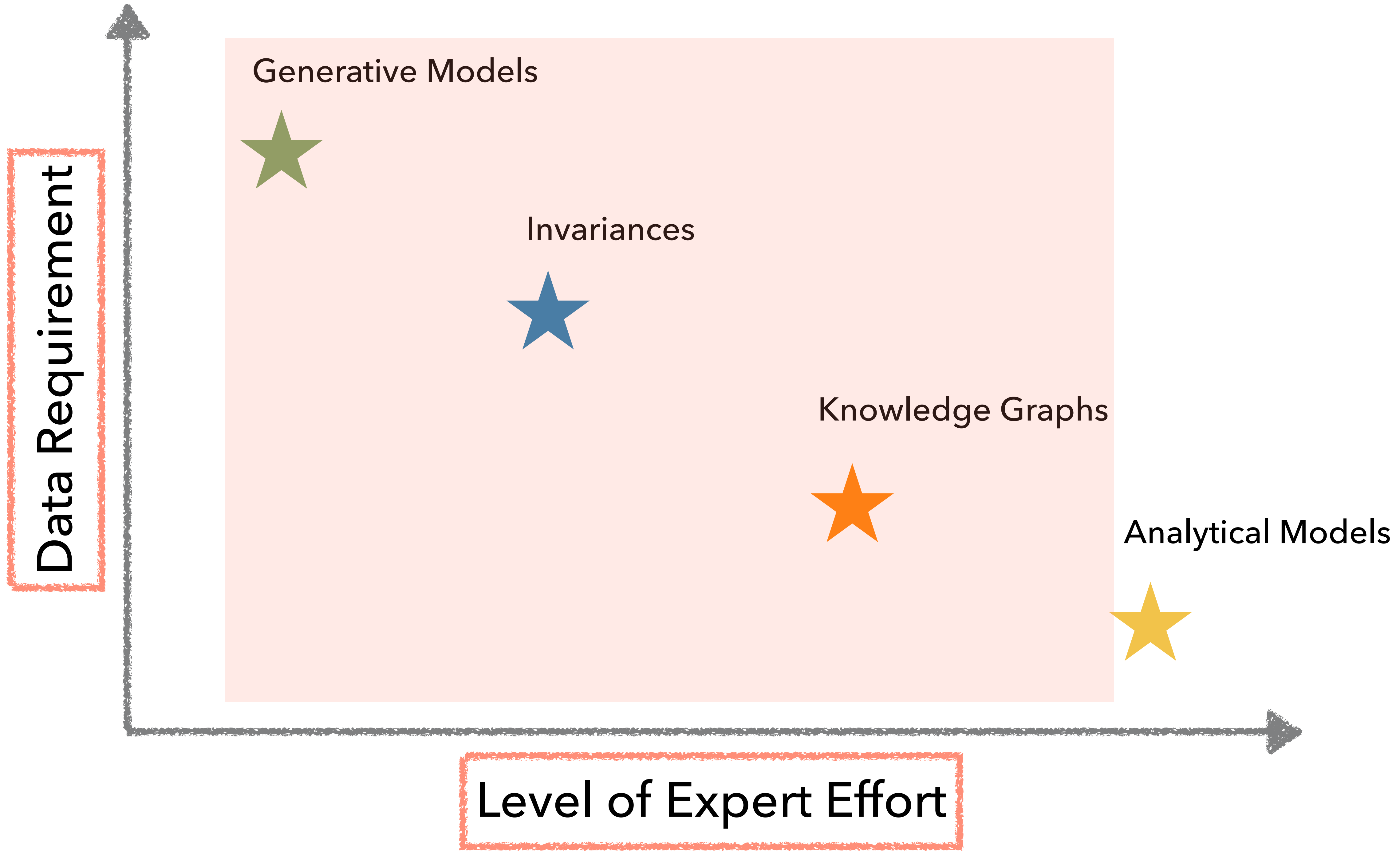
Rakshith S
(ASU)

Project Overview



Need for physically grounded, data-efficient and consistent AI models

How does one express domain knowledge for ML?



In this project, we develop the KDML framework to support machine learning with different levels of knowledge availability

Thrust 1: Providing Inductive Biases via Knowledge Graphs

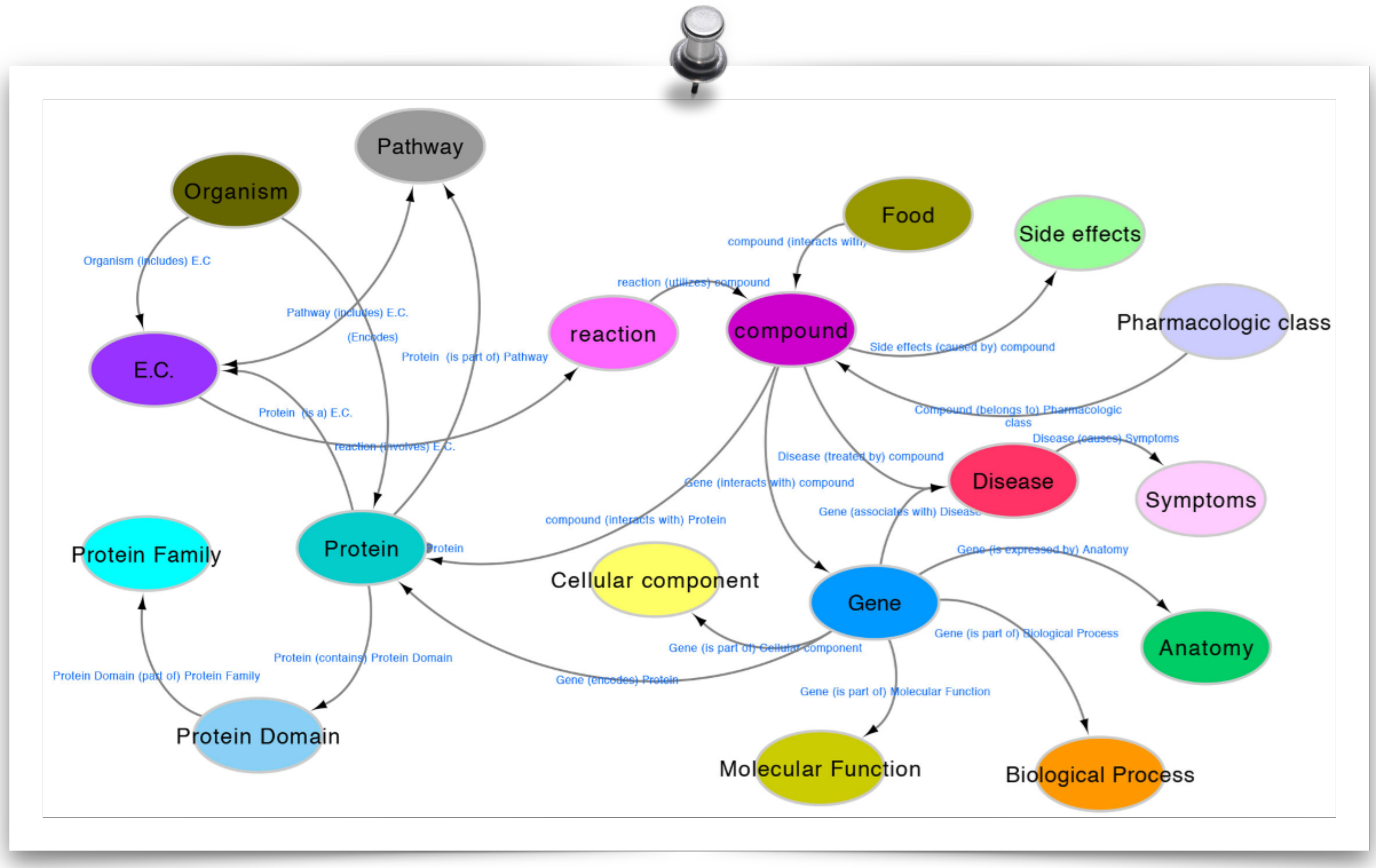


Image source: SPOKE (https://spoke.ucsf.edu/)

Knowledge graphs provide a convenient way to specify domain knowledge without requiring construction of analytical descriptions of a system.

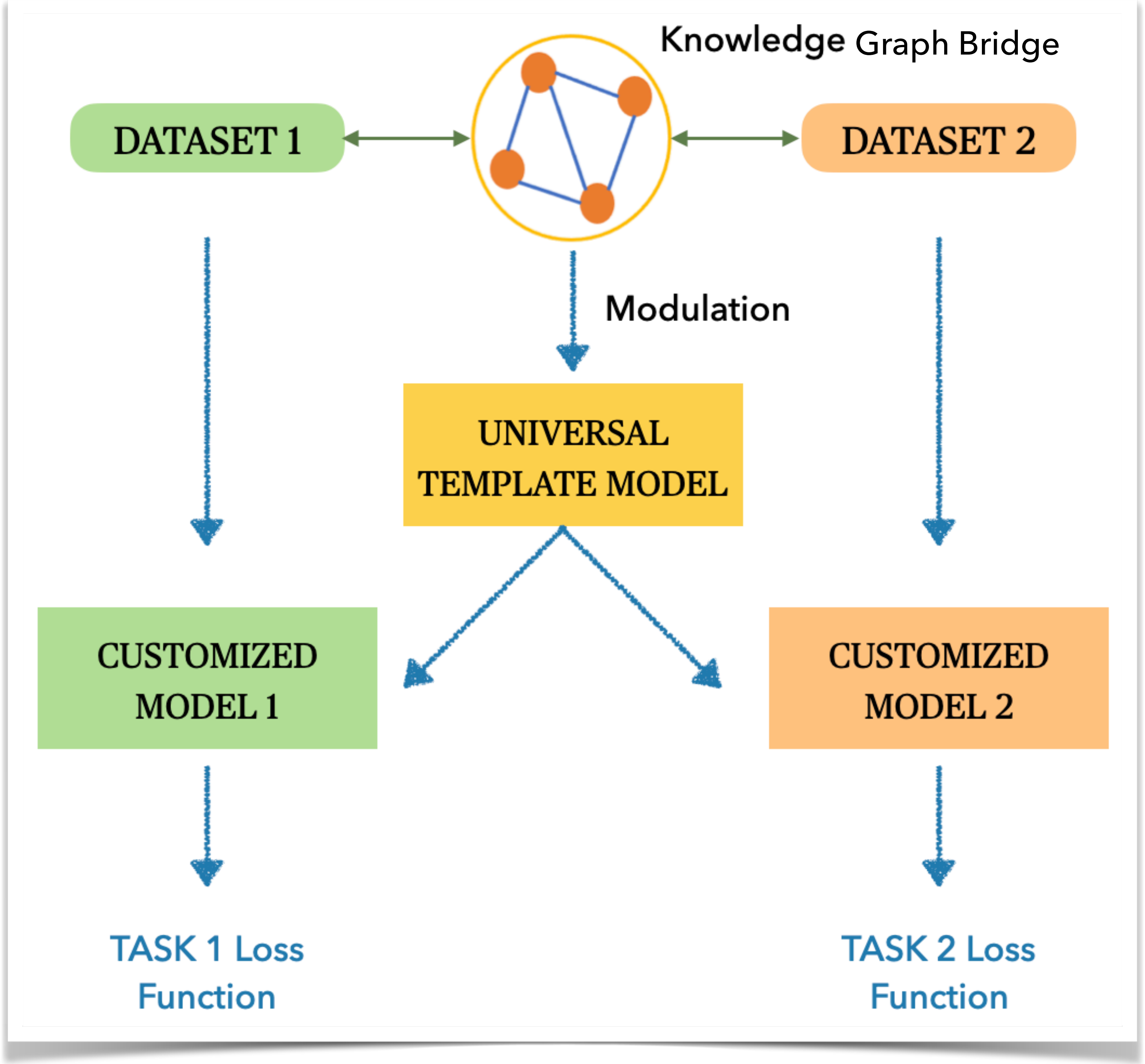
We create representations for entities and relationships in large-scale KGs in order to solve downstream ML tasks efficiently (e.g., gene interaction prediction in SPOKE).

Challenges: (i) long-tailed distribution; (ii) consistency across multiple views for the same nodes; (iii) loops.

Key Findings: Contrastive training and uncertainty modeling in the embedding space lead to more robust embeddings, and are resilient to noisy edges.

Outcomes: (i) A robust KGE algorithm for large-scale KGs, (ii) pre-trained embeddings for the SPOKE KG.

Thrust 1: Providing Inductive Biases via Knowledge Graphs



Knowledge Graph Bridges enable systematic transfer by modulating an universal template based on the dataset characteristics

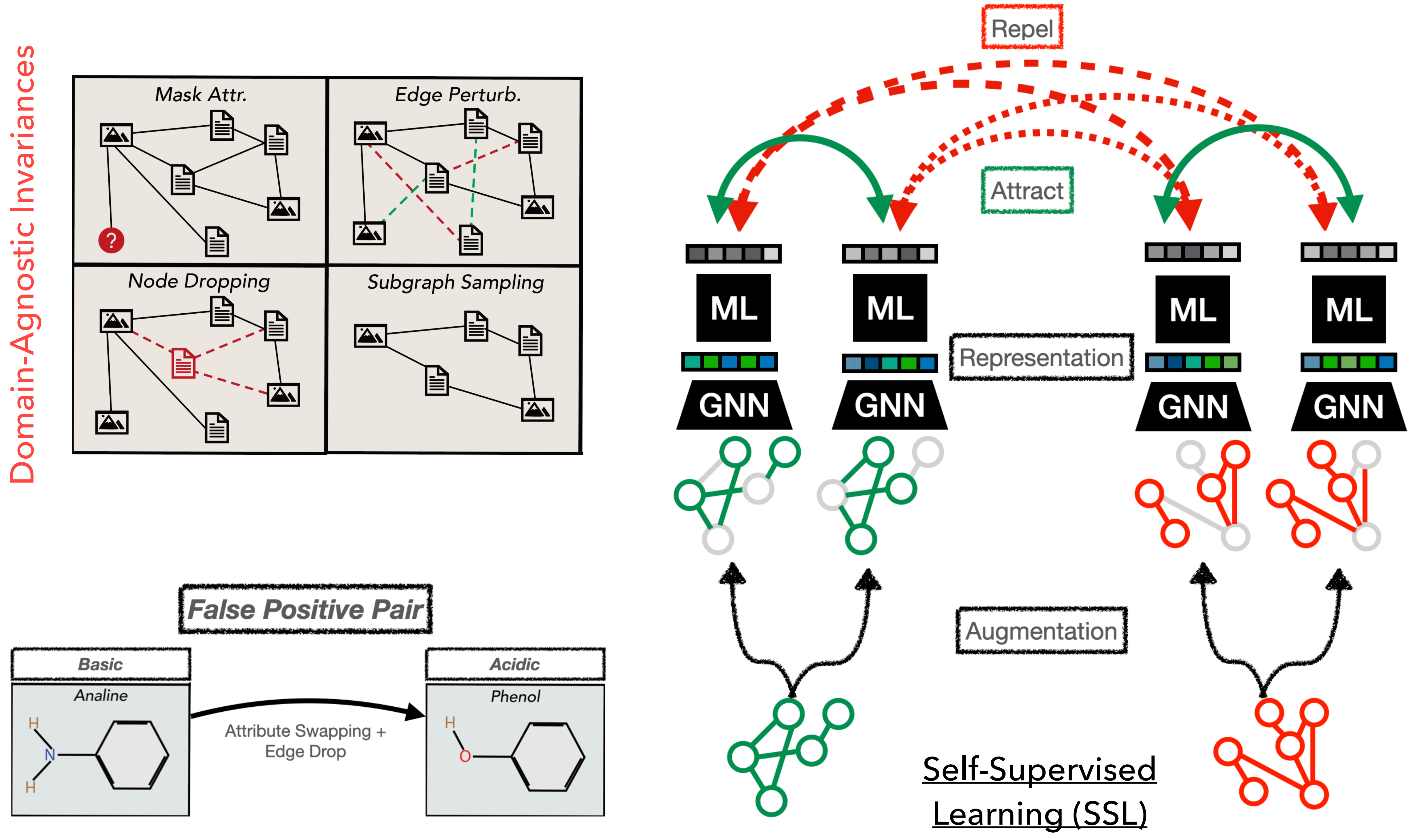
We develop a new transfer learning paradigm that relies on knowledge graph bridges

Challenges: (i) limited data for transfer; (ii) selectively transfer the most relevant information; (iii) automatic construction of KG bridges.

Key Findings: Jointly inferring the KG bridge and the universal template leads to advanced transfer learning capabilities in few-shot settings. We achieve 10-20% improvement in performance over SoTA.

Outcomes: (i) A structured meta-learning algorithm for transfer learning, (ii) demonstration on clinical modeling tasks, (iii) benchmarking few-shot learning for vision.

Thrust 2: Driving Representation Learning with Domain-Specific Invariances



Known invariances provide a powerful source of constraints

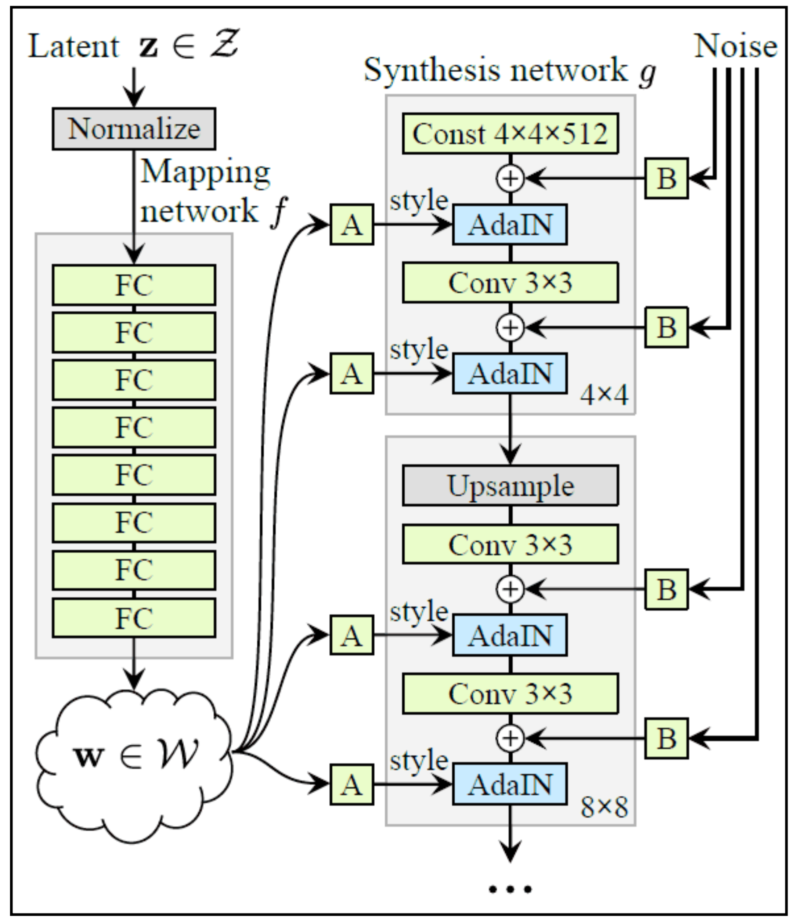
We develop a theoretical framework to understand the limitations of domain-agnostic invariances

Challenges: (i) no theoretical framework exists for establishing benefit of known invariances in unsupervised learning; (ii) benchmark datasets are insufficient.

Key Findings: We identify regimes where domain-agnostic invariances are harmful and known invariances are required!

Outcomes: (i) A first-of-its-kind spectral analysis framework, (ii) graph representation learning algorithms, (iii) a new benchmark dataset, (iv) demonstration in drug design.

Thrust 3: Generative Models as Knowledge Sources



ID Data: Face Images

Existing Ours

OOD Data: Cartoons

Input Frown Young Age

Face attributes applied to cartoons

We view pre-trained generative models trained on a data source (e.g., simulations) as a knowledge base and mine relevant information to solve tasks with out-of-distribution data (e.g., experiments).

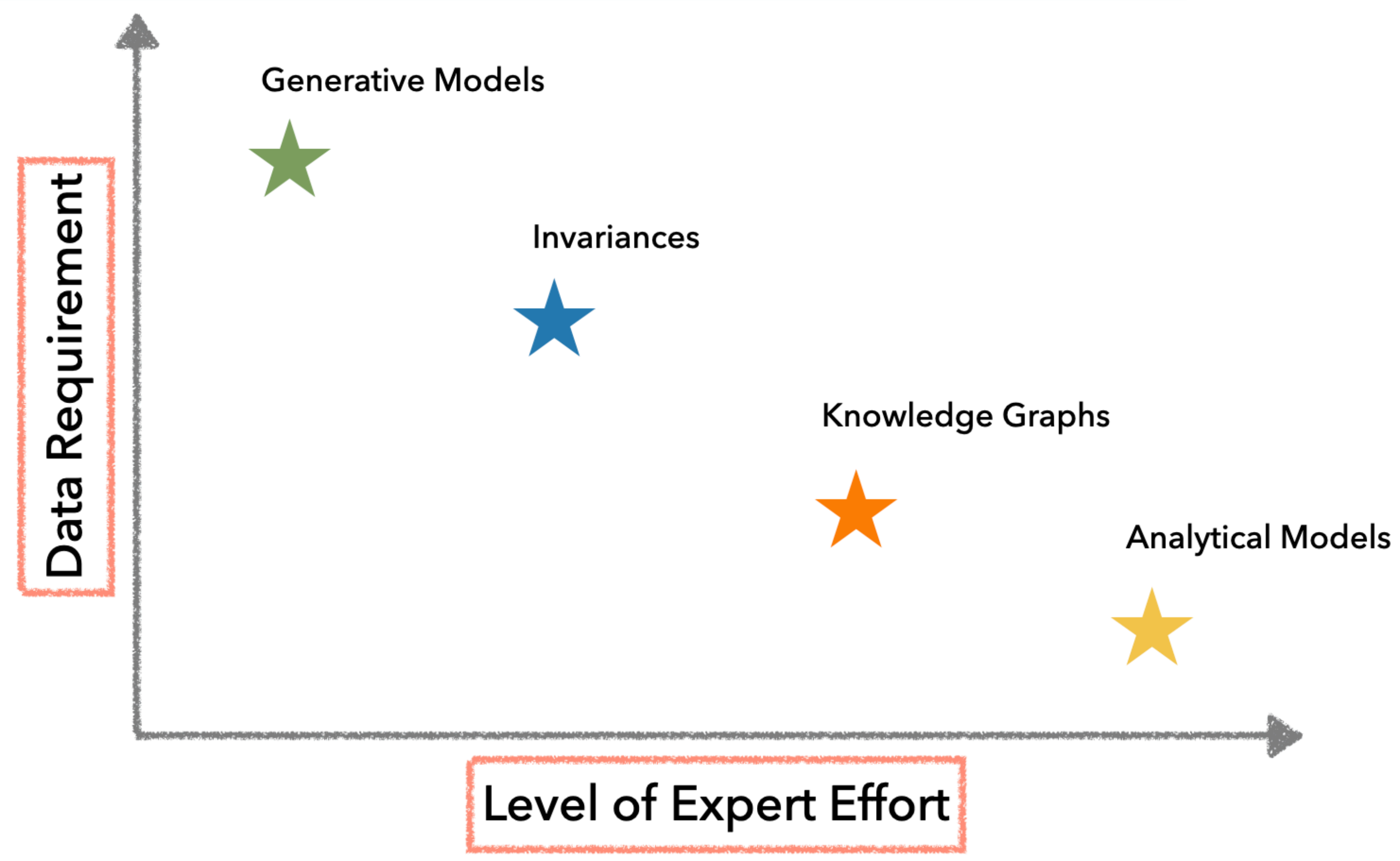
<u>Challenges</u>: (i) embedding OOD data into latent spaces of generative models is currently infeasible; (ii) transferring knowledge to OOD.

<u>Key Findings</u>: Using a novel optimization technique, our initial results with natural images show that we can effectively embed OOD data into StyleGAN latent spaces and transfer knowledge.

<u>Outcomes</u>: (i) A generic approach for leveraging pre-trained generative models for OOD data, (ii) demonstration on simulation –> experiment transfer.

Impact

- **New capabilities for domain-aware learning**:
  - KG-guided ML for clinical modeling.
  - KG bridges for data-efficient transfer in sciences.
  - A new theoretical framework for studying effect of domain-specific invariances.
  - Unsupervised learning algorithms for modeling small molecule graphs.
  - Generative models as a knowledge source.



- Collaboration with UCSF team (SPOKE KGs) as well as university partners for future engagements.
- Strongly aligned with DOE, DNN and NIH initiatives.