tnGPS

Discovering Unknown Tensor Network Structure Search Algorithms via Large Language Models (LLMs)

Presenter: *Junhua Zeng {jh.zenggdut@gmail.com}*

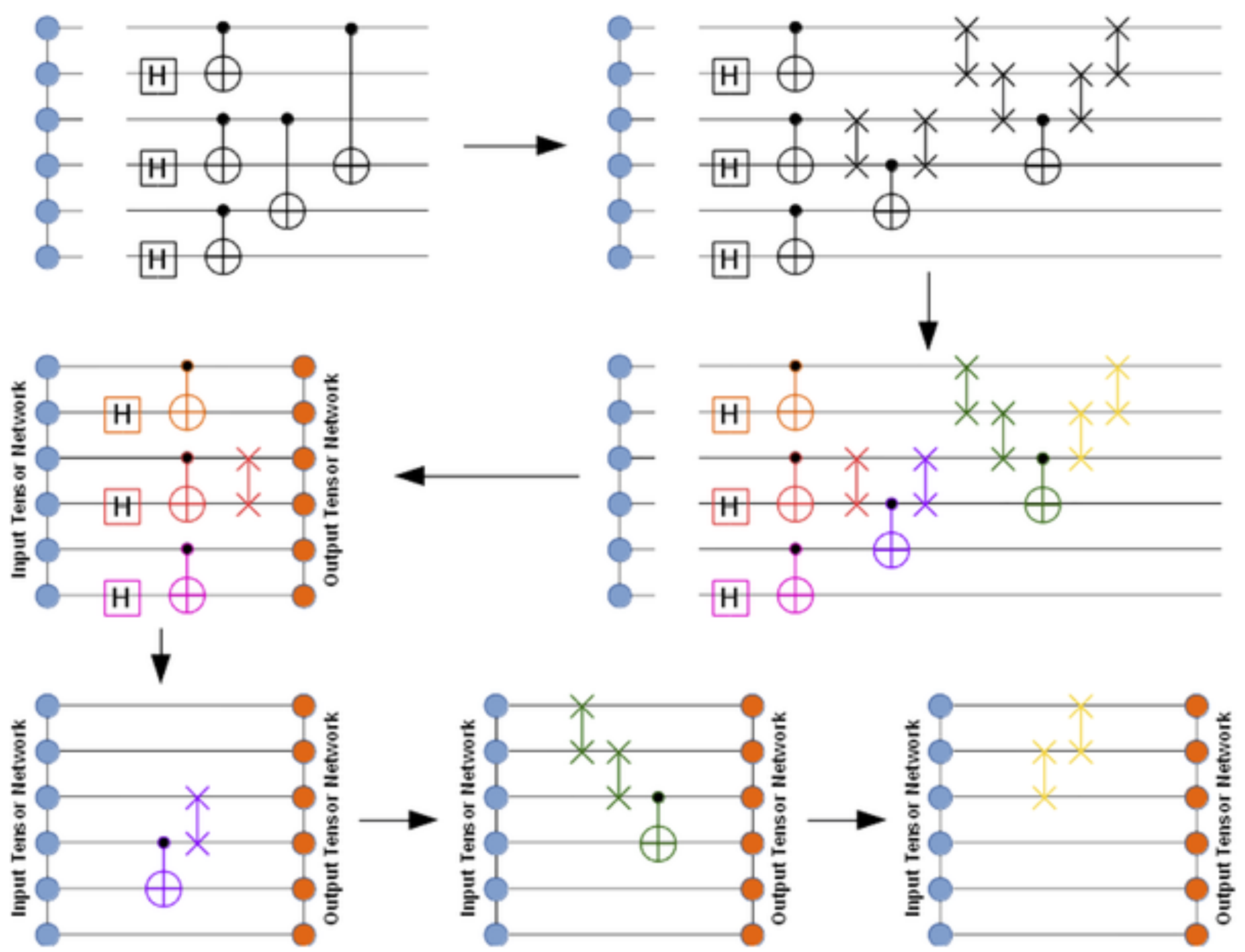
***Junhua Zeng**[1*]*, ***Chao Li**[2*]*, ***Zhun Sun**[3]*, ***Qibin Zhao**[2]* and ***Guoxu Zhou**[1,4]*

[1]*Guangdong University of Technology, China*
[2]*RIKEN-AIP, Japan*
[3]*Tencent Inc., China*
[4]*Key Laboratory of Intelligent Detection and the Internet of Things in Manufacturing, China*

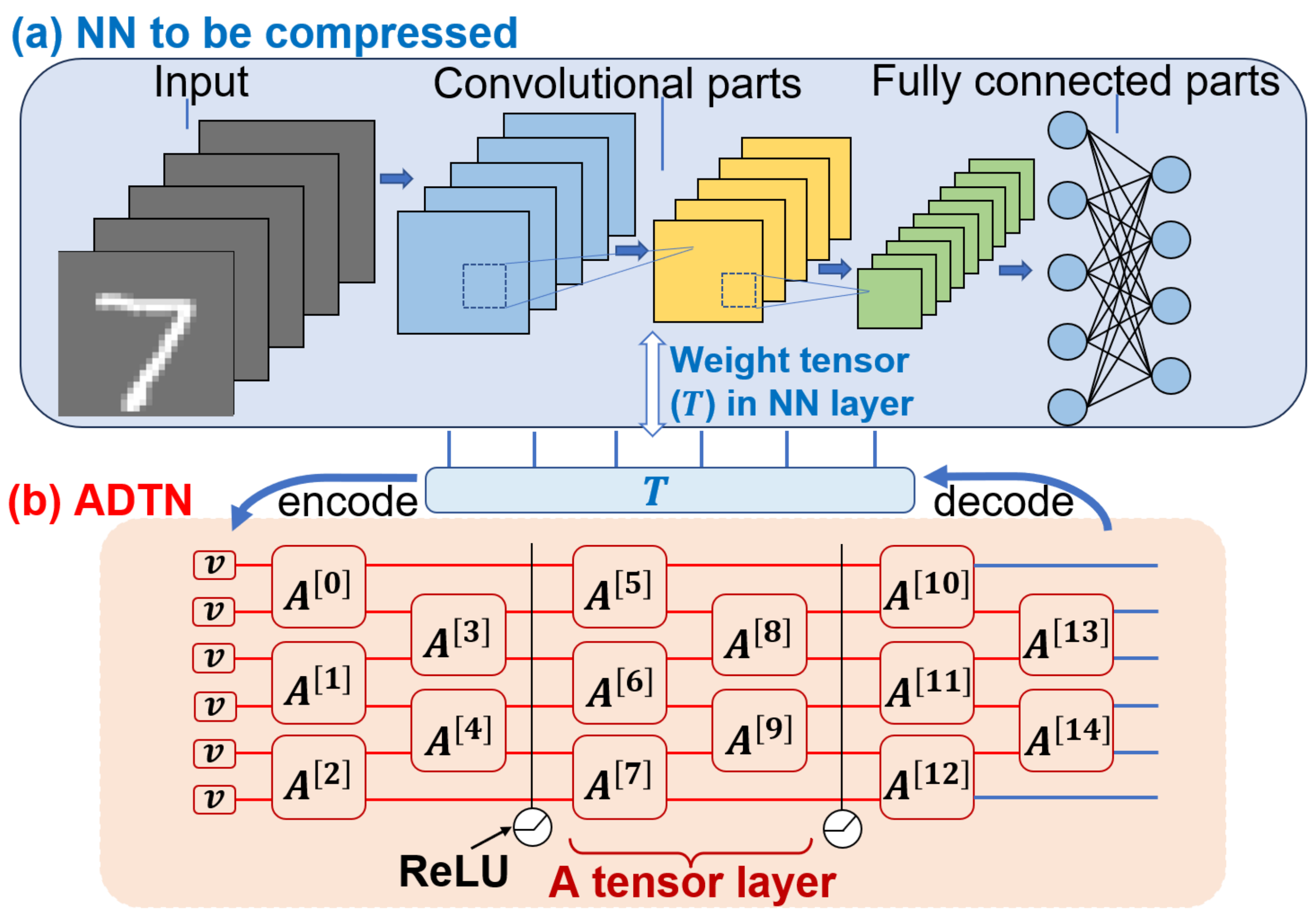
Background

Applications of Tensor Networks (TNs)

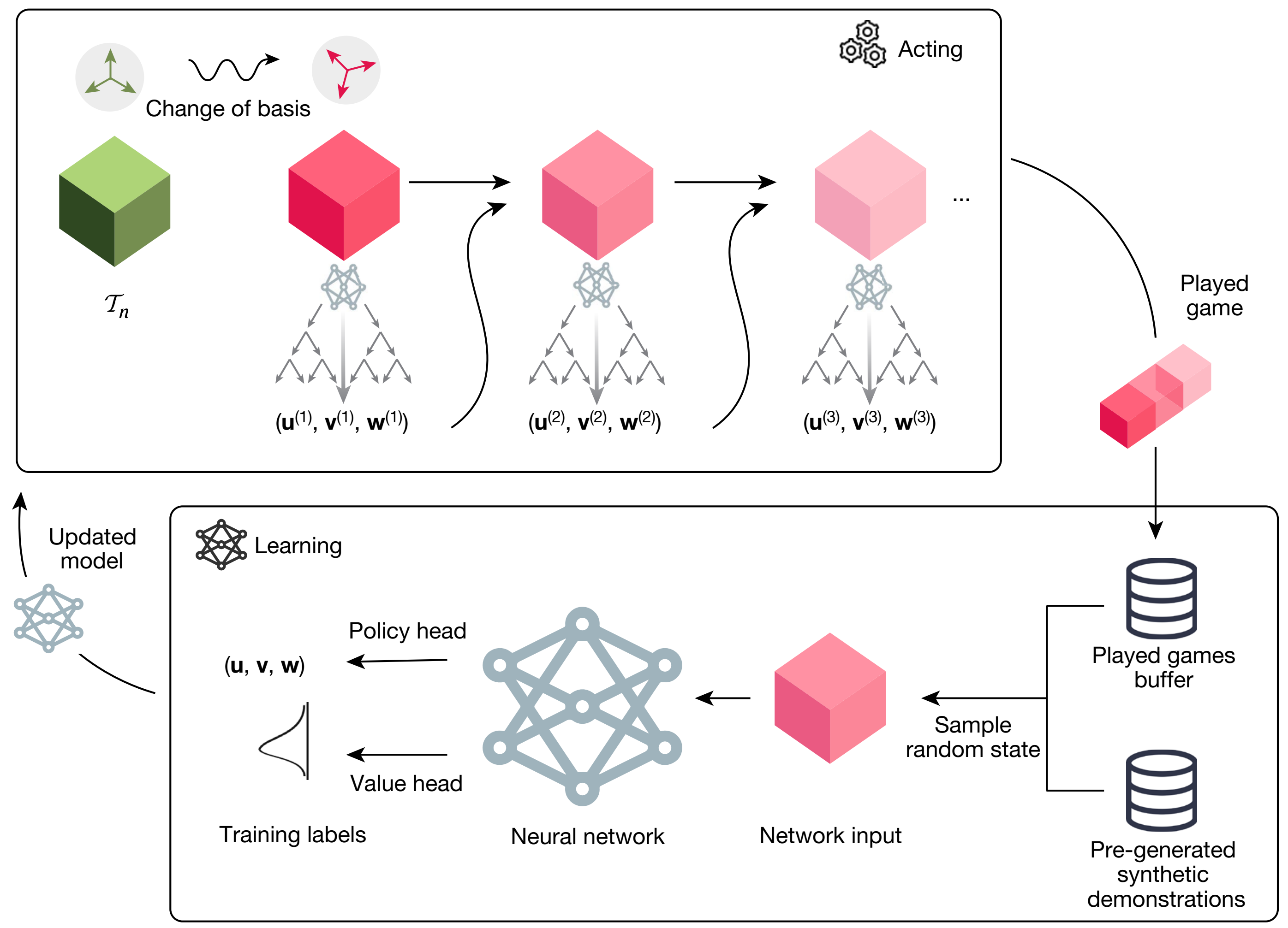
TN is an efficient framework for modeling complex systems by decomposing it into simpler, interconnected parts.



State representation in quantum physics



Model representation in machine learning



Discovering faster ***matrix multiplication***
(AlphaTensor, Fawzi et al., Nature'22)

Background

Why TN so Powerful?

Background

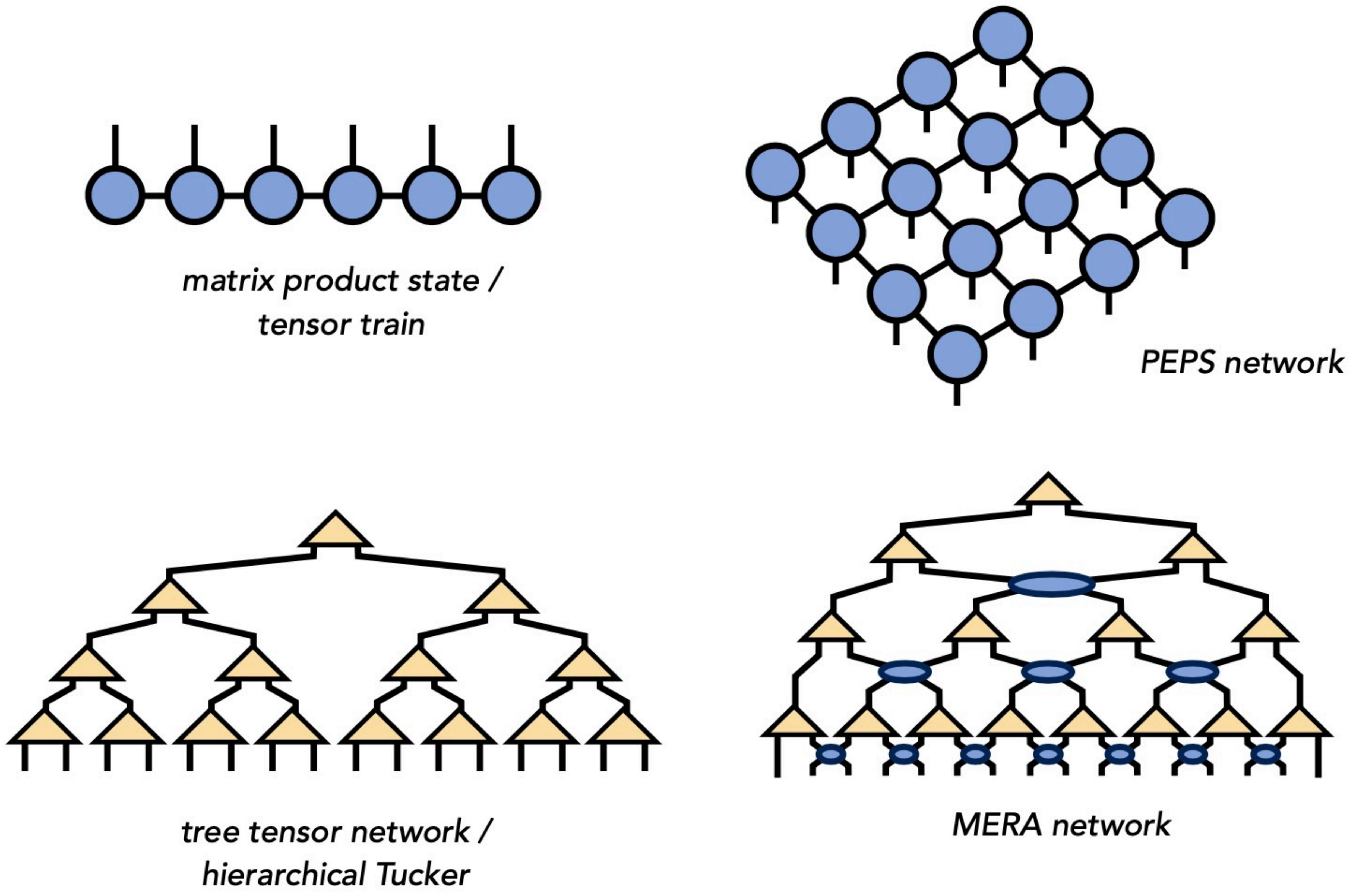
Why TN so Powerful?

Tensor networks can efficiently represent high-dimensional space.

Background

Why TN so Powerful?

Tensor networks can efficiently represent high-dimensional space.



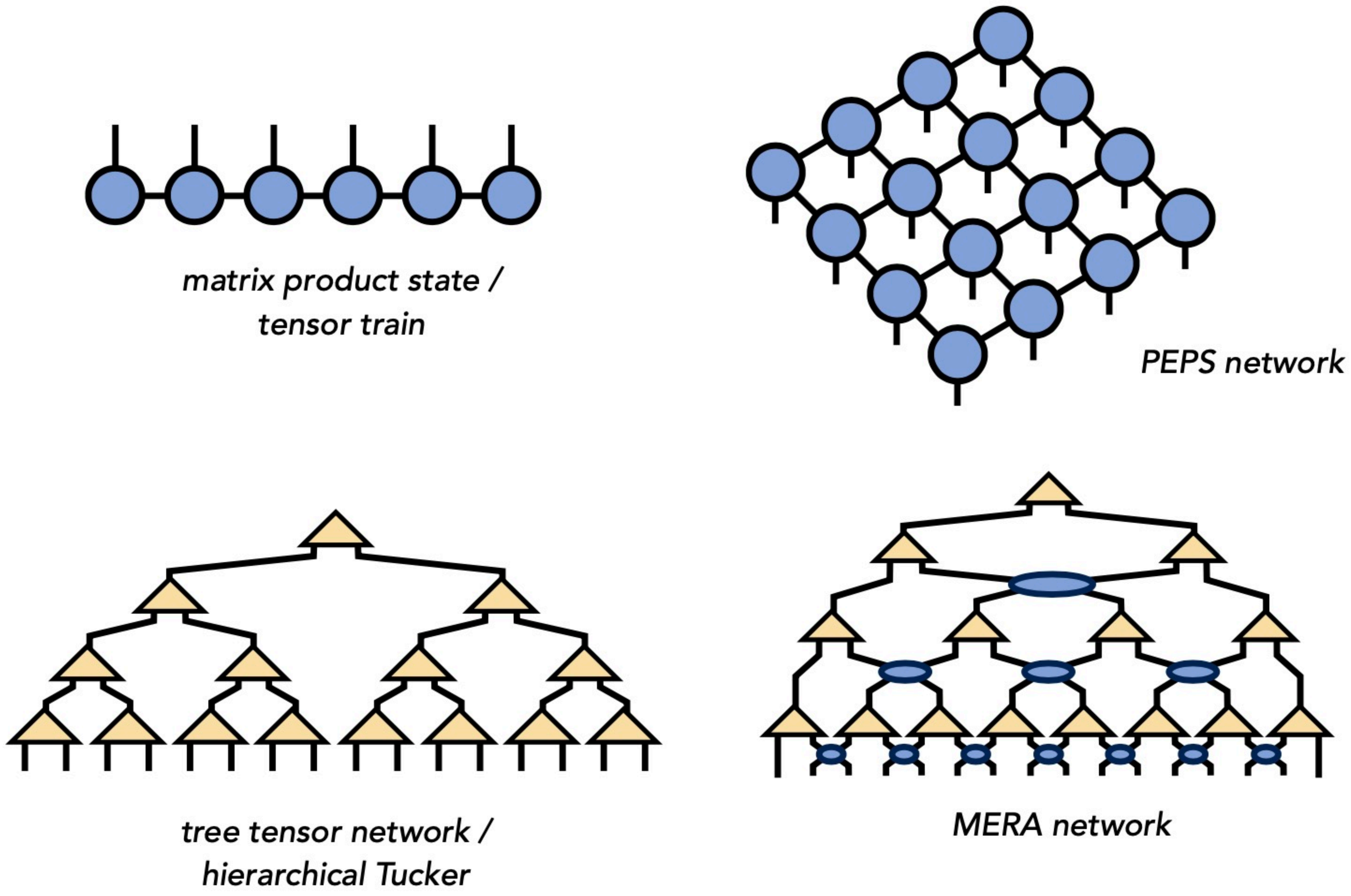
R. Orus, Ann. of Phys. 349, 117–158 (2014)

Background

Why TN so Powerful?

Tensor networks can efficiently represent high-dimensional space.

How to select the promising tensor network models?



R. Orus, Ann. of Phys. 349, 117–158 (2014)

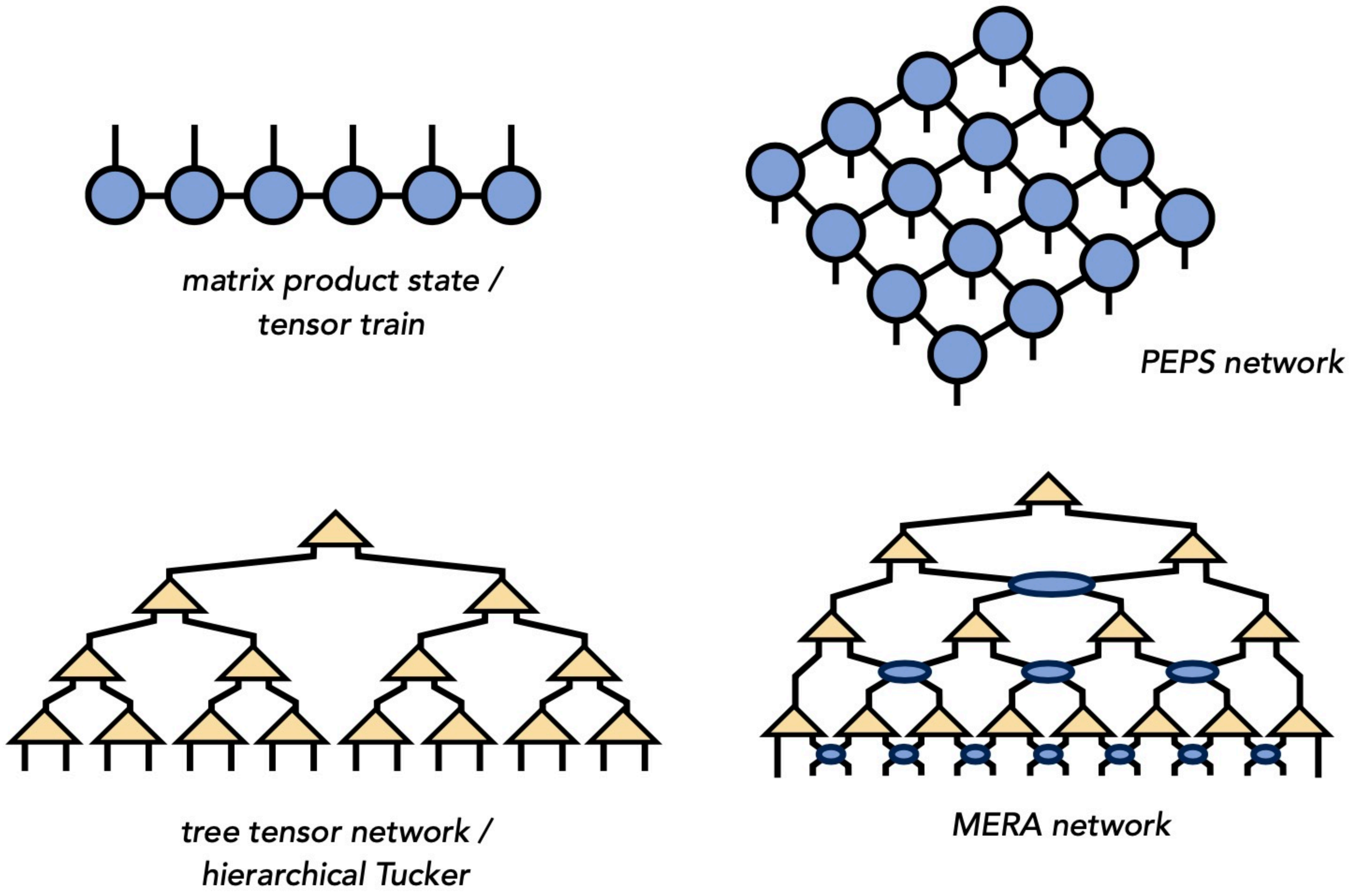
Background

Why TN so Powerful?

Tensor networks can efficiently represent high-dimensional space.

How to select the promising tensor network models?

Tensor Network Structure Search (TN-SS)



R. Orus, Ann. of Phys. 349, 117–158 (2014)

Background

Motivations

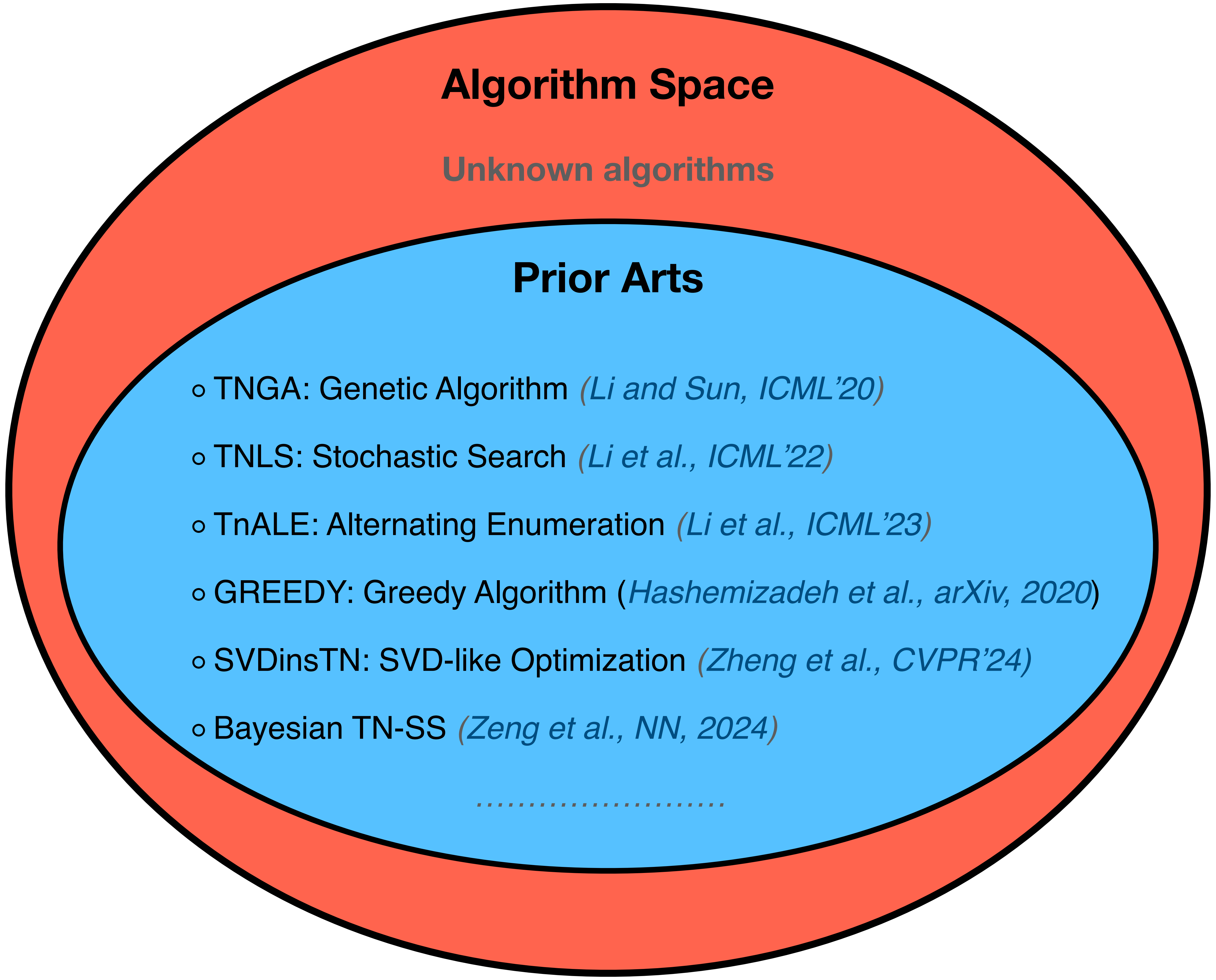
Prior Arts

- TNGA: Genetic Algorithm *(Li and Sun, ICML'20)*
- TNLS: Stochastic Search *(Li et al., ICML'22)*
- TnALE: Alternating Enumeration *(Li et al., ICML'23)*
- GREEDY: Greedy Algorithm (*Hashemizadeh et al., arXiv, 2020*)
- SVDinsTN: SVD-like Optimization *(Zheng et al., CVPR'24)*
- Bayesian TN-SS *(Zeng et al., NN, 2024)*

....

Background

Motivations



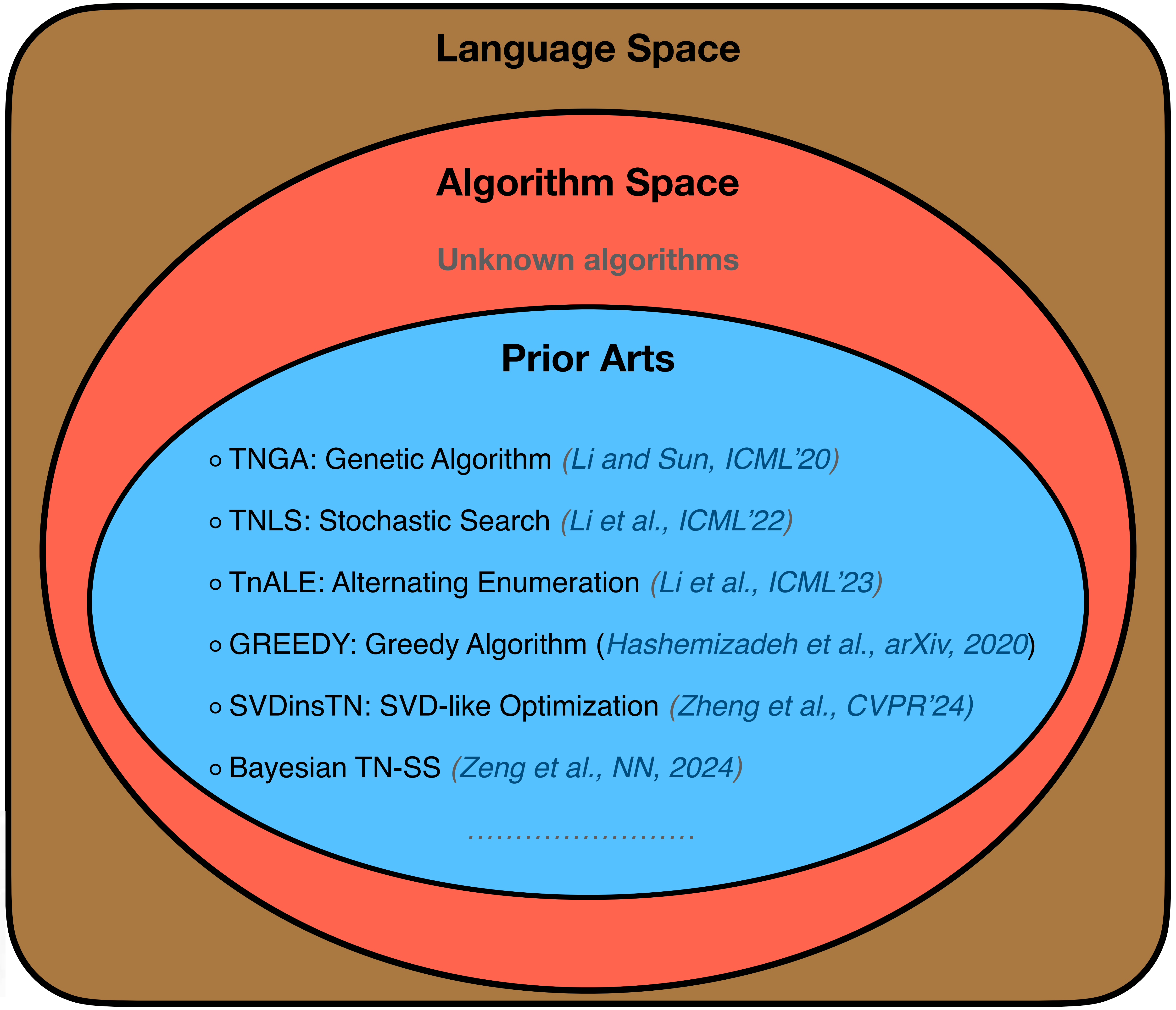
Background

Motivations

ChatGPT

Claude

Gemini



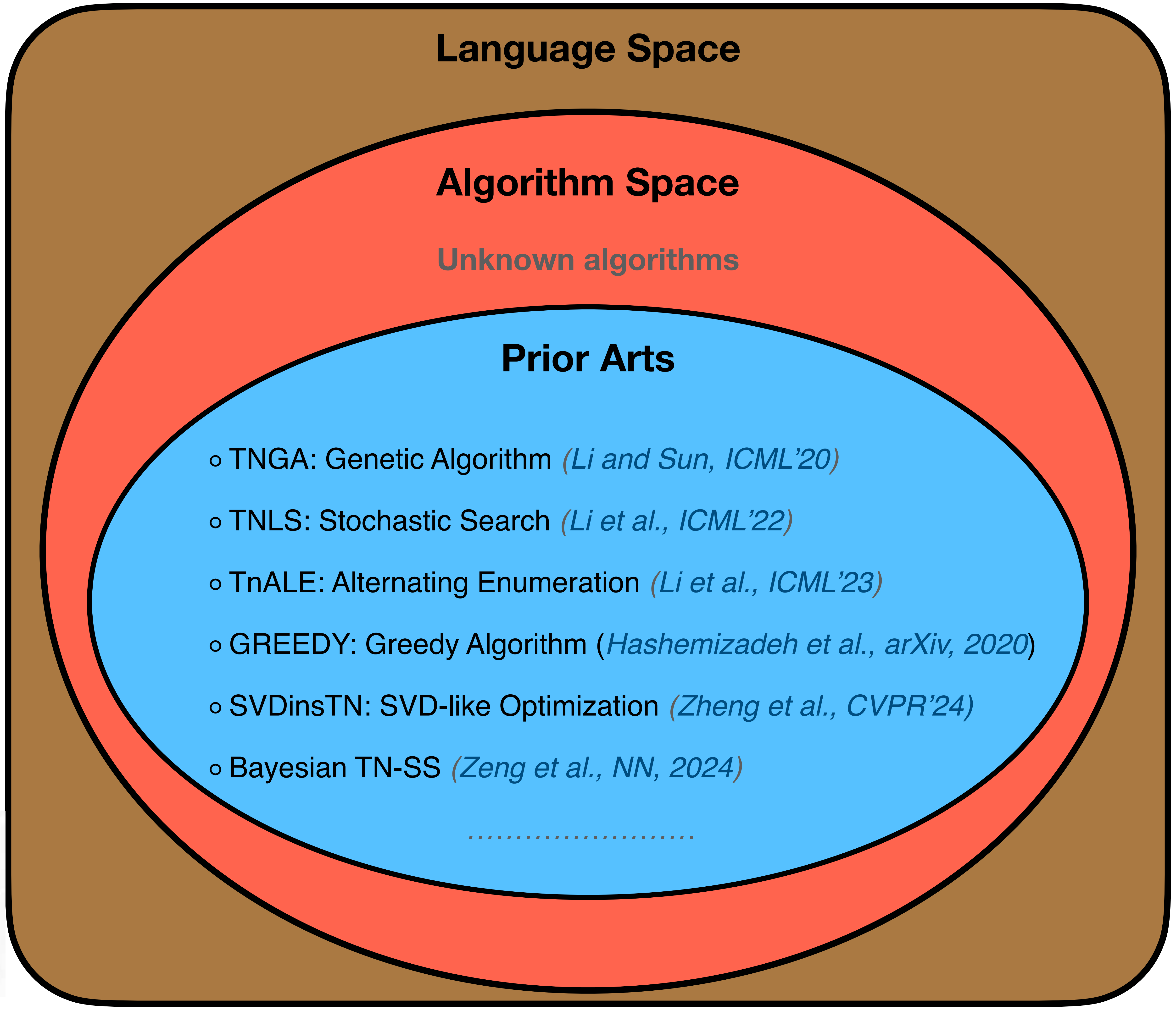
Background

Motivations

ChatGPT

Claude

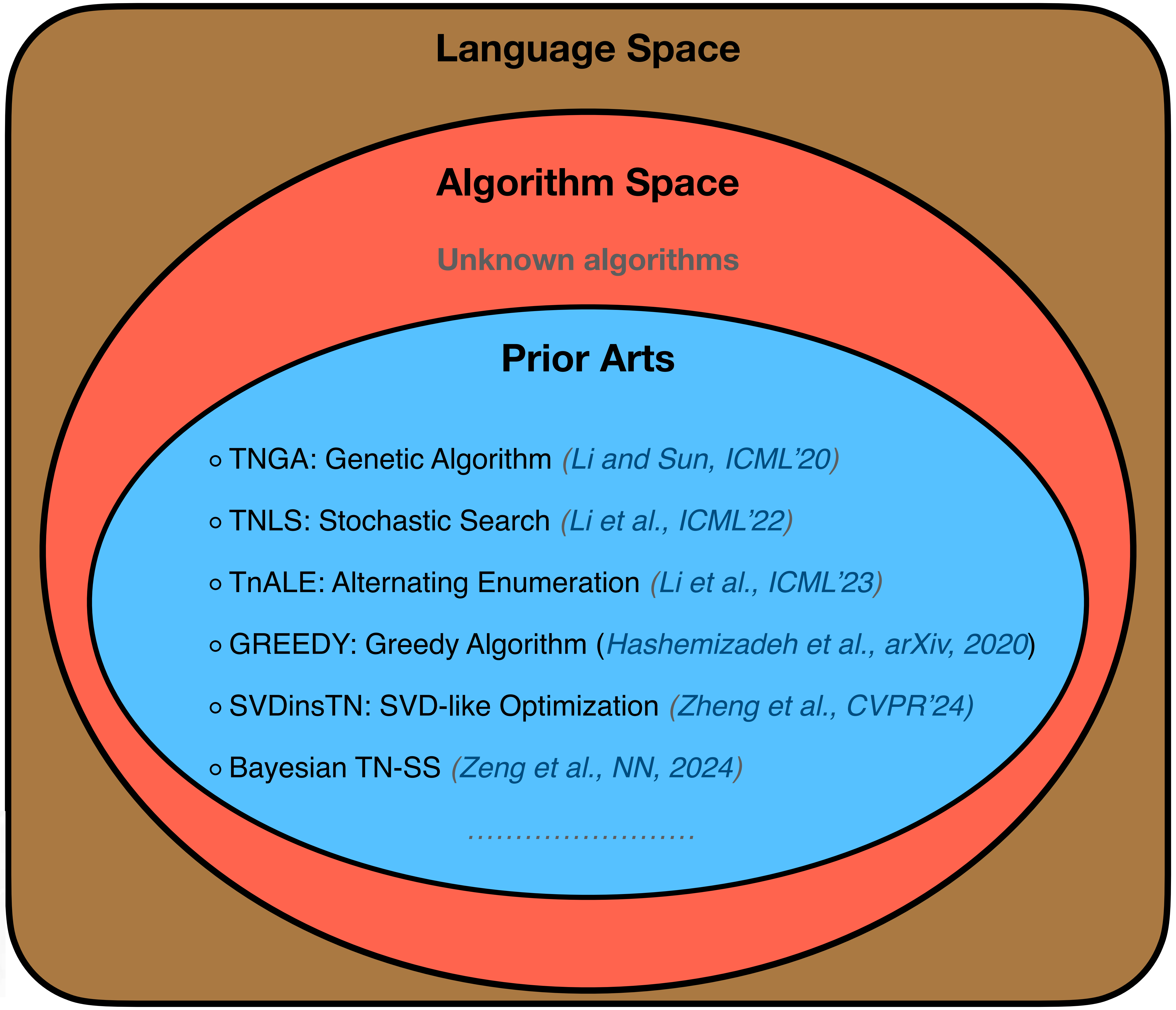
Gemini



Exploiting the *enormous language space* of LLMs for *autonomous TN-SS algorithm discovery*.

Background

Motivations



Exploiting the *enormous language space* of LLMs for *autonomous TN-SS algorithm discovery*.

Saving human experts from the labor-intensive algorithm design process and letting them focus on more challenging problems.

Background

Contributions

Background

Contributions

1. We *propose* *tensor-network-purposed GPT-driven structure search (tnGPS)*, a *LLM-driven* automation framework designed to *automatically generate novel and effective TN-SS algorithms* tailored to specific downstream tasks;

Background

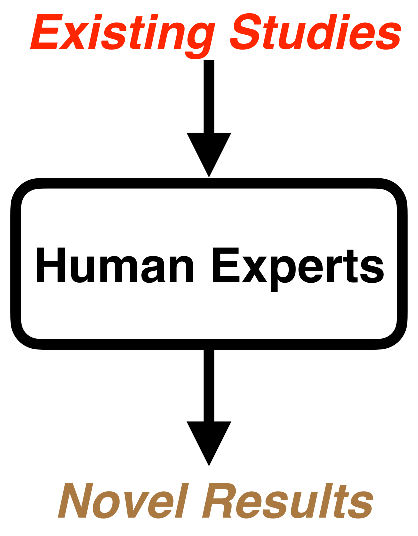
Contributions

1. We *propose* *tensor-network-purposed GPT-driven structure search (tnGPS)*, a *LLM-driven* automation framework designed to *automatically generate novel and effective TN-SS algorithms* tailored to specific downstream tasks;

2. Experimental results demonstrate that the algorithms discovered by tnGPS outperform existing TN-SS algorithms on benchmark data.

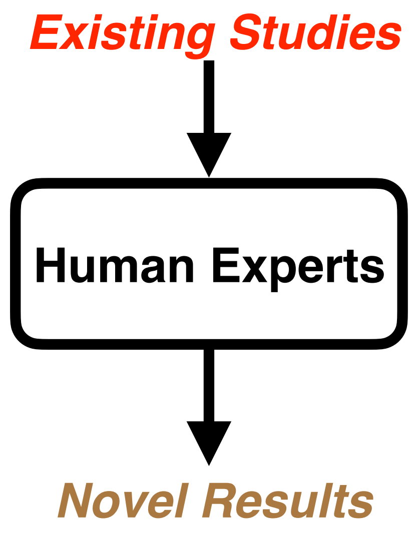
When TN-SS Meets LLMs

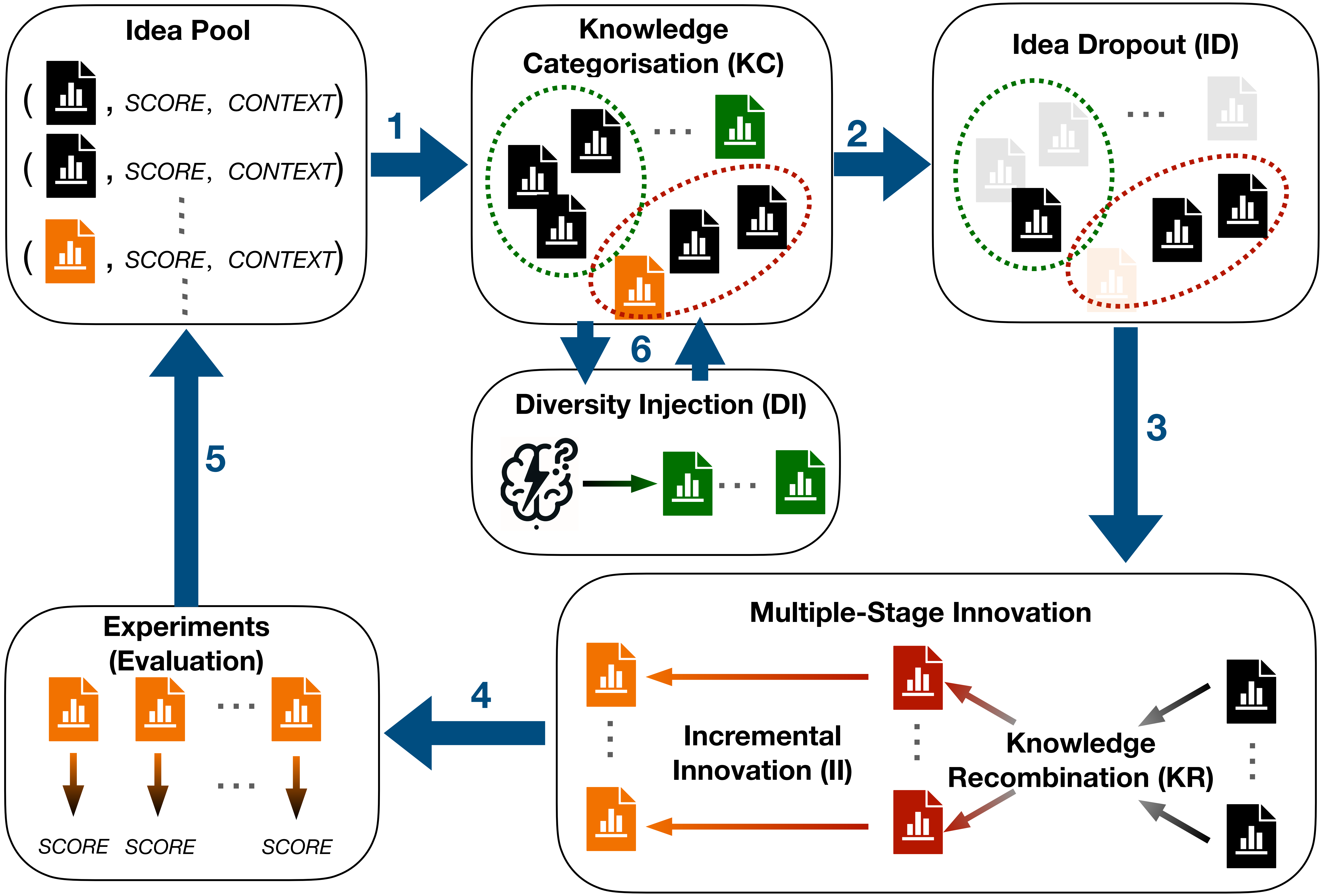
How Human Experts Conduct Innovative Research



When TN-SS Meets LLMs

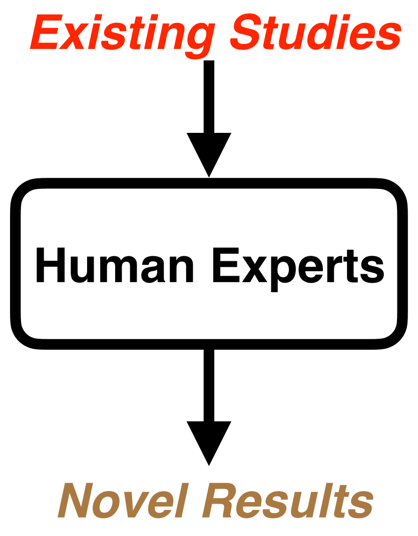
How Human Experts Conduct Innovative Research



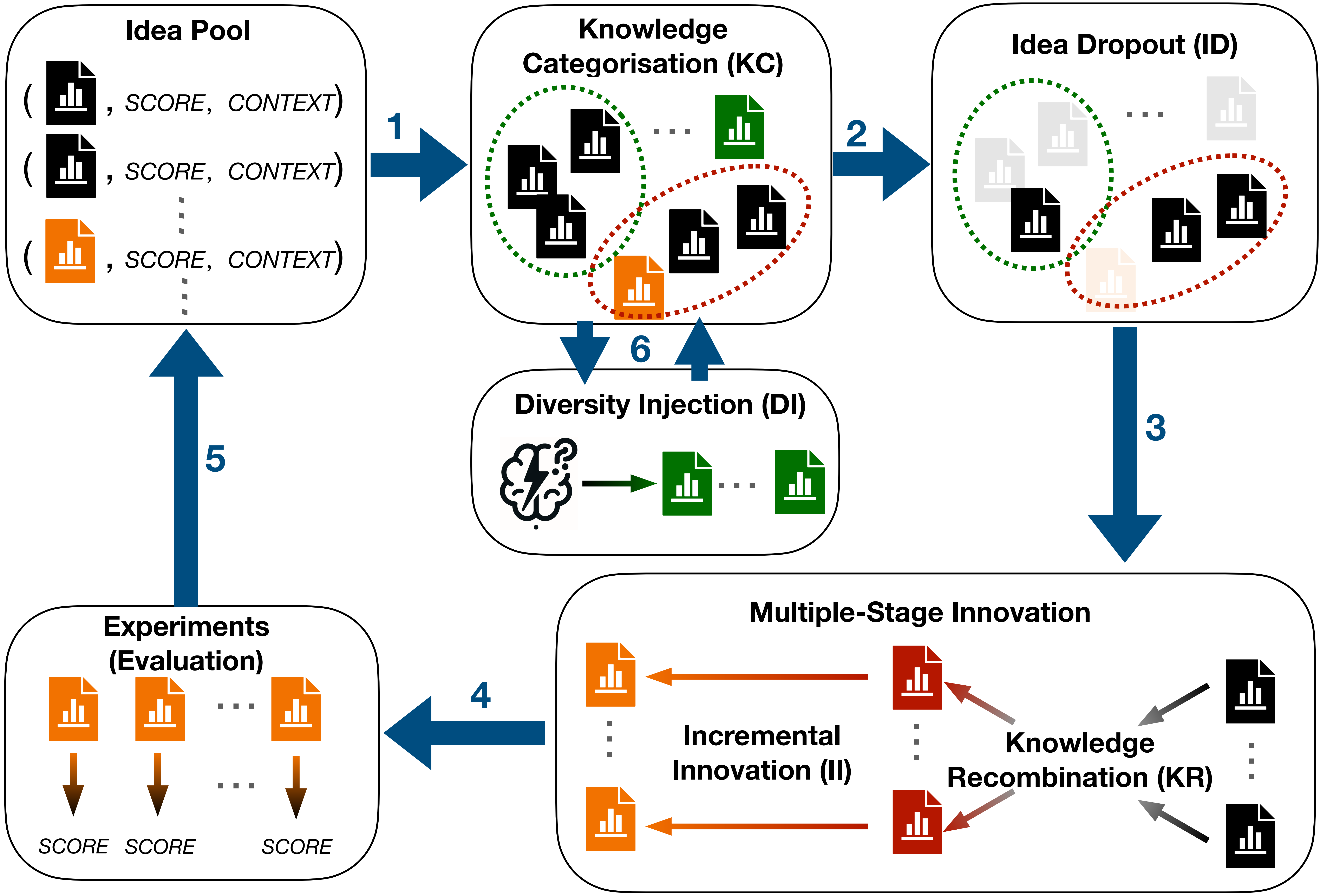


When TN-SS Meets LLMs

How Human Experts Conduct Innovative Research

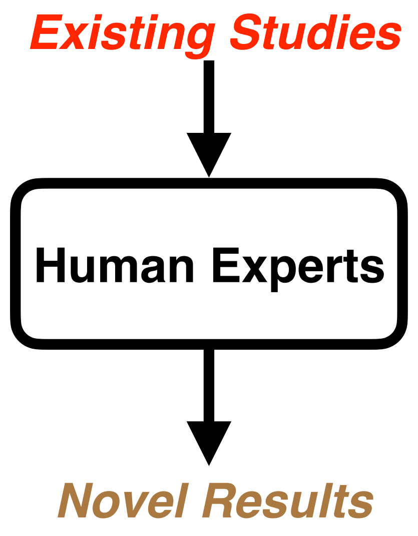


Idea Pool: Gather information through literature reviews and paper retrieval



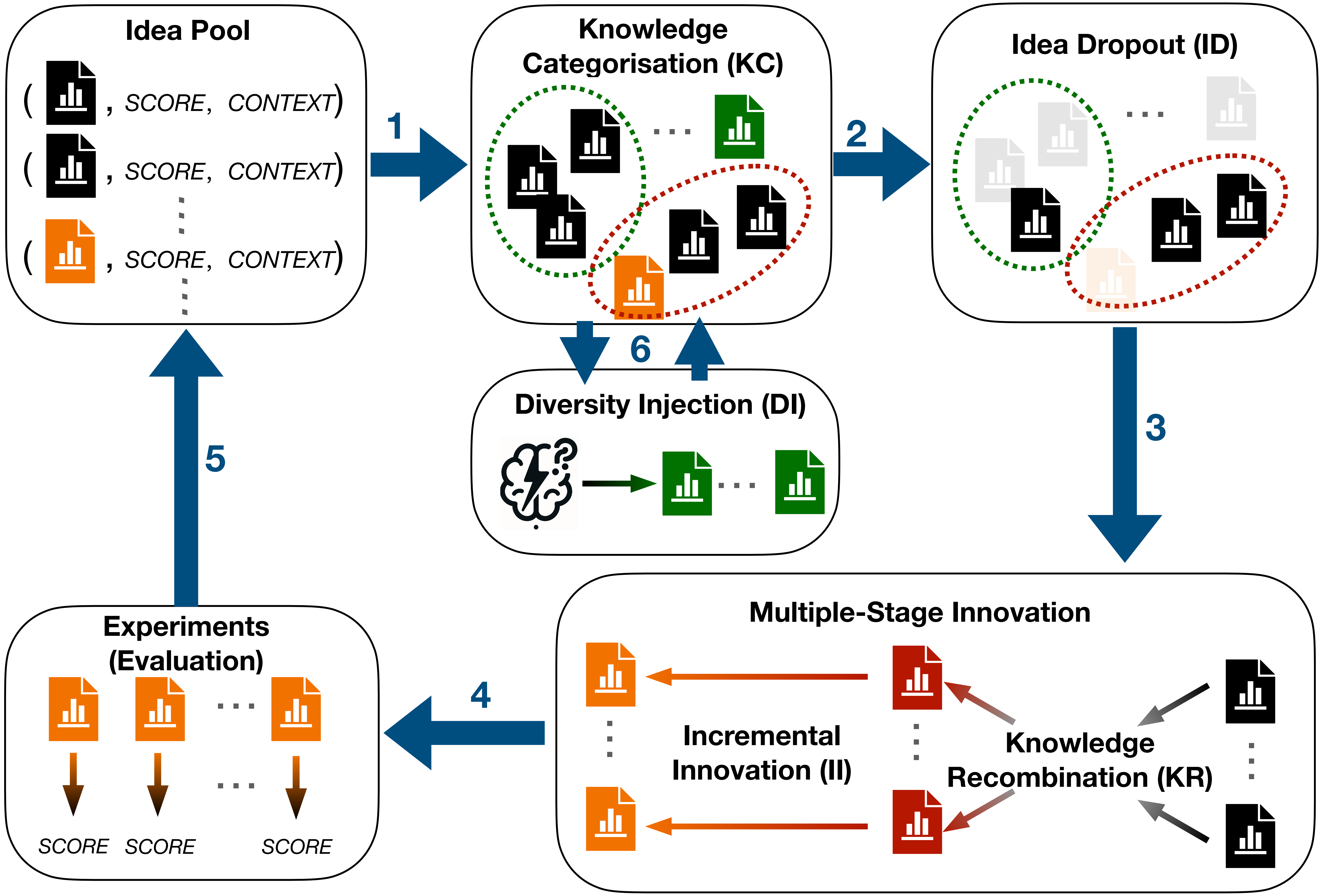
When TN-SS Meets LLMs

How Human Experts Conduct Innovative Research



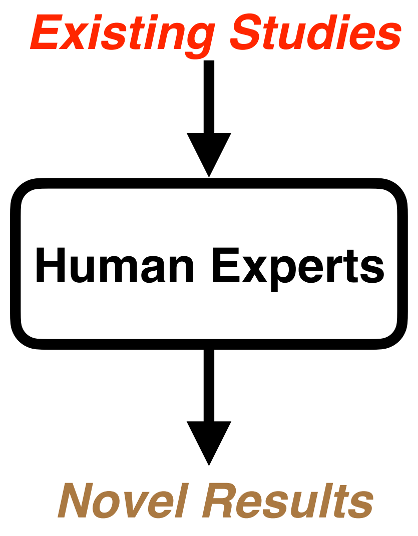
Idea Pool: Gather information through literature reviews and paper retrieval

Knowledge Categorization (KC): Refine ideas into knowledge clusters.



When TN-SS Meets LLMs

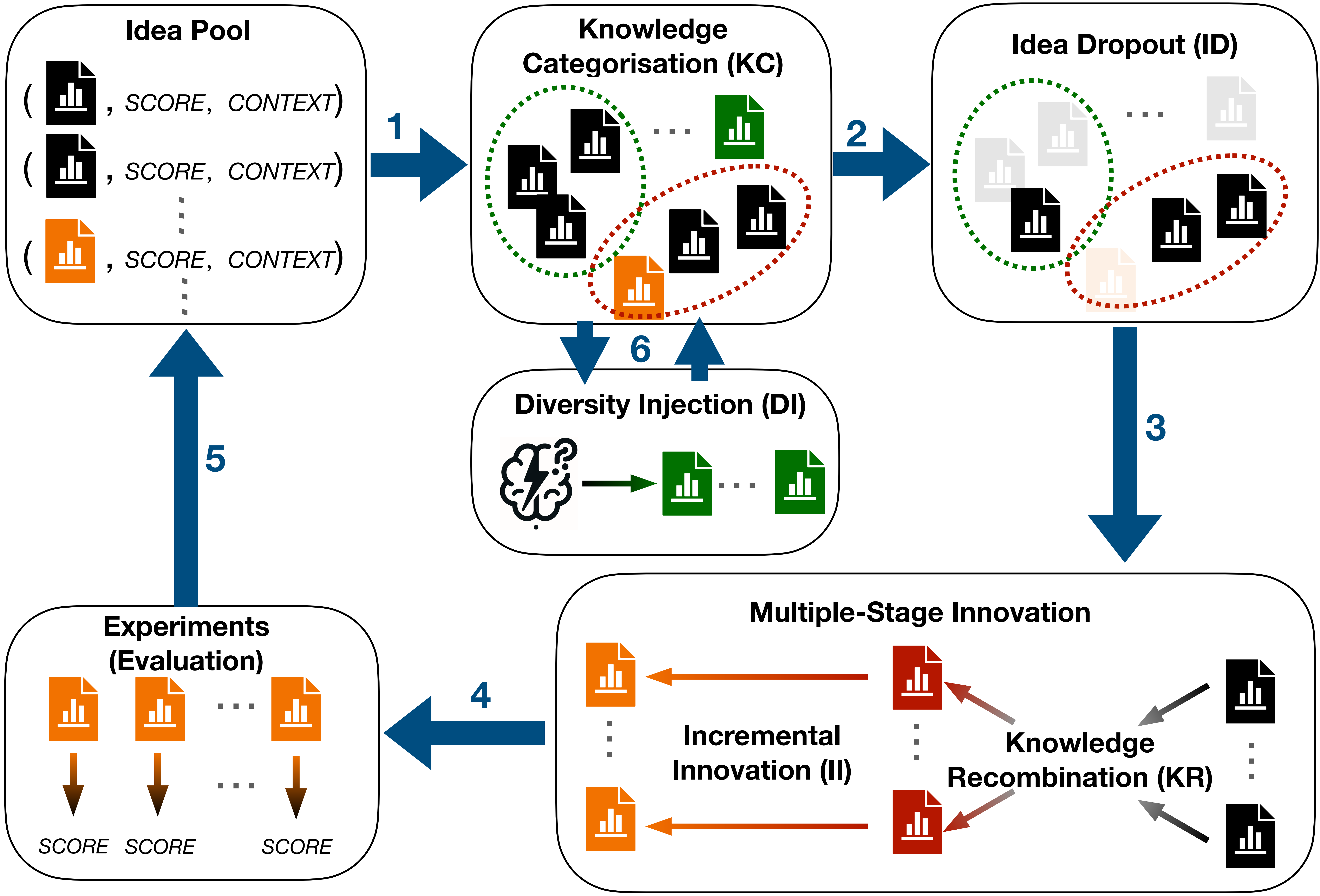
How Human Experts Conduct Innovative Research



Idea Pool: Gather information through literature reviews and paper retrieval

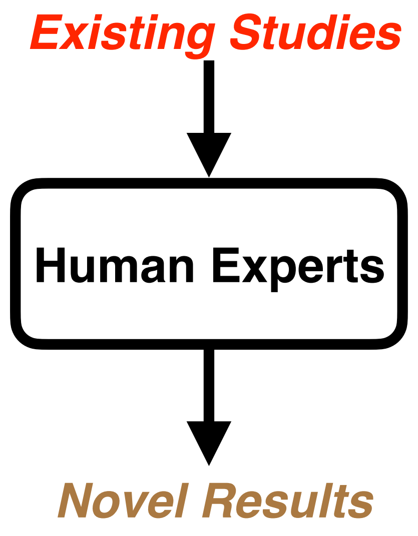
Knowledge Categorization (KC): Refine ideas into knowledge clusters.

Idea Dropout (ID): Filter out less interesting ideas to focus on the most promising ones.



When TN-SS Meets LLMs

How Human Experts Conduct Innovative Research

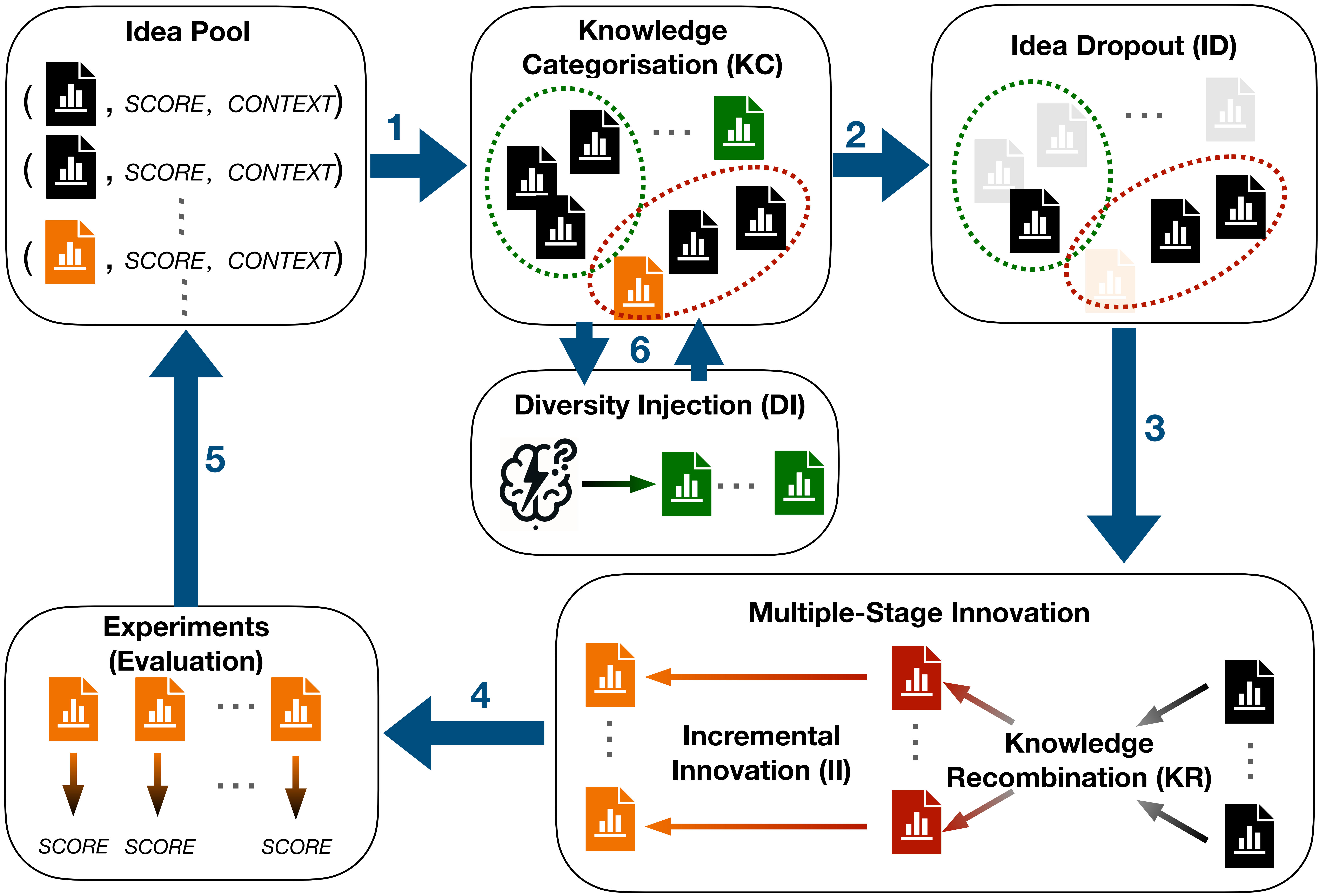


Idea Pool: Gather information through literature reviews and paper retrieval

Knowledge Categorization (KC): Refine ideas into knowledge clusters.

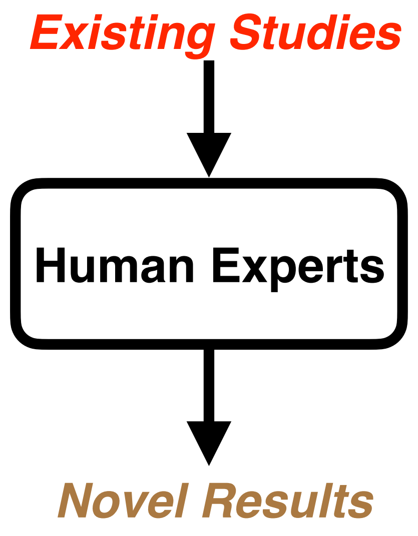
Idea Dropout (ID): Filter out less interesting ideas to focus on the most promising ones.

Knowledge Recombination (KR): Generate new ideas by merging existing ones.



When TN-SS Meets LLMs

How Human Experts Conduct Innovative Research



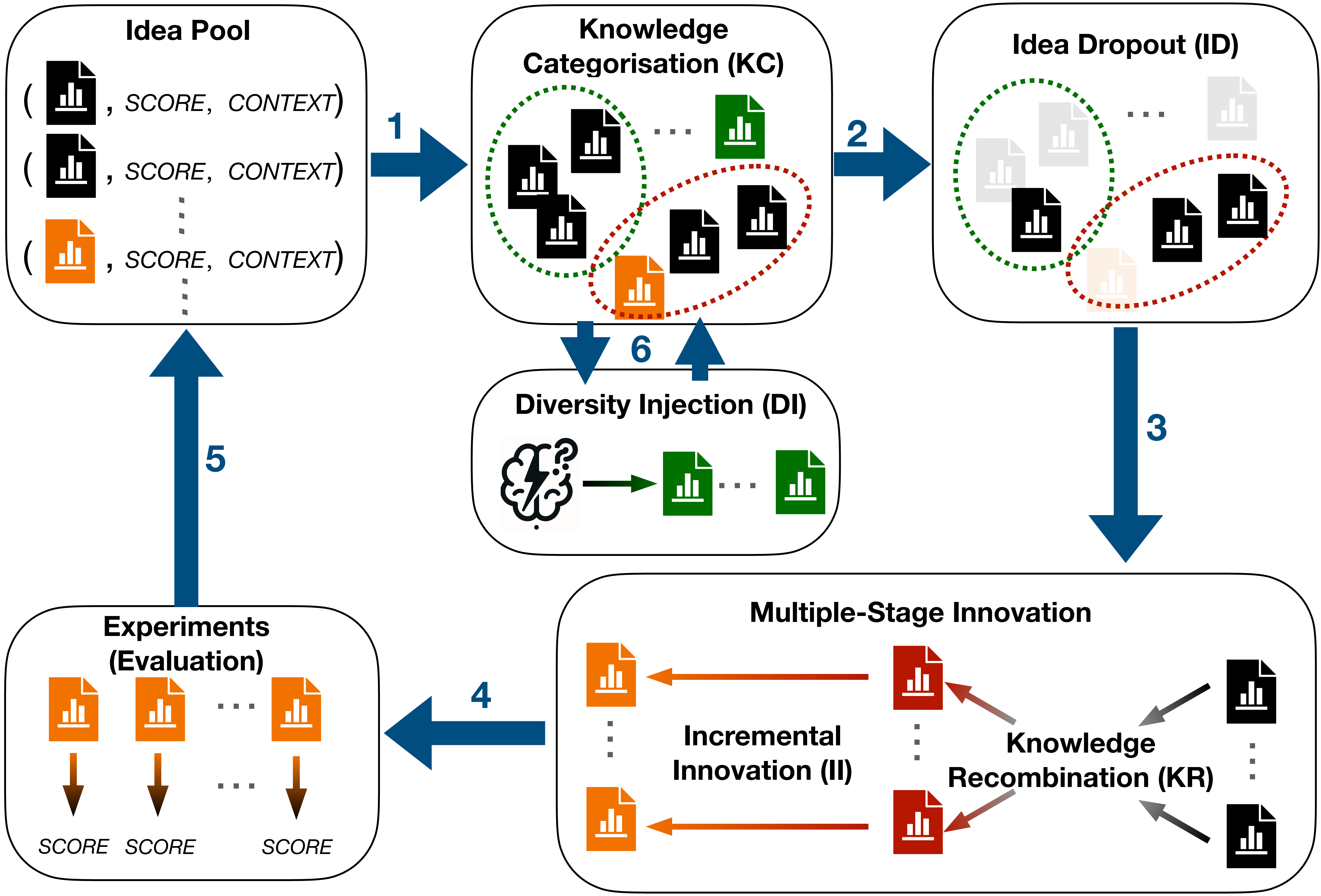
Idea Pool: Gather information through literature reviews and paper retrieval

Knowledge Categorization (KC): Refine ideas into knowledge clusters.

Idea Dropout (ID): Filter out less interesting ideas to focus on the most promising ones.

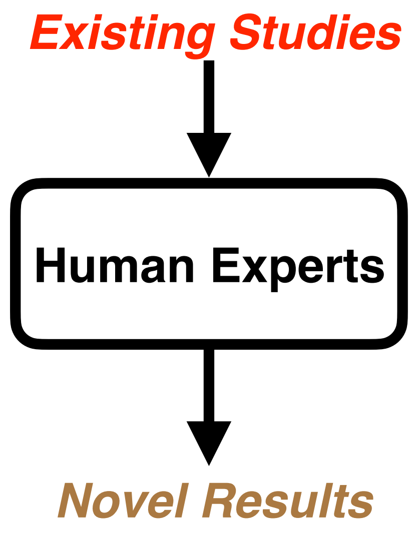
Knowledge Recombination (KR): Generate new ideas by merging existing ones.

Incremental Innovation (II): Make gradual improvements to existing ideas.



When TN-SS Meets LLMs

How Human Experts Conduct Innovative Research



Idea Pool: Gather information through literature reviews and paper retrieval

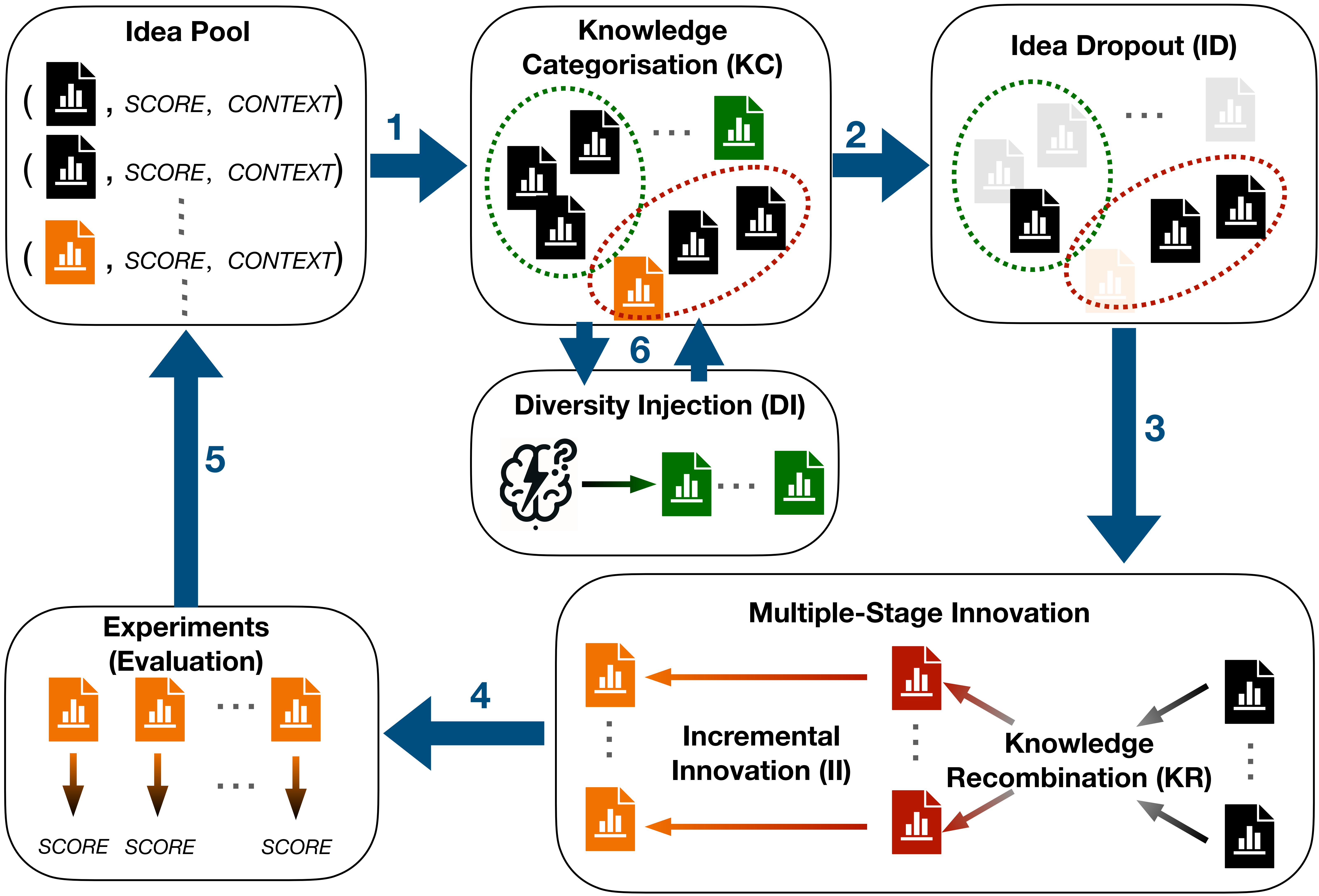
Knowledge Categorization (KC): Refine ideas into knowledge clusters.

Idea Dropout (ID): Filter out less interesting ideas to focus on the most promising ones.

Knowledge Recombination (KR): Generate new ideas by merging existing ones.

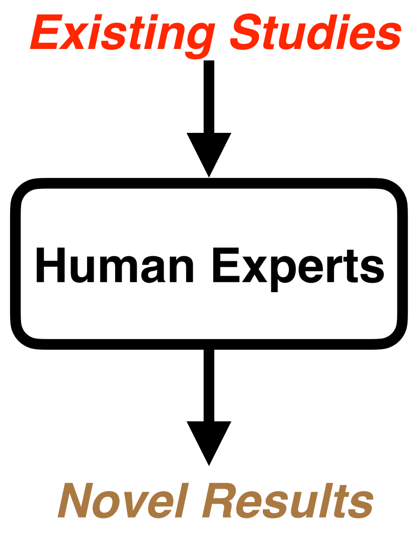
Incremental Innovation (II): Make gradual improvements to existing ideas.

Experiments (Evaluation): Test and score ideas to validate their potential.



When TN-SS Meets LLMs

How Human Experts Conduct Innovative Research



Idea Pool: Gather information through literature reviews and paper retrieval

Knowledge Categorization (KC): Refine ideas into knowledge clusters.

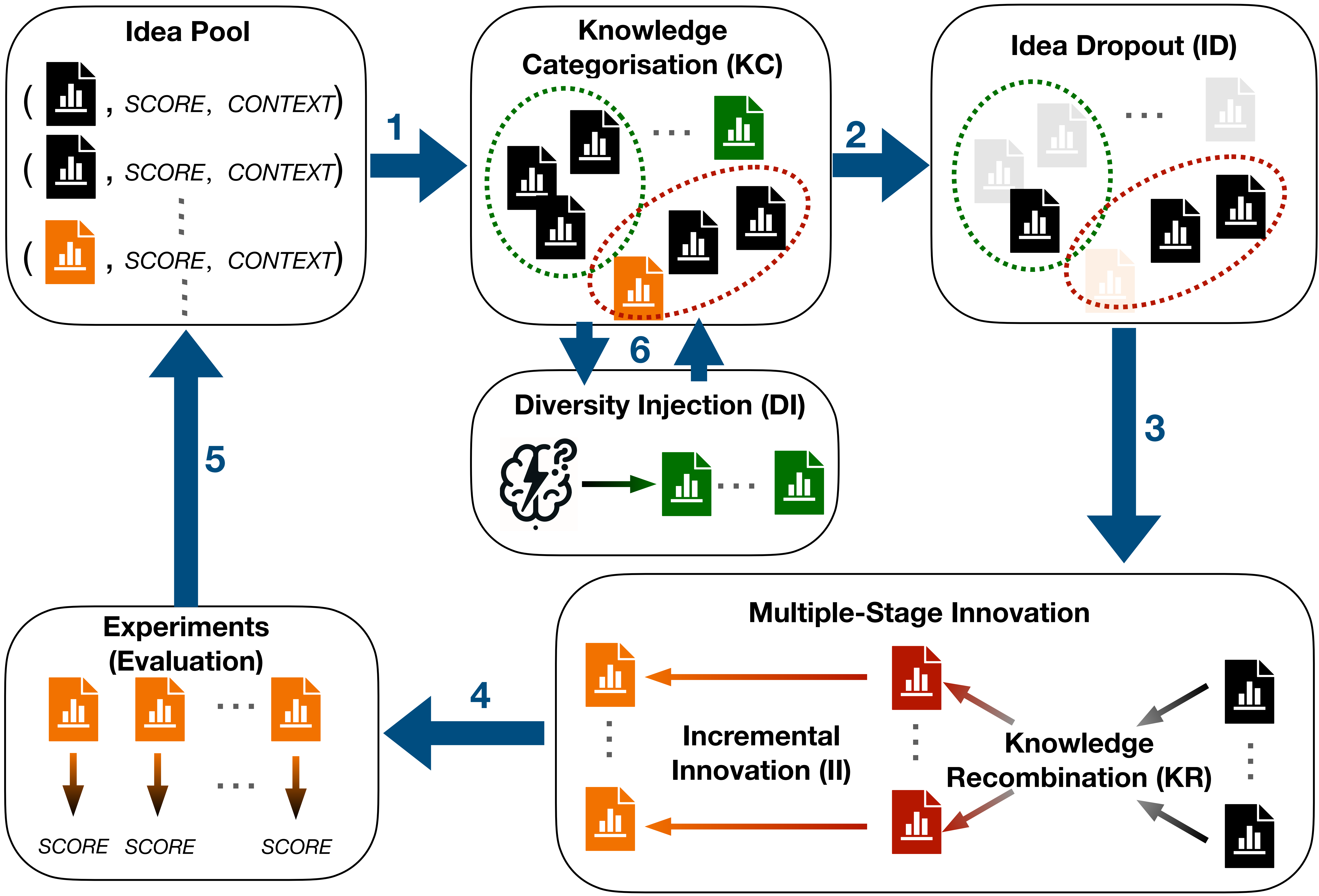
Idea Dropout (ID): Filter out less interesting ideas to focus on the most promising ones.

Knowledge Recombination (KR): Generate new ideas by merging existing ones.

Incremental Innovation (II): Make gradual improvements to existing ideas.

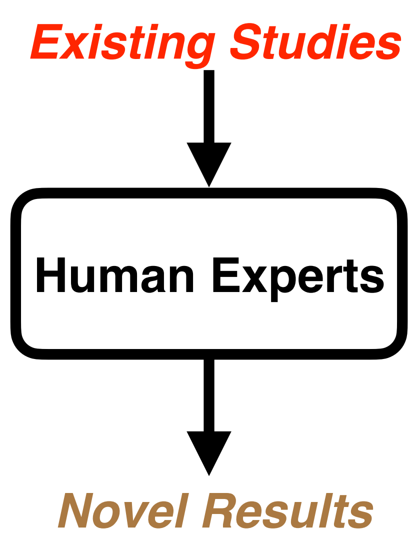
Experiments (Evaluation): Test and score ideas to validate their potential.

Diversity Injection (DI): Introduce new, orthogonal ideas through brainstorming or external feedback.



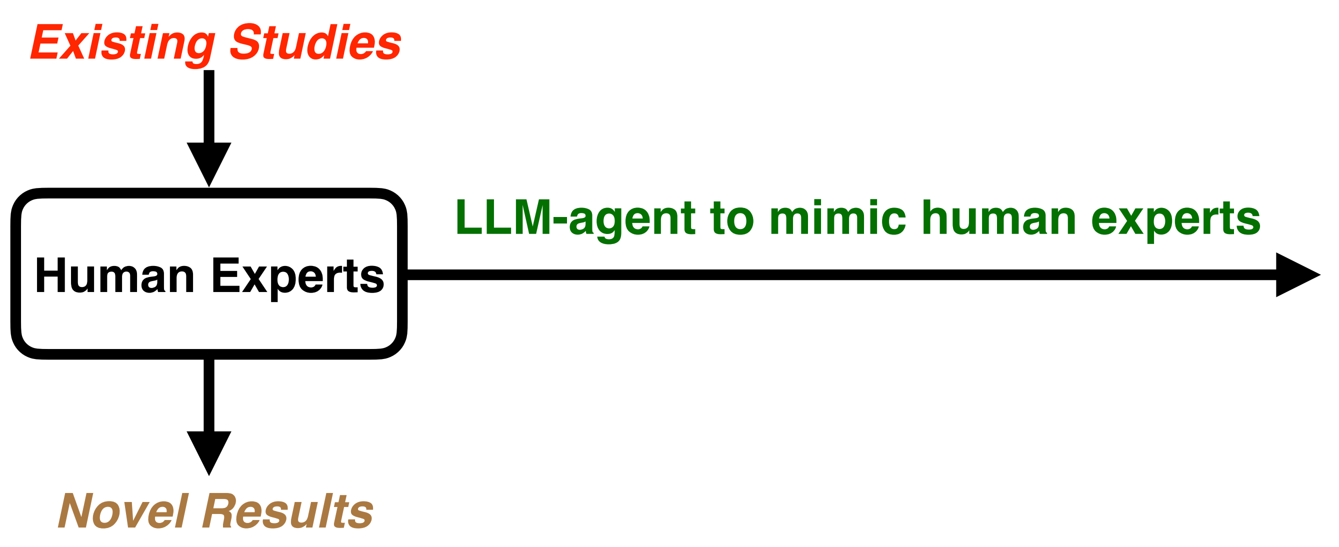
When TN-SS Meets LLMs

LLM-agent Driven Innovative Research



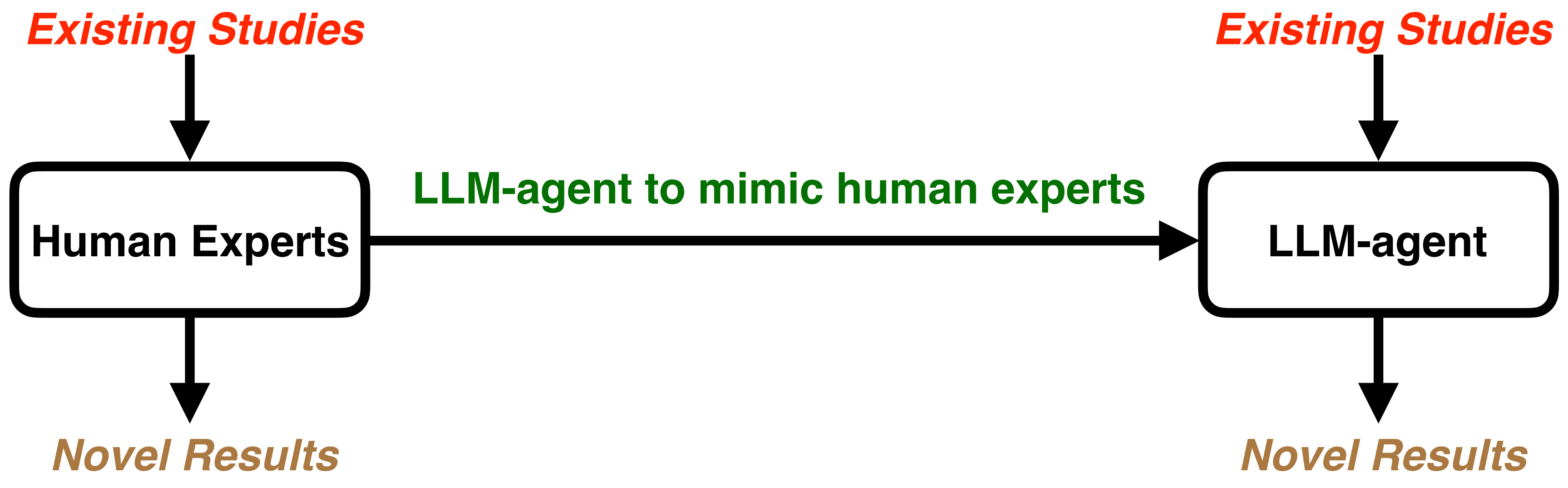
When TN-SS Meets LLMs

LLM-agent Driven Innovative Research



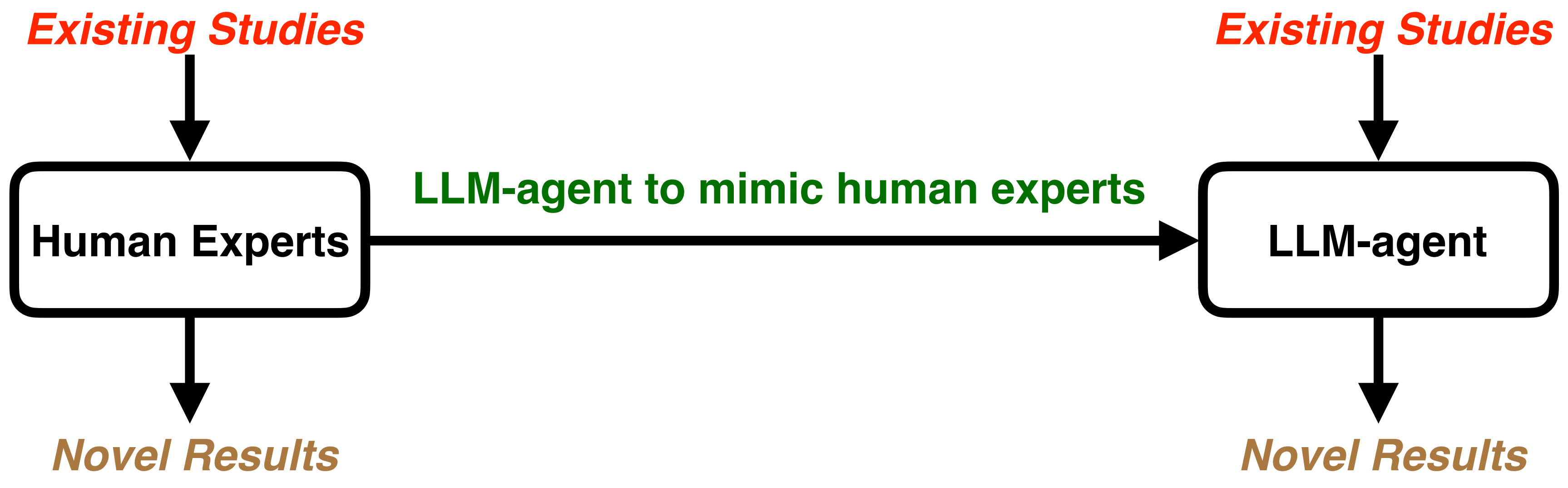
When TN-SS Meets LLMs

LLM-agent Driven Innovative Research



When TN-SS Meets LLMs

LLM-agent Driven Innovative Research



Algorithm 1: # centroid
. . .
⋮ (omitted)
Algorithm *N*: # centroid
. . .
=============
New Algorithm:
. . .
==============
Which algorithm in the above is methodologically most similar to the new algorithm? Just give me the function number with no other words.

Knowledge Categorization (KC) Prompt

Algorithm 1: . . .
. . .
Algorithm 1 score:
⋮ (omitted)
Algorithm *N*: . . .
. . .
Algorithm *N* score:
Algorithms 1 to *N* are implementations of the 'GenerateSample' function. A lower score implies better performance.

Learning from their results, think about what works and what doesn't, provide *M* novel methods with lower scores. You are encouraged to be creative to incorporate novel ideas but do not simply stack methods together.

Knowledge Recombination (KR) Prompt

Algorithm 1:
. . .
⋮ (omitted)
Algorithm *N*:
. . .
Independently make improvements over these Algorithms that will increase their practical performance (not on the code efficiency, readability and parallel processing level). You are encouraged to be creative to incorporate novel ideas.

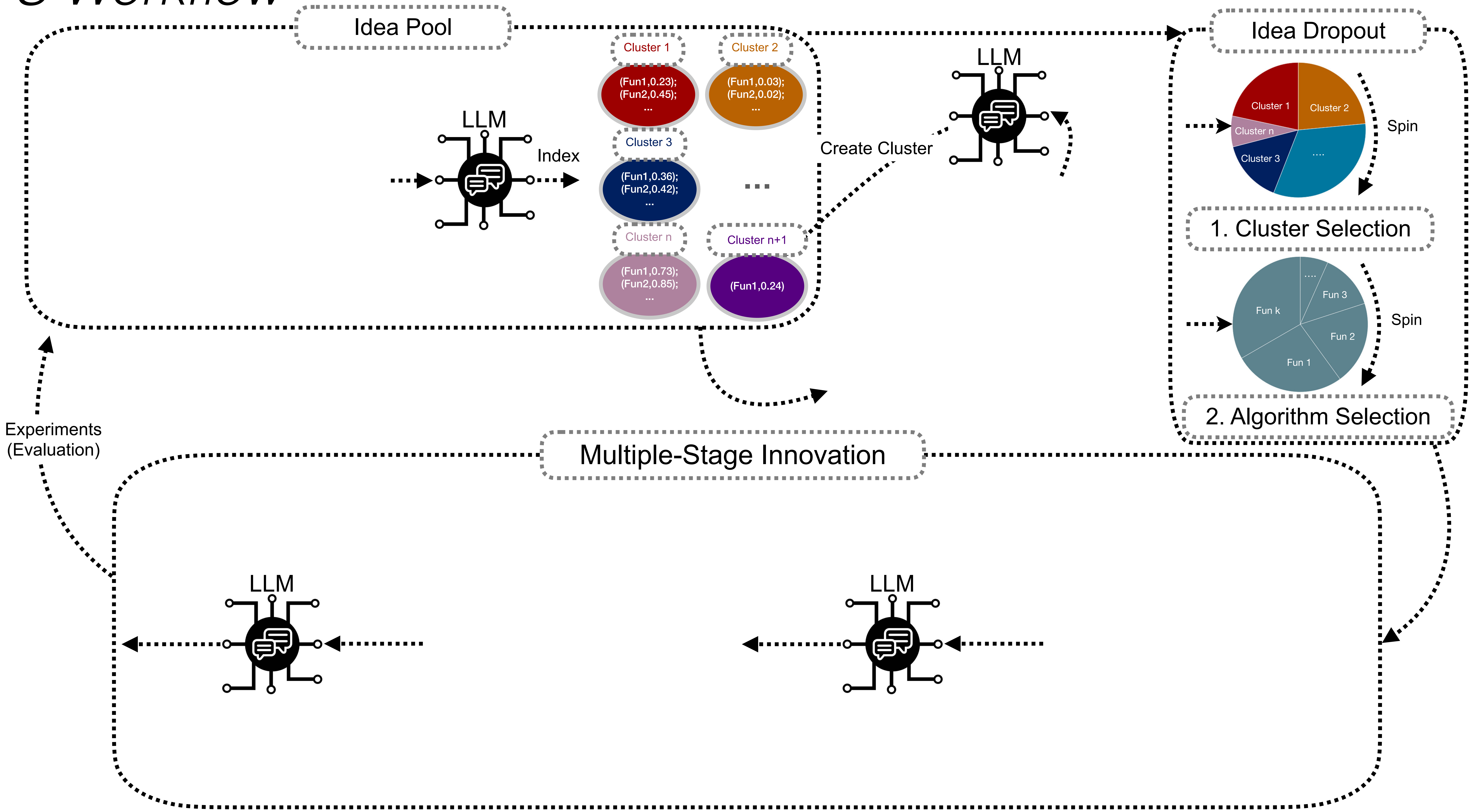
Incremental Innovation (II) Prompt

Algorithm 1: # centroid
. . .
⋮ (omitted)
Algorithm *N*: # centroid
. . .
Give me a novel 'GenerateSample' that is methodologically different from the above algorithms. You are encouraged to be creative to incorporate novel ideas but do not simply stack methods together.

Diversity Injection (DI) Prompt

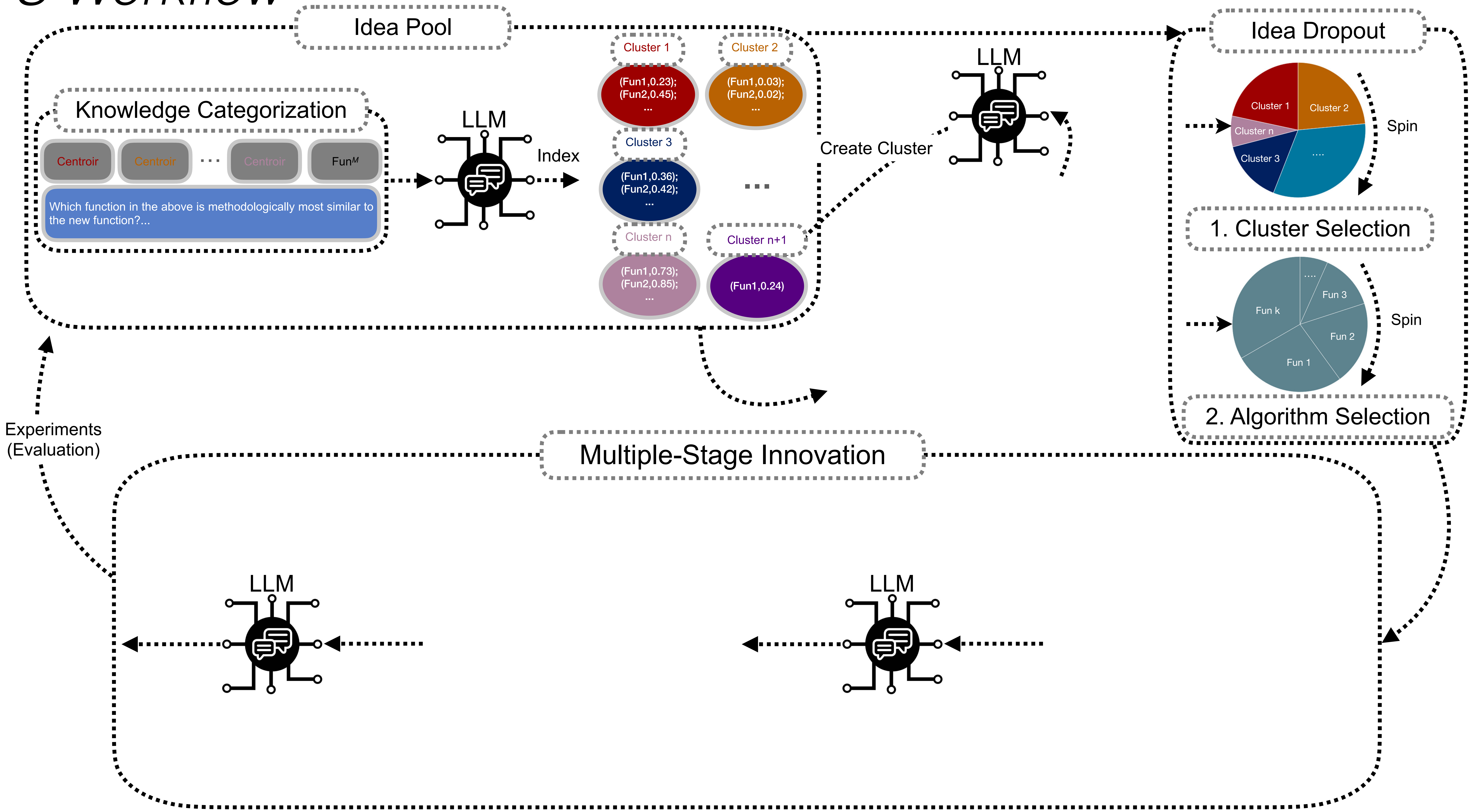
When TN-SS Meets LLMs

tnGPS Workflow



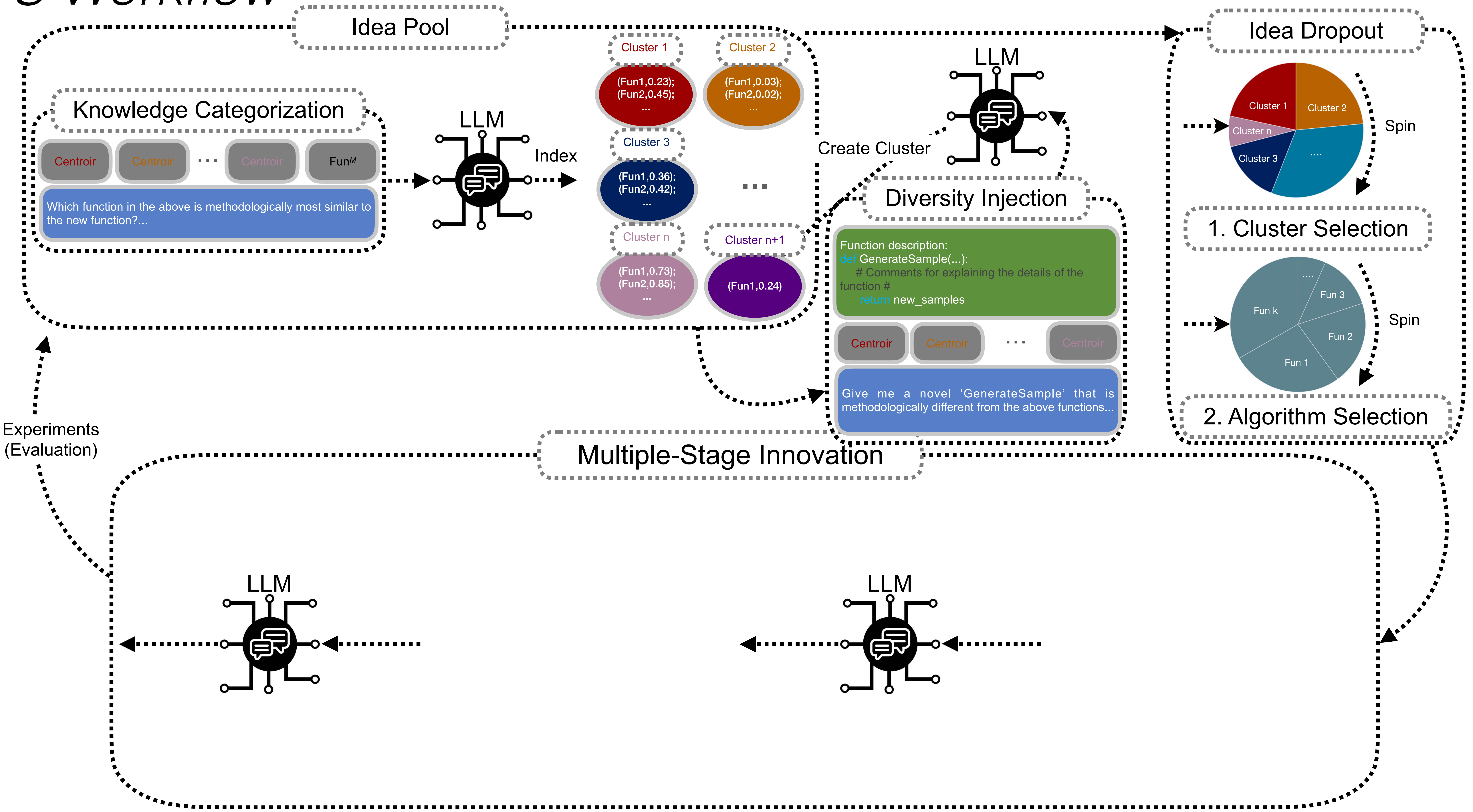
When TN-SS Meets LLMs

tnGPS Workflow



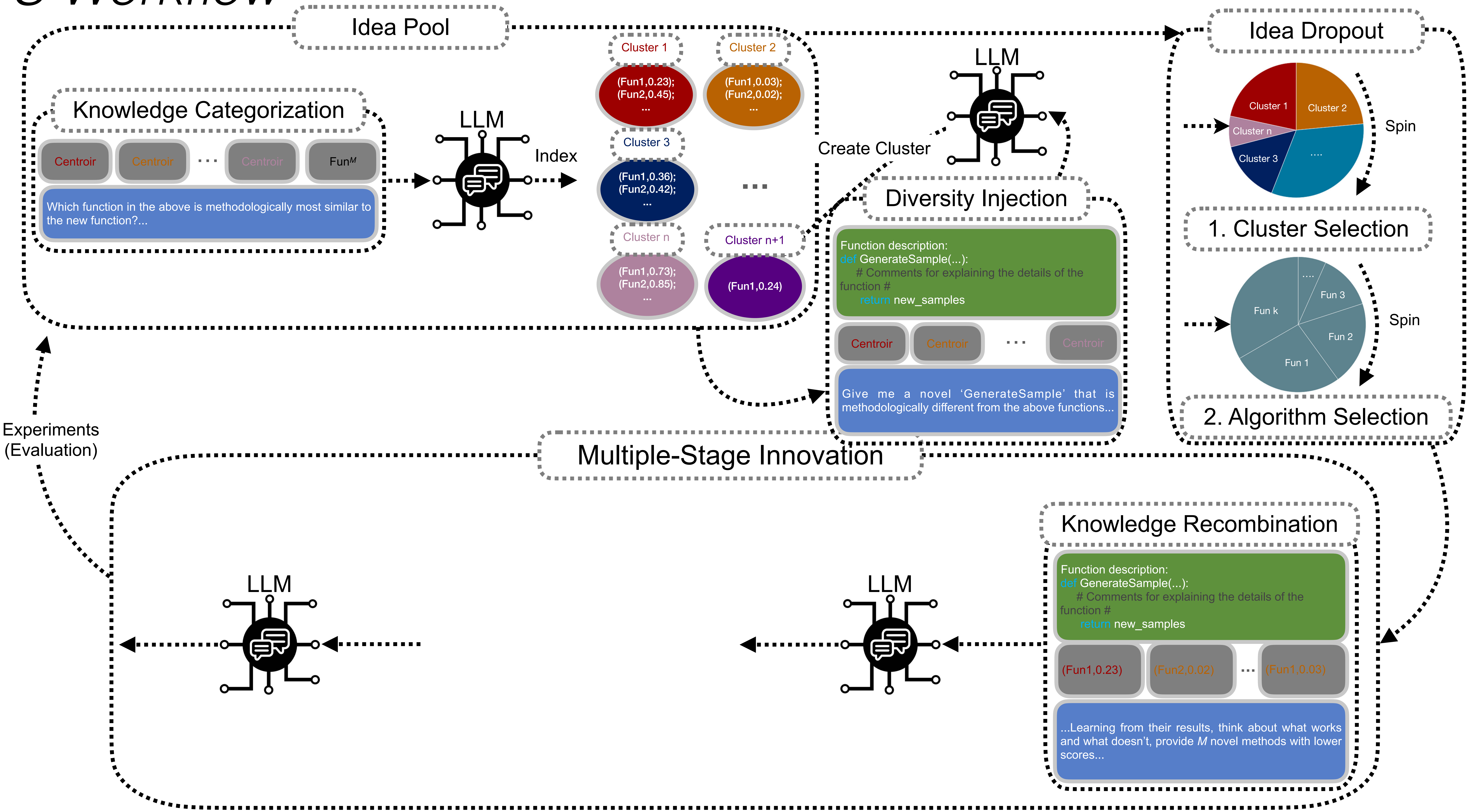
When TN-SS Meets LLMs

tnGPS Workflow



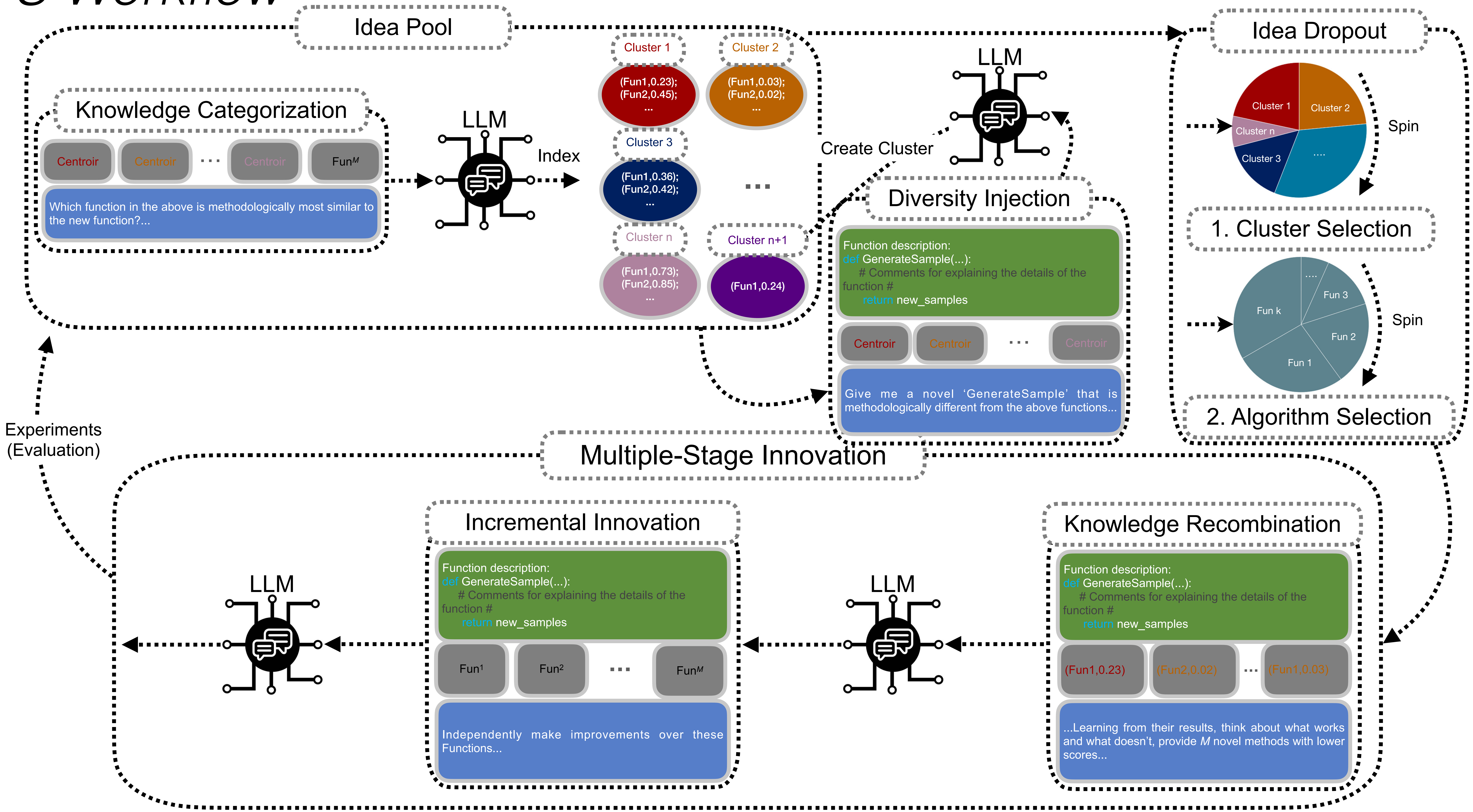
When TN-SS Meets LLMs

tnGPS Workflow



When TN-SS Meets LLMs

tnGPS Workflow



Experimental Results

Natural Images Compression

Objective:

$$F(G,\mathbf{r}) = \underbrace{\frac{1}{\epsilon(G,\mathbf{r})}}_{\text{compression ratio (CR)}} + \lambda \cdot \underbrace{\min_{\mathcal{Z}\in TNS(G,\mathbf{r})} \|\mathcal{X}-\mathcal{Z}\|^2 / \|\mathcal{X}\|^2}_{\text{relative square error (RSE)}},$$

Experimental Results

Natural Images Compression

Objective:

$$F(G,\mathbf{r}) = \underbrace{\frac{1}{\epsilon(G,\mathbf{r})}}_{\text{compression ratio (CR)}} + \lambda \cdot \underbrace{\min_{\mathcal{Z}\in TNS(G,\mathbf{r})} \left\|\mathcal{X}-\mathcal{Z}\right\|^2 / \left\|\mathcal{X}\right\|^2}_{\text{relative square error (RSE)}},$$

Settings of tnGPS: ***four images for training***, ***ten images for testing***.

Experimental Results

Natural Images Compression

Objective:

$$F(G,\mathbf{r}) = \underbrace{\frac{1}{\epsilon(G,\mathbf{r})}}_{\text{compression ratio (CR)}} + \lambda \cdot \underbrace{\min_{\mathcal{Z}\in TNS(G,\mathbf{r})} \|\mathcal{X}-\mathcal{Z}\|^2 / \|\mathcal{X}\|^2}_{\text{relative square error (RSE)}},$$

Settings of tnGPS: ***four images for training***, ***ten images for testing***.

Experimental Results

Natural Images Compression

Objective:

$$F(G,\mathbf{r}) = \underbrace{\frac{1}{\epsilon(G,\mathbf{r})}}_{\text{compression ratio (CR)}} + \lambda \cdot \underbrace{\min_{\mathcal{Z}\in TNS(G,\mathbf{r})} \|\mathcal{X}-\mathcal{Z}\|^2 / \|\mathcal{X}\|^2}_{\text{relative square error (RSE)}},$$

Settings of tnGPS: *four images for training*, *ten images for testing*.

Ho-1, Ho-2 and Ho-3 are algorithms discovered by tnGPS

Experimental Results

Ablation Studies

Experimental Results

Ablation Studies

	baseline	**tnGPS**	KR	II	DI
Objective	0.1558	**0.1102**	0.1308	0.1273	0.1239

The results highlight the importance of the 'KR', 'II', and 'DI' components.

Experimental Results

Ablation Studies

	baseline	**tnGPS**	KR	II	DI
Objective	0.1558	**0.1102**	0.1308	0.1273	0.1239

The results highlight the importance of the 'KR', 'II', and 'DI' components.

baseline	**GPT-4**	**GPT-3.5**	**Claude-1**	**Claude-2**	**Incomplete descriptions**
0.1847	**0.1813**	0.1842	0.1840	0.1834	0.1819

More powerful LLMs like GPT-4 enhance tnGPS performance.

Experimental Results

Insights Gained from the Generated Algorithms

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

- *Non-Markovian searching dynamic (Ho-1, Ho-2, Ho-3)*

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

- *Non-Markovian searching dynamic (Ho-1, Ho-2, Ho-3)*
- *Inverse annealing for mutation (Ho-1)*

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

- *Non-Markovian searching dynamic (Ho-1, Ho-2, Ho-3)*
- *Inverse annealing for mutation (Ho-1)*
- *Gaussian perturbation mutation (Ho-2)*

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

- *Non-Markovian searching dynamic (Ho-1, Ho-2, Ho-3)*
- *Inverse annealing for mutation (Ho-1)*
- *Gaussian perturbation mutation (Ho-2)*
- *Best structure crossover (Ho-3)*

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

- *Non-Markovian searching dynamic (Ho-1, Ho-2, Ho-3)*
- *Inverse annealing for mutation (Ho-1)*
- *Gaussian perturbation mutation (Ho-2)*
- *Best structure crossover (Ho-3)*

The algorithm discovered by tnGPS implicitly reflects the hidden structure of the data

Experimental Results

Insights Gained from the Generated Algorithms

tnGPS can leverage insights gained from the existing algorithms and the embedded knowledge in LLMs for novel algorithm generation

- *Non-Markovian searching dynamic (Ho-1, Ho-2, Ho-3)*
- *Inverse annealing for mutation (Ho-1)*
- *Gaussian perturbation mutation (Ho-2)*
- *Best structure crossover (Ho-3)*

The algorithm discovered by tnGPS implicitly reflects the hidden structure of the data

- *Boundary mutation (Ablation 1)*

Concluding Remarks

- **tnGPS: a LLM-driven framework for discovering new TN-SS algorithms.**
- **tnGPS is designed by prompting LLMs to mimic human experts.**
- **LLMs provide us new ideas of solving more broad tensor problems.**

Thank You