Automated Theorem Proving

Survey

Agenda

- Self introduction
- Large Language Models (LLM): What and How
- Automatic Theorem Provers (ATP): What and How
- How LLM benefits the task
  - Dense Passage Retrieval
  - Dense Passage Retrieval in Automatic Theorem Provers

Agenda

- Self introduction
- Large Language Models (LLM): What and How
- Automatic Theorem Provers (ATP): What and How
- How LLM benefits the task
  - Dense Passage Retrieval
  - Dense Passage Retrieval in Automatic Theorem Provers

What kept me busy

Hardware engineer 2015~

- Digital circuit+PCB design.
- Bootloader, kernel, driver customization

❑ National certification of embedding system (ARM9 arch, 2017)

Computer Vision & algorithm engineer 2017~

- CV algorithms, MPI @ CUDA
- H.264 streaming @ RTMP

❑ Graduation project: panoramic video living system

AI Researcher In CV domain 2019~

TOHOKU UNIVERSITY

- Oversea research
- AI in NLP, CV, time series forecasting, etc.
- HPC @ SuperComputer

❑ ***Kamei*** Scholarship

❑ PRUM 2021: Moving Scene Text Detection Using Synthetic Scene Text Video for Training
  - Novelty: trajectory synthesis algorithm
  - Novelty: amended representation for edge vibration and supporting algorithm

R&D engineer 2021~

CORPY&Co.

- Anomaly detection (2 proj.)
- Optical Flow (1 proj.)
- Object tracking (2 proj.)
- Depth estimation (2 proj.)
- Time series forecasting (1 proj.)
- Unsupervised learning (2 proj.)
- Multimodal learning (3 proj.)

❑ Main engineer

❑ R&D leader (1/10)

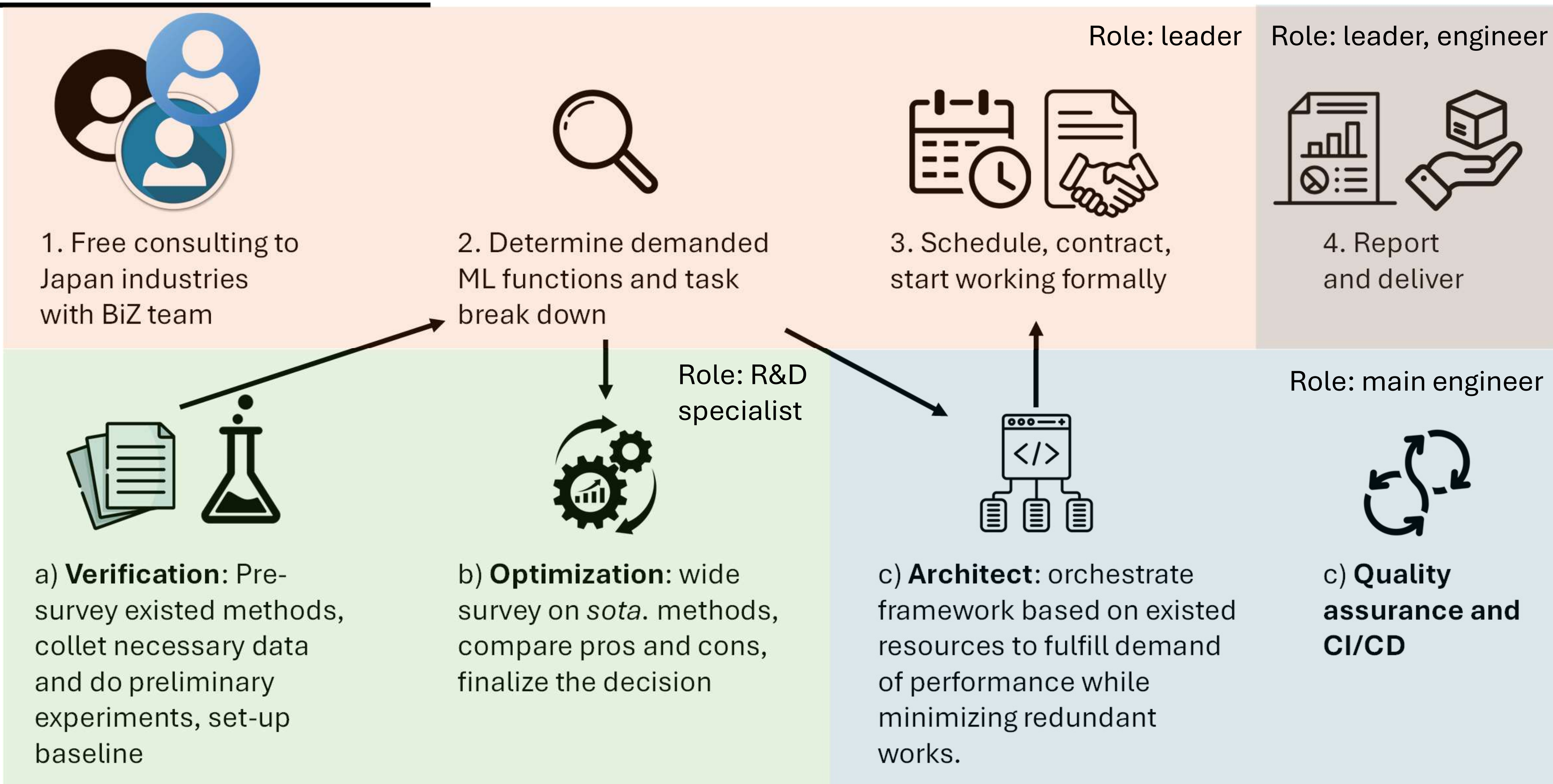
Cloud engineer 2022~

Cloud DevOps Engineer GOOGLE CLOUD CERTIFIED · PROFESSIONAL

- Backend dev. via Golang/Python
- Database Operation.
- MLOpt.
- Pipeline construction

❑ Google certificated ML eng.

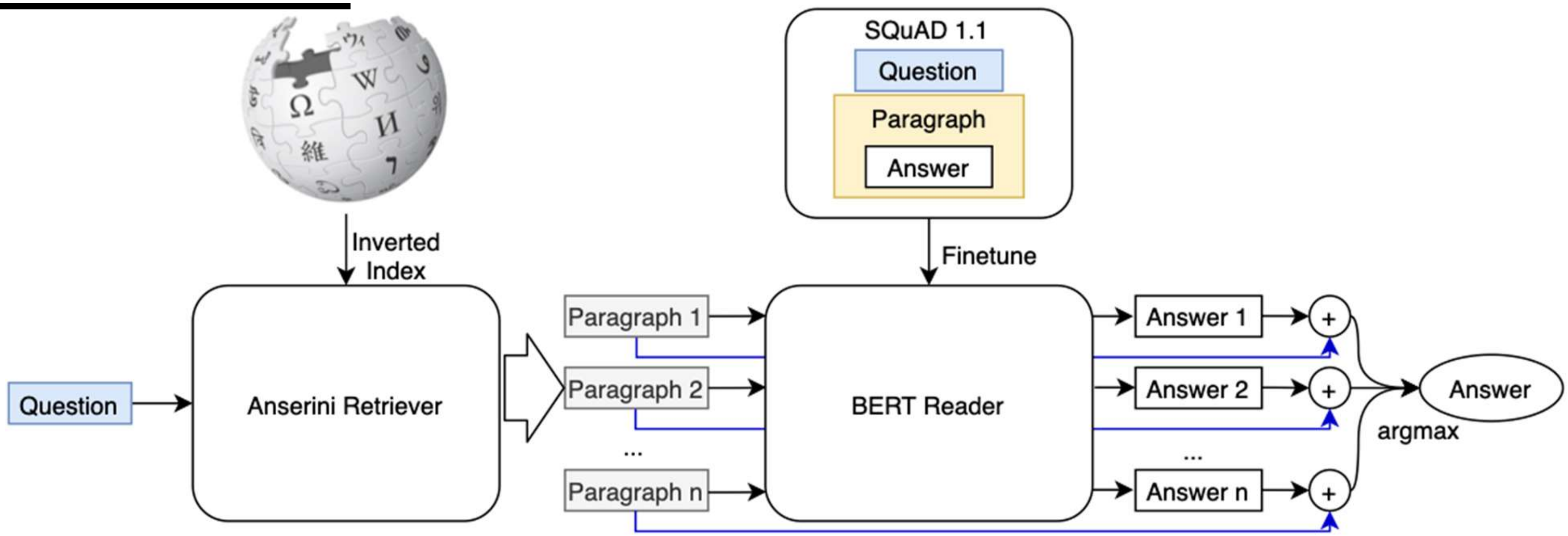
Workflow


Role: leader
Role: leader, engineer
1. Free consulting to Japan industries with BiZ team
2. Determine demanded ML functions and task break down
3. Schedule, contract, start working formally
4. Report and deliver
Role: R&D specialist
Role: main engineer
a) **Verification**: Pre-survey existed methods, collet necessary data and do preliminary experiments, set-up baseline
b) **Optimization**: wide survey on *sota*. methods, compare pros and cons, finalize the decision
c) **Architect**: orchestrate framework based on existed resources to fulfill demand of performance while minimizing redundant works.
c) **Quality assurance and CI/CD**

Agenda

- Self introduction
- Large Language Models (LLM): What and How
- Automatic Theorem Provers (ATP): What and How
- How LLM benefits the task
  - Dense Passage Retrieval
  - Dense Passage Retrieval in Automatic Theorem Provers

LLM: A tool to solve Open-Domain Question Answering



Common knowledge **Pre**

Articles

+ Weakly Supervised Learning

Pretrained model

Domain knowledge **T**

Questions Answers

+ Supervised Learning / Fine turning

Final model

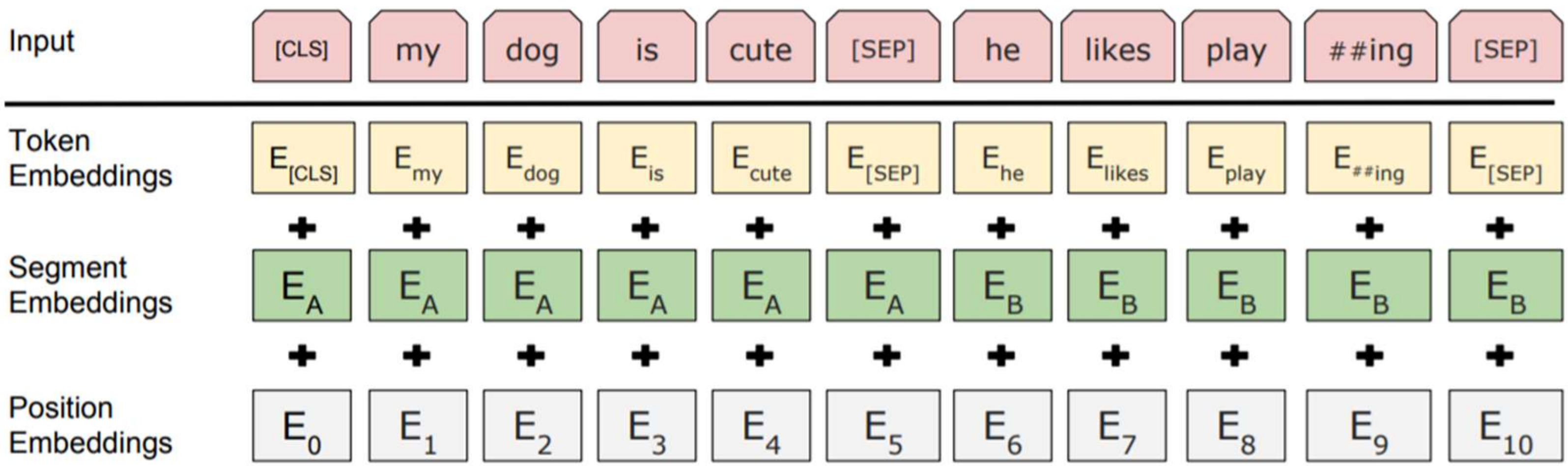
I

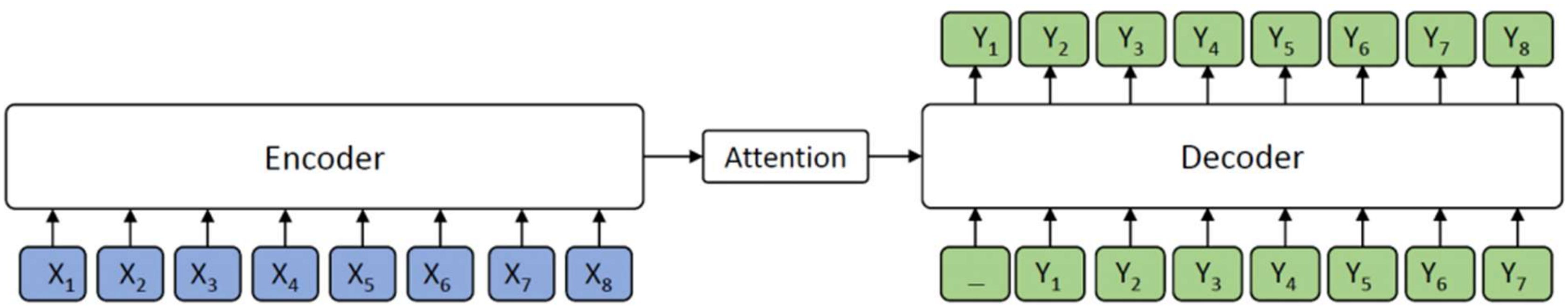
Questions > Final model

Answers

Milestone of LLM model: BERT

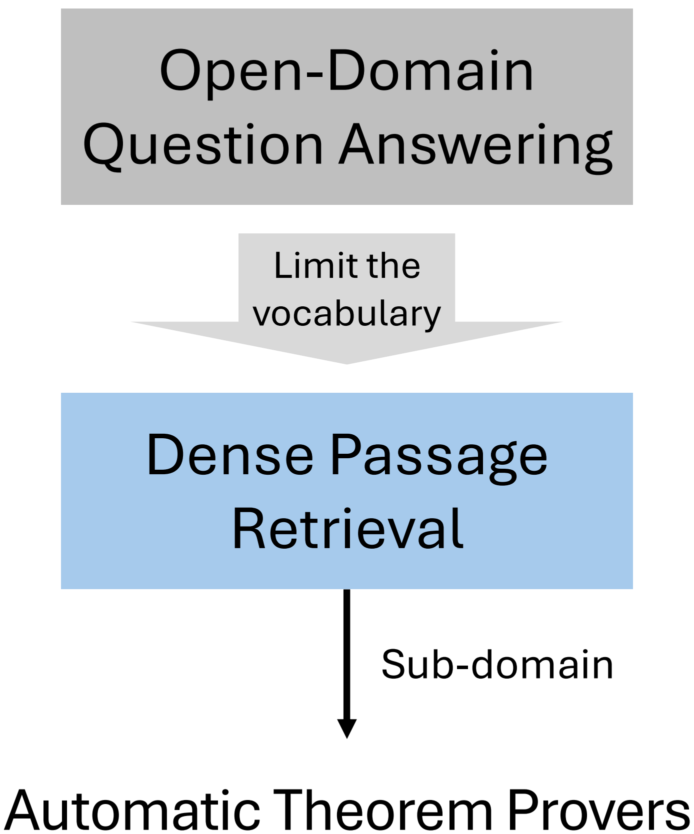
Encoder





BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NACCL 2019, Google AI

LLM: A tool to solve Open-Domain Question Answering (ODQA)



Question:
Goal to proof

Objects:

x y : ℕ

Assumptions:

hx : succ x ≤ succ y

Goal:

x ≤ y

Answer: each step to proof the goal

```
1. cases hx with d hd
2. use d
3. rw [succ_add] at hd
4. apply succ_inj at hd
5. exact hd
```

Agenda

- Self introduction
- Large Language Models (LLM): What and How
- Automatic Theorem Provers (ATP): What and How
- How LLM benefits the task
  - Dense Passage Retrieval
  - Dense Passage Retrieval in Automatic Theorem Provers

ATP: what is it, and how it benefit the BiZ & Engineering

Demand from development, simulation, etc.

Formulize

Objects:

$x\ y : \mathbb{N}$

Assumptions:

hx : succ x ≤ succ y

Goal:

x ≤ y

(Domain specific) Questions

~10x / day?

Very hard cases?

Theorem Provers

Automatic

Theorem Provers

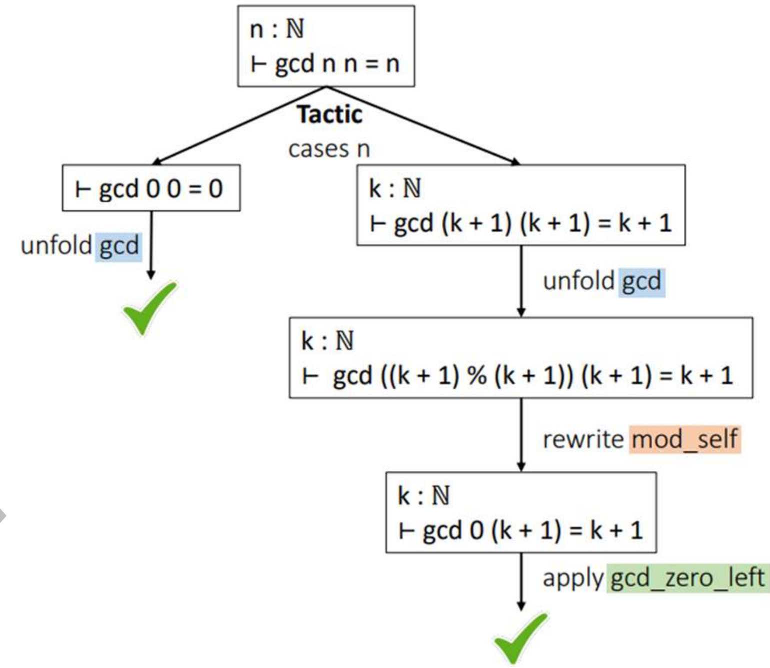
~100x

7*24

\+

ML model (LLM)

Human can concentrate on high-level works



Approximate / Statistic

-> theorem -> ***reliability***

-> compulsory requirements in medical, engineering, bionic, and other domains

-> Open the domain

ATP: reason to choose LLM

- ☑ Necessary demand
- ☑ Available resources for training framework
  - ☑ Open-source engine for logical inferences: Lean
  - ☑ Datasets
- ☑ Feasibility for functional
  - ☑ Methods verified on this domain (specialistic)
  - ☑ Methods verified on other domains (generalization and robustness)
  - ☑ Pre-trained networks for agile development

Agenda

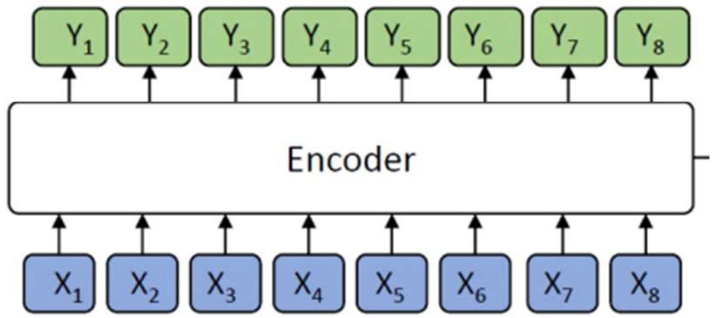
- Self introduction
- Large Language Models (LLM): What and How
- Automatic Theorem Provers (ATP): What and How
- How LLM benefits the task
  - Dense Passage Retrieval
  - Dense Passage Retrieval in Automatic Theorem Provers

Dense Passage Retrieval for Open-Domain Question Answering

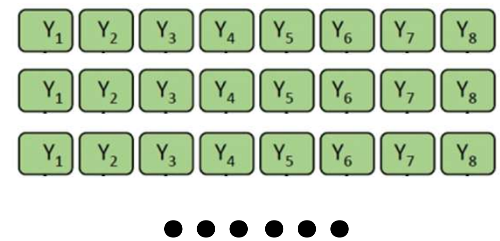
Question q_i

GT answer p_i^+

Irrelative answers $p_{i,n}^-$



Backbone, aka. **encoder**



Feature vectors
In latent space

T | **I**

$$L(q_i, p_i^+, p_{i,1}^-, \cdots, p_{i,n}^-)$$
$$= -\log \frac{e^{\mathrm{sim}(q_i, p_i^+)}}{e^{\mathrm{sim}(q_i, p_i^+)} + \sum_{j=1}^{n} e^{\mathrm{sim}(q_i, p_{i,j}^-)}}.$$

Loss function: grouping In latent space

$$\mathrm{sim}(q, p) = E_Q(q)^\intercal E_P(p).$$

Inference: find the Top-k answers
Based on L2 similarity

Dense Passage Retrieval for Open-Domain Question Answering. Apr 2020.

Dense Passage Retrieval for Open-Domain Question Answering

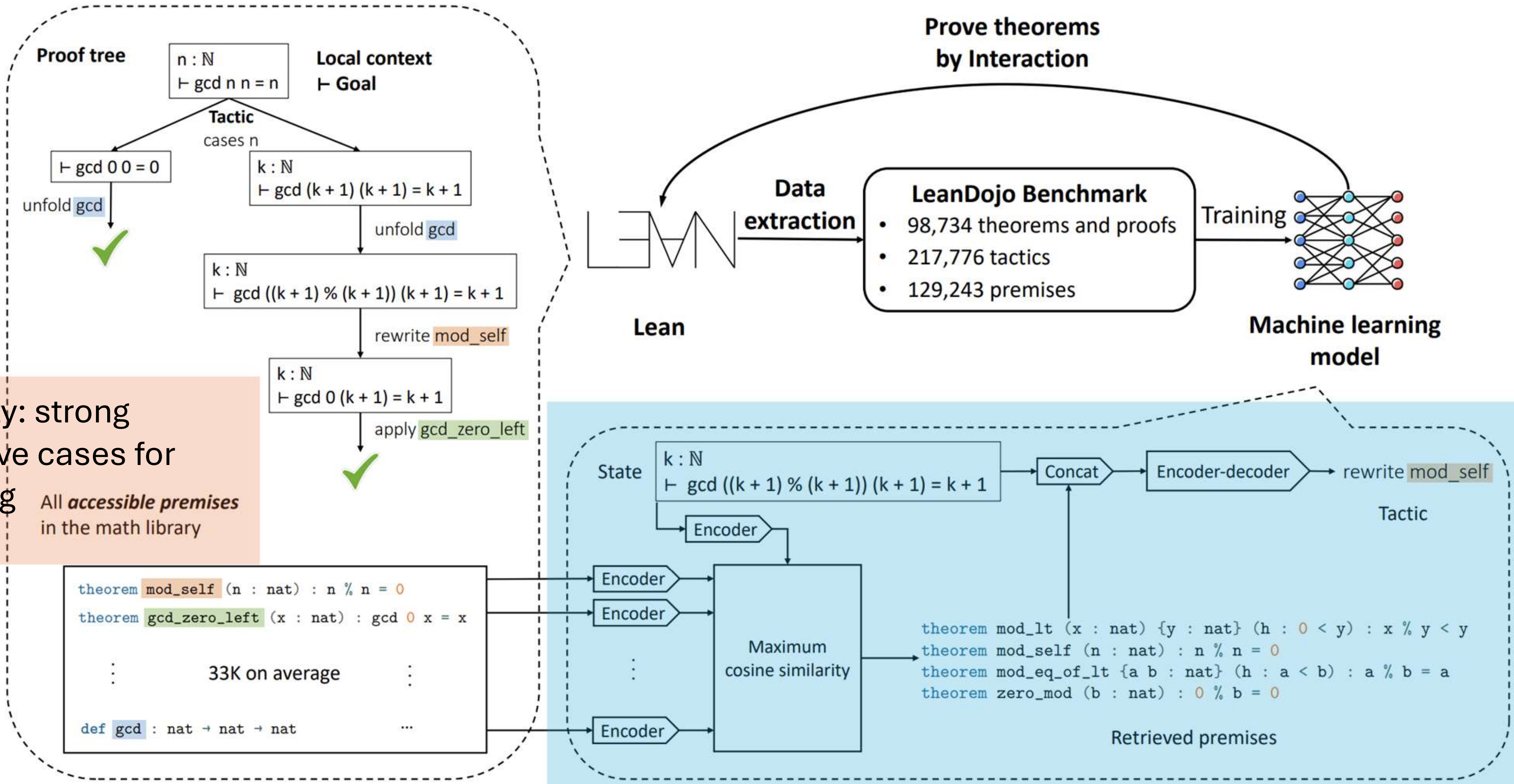
Construction / Generation **NLP**	**DPR** **Selection / Optimization**
Predict sequence of words to construct answer.	Select from pre-built answer set.
High degrees of freedom to construct answers.	Finite, controllable answer set.
Diversity.	Precise.
	Suit for graph construction with finite types of edge.

Dense Passage Retrieval for Open-Domain Question Answering. Apr 2020.

Agenda

- Self introduction
- Large Language Models (LLM): What and How
- Automatic Theorem Provers (ATP): What and How
- How LLM benefits the task
    - Dense Passage Retrieval
    - Dense Passage Retrieval in Automatic Theorem Provers

LeanDojo: Theorem Proving with Retrieval-Augmented Language



Novelty: strong negative cases for training

Limit searching space to accessible premises

Model: ByT5

Novelty: retrieval searching mechanism (**D**ense **P**assage **R**etrieval @ math)

LeanDojo: Theorem Proving with Retrieval-Augmented Language Models. Jun 2023

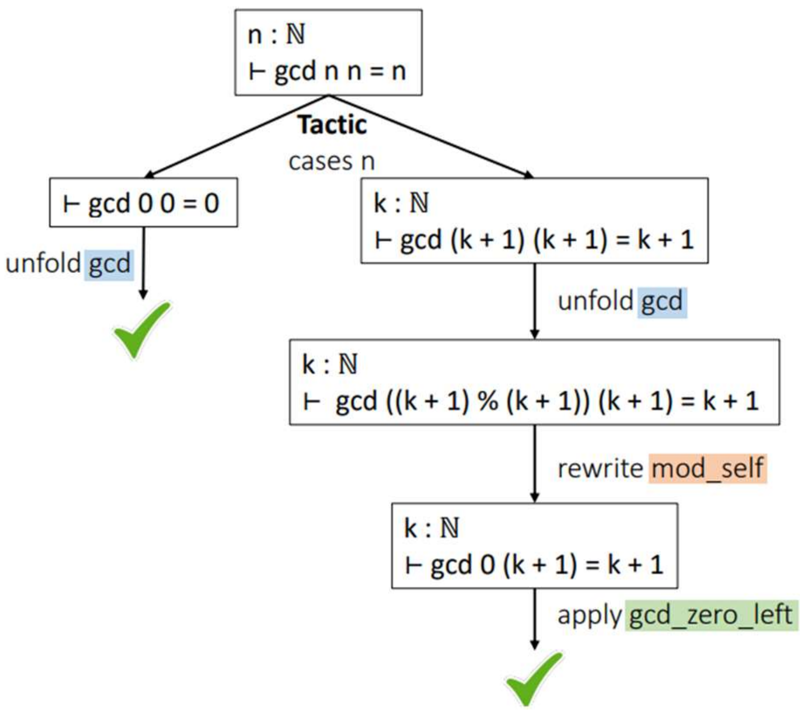
Appendix 1

Survey of Automatic Theorem Provers

ATP: introduction - survey method

1. Survey paper (published within 2 years, if available)
   - Objective:
     a) Understand categorization of methods
     b) Clarify the current state of development of domain
     c) List challenges
   - *Learning Guided Automated Reasoning: A Brief Survey*. Mar 2024
2. *Sota*. methods with pre-trained models (if available)
   - *LeanDojo: Theorem Proving with Retrieval-Augmented Language Models*. Oct 2023.
   - *Llemma: An Open Language Model For Mathematics*. Oct 2023.
   - An In-Context Learning Agent for Formal Theorem-Proving. Feb 2024.

ATP: introduction - task objective



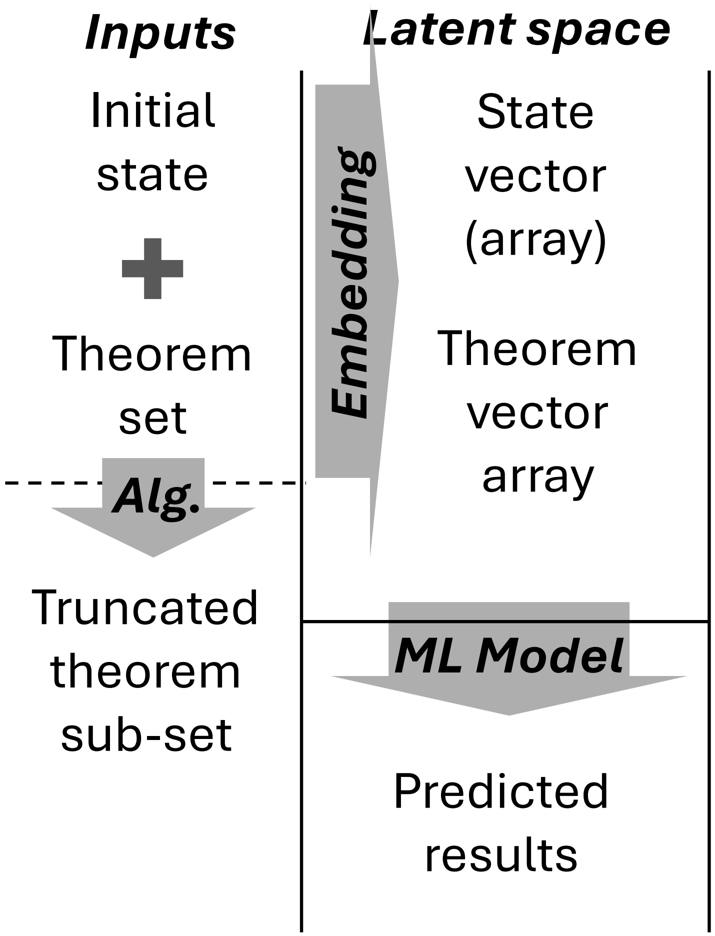
- Training input: tree / state machine + available theorems
  - Nodes: states
  - Edges: theorems
- Regression target: classification / top-k selection
  - which theorem to use in each step

T

E

- Inference input: root node / initial state + available theorems
  - What to proof, and conditions
- Inference target: classification / top-k selection
  - which theorem to use in each step

ATP: introduction - workflow and major challenges



** For machine learning family only*

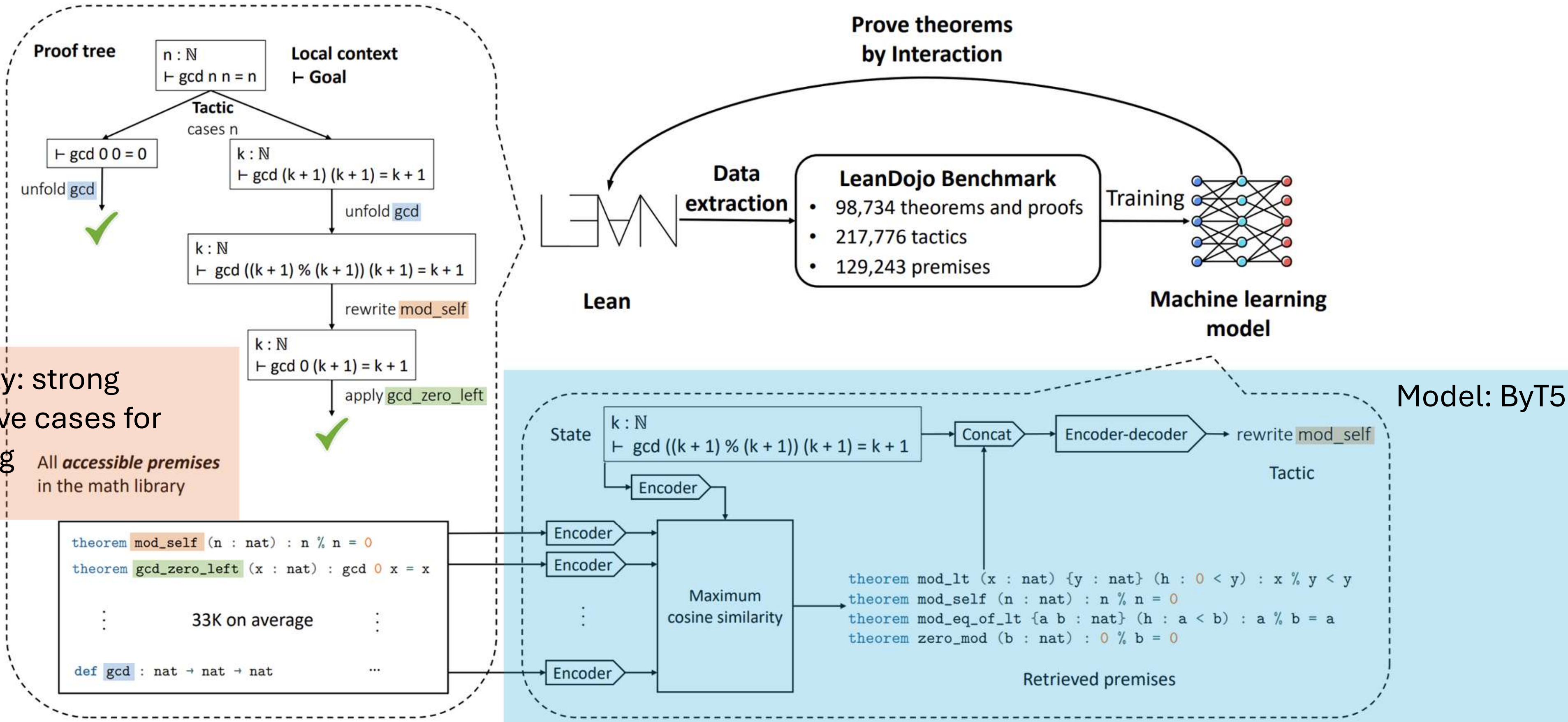
Consents:

1. Inputs of each need state, the longer state reserved the better context the model can comprehend.
2. Fix-size latent space.
3. NLP model for token suggestion task.

Challenges:

1. Pre-processing alg. for searching space (available theorems in set)
2. Embedding methods
   a) Finite set vs infinite set of unit of theorems.
   b) Encoding method for theorems with variable, location of variable, and operators.

LeanDojo: Theorem Proving with Retrieval-Augmented Language



Novelty: strong negative cases for training

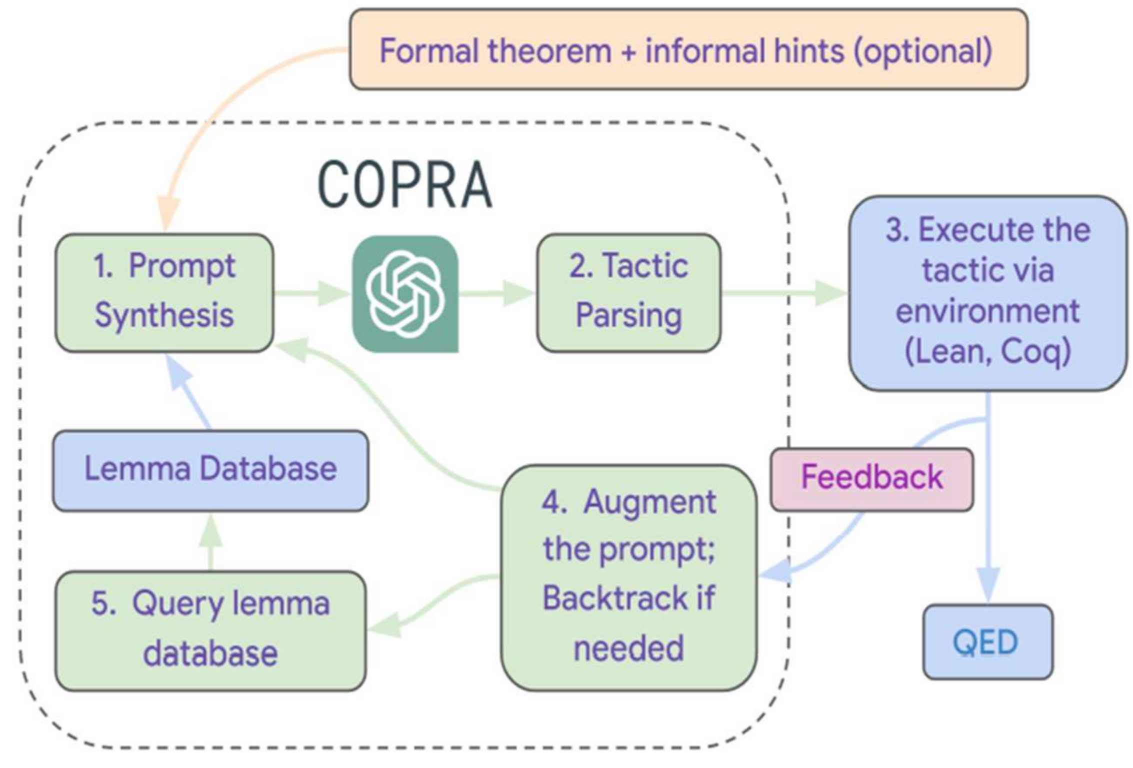
Limit searching space to accessible premises

Model: ByT5

Novelty: retrieval searching mechanism (**D**ense **P**assage **R**etrieval @ math)

LeanDojo: Theorem Proving with Retrieval-Augmented Language Models. Jun 2023

An In-Context Learning Agent for Formal Theorem-Proving



An In-Context Learning Agent for Formal Theorem-Proving. Feb 2024

GPT-4 + Retrieval + Informal

- Better performance on miniF2F.
- Need token for GPT-4 API.

(a)

```
theorem mod_arith_2
(x : ℕ) : x % 2 = 0
  → (x * x) % 2 = 0
:=
begin
  intro h,
  rw nat.mul_mod,
  rw h,
  rw nat.zero_mul,
  refl,
end
```

(b)

```
x: ℕ
h: x % 2 = 0
⊢ x * x % 2 = 0
```

(c)

```
begin
  intro h,
  have h1 : x = 2 * (x
     / 2)
     := (nat.
    mul_div_cancel' h)
    .symm,
  rw h1,
  rw nat.mul_div_assoc
    _
    (show 2 | 2, from
    dvd_refl _),
  rw [mul_assoc, nat.
    mul_mod_right],
end
```

Figure 2: (a) A Lean theorem and a correct proof found by COPRA. (b) Proof state after the first tactic. (c) An incorrect proof generated by GPT-4.

ATP: summary

Title	Year	Pass @1	Size	Novelty
LeanDojo: Theorem Provin g with Retrieval-Augmented Lan guage Models	Oct 2023	26.5		Alg. for retrieval The 1st retrieval architecture model in this domain Introduced strong negative for training
LLEMMA: AN OPEN LANG UAGE MODEL FOR MATHEMATICS	Mar 2024	25.8 26.2	34B 7B	Code Llama -> domain specified in math
An In-Context Learning Ag ent for Formal Theorem-Pr oving	Feb 2024	30.7		GPT-4 + Retrieval + Informal