RESEARCH STATEMENT

Haotian Chen (htchen18@fudan.edu.cn)

"I have not failed. I've just found 10,000 ways that won't work."

— *Thomas Alva Edison*

The progress of humanity is driven by those successful discoveries accompanied by countless failed experiments. Researchers often seek the potential research directions by reading and then verifying them through experiments. The process imposes a significant burden on researchers. In the past decade, the data-driven black-box deep learning method has demonstrated its effectiveness in a wide range of real-world scenarios, which exacerbates the experimental burden of researchers and thus renders the potential successful discoveries veiled. The presence of the issue raises a question at the heart of my research: **How do we automate such a research and development (R&D) process by ML systems (e.g., AI agent) in the hope of realizing the revolution of human productivity and living standards?** I conduct both theoretical (Section 4) and practical (Section 1,2,3) research to promote each other.

The overall assumption of my research is that knowledge **will facilitate the performance of AI Agent in Automatic R&D**. Here are three questions. **Q1**: *What's the mission of AI Agent in automatic R&D?* **Q2**: *What is knowledge?* **Q3**: *How to use knowledge to empower AI Agent?*

A1. As shown in the orange blocks in Figure 1, I assume three crucial (even necessary) steps in realizing automatic R&D: Idea generation, method implementation, and making decisions based on experimental results. The three steps form the life cycle of automatic R&D. I hope that AI Agents can follow the cycle and be well-performed in each step to push the frontiers of R&D. The detailed definitions of the three steps are as follows.

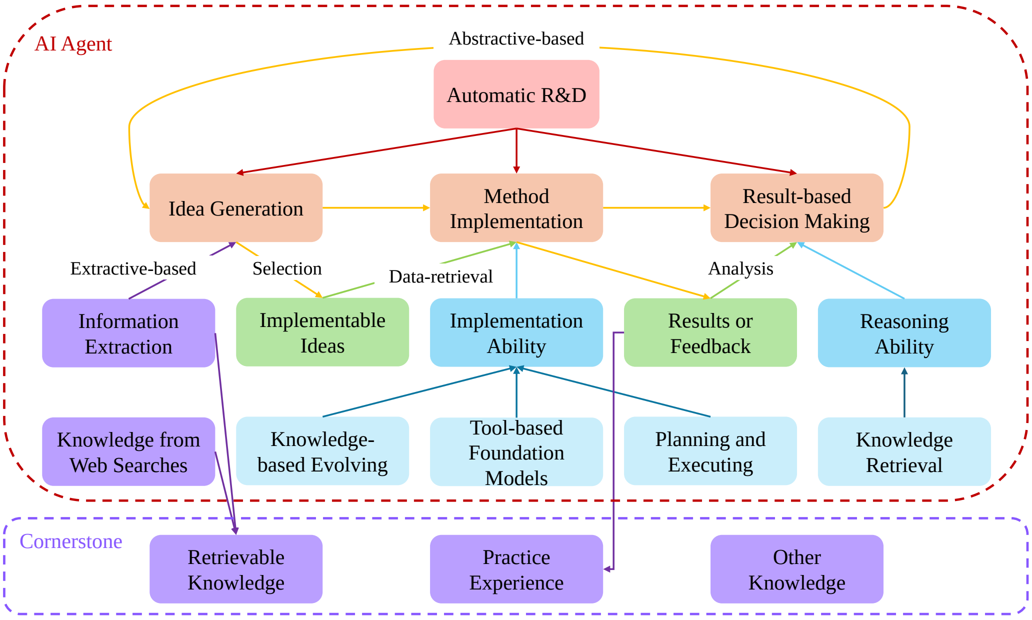

Figure 1: The overview of my research proposal in automatic R&D.

1. Idea Generation: **I consider all the ideas to be generated in two ways: extractive or abstractive**. The taxonomy is similar to that in text generation methods. **Extractive-based idea generation** means that we can find ideas from the raw data (e.g., papers, websites, reports, etc.) by information extraction: All the ideas are recorded or mentioned in the raw data. What AI agents should do is to distinguish and then extract them (sometimes even complete/create ideas by "hallucination"). **Abstractive-based idea generation** aims to find new ideas from the experimental results. AI agents will make decisions about what to implement in the next round for more "**informative**" experimental results. For example, the more

informative the results are, the greater information entropy or variance they will reduce. More details are introduced in Section 1.

2. Method Implementation: **I regard "methods" as the "implementable" ideas expressed by formulations and models**. To successfully implement these methods, AI agents should first accurately find the corresponding data according to the given data introduction or background description. Second, they use their **implementation ability** to obtain the results of these methods. Note that not all experimental attempts will be successful, so sometimes they get feedback from the environment (e.g., the error reported by Python or code evaluator) instead of final results. More details are introduced in Section 3.

3. Result-based Decision Making: After obtaining all the results or feedback, **AI Agents will decide what to implement (come up with some ideas) in the next round (i.e., an attempt) or next several rounds (i.e., a plan)**. They will use their **reasoning ability** to find clues/logic/evidence information that can reveal what to do next.

A2. As I mentioned in A1, AI Agents have to possess many strong abilities to tackle automatic R&D elegantly because the task is so difficult. Given that humans have accumulated thousands of years of experience (knowledge) in this task, equipping AI agents with knowledge ideally improves their reasoning ability, language understanding ability, and implementation ability. Therefore, we need to collect and govern the valuable knowledge. Details are introduced in Section 2.

A3. Current SOTA solutions focus on pretraining (Tool-based LLMs) or planning (e.g., CoT, React, etc.), neglecting to 1) collect the interaction results (experience) with environments in the past attempts; 2) retrieve the corresponding knowledge in the current step to make a step further towards the correct answer instead of giving up the current dialogue; 3) retrieve knowledge to customize the prompt in each interaction. **Our proposed knowledge-based evolving strategy depicts a promising framework to unify and extend the logic of current solutions**. *This work is in preparation.*

1 Idea Generation: Reasoning Ability

First, I give my definition and evaluation of the ideas within the scope of my research. Second, I elaborate on our current research in both extractive and abstractive idea generation.

Definition and Evaluation of Ideas. Since a mature idea should reveal potential solutions or improve existing solutions to a problem, it should satisfy the following requirements corresponding to its content and contribution, respectively.

1. **Content**. I consider two aspects for ideas' content evaluation, namely "**Soundness**" and "**Vision**". Specifically, the generated information of a certain idea should be comprehensive and rigorous enough (soundness) for further implementation. This aspect is often represented by the Methodology section of a research paper. Also, the proposed ideas should be well-rounded and balanced. This aspect is usually reflected by the vision (expertise) of a researcher: they offer their own insights and solutions across different scenarios.

2. **Contribution**. I consider two aspects for ideas' contribution evaluation: "**Novelty**" and "**Effectiveness**". Specifically, reviewers evaluate the novelty of ideas based on their knowledge, which is represented by the volume of related work they have read and the amount of information they have gathered from presentations or discussions. Besides, ideas should work or be effective in some scenarios (e.g., real-world scenarios), which is often demonstrated and discussed in the Experimental section of a research paper.

Note that immature (e.g., not implementable) ideas can be valuable in some conditions. They usually require continually evolving. We depict the gap between immature and mature ideas as "**Condition**". For example, immature ideas include **issue discovery** (data insights such as spurious correlations), **hypotheses** (a plan that needs feedback), **analysis of experimental results**, and so on. Evolution bridges the gap and makes immature ideas satisfy the Conditions. **To sum up, 3C standards accurately pinpoint the quality of a given idea.** I conduct the research by formalizing the evaluation metrics, constructing a benchmark, and analyzing the bottleneck and potential solutions. *This work is in preparation.*

Extractive-based Idea Generation. I consider the scenario where the relevant information of a certain implementable idea is scattered across the given documents. Extractive-based idea generation aims to aggregate all the relevant information and then output it according to the requirements, including unstructured and structured textual output. To this end, I conduct research in multi-document summarization, proposing aspect-supervised text generation to recover the missing aspect information of ideas in summaries [6, 4]. Meanwhile, I formalize the task as controllable text summarization and tailor learning models toward generating idea-specific summaries [7]. The proposed entity-oriented generator (EoG) surpasses the previous SOTA models and opens a new paradigm in information extraction.
Abstractive-based Idea Generation. Abstractive-based Idea Generation aims to form new ideas that can't be extracted from the existing raw materials. Most of these ideas are in the "Condition" state: they are immature and require evolution in order to be implementable to satisfy the "Content" and "Contribution" requirements. Therefore, **following the generating process of human-proposed good ideas, I base the development of immature ideas on multiple rounds of evolving and available supporting knowledge**. Specifically, the available knowledge includes the ideas generated by extractive-based methods, available information on the web, manually given expert knowledge, existing domain knowledge base, and feedback from interactions with environments (e.g., experimental results, python error information, etc.). With the help of knowledge, AI agents' reasoning ability gets improved and thus generates mature ideas gradually. *Both the demo and this work are in preparation.*

2 Knowledge Collection & Retrieval: Language Understanding Ability

We can observe from the previous sections that knowledge plays a vital role in improving the reasoning ability of AI agents. **knowledge is the cornerstone of idea generation, idea implementation, and even knowledge collection**. Here are two crucial questions about using knowledge to empower AI agents: 1) How to possess up-to-date, comprehensive, and accurate knowledge? 2) How to retrieve the "useful" knowledge according to the scenario? The solutions to the two questions share a common background: **improving the language understanding ability of AI**.
Knowledge Collection. The sources of knowledge required in automatic R&D can be roughly divided into three kinds: (1) **Retrievable knowledge**. We can get retrievable knowledge directly from the publicly available information. I give two examples here. One is the publicly available knowledge graph. The other is knowledge from web searches. Many LLMs perform well in collecting the two kinds of knowledge. (2) **Practice Experience**. I think an example of the typical practice experience is the feedbacks from environments: Different types of errors raised by python or reported in the comparison with ground-truth results. (3) **Other Knowledge**. There's also some knowledge that can neither be retrieved directly from raw information nor practice experience, such as private data, industry-specific knowledge graph, or the deep insights and understandings formed by reading hundreds of papers. Such kind of knowledge need to be manually instilled into AI agents. **Previous work in collecting structured knowledge (e.g., relation extraction) suffers from a fatal pitfall: making predictions according to spurious patterns instead of rationales**. I first conduct a comprehensive investigation by feature attribution to visualize the decision rules of existing SOTA long-text relation extraction (RE) methods and then demonstrate that their biased decision rules significantly weaken their robustness, generalization ability, and trustworthiness [3]. By proposing a conceptual framework based on causal inference, I appeal to future work to rethink the whole life cycle of developing LLM, including data governance, instilling human knowledge (causality) into model structure, and constructing causality-aware test data based on attacks [1]. In the RE task, specifically, I conduct research on benchmarking the robustness, out-of-distribution generalization ability, and language understanding ability of the SOTA long-text RE models, which brings new insights into long-text relation extraction [2, 3]. Meanwhile, I theoretically and experimentally demonstrate that the redundant information degrades the performance of relation extraction methods and impedes them from learning causality [7]. Based on new insights, my proposed methods adopt text summarization methods to filter redundant information and instill human knowledge into the model structure (e.g., transformer), which significantly improves their causality-learning ability [1].
Knowledge Retrieval. We could retrieve either structured knowledge (e.g., knowledge graph) or unstructured knowledge (e.g., ideas and their feedback, domain-specific experience and insights, issues and their solutions,

knowledge from AI Agent). Currently, we embed all kinds of easily collected knowledge (e.g., types of errors, human-crafted tools, etc.) into a graph consisting of nodes and their attributes. We retrieve the relevant knowledge by calculating the distance between the embeddings (generated by LLMs) of graph nodes and the historical dialogues. The retrieved knowledge forms new information to improve the answer of LLMs. However, the retrieval method sometimes introduces redundant information (irrelevant knowledge) which may degrade LLMs' performance. The issue can be solved by fine-grained evidence information. *This work is in preparation.*

3 The Broader Action Space: Implementation Ability

The data-driven R&D cycle can't be automated without the strong implementation ability of AI agents even if the generated idea performs well on 2C (Content and Contribution) standards. The prerequisite for developing the implementation ability of agents is to construct an accurate and comprehensive benchmark for their real-world deployment.

We serve as the first effort to benchmark the SOTA LLMs and upcoming LLM agents in data-driven automatic R&D [5]. We build the code framework to realize the overall automatic R&D cycle and propose a benchmark, which helps us foresee the challenges and bring new insights. Our ground-breaking work reveals the promising future of LLM agents and sheds some light on the future work in data-driven automatic R&D. Based on this work, we exploit one of the effective technical routes: knowledge-based evolving strategy. The current experimental results have demonstrated its significant effectiveness. *This work is in preparation.* With the further enrichment of knowledge, more effective methods (the combination of the next generation of knowledge collection, knowledge retrieval, and evolving) will be developed.

4 Why Does ML Systems Work or Fail: Conceptual Framework

I'm interested in conceptual framework induction and development to deepen our insights and understanding of AI. My research experience includes rethinking the whole developing life cycle of LLMs from the perspective of causal inference [1]. We find that most of LLMs fail to learn the causality from correlation, which disables them from learning rationales for predicting. To solve the issue, we first explain the underlying theoretical mechanism of their failure and argue that both the data imbalance and the omission of causality in model design and selection render the current training-testing paradigm failed to select the unique causality-based model from correlation-based models. Second, we take the legal text prediction task as the testbed and reconstruct the developing process of LLMs by simultaneously infusing causality into model architectures and organizing causality-based adversarial attacks for evaluation.

5 Future Research Agenda

In future, my research agenda includes: (1) complete all the work in preparation. (2) proposing R&D-specific pre-trained LLM if current LLMs touch the boundary of their abilities in automatic R&D. (3) Exploiting the potency of SOTA LLMs for extractive-based idea generation (especially mitigating their hallucination issue). (4) Exploiting the potency of SOTA LLMs for knowledge collection: Current LLMs are poor of generating the structured knowledge according to the requirements of users. (5) Proposing more conceptual frameworks based on the findings from automatic R&D research. Currently, we base the explanation of why ML systems work or fail on causal analysis. We can also develop other mathematical conceptual frameworks to explain the reason why prompt engineering is crucial (based on information theory) and why R&D benchmark selects the more trustworthy methods (based on probability inference). (6) Proposing pre-trained embodied AI agent. After developing LLM-based R&D agent into a real-world deployable level, we can consider developing **the embodied R&D AI agents** by changing RLHF into **reinforcement learning from world (environment) and human feedback**, which replicates the growth process of a child.

References

[1] **Haotian Chen**, Lingwei Zhang, Yiran Liu, Yang Yu. "Rethinking the Development of Large Language Models from the Causal Perspective: A Legal Text Prediction Case Study", *The 38th Annual AAAI Conference on Artificial Intelligence (AAAI 2024)*, Vancouver, Canada, Feb 2024.

[2] **Haotian Chen**, Houjing Guo, Bingsheng Chen, Xiangdong Zhou. "OODREB: Benchmarking State-of-the-Art Methods for Out-of-Distribution Generalization on Relation Extraction", to appear in *International World Wide Web Conference (WWW 2024)*.

[3] **Haotian Chen**, Bingsheng Chen, Xiangdong Zhou. "Did the Models Understand Documents? Benchmarking Models for Language Understanding in Document-Level Relation Extraction", *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, Toronto, Canada, Aug 2023.

[4] **Haotian Chen**, Han Zhang, Houjing Guo, Shuchang Yi, Bingsheng Chen, Xiangdong Zhou. SALAS: Supervised Aspect Learning Improves Abstractive Multi-document Summarization Through Aspect Information Loss, *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2023)*, Torino, Italy, Sep 2023.

[5] **Haotian Chen***, Xinjie Shen*, Zeqi Ye, Xiao Yang, Xu Yang, Weiqing Liu, Jiang Bian. RD2Bench: Toward Data-Centric Automatic R&D, *ICLR 2024 Workshop: How Far Are We From AGI*, Vienna, Austria, May 2024.

[6] **Haotian Chen**, Han Zhang, Houjing Guo, Shuchang Yi, Bingsheng Chen, Xiangdong Zhou. Recovering Missing Key Information: An Aspect-guided Generator for Abstractive Multi-document Summarization, *28th International Conference on Database Systems for Advanced Applications (DASFAA 2023)*.

[7] **Haotian Chen**, Bingsheng Chen, Xiangdong Zhou. A Novel Entity-Aware Generate-then-Extract Paradigm for Cross-Document Relation Extraction, *KDD 2024 Under Review*.