CHIMERA: A Knowledge Base of Idea Recombination in Scientific Literature

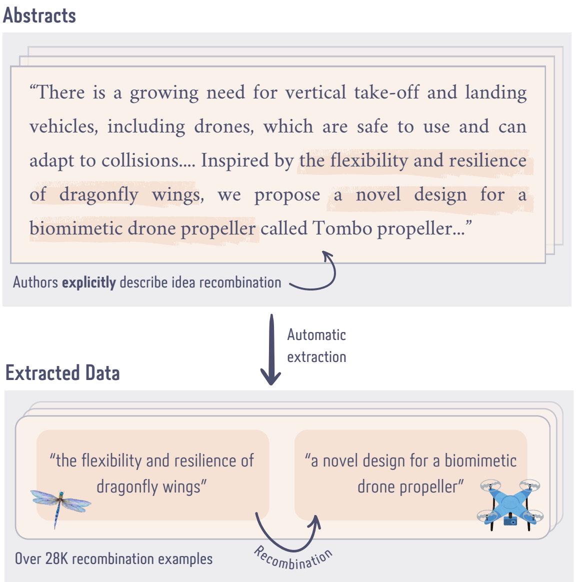
Anonymous ACL submission

Abstract

A hallmark of human innovation is the process of *recombination*—creating original ideas by integrating elements of existing mechanisms and concepts. In this work, we automatically mine the scientific literature and build `CHIMERA`: a large-scale knowledge base (KB) of recombination examples. `CHIMERA` can be used to empirically explore at scale how scientists recombine concepts and take inspirations from different areas, or to train supervised machine learning models that learn to predict new creative cross-domain directions. To build this KB, we present a novel information extraction task of extracting recombination from scientific paper abstracts, collect a high-quality corpus of hundreds of manually annotated abstracts, and use it to train an LLM-based extraction model. The model is applied to a large corpus of papers in the AI domain, yielding a KB of over 28K recombination examples. We analyze `CHIMERA` to explore the properties of recombination in different subareas of AI. Finally, we train a scientific hypothesis generation model using the KB, which predicts new recombination directions that real-world researchers find inspiring. Our data and code are available at `https://anonymous.4open.science/r/CHIMERA-0510`

1 Introduction

Recombination—creating original conceptual and physical combinations of existing mechanisms, methods, perspectives and approaches to address problems—is a common form of ideation (Uzzi et al., 2013; Youn et al., 2015; Shi and Evans, 2023). Recombination involves re-representing past ideas by decomposing them into conceptual chunks and then blending them into new solutions (Knoblich et al., 1999; McCaffrey, 2012), and also often involves forming abstract structural connections across domains (Gentner et al., 1997; Gentner and Markman, 1997; Gentner and Kurtz, 2005; Chan et al., 2011; Frich et al., 2019)—for example, nature-inspired optimization algorithms that recombine concepts from nature and optimization.



“There is a growing need for vertical take-off and landing vehicles, including drones, which are safe to use and can adapt to collisions.... Inspired by the flexibility and resilience of dragonfly wings, we propose a novel design for a biomimetic drone propeller called Tombo propeller...”

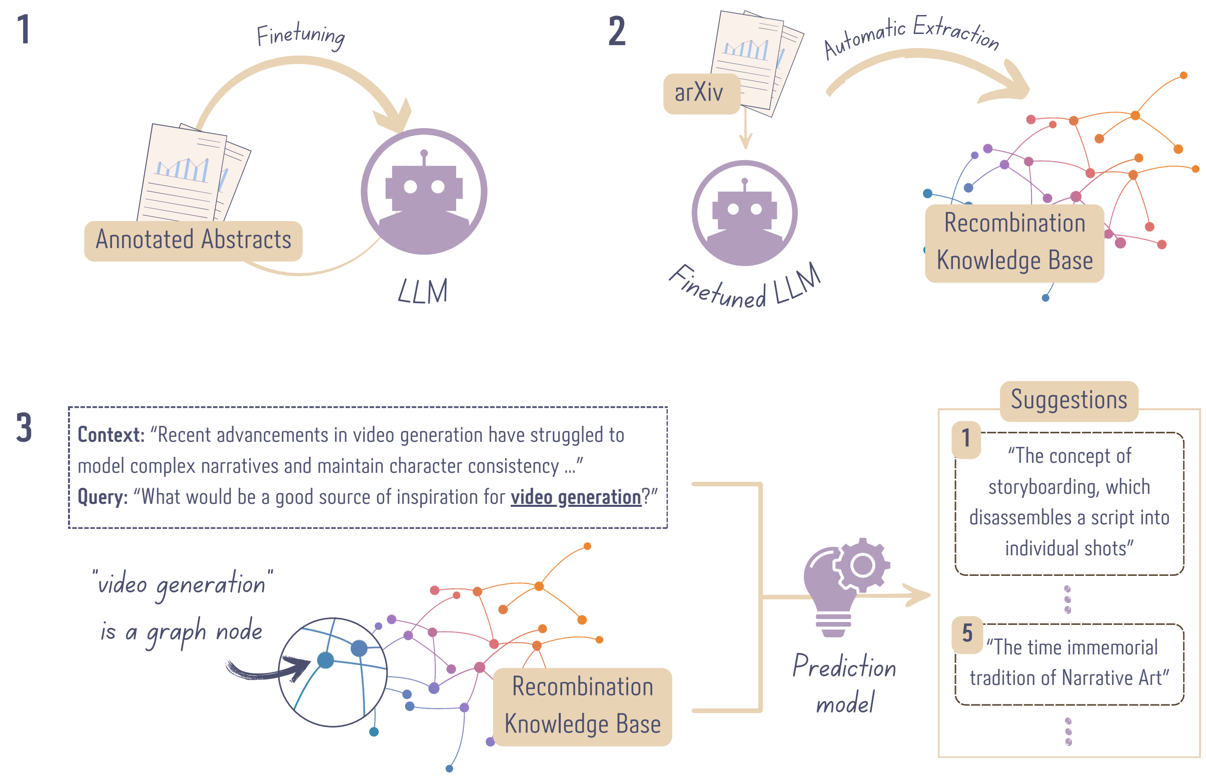
“the flexibility and resilience of dragonfly wings”

“a novel design for a biomimetic drone propeller”

Figure 1: We automatically extract thousands of recombination examples describing how scientists connect ideas in novel ways.

In this work, we automatically mine `CHIMERA`, a knowledge base of recombination examples from across the scientific literature. We focus on two recombination types, which we name *blends* and *inspirations*. Blends combine multiple concepts to create new approaches (e.g., boosting classical machine learning algorithms using quantum computing), while inspirations involve adaption of ideas from existing concepts to spark insight (e.g., applying bird flock behavior to coordinate drone swarms).

`CHIMERA` includes examples of blends of concepts within and across domains, and also inspirations in the form of analogies, reductions, and abstractions. Unlike simpler concept co-occurrence approaches (Krenn et al., 2022) or more generic



Context: "Recent advancements in video generation have struggled to model complex narratives and maintain character consistency ..."
Query: "What would be a good source of inspiration for **video generation**?"

1 "The concept of storyboarding, which disassembles a script into individual shots"

5 "The time immemorial tradition of Narrative Art"

Figure 2: 1) We start by collecting human-annotated recombination examples and use them to finetune an LLM for information extraction. 2) Next, we apply the fine-tuned LLM on the arXiv corpus and build a knowledge base of recombination examples. 3) Given a context string and a query concerning the recombination of a certain graph node, our recombination model suggests directions based on knowledge learned from the KB.

scientific extraction schema (Luan et al., 2018), `CHIMERA` contains examples in which the authors *explicitly* mention a recombination as one of the core contributions of their work. Figure 1 presents a recombination example automatically extracted using our methods.

Our methods and collected data have broad potential applications in the field of Science of Science (Shi and Evans, 2019). The collected data enables empirical studies of innovation in novel ways—for example, examining how fields draw inspiration from one another and investigating how the blending of mechanisms across different domains emerges and evolves. Using our knowledge base, we present an analysis of sources of inspiration in the AI and NLP communities. We also demonstrate how to use this data to build a supervised learning framework for recombinant ideation. While existing work in the human-computer interaction (HCI) community provides researchers with tools for exploring idea recombinations (Radensky et al., 2024a; Kang et al., 2022), `CHIMERA` allows us to take a different approach: training supervised models that *learn* from past examples how to recombine concepts for generating new scientific ideas. Figure 2 presents an overview of this process.

2 Related Work

Recombinant creativity Concept blending and finding inspiration through analogies are a key way to create new ideas (McKeown, 2014; 201, 2019; Holyoak and Thagard, 1994). Recent research explores how idea recombination can enhance LLM-powered ideation tools. For example, `CreativeConnect` (Choi et al., 2023) enables users to recombine keywords to generate graphic sketches, while `Luminate` (Suh et al., 2023) facilitates the recombination of dimensional values to produce diverse LLM responses. `Scideator` (Radensky et al., 2024b) is another recent work that allows researchers to explore new ideas by interactively recombining scientific concepts. Researchers have also investigated combining elements from input and analogous artifacts to create new ideas (Srinivasan and Chan, 2024; Chilton et al., 2019). In our work, we aim to extract a knowledge base of real recombinations in scientific papers, which

Recombination extraction examples
Abstract: "...Current archaeology depends on trained experts to carry out bronze dating, which is time-consuming and labor-intensive. For such dating, in this study, we propose a learning-based approach to integrate advanced deep learning techniques and archaeological knowledge..." **Blend**: "*advanced deep learning techniques*" ⟷ "*archaeological knowledge*"
Abstract: "...Traditional approaches to enhance dialogue planning in LLMs, ... either face efficiency issues or deliver suboptimal performance. Inspired by the dual-process theory in psychology, which identifies two distinct modes of thinking - intuitive (fast) and analytical (slow), we propose the Dual-Process Dialogue Planning (DPDP) framework..." **Inspiration**: "*the dual-process theory in psychology*" ⟶ "*enhance dialogue planning in LLMs*".

Table 1: Example blend and inspiration. Note that blend is a symmetric relation, while inspiration is not.

can be used to facilitate research on recombination. For example, we train models that *learn* from past examples of recombination how to predict new recombinations (Figure 2).

Scientific information extraction Much prior work has focused on information extraction (IE) from scientific papers. A notable example is the `SciERC` dataset (Luan et al., 2018), which includes annotations of scientific entities (e.g., methods, tasks, metrics) and relations (e.g., conjunction) for 500 abstracts. Other, more recent examples are `SciREX` (Jain et al., 2020), and `SciDMTAL` (Pan et al., 2024), introducing document-level scientific IE datasets covering similar entity types. However, existing extraction approaches do not focus on recombination relationships, as we demonstrate in Appendix L, Figure 18. In our work we design a simple information extraction schema that is focused on idea recombination, to enable exploration of this important form of scientific innovation. For example, as part of our knowledge base we are able to extract many examples of analogical inspirations used by the AI community (Figure 1), which was not possible using existing scientific IE schema.

3 Extracting Recombinations

3.1 Problem Definition

We focus on cases where paper abstracts clearly say that a contributed idea is rooted in an original combination of concepts or explicitly mention a source of inspiration. As discussed in the introduction, we define two coarse-grained relation types to capture the notion of recombination: blend and inspiration. **Blend** refers to joining multiple concepts (e.g., methods, models, theories). Note that we use the terms concept blend and concept combination interchangeably. **Inspiration** refers to using knowledge from a source entity and implementing it in a different target entity. This could involve, for example, using an analogy or an abstraction as a source of inspiration, or more generally being influenced by another line of work. Relations are defined between free-form spans of text representing scientific concepts (see Figure 1, and additional examples in Table 1). We refer to the entities in a blend relation as *combination-elements* and those in an inspiration relation as *inspiration-source* and *inspiration-target*.

3.2 Recombination Mining

Our approach to mining recombination examples begins with building a curated dataset of annotated examples. We then use this dataset to train an information extraction model. Finally, we apply the trained model to collect recombination examples at scale. This process is illustrated in Figure 2.

Data sourcing We use AI-related papers from the unarXive corpus (Saier and Färber, 2020) as a source of annotation examples[1]. The data undergo an initial keyword-based filtering to identify works that are more likely to specify idea recombination. Table 7 in Appendix A lists the keywords used in this process. We then assign the remaining abstracts to annotators.

Annotation process After a thorough screening session, we recruit two annotators with scientific

---

[1] We focus on the following arXiv categories: cs.AI, cs.CL, cs.CV, cs.CY, cs.HC, cs.IR, cs.LG, cs.RO, cs.SI

Example type	**# Train**	**# Test**	**# Total**
blend	124	76	100
inspiration	45	24	69
not-present	195	116	311
All	364	216	580

Table 2: Human-annotated corpus. We use examples with and without recombinations ("*not-present*"), simulating extraction over real-world data.

Category	**# Interdisciplinary**	**# Total**
Inspiration Edges	5,182 (54.1%)	9,578
Blend Edges	1,792 (9.6%)	18,586
Edges (total)	6,974 (24.8%)	28,164
Nodes (total)	n/a	43,393

Table 3: The `CHIMERA` knowledge base summary. Our knowledge base encompasses over 28K recombination examples, a quarter of which are interdisciplinary.

PhDs from Upwork[2]. The screening includes annotating examples following a detailed guidelines document[3]. Annotators who passed the screening had a personal one-hour training session to discuss additional examples. The annotators conduct the annotation through `LightTag` (Perry, 2021), a web-based text annotation platform (no longer operational). We collect a total of 580 annotated abstracts in this manner, as presented in Table 2.

To ensure annotation quality, we assign 10% of the examples to both annotators and review the shared section at the end of each batch. During the review, we discuss disagreements and provide the annotators with feedback, after which they revise their work. The annotations undergo an additional review by an NLP expert. The expert verifies correctness, adjusts span boundaries, and merges annotations from different annotators.

Automatic recombination mining We next use the collected data to fine-tune an LLM-based extraction model. Given the text and annotation schema, we instruct the model to extract the most salient recombination from the text, if one exists. The model must determine whether the text discusses recombination, infer its type, and identify entities in a single query.

We devise the test set from examples where at least two annotators (out of three) agree on the recombination type (or lack of recombination). This provides high-quality test data with fewer ambiguities. Table 2 presents the distribution of example types in both the train and test sets. Additional implementation details related to the extraction baselines are discussed in Appendix B.

3.3 The `CHIMERA` Knowledge Base

This section describes building the `CHIMERA` knowledge base. We first use our extraction method to mine recombination examples, categorize them, and build a KB in which nodes represent scientific concepts, and edges represent recombination relations between them.

Large-scale mining We source abstracts from the arXiv dataset[4]. This dataset updates monthly and includes newer examples than unarXive (Saier and Färber, 2020). We apply our fine-tuned extraction model over publications from 2019 to 2024 within the same CS categories used for the annotation task. After applying the model we filter predictions that fail to comply with the data schema or could not be properly parsed.

Categorization In addition to extracting the relations, we apply `GPT-4o` to identify the scientific domain of each extracted entity given the abstract. This enables analyses we perform in Section 4.2. The model is instructed to select the most appropriate arXiv category from either the arXiv taxonomy list or a supplementary list of non-arXiv domains (e.g., "*psychology*"). In cases where no given label captures the entity's scientific taxonomy, we classify the node's domain as `Other`. The analysis in Section 4.2 omits nodes from this domain, as we find they might be too noisy, too general or miscellaneous. Examples of such nodes, along with the used prompts and additional technical information about this step, are available in Appendix D.

We further assign the graph nodes a higher-level scientific discipline. The discipline is either the arXiv group name if available ("*computer-science*" for cs.AI), or the relevant non-arXiv domain.

KB building We normalize the knowledge base entities by clustering semantically similar ones, and further enrich each edge in the graph with the publication date and arXiv categories of the paper citing it. For simplicity, we focus on binary relations

---

[2] https://www.upwork.com/

[3] https://tinyurl.com/4mfdrx2f

[4] https://tinyurl.com/mrzksbky

Task	Baseline	P	R	F1
***Abstract classification**:* *Does it discusses recombination?*	Human-agreement	0.757	**0.765**	0.760
	E2E$_{Mistral-7B-Instruct-v0.3}$	**0.815**	0.762	**0.763**
	E2E$_{Llama-3.1-8B-Instruct}$	0.630	0.628	0.620
	E2E$_{GoLLIE-13B}$	0.677	0.667	0.667
	E2E$_{GPT-4o}$	0.720	0.580	0.572
	Abstract-classifier$_{Mistral-7B-Instruct-v0.3}$	0.622	0.607	0.602
	Abstract-classifier-CoT$_{Mistral-7B-Instruct-v0.3}$	0.774	0.748	0.749
***Entity extraction**:* *What are the relevant entities?*	Human-agreement	**0.876**	**0.591**	**0.675**
	E2E$_{Mistral-7B-Instruct-v0.3}$	0.587	0.352	0.440
	E2E$_{Llama-3.1-8B-Instruct}$	0.249	0.259	0.252
	E2E$_{GoLLIE-13B}$	0.259	0.187	0.217
	E2E$_{GPT-4o}$	0.138	0.293	0.217
	Entity-extractor$_{GPT-4o}$	0.268	0.263	0.247
	Entity-extractor$_{SciBERT}$	0.324	0.248	0.276
	Entity-extractor$_{PURE_{SciBERT}}$	0.187	0.536	0.271
***Relation extraction**:* *What is the recombination?*	Human-agreement	**0.805**	**0.581**	**0.651**
	E2E$_{Mistral-7B-Instruct-v0.3}$	0.598	0.366	0.454
	E2E$_{Llama-3.1-8B-Instruct}$	0.264	0.294	0.276
	E2E$_{GoLLIE-13B}$	0.301	0.219	0.253
	E2E$_{ICL-GPT-4o}$	0.223	0.385	0.244

Table 4: Recombination extraction results. **Bold** text signifies the best result, while underlined text signifies the second-best. We observe that surprisingly large and capable models struggle with the extraction tasks.

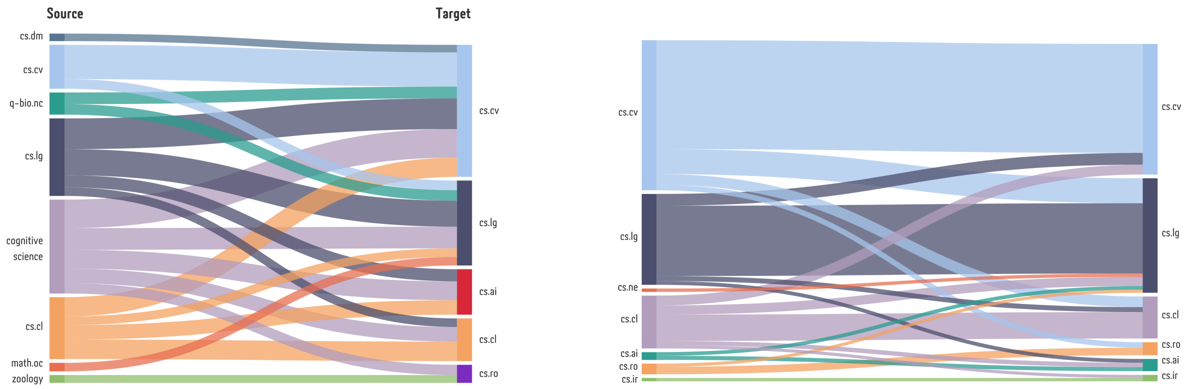
when building the graph. Table 3 reports the summary of the final KB, including the number of interdisciplinary blends and inspirations.

4 Results

4.1 Recombination extraction evaluation

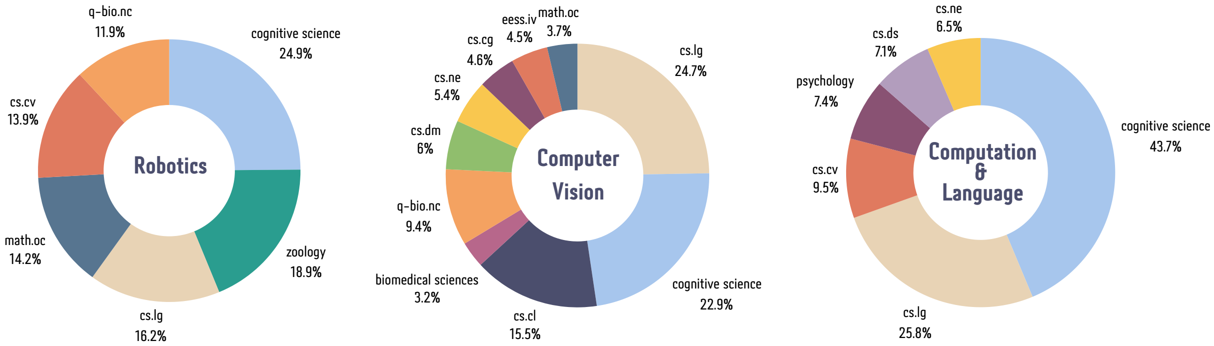
We evaluate the recombination extraction process in three different levels of increasing difficulty: **abstract classification** (whether the text discusses recombination), **entity extraction** (what entities are described in the text) and **relation extraction** (what the discussed relation is). To evaluate abstract classification, we use precision, recall and F1. We use a soft evaluation metric for entity and relation extraction tasks, where two entities of the same type match if they refer to a semantically similar concept. We utilize `GPT-4o-mini` as a judge of content similarity (Figure 12 in Appendix E presents our prompt). We select `GPT-4o-mini` over `GPT-4o` after conducting a qualitative examination and finding only a handful of cases in which the model judgment differs (3 span pairs in the entire test set). To avoid position bias, we run the judge twice for each pair, reversing the span order each time. We consider two spans equivalent only when the judge returns a positive answer on both runs. Each predicted entity can match with exactly one gold entity and vice versa, with any additional matches being ignored. Under this definition of soft entity matching, we compute the entity extraction quality of a model using standard precision, recall and F1. For recombination relations, we apply the same metrics using partial relation matching: a predicted relation's contribution to the true-positive score depends on how many entities match with a ground-truth relation of the same type.

Extraction results Table 4 reports results for different levels of recombination extraction (abstract classification, entity extraction and relation extraction). We experiment with end-to-end (E2E) extraction approaches, inferring whether the text discusses recombination, its type, and what entities are involved - all at once. Note that when handling E2E approaches, we regard any extracted relation as a positive abstract classification. We also study models specialized in classification (Abstract-classifiers) or entity extraction (Entity-extractors). Appendix B describes implementation details re-



(a) Frequent inspiration source and target domains.

(b) Frequent domains of blend nodes.



(c) Common sources of inspiration for leading domains.

Figure 3: Recombinations between areas. cs.*, q-bio.nc and math.oc are arXiv categories. Inspirational connections are often cross-domain (Figure 3a), while blends more often stay within-domain (Figure 3b). Figure 3c zooms in on a few domains, for example revealing that robotics often draws inspiration from zoology).

garding extraction.

Human agreement has F1 scores of 0.760, 0.675, and 0.651 for classification, entity extraction, and relation extraction, respectively. These values are comparable with other works measuring soft annotators' agreement in complex extraction tasks (Naik et al., 2023; Sharif et al., 2024).

Generally, fine-tuning `Mistral-7B` using our data obtains the best performance in all subtasks. We observe that entity and relation extraction appear more challenging than classification for both humans and SOTA LLMs. However, humans still significantly outperform automatic extraction approaches. We present error analysis in Appendix F. Our findings indicate that focusing on a smaller portion of the recombination extraction task is not necessarily easier than performing it end-to-end, as seen in the lower performance of abstract classifiers, and discuss this point further in Appendix C.

4.2 Knowledge base analysis

Blends vs. inspirations Figures 3a and 3a present the predominant domain pairs for inspiration and blend relations in `CHIMERA` (with frequency above the 0.9 quantile). Inspirations display a larger selection of domains than blends. We also observe that blends connect the same or similar domains, while inspirations are often between different and further domains. Of note is the volume of inspiration drawn from brain-related sources, such as *cognitive science* and *q-bio.nc*. A possible explanation might be that many of our arXiv categories of interest are related to machine learning, where the human brain historically serves as a general source of inspiration. Table 14 in Appendix H presents the same information in tabular format for better readability.

Inspiration analysis We next analyze how different fields draw inspiration from each other. Figure 3c shows the top 10% cross-domain inspiration sources for three prevalent domains in the graph: **cs.RO** (Robotics), **cs.CV** (Computer Vi-

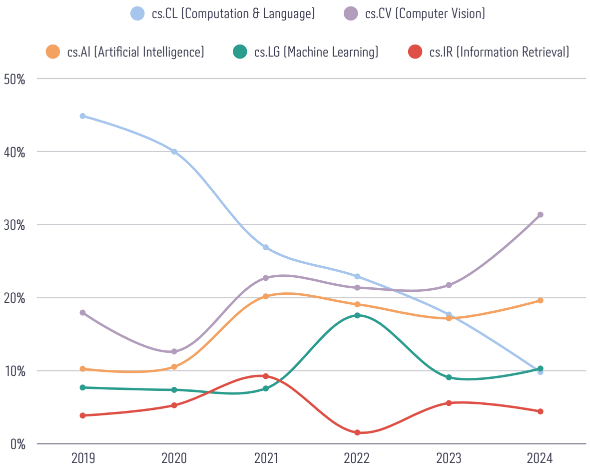


Figure 4: Prevalent domains inspired by cs.CL concepts (NLP). Note the decrease in within-domain inspiration.

sion) and **cs.CL** (Computation and Language). We observe that while some sources of inspiration (like *cognitive-science*) are commonly shared across related fields, domains may draw inspiration from unique sources (e.g., from *zoology* to *cs.RO*). Examples of these interdisciplinary inspirations can be found in Appendix G, Table 13.

Figure 4 shows the percentage of target nodes in domains drawing inspiration from cs.CL (Computation and Language) over the past five years. We observe two trends: a decrease in intra-domain inspiration (where cs.CL concepts inspire other cs.CL concepts), and an increase in cs.CV (Computer Vision) concepts drawing inspiration from cs.CL.

5 Recombination Prediction

This section gives an example of a possible use case of the `CHIMERA` knowledge base. Using this data, we train supervised models that learn how to recombine concepts for predicting new scientific ideas. This process is illustrated in Figure 2.

5.1 Problem Definition

Split	# Inspiration	# Blend	# Total
Train	5,408	19,909	25,317
Validation	119	411	530
Test	2,026	8,591	10,617

Table 5: Prediction data splits. We divide our data by the publication years associated with each query (training and validation sets $<$ 2024, test set $\geq$ 2024) to avoid contamination.

Figure 2 shows an example of the inputs and outputs of this task: given a context string ("*Recent advancements in video generation have struggled to model complex narratives...*") and a query regarding the recombination of a graph entity ("*What would be a good source of **inspiration** for **video generation**?*") we aim to answer this question with an additional graph entity (*"The concept of storyboarding..."*). More formally, given a context string (e.g., a problem, experimental settings, goals), an entity e and a recombination type τ, we aim to find a different graph entity e' such that (e, τ, e') is an edge in `CHIMERA`.

5.2 Recombination prediction

Data preparation We start by converting edges to pairs of queries and answers. The queries describe the task inputs: a single graph node, the edge recombination type, and a context string, which we extract from the corresponding abstract using `GPT-4o-mini`. Note that this process might leak information regarding the answer (the other graph node) into the query. Therefore, we follow it by applying `GPT-4o-mini` again to the result to identify leakages in the generated queries (see examples and additional implementation details for this step in Appendix I). We discard bad query-answer pairs (approximately 22% of the pairs, mostly due to leaks) and divide the remaining pairs into data splits based on the publication year. Our test set includes all pairs associated with papers published after 2024. Table 5 shows a summary of the resulting data splits.

Prediction We experiment with zero-shot and finetuned retrievers based on encoders trained before the test set cutoff year (2024). We next explore applying a `GPT-4o`-based reranker (Sun et al., 2023) to the top 20 retrieved results to improve our predictions further. The GPT-4o data cutoff is October 2023, meaning the reranker is also unfamiliar with our test set. Appendix J provides additional implementation details for the prediction baselines.

5.3 Prediction results

We present our results in Table 6. We observe that fine-tuning greatly helps to improve retrievers, decreasing the median rank of the gold answer by one order of magnitude. The last row of Table 6 shows results for applying RankGPT (Sun et al., 2023) with `GPT-4o` over the top-20 results of the best-performing retriever (all-mpnet-base-v2$_{finetuned}$).

Baseline	**H@3**	**H@5**	**H@10**	**H@50**	**H@100**	**MRR**	**MedR**
all-mpnet-base-v2	0.033	0.042	0.061	0.126	0.170	0.033	1305
bge-large-en-v1.5	0.041	0.053	0.076	0.151	0.199	0.041	1135
e5-large-v2	0.024	0.033	0.050	0.113	0.155	0.026	1590
all-mpnet-base-v2$_{finetuned}$	**0.110**	**0.135**	0.178	**0.320**	**0.402**	**0.106**	**194**
bge-large-en-v1.5$_{finetuned}$	0.104	0.130	0.168	0.306	0.392	0.102	222
e5-large-v2$_{finetuned}$	0.107	0.133	0.173	0.317	0.397	0.103	212
all-mpnet-base-v2$_{finetuned}$ + RankGPT	0.100	0.130	**0.192**	**0.320**	**0.402**	0.097	**194**

Table 6: Recombination prediction results. **MedR** stands for "Median Rank". Using `CHIMERA` data to fine-tune the models improves the median rank by a factor of 10. Reranking the top-20 answers using RankGPT boosts the H@10 but slightly reduces other metrics (H@3,5 and MRR).

While the reranker improves H@10, it degrades H@k for $k = 3, 5$ and MRR values. We hypothesize the reranker might lower the gold answer rank if there are many possible answers.

User study We present a human evaluation study exploring our ideation approach compared to other baselines. We recruit three volunteers with verifiable research experience (the authors of at least one research paper) and assign them examples based on their area of expertise. The examples are inspiration queries from our test set, along with inspiration source suggestions from different baselines: (1) **Ours**: our method, including reranking (2) **Gold**: the gold answer, (3) **Random**: a random test-set node, (4) **GPT-4o**: a suggestion generated by GPT-4o, (5) **ZS-`CHIMERA`**: a zero-shot prediction model using our test nodes as candidates, and (6) **ZS-SciERC**: a zero-shot prediction model using candidates extracted from test set abstracts with the SciERC (Luan et al., 2018) schema. Note that we use the highest ranked answer (k=1) for baselines returning a ranked list of candidates.

We request the annotators to rank baseline suggestions based on their *helpfulness* in inspiring interesting ideas. Figure 5 presents each baseline's median and average rank over 70 annotated examples. Since the most helpful suggestions are ranked first, a lower median and average rank signifies a more helpful baseline. Our approach receives a similar median and average rank as the gold answer, and annotators prefer it to all other baselines. This gives as an additional, complementary signal to the automatic evaluation, showing that our recombination prediction approach learns to create helpful recombinations. Appendix K presents additional details regarding the task and interface we used to conduct the study.

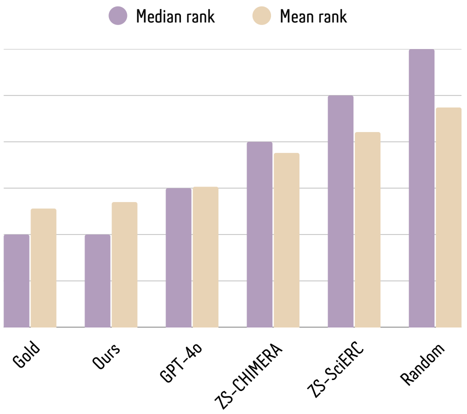


Figure 5: Researchers find our recombination suggestions almost as helpful as the gold answer in inspiring ideas, providing additional verification of our automated evaluation.

Conclusions

We automatically mine the scientific literature to create `CHIMERA`, a novel knowledge base spanning over 28K examples of how scientists blend concepts and draw inspiration from different areas. This knowledge base offers broad applications, and we demonstrate how it can be used to empirically study idea recombination across domains and to fine-tune models that predict new recombination directions that researchers find inspiring.

Limitations

Extraction quality While our information extraction model improves the quality of mined recombinations, it is still far from being perfect. Our qualitative error analysis shows the extraction model struggles to identify and extract more subtle recombination descriptions, and it still falls significantly short of human performance on the same task. Improving the extraction model remains a challenging and interesting direction for future work.

Recombination prediction evaluation One particularly challenging aspect of the recombination prediction task is the lack of a single correct response. Given a problem description, there are numerous ways to blend ideas and take inspiration that can lead to a novel, recombinant solution. This could lead to many false negatives, resulting in an overly pessimistic estimate of the models. We partially handle this through an additional human evaluation. However, given the high level of expertise required from the participants, the scope and thoroughness of such an evaluation are limited.

Experimenting with additional models We use models from the GPT-4o series for evaluation (judging entity span similarity), analysis (identifying entity's scientific domain), and to enrich our data (generating a context string for the extracted recombinations). Since this work prioritises the extraction and prediction of recombinations, we only experiment with those models. We leave experimenting with a larger range of models for these tasks for future work.

Ethics Statement

To collect human-annotated recombination examples, we recruited crowdworkers via Upwork. Annotators were informed of the nature and purpose of the task and were compensated at an hourly rate of $26–$30. Additionally, three volunteers participated in our human study. No personal information about the annotators or volunteers is disclosed.

To promote transparency and reproducibility, we release our code and model checkpoints. The collected data is shared under an open license to facilitate further research. We used AI assistants for grammatical corrections and code writing (e.g., GitHub Copilot).

References

2019. The cambridge handbook of creativity.

Joel Chan, Katherine Fu, Christian Schunn, Jonathan Cagan, Kristin Wood, and Kenneth Kotovsky. 2011. On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of mechanical design*, 133(8):081004.

Lydia B. Chilton, S. Petridis, and Maneesh Agrawala. 2019. Visiblends: A flexible workflow for visual blends. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2023. Creativeconnect: Supporting reference recombination for graphic design ideation with generative ai. *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the landscape of creativity support tools in hci. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–18.

Dedre Gentner, Sarah Brem, Ron Ferguson, Philip Wolff, Arthur B Markman, and KD Forbus. 1997. Analogy and creativity in the works of johannes kepler. *Creative thought: An investigation of conceptual structures and processes*, pages 403–459.

Dedre Gentner and Kenneth J. Kurtz. 2005. Relational Categories. In *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin*, APA decade of behavior series. American Psychological Association, Washington, DC, US.

Dedre Gentner and Arthur B Markman. 1997. Structure mapping in analogy and similarity. *American psychologist*, 52(1):45.

Moritz Hennen, Florian Babl, and Michaela Geierhos. 2024. Iter: Iterative transformer-based entity recognition and relation extraction. In *Conference on Empirical Methods in Natural Language Processing*.

Keith J. Holyoak and Paul Thagard. 1994. Mental leaps: Analogy in creative thought.

J. Edward Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *ArXiv*, abs/2106.09685.

Sarthak Jain, Madeleine van Zuylen, Hannaneh Hajishirzi, and Iz Beltagy. 2020. Scirex: A challenge dataset for document-level information extraction. In *Annual Meeting of the Association for Computational Linguistics*.

Hyeonsu B Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. 2022. Augmenting scientific creativity with an analogical search engine. *ACM Transactions on Computer-Human Interaction*, 29:1 – 36.

G. Knoblich, S. Ohlsson, H. Haider, and D. Rhenius. 1999. Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6):1534–1555. 00691.

Mario Krenn, Lorenzo Buffoni, Bruno Coutinho, Sagi Eppel, Jacob Gates Foster, Andrew Gritsevskiy, Harlin Lee, Yichao Lu, João P. Moutinho, Nima Sanjabi, Rishi Sonthalia, Ngoc M. Tran, Francisco Valente, Yangxinyu Xie, Rose Yu, and Michael Kopp. 2022. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. *Nature Machine Intelligence*, 5:1326–1335.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. *ArXiv*, abs/1808.09602.

T. McCaffrey. 2012. Innovation Relies on the Obscure: A Key to Overcoming the Classic Problem of Functional Fixedness. *Psychological Science*, 23(3):215–218. 00117.

Céline McKeown. 2014. The cognitive science of science: explanation, discovery, and conceptual change. *Ergonomics*, 57:632 – 633.

Aakanksha Naik, Bailey Kuehl, Erin Bransom, Doug Downey, and Tom Hope. 2023. Care: Extracting experimental findings from clinical literature. *ArXiv*, abs/2311.09736.

Huitong Pan, Qi Zhang, Cornelia Caragea, Eduard Constantin Dragut, and Longin Jan Latecki. 2024. Scidmt: A large-scale corpus for detecting scientific mentions. In *International Conference on Language Resources and Evaluation*.

Tal Perry. 2021. Lighttag: Text annotation platform. In *Conference on Empirical Methods in Natural Language Processing*.

Marissa Radensky, Simra Shahid, Raymond Fok, Pao Siangliulue, Tom Hope, and Daniel S. Weld. 2024a. Scideator: Human-llm scientific idea generation grounded in research-paper facet recombination. *ArXiv*, abs/2409.14634.

Marissa Radensky, Simra Shahid, Raymond Fok, Pao Siangliulue, Daniel S Weld, and Tom Hope. 2024b. Scideator: Human-llm scientific idea generation grounded in research-paper facet recombination. *arXiv preprint arXiv:2409.14634*.

Tarek Saier and Michael Färber. 2020. unarxive: a large scholarly data set with publications' full-text, annotated in-text citations, and links to metadata. *Scientometrics*, 125:3085 – 3108.

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier López de Lacalle, German Rigau, and Eneko Agirre. 2023. Gollie: Annotation guidelines improve zero-shot information-extraction. *ArXiv*, abs/2310.03668.

Omar Sharif, Joseph Gatto, Madhusudan Basak, and Sarah Masud Preum. 2024. Explicit, implicit, and scattered: Revisiting event extraction to capture complex arguments. In *Conference on Empirical Methods in Natural Language Processing*.

Feng Shi and James Evans. 2023. Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, 14(1):1641.

Feng Shi and James Allen Evans. 2019. Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, 14.

Arvind Srinivasan and Joel Chan. 2024. Improving selection of analogical inspirations through chunking and recombination. *Proceedings of the 16th Conference on Creativity & Cognition*.

Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2023. Luminate: Structured generation and exploration of design space with large language models for human-ai co-creation. *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

Weiwei Sun, Lingyong Yan, Xinyu Ma, Pengjie Ren, Dawei Yin, and Zhaochun Ren. 2023. Is chatgpt good at search? investigating large language models as re-ranking agent. *ArXiv*, abs/2304.09542.

You Can Teach, Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. 2020. You can teach an old dog new tricks! on training knowledge graph embeddings. In *International Conference on Learning Representations*.

Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical combinations and scientific impact. *Science*, 342(6157):468–472.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. 2022. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-pack: Packaged resources to advance general chinese embedding. *Preprint*, arXiv:2309.07597.

Hyejin Youn, Deborah Strumsky, Luis MA Bettencourt, and José Lobo. 2015. Invention as a combinatorial process: evidence from us patents. *Journal of the Royal Society interface*, 12(106):20150272.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In *North American Chapter of the Association for Computational Linguistics*.

Recombination keywords					
combines	analogies	aggregate	intermingle	unify	blending
combined	equivalence	aggregation	intermingling	unification	blends
combine	equivalent	align	join	weave	blend
combination	reduction	alignment	joining	weaving	blends
combinations	reframing	amalgamate	juxtapose	hybrid	merge
combining	reframe	amalgamation	juxtaposition	merge	merges
mixing	reformulating	assemble	link	merges	unites
mixture	casting	assembling	linkage	merging	analogy
mix	cast	associate	meld	merged	analogize
mixed	casts	association	melding	conflation	analogies
integrates	viewing	bond	mesh	couple	equivalence
integrating	viewed	bonding	meshing	unite	equivalent
integrate	view	bridge	perceive	unites	correlate
integrated	inspire	bridging	perception	interplay	correlation
connection	inspired	coalesce	relate	interconnect	envision
synergy	inspiration	coalescence	relation	harmonize	envisioning
fusion	inspires	compose	splice	harmony	harmonize
fuses	inspiring	composition	splicing	incorporate	harmony
unify	interconnect	incorporation	synthesis	reduction	synthesis
aggregate	align	inspiring	inspire	couple	conjunction
aggregation	reframing	inspiration	fuse	unite	conjoin
alignment	reframe	inspires	synthesis		

Table 7: Recombination keywords. We use a predefined list of keywords to identify works that are more likely to discuss idea recombination.

A Recombination keywords

We use keyword-based filtering to identify works that are more likely to discuss recombination before assigning papers to human annotators. Table 7 presents the list of keywords used for this step.

B Extraction baselines implementation

Annotators' agreement We calculate the annotators' agreement by treating one annotator's work as the ground truth and the other as predictions. In addition to entity-level and relation-level agreement, we also measure agreement on the recombination presence - whether the text presents a recombination, regardless of type. We apply the soft entity and relation matching described in 4.1 to measure the relation and entity agreement. The agreement is computed over the 52 documents shared by both annotators (approximately 10% of all annotated data). We regard the annotators' agreement as an approximation of human performance measurement over this task.

E2E recombination extraction We use `Mistral-7B` as the backbone for our recombination extraction baseline. We fine-tune the model using `mistral-finetune`[5] on a single NVIDIA RTX A6000 48GB GPU over 500 steps. The training was conducted using the default learning rate of $6.e - 5$ and weight decay of 0.1. We use a batch size of 1 and a maximum sequence length of 4096 tokens. `mistral-finetune` implements Low-Rank Adaptation of LLM (LoRA), a parameter efficient fine-tuning method (Hu et al., 2021), which we use with the default rank of 64. The evaluation uses the corresponding repository, mistral-inference[6]. We rerun the same experiment using `Llama-3.1-8B` as a backbone, using an additional 500 warm-up steps, a learning rate of $2e - 5$ and a weight decay of 0.01. Figure 6 presents the prompt for these experiments.

In addition to fine-tuning LLMs on our data, we experiment with `GoLLIE` (Sainz et al., 2023), a general IE model fine-tuned to follow any annotation guidelines in a zero-shot fashion. We apply `GollIE-13B` on our data, using a single NVIDIA

---

[5] https://github.com/mistralai/mistral-finetune

[6] https://github.com/mistralai/mistral-inference

You are an AI assistant tasked with analyzing scientific abstracts for idea recombination. Your goal is to identify the most salient recombination in the given abstract and format it as a JSON string. Follow these instructions carefully:

1. First, familiarize yourself with the possible entity types for recombinations:

<entity_types>
combination-element: An idea, method, model, technique, or approach combined in the text with other elements.
inspiration-source: A concept, idea, problem, approach, or domain the authors drew inspiration from.
inspiration-target: A concept, idea, problem, approach, or domain in which the authors utilize the inspiration they drew from the inspiration source.
</entity_types>

2. Now, carefully read the following scientific abstract: <abstract>{TEXT}</abstract>

3. Your task is to extract the most salient recombination from this abstract. A recombination can be either:
a) Combination: The authors combine two or more ideas, methods, models, techniques, or approaches to obtain a certain goal.
b) Inspiration: The authors draw inspiration or similarities from one concept, idea, problem, approach, or domain and implement it in another.

4. After identifying the recombination, you will format it as a JSON string in the following structure:

<recombination>{recombination_type: {entity_type_1: [ent_1, ent_2], entity_type_2: [ent_3],...}}</recombination>

If you don't think the text discusses a recombination, or that the recombination is not a central part of the work, return an empty JSON object: {}.

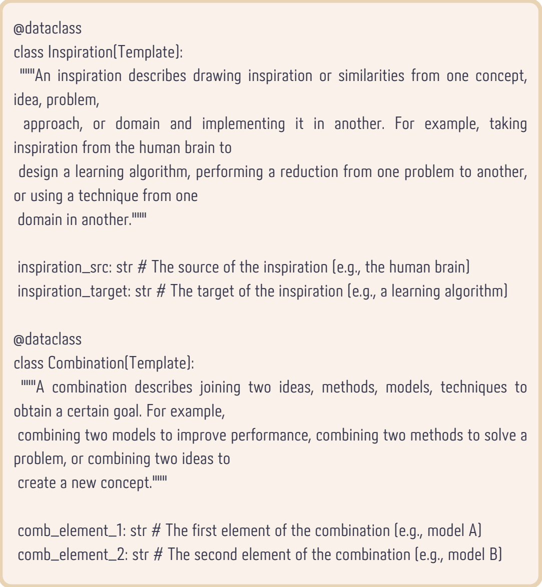
5. Before providing your final answer, use the following scratchpad to think through the process:

<scratchpad>
1. Identify the main ideas, methods, or approaches discussed in the abstract.
2. Determine if there is a clear combination of ideas or if one idea inspired the application in another domain.
3. Identify the specific entities involved in the recombination.
4. Classify the entities according to the provided entity types.
5. Determine the recombination type (combination or inspiration).
</scratchpad>

6. Now, provide your final output in the specified JSON format. Ensure that the output is a valid JSON string. If the output is empty, return {}. Place your answer within <answer> tags.

Remember to carefully analyze the abstract and only identify a recombination if it is clearly present and central to the work described.

Figure 6: E2E extraction prompt. {TEXT} is the placeholder for the input abstract text.



```
@dataclass
class Inspiration(Template):
 """An inspiration describes drawing inspiration or similarities from one concept,
idea, problem,
 approach, or domain and implementing it in another. For example, taking
inspiration from the human brain to
 design a learning algorithm, performing a reduction from one problem to another,
or using a technique from one
 domain in another."""

 inspiration_src: str # The source of the inspiration (e.g., the human brain)
 inspiration_target: str # The target of the inspiration (e.g., a learning algorithm)

@dataclass
class Combination(Template):
 """A combination describes joining two ideas, methods, models, techniques to
obtain a certain goal. For example,
 combining two models to improve performance, combining two methods to solve a
problem, or combining two ideas to
 create a new concept."""

 comb_element_1: str # The first element of the combination (e.g., model A)
 comb_element_2: str # The second element of the combination (e.g., model B)
```

Figure 7: GoLLIE guidelines.

RTX A6000 48GB GPU, 1-beam search, and limit the new token number to 128. `GoLLIE` is finetuned from `CODE-LLaMA2`, and receives guidelines in the form of data classes describing what objects and properties the model should extract. Figure 7 depicts the guidelines we used to test `GoLLIE` as an E2E recombination extraction model. In the rare cases where the model returns more than a single recombination type (< 10), we select the first.

We also experiment with `GPT-4o` in few-shot settings. We select 45 examples for each example type (*blend*, *inspiration*, *not-present*) from the training data (a total of 135). As Table 2 describes, the training set only has 45 *inspiration* examples (as opposed to > 100 *blend* and *not-present* examples). 45 is, therefore, the maximal number of examples per class we can sample while keeping the ICL set balanced. We run each experiment 5 times, sampling a new set of few-shot examples in each, and report the average. Figure 8 presents the prompt for this experiment.

Specialized baselines The recombination extraction model has to execute multiple tasks at once (classifying the document, extracting entities, inferring relations), which might be more challenging than performing them separately. To explore this question, we examine our model classification and extraction abilities against designated models for each task. We use `Mistral-7B` as a specialized classifier and experiment with two versions of the training data. The first includes binary responses (*present*, *not-present*), while the other contains a short CoT-style analysis string as well as the gold class. We construct the analysis string by incorporating the human entity annotations into predetermined templates (e.g., "*This paper discusses a recombination since the authors take inspiration from [inspiration-source] and implement it in [inspiration-target]*").

To evaluate entity extraction, we compare our model against `GPT-4o` in few-shot settings and include 45 cases per example type, similarly to the E2E experiment. To account for variability due to example selection, we run each experiment 5 times, sampling a new set of few-shot examples in each, and report the average. The total cost of this process sums up to 50$. The prompt template for this experiment is available on Figure 9.

We experiment with non-generative approaches as well, and compare our model to a `SciBERT` (Zhong and Chen, 2021) based token classifier. The encoder uses a standard Hugging-Face implementation of `SciBERT`, which we train on a single NVIDIA RTX A6000 48GB GPU over 500 steps. We use a weight decay of 0.1, a learning rate of $6.e-5$ and a batch size of 1. We also experiment with PURE (Zhong and Chen, 2021), a well-known information extraction baseline. We finetune PURE over our train set using the default parameters, except for max_span_length, which we set to 40 to accommodate for the longer entities in our data.

C E2E vs Specialized extraction

This section reflects on the results described in Section 4, drawing on implementation details of the baselines (described in Appendix B). In Section 4, we observe that narrowing the focus to a smaller portion of the recombination extraction task does not always improve performance - in fact, it can lead to worse results. This pattern emerges across three Mistral-based classifiers: the end-to-end version (E2E), the specialized version (Abstract-classifier), and the specialized version trained with synthetic CoT strings (Abstract-classifier-CoT). We hypothesize that identifying recombination relations in text may be analogous to Chain-of-Thought prompting (CoT), a technique known to enhance LLM performance across various tasks (Wei et al., 2022). This hypothesis is supported by the superior performance of Abstract-classifier-CoT compared to its non-CoT counterpart.

You are an AI assistant tasked with analyzing scientific abstracts for idea recombination. Your goal is to identify the most salient recombination in a given abstract and format it as a JSON string. Follow these instructions carefully:

1. First, familiarize yourself with the possible entity types for recombinations:

<entity_types>
comb-element: An idea, method, model, technique, or approach combined in the text with other elements.
inspiration-src: A concept, idea, problem, approach, or domain the authors drew inspiration from.
inspiration-target: A concept, idea, problem, approach, or domain in which the authors utilize the inspiration they drew from the inspiration source.
</entity_types>

2. Review the following examples to understand the expected output format and the process of identifying recombinations:

<examples>{EXAMPLES}</examples>

3. Now, carefully read the following scientific abstract: <abstract>{TEXT}</abstract>

4. Your task is to extract the most salient recombination from this abstract. A recombination can be either:
a) Combination: The authors combine two or more ideas, methods, models, techniques, or approaches to obtain a certain goal.
b) Inspiration: The authors draw inspiration or similarities from one concept, idea, problem, approach, or domain and implement it in another.

5. After identifying the recombination, you will format it as a JSON string in the following structure:

<recombination>{recombination_type: {entity_type_1: [ent_1, ent_2], entity_type_2: [ent_3],...}}</recombination>

If you don't think the text discusses a recombination, or that the recombination is not a central part of the work, return an empty JSON object: {}.

6. Before providing your final answer, use the following scratchpad to think through the process:

<scratchpad>
1. Identify the main ideas, methods, or approaches discussed in the abstract.
2. Determine if there is a clear combination of ideas or if one idea inspired the application in another domain.
3. Identify the specific entities involved in the recombination.
4. Classify the entities according to the provided entity types.
5. Determine the recombination type (combination or inspiration).
</scratchpad>

7. Now, provide your final output in the specified JSON format. Ensure that the output is a valid JSON string. If the output is empty, return {}. Place your answer within <recombination> tags.

Remember to carefully analyze the abstract and only identify a recombination if it is clearly present and central to the work described.

Figure 8: E2E ICL prompt. {TEXT} is a placeholder for the abstract text, and {EXAMPLES} for the ICL examples.

```
You are tasked with identifying specific types of entities in a given scientific abstract. The entity types you need to identify are:

1. comb-element: An idea, method, model, technique, or approach combined in the text with other elements.
2. inspiration-src: A concept, idea, problem, approach, or domain the authors drew inspiration from.
3. inspiration-target: A concept, idea, problem, approach, or domain in which the authors utilize the inspiration they drew from the inspiration source.

Here is the text you need to analyze:

<text>{TEXT}</text>

Please read the text carefully and identify all entities that belong to the types listed above. Pay close attention to the context and relationships between concepts to accurately categorize each entity.

After identifying the entities, you should output them in a valid JSON format. Use the entity types as keys and lists of entities as values. For example:

{"comb-element": ["entity1", "entity2"],
 "inspiration-src": ["entity3"],
 "inspiration-target": ["entity4", "entity5"]}

Ensure that your JSON output is valid:
- Use double quotes around strings
- Do not include a trailing comma after the last item in a list or object
- Escape any double quotes that appear within entity names

Enclose your final JSON output in <output_json> tags.

Remember to review your output for accuracy and completeness before submitting your final answer.
```

Figure 9: Entity extraction prompt. {TEXT} is a placeholder for the input abstract.

D Graph nodes domains

We identify entities' scientific domain using `GPT-4o` in zero-shot settings. Given the abstract and recombination entities extracted from it, the model has to assign each an arXiv category and a scientific branch. In case the model manages to assign the entity an arxive category, the scientific branch is the category's full name (e.g., "Artificial Intelligence" for cs.AI). Otherwise, the models assign the branch from a list of outer-arXiv domains described in Table 8. If the model can assign the entity a standard arXiv category, we use it as the domain. Otherwise, we use the branch (an outer-arXiv domain). Entities with neither are assigned to the `Other` domain. Figures 10 and 11 present our analysis prompts for blend and inspiration relations, respectively. The cost of running the analysis over the collected corpus was 250$.

The `Other` domain We use the `Other` domain for nodes the model fails to analyze, and 2127 of the graph nodes are assigned to this category. We examine a sample of 150 such nodes and observe that many are too noisy or overly general to classify. Interestingly, some nodes in this domain describe nonacademic or niche concepts (see examples in Table 9).

domain grouping To avoid sparsity, we group similar domains as displayed in Table 10. Table 11 presents the node distribution of common domains after applying this grouping process.

E Span similarity prompt

We provide our span similarity prompt in Figure E. We use it in the extraction evaluation process as discussed in Section 4.1.

F Error analysis

We perform analysis over the test set, revealing different sources of error which may inspire future improvements. Our focus is on understanding how different types of input texts can influence the result, specifically, in cases where the extraction model struggles. We use our best-performing fine-tuned E2E model for this analysis.

Context dependent or implicit phrasing We observe that, unsurprisingly, cases in which the recombination is implied or subtle are more challenging for the model. For instance (see also Table 12, row 1), *"Kahneman & Tversky's prospect theory"* inspires the design of a loss function that *"directly maximizes the utility of generations"*, but this is not stated explicitly. Moreover, abstracts that explicitly express idea recombination while referencing previously mentioned entities are also harder to detect.

Multiple recombinations Some papers present a salient recombination along with other insignificant ones. We notice that in those cases, the model might extract a non-salient recombination or mix multiple ones (see Table 12, row 2 for such a case).

Borderline cases The role of a recombination as a core element in the work is sometimes debatable. Table 12, row 3 presents an example of such a case where the authors explicitly mention integrating *"embedding space comparison"* with *"computational notebook environment"*, which may be interpreted as a recombination (the usage of notebook in these environments is completely new and novel), or simply as a way to present the tool's environment. We notice that the extraction model tends to miss those cases.

Non-arXiv scientific domains		
Agricultural Science	Anatomy	Animal Science
Anthropology	Archaeology	Behavioral Science
Biochemistry	Bioinformatics	Bioclimatology
Biomedical Engineering	Biophysics	Biotechnology
Botany	Cardiology	Chemical Engineering
Civil Engineering	Clinical Psychology	Cognitive Science
Criminology	Cryosphere Science	Cytology
Demography	Dentistry	Dermatology
Developmental Biology	Ecology	Ecotoxicology
Economics	Educational Psychology	Electrical Engineering
Emergency Medicine	Endocrinology	Energy Science
Engineering Science	Entomology	Environmental Engineering
Environmental Science	Epidemiology	Ethology
Food Science	Forestry	Gastroenterology
Genetics	Genomics	Geography
Geology	Geophysics	Glaciology
Health Informatics	Histopathology	Hydrodynamics
Hydrogeology	Hydrology	Immunogenetics
Immunology	Industrial/Organizational Psychology	Landscape Architecture
Linguistics	Marine Biology	Materials Science
Mechanical Engineering	Medical Microbiology	Meteorology
Microbiology	Mineralogy	Molecular Biology
Mycology	Nanotechnology	Neurology
Neuroscience	Nuclear Engineering	Nutritional Science
Obstetrics	Oceanography	Oncology
Ophthalmology	Ornithology	Orthopedics
Otology	Paleoclimatology	Paleontology
Pathobiology	Pathology	Pediatric Medicine
Pedagogy	Petrology	Pharmacogenomics
Pharmacology	Philosophy	Physiology
Political Science	Proteomics	Psychiatry
Psychology	Psychopathology	Public Health
Pulmonology	Radiology	Rheumatology
Seismology	Social Psychology	Sociology
Surgery	Systems Biology	Thermodynamics
Toxicology	Urban Planning	Urology
Veterinary Science	Virology	Volcanology
Wildlife Biology	Zoology	

Table 8: Non-arXiv scientific domains. We complement arXiv category taxonomy using a broader list of scientific fields.

You are an AI assistant tasked with analyzing a scientific abstract to determine the arXiv categories and scientific branches of combined elements. Your goal is to identify the most appropriate arxiv taxonomy category and most suitable scientific domain for each element provided.

Here is the abstract you will be analyzing:
<abstract>{ABSTRACT}</abstract>

And here is the list of combined elements identified from the abstract:
<elements>{ELEMENTS}</elements>

Here is a list of the standard arXiv categories:
<arxiv>{ARXIV}</arxiv>

And here is a list of scientific branches:
<branches>{BRANCHES}</branches>

For each element in the list, you need to:
1. Identify the best matching arXiv taxonomy category from the provided list. If it doesn't match any category, use "other". If there's insufficient information, use "insufficient-info".
2. Identify the scientific branch from the provided branches list. If there's insufficient information, use "insufficient-info". If no branch name in the list describes the source properly, use "other".

Return your output in the following format:
<output>
[{"text": "element1",
 "arxiv_category": "category1",
 "scientific_branch": "branch1"},
 {"text": "element2",
 "arxiv_category": "category2",
 "scientific_branch": "branch2"}, ...]
</output>

Format your response as a valid JSON string.

Now, analyze the provided elements from the abstract and generate your response in the specified JSON format. Make sure to include all elements from the provided list, and ensure that your output is properly formatted as a valid JSON string.

Figure 10: blend domain analysis prompt. {ELEMENTS} is a placeholder for the recombination entities extracted from {ABSTRACT}. {ARXIV} is a placeholder for full arXiv category names and their descriptions. {BRANCHES} is a placeholder for the list of non-arXiv domains given in Appendix D, Table 8.

```
You will be analyzing the scientific branches and arXiv taxonomy categories of an inspiration source and target based on an abstract from a scientific paper. Here's the information you'll be working with:

<abstract>{ABSTRACT}</abstract>

<inspiration_source>{INSPIRATION_SOURCE}</inspiration_source>

<inspiration_target>{INSPIRATION_TARGET}</inspiration_target>

<arxiv>{ARXIV}</arxiv>

<branches>{BRANCHES}</branches>

Your task is to identify the arXiv taxonomy category and most suitable scientific branch for both the inspiration source and the inspiration target.

For the inspiration source:
1. Identify the best matching arXiv taxonomy category from the provided list. If it doesn't match any category, use "other". If there's insufficient information, use "insufficient-info".
2. Identify the scientific branch from the provided branches list. If there's insufficient information, use "insufficient-info". If no branch name in the list describes the source properly, use "other".

Repeat the same process for the inspiration target.

Provide your analysis in the following format:

<source-branch>[Insert the scientific branch of the inspiration source here]</source-branch>
<source-arXiv>[Insert the arXiv taxonomy category of the inspiration source here]</source-arXiv>
<target-branch>[Insert the scientific branch of the inspiration target here]</target-branch>
<target-arXiv>[Insert the arXiv taxonomy category of the inspiration target here]</target-arXiv>

Ensure that you only include the requested information within each tag, without any additional explanation or reasoning.
```

Figure 11: inspiration domain analysis prompt. {INSPIRATION_SOURCE} and {INSPIRATION_TARGET} are placeholders for the inspiration entities extracted from {ABSTRACT}. {ARXIV} is a placeholder for full arXiv category names and their descriptions. {BRANCHES} is a placeholder for the list of non-arXiv domains given in Appendix D, Table 8.

Type	**Examples**
Non-Academic	"the snap-through action of a steel hairclip", "yoga", "origami, the traditional Japanese paper-folding technique, is a powerful metaphor for design and fabrication of reconfigurable structures", "Tangram, a game that requires replicating an abstract pattern from seven dissected shapes"
Noisy	"a deep", "word-", "at the context level", "a neural part", "post", "text–audio", "end-to-end multi-modal model only X-VLM only X-VLM only X-VLM only X-VLM only X-VLM only X-VLM only X-VLM only X-VLM only X-VLMs", "a user's long-term"
Overly-general	"human experiences", "a styling method", "local search method", "a pipeline inspired by experts' work", "a new modality", "feature based approaches"
Misclassified	"Reinforcement learning, or RL", "Facial Expressions Recognition(FER)", "a Kullback-Liebler regularization function", "K-nearest neighbors algorithm", "Shapley values from game theory", "Gaussian Stochastic Weight Averaging"

Table 9: Examples of graph nodes in the "other" domain. We analyze a sample of 150 nodes in this domain and identify groups with common traits, as shown in the table.

Group	Scientific domains
Geosciences	Geology, Geophysics, Petrology, Mineralogy, Hydrology, Hydrogeology, Seismology, Volcanology, Cryosphere Science, Glaciology, Geography
Environmental Sciences	Environmental Science, Environmental Engineering, Ecology, Ecotoxicology
Biomedical Sciences	Biochemistry, Immunology, Immunogenetics, Neuroscience, Oncology, Pathology, Pathobiology, Pharmacology, Toxicology
Health and Medicine	Cardiology, Neurology, Urology, Gastroenterology, Obstetrics, Pediatric Medicine, Rheumatology, Dermatology, Ophthalmology, Otology, Pulmonology, Emergency Medicine, Surgery, Radiology, Orthopedics, Psychiatry, Dentistry, Public Health, Epidemiology, Health Informatics, Clinical Psychology, Psychopathology
Zoology	Zoology, Entomology, Ornithology, Wildlife Biology, Animal Science, Veterinary Science, Ethology
Agriculture	Agricultural Science, Forestry
Food Sciences	Nutritional Science, Food Science
Psychology	Educational Psychology, Social Psychology, Psychology, Industrial/Organizational Psychology
Microbiology	Microbiology, Medical Microbiology
Humanities	Linguistics, Philosophy, Pedagogy
Social Sciences	Sociology, Anthropology, Political Science, Demography

Table 10: Scientific domains grouped by category. We group similar non-arXiv scientific domains (see Table 8) to thicken infrequent ones.

G Extraction examples

Table 13 presents examples of interdisciplinary, automatically extracted inspiration recombinations.

H Predominant inspiration and blend relations

We provide a tabular version of Figure 3 in Section 4.2 on Table 14 for better readability.

I Prediction data preprocessing

Context extraction and leakage filtering We use GPT-4o-mini to extract a few sentences from each abstract describing the background or motivation of the authors using recombination (See prompt on Figure 13). Adding these contexts to the queries helps them be more specific and limits the search space. However, this might introduce leaks into the queries - cases where the extracted context reveals the answer. Table 15 presents leak examples. We utilize GPT-4o-mini again to filter out such cases from the data, using the prompt shown in Figure 14. In a qualitative analysis of 50 randomly sampled query-answer pairs, we find that a human annotator agrees with 87% of the model's predictions (whether there is a leak). Finally, we divide the remaining query-answer pairs into splits as described in Table 5 is Section 5.2.

J Prediction baselines

We use a bi-encoder architecture for recombination prediction and experiment with three popular encoders as backbones: `all-mpnet-base-v2` (109M parameters), `bge-large-en-v1.5` (Xiao et al., 2023) (335M parameters) and `e5-large-v2` (Wang et al., 2022) (335M parameters). These models' checkpoints predate 2024, meaning they are unfamiliar with our test set. The model receives a query string composed of a context description, a graph entity, and a relation type and returns a ranked list of answers (other graph nodes). We perform HPO (random grid search of 10 trails) to select the number of training epochs, warmup ratio and learning rate for each model. We use contrastive loss and generate 30 negatives per positive example. Following the literature standard (Teach et al., 2020), we report metrics in the filtered settings to avoid false negatives. Given the difficulty of the task we focus on ranking only the 12751 test

Domain	**Count**	**Domain**	**Count**	**Domain**	**Count**
cs.cv	12504	cs.lg	8440	cs.cl	4697
cs.ro	2241	cs.ai	2091	cognitive science	936
cs.ir	884	cs.ne	864	cs.si	655
cs.hc	645	q-bio.nc	441	cs.ds	409
cs.cg	382	cs.cy	378	cs.gr	367
math.oc	356	eess.iv	278	cs.dm	269
cs.db	254	eess.sp	242	cs.lo	204
cs.ma	203	cs.ce	185	cs.sy	177
cs.cr	164	stat.me	138	cs.gt	132
psychology	116	eess.sy	108	cs.se	104
zoology	101	cs.it	100	math.pr	96
cs.dc	89	behavioral science	88	cs.mm	82
eess.as	79	nlin.ao	79	cs.ar	74
cs.na	66	cs.pl	65	biomedical sciences	63
physics.med-ph	60	stat.ml	56	health and medicine	56
physics.bio-ph	52	cs.ni	48	physics.ao-ph	44
stat.th	43	anatomy	41	math.na	40
math.ds	39	cs.fl	38	humanities	38
q-bio.pe	32	cs.dl	32	cs.sc	30
math-ph	27	cond-mat.stat-mech	25	math.ap	24
math.dg	22	physics.class-ph	22	cs.sd	22
econ.th	21	math.ca	21	math.mg	20
physics.comp-ph	20	physics.optics	20	cs.et	20

Table 11: Node domains distribution. The table presents the number of graph nodes from each domain with above-median frequency.

Bad extraction examples
Abstract: "...Kahneman & Tversky's prospect theory tells us that humans perceive random variables in a biased but well-defined manner (1992) ... Using a Kahneman-Tversky model of human utility, we propose a HALO [Human Aware Loss Function] that directly maximizes the utility of generations instead of maximizing the log-likelihood of preferences, as current methods do..." **Gold** = [**Inspiration**: "*Kahneman & Tversky's prospect theory*" ⟶ "*a HALO*"] **Pred** = []
Abstract: "...We address the problem by proposing a Wasserstein GAN combined with a new reverse mask operator, namely Reverse Masking Network (R-MNet), a perceptual adversarial network for image inpainting ... Additionally, we propose a new loss function computed in feature space to target only valid pixels combined with adversarial training..." **Gold** = [**Blend**: "*a Wasserstein GAN*" ⟷ "*...R-MNet*"] **Pred** = [**Blend**: "*a Wasserstein GAN*" ⟷ "*...R-MNet*" ⟷ "*a new loss function*"]
Abstract: "... In order to characterize model flaws and choose a desirable representation, model builders often need to compare across multiple embedding spaces, a challenging analytical task supported by few existing tools. We first interviewed nine embedding experts in a variety of fields to characterize the diverse challenges they face and techniques they use when analyzing embedding spaces. Informed by these perspectives, we developed a novel system called Emblaze that integrates embedding space comparison within a computational notebook environment..." **Gold** = [**Blend**: "*embedding space comparison*" ⟷ "*...notebook environment*"] **Pred** = []

Table 12: In the first row, the extraction model misses an inspiration relation because of subtle phrasing. In the second row, when analyzing an abstract with multiple recombinations, the model fails to identify the most important one and confuses entities across different relations. In the third row, the model fails to detect a weak recombination example.

Detected Inspiration	**Abstract**
Inspiration-Source: "the shepherding behavior of herding dogs" [zoology] *Inspiration-Target*: "Frontier exploration" [cs.ro]	"Efficient exploration of large-scale environments remains a critical challenge in robotics... The presented bio-inspired framework heuristically models **frontier exploration** similar to the **shepherding behavior of herding dogs**. This is achieved by modeling frontiers as a sheep swarm reacting to robots modeled as shepherding dogs..."
Inspiration-Source: "the multi-granular diagnostic approach of pathologists" [biomedical sciences] *Inspiration-Target*: "Histopathological image classification" [cs.cv]	"**Histopathological image classification** constitutes a pivotal task in computer-aided diagnostics... In the diagnostic process of pathologists, a multi-tiered approach is typically employed to assess abnormalities in cell regions at different magnifications...Inspired by **the multi-granular diagnostic approach of pathologists**, we perform feature extraction on cell structures at coarse, medium, and fine granularity, enabling the model to fully harness the information in histopathological images..."
Inspiration-Source: "the dual-process theory in psychology" [psychology] *Inspiration-Target*: "enhance dialogue planning in LLMs" [cs.cl]	"...Traditional approaches to **enhance dialogue planning in LLMs**, ... either face efficiency issues or deliver suboptimal performance. Inspired by **the dualprocess theory in psychology**, which identifies two distinct modes of thinking - intuitive (fast) and analytical (slow), we propose the Dual-Process Dialogue Planning (DPDP) framework. DPDP embodies this theory through two complementary planning systems: an instinctive policy model for familiar contexts and a deliberative Monte Carlo Tree Search (MCTS) mechanism for complex, novel scenarios..."
Inspiration-Source: "the Global Workspace Theory in conscious processing" [cognitive science] *Inspiration-Target*: "learning effective feature embeddings for CTR prediction" [cs.lg]	"**Click-Through Rate (CTR) prediction is a pivotal task in product and content recommendation, where learning effective feature embeddings is of great significance**....inspired by **the Global Workspace Theory in conscious processing**, which posits that only a specific subset of the product features are pertinent while the rest can be noisy and even detrimental to human-click behaviors, we propose a CTR model that enables Dynamic Embedding Learning with Truncated Conscious Attention for CTR prediction, termed DELTA..."

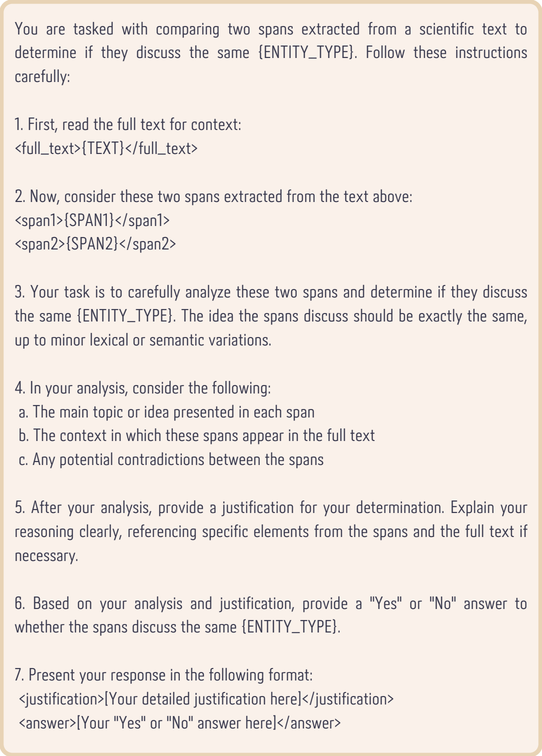
Table 13: Inter-domain inspiration examples from the `CHIMERA` knowledge graph.

Inspirations			*Blends*		
Source	**Target**	**Count**	**Source**	**Target**	**Count**
cs.cv	cs.cv	334	cs.cv	cs.cv	4329
cs.lg	cs.cv	300	cs.lg	cs.lg	2793
cognitive science	cs.cv	278	cs.cl	cs.cl	1049
cs.lg	cs.lg	254	cs.lg	cs.cv	992
cognitive science	cs.lg	211	cs.cv	cs.lg	470
cs.cl	cs.cl	190	cs.cl	cs.cv	422
cs.cl	cs.cv	188	cs.cv	cs.cl	391
cognitive science	cs.ai	184	cs.lg	cs.cl	363
cognitive science	cs.cl	142	cs.ro	cs.ro	299
cs.cl	cs.ai	141	cs.ro	cs.cv	218
cs.lg	cs.ai	118	cs.cl	cs.lg	197
q-bio.nc	cs.cv	114	cs.ai	cs.cl	174
q-bio.nc	cs.lg	102	cs.ai	cs.ai	161
cognitive science	cs.ro	100	cs.ai	cs.lg	151
cs.cv	cs.lg	94	cs.lg	cs.ai	146
cs.lg	cs.cl	84	cs.lg	cs.ne	133
cs.cl	cs.lg	84	cs.ir	cs.ir	132
math.oc	cs.lg	83	cs.lg	cs.ro	124
zoology	cs.ro	76			

Table 14: Predominant inspiration and blend relations. The above is a tabular version of Figures 3b, 3a in Section 4.2. It presents edges with (source-domain, target-domain) pairs frequency above the 0.98 quantile.

Query	**Answer**
Understanding the human brain's processing capabilities can inspire advancements in machine learning algorithms and architectures. Previous methods in brain research were limited to identifying regions of interest for one subject at a time, restricting their applicability and scalability across multiple subjects. What would be a good source of inspiration for "**a highly efficient processing unit**"?	The human brain
Existing models for link prediction in knowledge graphs primarily focus on **representing triplets in either distance or semantic space**, which limits their ability to fully capture the information of head and tail entities and utilize hierarchical level information effectively. **This indicates a need for improved methods that can leverage both types of information** for better representation learning in knowledge graphs. What could we blend with "**distance measurement space**" to address the described settings?	Semantic measurement space

Table 15: Leakages examples. Examples of leaks - queries that reveal or strongly imply the answer.


You are tasked with comparing two spans extracted from a scientific text to determine if they discuss the same {ENTITY_TYPE}. Follow these instructions carefully:

1. First, read the full text for context:
<full_text>{TEXT}</full_text>

2. Now, consider these two spans extracted from the text above:
<span1>{SPAN1}</span1>
<span2>{SPAN2}</span2>

3. Your task is to carefully analyze these two spans and determine if they discuss the same {ENTITY_TYPE}. The idea the spans discuss should be exactly the same, up to minor lexical or semantic variations.

4. In your analysis, consider the following:
 a. The main topic or idea presented in each span
 b. The context in which these spans appear in the full text
 c. Any potential contradictions between the spans

5. After your analysis, provide a justification for your determination. Explain your reasoning clearly, referencing specific elements from the spans and the full text if necessary.

6. Based on your analysis and justification, provide a "Yes" or "No" answer to whether the spans discuss the same {ENTITY_TYPE}.

7. Present your response in the following format:
 <justification>[Your detailed justification here]</justification>
 <answer>[Your "Yes" or "No" answer here]</answer>

Figure 12: Span similarity prompt. {ENTITY_TYPE} is either "combination-element", "inspiration-source" or "inspiration-target". {TEXT} is a placeholder for the paper's abstract. {SPAN1}, {SPAN2} are placeholders for the compared spans.

set entities. A full summary of our data splits is available on 5. The examples we use to train and evaluate our prediction models contain all collected nodes, including those classified as belonging to the "other" domain.

We utilize RankGPT (Sun et al., 2023) as a strong reranker and apply it to rerank the top-20 predicted results. We employ RankGPT with GPT-4o, a window size of 10 and a step size of 5. Note the information cutoff of GPT-4o is October 2023 [7], meaning it is unfamiliar with our test set as well. We use the implementation available in [8]. However, we find that adjusting the default prompt works better for our task. Figure 15 shows the modified reranking prompt. The cost of applying the reranker to our data was 60$.

K User study additional details

We request each to fill out a form asking in what scientific domains they feel comfortable reading papers and a short description of their research area. We then used `granite-embedding-125m-english` to retrieve semantically similar contexts to this description from the relevant arXiv categories. We manually verify that the retrieved contexts match the description and discard examples with poorly extracted information (e.g., the context begins with "*This study reviews the problem of...*" instead of directly describing the source study problem). In addition, we let the volunteers mark an example as "ill-defined", in which case we ignore their inputs. We conduct a 10-minute training session with each volunteer, requesting them to read the instructions and explain the task. Figure 16 presents the instructions given to the participants in the study. Figure 17 presents the web interface of the annotation platform.

L Comparison to other information extraction methods

Both general scientific extraction and concept co-occurrence struggle to capture concise and accurate recombination relations, as can be seen in Figure 18. Figure 18a presents how general scientific IE schemas lack relation types to model recombinations. The figure presents the results of our specialized extraction method besides a transformer-based extraction model (Hennen et al., 2024) fine-

---

[7] As stated in https://platform.openai.com/docs/models/gpt-4o

[8] https://github.com/sunnweiwei/RankGPT/tree/main

You are tasked with extracting the rationale behind the selection of a specific methodology in a scientific study. You will be provided with an abstract and a statement about the methodology used. Your goal is to extract the reasons for choosing this methodology from the abstract.

First, carefully read the following abstract: <abstract>{{ABSTRACT}}</abstract>

Next, inspect the following examples of background descriptions:

1. Large language models (LLMs) commonly employ autoregressive generation during inference, leading to high memory bandwidth demand and consequently extended latency.
2. Reconstructing deformable tissues from endoscopic videos is essential in many downstream surgical applications. However, existing methods suffer from slow rendering speed, greatly limiting their practical use.
3. Many industrial tasks-such as sanding, installing fasteners, and wire harnessing-are difficult to automate due to task complexity and variability.
4. Multi-legged robots offer enhanced stability in complex terrains, yet autonomously learning natural and robust motions in such environments remains challenging.

Now, consider this methodology statement: <methodology_statement>{{METHODOLOGY_STATEMENT}}</methodology_statement>

To complete this task, follow these steps:

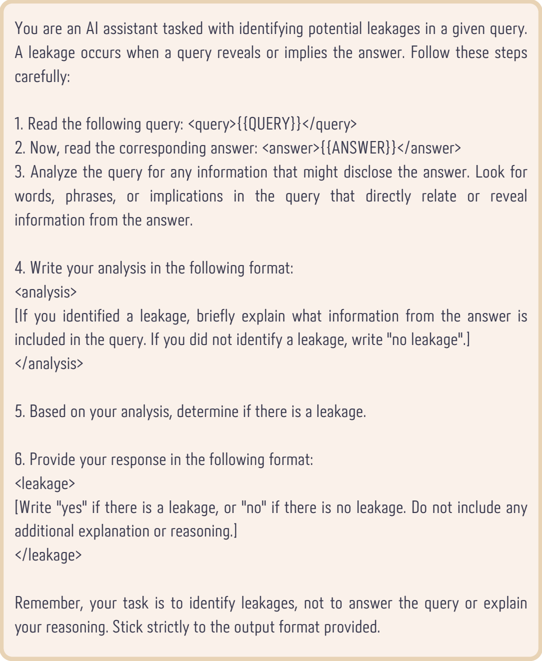
1. Analyze the abstract thoroughly, focusing on:
   - The context or reasons that justify the methodology choice
   - Any challenges, limitations, or research needs the methodology addresses
   - Mentions of previous research or knowledge gaps that the methodology aims to target

2. When formulating your response:
   - Phrase your response as a general 1-2 sentence description of a challenge, limitation research needs, etc.
   - Use exclusively the information from the abstract. Do not incorporate external knowledge or assumptions.
   - Minimize including information from the methodology statement in your answer.
   - Do not include information about the used methodology in your answer.
   - If the background details are unclear, return an empty response.

3. Format your response as follows:
<background>
[1-2 background sentences]
</background>

Remember to base your response strictly on the provided abstract and statement. Do not include additional information or assumptions.

Figure 13: Context extraction prompt. {{ABSTRACT}} is a placeholder for the input abstract. {{METHODOLOGY_STATEMENT}} is a sentence describing the recombination. We build it by filling one of the following templates with the extracted recombination entities: "*Combine <source-entity> and <target-entity>*" for blends and "*Take inspiration from <source-entity> and apply it to <target-entity>*" for inspirtions.



You are an AI assistant tasked with identifying potential leakages in a given query. A leakage occurs when a query reveals or implies the answer. Follow these steps carefully:

1. Read the following query: <query>{{QUERY}}</query>
2. Now, read the corresponding answer: <answer>{{ANSWER}}</answer>
3. Analyze the query for any information that might disclose the answer. Look for words, phrases, or implications in the query that directly relate or reveal information from the answer.

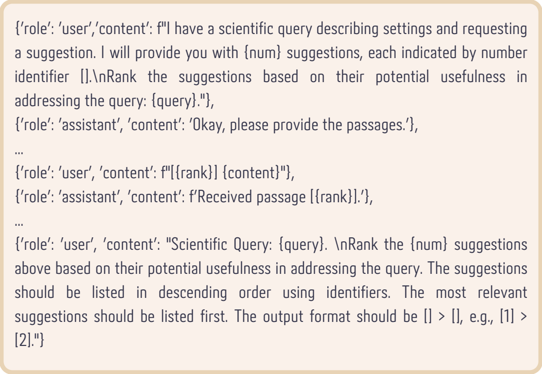
4. Write your analysis in the following format:
<analysis>
[If you identified a leakage, briefly explain what information from the answer is included in the query. If you did not identify a leakage, write "no leakage".]
</analysis>

5. Based on your analysis, determine if there is a leakage.

6. Provide your response in the following format:
<leakage>
[Write "yes" if there is a leakage, or "no" if there is no leakage. Do not include any additional explanation or reasoning.]
</leakage>

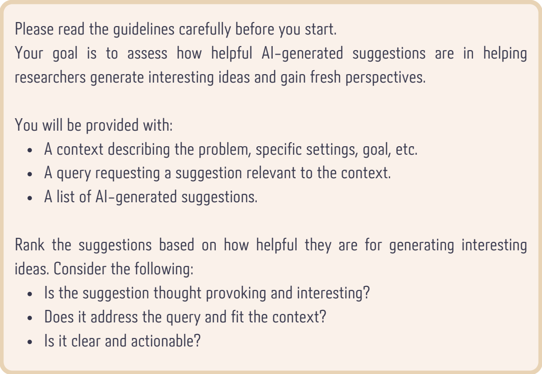
Remember, your task is to identify leakages, not to answer the query or explain your reasoning. Stick strictly to the output format provided.

Figure 14: Leak detection prompt.



{‘role’: ‘user’,‘content’: f"I have a scientific query describing settings and requesting a suggestion. I will provide you with {num} suggestions, each indicated by number identifier [].\nRank the suggestions based on their potential usefulness in addressing the query: {query}."},
{‘role’: ‘assistant’, ‘content’: ‘Okay, please provide the passages.’},
...
{‘role’: ‘user’, ‘content’: f"[{rank}] {content}"},
{‘role’: ‘assistant’, ‘content’: f‘Received passage [{rank}].’},
...
{‘role’: ‘user’, ‘content’: "Scientific Query: {query}. \nRank the {num} suggestions above based on their potential usefulness in addressing the query. The suggestions should be listed in descending order using identifiers. The most relevant suggestions should be listed first. The output format should be [] > [], e.g., [1] > [2]."}

Figure 15: Adjusted RankGPT prompt.



Please read the guidelines carefully before you start.
Your goal is to assess how helpful AI-generated suggestions are in helping researchers generate interesting ideas and gain fresh perspectives.

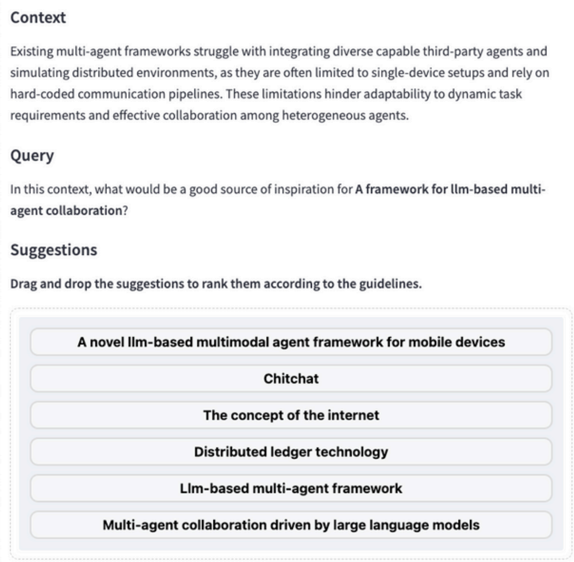
You will be provided with:

- A context describing the problem, specific settings, goal, etc.
- A query requesting a suggestion relevant to the context.
- A list of AI-generated suggestions.

Rank the suggestions based on how helpful they are for generating interesting ideas. Consider the following:

- Is the suggestion thought provoking and interesting?
- Does it address the query and fit the context?
- Is it clear and actionable?

Figure 16: User study guidelines.



Context

Existing multi-agent frameworks struggle with integrating diverse capable third-party agents and simulating distributed environments, as they are often limited to single-device setups and rely on hard-coded communication pipelines. These limitations hinder adaptability to dynamic task requirements and effective collaboration among heterogeneous agents.

Query

In this context, what would be a good source of inspiration for **A framework for llm-based multi-agent collaboration**?

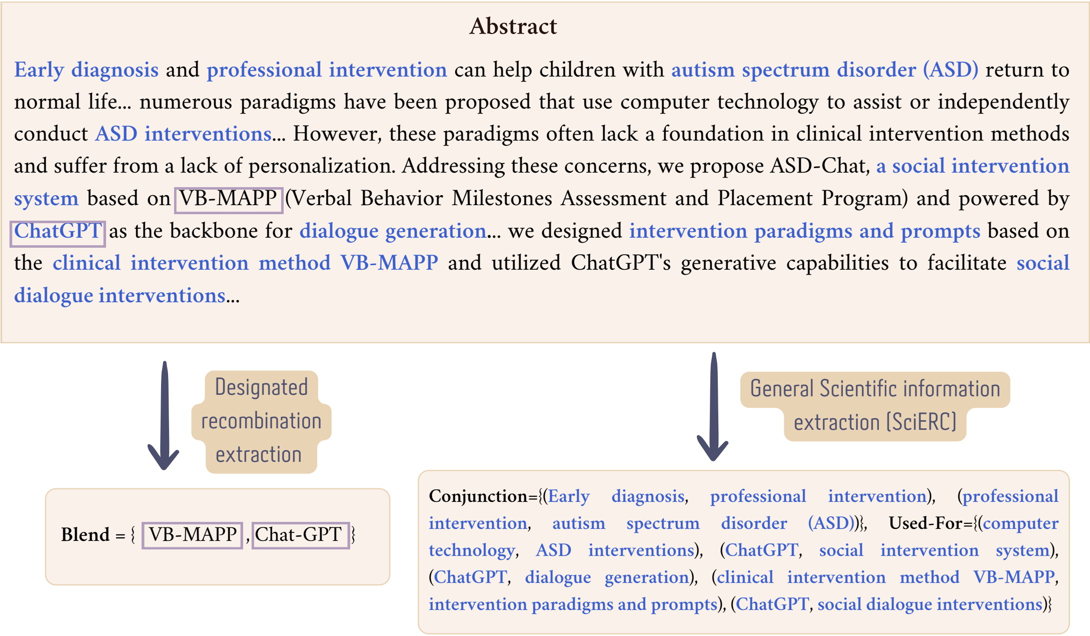
Suggestions

Drag and drop the suggestions to rank them according to the guidelines.

A novel llm-based multimodal agent framework for mobile devices
Chitchat
The concept of the internet
Distributed ledger technology
Llm-based multi-agent framework
Multi-agent collaboration driven by large language models

Figure 17: User study interface.

tuned on SciERC (Luan et al., 2018), a general IE schema. While our new data schema easily models the recombinant connection between two techniques: "*BV-MAPP (Verbal Behavior Milestones Assessment and Placement Program)*", "*ChatGPT*" as a concept *blend*, the SciERC extraction schema isn't equipped with proper relation types for this. As a result, it captures mostly irrelevant information for our task (e.g background details as "*Early diagnosis*" or "*professional intervention*"). Figure 18b shows how recombination extraction using concept co-occurrence might be misleading. In this method, each pair of canonical scientific concepts (e.g, *neural networks*) that co-occur within the same abstract are considered a recombination. The figure presents an example of using AI-related concepts curated by Krenn et al. (2022) for recombination extraction, alongside recombination extracted using our designated approach. Note that when using concept co-occurrence, the extracted recombinations are essentially $\{concepts\}^2$, which might be imprecise, and capture meaningless recombinations (e.g., "*wide application*" recombined with "*final prediction*") or misleading recombinations (e.g., "*question answering*" with "*language models*", which explicitly presented by the authors as a lacking approach for the task). In comparison, our new extraction schema neatly models the main recombiant relation presented in the text as taking inspiration from "*the step-by-step reasoning behavior of humans*" for "*temporal question answering*."



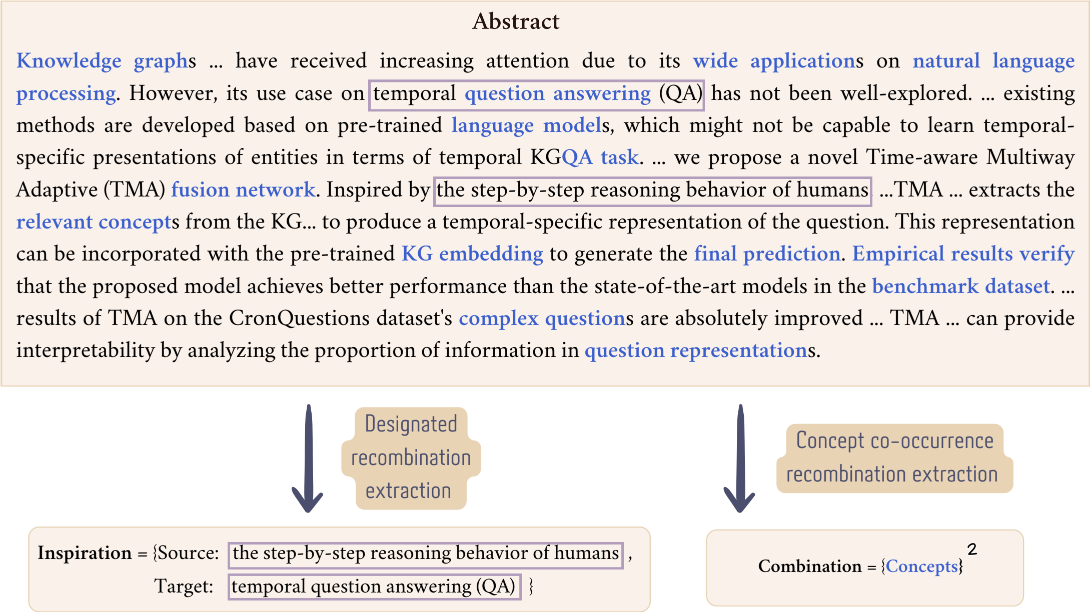
Abstract

Early diagnosis and professional intervention can help children with autism spectrum disorder (ASD) return to normal life... numerous paradigms have been proposed that use computer technology to assist or independently conduct ASD interventions... However, these paradigms often lack a foundation in clinical intervention methods and suffer from a lack of personalization. Addressing these concerns, we propose ASD-Chat, a social intervention system based on VB-MAPP (Verbal Behavior Milestones Assessment and Placement Program) and powered by ChatGPT as the backbone for dialogue generation... we designed intervention paradigms and prompts based on the clinical intervention method VB-MAPP and utilized ChatGPT's generative capabilities to facilitate social dialogue interventions...

Blend = { VB-MAPP , Chat-GPT }

Conjunction={(Early diagnosis, professional intervention), (professional intervention, autism spectrum disorder (ASD))}, **Used-For**={(computer technology, ASD interventions), (ChatGPT, social intervention system), (ChatGPT, dialogue generation), (clinical intervention method VB-MAPP, intervention paradigms and prompts), (ChatGPT, social dialogue interventions)}

(a) Comparison to recombination extraction using a general scientific IE schema (SciERC)



Abstract

Knowledge graphs ... have received increasing attention due to its wide applications on natural language processing. However, its use case on temporal question answering (QA) has not been well-explored. ... existing methods are developed based on pre-trained language models, which might not be capable to learn temporal-specific presentations of entities in terms of temporal KGQA task. ... we propose a novel Time-aware Multiway Adaptive (TMA) fusion network. Inspired by the step-by-step reasoning behavior of humans ...TMA ... extracts the relevant concepts from the KG... to produce a temporal-specific representation of the question. This representation can be incorporated with the pre-trained KG embedding to generate the final prediction. Empirical results verify that the proposed model achieves better performance than the state-of-the-art models in the benchmark dataset. ... results of TMA on the CronQuestions dataset's complex questions are absolutely improved ... TMA ... can provide interpretability by analyzing the proportion of information in question representations.

Inspiration = {Source: the step-by-step reasoning behavior of humans ,
Target: temporal question answering (QA) }

Combination = $\{\text{Concepts}\}^2$

(b) Comparison to recombination extraction using concept co-occurrence.

Figure 18: Comparison of our designate recombination extraction method to alternative approaches. Figure 18a: General recombination extraction schemas lack fitting relation types to capture recombinations, which results in capturing plenty of irrelevant relations ("*Early diagnosis*" ⟷ "*professional intervention*"). Figure 18b: Recombination extraction using concept co-occurrence might be nonsensical ("*wide application*" ⟷ "*final prediction*") or even misleading ("*question answering*" ⟷ "*language models*")).