Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

PRE-TRAINED LANGUAGE MODEL AUGMENTED WITH KNOWLEDGE (PLMS-AWK): COMPREHENSIVE OVERVIEW AND INSIGHTS

Hrishita Singh
Computer Science & Engineering, Asansol Engineering College, West Bengal, India

Gungun Srivastava
Computer Science & Engineering (AI&ML), Asansol Engineering College, West Bengal, India

Akansha Sharma
Information Technology, Asansol Engineering College, West Bengal, India

Dr. Abhishek Bandyopadhyay
HOD of Computer Science & Engineering (AI&ML), Asansol Engineering College, West Bengal, India

Abstract –
The progress regarding Pre-trained Language Models (PLMs) has brought about a transformative impact on the field of Natural Language Processing (NLP). through self-supervised learning, yet their extensive parameters often lack external knowledge. To bridge this gap, Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) have emerged, aiming to integrate external information. This review thoroughly explores the development of KE-PLMs in Natural Language Understanding (NLU), incorporating linguistic, textual, KG, and rule-based knowledge. Additionally, in Natural Language Generation (NLG), it investigates the utilization of KG-based and retrieval-based techniques. Moreover, the exploration emphasizes advancements propelling KE-PLMs toward superior language comprehension and generation. These strides are focused on refining KE-PLMs, enhancing their language processing capabilities to achieve more sophisticated levels of understanding and generation.

Keywords – natural language production and pre-trained language models, natural language processing, continual learning Domain-specific knowledge, Efficient fine-tuning, User-friendly interfaces, Robustness and security.

1. INTRODUCTION

[1] Recent years have seen a broad acceptance of Pre-trained Language Models (PLMs) in Natural Language Processing due to the ongoing development of deep learning technology (NLP). Self-supervised learning is used by models like BERT [1], GPT [2], and T5 [3], which pre-train on large amounts of unlabeled data and then fine-tune on smaller amounts of labelled data. These PLMs have excelled in various NLP tasks, marking a paradigm shift from supervised to self-supervised learning. As PLMs grow in size, those with hundreds of millions of parameters showcase remarkable capabilities in capturing linguistic nuances and factual knowledge. However, their weak symbolic reasoning capabilities derive from

the lack of explicit knowledge representation in raw data, which limits their performance on downstream tasks [9]. Recognizing this limitation, researchers emphasize the incorporation of knowledge into PLMs to enhance memorization and reasoning abilities [10]. This observation underscores the need to augment PLMs with external knowledge for more effective language understanding and reasoning [10].

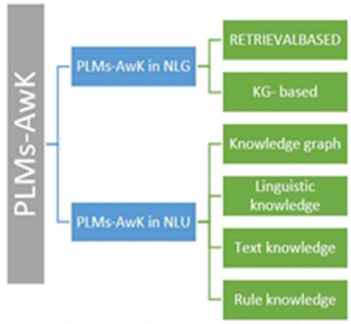
[2] Existing PLMs often lack human commonsense due to their oversight of external world knowledge [11]. Recent studies advocate for the explicit incorporation of knowledge into PLMs [12], with classifications based on knowledge sources, granularity, and applications [12]. This survey provides a thorough analysis of various knowledge sources, tasks, and fusion techniques with a focus on Knowledge-Enhanced Pre-trained Language Models (KE-PLMs) and their effects on NLU and NLG.Our taxonomies for NLU and NLG, depicted in Fig. 1, categorize KE-PLMs based on knowledge types for language understanding and retrieval-based and KG-based methods for language generation. Furthermore, we discuss potential research directions to address current challenges and advance KE-PLMs' capabilities. By considering future innovations, we contribute to the ongoing development of KE-PLMs for more robust language understanding and generation in NLP [13]. The paper concludes by outlining PLMs and training paradigms, presenting taxonomies for KE-PLMs in NLU and NLG, discussing representative works, and proposing future research directions [1-3, 9-13].

[3] 2. BACKGROUND

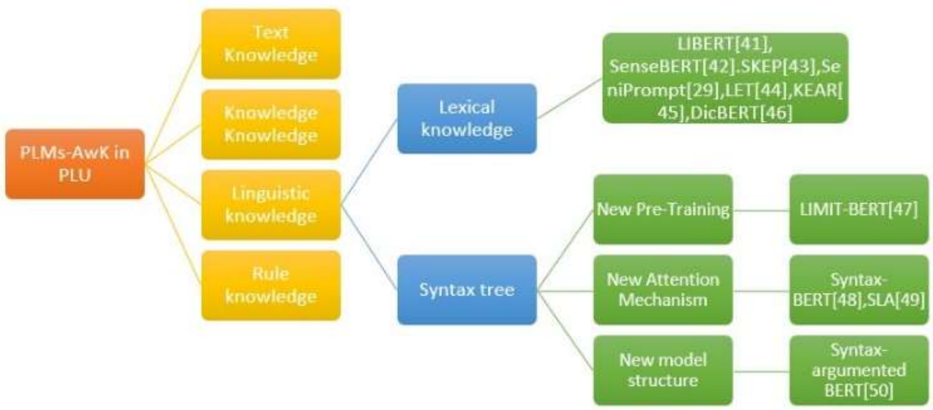
[4] Pre-trained Language Models (PLMs) have undergone a transformative shift in natural language processing (NLP), moving from traditional language modeling to a pre-train and fine-tune methodology. This evolution is exemplified by models like ELMo and ULMFiT, which leverage LSTM architecture and emphasize layer-by-layer fine-tuning for downstream tasks. Departing from early data-constrained supervised methods, PLMs now undergo extensive training on unprocessed textual data to acquire adaptable, all-purpose representations.

[5] The pre-train and fine-tune paradigm involves initial pre-training on large amounts of unprocessed textual data, followed by fine-tuning with task-specific objectives during the refining stage. UNILM is one such model that integrates multiple language modeling objectives, ensuring flexibility in tasks related to natural language generation (NLG) and comprehension (NLU). This approach, with its focus on layer-by-layer fine-tuning, enhances

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904



[6] Figure 1"Classification of KE-PLMs by NLU and NLG tasks in NLP."
[7]



[8] Figure 2 Linguistic Knowledge Categorization

[9] the generalization capabilities of models and makes them competitive across a spectrum of natural language processing applications.

[10]

[11] Transformer architecture with multi-head self-attention is the cornerstone of many contemporary language models, such as GPT-2, BERT, BART, and T5. This approach helps these models perform well in a variety of tasks by allowing the capturing of long-range relationships and the creation of expressive representations. While BERT predicts words bidirectionally using masked language modeling, GPT generates sequences sequentially. BART and T5 employ encoder-decoder architectures to create sequences based on input sequences.

[12] By proposing a fresh approach in NLP, prompt learning challenges the traditional pre-train and fine-tune paradigm. The "pre-train, prompt, and predict" paradigm bypasses fine-tuning, allowing PLMs to predict outputs based on textual prompts. This departure from traditional methods aims to overcome constraints and align objectives, presenting a transformative shift in NLP.

[13] Despite the promising outcomes of prompt-based learning, challenges persist in selecting optimal prompt verbalizers and templates, influencing model performance. Recent endeavors advocate for leveraging knowledge as a prompt during fine-tuning, infusing

China Petroleum Processing and Petrochemical Technology
Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

domain/task-specific knowledge to enhance PLMs' adaptability and performance in downstream tasks. This strategic move signifies an elevation of the prompt-tuning process.

[14] The survey delves into the realm of prompt-based learning, exploring its landscape and revealing both progress and challenges. The infusion of domain-specific knowledge during fine-tuning is highlighted as a strategic enhancement, contributing to heightened performance across various NLP tasks. The selection of optimal prompt verbalizers and templates remains a critical consideration, and recent efforts emphasize leveraging knowledge as a prompt to address these challenges.

[15] Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) represent a rapidly evolving field. This paradigm shift reflects ongoing efforts to incorporate external knowledge in innovative ways. The survey investigates how injecting relevant knowledge during the fine-tuning stage enhances model adaptability and performance, shaping the future of PLMs in the dynamic field of natural language processing.

[16]

3. KE-PLMS FOR NLU

Decoding textual content, the machine is driven by Natural Language Understanding (NLU). Guidance is essential for tasks like text categorization and NLU (Named Entity Recognition, NLU, and Natural Language Understanding). To facilitate a comprehensive comprehension of natural language, Figure 1 illustrates how Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) integrate linguistic, text, information graph, and rule knowledge for natural language understanding tasks.

3.1 Integrating Information About Linguistics

Linguistic knowledge plays a pivotal role in augmenting Pre-trained Language Models (PLMs), encompassing lexical knowledge and syntax tree information [40]. LIBERT introduces Lexical Relation Classification (LRC) within the BERT framework, predicting semantic relations by utilizing synonyms and hypernym-hyponym pairs [41]. SenseBERT integrates word-supersense information to predict corresponding supersenses, enriching semantic understanding [42]. SKEP enhances sentiment analysis in PLMs by incorporating sentiment knowledge, including sentiment words and aspect-sentiment pairs [43]. SentiPrompt takes a step further by integrating sentiment knowledge about aspects, opinions, and polarities into prompts, enhancing task-related knowledge through prompt-tuning methods [28]. LET leverages HowNet's semantic information for Chinese sentence matching tasks [44], while KEAR excels in commonsense knowledge question answering by combining ConceptNet, dictionary entries, and labeled training data [45]. DictBERT adopts a unique strategy by employing dictionary knowledge as an external source, enhancing knowledge through contrastive learning [46].

In the realm of Knowledge-Enhanced Pre-trained Language Models (KE-PLMs), syntax tree knowledge integration employs diverse strategies. LIMIT-BERT employs multi-task learning, amalgamating task-specific losses related to linguistic knowledge during model training [47By using a syntax tree parser and a novel attention strategy, Syntax-BERT improves syntactic information [48]. By combining syntax information through a graph-based structure, syntax-augmented BERT presents a syntax-based graph neural network [50]. These approaches showcase the versatility of KE-PLMs in accommodating linguistic structures. The integration of linguistic knowledge occurs at different stages: LIBERT, SenseBERT, Fig.2 SKEP, LIMIT-BERT, Syntax-BERT, and DictBERT fuse linguistic knowledge during the pre-training stage [41, 42, 43, 47, 48, 46], while SentiPrompt, LET, SLA, Syntax-augmented BERT, and KEAR incorporate knowledge during the fine-tuning stage, aiming to improve task performance [28, 44, 49, 50, 45].

3.2 Infusing Textual Knowledge into PLMs

Pre-trained Language Models (PLMs) derive textual knowledge from diverse sources, including general-domain text collections like Wiki Text and Wiktionary, along with extensive corpora such as Wikipedia. KNN-LM incorporates knowledge by selecting the nearest neighbors from training samples, inspired by cache-LM. REALM utilizes a text retriever trained on a corpus, extracting information from external knowledge bases like Wikipedia to predict masked tokens. Textual description integration improves performance in models such as ExpBERT and KEAR. Using a novel strategy, OK-Transformer integrates extensive out-of-domain commonsense descriptions from the ATOMIC2020 knowledge base and uses Transformer to fuse them with input text. Kformer retrieves text knowledge from external sources, embedding it into the FeedForward Network (FFN) layer of the Transformer. REINA leverages knowledge by retrieving training samples similar to the input from external datasets, while UniK-QA, UDT-QA, and KiC integrate text, knowledge graph, and table knowledge. For domain-specific knowledge, BioBERT and SciBERT focus on scientific domain corpora for pre-training tasks. S2ORC-BERT employs a methodology similar to SciBERT but operates on a larger corpus covering diverse academic papers, contributing to enhanced performance on various downstream tasks associated with academic literature. In the knowledge fusion process, While KNN-LM, ExpBERT, KEAR, OK-Transformer, Kformer, REINA, UniK-QA, and UDT-QA incorporate knowledge during the fine-tuning stage, optimizing models for particular downstream tasks, REALM, SciBERT, BioBERT, and S2ORC-BERT integrate text knowledge during pre-training [51–63].

3.3 Infusing Textual Knowledge into PLMs

A knowledge graph, in which nodes stand for entities and edges for relations, is a powerful tool for representing real-world data in a graph structure. [64], [65]. Unlike unstructured text knowledge, knowledge graphs offer richer structured information, making them valuable for enhancing the learning capacity of models [66], [67], [68], [69]. Within the knowledge graph, we distinguish entity knowledge and triplet knowledge, as illustrated in Fig. 3. This section offers a sophisticated interpretation of structured data, highlighting its ability to enhance language models' capabilities.

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

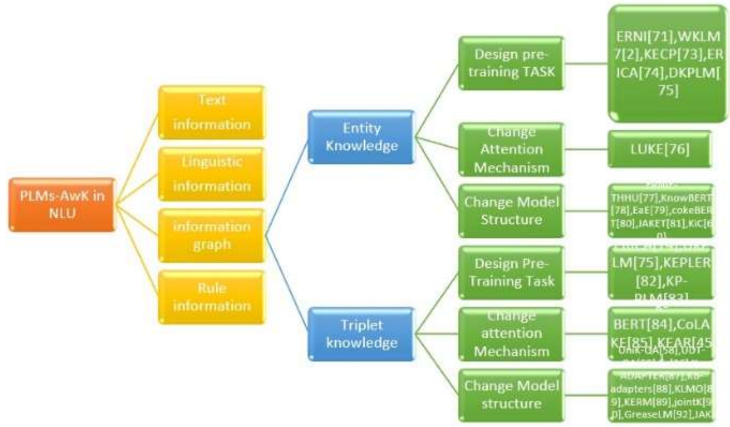


Figure 3:Information Graphs: Categories and Methods for Integration

3.3.1 Entity Knowledge

Knowledge graphs include several relational triples in addition to entity knowledge, which provide structured information to improve the semantic comprehension of Pre-trained Language Models (PLMs) and make them stronger. Techniques for triplets in KE-PLMs can be divided into three sub-types, much like entity knowledge incorporation.

Triplet pre-training assignments, such as ERICA [74], incorporate entity and relation discrimination exercises to improve PLM understanding via contrastive learning. KEPLER [82] elevates knowledge representation by teaching veiled language modelling aims and knowledge embedding at the same time.

Method	Pre -training Tasks /Objectives
ERNIE [71]	Token -level ,Phrase -level and Entity -level MLM
WKLM [72]	Entity replacement ,MLM
KECP [73]	Token -level MLM ,Span -level contrastive learning
ERICA [74]	Entity and relation discrimination tasks ,MLM
DKPLM [75]	Relational knowledge decoding ,Token -level MLM
KP -PLM [83]	Prompt relevance inspection ,Masked prompt modeling
KEPLER [82]	Knowledge embedding ,MLM

Table 2: Pre-training Tasks for Entity/Triplet with MLM (Masked Language Modeling)

DKPLM [75] substitutes long-tail entities with knowledge triplets by using a fig .4 pre-training activity to predict the relations and entities that will be replaced. In addition to producing pertinent knowledge sub-graphs for phrases and knowledge-based pre-training activities, KP-PLM [83] also provides knowledgeable prompts. These prompts are produced from mapping rules and linked with the original input to provide PLMs with a single input.

Pre-trained Language Models (PLMs) are improved by a variety of approaches that change model structures, add external knowledge, and tweak attention mechanisms. As an example, KEAR [45] suggests an external attention mechanism that integrates external knowledge into the Transformer design, as seen in Table 2. Model structure alterations involve the introduction of knowledge

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904
fusion modules, as shown by the works of FaE [86], KLMO [89], KB-adapters [88], KERM [90], and K-ADAPTER [87]. Using unified architectures.

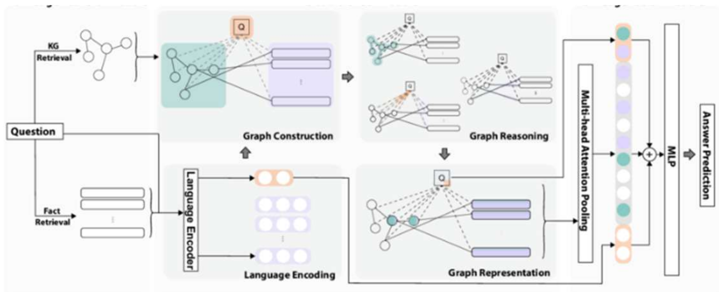


Figure 4: Overview approach involves three modules: Knowledge Retrieval, Double Check, and Knowledge Fusion. It retrieves relevant local KG and facts, enhances heterogeneous knowledge, and deduces the final answer prediction.[153]

3.3.2 Triplet Knowledge

Pre-trained Language Models (PLMs) have contributed significantly to the advancement of Natural Language Processing (NLP), but knowledge graphs' structured knowledge can be used to maximize their potential, particularly entity and relational triples. This integration is systematically classified into three sub-categories, each contributing unique methodologies to augment PLMs.In the first category, pre-training tasks pertaining to triplets are designed to specifically target entity and relational triples. ERICA introduces discrimination tasks for entities and relations, intensifying PLMs' comprehension through contrastive learning. KEPLER takes a comprehensive approach, simultaneously to improve knowledge representation, masked language modelling goals and training knowledge embedding are used. DKPLM, on the other hand, concentrates on long-tail entities and augments semantic information by pre-training them with appropriate knowledge triplets. KP-PLM is a leader in knowledge-aware pre-training tasks, enhancing performance on a variety triplet into natural language cues with ease, improving performance in Natural Language Understanding (NLU) challenges.

The second group is concerned with altering PLMs' attention mechanisms. To avoid meaning changes, K-BERT incorporates a knowledge layer into the input sentence to include relevant triplets. CoLAKE extends the input context of word-knowledge graphs and leverages masked self-attention to extract insightful information. KEAR offers an external attention mechanism, enhancing the Transformer architecture's possibilities. The PLM model's structural modifications fall under the third category. In order to efficiently combine information from symbolic knowledge graphs, FaE has a facts memory module. External adapter modules are used by K-ADAPTER and KB-adapters for smooth knowledge inclusion, while a knowledge aggregator component is used by KLMO for interactive modelling. A knowledge injector module is introduced by KERM for activities like passage re-ranking. Graph Neural Networks (GNNs) and PLMs are coupled in JointLK and GreaseLM to enable joint reasoning in commonsense reasoning. JAKET introduces a comprehensive joint training framework, On the other hand, for open-domain question answering, UniK-QA and UDT-QA use unified knowledge representation architectures. Significant progress in integrating triplet knowledge into PLMs has been made with Know Prompt and Onto Prompt, which incorporate entity and relation knowledge into prompts. These programmers highlight emerging research trends in the application of AI and ML [74, 82, 75, 83, 84, 85, 45, 30, 31, 58, 59, 81, 86, 87, 88, 89, 90, 91, 92].

Table 1:Examples of Modified Attention for Triplet Knowledge Integration.

Method	Pre -training Tasks /Objectives
ERNIE [71]	Token -level ,Phrase -level and Entity -level MLM
WKLM [72]	Entity replacement ,MLM
KECP [73]	Token -level MLM ,Span -level contrastive learning
ERICA [74]	Entity and relation discrimination tasks ,MLM
DKPLM [75]	Relational knowledge decoding ,Token -level MLM

3.3.3 Fusion stage

Pre-fusion is a large category of tactics that pre-trained Language Models (PLMs) benefit from and post-fusion methods, with some models adopting a hybrid approach combining both stages. In the pre-fusion category, exemplified by ERNIE [71], WKLM [72], LUKE [76], and CoLAKE [85], knowledge integration occurs during the pre-training stage. Conversely, post-fusion methods like K-BERT [84], K-ADAPTER [87], and JointLK [91] introduce knowledge during the fine-tuning stage. ERNIE-THU [77] and KnowBert [78] exemplify hybrid models, adeptly integrating knowledge across pre-training and fine-tuning phases. Notable examples include CoLAKE [85], engaging in joint learning of entity and relation embeddings during the training phase, and K-BERT [84], introducing knowledge graph triples during the reasoning phase. This array of fusion methods collectively bolsters PLMs, fortifying their language understanding and reasoning capacities for a range of natural language processing tasks [71, 72, 76, 85, 84, 87, 91, 77, 78].

3.4 Rule Knowledge in PLMs

Incorporating clear logical reasoning processes from logic rules, which formalize knowledge from external sources, into Pre-trained Language Models (PLMs) enhances interpretability. For instance, RuleBERT [94] leverages Horn rules from existing corpora, creating a training dataset and subsequently fine-tuning the model. Employing probabilistic answer set programming, it predicts event probabilities and endeavors to learn soft rules from PLMs, demonstrating improved deductive reasoning performance. In the realm of incorporating logic rules, PTR [27] stands out for integrating them during fine-tuning through manually created sub-prompts within task-specific prompts. More interpretable prompts result from this technique. This additionally allows the model to encode prior information relevant to a given job. [94, 27].

Table 3: Comparisons of PLMs and KE-PLMs on LAMA and LAMA-UHN Benchmarks

Method	LAMA		LAMA-UHN	
	Google-RE	T-REx	UHN-Google-RE	UHN-T-REx
ELMo [14]	2.2	0.2	2.3	0.2
BERT [1]	11.4	32.5	5.7	23.3
EaE [79]	9.	37.		
CoLAKE [85]	10.	29.	5.	20.
DKPLM [75]	11.	32.0	5.	23.
KP-PLM [83]	11.0	32.	6.	23.

Table 4: Evaluation of KE-PLMs on TACRED and OpenEntity benchmark

Method	TACRED			OpenEntity		
	P	R	F1	P	R	F1
ERNIE-THU [77]	70.	66.	68.	78.	72.90	76.
KnowBert [78]	71.60	71.40	71.50	78.60	73.70	76.10
KEPLER [82]	71.50	72.50	72.00	77.80	74.60	76.20
CoLAKE [85]				77.00	75.70	76.40
K-ADAPTER [87]	70.	74.	72.	79.	76.	78.
DKPLM [75]	73.	74.	73.	79.20	75.90	77.50

3.5 NLU Benchmarks Overview

The Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) collection assesses Natural Language Understanding (NLU) tasks that are cognizant of previous information by testing their performance on pre-established benchmarks Table 5 and specialized datasets. The goal of these assessments is to offer a comprehensive picture of KE-PLM performance across a range of language areas.

The General Language Understanding Evaluation, or GLUE [95], is a baseline measure that includes nine NLU tasks such as sentiment analysis and textual entailment. Super GLUE [96] is an expansion of it that presents eight tasks with increasingly complex language understanding problems, pushing models to higher comprehension levels.

Using cloze-style statements, LAMA (Language Model Analysis) [7] evaluates factual knowledge in language models in a knowledge-centric manner. It evaluates models for a range of tasks that are intended to weed out examples with simple answers.

A large supervised relation extraction dataset called TACRED [98] is used to evaluate models' ability to extract relations from text. The FewRel [99] few-shot relation classification dataset is used to evaluate a model's capacity to categories relations using few training examples.

While CoNLL-2003 [102] concentrates on named entity recognition, Open Entity [100] and FIGER [101] evaluate entity typing. Commonsense QA [103] and OpenBook QA [104] assess models' commonsense reasoning skills through multiple-choice question responding.

WebQuestions [106] concentrates on Freebase entities, while

Task	Dataset	# Train	# Dev.	# Test	Model
Relation classification /extraction	TACRED [98]	75,050	25,764	18,660	DictBERT [46], ExpBERT [541, ERICA [74], DKPLM [75], KnowBert [78], KP-PLM [83], LUKE [76], KEPLER [82], ERNIE-THU [77], K-ADAPTER [87], PTR [27]
	FewRel [99]	4,48,000	1,12,000	1,40,000	KP-PLM [83], CoLAKE [85], ERNIE-THU [77], JAKET [81], KEPLER [82]
Entity ping typing	OpenEntity [100]	2,0(10	2,000	2,000	DKPLM [75], LUKE [76], ERNIE-THU [77], KnowBert [78], KEPLER [82], KP-PLM [83], CoLAKE [85], K-ADAPTER [87]
	FIGER [101]	20,00,000	10,000	563	WKLM [72], ERICA [74], K-ADAPTER [87]
Named entity recognition	CoNLL-2003 [102]	2,03,621	51,362	46,435	DictBERT [46], SLA [49], LUKE [76]
Commonsense reasoning	CommonsenseQA [103]	9,741	1221 ,	1,140	KEAR [45], DictBERT [46], OK-Transformer [55], JointLK [91], GreaseLM [92]
	OpenBookQA [104]	4,957	500	500	DictBERT [46], JointLK [91], GreaseLM [92]
	Natural Questions [105]	3,07,373	7,830	7,842	REALM [52], UniK-QA [58], UDT-QA [59], ERICA [74]
Question answering	WebQuestions [106]	3,778		2,032	REALM [52], UniK-QA [581, UDT-QA [59], WKLM [72], EaE [79], FaE [86]
	TriviaQA [107]	87,291	11,274	10,790	UniK-QA [58], WKLM [72],ERICA [74], EaE [79]
Sentiment classification	SST-2 [108]	8544	1101	2210	SKEP [43], Syntax-BERT [48], SLA [49]
Knowledge probing	LAMA [7]				DKPLM [75], EaE [79], KP-PLM [83], CoLAKE [85]

Table 5: Benchmarking Knowledge-Grounded NLU Datasets with Variable Splits in Studies

Natural Questions [105] tests models using open-domain queries from the Google search engine. With trivia and quiz-league questions, TriviaQA [107] expands on open-domain question-answering.

The Stanford Sentiment Treebank, or SST-2 [108], evaluates how well models capture the subtleties of sentiment at the phrase level. This comprehensive assessment provides an overview of the capabilities of KE-PLMs [95-108].

4. PLMS-AWK FOR NLG:

Making it possible for machines to produce understandable linguistic documents that resemble human expressiveness is the aim of natural language generation, or NLG. KE-PLMs use a variety of knowledge forms other than input sequences to increase text production efficiency. These strategies, which offer a sophisticated approach to Natural Language Generation challenges, fall into two categories: KG-based and retrieval-based techniques. They are informed by a knowledge-enhanced text generation survey [32]. This classification offers a formalized summary of how KE-PLMs use external knowledge, obtained from retrieval methods or knowledge graphs, to enhance language generation models and improve their ability to produce coherent, contextually relevant, and human-readable texts.

4.1 Retrieval-Based Knowledge Integration with PLMs

In order to advance Pre-trained Language Models Augmented with Knowledge (PLMs-AwK), retrieval-based techniques are essential because they enable skilled retrieval mechanisms to add extra knowledge to input sequences [32]. In addition to the input, they retrieve additional data from a variety of outside sources, such as training sets, big databases, and internet search engines. The main objective is to direct the process of creating content and improve the produced output's contextual relevance. Retrieval augmented generation techniques, which priorities retrieving materials and rearranging them for generation, are part of these approaches. Retrieval augmented generation techniques improve generation by integrating acquired knowledge [49, 85-92].

These sub-methods are graphically categorized in Figure 5, which also shows the different tactics used in each. Figure 6 illustrates how KE-PLMs integrate outside knowledge during the fine-tuning stage, enhancing their performance on a range of downstream

In the retrieval augmented generation space, MemNet [85] creates the first Transformer memory network that can retrieve significant conversational data from the past pertaining to a certain topic, enhancing the subsequent words that are stated. During prediction creation, RAG [86] uses a retriever to identify the top K relevant documents and seamlessly combines them for more data. KFCNet [87] sets itself apart by obtaining prototypes that capture ideas from a given set and ensuring that they are semantically consistent with target sentences. This model captures global target information and effectively extracts general features from recovered models by incorporating contrastive learning modules into both the encoder and the decoder. To improve machine translation, REINA (49), selects training examples that are similar to the input text KGR4 [88] utilizes. This improvement substantially raises the bar for open-domain question answering ability. See A three-module system (search, knowledge production, and final answer) is used by KeR [90] to produce knowledge from articles retrieved through search engine results. The generated responses are timelier and more relevant when current data is used, which sets them apart from the competitors [49, 88-90]. In terms of retrieval strategies for knowledge-enhanced text synthesis, RETRO [91] creatively gathers information from vast text databases using trillions of tokens as retrieval inputs to enrich the language model. In order to improve performance on these diverse tasks, Unified SKG [92] employs a comprehensive technique that combines six task categories into a text-to-text format and introduces linearized knowledge. The various abilities of knowledge-enhanced pre-trained language models (KE-PLMs) to boost language creation processes are demonstrated by these creative retrieval strategies.

In The domain of the retrieve, rerack, and rewrite process SKT [93] is one of the notable studies that views the process of choosing information as a sequence of choices. SKT makes use of a A sequential latent variable model can be used to gradually increase the accuracy of knowledge selection through multiple discussion rounds. PLUG [94] uses a multimodal retrieval approach, gathering relevant information from dictionaries, Wikipedia, and knowledge graphs, among other sources. PLUG then uses both statistical and semantic criteria to rank this material, laying the groundwork for knowledge-grounded discourse production [91–94]. The diverse array of retrieval-based methods showcases several methods for integrating external knowledge and highlights the adaptability of these models for an array of natural language production uses. The subtle methods used by these

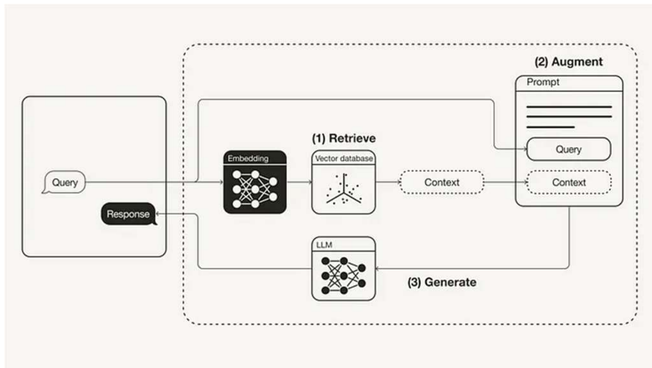


Figure 6: Retrieval-Augmented Generation Workflow

4.2 Integrating Knowledge Graphs into Pre-trained Language Models (PLMs):

In order to categories the degree of detail in We categories the information found in Pre-trained Language Models Enhanced with Knowledge (PLMs-AwK) into three groups, as indicated on the right: understanding acquired through Information Graphs (KG): triplet knowledge, subgraph knowledge, and knowledge obtained through path finding. Figure 5 In the first category, route discovery is used to extract information [96, 97], and careful consideration of the related path is made to facilitate reliable decision-making. This detailed classification offers a thorough grasp of how KE-PLMs use various types of knowledge to improve their capacity for language comprehension and reasoning [96–97]. This technique is demonstrated by Fig. 7 CE-PR [96] and MRG [97], which engage in explicit reasoning on relation routes, resulting in a significant increase in text output efficiency. To maintain nodes with greater scores, CE-PR creates a subgraph based.

In terms of subgraph integration, certain KE-PLMs (e.g., GRF [101]) include this knowledge into the decoder. This method improves the interpretability of the model by allowing the tracking of each decoding step. Fig.8 BART increases language understanding and creation across several phases by introducing subgraph information into both the encoder and the decoder [11].

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

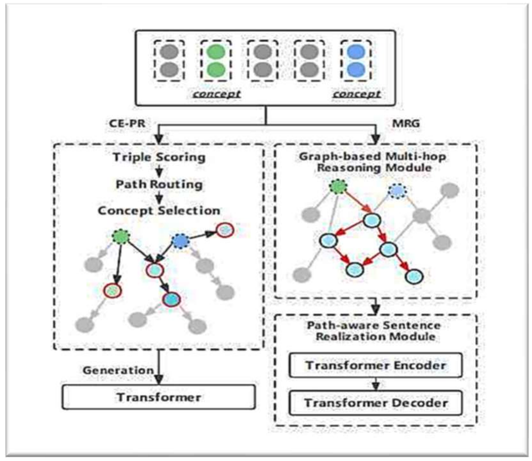


Fig.7: Comparison between MRG and CE-PR using information gleaned via path finding.

Use pre-fusion strategies for Natural Language Generation in the context of methods that include data before instruction, such as JointGT [102], PLUG [94], and UnifiedSKG [92]. (NLG). On the other hand, post-fusion techniques like KG-BART [11], MoKGE [103], MixGEN [99], CNTF [100], MRG [97], KEPM [98], GRF [101], and CE-PR [96] all include knowledge when fine-tuning. The contrast between these two integration strategies provides information on the many ways that KE-PLMs use knowledge to improve NLG performance [11, 92].

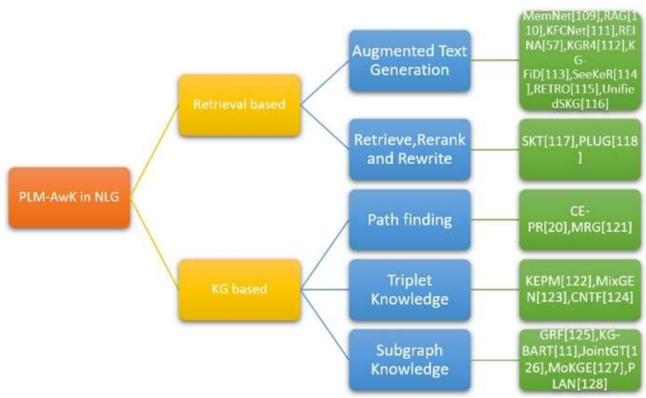


Figure 5: Refined Classification of Retrieval-Based & KG-Based Methods.

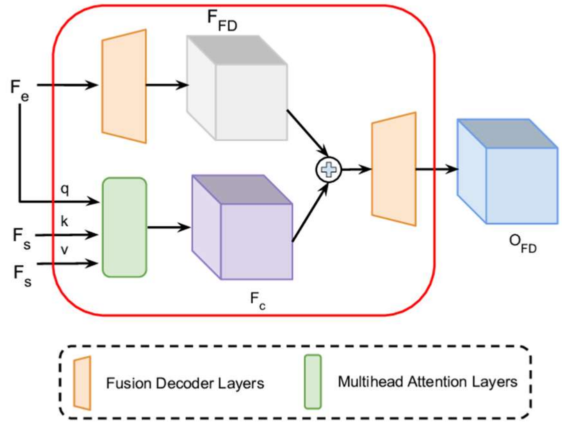


Figure 8 The fusion decoder, enclosed in a red box, combines encoder feature map (Fe) and self-attended decoder volume (Fs) using multi-layered cross-attention, producing CGL segmentation mask (OF D).

4.3 NLG Benchmarks

Establishing comprehensive benchmarks for evaluation the topic of natural language generation (NLG) requires further research. Although well-known benchmarks like as GLGE [136] and KilT [137] offer insightful information, it's possible that they aren't explicitly designed to meet the needs of knowledge-enhanced generation [40]. Table 6 shows that pre-trained Language Models Augmented with Knowledge (PLMs-AwK) are frequently used for particular NLG tasks, which means that a summary of pertinent dataset benchmarks is required. To assess the effectiveness of Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) on a range of natural language tasks, for example, datasets are crucial. Benchmark datasets include the Wizard of Wikipedia [109], CMU DoG [129], Natural Questions [105], TriviaQA [107], and CommonGen [130], which evaluate KE-PLMs on a range of linguistic tasks to demonstrate their adaptability and skill [109, 129, 105, 107, 130].

Table 6: Benchmarking Knowledge-Grounded NLG Datasets with Variable Splits in Studies

Task	Dataset	# Train	# Dev.	# Test	Model
Dialogue generation	Wizard of Wikipedia [109]	18,430	1,948	1,933	MemNet [109], CNTF [124], PLUG [118], SKT [117]
	CMU_DoG [129]	3,373	229	619	CNTF [124]
Question answering	Natural Questions [105]	3,07,373	7,830	7,842	KG-FiD [113], RAG [110], RETRO [115]
	TriviaQA [107]	87,291	11,274	10,790	KG-FiD [113], RAG [110]
Commonsense reasoning	CommonGen [130]	67,389	4,018	6,042	KG-BART Mk KFCNet [111], KGR4 [112]
	aNLG-ART [131]	50,481	7,252	2,976	GRF [125], MoKGE [127]
Explanation reasoning	ComVE [132]	25,596	1,428	2,976	GRF [125], MoKGE [127], CE-PR [120]
Story generation	ROCStories [133]	90,000	4,081	4,081	KEPM [122], GRF [125]
Summarization	CNN/Dailymail [134]	2,87,226	13,368	11,490	REINA [57]
	XSum [135]	2,04,045	11,332	11,334	REINA [57]

5. PROSPECTIVE PATHS

This section describes prospective avenues for KE-PLMs research in the future, with the goal of resolving persistent problems and difficulties in the field. The suggested paths are meant to advance the research and development of PLMs-AwKs, or Pre-trained Language Models Augmented with Knowledge, resulting in new discoveries and enhanced capabilities.

5.1 Knowledge Fusion from Various Sources

China Petroleum Processing and Petrochemical Technology

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

There is a lot of potential for investigating the integration of knowledge from varied and heterogeneous sources because current approaches primarily rely on knowledge from unique sources, like knowledge graphs or web repositories. In order to improve the capabilities and performance of language models on a variety of tasks, future research should give priority to developing strategies that efficiently combine insights from different knowledge stores. UniK-QA [58] integrates text, tables, and relational triplets, harmonizing structured KBQA knowledge with open-domain unstructured information. UDT-QA [59] brings structured knowledge, including graphs and tables, into linear sequences for text generation. Pre-trained Language Models (PLMs) can respond to open-domain questions more reliably and with greater coverage of the subject matter when they are fed a diverse range of information sources.

TABLE 7
Comparisons between PLMs and KE-PLMs on CommonGen benchmarks.

Method	BLEU-3/4		ROUGE-2/L		METEOR	CIDEr	SPICE
T5 [3]	39.00	28.60	22.01	42.97	30.10	14.96	31.60
BART [20]	36.30	26.30	22.23	41.98	30.90	13.92	30.60
KG-BART [11]	42.10	30.90	23.38	44.54	32.40	16.83	32.70
KFCNet [111]	57.33	51.46	26.81	47.52	38.92	20.98	39.15
KGR4 [112]	-	42.82	-	-	-	18.42	39.70

TABLE 8
Evaluation of KE-PLMs on Natural Questions, WebQuestions, and TriviaQA benchmarks. Exact match scores are reported.

Method	# params	Natural Q.	Web Q.	TriviaQA
T5 [3]	11,318M	34.5	37.4	50.1
EaE [79]	367M		39.0	53.4
REALM [52]	330M	40.4	40.7	
UniK-QA [58]	990M	54.0	57.8	64.1
UDT-QA [59]	1,320M	55.2	52.0	-
RAG [110]	626M	44.5	45.2	56.8
KG-FiD [113]	994M	53.4	-	69.8

5.2 Two-Modal Knowledge Integration Exploration

Although most of the research being done presently focuses on textual knowledge, there is a growing recognition of the potential that remains unexplored in multi-modal sources. Beyond textual and tabular data, photos, videos, and audio appear as unexplored knowledge sources for Pre-trained Language Models (PLMs), offering an opportunity to further enhance the capabilities of PLMs-AwK, or Pre-trained Language Models Augmented with Knowledge.

ERNIE-VIL [140] and KB-VLP [139] were the first to integrate multi-modal knowledge. By utilising both text and images from other sources, KB-VLP improves semantic alignment. In order to achieve precise semantic alignment across vision and language modalities, ERNIE-VIL parses input descriptions and fine-tunes models.

Combining various modalities together, including written explanations and visual aids like pictures, encourages complementarity and synergy, where each modality enhances the understanding provided by the others.

5.3 Presenting Evidence for Interpretability

While KE-PLMs do well in a variety of text creation tasks, they struggle with tasks that call for commonsense knowledge reasoning. Projects such as GRF [125] mitigate this problem by using external knowledge graphs to support explicit commonsense reasoning. GRF stands out for incorporating comprehensive structural information, enabling dynamic multi-hop reasoning across various relational paths. This not only offers a theoretical basis for result generation but also emphasizes the significance of explicit reasoning paths to enhance model interpretability and ensure more rational predictions.

5.4 Continuous Knowledge Learning in Model

Contemporary models, primarily trained on static or non-updated data during pre-training, confront a significant challenge known as the catastrophic forgetting problem when confronted with new tasks, wherein initial knowledge tends to be overlooked [141]. As information from diverse sources continues to expand, there is a demand for strategies enabling models to assimilate new information while retaining their existing knowledge—a paradigm referred to as continual learning or life-long learning.

In response to this challenge, ELLE [142] presents an extension module that expands the breadth and depth of the model, allowing for the efficient learning of new information while maintaining the accuracy of preexisting knowledge. K-ADAPTER [87] and KB-adapters [88] incorporate adapters into pre-trained language models concurrently (PLMs) to store both factual and linguistic knowledge, ensuring the continual assimilation of additional information. Prioritizing continuous knowledge integration stands out as a promising research avenue [40]. The adoption of continuous and expanding pre-training methodologies holds the ability to increase PLMs' universality, solve the catastrophic forgetting issue, and guarantee that these models adapt to new information. Strategies for continuous learning emerge as crucial in optimizing PLMs for evolving tasks and knowledge landscapes, emphasizing the necessity for adaptability and sustained knowledge incorporation.

5.5 Efficiency in Knowledge Integration for Large Models

The proliferation of knowledge and trained models' integration has led to heightened computational challenges [143]. Although there has been improvement in the pre-training duties, there is a notable lack of focus on the computational implications of knowledge fusion. To tackle this issue, we suggest two investigational paths: firstly, improving efficiency in knowledge acquisition and filtering, and secondly, optimizing the computational load. These approaches present promising strategies for addressing the escalating computational requirements linked to the increasing the size of previously trained models and introducing new information.

5.6 Enhancing Result Diversity

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

Diversifying output creation is a crucial task in the field of Natural Language Generation (NLG)., especially in generative commonsense reasoning, is addressed by MoKGE [127]. This method strategically employs diversified knowledge reasoning sourced from commonsense knowledge graphs, enhancing NLG diversity. By integrating concepts linked to the initial input, guided by human annotations, MoKGE employs a mixture of expert methods to generate diverse and plausible outputs. This approach significantly contributes to amplifying the range of results in NLG tasks pertaining to generative commonsense reasoning.

Table 9: KE-PLMs Summarized: RC, RE, ET, NER, QA,

Method	Knowledge Type	Fusion in Pre-training	Fusion in Fine-tuning	NLU or NLG tasks
LIBERT [41]	lexical	Yes		lexical simplification, sentence/sentence-pair classification, NLI
SenseBERT [42]	lexical	Yes		word supersense disambiguation, WiC
SKEP [43]	lexical	Yes		sentence/aspect-level sentiment classification, opinion role labeling
SentiPrompt [28]	lexical		Yes	triplet extraction, pair extraction, aspect term extraction
LET [44]	lexical		Yes	Chinese short text matching
KEAR [45]	lexical, general text, triplet		Yes	commonsense reasoning
DictBERT [46]	lexical	Yes	Yes	NER, RE, commonsense reasoning, GLUE
LIMIT-BERT [47]	syntax tree	Yes		syntactic parsing, semantic parsing
Syntax-BERT [48]	syntax tree	Yes		sentiment classification, NLI, GLUE
SLA [49]	syntax tree		Yes	sentiment classification, NER, grammatical error detection
Syntax-augmented BERT [50]	syntax tree		Yes	semantic role labeling, NER, RE
KNN-LM [51]	general text		Yes	language modeling
REALM [52]	general text	Yes		open-QA
ExpBERT [54]	general text		Yes	RE
OK-Transformer [55]	general text		Yes	commonsense reasoning, text classification
Kformer [56]	general text		Yes	commonsense reasoning, medical QA
UniK-QA [58]	general text, triplet		Yes	multi-source QA
UDT-QA [59]	general text, triplet		Yes	open-domain QA
SciBERT [62]	domain-specific text	Yes		sequence tagging, sentence classification, dependency parsing
BioBERT [61]	domain-specific text	Yes		biomedical NER, RE, QA
S2ORC-BERT [63]	domain-specific text	Yes		inline citation detection, bibliography parsing, bibliography linking
ERNIE [71]	entity	Yes		NLI, semantic similarity, NER, sentiment analysis, QA
WKLM [72]	entity	Yes		QA, ET
ERICA [74]	entity, triplet	Yes		RE, ET, QA
KECP [73]	entity		Yes	extractive QA
DKPLM [75]	entity, triplet	Yes		knowledge probing, RE, ET
LUKE [76]	entity	Yes		ET, RC, NER, cloze-style/extractive QA
ERNIE-THU [77]	entity	Yes	Yes	ET, RC, GLUE
KnowBert [78]	entity	Yes	Yes	RE, WiC, ET
EaE [79]	entity	Yes		knowledge probing, open-domain QA, RE
JAKET [81]	entity, triplet	Yes		RC, QA over KG, entity classification
KEPLER [82]	entity, triplet	Yes		RC, ET, GLUE, link prediction
KP-PLM [83]	triplet	Yes		knowledge probing, RE, ET
K-BERT [84]	triplet		Yes	sentence classification, QA, NER
CoLAKE [85]	triplet	Yes		ET, RE, knowledge probing, GLUE, knowledge graph completion
FaE [86]	entity, triplet	Yes		open-domain QA
KLMO [89]	entity, triplet	Yes		ET, RC
K-ADAPTER [87]	general text, triplet		Yes	ET, QA, RC
KB-adapters [88]	entity, triplet		Yes	knowledge-probing using response selection, fact memorization, response generation

Method	Knowledge Source	Fusion in Pre-training	Fusion in Fine-tuning	NLU or NLG tasks
KERM [90]	triplet	Yes		passage re-ranking
JointLK [91]	triplet		Yes	commonsense reasoning
GreaseLM [92]	triplet		Yes	commonsense reasoning, medical QA
RuleBERT [94]	rule		Yes	rule reasoning
PTR [27]	rule		Yes	RC
MemNet [109]	retrieved text		Yes	open-domain dialogue generation
RAG [110]	retrieved text		Yes	open-domain/abstractive QA, Jeopardy question generation, fact verification
KFCNet [111]	retrieved text		Yes	commonsense/keyword generation
REINA [57]	general text, retrieval augmented generation		Yes	summarization, language modeling, machine translation, QA
KGR4 [112]	retrieved text		Yes	commonsense reasoning
KG-FiD [113]	retrieved text		Yes	open-domain QA
SeeKeR [114]	retrieved text		Yes	open-domain dialogue, prompt completion
RETRO [115]	retrieved text		Yes	language modelling, QA
UnifiedSKG [116]	structured knowledge	Yes		structured knowledge grounding
SKT [117]	retrieved text		Yes	knowledge-grounded dialogue
PLUG [118]	retrieved text, knowledge graph	Yes		knowledge-grounded dialogue, conversational recommendation
CE-PR [120]	knowledge graph		Yes	commonsense explanation generation
MRG [121]	knowledge graph		Yes	story/review/description generation
KEPM [122]	knowledge graph		Yes	story generation
MixGEN [123]	knowledge graph, expert knowledge		Yes	toxicity explanation
CNTF [124]	knowledge graph		Yes	dialogue generation
GRF [125]	knowledge graph		Yes	story ending generation, abductive NLG, explanation generation
KG-BART [11]	knowledge graph		Yes	commonsense reasoning/QA
JointGT [126]	knowledge graph	Yes		KG-to-text generation
MoKGE [127]	knowledge graph		Yes	commonsense explanation generation, abductive commonsense reasoning

KEPLMS: UNVEILING FUTURE TRAJECTORIES FOR ADVANCEMENTS

Revolutionizing ML and DL techniques in 2022-23, Knowledge-Enriched Pre-trained Language Models (KEPLMs) excel with advanced training strategies, leveraging diverse datasets and optimized model architectures.

6.1 Multimodal Learning Integration:

The recent strides in multimodal architectures, notably exemplified by innovations like DALL-E, showcase remarkable progress in image generation from textual prompts. Over the past two years, these advancements have stirred significant interest on platforms like Twitter, attracting attention from both researchers and the broader online and art communities.

The commercial applications in film production and gaming are evident, although copyright challenges persist. The precision achieved by these architectures is noteworthy, yet concerns about potential misuse for deep fakes linger. Multimodal architectures, like GPT-3 with 175 billion parameters, are not immune to biases, necessitating ongoing investigation, while their "black-box"

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904

nature poses transparency challenges. Integrating Multimodal Learning with Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) , exemplified by Fig.9 GANs, emerges as a futuristic research strength. However, as architectures grow in complexity, addressing environmental costs, computational demands, and promoting equitable access are critical for responsible AI development.

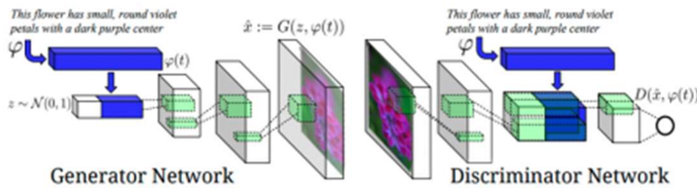


Figure 9: The text-conditioned convolutional GAN architecture that is being suggested. The Discriminator and the Generator are both fed text encoding υ(t). First, in fully-convolutional neural networks, it is projected to a lower dimensionality before

6.2 Continual Pre-Training:

[146] presented TCPL, a two-phase learning system for concept prerequisites. Pre-training that is ongoing improves language models through relationship discrimination. A resource–concept graph is used in joint learning to maximise R-GCN and the language model. Outperformed baselines but has limitations. Future work includes heterogeneous graph modeling and diverse relationship consideration.

By training on During the continuous pre-training phase, domain-related knowledge is integrated into the pre-trained language model using a masked language model and relationship discrimination tasks in the domain corpus.[146] The stage of collaborative learning makes use of the connections between structural and textual data. During the continuous pre-training phase, idea pairs improve the BERT model for textual representation, and a resource–concept heterogeneous graph is created. Concept-related knowledge is also into the BERT framework.

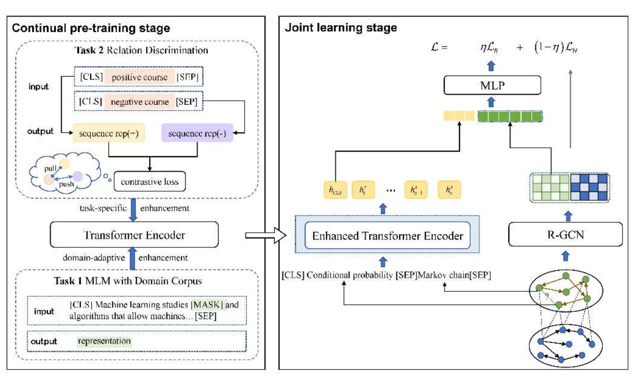


Figure 10: The design of the method we suggest. The pretrained language model is improved and adjusted with R-GCN at the joint learning stage of the continuous pre-training phase.

inputs for structural representations in the R-GCN model. [Fig.9] Concatenating these representations, the classification layer processes the input. BERT, Fig. 10 R-GCN, and the classification layer are all optimized at the same time in order to lower the training objective TM.

6.3 Dynamic Knowledge Graphs:

Innovating the conceptual design phase, current CAD tools lack ample support. Future research proposes a dynamic knowledge graph, a futuristic approach leveraging pre-trained language models. This envisions architects developing abstract conceptual sketches, divorcing details, supported by real-time domain ontology. The dynamic knowledge graph allows experts to formalize knowledge, with runtime adaptability for building class-specific design rules. Tool support showcases these concepts, exemplifying the cutting-edge consistency analysis between knowledge and conceptual design. This groundbreaking study [147] aims to produce pre-trained language models and dynamic knowledge representation, which will transform the field of conceptual design in building construction.

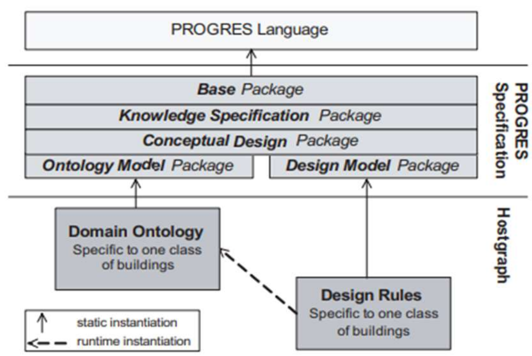


Figure 11: Multi-layered system architecture for dynamic knowledge formalization [147]

6.4 Transfer Learning Advancements:

In the realm of image forensics, a forward-looking research approach pioneers the fusion of Fig. 12 DeepLabv3+ [148] and Error Level Analysis (ELA)[148]. This innovative method employs ELA to accentuate error level variations induced during image compression, facilitating precise localization of tampered areas. Leveraging ResNet50 in the transfer learning framework minimizes training time. Despite a constrained dataset, the network is fine-tuned for accurate forged region localization. Comparative assessments highlight superior precision compared to existing techniques, regardless of forging methods. Future research endeavors aim to elevate performance by integrating other pretrained models in Pre-trained Language Models Augmented with Knowledge (PLMs-AwK). This strategic leap embraces the transformative potential of applications for deep learning (DL) and machine learning (ML)within the evolving landscape of digital forensics.

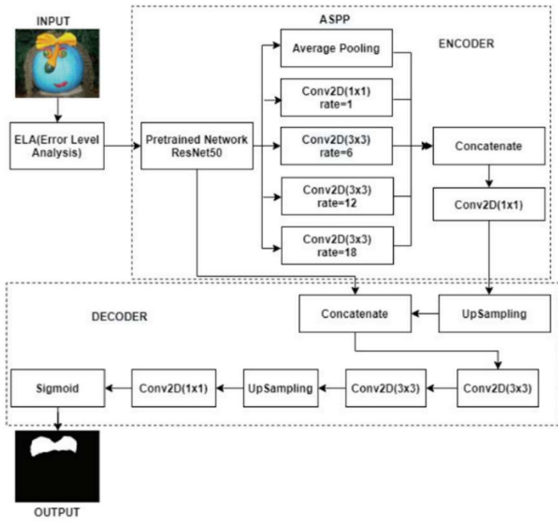


Figure 12: DeepLabV3+ [148] based method for localising image forgeries. Batch normalisation and ReLU are used in each convolution block in the diagram.

6.5 Ethical and Bias Mitigation:

In AI and ML applications, the synergy between Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) and bias mitigation is pivotal for future research trends. Mitigating biases is crucial for fair outcomes, and KE-PLMs, enriched by external knowledge, play a significant role. They actively contribute to bias mitigation by leveraging diverse knowledge and training on datasets emphasizing fairness, diversity, and representation. KE-PLMs integrate cultural nuances, historical contexts, and societal factors to navigate biases in language and data, fostering context-aware and unbiased predictions. Future research may explore advanced techniques within KE-PLMs for both bias identification and active mitigation during decision-making, ensuring adaptability to evolving societal norms. The proposed interactive approach [149], integrating human intuition, judgment, and objective bias measures, lays the foundation for ongoing investigations. Expanding user interactions and adapting sampling to address conflicting fairness criteria aligns with the evolving paradigm of multi-criteria decision-making, anticipating a future where KE-PLMs seamlessly integrate ethical considerations into AI and ML applications.

6.6 User-Friendly Interfaces:

[150] paper extensively reviews Fig.13 Real-Time Analytics (RTA) solutions: investigating tools, processing platforms, infrastructure, and analytics methodologies. It introduces a logical analytics stack, addressing key research questions (RQs) on recent concepts, architectures, and integrating machine learning/artificial intelligence (ML/AI) into RTA. Noteworthy novelties include practical insights from real-life finance and health case studies, emphasizing ML/AI integration. The paper anticipates future trends by discussing challenges, potential ML/AI incorporation, data quality, and user-friendly solutions. The connection with Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) lies in enhancing RTA fig.12 through advanced language understanding, potentially aiding in data quality and analytics insights. This comprehensive

Catalyst Research Volume 23, Issue 2, December 2023 Pp. 4871-4904
review, bridging AI/ML with RTA, is crucial for both academic and industrial collaboration, guiding future advancements in the field.

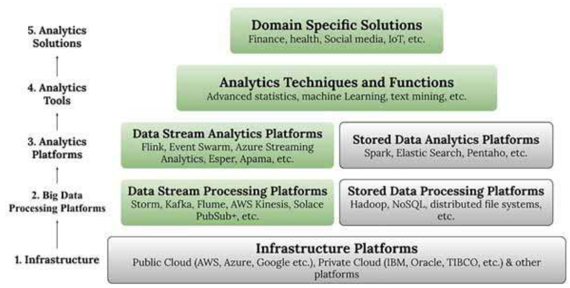


Figure 13: The real-time analytics stack [150]

6.7 Robustness and Security Measures:

Our exploration delves into large-scale pre-trained transformers' durability, highlighting their tenacious accuracy on difficult datasets intended to reveal false correlations hidden within the model. Notable results highlight how pre-training significantly increases robustness by skillfully extrapolating from a small number of samples, counteracting the predominately incorrect tendencies in the training set. To further improve resilience, the study recommends using bigger model sizes, augmenting pre-training data, and implementing Fig.14 Multi-Task Learning (MTL) with auxiliary data, especially in situations when minority cases are few. As we look ahead, we advocate a paradigm shift, focusing on fortifying Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) with enhanced security measures. Leveraging insights from data diversity expansion methods and principled out-of-distribution generalization strategies, this approach aims to anticipate challenges in future applications. Unraveling the subtleties of pre-trained models' resistance to overfitting and understanding the impact of diverse initializations are vital strides in designing robust fig. 13 and secure models, aligning with the trajectory of technology forecasting in Machine Learn

6.8 Collaborative Research Initiatives:

[151] article addresses five challenges in modeling Open Innovation (OI) settings, proposing requirements for an enhanced technique, FPM 2.0, derived from BPMN 2.0. FPM 2.0 caters specifically to Collaborative Exploration (CE) modeling, including elements crucial for OI, such as collaboration patterns and think Let utilization. While promising, the article has limitations, focusing on "outside-in" OI processes and requiring further exploration for broader applicability, especially in "inside-out" fig. 14 scenarios. The quality and practicality of FPM 2.0 need validation through assessment studies, testing ease of use and completeness. Connecting this with Fig.15 Pre-trained Language Models Augmented with Knowledge (PLMs-AwK) could enhance FPM 2.0 by providing advanced language understanding, potentially improving the accuracy and completeness of collaborative process representations in OI initiatives. Future research trends in AI and ML may involve integrating KE-PLMs to optimize modeling support for mass collaboration in OI.

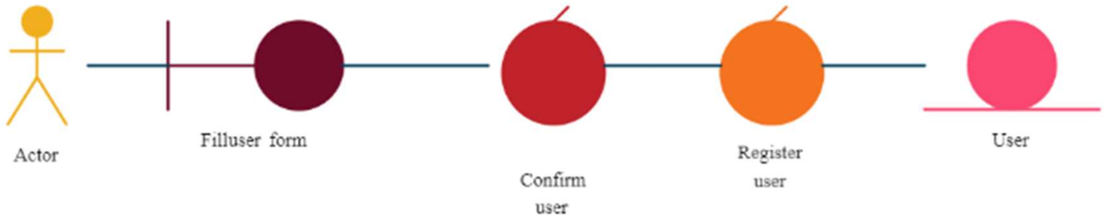


Figure 14: Use Creately's easy online diagram editor to edit this diagram, collaborate with others and export results to multiple image formats.

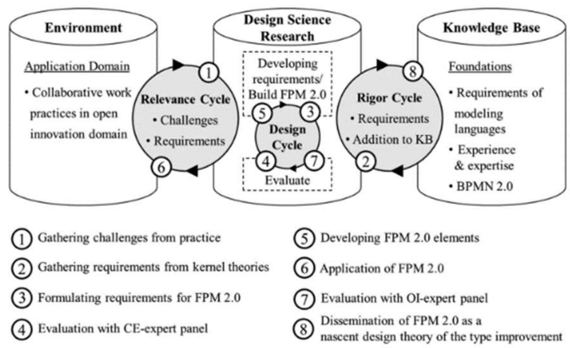


Figure 15:Research approach.[151]

CONCLUSION

Finally, our thorough analysis offers a thorough examination of Knowledge Enhanced Pre-Trained Language Models (KE-PLMs) in the fields of Natural Language Generation (NLG) and Natural Language Understanding (NLU) (NLG). The proposed meticulous taxonomies offer a nuanced understanding, discerning the unique focuses of KE-PLMs in each domain. By delving into representative works and methodologies, we illuminate the contributions of these models, accompanied by a critical examination of evaluation benchmarks.

Within NLU, our survey navigates through the nuanced integration of linguistic, textual, rule-based knowledge and knowledge graphs (KG) into KE-PLMs. Concurrently, Our division of models into retrieval-based and KG-based procedures in the field of NLG highlights the variety of methods used to incorporate outside information.

This survey transcends the current landscape to spotlight the prevailing challenges and problems encountered by KE-PLMs. Our insights into potential future research directions serve as a roadmap, guiding and stimulating further advancements in this promising field. As a valuable resource, this survey fosters a deeper appreciation of the intricacies of KE-PLMs, paving the way for future innovations and breakthroughs.

In shaping the future of research in KE-PLMs, this study contributes to the foundation of knowledge in the field. Researchers can draw upon our insights to inform their investigations, addressing existing challenges and forging new paths for exploration. The interdisciplinary nature of KE-PLMs, combining linguistic, textual, and knowledge-driven elements, underscores their potential for transformative applications across diverse sectors. This study lays the groundwork for a continued dialogue, encouraging scholars and practitioners to collaborate and push the boundaries of KE-PLMs, thereby propelling the field towards new frontiers of discovery and innovation.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language under-standing,” in NAACL-HLT, 2019, pp. 4171–4186.
[2] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever et al., “Improving language understanding by generative pre-training,”2018.
[3] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena,Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transferlearning with a unified text-to-text transformer,” Journal of Ma-chine Learning Research, vol. 21, pp. 140:1–140:67, 2020.
[4] N. F. Liu, M. Gardner, Y. Belinkov, M. E. Peters, and N. A. Smith, “Linguistic knowledge and transferability of contextual representations,” in NAACL-HLT, 2019, pp. 1073–1094.
[5] Y. Belinkov and J. R. Glass, “Analysis methods in neural language processing: A survey,” in NAACL-HLT, J. Burstein, C. Doran, and T. Solorio, Eds., 2019, pp. 3348–3354.
[6] I. Vulic, E. M. Ponti, R. Litschko, G. Glavas, and A. Korhonen, “Probing pretrained language models for lexical semantics,” in EMNLP, 2020, pp. 7222–7240.
[7] F. Petroni, T. Rockt¨aschel, S. Riedel, P. S. H. Lewis, A. Bakhtin, Y. Wu, and A. H. Miller, “Language models as knowledge bases?” in EMNLP-IJCNLP, 2019, pp. 2463–2473.
[8] S. Li, X. Li, L. Shang, Z. Dong, C. Sun, B. Liu, Z. Ji, X. Jiang, and Q. Liu, “How pre-trained language models capture factual knowledge? A causal-inspired analysis,” in ACL, 2022, pp. 1720–1732.
[9] A. Talmor, Y. Elazar, Y. Goldberg, and J. Berant, “olmpics – on what language model pre-training captures,” Transactions of the Association for Computational Linguistics, vol. 8, pp. 743–758, 2020.
[10] Y. Sun, S. Wang, S. Feng, S. Ding, C. Pang, J. Shang, J. Liu, X. Chen, Y. Zhao, Y. Lu et al., “ERNIE 3.0: Large-scale knowledge enhanced pre-training for language understanding and genera-tion,” arXiv preprint arXiv:2107.02137, 2021.
[11] Y. Liu, Y. Wan, L. He, H. Peng, and S. Y. Philip, “Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning,” in AAAI, vol. 35, no. 7, 2021, pp. 6418–6425.
[12] X. Wei, S. Wang, D. Zhang, P. Bhatia, and A. Arnold, “Knowledge enhanced pretrained language models: A comprehensive survey,” arXiv preprint arXiv:2110.08455, 2021.
[13] D. Yin, L. Dong, H. Cheng, X. Liu, K.-W. Chang, F. Wei, and J. Gao, “A survey of knowledge-intensive NLP with pre-trained language models,” arXiv preprint arXiv:2202.08772, 2022.
[14] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” in NAACL-HLT, 2018, pp. 2227–2237.

[15] J. Howard and S. Ruder, “Universal language model fine-tuning for text classification,” in ACL, 2018, pp. 328–339.
[16] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, “Self-supervised learning: Generative or contrastive,” IEEE Transactions on Knowledge and Data Engineering, 2021.
[17] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pre-train, prompt, and predict: A systematic survey of prompt-ing methods in natural language processing,” arXiv preprint arXiv:2107.13586, 2021.
[18] L. Dong, N. Yang, W. Wang, F. Wei, X. Liu, Y. Wang, J. Gao, M. Zhou, and H. Hon, “Unified language model pre-training for natural language understanding and generation,” in NeurIPS, 2019, pp. 13 042–13 054.
[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N.Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,”in NIPS, 2017, pp. 5998–6008.
[20] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, “BART: denoising sequence-to-sequence pre-training for natural language genera-tion, translation, and comprehension,” in ACL, 2020, pp. 7871–7880.
[21] T. Shin, Y. Razeghi, R. L. L. IV, E. Wallace, and S. Singh, “Autoprompt: Eliciting knowledge from language models with automatically generated prompts,” in EMNLP, 2020, pp. 4222–4235.
[22] T. Schick and H. Sch ̈utze, “Exploiting cloze-questions for few-shot text classification and natural language inference,” in EACL,2021, pp. 255–269.
[23] X. L. Li and P. Liang, “Prefix-tuning: Optimizing continuous prompts for generation,” in ACL-IJCNLP, 2021, pp. 4582–4597.
[24] Y. Gu, X. Han, Z. Liu, and M. Huang, “PPT: pre-trained prompt tuning for few-shot learning,” in ACL, 2022, pp. 8410–8423.
[25] N. Zhang, L. Li, X. Chen, S. Deng, Z. Bi, C. Tan, F. Huang, and H. Chen, “Differentiable prompt makes pre-trained language models better few-shot learners,” in ICLR, 2022.
[26] N. Ding, S. Hu, W. Zhao, Y. Chen, Z. Liu, H. Zheng, and M. Sun, “Openprompt: An open-source framework for prompt-learning,” in ACL, 2022, pp. 105–113.
[27] X. Han, W. Zhao, N. Ding, Z. Liu, and M. Sun, “PTR: prompt tuning with rules for text classification,” arXiv preprint arXiv:2105.11259, 2021.
[28] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge en-hanced prompt-tuning for aspect-based analysis,” arXiv preprint arXiv:2109.08306, 2021.
[29] S. Hu, N. Ding, H. Wang, Z. Liu, J. Wang, J. Li, W. Wu, and M. Sun, “Knowledgeable prompt-tuning: Incorporating knowl-edge into prompt verbalizer for text classification,” in ACL, 2022, pp. 2225–2240.
[30] X. Chen, N. Zhang, X. Xie, S. Deng, Y. Yao, C. Tan, F. Huang, L. Si, and H. Chen, “Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction,” in ACM, 2022, pp. 2778–2788.

[31] H. Ye, N. Zhang, S. Deng, X. Chen, H. Chen, F. Xiong, X. Chen, and H. Chen, “Ontology-enhanced prompt-tuning for few-shot learning,” in ACM, 2022, pp. 778–787.
[32] W. Yu, C. Zhu, Z. Li, Z. Hu, Q. Wang, H. Ji, and M. Jiang, “A survey of knowledge-enhanced text generation,” ACM Computing Surveys (CSUR), 2022.
[33] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra,A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann,P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes,Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson,R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin,T. Duke, A. Levskaya, S. Ghemawat, S. Dev, H. Michalewski,X. Garcia, V. Misra, K. Robinson, L. Fedus, D. Zhou, D. Ip-polito, D. Luan, H. Lim, B. Zoph, A. Spiridonov, R. Sepassi,D. Dohan, S. Agrawal, M. Omernick, A. M. Dai, T. S. Pillai,M. Pellat, A. Lewkowycz, E. Moreira, R. Child, O. Polozov,K. Lee, Z. Zhou, X. Wang, B. Saeta, M. Diaz, O. Firat, M. Catasta,J. Wei, K. Meier-Hellstern, D. Eck, J. Dean, S. Petrov, and N. Fiedel, “Palm: Scaling language modeling with pathways,”arXiv preprint arXiv:2204.02311, 2022.
[34] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min,B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen,J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J. Nie, and J. Wen, “A survey of large language models,” arXiv preprint arXiv:2303.18223, 2023.
[35] OpenAI, “GPT-4 technical report,” OpenAI, 2023.
[36] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud,D.Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi,T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus,“Emergent abilities of large language models,” arXiv preprintarXiv:2206.07682, 2022.
[37] R. Kemker, M. McClure, A. Abitino, T. L. Hayes, and C. Kanan“Measuring catastrophic forgetting in neural networks,” in AAAI, 2018, pp. 3390–3398.
[38] G. Izacard, P. S. H. Lewis, M. Lomeli, L. Hosseini, F. Petroni,T. Schick, J. Dwivedi-Yu, A. Joulin, S. Riedel, and E. Grave, “Few-shot learning with retrieval augmented language models,” arXiv preprint arXiv:2208.03299, 2022.
[39] B. Peng, M. Galley, P. He, H. Cheng, Y. Xie, Y. Hu, Q. Huang,L. Liden, Z. Yu, W. Chen, and J. Gao, “Check your facts and try again: Improving large language models with external knowl-edge and automated feedback,” arXiv preprint arXiv:2302.12813, 2023.
[40] W. Yu, C. Zhu, Z. Li, Z. Hu, Q. Wang, H. Ji, and M. Jiang, “A survey of knowledge-enhanced text generation,” ACM Computing Surveys (CSUR), 2022.
[41] A. Lauscher, I. Vulic, E. M. Ponti, A. Korhonen, and G. Glavas, “Specializing unsupervised pretraining models for word-level semantic similarity,” in COLING, 2020, pp. 1371–1383.
[42] Y. Levine, B. Lenz, O. Dagan, O. Ram, D. Padnos, O. Sharir, S. Shalev-Shwartz, A. Shashua, and Y. Shoham, “Sensebert: Driving some sense into BERT,” in ACL, 2020, pp. 4656–4667.
[43] H. Tian, C. Gao, X. Xiao, H. Liu, B. He, H. Wu, H. Wang, and F. Wu, “SKEP: sentiment knowledge enhanced pre-training forsentiment analysis,” in ACL, 2020, pp. 4067–4076.
[44] B. Lyu, L. Chen, S. Zhu, and K. Yu, “Let: Linguistic knowledge enhanced graph transformer for chinese short text matching,” in AAAI, vol. 35, no. 15, 2021, pp. 13 498–13 506.

[45] Y. Xu, C. Zhu, S. Wang, S. Sun, H. Cheng, X. Liu, J. Gao, P. He,M. Zeng, and X. Huang, “Human parity on commonsenseqa: Augmenting self-attention with external attention,” in IJCAI,2022, pp. 2762–2768.

[46] Q. Chen, F. Li, G. Xu, M. Yan, J. Zhang, and Y. Zhang, “Dictbert: Dictionary description knowledge enhanced language model pre-training via contrastive learning,” in IJCAI, 2022, pp. 4086–4092.

[47] J. Zhou, Z. Zhang, H. Zhao, and S. Zhang, “LIMIT-BERT :Linguistics informed multi-task BERT,” in EMNLP, 2020, pp.4450–4461.

[48] J. Bai, Y. Wang, Y. Chen, Y. Yang, J. Bai, J. Yu, and Y. Tong, “Syntax-bert: Improving pre-trained transformers with syntax trees,” in EACL, 2021, pp. 3011–3020.

[49] Z. Li, Q. Zhou, C. Li, K. Xu, and Y. Cao, “Improving BERT with syntax-aware local attention,” in ACL-IJCNLP, 2021, pp. 645–653.

[50] D. S. Sachan, Y. Zhang, P. Qi, and W. L. Hamilton, “Do syn-tax trees help pre-trained transformers extract information?” in EACL, 2021, pp. 2647–2661.

[51] U. Khandelwal, O. Levy, D. Jurafsky, L. Zettlemoyer, andM. Lewis, “Generalization through memorization: Nearest neighbor language models,” in ICLR, 2020.

[52] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, “Retrieval augmented language model pre-training,” in ICML, 2020, pp.3929–3938.

[53] E. Grave, A. Joulin, and N. Usunier, “Improving neural language models with a continuous cache,” in ICLR, 2017.

[54] S. Murty, P. W. Koh, and P. Liang, “Expbert: Representation engineering with natural language explanations,” in ACL, 2020, pp. 2106–2113.

[55] W. Cui and X. Chen, “Enhancing natural language representation with large-scale out-of-domain commonsense,” in ACL, 2022, pp.1746–1756.

[56] Y. Yao, S. Huang, L. Dong, F. Wei, H. Chen, and N. Zhang, “Kformer: Knowledge injection in transformer feed-forward layers,” in NLPCC, vol. 13551, 2022, pp. 131–143.

[57] S. Wang, Y. Xu, Y. Fang, Y. Liu, S. Sun, R. Xu, C. Zhu, and M. Zeng, “Training data is more valuable than you think: A simple and effective method by retrieving from training data,” in ACL,2022, pp. 3170–3179.

[58] B. Oguz, X. Chen, V. Karpukhin, S. Peshterliev, D. Okhonko, M. S. Schlichtkrull, S. Gupta, Y. Mehdad, and S. Yih, “Unik-qa: Unified representations of structured and unstructured knowledge for open-domain question answering,” in NAACL, 2022, pp. 1535–1546.

[59] K. Ma, H. Cheng, X. Liu, E. Nyberg, and J. Gao, “Open domain question answering with a unified knowledge interface,” in ACL, 2022, pp. 1605–1620.

[60] X. Pan, W. Yao, H. Zhang, D. Yu, D. Yu, and J. Chen, “Knowledge-in-context: Towards knowledgeable semi-parametric language models,” ICLR, 2023.

[61] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, andJ. Kang, “Biobert: a pre-trained biomedical language represen-tation model for biomedical text mining,” Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.

[62] I. Beltagy, K. Lo, and A. Cohan, "Scibert: A pretrained language model for scientific text," in EMNLP-IJCNLP, 2019, pp. 3613–3618.
[63] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. S. Weld,"S2ORC: the semantic scholar open research corpus," in ACL,2020, pp. 4969–4983.
[64] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier et al., "Knowledge graphs," ACM Computing Surveys (CSUR),vol. 54, no. 4, pp. 1–37, 2021.
[65] B. Xue and L. Zou, "Knowledge graph quality management: a comprehensive survey," IEEE Transactions on Knowledge and Data Engineering, 2022.
[66] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 12, pp. 2724–2743, 2017.
[67] X. Wang, F. Luo, Q. Wu, and Z. Bao, "How context or knowledge can benefit healthcare question answering," IEEE Transactions on Knowledge and Data Engineering, 2021.
[68] L. Fang, Y. Luo, K. Feng, K. Zhao, and A. Hu, "A knowledge-enriched ensemble method for word embedding and multi-sense embedding," IEEE Transactions on Knowledge and Data Engineering, 2022.
[69] Q. Li, D. Wang, S. F. K. Song, Y. Zhang, and G. Yu, "Oerl:Enhanced representation learning via open knowledge graphs,"IEEE Transactions on Knowledge and Data Engineering, 2022.
[70] Y. Zhao, H. Zhou, A. Zhang, R. Xie, Q. Li, and F. Zhuang, "Connecting embeddings based on multiplex relational graph attention networks for knowledge graph entity typing," IEEE Transactions on Knowledge and Data Engineering, 2022.
[71] Y. Sun, S. Wang, Y. Li, S. Feng, X. Chen, H. Zhang, X. Tian, D. Zhu, H. Tian, and H. Wu, "ERNIE: enhanced representation through knowledge integration," arXiv preprint arXiv:1904.09223, 2019.
[72] W. Xiong, J. Du, W. Y. Wang, and V. Stoyanov, "Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model," in ICLR, 2020.
[73] J. Wang, C. Wang, M. Qiu, Q. Shi, H. Wang, J. Huang, and M. Gao, "Kecp: Knowledge enhanced contrastive prompting for few-shot extractive question answering," arXiv preprint arXiv:2205.03071, 2022.
[74] Y. Qin, Y. Lin, R. Takanobu, Z. Liu, P. Li, H. Ji, M. Huang, M. Sun, and J. Zhou, "ERICA: improving entity and relation understanding for pre-trained language models via contrastive learning," in ACL-IJCNL, 2021, pp. 3350–3363.
[75] T. Zhang, C. Wang, N. Hu, M. Qiu, C. Tang, X. He, and J. Huang, "Dkplm: Decomposable knowledge-enhanced pre-trained language model for natural language understanding," in AAAI, vol. 36, no. 10, 2022, pp. 11 703–11 711.
[76] I. Yamada, A. Asai, H. Shindo, H. Takeda, and Y. Matsumoto, "LUKE: deep contextualized entity representations with entity-aware self-attention," in EMNLP, 2020, pp. 6442–6454.
[77] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, "ERNIE: enhanced language representation with informative entities," in ACL, 2019, pp. 1441–1451.

78] M. E. Peters, M. Neumann, R. L. L. IV, R. Schwartz, V. Joshi, S. Singh, and N. A. Smith, "Knowledge enhanced contextual word representations," in EMNLP-IJCNLP, 2019, pp. 43–54.

[79] T. F´evry, L. B. Soares, N. FitzGerald, E. Choi, and T. Kwiatkowski, "Entities as experts: Sparse memory access with entity supervision," in EMNLP, 2020, pp. 4937–4951.

[80] Y. Su, X. Han, Z. Zhang, Y. Lin, P. Li, Z. Liu, J. Zhou, and M. Sun, "Cokebert: Contextual knowledge selection and embedding towards enhanced pre-trained language models," AI Open, vol. 2, pp. 127–134, 2021.

[81] D. Yu, C. Zhu, Y. Yang, and M. Zeng, "JAKET: joint pre-training of knowledge graph and language understanding," in AAAI, 2022, pp. 11 630–11 638.

[82] X. Wang, T. Gao, Z. Zhu, Z. Zhang, Z. Liu, J. Li, and J. Tang, "KEPLER: A unified model for knowledge embedding and pre-trained language representation," Transactions of the Association for Computational Linguistics, vol. 9, pp. 176–194, 2021.

[83] J. Wang, W. Huang, Q. Shi, H. Wang, M. Qiu, X. Li, and M. Gao, "Knowledge prompting in pre-trained language model for nat-ural language understanding," arXiv preprint arXiv:2210.08536,2022.

[84] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, andP. Wang, "K-bert: Enabling language representation with knowl-edge graph," in AAAI, vol. 34, no. 03, 2020, pp. 2901–2908.

[85] T. Sun, Y. Shao, X. Qiu, Q. Guo, Y. Hu, X. Huang, and Z. Zhang, "Colake: Contextualized language and knowledge embedding, "in COLING, 2020, pp. 3660–3670.

[86] P. Verga, H. Sun, L. B. Soares, and W. W. Cohen, "Facts as experts: Adaptable and interpretable neural memory over symbolic knowledge," arXiv preprint arXiv:2007.00849, 2020.

[87] R. Wang, D. Tang, N. Duan, Z. Wei, X. Huang, J. Ji, G. Cao, D. Jiang, and M. Zhou, "K-adapter: Infusing knowledge into pre-trained models with adapters," in ACL-IJCNLP, 2021, pp. 1405–1418.

[88] D. Emelin, D. Bonadiman, S. Alqahtani, Y. Zhang, and S. Mansour, "Injecting domain knowledge in language models for task oriented dialogue systems," in EMNLP, 2022.

[89] L. He, S. Zheng, T. Yang, and F. Zhang, "Klmo: Knowledge graph enhanced pretrained language model with fine-grained relationships," in EMNLP, 2021, pp. 4536–4542.

[90] Q. Dong, Y. Liu, S. Cheng, S. Wang, Z. Cheng, S. Niu, and D. Yin, "Incorporating explicit knowledge in pre-trained language models for passage re-ranking," in SIGIR, 2022, pp. 1490–1501.

[91] Y. Sun, Q. Shi, L. Qi, and Y. Zhang, "Jointlk: Joint reasoning with language models and knowledge graphs for commonsense question answering," in NAACL, 2022, pp. 5049–5060.

[92] X. Zhang, A. Bosselut, M. Yasunaga, H. Ren, P. Liang, C. D. Manning, and J. Leskovec, "Greaselm: Graph reasoning enhanced language models," in ICLR, 2022.

[93] L. Von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giessel-bach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy et al., "Informed machine learning–a taxonomy and

survey of integrating knowledge into learning systems," arXiv preprint arXiv:1903.12394, 2019.
[94] M. Saeed, N. Ahmadi, P. Nakov, and P. Papotti, "Rulebert: Teaching soft rules to pre-trained language models," in EMNLP, 2021, pp. 1460–1476.
[95] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bow-man, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," in EMNLP, 2018, pp. 353–355.
[96] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Superglue: A stickier bench-mark for general-purpose language understanding systems," in NeurIPS, 2019, pp. 3261–3275.
[97] N. P ¨orner, U. Waltinger, and H. Sch ¨utze, "BERT is not a knowledge base (yet): Factual knowledge vs. name-based reasoning in unsupervised QA," arXiv preprint arXiv:1911.03681, 2019.
[98] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning, "Position-aware attention and supervised data improve slot fill-ing," in EMNLP, 2017, pp. 35–45.
[99] X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun, "Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation," in EMNLP, 2018, pp. 4803–4809.
[100] E. Choi, O. Levy, Y. Choi, and L. Zettlemoyer, "Ultra-fine entity typing," in ACL, 2018, pp. 87–96.
[101] X. Ling, S. Singh, and D. S. Weld, "Design challenges for entity linking," Trans. Assoc. Comput. Linguistics, vol. 3, pp. 315–328, 2015.
[102] E. F. T. K. Sang and F. D. Meulder, "Introduction to the conll-2003 shared task: Language-independent named entity recognition," in HLT-NAACL, 2003, pp. 142–147.
[103] A. Talmor, J. Herzig, N. Lourie, and J. Berant, "Commonsenseqa: A question answering challenge targeting commonsense knowledge," in NAACL-HLT, 2019, pp. 4149–4158.
[104] T. Mihaylov, P. Clark, T. Khot, and A. Sabharwal, "Can a suit of armor conduct electricity? A new dataset for open book question answering," in EMNLP, 2018, pp. 2381–2391.
[105] T. Kwiatkowski, J. Palomaki, O. Redfield, M. Collins, A. P. Parikh, C. Alberti, D. Epstein, I. Polosukhin, J. Devlin, K. Lee, K. Toutanova, L. Jones, M. Kelcey, M. Chang, A. M. Dai, J. Uszko-reit, Q. Le, and S. Petrov, "Natural questions: a benchmark for question answering research," Trans. Assoc. Comput. Linguistics, vol. 7, pp. 452–466, 2019.
[106] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," in EMNLP, 2013, pp. 1533–1544.
[107] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer, "Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension," in ACL, 2017, pp. 1601–1611.
[108] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in EMNLP, 2013, pp. 1631–1642.

[109] E. Dinan, S. Roller, K. Shuster, A. Fan, M. Auli, and J. We-ston, “Wizard of wikipedia: Knowledge-powered conversational agents,” in ICLR, 2019.

[110] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. K ¨uttler, M. Lewis, W.-t. Yih, T. Rockt¨aschel et al., “Retrieval-augmented generation for knowledge-intensive nlp tasks,” Advances in Neural Information Processing Systems, vol. 33, pp. 9459– 9474, 2020.

[111] H. Li, Y. Gong, J. Jiao, R. Zhang, T. Baldwin, and N. Duan, “Kfcnet: Knowledge filtering and contrastive learning net-work for generative commonsense reasoning,” arXiv preprint arXiv:2109.06704, 2021.

[112] X. Liu, D. Liu, B. Yang, H. Zhang, J. Ding, W. Yao, W. Luo, H. Zhang, and J. Su, “Kgr4: Retrieval, retrospect, refine and rethink for commonsense generation,” in AAAI, vol. 36, no. 10,2022, pp. 11 029–11 037.

[113] D. Yu, C. Zhu, Y. Fang, W. Yu, S. Wang, Y. Xu, X. Ren, Y. Yang, and M. Zeng, “Kg-fid: Infusing knowledge graph in fusion-in-decoder for open-domain question answering,” in ACL, 2022, pp. 4961–4974.

[114] K. Shuster, M. Komeili, L. Adolphs, S. Roller, A. Szlam, and J. Weston, “Language models that seek for knowledge: Modular search & generation for dialogue and prompt completion,” arXiv preprint arXiv:2203.13224, 2022.

[115] S. Borgeaud, A. Mensch, J. Hoffmann, T. Cai, E. Rutherford, K. Millican, G. B. Van Den Driessche, J.-B. Lespiau, B. Damoc, A. Clark et al., “Improving language models by retrieving from trillions of tokens,” in ICML, 2022, pp. 2206–2240.

[116] T. Xie, C. H. Wu, P. Shi, R. Zhong, T. Scholak, M. Yasunaga,C. Wu, M. Zhong, P. Yin, S. I. Wang, V. Zhong, B. Wang, C. Li,C. Boyle, A. Ni, Z. Yao, D. R. Radev, C. Xiong, L. Kong, R. Zhang,N. A. Smith, L. Zettlemoyer, and T. Yu, “Unifiedskg: Unifying and multi-tasking structured knowledge grounding with text-to-text language models,” arXiv preprint arXiv:2201.05966, 2022.

[117] B. Kim, J. Ahn, and G. Kim, “Sequential latent knowledge selection for knowledge-grounded dialogue,” in ICLR, 2020.

[118] Y. Li, B. Peng, Y. Shen, Y. Mao, L. Liden, Z. Yu, and J. Gao, “Knowledge-grounded dialogue generation with a unified knowledge representation,” in NAACL, 2022, pp. 206–218.

[119] G. Izacard and E. Grave, “Leveraging passage retrieval with generative models for open domain question answering,” in EACL, 2021, pp. 874–880.

[120] H. Ji, P. Ke, S. Huang, F. Wei, and M. Huang, “Generating commonsense explanation by extracting bridge concepts from reasoning paths,” in AACL-IJCNLP, 2020, pp. 248–257.

[121] L. Zhao, J. Xu, J. Lin, Y. Zhang, H. Yang, and X. Sun, “Graph-based multi-hop reasoning for long text generation,” arXiv preprint arXiv:2009.13282, 2020.

[122] J. Guan, F. Huang, M. Huang, Z. Zhao, and X. Zhu, “A knowledge-enhanced pretraining model for commonsense story generation,” Transactions of the Association for Computational Linguistics, vol. 8, pp. 93–108, 2020.

[123] R. Sridhar and D. Yang, "Explaining toxic text via knowledge enhanced text generation," in NAACL, 2022, pp. 811–826.
[124] D. Varshney, A. Prabhakar, and A. Ekbal, "Commonsense and named entity aware knowledge grounded dialogue generation," in NAACL, 2022, pp. 1322–1335.
[125] H. Ji, P. Ke, S. Huang, F. Wei, X. Zhu, and M. Huang, "Language generation with multi-hop reasoning on commonsense knowledge graph," in EMNLP, 2020, pp. 725–736.
[126] P. Ke, H. Ji, Y. Ran, X. Cui, L. Wang, L. Song, X. Zhu, andM. Huang, "Jointgt: Graph-text joint representation learning fortext generation from knowledge graphs," in ACL-IJCNLP, 2021,pp . 2526–2538.
[127] W. Yu, C. Zhu, L. Qin, Z. Zhang, T. Zhao, and M. Jiang, "Di-versifying content generation for commonsense reasoning withmixture of knowledge graph experts," in ACL, 2022, pp. 1896–1906.
[128] Y. Lu, W. Feng, W. Zhu, W. Xu, X. E. Wang, M. Eckstein, and W. Y. Wang, "Neuro-symbolic procedural planning with commonsense prompting," ICLR, 2023.
[129] K. Zhou, S. Prabhumoye, and A. W. Black, "A dataset for document grounded conversations," in EMNLP, 2018, pp. 708–713.
[130] B. Y. Lin, W. Zhou, M. Shen, P. Zhou, C. Bhagavatula, Y. Choi, and X. Ren, "Commongen: A constrained text generation challenge for generative commonsense reasoning," in EMNLP-Findings, 2020, pp. 1823–1840.
[131] C. Bhagavatula, R. L. Bras, C. Malaviya, K. Sakaguchi, A. Holtzman, H. Rashkin, D. Downey, W. Yih, and Y. Choi, "Abductive commonsense reasoning," in ICLR, 2020.
[132] C. Wang, S. Liang, Y. Zhang, X. Li, and T. Gao, "Does it make sense? and why? A pilot study for sense making and explana-tion," in ACL, 2019, pp. 4020–4026.
[133] N. Mostafazadeh, N. Chambers, X. He, D. Parikh, D. Batra,L. Vanderwende, P. Kohli, and J. F. Allen, "A corpus and cloze evaluation for deeper understanding of commonsense stories ,"in NAACL HLT, 2016, pp. 839–849.
[134] R. Nallapati, B. Zhou, C. N. dos Santos, C¸. G ¨ulc¸ehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," in CoNLL, 2016, pp. 280–290.
[135] S. Narayan, S. B. Cohen, and M. Lapata, "Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization," in EMNLP, 2018, pp. 1797–1807.
[136] D. Liu, Y. Yan, Y. Gong, W. Qi, H. Zhang, J. Jiao, W. Chen, J. Fu, L. Shou, M. Gong, P. Wang, J. Chen, D. Jiang, J. Lv, R. Zhang,W. Wu, M. Zhou, and N. Duan, "GLGE: A new general language generation evaluation benchmark," in ACL/IJCNLP, ser. Findings of ACL, vol. ACL/IJCNLP 2021, 2021, pp. 408–420.
[137] F. Petroni, A. Piktus, A. Fan, P. S. H. Lewis, M. Yazdani, N. D.Cao, J. Thorne, Y. Jernite, V. Karpukhin, J. Maillard, V. Plachouras, T. Rockt¨aschel, and S. Riedel, "KILT: a benchmark for knowledge intensive language tasks," in NAACL-HLT, 2021, pp. 2523–2544.

[138] C. Wang, S. Liang, Y. Jin, Y. Wang, X. Zhu, and Y. Zhang,"Semeval-2020 task 4: Commonsense validation and explanation," in SemEval@COLING, 2020, pp. 307–321.

[139] K. Chen, Q. Huang, Y. Bisk, D. McDuff, and J. Gao, "Kb-vlp: Knowledge based vision and language pretraining," in ICML workshop, July 2021.

[140] F. Yu, J. Tang, W. Yin, Y. Sun, H. Tian, H. Wu, and H. Wang,"Ernie-vil: Knowledge enhanced vision-language representations through scene graphs," in AAAI, vol. 35, no. 4, 2021, pp. 3208–3216.

[141] G. M. Van de Ven and A. S. Tolias, "Three scenarios for continual learning," arXiv preprint arXiv:1904.07734, 2019.

[142] Y. Qin, J. Zhang, Y. Lin, Z. Liu, P. Li, M. Sun, and J. Zhou, "ELLE: efficient lifelong pre-training for emerging data," in ACL, 2022, pp. 2789–2810.

[143] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, L. Zhang, W. Han, M. Huang, Q. Jin, Y. Lan, Y. Liu,Z. Liu, Z. Lu, X. Qiu, R. Song, J. Tang, J. Wen, J. Yuan, W. X. Zhao,and J. Zhu, "Pre-trained models: Past, present and future," AI Open, vol. 2, pp. 225–250, 2021.

[144] S. Rajbhandari, J. Rasley, O. Ruwase, and Y. He, "Zero: memory optimizations toward training trillion parameter models," in SC20: International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, 2020, pp. 1–16.

[145] Cem Akkus, undefined., et al, "Multimodal Deep Learning,"2023.

[146] X. Tang, K. Liu, H. Xu, W. Xiao, and Z. Tan, 'Continual Pre-Training of Language Models for Concept Prerequisite Learning with Graph Neural Networks', Mathematics, vol. 11, no. 12, 2023.

[147] "B. Kraft and D. Retkowitz, "Graph Transformations for Dynamic Knowledge Processing," Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), Kauia, HI, USA, 2006, pp. 154b-154b, doi: 10.1109/HICSS.2006.200.,"

[148] N. Hebbar and A. Kunte, "A Deep Learning Framework with Transfer Learning Approach for Image Forgery Localization," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022, pp. 1-6,doi:10.1109/GCAT55367.2022. 9971986.

[149] H. Wang, S. Mukhopadhyay, Y. Xiao and S. Fang, "An Interactive Approach to Bias Mitigation in Machine Learning," 2021 IEEE 20th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Banff, AB, Canada,2021, pp199205,doi:10.1109/ICCICC53683.2021.9811333.

[150] W. Chen, Z. Milosevic, F. A. Rabhi and A. Berry, "Real-Time Analytics: Concepts, Architectures, and ML/AI Considerations," in IEEE Access, vol. 11, pp. 71634-71657, 2023, doi: 10.1109/ACCESS.2023.3295694.

[151] R. Winkler, R. O. Briggs, G. -J. de Vreede, J. M. Leimeister, S. Oeste-Reiss and M. Söllner, "Modeling Support for Mass Collaboration in Open Innovation Initiatives—The Facilitation Process Model 2.0," in IEEE Transactions on Engineering Management, vol. 69, no. 2, pp. 423-437, April2022,doi:10.1109/TEM.2020.2975938.

[152] undefined. Tobias Müller, undefined. Milena Zahn, undefined. Florian Matthes, "Unlocking the Potential of Collaborative AI – On the Socio-technical Challenges of Federated Machine Learning," 2023.

[153] Zou, Anni & Zhang, Zhuosheng & Zhao, Hai. (2023). Decker: Double Check with Heterogeneous Knowledge for Commonsense Fact Verification.