LLM-Empowered Literature Mining for Material Substitution Studies in Sustainable Concrete

Yifei Duan [a,b], Yixi Tian [a], Soumya Ghosh [c,d], Vineeth Venugopal [a], Jie Chen [c,d],

Elsa A. Olivetti [a*]

[a] Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[b] Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[c] IBM Research, Cambridge, MA 02142, USA

[d] MIT-IBM Watson AI Lab, Cambridge, MA 02142, USA

* Corresponding author.

Email address: elsao@mit.edu (Elsa A. Olivetti)

Full postal address: 77 Massachusetts Avenue Cambridge, MA 02139, USA

Abstract

Substituting constituents within concrete with lower impact materials is of utmost importance for the sustainable transition of the concrete industry. Systematic analyses of knowledge within the published literature can facilitate industrial practice and focus research inquiry. To address the prohibitive workload of manual review and the multifaceted linguistic complexity of communication within the domain, this study develops an automatic literature mining framework combining lightweight large language models (LLMs) (pythia-2.8B) with multiple-choice instructions. The current landscape, temporal trends, and future directions of concrete material substitution studies were analyzed using the extracted information. Although supplementary cementitious materials (SCMs) have remained a research hotspot, results revealed a systematic shift in recent studies from commercial SCMs to other materials. Geopolymer and fine aggregate studies have surged in the recent period, while clinker feedstock and filler studies have declined. Lime-pozzolan cement has been an underexplored application but emerges as a potentially promising future research direction.

Keywords: concrete, beneficial uses, alternative materials, literature mining, large language models, knowledge graph

1. Introduction

To reduce greenhouse emissions and limit raw material extraction, the concrete industry has committed to increasing use of natural or byproduct materials to substitute for concrete constituents. This is because of the industry's substantial carbon emissions from production (constituting 8 ~ 9% of global annual emissions) (Ellis et al., 2020) and the need for large-scale raw material mining that can lead to top-soil loss, deforestation, and resource depletion (Habert et al., 2020; Mehta, 2001). Both within industry and across the research community, work has involved substituting the constituents of concrete such as Portland cement, fine aggregate, and coarse aggregate, with processed natural mineral materials (e.g., metakaolin), recycled demolition and construction waste, industrial residues (e.g., silica fume, coal ashes, metallurgical slags), and agricultural and municipal solid waste incineration (MSWI) residues (Juenger et al., 2019; Kurniati et al., 2023; Snellings et al., 2023). To effectively navigate the diverse realm of raw materials and potential beneficial uses and identify research gaps, the community analyzes published literature to identify valuable research directions. Doing so in an automated fashion has become an attractive approach in recent years.

Natural language processing (NLP) methods have been used in a few previous works to automate literature mining on the topics related to resource use, recycling, and waste management, adopting deep neural networks for named entity recognition (NER) (Zhu et al., 2021; Zhu and Ren, 2023) that relies on the specification of noun phrases (NP). The previous efforts performed statistical summaries and trend analyses of paper topics related to waste-based materials, industrial management, human behaviors, climate and energy, etc. These efforts did not categorize relationships across entities so the interrelations in the extracted topics terms are lost (e.g. “coal”, “energy”, “water”, “land use” are all extracted from a same paper, but it’s difficult to automatically determine how they relate to each other or form a specific research direction). Kumar and coauthors (Kumar et al., 2023) focused on waste plastics recycling and adopted BERT architectures for question answering to extract knowledge such as reactants, products, and catalysts, achieving F1 metrics around 80%. A shared limitation of the methods adopted in these works is that the extracted information is limited to text strings that can be directly located within a span in the original literature text, as NP spanning text strings for NER are recognized as a foundation for literature mining tasks in most scientific topics (Li et al., 2023; Nasar et al., 2022).

Due to the complex linguistic structure of papers describing waste valorization, current approaches provide limited ability to perform logical inference from such complex source strings or formulate structured relational databases. Papers in these fields use non-standardized terminology for both feedstocks and their applications, and possess indirect non-NP source strings, non-local and non-sequential syntactic dependency, and the non-injective mapping relations (e.g. one paper mapped to multiple materials, one material to multiple applications). NLP researchers have made efforts to address these issues in other domains through methods to represent language or reframe tasks. Named entity normalization (NEN) was proposed for NER with non-standard terminology and has been adopted in scientific literature mining, using either a rule-based system or a trained synonym classifier for post-NER normalization (Leaman et al., 2015; Weston et al., 2019) or a semi-Markov model for joint NER-NEN (Leaman and Lu, 2016). Meanwhile, graph and hierarchical language representations were developed to model syntactic structures that can address certain non-local or non-sequential dependencies in information extraction (IE) like coreference or identical-mentions (Qian et al., 2019; Wilcox et al., 2019). However, all of these NEN and non-local IE methods still rely on the basic assumption of direct text sources for word- or phrase-level tagging (e.g. NP as named entity), and thus face obstacles for application in waste valorization. In certain scenarios, the tasks requiring indirect answers through contextualized inference were completed with text classification techniques using vectorized text embeddings for classifier inputs (Jindal et al., 2015; Krallinger et al., 2017). Despite its success, such methods lack the flexibility of producing an arbitrarily sized set of multiple classes (e.g. multiple materials) and struggle to handle the conditional problems (e.g. classify application conditioned on the material).

Despite these methodological innovations that address certain aspects of linguistic complexity individually, these approaches have not been integrated to tackle literature mining tasks that include several of these issues. Previous studies by the cement and concrete community have also made attempts to automate literature-mining, but in light of the aforementioned challenges these were confined to a limited scope with a suitable specialized solution. One notable work focused on extracting reactivity data from tables in literature (Uvegi et al., 2021), but the structured tabular sources are not applicable in more general research questions related to concrete constituents. Another effort classified cement manufacturing operation failure types based on text reports (Wang et al., 2023), but it features a single-task with injective mapping like other text classification problems. To infer indirect information from source text containing non-standard terminology and non-local dependencies, and simultaneously account for the non-injective mapping relations, we propose an integrated methodology leveraging large language models (LLMs). The advancement of generative LLMs introduces greater flexibility in both inputs and outputs to accommodate such complexity in constituent substitution literature mining. LLMs can ingest detailed instructions provided alongside the paper context as inputs, and they can output classes of materials and applications with the flexibility to determine the number of classes autonomously. The chemistry and materials science communities have explored LLMs for literature mining (Dagdelen et al., 2024; Walker et al., 2023; Xie et al., 2023), but the methods still centered on NER with NP text sources as a first step and used large, computationally expensive models like GPT-3.5 (175B).

This study developed multiple-choice instructions to leverage computationally cheaper LLMs (2.8B parameters) for contextualized answer extraction. Multiple-choice problem solving is a commonly studied task for LLMs (Hendrycks et al., 2021; Pezeshkpour and Hruschka, 2023; Savelka et al., 2023; Talmor et al., 2019; Zheng et al., 2024). This approach was used to formulate question answering in a multitask LLM benchmarking work (Rae et al., 2022) on a variety of commonsense and test-style general math and science knowledge questions, but the potential of transferring such problem formulation to literature understanding was underexplored. To the best of our knowledge, this study is the first to explore the use of multiple-choice problems as instructions in literature mining tasks. We applied this literature mining method on a collection of research papers on concrete constituent substitution, and the literature-mined knowledge summary was used for descriptive analysis on current research landscape and topic trends, as well as predictive analysis on underexplored material-application links. This framework can assist industrial efforts to identify well-studied substitution strategies for deployment and academic efforts to pinpoint promising directions for future experimentation.

2. Methodology

The overall literature mining approach is shown in Figure 1, synergizing NLP methods, general data analysis techniques, and domain expertise. We retrieved papers related to the defined problem scope from a literature database and designed templated instruction-completion schemes for the specific information extraction tasks. We formulated these as multiple-choice problems to accommodate the complex one-to-many relations and enable entity inference from complex linguistic settings (detailed in section 2.1). We selected and annotated 102 representative papers based on designed schemes to form the training and testing sets comprising 441 unique data points. We fine-tuned lightweight, open-source LLMs by using instruction and completion pairs in a supervised manner with the training set and the LLM performance was evaluated on the testing set, compared with the baseline GPT-3.5 in-context few-shot learning. We used the best-performing fine-tuned model to extract information from the entire corpus of ~7,000 relevant papers. We constructed knowledge graphs based on extracted data and used these graphs for subsequent data analysis including graph statistics, temporal trends, and link prediction.

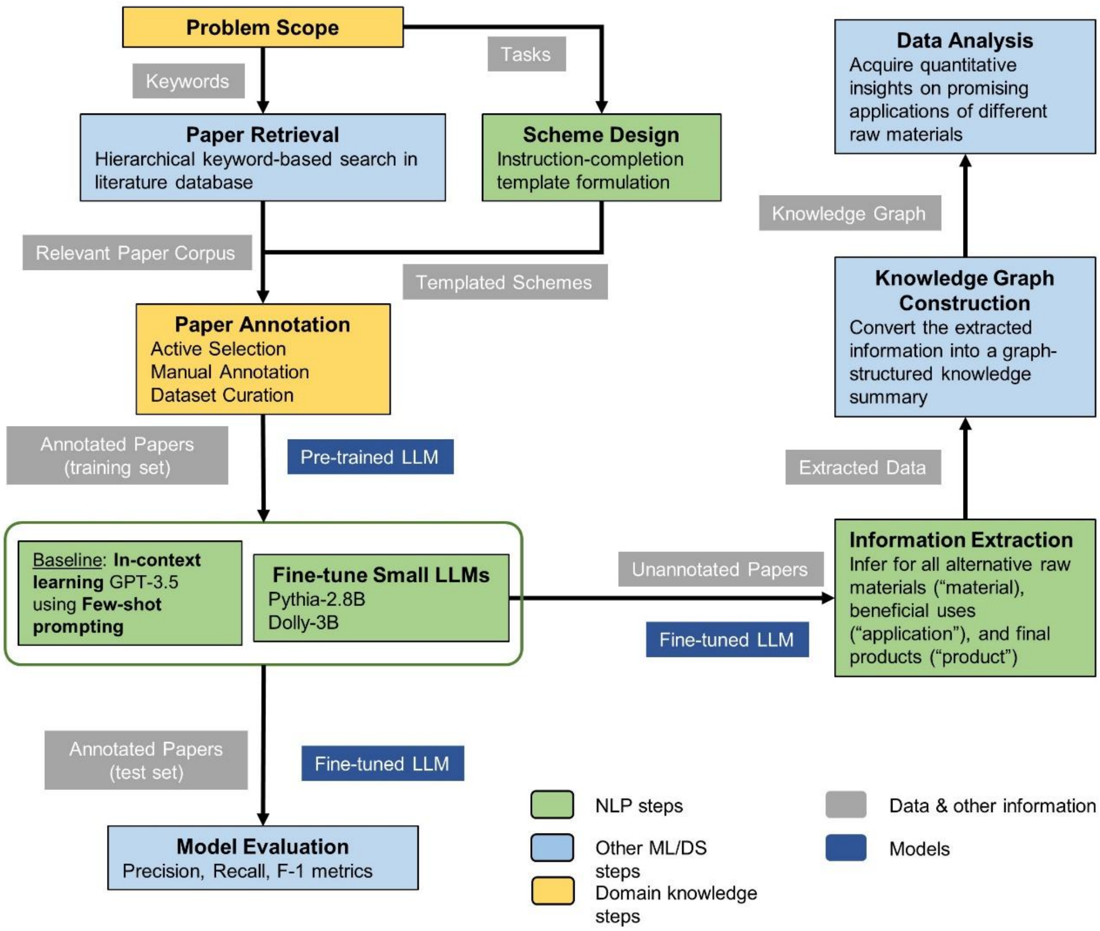

Figure 1. Overall methodological framework and workflow for literature mining and analysis

2.1 Instruction-based Entity Inference Schemes for Complex Linguistic Settings

The linguistic complexity that poses challenges to literature mining on concrete constituent substitution studies and other waste valorization topics includes non-standardized terminology, indirect source information, non-local and non-sequential dependency, and the non-injective mapping relations. Widely-adopted methods for automated scientific literature mining related to materials and chemicals fall short of addressing such issues as they relied on NER from direct NP sources, and the requirement for clear compound identification is considered fundamental in such endeavors (Gupta et al., 2022; Krallinger et al., 2017). Firstly, as the alternative materials explored for beneficial uses are mostly secondary materials or natural composites with variable chemical compositions, the reference to both the materials and the types of applications in constituent substitution are not standardized, rendering direct NER ill-suited. In fact, essential information necessary for determining the specific materials (Figure 2 (b)) and applications (Figure 2 (a)) is commonly delineated through descriptive sentences instead of direct NPs. Meanwhile, the non-local and non-sequential dependencies of syntactic components across sentences are necessary for information extraction due to conditional and referential relations (Figure 2 (b)). The complexity of such relational dependency is further exacerbated by non-injective mapping, as it is common for a paper to contain a material that serves multiple applications (Figure 2 (a)) or multiple materials each serving a different application (Figure 2 (b)).

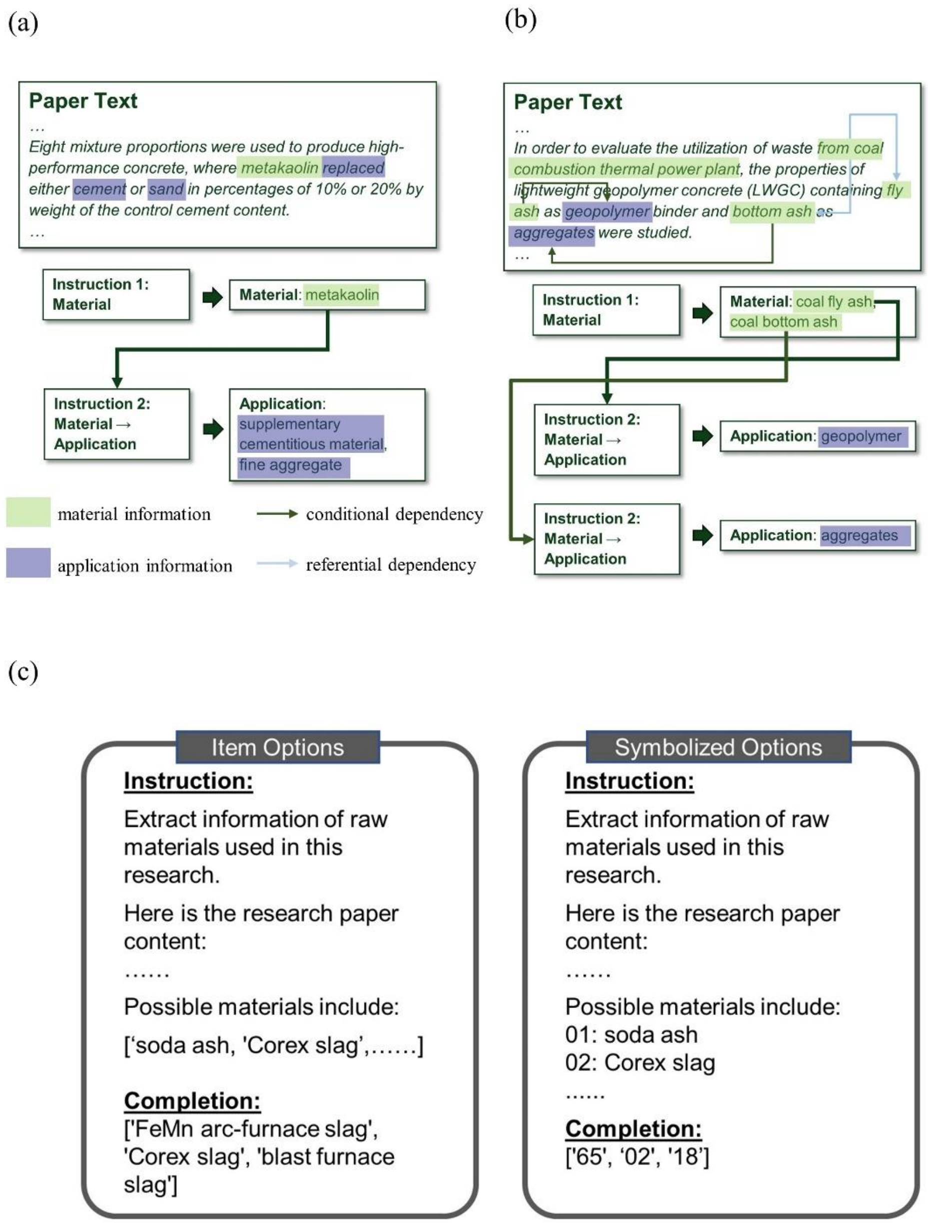

Figure 2. (a) Examples of complex linguistic settings that require indirect information inference from relevant papers; (b) Two instruction-completion schemes with a common multiple-choice formulation but different notation systems

We formulated the information inference tasks as multiple-choice problems to accommodate the need for logical inference from complex linguistic settings by providing options in the instructions to guide model inference. We developed distinct instructions for extracting the three entities, namely material (e.g. "coal fly ash", "silica fume", "waste glass"), application (e.g.

“supplementary cementitious material”, “geopolymer”, “fine aggregate”), and product (e.g. “cement mortar”, “structural concrete”, “concrete pavement”), where the extraction of applications relies on the extraction of materials. We established a set of predefined options containing 76 materials, 13 applications, and 16 products, as shown in Supplementary Table 1. An "unknown" option for negative examples (papers where no answer can be found) was included in each category to address potential model hallucination, a common issue where LLMs generate formally sensible but contextually irrelevant responses (Ji et al., 2023; Zhang et al., 2023). The research community of sustainable concrete primarily focuses on secondary materials and natural composites as alternative raw materials, instead of synthesized compounds, and therefore the introduction of new materials not covered by any previous works is rare. This work thus focuses on statistically summarizing potential types of applications for different alternative raw materials with automated information extraction from existing literature, and the subsequent link prediction aims to reveal potential under-explored applications of such materials, rather than discovering new materials.

Due to the token-based processing, LLMs break down strings of long words or multiple words into multiple tokens during tokenization, meaning the subsequent model behavior depends on the choice of notation. Therefore, a multiple-choice problem formulation was developed by two distinct instruction-completion schemes as shown in Figure 2 (c) using different notations. The two schemes are:

- Item Options: the options in the multiple-choice problem were presented as a list of string items, resulting in each option being tokenized into multiple tokens.
- Symbolized Options: the options were annotated with double-digit notations, allowing each option to be represented by a single token corresponding to its notation.

We permuted the order of choices for all examples to address LLM sensitivity to option ordering (Pezeshkpour and Hruschka, 2023), enhancing the permutation invariance of the model and expanding its learning dataset. We allocated permuted examples from each paper exclusively to either the training or testing set, avoiding data leakage by ensuring that instances with similar contexts to not overlap between training and testing.

2.2 LLM Adaptation and Evaluation

We applied both supervised fine-tuning and in-context few-shot learning on pre-trained LLMs for adaptation to perform the information extraction task based on instruction-completion schemes (Figure 2 (c)). We performed fully supervised fine-tuning on two open-source LLMs, pythia-2.8B (Luo et al., 2023) and dolly-3B (Conver et al., 2023) as they make an ideal contrastive pair, sharing identical tokenizers, overall model architectures, and sizes (by its naming convention, dolly rounds up model sizes to integers), with the primary distinction being that dolly models underwent additional pre-training with common-sense instruction-following data. Meanwhile, in-context, few-shot learning was applied to GPT-3.5 (175B) in this study as a baseline to compare the fine-tuned 2.8B models against. In contrast to fine-tuning, in-context, few-shot learning provides instructions along with demonstration examples to LLMs without modifying the model parameters. Its applicability in smaller models faces serious challenges, but has emerged as the standard practice for the most advanced models with considerably larger sizes and potentially restricted accessibility (Brown et al., 2020; Chowdhery et al., 2022; Dong et al., 2024; OpenAI, 2023). The information extraction performance was evaluated using entity-level precision, recall, and F1 score.

2.3 Knowledge Graph Construction and Temporal Trend Analysis

We used extracted information to build a knowledge graph, linking materials with their applications and products, as well as paper DOIs, where node and edge attributes capture frequencies of entities (materials, applications, products) and their relations for quantitative analysis. We statistically analyzed material-application relations by normalizing pair-level frequencies by material and application respectively. We analyzed temporal trends to compare the research intensity of a certain topic between early (before 2010) and recent (since 2010) periods, applying the concept of trending factor (Eq.(2)) (Zhu et al., 2021), in line with previous works (Zhu and Ren, 2023).

$$F_{early} = 1000 \times \frac{f_{yr \leq 2010}}{N_{yr \leq 2010}};\ F_{recent} = 1000 \times \frac{f_{yr > 2010}}{N_{yr > 2010}} \quad (1)$$

$$TF = log\left(\frac{F_{recent}}{F_{early}}\right) \quad (2)$$

F_{early} and F_{recent} denote the frequencies for both periods normalized per 1000 papers for better communication of results, f_{period} and N_{period} are the raw count of papers related a certain topic during the period and the total amount of collected papers within that period, respectively. TF is the trending factor for the topic.

In addition to the binary-time trending factors, continuous-time topic rank evolution has been used in previous study on topic association network of materials science literature to track annual changes of main topics (Choi and Lee, 2024), which is also performed for this study on material and application topics to reveal the evolution of research landscape.

2.4 Link Prediction and Uncertainty Quantification

To perform link prediction to identify potential new material-application pairs, we developed an interpretable algorithm (Eq.(4)), leveraging weighted Jaccard node similarity (Eq.(3)) (modified from (Li and Li, 2021)) based on the original local graph structure.

$$J(m_i, m_j) = \frac{\sum_{a \in N_{APP}(m_i) \cap N_{APP}(m_j)} \min\left(\hat{w}(m_i,a), \hat{w}(m_j,a)\right)}{\sum_{a \in N_{APP}(m_i) \cup N_{APP}(m_j)} \max\left(\hat{w}(m_i,a), \hat{w}(m_j,a)\right)} \quad (3)$$

$$S(m_i, a_k)|_{\tilde{w}(m_i,a_k)=0} = \sum_{j=1}^{|MAT|} J(m_i, m_j) \cdot \tilde{w}(m_j, a_k) \quad (4)$$

$J(m_i, m_j)$ is the weighted Jaccard coefficient between two material nodes m_i, m_j, $N_{APP}(m_i)$ denotes the set of all application node neighbors of material node m_i, with ∩ and ∪ denoting set intersection and union respectively. $\hat{w}(m_i, a)$ denotes the edge weight between a material-application node pair (m_i, a) normalized by material (so that $\sum_{a \in N_{APP}(m_i)} \hat{w}(m_i, a) = 1$)). $S(m_i, a_k)|_{\tilde{w}(m_i,a_k)=0}$ denotes the predicted link **S**core between the material node m_i and application node a_k, under the condition that the two nodes are not connected in the original knowledge graph. MAT denotes the set of all material nodes, and $\tilde{w}(m_j, a_k)$ denotes the edge weight between node pair (m_j, a_k) normalized by application (so that $\sum_{j=1}^{|MAT|} \tilde{w}(m_j, a_k) = 1$). Intuitively, the method attributes higher potential to the new material-application links where materials similar to the material of interest are commonly studied for the application, providing insights on potentially promising new research directions.

As noises in the literature-mined knowledge graph may render uncertainty cascaded to link prediction results, we considered combining the ideas of Monte Carlo sampling and graph perturbation for robustness improvement and uncertainty quantification. A perturbation-based link prediction framework was proposed in a previous work (Wang et al., 2016), but its latent space perturbation method fails to accommodate the bipartite nature of our knowledge graph or our link prediction method based on graph-contextualized material node similarity. Inspired by the idea of graph perturbation with random noises, we designed the following algorithm (Eq. (5)-(8)).

$$G_r(MAT, APP, E_r) = f(G(MAT, APP, E), r, p), \forall r = 1,2, \ldots \quad (5)$$

$$f\text{: for edge } E_r(m,a)\text{, set its weight } w^{(G_r)}(m,a) = \begin{cases} w(m,a), & \text{with prob. } p \\ w(m,a) + v, & \text{with prob. } (1-p) \end{cases} \quad (6)$$

$$J^{(G_r)}(m_i, m_j) = \frac{\sum_{a \in N_{APP}^{(G_r)}(m_i) \cap N_{APP}^{(G_r)}(m_j)} \min\left(\widehat{w}^{(G_r)}(m_i,a), \widehat{w}^{(G_r)}(m_j,a)\right)}{\sum_{a \in N_{APP}^{(G_r)}(m_i) \cup N_{APP}^{(G_r)}(m_j)} \max\left(\widehat{w}^{(G_r)}(m_i,a)\ \widehat{w}^{(G_r)}(m_j,a)\right)} \quad (7)$$

$$S^{(G_r)}(m_i, a_k)\big|_{\widetilde{w}^{(G_r)}(m_i,a_k)=0} = \sum_{j=1}^{|MAT|} J^{(G_r)}(m_i, m_j) \cdot \widetilde{w}^{(G_r)}(m_j, a_k) \quad (8)$$

$G_r(MAT, APP, E_r)$ (or simply G_r) denotes a perturbed realization of the original knowledge graph $G(MAT, APP, E)$ (Eq. (5)). Perturbation happens on any edge with probability $1-p$, and each edge-wise perturbation is the addition of a noise v onto the original edge weight (Eq. (6)). The weighted Jaccard coefficients and link prediction scores are subsequently calculated for each perturbed graph G_r (Eq. (7)-(8)). In practice, we perturb the original knowledge graph by adding unit edge weight increments ($v = 1$) to randomly sampled material-application node pairs, including both current existing links (edge weight +1) and non-existing links (set edge with weight 1). This approach can be translated into having equivalent effects of hypothetically adding into our knowledge base an example containing the corresponding material-application pair, which models the uncertainty stemming from potentially missed papers in literature collection and missed extractions from collected papers. The random sampling-perturbation workflow was repeated 100 times and resulted in 100 different realizations of perturbed knowledge graphs to calculate modified weighted Jaccard coefficients and subsequent link predictions scores, with each material-application node pair having 20% probability (i.e. $p = 0.8$) of being perturbed in every random realization. The mean Jaccard coefficients and link prediction scores are computed across the 100 perturbed graph realizations for robust link prediction, with uncertainty quantified through standard deviations across realizations.

3. Results and Analysis

We evaluate and compare the LLM performance for entity inference across different models' adaptation with different instruction schemes. We analyze the most well-studied materials and applications within the current research landscape and investigate the temporal trends reflecting the evolution of topic popularity. Finally, we use link prediction on the knowledge graph built from literature-mined information.

3.1 Entity-Inference Performance

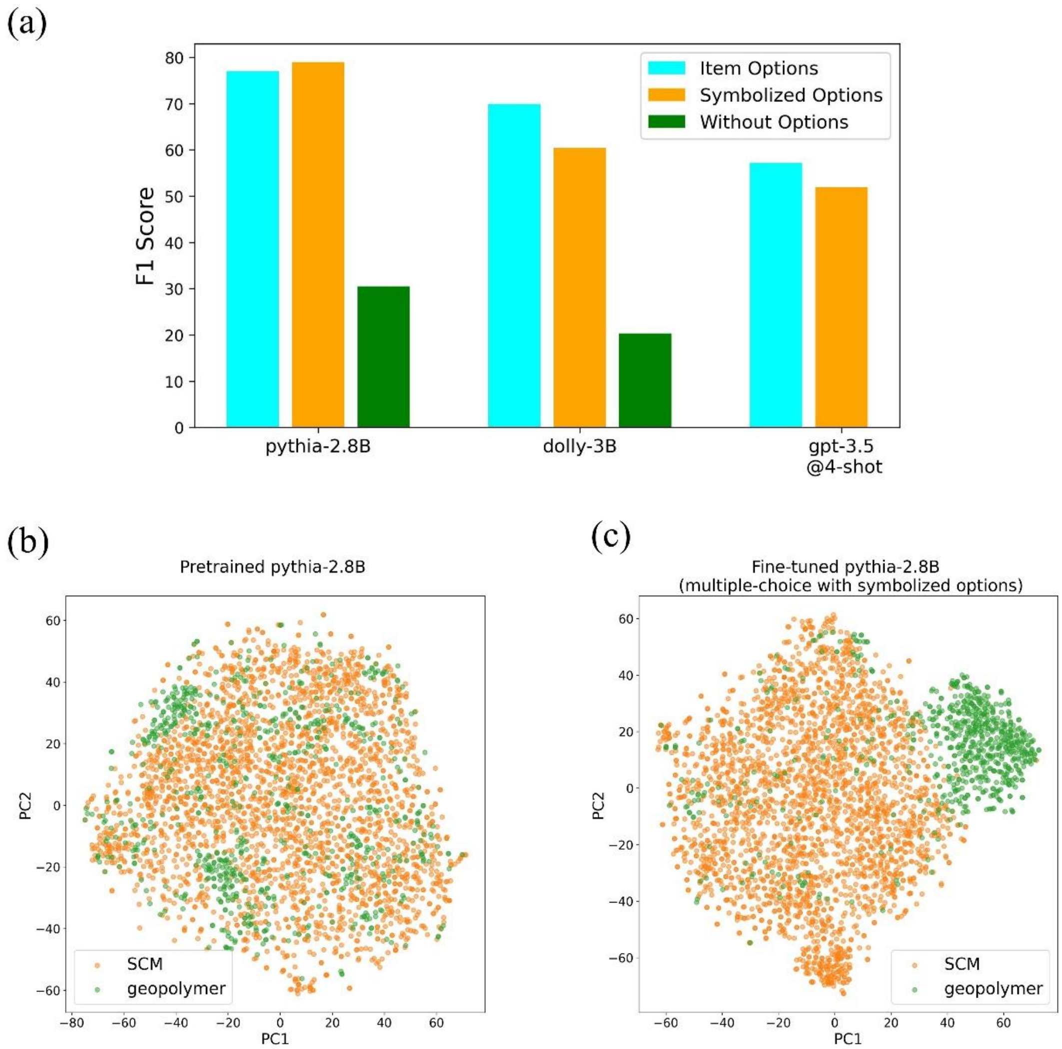

Figure 3. (a) Model F1 scores under different instruction schemes; (b)(c) Text embeddings for paper paragraphs on supplementary cementitious materials (SCM, orange) and geopolymer (green), with pretrained and fine-tuned pythia-2.8B, respectively

We evaluated LLM performance on the entity-level, comparing the LLM-generated completions (lists of materials, applications, products, in the form of strings or symbolized notations) with ground truth annotations. The results revealed that, by combining multiple-choice instruction schemes with instruction-following fine-tuning of small LLMs of just 2.8B parameters, the proposed method yields a F1-score of 79.0%, precision of 81.2% and recall of 77.0%. As shown in Figure 3 (a), the best F1 scores for both pythia and dolly following supervised fine-tuning outperform the in-context learning baseline using the advanced large models (GPT-3.5 of 175B parameters), saving memory use and training time by over 95% compared to previous scientific literature mining works that relied on fine-tuning large models over 70B to achieve F1 scores above 80% (Dagdelen et al., 2024; Walker et al., 2023). For pythia-2.8B, the comparison of different instruction schemes reveals that symbolized notations prove to further boost accuracy by over 2% than plain natural language in the multiple-choice instruction design. The fact that pythia outperforms dolly post-fine-tuning in this study suggests the importance of task-specific model selection, as the common-sense instruction-tuning of dolly does not necessarily enhance its performance in scientific tasks.

To further understand the performance improvement after fine-tuning through language representations, using the examples of SCMs versus geopolymer, Figure 3 (b)-(c) visualizes the final vectorized representations of each textual example from the best-performing LLM (pythia-2.8B) and projects them to 2-D space with tSNE dimensionality reduction (van der Maaten and Hinton, 2008). Compared to the pre-trained model (Figure 3 (b)), the model fine-tuned under multiple-choice instructions with symbolized options (Figure 3 (c)) can partition the papers focusing on the two alternative binder mechanisms.

3.2 Quantitative Summary of Constituent Substitution Research Landscape

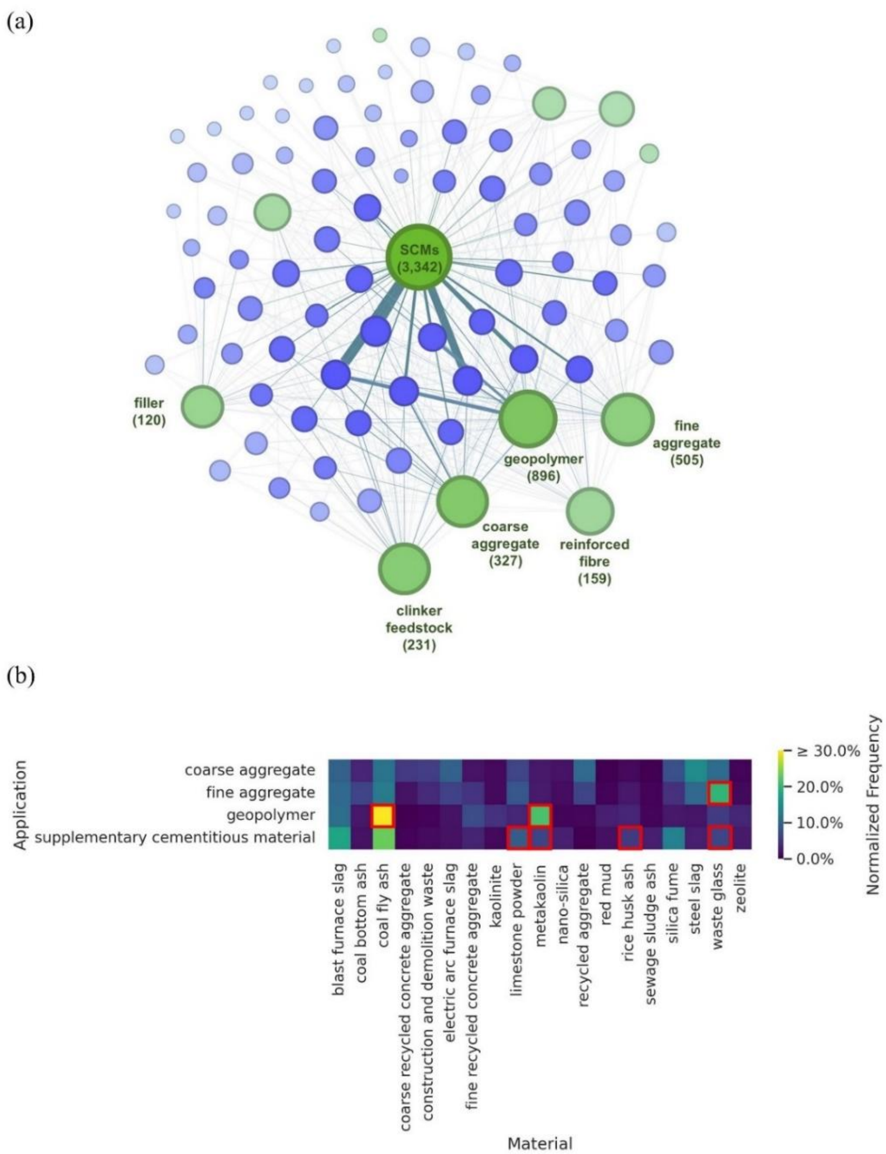

Figure 4. (a) Knowledge graph containing applications (green) and materials (blue) nodes (node sizes and edge weights shown in log scale); (b) Frequencies of top materials for selected applications (normalized by application)

This analysis provides quantitative indication of the overall research landscape. Figure 4 (a) visualizes a subgraph of the overall knowledge graph presenting the high-frequency applications mentioned in the corpus and their most frequently investigated raw materials. SCMs are the most widely studied application with a dominating frequency of 3,342, followed by geopolymer with 896 mentions, indicating ubiquity of alternative binder study within the literature. Given that the chemical processes used in cement production are a main driver of emissions in the concrete industry, constituting at least 70% of its GHG emissions (Habert et al., 2020), alongside high temperature heating processes (Miller et al., 2021), material substitution is desirable. Alternative clinker feedstocks, which reduce emissions from cement manufacturing life cycle instead of directly replacing ordinary cement, appears 231 times within the corpus. Other commonly studied applications of alternative raw materials include fine aggregate (505), coarse aggregate (327), reinforced fibre (159) and filler (120), trailing behind cement-related applications because they offer less potential for CO_2 emission reduction (Plaza et al., 2021; Sabău et al., 2021) and generally have lower economic benefits (Kirthika et al., 2020).

The most prominent materials for each application are further illustrated in Figure 4 (b) (the normalized frequencies for all material-application pairs are shown in Supplementary Figure 1). Coal fly ash (coal FA), blast furnace slag (BFS), and silica fume (SF), the three conventional residues accepted for use as SCMs in current industrial practice, account for a total of 41.2% of studies (coal FA 17.6%, BFS 13.5%, SF 10.1%). In addition, natural minerals and other secondary materials including limestone powder, waste glass, metakaolin, and rice husk ash each account for 4.0-6.7% of studies on SCMs. For applications in geopolymers, coal fly ash and metakaolin stand out as the most extensively studied raw materials, accounting for 23.5% and 16.9% of related papers respectively, followed by BFS (8.6%) and waste glass (4.2%). Geopolymers have garnered significant research interest due to their potential benefits of reducing carbon footprint and improving durability. However, geopolymers have not been widely adopted in the concrete industry because of the technical challenges related to the complex alkaline activation processes with harsh corrosive chemicals, which complicates large-scale onsite handling for consistency and setting time, along with economic barriers stemming from the costs of with chemical activators and operations (Upadhyay et al., 2022; Van Deventer et al., 2012; Wu et al., 2019; Y. Zhang et al., 2024). Furthermore, the studies on waste glass for fine aggregate stands out with a normalized frequency of 16.4%, which matches domain knowledge as the advantages of incorporating waste glass as fine aggregate have been studied due to its pozzolanic reaction on surface that enhances the mechanical properties, while it is not suitable for use as coarse aggregate because of its smooth surface (Harrison et al., 2020).

In light of the expected decline of industrial residue supply for already-commercialized SCMs (Juenger et al., 2019; Snellings et al., 2023), these findings pinpoint other constituent substitution strategies as alternative long-term sustainable solutions in the concrete industry, including the use of limestone powder and rice husk ash for SCMs, metakaolin for SCMs and geopolymer, and waste glass for SCMs and fine aggregate.

3.3 Temporal Trends of Research Topics

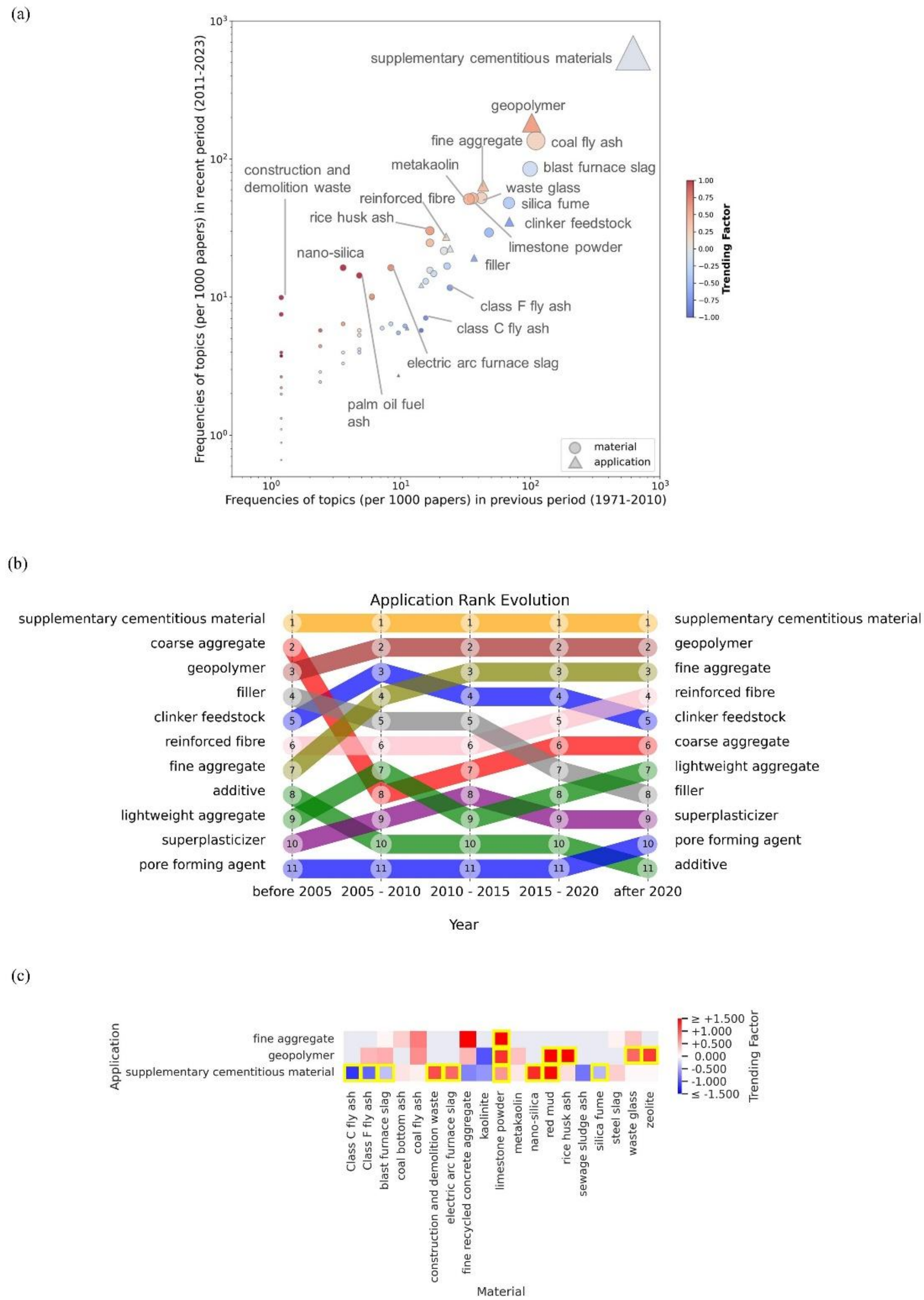

Figure 5. (a) Trending factors of each material and application topic comparing previous (1971-2010) and recent (2011-2023) periods, with frequencies per 1000 papers; (b) Rank evolution of application topics over time; (c) Trending factors for notable material-application pairs associated with SCMs, geopolymer and fine aggregate. (The links with fewer than 10 mentions in both pre-2010 and post-2010 studies are greyed out)

We present temporal shifts across research topics in Figure 5 (a) through trending factors as defined in Eq.(1)-(2), with a greater positive (negative) value indicating a greater increase (decrease) of research interests in recent years compared to the earlier period. For applications, the general interest for geopolymer (with a trending factor of +0.578) and fine aggregate (+0.391) increased dramatically, accompanied by a moderate increase of reinforced fibre (+0.195), whereas the studies on clinker feedstock (-0.675) and filler (-0.653) declined significantly in recent years. SCM mentions declined slightly after 2010 with a trending factor of -0.058, although it remains the most popular in academic research. Pertaining to raw materials, widely applied residues including SF (-0.354), BFS (-0.164), class C FA (-0.794) and class F FA (-0.309) are less studied in the recent period, while an increased interest is observed in rice husk ash (+0.585), palm oil fuel ash (+1.091), limestone powder (+0.363), waste glass (+0.217), metakaolin (+0.419), nano-silica (+1.508), electric arc furnace slag (+0.662), construction and demolition waste (+2.104), etc. The findings indicate a systematic shift toward exploring a broader range of alternative raw materials for sustainable applications, reflecting efforts to address the anticipated decline of supply in coal FA and BFS. (Juenger et al., 2019; Snellings et al., 2023).

Whereas the trending factors quantify the relative importance change of research topics over time in a binary time period definition, topic rank evolution (Figure 5(b)) depicts the changes of main topics over a continuous timeline. The ranking of fine aggregate continuously rises over time, from the 7th-ranked to the 3rd-ranked application topic, whereas geopolymer and reinforced fibre have risen from the 3rd place to the 2nd place, and from the 6th place to the 4th place, respectively. Meanwhile, the ranking of filler significantly dropped through the years from the 4th most widely investigated topic to the 8th, and clinker feedstock dropped from the 3rd place in 2005-2010 to the 5th place after 2020. Aside from the changes that reaffirm the trending factor results, rank evolution analysis further reveals that the research popularity of coarse aggregate has significantly dropped from the 2nd place before 2005 to 6th-8th places in following years.

Despite the continued popularity of SCMs studies, Figure 5(c) reveals a shift to a broader range of raw materials (complete trending factor results of material-application pairs are shown in Supplementary Figure 2). Links between SCM and class C FA (-1.193), class F FA (-0.904), SF (-0.417), and BFS (-0.375) have been trending down, as they have been extensively studied (Figure 4 (b)) and may face limited supply. Meanwhile, SCM studies of red mud (+1.462), nano-silica (+1.252), construction and demolition waste (+1.002), electric arc furnace slag (+0.881) and limestone powder (+0.619) have increased. Among all materials on geopolymer research, 42 of 75 trend up and only 6 trend down, with rice husk ash (+6.305), red mud (+6.134), limestone powder (+1.200), zeolite (+1.148) and waste glass (+0.880) leading the increases. Despite not typically participating in pozzolanic or geopolymerization reactions, limestone powder improves cement hydration through nucleation effects as SCM (Ji et al., 2024; Wang et al., 2019) and enhances properties including strength and workability when blended to geopolymer mixtures (Rashad, 2022; Rashad et al., 2023), and it also leads the recent surge in fine aggregate studies (+6.406), rendering it a promising material in industrial practice across different types of applications.

3.4 Link Prediction for Under-studied Material-Application Pairs

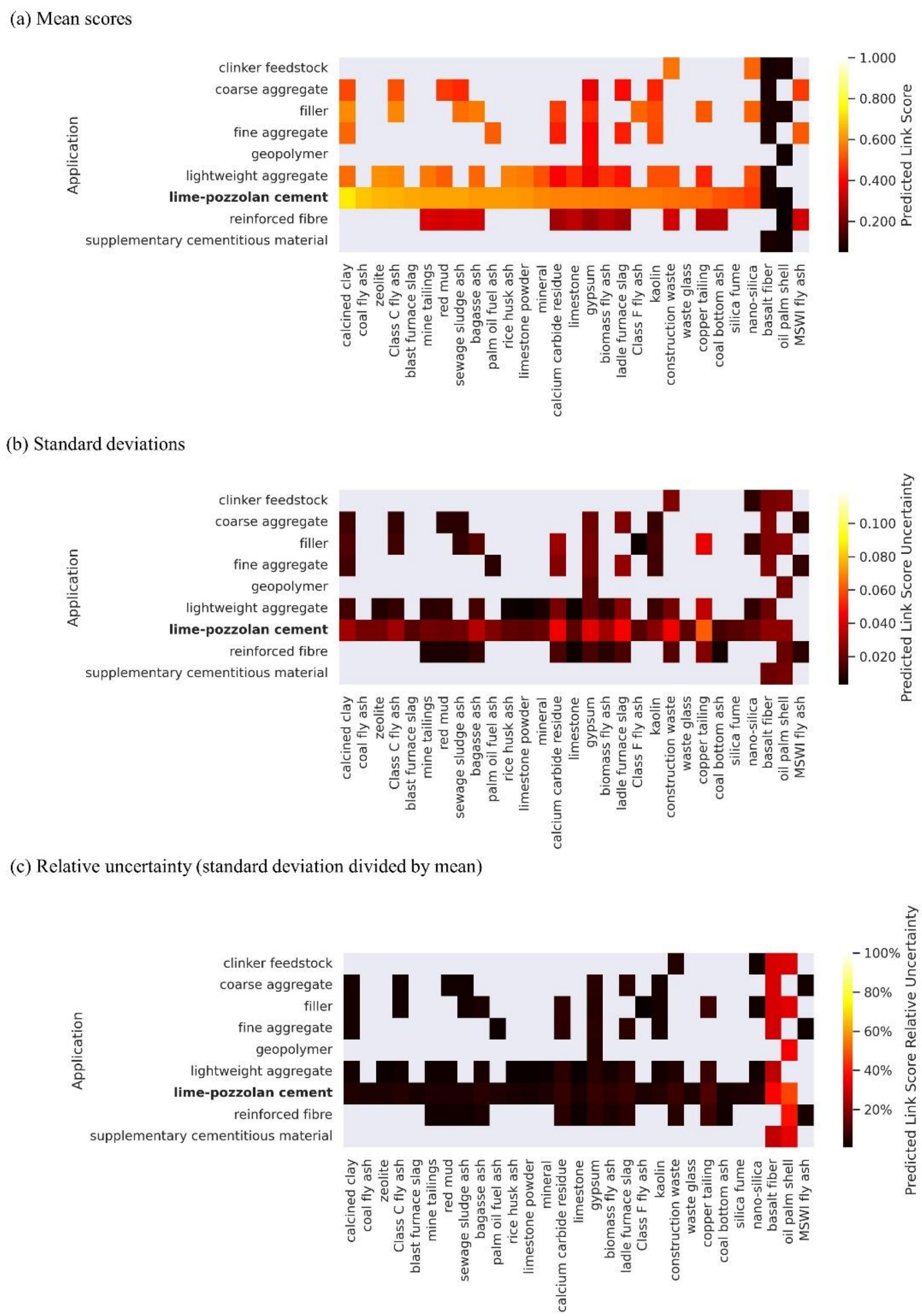

Figure 6. Predicted link results for material-application pairs underreported in retrieved corpus: (a) Mean link prediction scores across 100 randomly perturbed knowledge graph realizations; (b) Standard deviations of link prediction scores; (c) Relative uncertainty of link prediction scores (link-wise standard deviation divided by link-wise mean score)

The local subgraph structures within the knowledge graph containing literature-mined material-application links were used to quantify material node similarity (see Section 2.3 and

Supplementary Figure 4), which subsequently yielded predicted link scores for missing material-application pairs. The application of lime-pozzolan cement is shown to be associated with most high-score predicted links in Figure 6(a) (about 95% of the material-application pairs with a predicted score above 0.6), indicating that lime-pozzolan cement is currently under-explored but promising to investigate. The lime-pozzolan cement refers to the lime reacting with water and pozzolanic materials to form calcium silicate/aluminate hydrates (C-S-H or C-A-H) to provide binding properties (Malathy et al., 2023; Zhang et al., 2020). Although lime-pozzolan cement presents a lower carbon footprint than OPC, the slow strength gains and lower early-stage strength have limited its application in the concrete industry, suggesting the research direction of mixing lime-pozzolan with OPC for engineering practice (Baghabra Al-Amoudi et al., 2022; Grist et al., 2016; Malathy et al., 2023).

The prominent predicted materials for future lime-pozzolan cement studies include calcined clay (with a score of 0.730±0.030), coal FA (0.677±0.022), zeolite (0.659±0.022), Class C fly ash (0.651±0.030), blast furnace slag (0.633±0.017), mine tailings (0.632±0.021), red mud (0.632±0.020), sewage sludge ash (0.630±0.021), and bagasse ash (0.618±0.033), all of which were previously mentioned as pozzolanic materials in the publications (Chusilp et al., 2009; Çokça, 2001; Liu et al., 2020; Lynn et al., 2015; Najimi et al., 2012; Nedunuri et al., 2020; Tironi et al., 2013; Yang et al., 2019). Only municipal solid waste incinerator (MSWI) fly ash is greyed out in Figure 6(a), as it has been previously studied for blending with lime-pozzolan cement to stabilize/solidify heavy metals in ash with low-cost (Ubbriaco, 1996), while the high content of hydrated lime in ash and active pozzolan components suggests its potential uses (Chen et al., 2023; Marieta et al., 2021; Tang et al., 2016).

Model uncertainty was further quantified through standard deviation of link prediction scores (Figure 6(b)) and the relative uncertainty (ratios of link-wise standard deviations to mean scores) (Figure 6(c)) across the 100 realizations. It is observed that the standard deviations for all prediction scores are under 0.120, with over 80% of the standard deviations under 0.040. Meanwhile, all of the top 20 predicted links for lime-pozzolan cement have a relative uncertainty below 10% (Figure 6(c)), highlighting the robustness of the prediction results.

4. Limitation and Discussion

Despite the promising results, there are certain challenges and limitations that warrant discussion to contextualize our findings and clarify our contributions, with future work expected to address relevant open questions.

Challenges in Extracting More Detailed Information The current work summarizes existing knowledge and conducts link prediction to assess materials for constituent substitution, accommodating named entity inference from complex contexts. Extracting comprehensive material characteristics and mixture factors could enhance the predictions accuracy and optimize the product performance. Material substitution often involves detailed factors beyond material-application pairing, such as chemical composition, physical properties, mixture proportions, performance and material substitution tradeoffs, reaction kinetics, and thermodynamic constraints, which can be valuable in decision-making. Nevertheless, in many cases covering fine and coarse aggregates substitutions, reinforced fibers, additives, etc., chemical composition is usually not reported in the article context (Khan and Ali, 2016; Wang et al., 2020; Zeng et al., 2020). Furthermore, some detailed materials data such as chemical composition, mixture proportions and product performance are provided in tables and figures instead of text (Baeza et al., 2014; Copetti et al., 2020; Mohebi et al., 2015; Praneeth et al., 2020; Tanyildizi and Coskun, 2008; Thongsanitgarn et al., 2014; Van Den Heede et al., 2010; Wang et al., 2008). Additionally, time-dependent kinetic and thermodynamic phase equilibrium are typically reported in figures (Cheung et al., 2011; Kamali et al., 2003; Lothenbach, 2010; Saillio et al., 2021; Thomas et al., 2011), and are generally not relevant to non-SCM application categories covered in this work. Automated table understanding (Sarkar and Lausen, 2023; Sui et al., 2024) and chart understanding (Han et al., 2023; Masry et al., 2024) using LLMs are advanced active research questions for foundational NLP works, and require dedicated efforts in developing specialized methodologies such as table parsing pretraining, visual instruction tuning, and multimodal chain-of-thought (Herzig et al., 2020; Liu et al., 2023; Z. Zhang et al., 2024). Adaptation of these methods for scientific domain questions may be incorporated in future studies to extract more comprehensive information from multimodal sources. Despite not incorporating aforementioned detailed material characteristics and behavior information, the analysis and findings of this work are expected to be valuable for practical applied scenarios. The quantitative summary of material-application pairs can be coupled with experimental trials to diversify industrial sustainable strategies, as manufacturers can't optimize the characteristics of secondary materials they receive, and data points from experimental research can't cover the vast compositional space of real-world composites. Meanwhile, experimentalists can draw inspiration from our results and determine the underexplored material-application pairs to prioritize for future research development.

Challenges in Distinguishing Reaction Mechanisms The research scope of this work covers material substitution not only limited to SCMs, but also other construction applications including coarse or fine aggregate, geopolymer binder, filler, reinforced fiber, etc. Currently, the terminology for concrete constituent substitution was standardized based on mix formulation, so materials substituting cement were classified as SCMs, despite the specific reaction mechanisms may differ, which were not further addressed due to the research scope of this study and the ambiguity of classification in certain cases. From a domain knowledge perspective, the SCMs are classified to inert, pozzolanic, and hydraulic materials in cementitious systems according to isothermal calorimetry heat release and $Ca(OH)_2$ consumption, where the most commonly used measurement methods are summarized in the recent technote from U.S. Department of Transportation with quantitative assessment (Skibsted and Snellings, 2019; Suraneni and Weiss, 2017; US Department of Transportation Federal Highway Administration, 2025). For instance, limestone powder (Section 3.3) typically does not chemically react when it is for SCM use, but primarily promotes cement hydration through nucleation effects, where the fine limestone particles improve workability, increase packing density, and reduce voids (Moon et al., 2017; Wang et al., 2018). However, the nucleation effects on hydration and pozzolanic reactions can be overlapped in the calorimetry curves, posing challenges to establish distinctive boundaries in the terminology.

Limitation of Literature Collection Due to the limited access to the initial literature database and the constraints of keyword-based retrieval, some important papers may not be fully covered in our collection. For example, materials including calcined clay and BFS were studied for lime-pozzolan cement in a previous paper not included in the database (Walker and Pavía, 2011). The maintenance of an up-to-date literature database with access to papers from all major publishers is expected to remain a fundamental challenge, but future works may explore more comprehensive topic-specific paper collection approaches within existing literature databases, by employing crawler-based or embedding-based strategies to complement the traditional keyword-based ones.

Need for Future Knowledge Base Enrichment and Experimental Validation Despite materials outside the pre-defined multichoice options can be rare (see Section 2.1), future research may introduce new materials in marginal cases. By incorporating the “unknown” option, the literature mining framework can identify marginal cases without hallucinating materials from the known list. When treating research papers, “unknown” materials reported by the model will require a case-by-case analysis by domain experts, and further annotation may be performed to enrich the knowledge base. Additionally, this work builds upon the previous automated literature mining works in materials domains that focused on knowledge base construction with evaluation through computational metrics (Dagdelen et al., 2024; Kim et al., 2017; Kumar et al., 2023; Venugopal and Olivetti, 2024; Walker et al., 2023). To extend the efforts of this work, experimental research is needed to validate the predicted functionality of materials.

5. Conclusion

We developed an LLM-empowered literature-mining method to complete entity inference from domain-specific complex linguistic settings, using symbolized multiple-choice instructions and supervised fine-tuning of computationally efficient small LLMs. This strategy is time- and memory-efficient in model tuning and addresses a wide range of problems including non-standardized terminology, indirect text mentions, non-local and non-sequential dependency, and the non-injective mapping relations. As such complexity is prevalent in papers related to resources management and beneficial uses of alternative materials, the developed methodology could support future literature mining works in related fields as a generalizable framework.

In this study, the method was applied to acquire a systematic knowledge summary of studies on concrete constituent substitution. The literature-mined information was analyzed through statistical and graph-based quantitative methods to identify the hotspots within the research landscape, assisting the industry to prioritize areas for further deployment. SCMs are the most widely studied application, followed by geopolymers. Most prominent materials for SCMs include 3 industrial residues, namely coal FA, BFS and SF, as well as natural minerals and other secondary materials including limestone powder, waste glass, metakaolin, and rice husk ash. Coal FA and metakaolin are the most extensively studied materials as geopolymers, while studies on waste glass for fine aggregate purposes are also highlighted.

Temporal trends of different research topics were further analyzed, revealing a systematic shift of research interest in the recent period. Among all applications, geopolymer and fine aggregate studies have become significantly more popular in recent years across different raw materials, while clinker feedstock and filler studies have been in decline over time. SCMs remained popular over time, but the materials studied have been significantly diversified, with nano-silica, red mud and rice husk ash trending up and the well-studied industrial residues facing a declining supply (coal FA, BFS, SF) trending down. Limestone powder was also found to be a promising raw material for alternative binders with its SCM, geopolymer and fine aggregate studies all surging after 2010.

Meanwhile, some of the currently underexplored material-application links have been predicted to be potentially promising directions for future research. Lime-pozzolan cement stands out to be a notably potential use for several different materials including calcined clay, coal FA, zeolite, Class C FA, BFS, mine tailings, red mud, sewage sludge ash, and bagasse ash. The potential of such materials in lime-pozzolan cement mostly come from the pozzolanic reactivity that contributes to the formation of strength-providing calcium silicate/aluminate hydrate.

Acknowledgements

The work is based in part on the financial support of the MIT-IBM Watson AI Lab. We also acknowledge support from NSF CAREER #1751925.

References

Baeza, F., Payá, J., Galao, O., Saval, J.M., Garcés, P., 2014. Blending of industrial waste from different sources as partial substitution of Portland cement in pastes and mortars. Constr. Build. Mater. 66, 645–653. https://doi.org/10.1016/j.conbuildmat.2014.05.089

Baghabra Al-Amoudi, O.S., Ahmad, S., Maslehuddin, M., Khan, S.M.S., 2022. Lime-activation of natural pozzolan for use as supplementary cementitious material in concrete. Ain Shams Eng. J. 13, 101602. https://doi.org/10.1016/j.asej.2021.09.029

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., 2020. Language Models are Few-Shot Learners. 34th Conf. Neural Inf. Process. Syst. NeurIPS 2020.

Chen, Z., Li, J.-S., Poon, C.-S., Jiang, W.-H., Ma, Z.-H., Chen, X., Lu, J.-X., Dong, H.-X., 2023. Physicochemical and pozzolanic properties of municipal solid waste incineration fly ash with different pretreatments. Waste Manag. 160, 146–155. https://doi.org/10.1016/j.wasman.2023.02.016

Cheung, J., Jeknavorian, A., Roberts, L., Silva, D., 2011. Impact of admixtures on the hydration kinetics of Portland cement. Cem. Concr. Res. 41, 1289–1309. https://doi.org/10.1016/j.cemconres.2011.03.005

Choi, J., Lee, B., 2024. Quantitative Topic Analysis of Materials Science Literature Using Natural Language Processing. ACS Appl. Mater. Interfaces 16, 1957–1968. https://doi.org/10.1021/acsami.3c12301

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H.W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., Reif, E., Du, N., Hutchinson, B., Pope, R., Bradbury, J., Austin, J., Isard, M., Gur-Ari, G., Yin, P., Duke, T., Levskaya, A., Ghemawat, S., Dev, S., Michalewski, H., Garcia, X., Misra, V., Robinson, K., Fedus, L., Zhou, D., Ippolito, D., Luan, D., Lim, H., Zoph, B., Spiridonov, A., Sepassi, R., Dohan, D., Agrawal, S., Omernick, M., Dai, A.M., Pillai, T.S., Pellat, M., Lewkowycz, A., Moreira, E., Child, R., Polozov, O., Lee, K., Zhou, Z., Wang, X., Saeta, B., Diaz, M., Firat, O., Catasta, M., Wei, J., Meier-Hellstern, K., Eck, D., Dean, J., Petrov, S., Fiedel, N., 2022. PaLM: Scaling Language Modeling with Pathways.

Chusilp, N., Jaturapitakkul, C., Kiattikomol, K., 2009. Utilization of bagasse ash as a pozzolanic material in concrete. Constr. Build. Mater. 23, 3352–3358. https://doi.org/10.1016/j.conbuildmat.2009.06.030

Çokça, E., 2001. Use of Class C Fly Ashes for the Stabilizationof an Expansive Soil. J. Geotech. Geoenvironmental Eng. 127, 568–573. https://doi.org/10.1061/(ASCE)1090-0241(2001)127:7(568)

Conver, M., Hayes, M., Mathur, A., 2023. Free Dolly: Introducing the World's First Truly Open Instruction-Tuned LLM.

Copetti, C.M., Borges, P.M., Squiavon, J.Z., Da Silva, S.R., De Oliveira Andrade, J.J., 2020. Evaluation of tire rubber surface pre-treatment and silica fume on physical-mechanical behavior and microstructural properties of concrete. J. Clean. Prod. 256, 120670. https://doi.org/10.1016/j.jclepro.2020.120670

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A.S., Ceder, G., Persson, K.A., Jain, A., 2024. Structured information extraction from scientific text with large language models. Nat. Commun. 15, 1418. https://doi.org/10.1038/s41467-024-45563-x

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Chang, B., Sun, X., Li, L., Sui, Z., 2024. A Survey on In-context Learning.

Ellis, L.D., Badel, A.F., Chiang, M.L., Park, R.J.-Y., Chiang, Y.-M., 2020. Toward electrochemical synthesis of cement—An electrolyzer-based process for decarbonating $CaCO_3$ while producing useful gas streams. Proc. Natl. Acad. Sci. 117, 12584–12591. https://doi.org/10.1073/pnas.1821673116

Grist, E.R., Paine, K.A., Heath, A., 2016. HYDRAULIC LIME – POZZOLANS: PROPERTIES, USES AND RESEARCH NEEDS. 9th Int. Concr. Conf. 2016 Environ. Effic. Econ. Chall. Concr.
Gupta, T., Zaki, M., Krishnan, N.M.A., Mausam, 2022. MatSciBERT: A materials domain language model for text mining and information extraction. Npj Comput. Mater. 8, 102. https://doi.org/10.1038/s41524-022-00784-w
Habert, G., Miller, S.A., John, V.M., Provis, J.L., Favier, A., Horvath, A., Scrivener, K.L., 2020. Environmental impacts and decarbonization strategies in the cement and concrete industries. Nat. Rev. Earth Environ. 1, 559–573. https://doi.org/10.1038/s43017-020-0093-3
Han, Y., Zhang, C., Chen, X., Yang, X., Wang, Z., Yu, G., Fu, B., Zhang, H., 2023. ChartLlama: A Multimodal LLM for Chart Understanding and Generation. https://doi.org/10.48550/arXiv.2311.16483
Harrison, E., Berenjian, A., Seifan, M., 2020. Recycling of waste glass as aggregate in cement-based materials. Environ. Sci. Ecotechnology 4, 100064. https://doi.org/10.1016/j.ese.2020.100064
Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., Steinhardt, J., 2021. Measuring Massive Multitask Language Understanding.
Herzig, J., Nowak, P.K., Müller, T., Piccinno, F., Eisenschlos, J.M., 2020. TAPAS: Weakly Supervised Table Parsing via Pre-training, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 4320–4333. https://doi.org/10.18653/v1/2020.acl-main.398
Ji, G., Chi, H., Sun, K., Peng, X., Cai, Y., 2024. Effect of limestone waste on the hydration and microstructural properties of cement-based materials. Constr. Build. Mater. 443, 137784. https://doi.org/10.1016/j.conbuildmat.2024.137784
Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P., 2023. Survey of Hallucination in Natural Language Generation. ACM Comput. Surv. 55, 1–38. https://doi.org/10.1145/3571730
Jindal, R., Malhotra, R., Jain, A., 2015. Techniques for text classification: Literature review and current trends. webology.
Juenger, M.C.G., Snellings, R., Bernal, S.A., 2019. Supplementary cementitious materials: New sources, characterization, and performance insights. Cem. Concr. Res. 122, 257–273. https://doi.org/10.1016/j.cemconres.2019.05.008
Kamali, S., Gérard, B., Moranville, M., 2003. Modelling the leaching kinetics of cement-based materials––influence of materials and environment. Cem. Concr. Compos. 25, 451–458. https://doi.org/10.1016/S0958-9465(02)00085-9
Khan, M., Ali, M., 2016. Use of glass and nylon fibers in concrete for controlling early age micro cracking in bridge decks. Constr. Build. Mater. 125, 800–808. https://doi.org/10.1016/j.conbuildmat.2016.08.111
Kim, E., Huang, K., Saunders, A., McCallum, A., Ceder, G., Olivetti, E., 2017. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. Chem. Mater. 29, 9436–9444. https://doi.org/10.1021/acs.chemmater.7b03500
Kirthika, S.K., Singh, S.K., Chourasia, A., 2020. Alternative fine aggregates in production of sustainable concrete- A review. J. Clean. Prod. 268, 122089. https://doi.org/10.1016/j.jclepro.2020.122089
Krallinger, M., Rabal, O., Lourenço, A., Oyarzabal, J., Valencia, A., 2017. Information Retrieval and Text Mining Technologies for Chemistry. Chem. Rev. 117, 7673–7761. https://doi.org/10.1021/acs.chemrev.6b00851
Kumar, A., Bakshi, B.R., Ramteke, M., Kodamana, H., 2023. Recycle-BERT: Extracting Knowledge about Plastic Waste Recycling by Natural Language Processing. ACS Sustain. Chem. Eng. 11, 12123–12134. https://doi.org/10.1021/acssuschemeng.3c03162
Kurniati, E.O., Pederson, F., Kim, H.-J., 2023. Application of steel slags, ferronickel slags, and copper mining waste as construction materials: A review. Resour. Conserv. Recycl. 198. https://doi.org/10.1016/j.resconrec.2023.107175
Leaman, R., Lu, Z., 2016. TaggerOne: joint named entity recognition and normalization with semi-Markov Models. Bioinformatics 32, 2839–2846. https://doi.org/10.1093/bioinformatics/btw343

Leaman, R., Wei, C.-H., Lu, Z., 2015. tmChem: a high performance approach for chemical named entity recognition and normalization. J. Cheminformatics 7, S3. https://doi.org/10.1186/1758-2946-7-S1-S3
Li, J., Sun, A., Han, J., Li, C., 2023. A Survey on Deep Learning for Named Entity Recognition : Extended Abstract, in: 2023 IEEE 39th International Conference on Data Engineering (ICDE). Presented at the 2023 IEEE 39th International Conference on Data Engineering (ICDE), IEEE, Anaheim, CA, USA, pp. 3817–3818. https://doi.org/10.1109/ICDE55515.2023.00335
Li, X., Li, P., 2021. Rejection Sampling for Weighted Jaccard Similarity Revisited. Proc. AAAI Conf. Artif. Intell. 35, 4197–4205. https://doi.org/10.1609/aaai.v35i5.16543
Liu, H., Li, C., Wu, Q., Lee, Y.J., 2023. Visual Instruction Tuning. https://doi.org/10.48550/arXiv.2304.08485
Liu, S., Wang, L., Li, Q., Song, J., 2020. Hydration properties of Portland cement-copper tailing powder composite binder. Constr. Build. Mater. 251, 118882. https://doi.org/10.1016/j.conbuildmat.2020.118882
Lothenbach, B., 2010. Thermodynamic equilibrium calculations in cementitious systems. Mater. Struct. 43, 1413–1433. https://doi.org/10.1617/s11527-010-9592-x
Luo, Z., Xu, C., Zhao, P., Sun, Q., Geng, X., Hu, W., Tao, C., Ma, J., Lin, Q., Jiang, D., 2023. WizardCoder: Empowering Code Large Language Models with Evol-Instruct.
Lynn, C.J., Dhir, R.K., Ghataora, G.S., West, R.P., 2015. Sewage sludge ash characteristics and potential for use in concrete. Constr. Build. Mater. 98, 767–779. https://doi.org/10.1016/j.conbuildmat.2015.08.122
Malathy, R., Shanmugam, R., Dhamotharan, D., Kamaraj, D., Prabakaran, M., Kim, J., 2023. Lime based concrete and mortar enhanced with pozzolanic materials – State of art. Constr. Build. Mater. 390, 131415. https://doi.org/10.1016/j.conbuildmat.2023.131415
Marieta, C., Guerrero, A., Leon, I., 2021. Municipal solid waste incineration fly ash to produce eco-friendly binders for sustainable building construction. Waste Manag. 120, 114–124. https://doi.org/10.1016/j.wasman.2020.11.034
Masry, A., Shahmohammadi, M., Parvez, M.R., Hoque, E., Joty, S., 2024. ChartInstruct: Instruction Tuning for Chart Comprehension and Reasoning. https://doi.org/10.48550/arXiv.2403.09028
Mehta, P.K., 2001. Reducing the Environmental Impact of Concrete. Concr. Int.
Miller, S.A., Habert, G., Myers, R.J., Harvey, J.T., 2021. Achieving net zero greenhouse gas emissions in the cement industry via value chain mitigation strategies. One Earth 4, 1398–1411. https://doi.org/10.1016/j.oneear.2021.09.011
Mohebi, R., Behfarnia, K., Shojaei, M., 2015. Abrasion resistance of alkali-activated slag concrete designed by Taguchi method. Constr. Build. Mater. 98, 792–798. https://doi.org/10.1016/j.conbuildmat.2015.08.128
Moon, G.D., Oh, S., Jung, S.H., Choi, Y.C., 2017. Effects of the fineness of limestone powder and cement on the hydration and strength development of PLC concrete. Constr. Build. Mater. 135, 129–136. https://doi.org/10.1016/j.conbuildmat.2016.12.189
Najimi, M., Sobhani, J., Ahmadi, B., Shekarchi, M., 2012. An experimental study on durability properties of concrete containing zeolite as a highly reactive natural pozzolan. Constr. Build. Mater. 35, 1023–1033. https://doi.org/10.1016/j.conbuildmat.2012.04.038
Nasar, Z., Jaffry, S.W., Malik, M.K., 2022. Named Entity Recognition and Relation Extraction: State-of-the-Art. ACM Comput. Surv. 54, 1–39. https://doi.org/10.1145/3445965
Nedunuri, S.S.S.A., Sertse, S.G., Muhammad, S., 2020. Microstructural study of Portland cement partially replaced with fly ash, ground granulated blast furnace slag and silica fume as determined by pozzolanic activity. Constr. Build. Mater. 238, 117561. https://doi.org/10.1016/j.conbuildmat.2019.117561
OpenAI, 2023. GPT-4 Technical Report.
Pezeshkpour, P., Hruschka, E., 2023. Large Language Models Sensitivity to The Order of Options in Multiple-Choice Questions.

Plaza, P., Sáez del Bosque, I.F., Frías, M., Sánchez de Rojas, M.I., Medina, C., 2021. Use of recycled coarse and fine aggregates in structural eco-concretes. Physical and mechanical properties and CO2 emissions. Constr. Build. Mater. 285, 122926. https://doi.org/10.1016/j.conbuildmat.2021.122926

Praneeth, S., Guo, R., Wang, T., Dubey, B.K., Sarmah, A.K., 2020. Accelerated carbonation of biochar reinforced cement-fly ash composites: Enhancing and sequestering CO2 in building materials. Constr. Build. Mater. 244, 118363. https://doi.org/10.1016/j.conbuildmat.2020.118363

Qian, Y., Santus, E., Jin, Z., Guo, J., Barzilay, R., 2019. GraphIE: A Graph-Based Framework for Information Extraction.

Rae, J.W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S., Rutherford, E., Hennigan, T., Menick, J., Cassirer, A., Powell, R., Welbl, J., Dathathri, S., Huang, S., Uesato, J., Mellor, J., Higgins, I., Creswell, A., McAleese, N., Wu, A., Elsen, E., Jayakumar, S., Buchatskaya, E., Budden, D., Sutherland, E., Simonyan, K., Paganini, M., Sifre, L., Martens, L., Li, X.L., Kuncoro, A., Nematzadeh, A., Gribovskaya, E., Donato, D., Lazaridou, A., Mensch, A., Lespiau, J.-B., Tsimpoukelli, M., Grigorev, N., Fritz, D., Sottiaux, T., Pajarskas, M., Pohlen, T., Gong, Z., Toyama, D., 2022. Scaling Language Models: Methods, Analysis & Insights from Training Gopher.

Rashad, A.M., 2022. Effect of limestone powder on the properties of alkali-activated materials – A critical overview. Constr. Build. Mater. 356, 129188. https://doi.org/10.1016/j.conbuildmat.2022.129188

Rashad, A.M., Ezzat, M., ElNagar, A.M., El-Nashar, M.H., 2023. Valorization of limestone powder as an additive for fly ash geopolymer cement under the effect of the simulated tidal zone and seawater attack. Constr. Build. Mater. 369, 130616. https://doi.org/10.1016/j.conbuildmat.2023.130616

Sabău, M., Bompa, D.V., Silva, L.F.O., 2021. Comparative carbon emission assessments of recycled and natural aggregate concrete: Environmental influence of cement content. Geosci. Front. 12, 101235. https://doi.org/10.1016/j.gsf.2021.101235

Saillio, M., Baroghel-Bouny, V., Pradelle, S., Bertin, M., Vincent, J., d’Espinose de Lacaillerie, J.-B., 2021. Effect of supplementary cementitious materials on carbonation of cement pastes. Cem. Concr. Res. 142, 106358. https://doi.org/10.1016/j.cemconres.2021.106358

Sarkar, S., Lausen, L., 2023. Testing the Limits of Unified Sequence to Sequence LLM Pretraining on Diverse Table Data Tasks. https://doi.org/10.48550/arXiv.2310.00789

Savelka, J., Agarwal, A., Bogart, C., Sakr, M., 2023. Large Language Models (GPT) Struggle to Answer Multiple-Choice Questions about Code.

Skibsted, J., Snellings, R., 2019. Reactivity of supplementary cementitious materials (SCMs) in cement blends. Cem. Concr. Res. 124, 105799. https://doi.org/10.1016/j.cemconres.2019.105799

Snellings, R., Suraneni, P., Skibsted, J., 2023. Future and emerging supplementary cementitious materials. Cem. Concr. Res. 171. https://doi.org/10.1016/j.cemconres.2023.107199

Sui, Y., Zhou, Mengyu, Zhou, Mingjie, Han, S., Zhang, D., 2024. Table Meets LLM: Can Large Language Models Understand Structured Table Data? A Benchmark and Empirical Study, in: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. Presented at the WSDM ’24: The 17th ACM International Conference on Web Search and Data Mining, ACM, Merida Mexico, pp. 645–654. https://doi.org/10.1145/3616855.3635752

Suraneni, P., Weiss, J., 2017. Examining the pozzolanicity of supplementary cementitious materials using isothermal calorimetry and thermogravimetric analysis. Cem. Concr. Compos. 83, 273–278. https://doi.org/10.1016/j.cemconcomp.2017.07.009

Talmor, A., Herzig, J., Lourie, N., Berant, J., 2019. CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge.

Tang, Q., Liu, Y., Gu, F., Zhou, T., 2016. Solidification/Stabilization of Fly Ash from a Municipal Solid Waste Incineration Facility Using Portland Cement. Adv. Mater. Sci. Eng. 2016, 1–10. https://doi.org/10.1155/2016/7101243

Tanyildizi, H., Coskun, A., 2008. The effect of high temperature on compressive strength and splitting tensile strength of structural lightweight concrete containing fly ash. Constr. Build. Mater. 22, 2269–2275. https://doi.org/10.1016/j.conbuildmat.2007.07.033

Thomas, J.J., Biernacki, J.J., Bullard, J.W., Bishnoi, S., Dolado, J.S., Scherer, G.W., Luttge, A., 2011. Modeling and simulation of cement hydration kinetics and microstructure development. Cem. Concr. Res. 41, 1257–1278. https://doi.org/10.1016/j.cemconres.2010.10.004

Thongsanitgarn, P., Wongkeo, W., Chaipanich, A., Poon, C.S., 2014. Heat of hydration of Portland high-calcium fly ash cement incorporating limestone powder: Effect of limestone particle size. Constr. Build. Mater. 66, 410–417. https://doi.org/10.1016/j.conbuildmat.2014.05.060

Tironi, A., Trezza, M.A., Scian, A.N., Irassar, E.F., 2013. Assessment of pozzolanic activity of different calcined clays. Cem. Concr. Compos. 37, 319–327. https://doi.org/10.1016/j.cemconcomp.2013.01.002

Ubbriaco, P., 1996. Hydration behaviour of a lime-pozzolan cement containing MSW incinerator fly ash. J. Therm. Anal. 47, 7–16. https://doi.org/10.1007/BF01982681

Upadhyay, H., Mungule, M., K. R. Iyer, K., 2022. Issues and challenges for development of geopolymer concrete. Int. Conf. Adv. Constr. Mater. Struct. 65, 1567–1574. https://doi.org/10.1016/j.matpr.2022.04.520

US Department of Transportation Federal Highway Administration, 2025. Reactivity of Supplementary Cementitious Materials.

Uvegi, H., Jensen, Z., Hoang, T.N., Traynor, B., Aytaş, T., Goodwin, R.T., Olivetti, E.A., 2021. Literature mining for alternative cementitious precursors and dissolution rate modeling of glassy phases. J. Am. Ceram. Soc. 104, 3042–3057. https://doi.org/10.1111/jace.17631

Van Den Heede, P., Gruyaert, E., De Belie, N., 2010. Transport properties of high-volume fly ash concrete: Capillary water sorption, water sorption under vacuum and gas permeability. Cem. Concr. Compos. 32, 749–756. https://doi.org/10.1016/j.cemconcomp.2010.08.006

van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Van Deventer, J.S.J., Provis, J.L., Duxson, P., 2012. Technical and commercial progress in the adoption of geopolymer cement. Sustain. Resour. Conserv. Recycl. 29, 89–104. https://doi.org/10.1016/j.mineng.2011.09.009

Venugopal, V., Olivetti, E., 2024. MatKG: An autonomously generated knowledge graph in Material Science. Sci. Data 11, 217. https://doi.org/10.1038/s41597-024-03039-z

Walker, N., Lee, S., Dagdelen, J., Cruse, K., Gleason, S., Dunn, A., Ceder, G., Alivisatos, A.P., Persson, K.A., Jain, A., 2023. Extracting structured seed-mediated gold nanorod growth procedures from scientific text with LLMs. Digit. Discov. 2, 1768–1782. https://doi.org/10.1039/D3DD00019B

Walker, R., Pavía, S., 2011. Physical properties and reactivity of pozzolans, and their influence on the properties of lime–pozzolan pastes. Mater. Struct. 44, 1139–1150. https://doi.org/10.1617/s11527-010-9689-2

Wang, B., Gong, Y., Zhou, J., 2023. Text mining and association rules-based analysis of 245 cement production accidents in a cement manufacturing plant. https://doi.org/10.21203/rs.3.rs-2976063/v1

Wang, D., Shi, C., Farzadnia, N., Jia, H., Zeng, R., Wu, Y., Lao, L., 2019. A quantitative study on physical and chemical effects of limestone powder on properties of cement pastes. Constr. Build. Mater. 204, 58–69. https://doi.org/10.1016/j.conbuildmat.2019.01.154

Wang, D., Shi, C., Farzadnia, N., Shi, Z., Jia, H., Ou, Z., 2018. A review on use of limestone powder in cement-based materials: Mechanism, hydration and microstructures. Constr. Build. Mater. 181, 659–672. https://doi.org/10.1016/j.conbuildmat.2018.06.075

Wang, S., Miller, A., Llamazos, E., Fonseca, F., Baxter, L., 2008. Biomass fly ash in concrete: Mixture proportioning and mechanical properties. Fuel 87, 365–371. https://doi.org/10.1016/j.fuel.2007.05.026

Wang, W., Cai, F., Jiao, P., Pan, L., 2016. A perturbation-based framework for link prediction via non-negative matrix factorization. Sci. Rep. 6, 38938. https://doi.org/10.1038/srep38938

Wang, W., Zhang, X., Mo, Z., Chouw, N., Li, Z., Xu, Z.-D., 2020. A comparative study of impact behaviour between natural flax and glass FRP confined concrete composites. Constr. Build. Mater. 241, 117997. https://doi.org/10.1016/j.conbuildmat.2020.117997

Weston, L., Tshitoyan, V., Dagdelen, J., Kononova, O., Trewartha, A., Persson, K.A., Ceder, G., Jain, A., 2019. Named Entity Recognition and Normalization Applied to Large-Scale Information Extraction from the Materials Science Literature. J. Chem. Inf. Model. 59, 3692–3702. https://doi.org/10.1021/acs.jcim.9b00470

Wilcox, E., Qian, P., Futrell, R., Ballesteros, M., Levy, R., 2019. Structural Supervision Improves Learning of Non-Local Grammatical Dependencies, in: Proceedings of the 2019 Conference of the North. Presented at the Proceedings of the 2019 Conference of the North, Association for Computational Linguistics, Minneapolis, Minnesota, pp. 3302–3312. https://doi.org/10.18653/v1/N19-1334

Wu, Y., Lu, B., Bai, T., Wang, H., Du, F., Zhang, Y., Cai, L., Jiang, C., Wang, W., 2019. Geopolymer, green alkali activated cementitious material: Synthesis, applications and challenges. Constr. Build. Mater. 224, 930–949. https://doi.org/10.1016/j.conbuildmat.2019.07.112

Xie, T., Wan, Y., Huang, W., Zhou, Y., Liu, Y., Linghu, Q., Wang, S., Kit, C., Grazian, C., Zhang, W., Hoex, B., 2023. Large Language Models as Master Key: Unlocking the Secrets of Materials Science with GPT.

Yang, J., Huang, J., Su, Y., He, X., Tan, H., Yang, W., Strnadel, B., 2019. Eco-friendly treatment of low-calcium coal fly ash for high pozzolanic reactivity: A step towards waste utilization in sustainable building material. J. Clean. Prod. 238, 117962. https://doi.org/10.1016/j.jclepro.2019.117962

Zeng, W., Zhao, Y., Zheng, H., Poon, C.S., 2020. Improvement in corrosion resistance of recycled aggregate concrete by nano silica suspension modification on recycled aggregates. Cem. Concr. Compos. 106, 103476. https://doi.org/10.1016/j.cemconcomp.2019.103476

Zhang, D., Zhao, J., Wang, D., Wang, Y., Ma, X., 2020. Influence of pozzolanic materials on the properties of natural hydraulic lime based mortars. Constr. Build. Mater. 244, 118360. https://doi.org/10.1016/j.conbuildmat.2020.118360

Zhang, Y., Liu, W., Liu, M., 2024. Setting time and mechanical properties of chemical admixtures modified FA/GGBS-based engineered geopolymer composites. Constr. Build. Mater. 431, 136473. https://doi.org/10.1016/j.conbuildmat.2024.136473

Zhang, Yue, Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Yu, Chen, Y., Wang, L., Luu, A.T., Bi, W., Shi, F., Shi, S., 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models.

Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., Smola, A., 2024. Multimodal Chain-of-Thought Reasoning in Language Models. https://doi.org/10.48550/arXiv.2302.00923

Zheng, C., Zhou, H., Meng, F., Zhou, J., Huang, M., 2024. LARGE LANGUAGE MODELS ARE NOT ROBUST MULTIPLE CHOICE SELECTORS.

Zhu, J.-J., Dressel, W., Pacion, K., Ren, Z.J., 2021. *ES&T* in the 21st Century: A Data-Driven Analysis of Research Topics, Interconnections, And Trends in the Past 20 Years. Environ. Sci. Technol. 55, 3453–3464. https://doi.org/10.1021/acs.est.0c07551

Zhu, J.-J., Ren, Z.J., 2023. The evolution of research in resources, conservation & recycling revealed by Word2vec-enhanced data mining. Resour. Conserv. Recycl. 190, 106876. https://doi.org/10.1016/j.resconrec.2023.106876