Robotics and Computer–Integrated Manufacturing 93 (2025) 102900

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Full length article

Knowledge extraction for additive manufacturing process via named entity recognition with LLMs

Xuan Liu [a], John Ahmet Erkoyuncu [b], Jerry Ying Hsi Fuh [a], Wen Feng Lu [a], Bingbing Li [a,c,*]

[a] Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, Singapore
[b] School of Aerospace, Transport systems and Manufacturing ad materials (SATM), Cranfield University, Cranfield, United Kingdom
[c] Department of Manufacturing Systems Engineering and Management, California State University, Northridge, USA

ARTICLE INFO

Keywords:
Domain knowledge management
Named entity recognition
Retrieval augmented generation
Large Language Models
Additive manufacturing
Fused deposition modeling

ABSTRACT

This paper proposes a novel NER framework, leveraging the advanced capabilities of Large Language Models (LLMs), to address the limitations of manually defined taxonomy. Our framework integrates the expert knowledge internalized in both academic materials and LLMs through retrieval-augmented generation (RAG) to automatically customize taxonomies for specific manufacturing processes and adopts two distinct strategies of using LLMs — In-Context Learning (ICL) and fine-tuning to complete manufacturing NER tasks with minimal training data. We demonstrate the framework efficiency through its superior ability to define precise taxonomies, identify and classify process-level entities related to the most popular additive manufacturing process fused deposition modeling (FDM) as case study, achieving a high F1 score of 0.9192.

1. Introduction

In the context of Industry 4.0, knowledge management (KM) systems, which serve the entire life cycle of group and organizational knowledge, from creation, storage, to transfer, and application, are recognized for their substantial contributions to various corporate facets, encompassing digital transformation, innovation ecosystems, decision-making, and production optimization [1]. With the ability to extract knowledge (e.g. named entities) from unstructured texts based on a predefined taxonomy, NER is one of the fundamental tasks for constructing a KM system [2]. NER models based on long short-term memory (LSTM) [3] or fine-tuned language models [4] have been used in extracting knowledge from manufacturing texts to mitigate the burden of manual search and parsing. Unlike general NER models focusing on extracting common terms, manufacturing NER is extended to more technical entities such as materials and process parameters. Despite these advancements, most works still follow the paradigm of general NER without considering the correlation between entities and specific manufacturing domains, resulting in a crucial drawback: these approaches can hardly distinguish entities related to different manufacturing processes or techniques. For example, within the knowledge domain of additive manufacturing, the differentiation of materials suitable for fused deposition modeling (FDM) and materials suitable for powder bed fusion (PBF) is quite difficult for these models.

Recognizing entities specific to a manufacturing process enables professionals to integrate process-level domain knowledge to address practical manufacturing issues, facilitates a swift onboarding process for beginners, and guides the decision making. The domain-specific entity list can serve as a checklist for manufacturing process design and problem troubleshooting. However, it is challenging to autonomously identify both the category of an entity and its related manufacturing process simultaneously. This problem not only roots in the model design, but also exacerbated by the lack of annotated process-level manufacturing dataset and the requirement of comprehensive process knowledge for taxonomy definition. Taxonomy is the foundation of knowledge extraction tasks including NER [5], which specifies the categories of entities that should be identified. Designing the taxonomy for a particular manufacturing process requires a comprehensive understanding of the process, placing higher demands on the experience and expert knowledge than classifying entities. In consequence, the conventional taxonomy design methods highly rely on domain experts, posing a significant barrier for novices, requiring considerable time, and lacking scalability [6]. Additionally, the manufacturing text is highly complex and unstructured, containing many entities of the same class but belonging to different manufacturing domains. Even the same entities can be related to different processes. When the training data is scarce, it is extremely difficult for traditional models to accurately learn the probability distribution for classifying the entities and finding the entities relevant to a specific process.

The emergence of LLMs, such as Generative Pre-trained Transformer 4 (GPT-4), marks a significant milestone in artificial intelli-

* Corresponding author at: Department of Manufacturing Systems Engineering and Management, California State University, Northridge, USA.
E-mail address: bingbing.li@csun.edu (B. Li).

https://doi.org/10.1016/j.rcim.2024.102900
Received 13 June 2024; Received in revised form 31 October 2024; Accepted 6 November 2024
Available online 21 November 2024

gence, bringing new opportunities to fill these gaps. LLMs internalize plenty of manufacturing domain knowledge, enabling their application in process-level manufacturing tasks. The retrieval-augmented generation (RAG) technique can further enhance their ability, mitigate hallucination, and support them in more professional tasks which need comprehensive domain knowledge [7]. LLMs showcase high generalization abilities through In-Context Learning (ICL) and fine-tuning, which promote their proficiency in not only semantic understanding and generation, but also various domain-specific tasks [8,9], with few-shot demonstrations or limited domain-specific training data. Meanwhile, the development of chain-of-thought (CoT) prompting enhance their performances on complex tasks by promoting reasoning [10]. These advantages enable LLMs to understand and extract information from unstructured natural language texts [11]. For example, LLM-based ICL has been used in general domain NER from individual sentences [12]. In addition, recent LLMs support handling lengthy context, facilitating with the cross-sentence information extraction ability.

With the identified research gaps and the advantages of these emerging techniques, this paper introduces a novel domain-specific manufacturing NER framework by leveraging the ability of LLMs to extract the process-level knowledge from natural language texts, assisting the optimization of specified manufacturing process. We verify the effectiveness of our framework on an NER task within the FDM domain. Firstly, we build a corpus for FDM process from unstructured FDM-related texts. Then the FDM taxonomy is customized by GPT-4 through RAG and candidate integration based on the corpus, containing 18 classes and corresponding subclasses. Among them, the most frequently mentioned entity classes of FDM are selected for recognition, including printing equipment, material types, printing parameters, software concepts, quality metrics, common issues, post-processing techniques, applications, standards and regulations, and 3D design and modeling. Meanwhile, the input data are obtained by extracting texts from the corpus and generating remaining texts by LLMs based on the taxonomy. We also optimize the labeling method introduced by Yan et al. [13] to label the dataset and help LLMs execute the NER task. To recognize and classify entities semantically related to FDM with minimal training data, both the ICL approach and the fine tuning approach of LLMs are adopted. We not only compare the performances of these two methods and the baseline model BERT-MRC [14], but also explore different strategies to optimize the results and achieve the highest F1 value of 0.9192. At last, we also provide an example of how to effectively absorb knowledge from the extracted entities. To the best of our knowledge, this is the first framework that is able to recognize process-level manufacturing entities without manually defined taxonomy and large amount of training data.

The main contributions of this work can be summarized as follows: (1) Leverage both the internalized knowledge of LLMs and expert information from academic materials through RAG to automatically customize the taxonomy of a manufacturing process, which mitigates the hallucination problem, reduces time and expert knowledge required by NER task, and builds a knowledge structure to help beginners understand the manufacturing process; (2) Develop two LLM-based NER approaches based on ICL and fine-tuning respectively, which both can classify entities into predefined taxonomy and identify the entities related to the particular manufacturing process simultaneously, with scarce training data. This framework showcase remarkable performance in process-level NER from unstructured text data.[1]

[1] Data are available at https://github.com/classic-byte/Knowledge-Extraction-for-AM-via-NER-with-LLMs

2. Related work

KM encompasses the activities of discovering, collecting, and exploiting the shared knowledge within an organization to boost its capacity for innovation, agility, and competitive advantage [15,16]. Deploying a KM system mitigates the poor generalization problem of data-driven models in handling a new domain without sufficient data [17]. 60 years ago, the development of KM started from the systematization and storage of explicit knowledge [18]. Nowadays, companies can create knowledge in production and absorb knowledge from suppliers and customers, which is essential for production optimization, customized design, and employee upskilling. The successful management of these data can accelerate the digital and intelligent transformation of companies in industry 4.0 [1,19].

As an important technique of KM, NER aims to identify mentions of different semantic classes in texts, such as location, organization, person, etc. [20]. In order to recognize domain-specific entities, NER models for specified domains such as material science [21], chemical reactions [22], biomedical [23], and manufacturing [3] have been developed. For manufacturing domain, researchers firstly used a bidirectional long-short term memory, plus a conditional random field (BiLSTM + CRF) to extract manufacturing entities from abstracts [3]. Then, Attention-based language models like BERT became popular in manufacturing NER, which improved the performances by fine-tuning different BERT-based models [4], enabled the recognition of Chinese fine-grained mechanical entities [24], and promoted knowledge extraction for degradation analysis [25]. Though these models work well in some widespread domains, they can hardly determine the semantic relation between an entity and a specific manufacturing process from the context. Their dependency on a vast amount of labeled data for training and comprehensive domain knowledge for taxonomy construction further limit their application in recognizing process-level manufacturing entities.

The recent outbreak of LLMs shows remarkable few or zero shots ability [26,27], becoming a promising solution to solve these problems. Most applications are based on ICL, in which a model learns from a few examples provided in the prompt or some data input during inference, without additional training or updates to its parameters. ICL demonstrations of high quality can prominently improve the performance of a language model in unfamiliar tasks [28]. LLMs are proved to be effective in general domain NER from individual sentences by marking desired entities with special symbols in ICL demonstrations, even when the training data is scarce [12,13]. Chen et al. models an LLM as a meta-function to recognize entities from Wikidata [29]. Moreover, Cai et al. achieve multimodal NER for a text-image pair with LLMs by ICL [30]. In addition, LLMs are applied in material NER [31], biomedical NER [32], and clinical NER [33] with a few ICL demonstrations. In contrast, fine-tuning LLMs for NER is less studied. Li et al. fine-tune Llama with discriminant labels, achieving in state-of-the-art performance in general-domain NER tasks [34], while most fine-tuning methods are used in the comparisons with ICL [31,32]. Furthermore, by creating embeddings from domain-specific professional texts and storing them in a vector database for RAG, the hallucination of LLMs can be mitigated and their performances in domain-specific tasks requiring comprehensive knowledge can be improved [7]. The CoT prompt style which has improved the performance of other LLM-based knowledge extraction techniques, is also promising in NER [35]. Nevertheless, LLMs with ICL or fine-tuning are seldom adopted in the manufacturing NER, though they are proven to be effective tools for specific manufacturing processes [36–38].

To summarize, domain-specific NER, including manufacturing NER, showcasing significant progress in categorizing domain-specific entities into taxonomies, but they are not suitable for distinguishing manufacturing entities that are semantically related to different processes based on the context. The reason includes unsuitable model design, the requirement of comprehensive process knowledge, and the lack

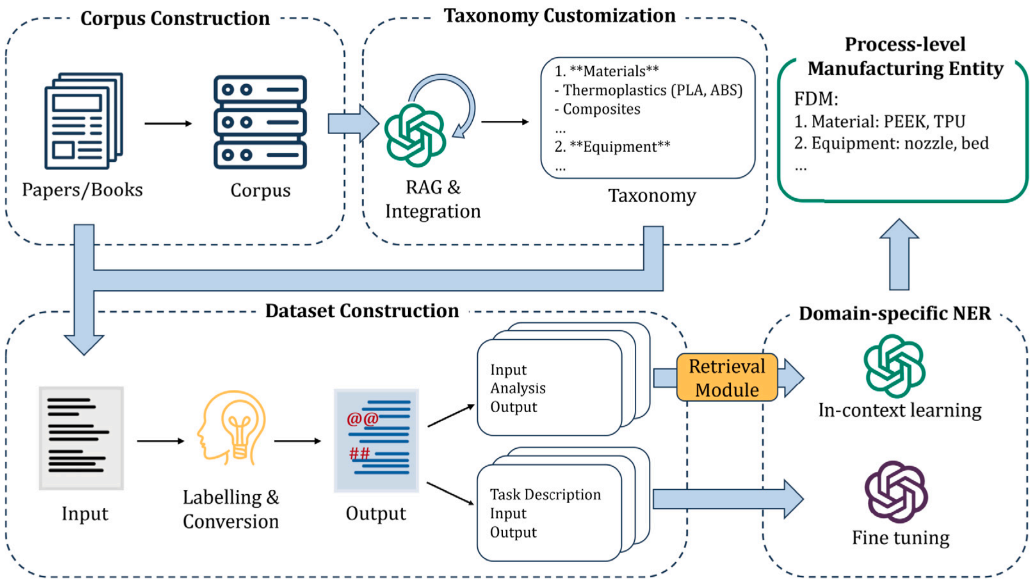

Fig. 1. The LLM-based domain-specific NER framework.

of process-level manufacturing dataset. Consequently, the extracted manufacturing entities cannot be used to guide specific production processes. LLMs have optimized the NER performance in the data-limited scenario already and shown their potentials in the domain-specific manufacturing tasks, owing to their internalized process knowledge and remarkable semantic understanding and generation abilities enhanced by suitable techniques. Therefore, developing an LLM-based NER approach to identify process-level entities with minimal training data and process knowledge is necessary and promising.

3. Methodology

3.1. Framework overview

The proposed framework is illustrated in Fig. 1, consisting of four elements: process corpus construction, process-level taxonomy customization, LLM-assisted dataset construction, and domain-specific manufacturing NER. We introduce two approaches for domain-specific NER in this paper: ICL-based method and fine tuning-based method. The framework is demonstrated with the use case of NER in the domain of FDM from a collection of published papers and books.

3.2. Corpus construction

Containing plenty of professional materials about various manufacturing processes, the academic databases such as SCOPUS, IEEE Xplore and ScienceDirect were used as the sources of corpus data. We searched and downloaded hundreds of papers and book chapters with “FDM”, “fused deposition modeling”, “material extrusion”, “FFF”, or “fused filament fabrication” in their metadata from these databases. Only the main body of each material which contains the most complete information was focused in our research, while the other parts such as references, tables, and figure captions are deleted. The citation numbers in the main body were also removed. The main body of each paper was split into paragraphs to preserve cross-sentence knowledge effectively while avoiding the poor performance of LLM in processing overlong context. A filter was applied to eliminate paragraphs without mention of “FDM” or the above synonyms of it. Finally, the filtered paragraphs formed a corpus for the FDM process. Since multiple manufacturing techniques could be mentioned in the same paragraph of professional materials, the corpus contained not only entities related to FDM, but also entities of desired categories related to other processes such as SLM, SLS, and general additive manufacturing, which had to be differentiated in the domain-specific NER.

3.3. RAG-based process-level taxonomy customization

The taxonomy of entities should be determined at first to guide the recognition, which is manually designed by domain experts in typical domain-specific NER approaches. In contrast, we used the internalized domain knowledge of LLMs and RAG technique to automatically generate the taxonomy of specified manufacturing processes, reducing cost and the demand of comprehensive domain knowledge. The two steps included are shown in Fig. 2.

Retrieval-augmented candidate generation — We firstly created embeddings for the corpus, stored them into a vector database, and connected the database with the LLM to complement its internalized domain knowledge. Then, the LLM was used to retrieve professional information from the corpus and integrate it with internalized knowledge in generating response to every prompts. Since the corpus data were extracted from academic sources, the hallucination problem of the LLM in the specific domain was significantly mitigated, while the integrity and depth of responses were improved, enabling the automatic generation of process-level taxonomies with proper prompts and reducing the manual involvement.

The taxonomy candidate generation prompt is a clear task description with zero shot, indicating the domain X, the purpose of the generated taxonomy P, and the maximum number of first-level categories N. N can avoid generating taxonomies that lacks hierarchy and generality, in which any subtle differences may cause entities to be categorized differently. The last sentence of the prompt encourages the LLM to produce taxonomies with varying numbers of first-level categories, enhancing the diversity of candidates in terms of scope and structure. In this instance, the domain was FDM process, the taxonomy was used for NER, and the maximum number of first-level classes was set as 20 to balance the generality and accuracy of the taxonomy. Since the responses content of the LLM varies in different attempts, this step was repeated n times to generate multiple candidate taxonomies.

GPT integration — This step enhances the completeness and stability of the process-level taxonomy by utilizing the LLM to combine

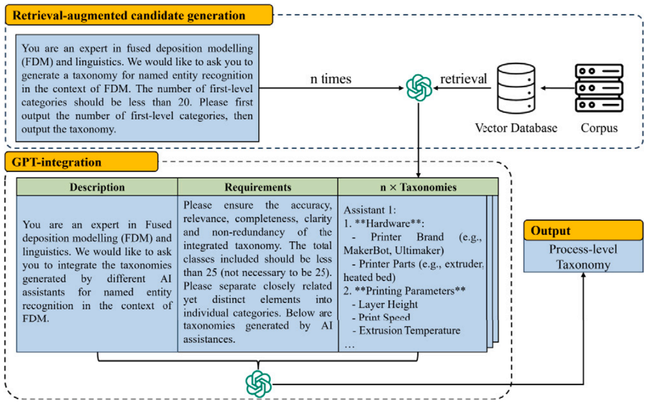

Fig. 2. The overall taxonomy customization process for the FDM domain.

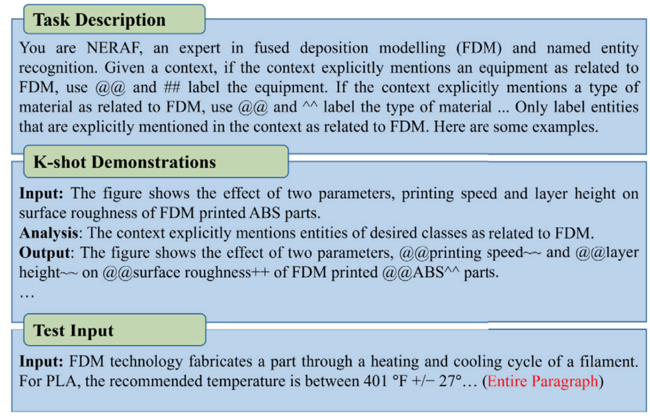

Fig. 3. An example of the CoT style prompt used in ICL-based NER.

the n generated taxonomies. The prompt is divided into three sections: (1) a general description of the integration task, (2) the requirements for the integrated taxonomy, and (3) the candidate taxonomies. The integrated taxonomy is reorganized from the candidates, including categories, subcategories, and examples of FDM entities. In multiple tests, the content of the integrated taxonomies was stable, with only slight fluctuations in the structure.

3.4. Dataset construction

In this framework, the samples of the fine-tuning approach and the ICL approach share the same input and output data. The input should be a short paragraph or a sentence, while the output should be a labeled sequence which at least contains the information including entity type, start position and end position of the entity. Task description is another component of each fine-tuning sample, while in ICL, it is the beginning of a whole prompt instead of a part of sample.

The texts from the corpus were taken as the major source of inputs. For training set, we split these texts into smaller pieces ranging from one to three sentences in length to improve the training efficiency and the demonstration quality, since for each test data, the cosine similarity between it and one of these pieces is usually higher than that between it and the entire paragraph. The pieces without mention of "FDM" or its synonyms, and the pieces with disorganized structure or grammatical errors were abandoned to further increase the quality of input data. Since the corpus we built for FDM process is small, we also used LLMs to generate some input texts based on the taxonomy to enhance the diversity of expression, which were only used as the input of training data. LLMs were asked to generate "a short paragraph" mentioning subcategories or examples as listed in the taxonomy, in the context of "FDM", "conform to the fact", and less than three sentences. The hallucination problem of LLMs is acceptable in generating these texts, since we hope the LLMs learn how to classify entities and determine the relation between process and entities by understanding the semantic information contained in the context through ICL or fine-tuning rather than just being accurately telling the classification of a specific noun or phrase and its relationship with FDM. Besides, these texts were not added into the test set, so the inaccurate information contained has no influence on the NER results. We need to manually verify that the input texts cover all types of entities need to be recognized and can be divided into 3 types: (1) Positive input — text containing entities belonging to varying desired categories and are relevant to FDM, (2) Negative input — texts that mention FDM but do not contain entities of needed categories, and (3) Mixed input — texts that mention FDM process and contain entities of needed categories but some of these entities are not explicitly relevant to the process. In comparison, obtaining the input for test set is easier, as it directly use the entire paragraphs from the remaining corpus.

For output, 2 graduate students with basic knowledge of FDM labeled the input data individually. They found the entities corresponding to different predefined classes from the text data and determined from the context whether an entity is semantically related to FDM. The well-known BIO style annotation of traditional NER models contains the label of each word from the input sentence. Since LLMs are generative models which are not specifically trained for NER, understanding and following such output style is a great burden to ICL. Inspired by Yan [13], we designed a novel output style in which the input sentence is repeated while the start and end position of each entity are marked with special symbols, to simplify the annotation task. The symbols varied with different classes of entities for differentiation. After labeling, we reviewed the annotation results together to deal with the disagreement. The key points of revision process include: (1) Check the classification accuracy and the relation between entities and FDM in the context. (2) Ensure labeling the whole proper nouns and compound nouns instead of a part of them. (3) Decide whether label a qualifier or not based on its impact to the meaning of the whole phrase. (4) Label the abbreviation and its full form as one entity if they occur together. (5) Make sure there are no grammatical or spelling errors in the texts. This form of output data is suitable for both ICL approach and fine-tuning approach.

The task description of each fine-tuning sample indicates the role of LLMs, describes the general task, and defines the tokens used to label different entities, while the task description of each ICL-based NER prompt adds "Here are some examples:" at last to introduce the examples.

Additionally, analysis is an optional part for the ICL training sample, which form a CoT from the input to the output, improving the differentiation ability of entities not related to the specified domain. The detailed architecture of data used in the ICL approach and the fine-tuning approach will be introduced in 3.5 and 3.6 respectively.

3.5. ICL-based domain-specific NER

In the ICL-based approach, domain-specific entities of multiple categories can be identified based on the examples in the prompt, which is extracted from the training set. As shown in Fig. 3, the prompt for recognizing FDM process-level entities involves three parts:

Task Description — This part provides an overview of the domain-specific NER task. In task description, we firstly named the LLM as "NERAF", which is the abbreviation of NER Assistant of FDM, and defined its role as an expert in both FDM and NER. Then we introduced the unique symbol pair assigned for each FDM-related entity class in output. For example, we used "@@" and "~~" to label the printing parameters of FDM and use "@@" and "++" to label the quality metrics

Table 1
The hyperparameters of LLMs.

Model hyperparameter	Value
max_token	4096
n	1
temperature	0.0
top_p	1

Table 2
The hyperparameters of fine-tuning.

Fine-tuning hyperparameter	Value
model	gpt-3.5-turbo-0125
num_epochs	4
batch_size	1
learning_rate_multiplier	2

of FDM. Before introducing examples, the LLM is requested not to label entities without an explicitly semantic relationship with FDM.

K-shot Demonstrations — This part aims to retrieve k samples from training set as demonstrations to regulate the format of output and provide the LLM with semantic clues and references for the NER task. Based on whether the samples in the training set contain analyses or not, the demonstrations can be divided into CoT style demonstrations and Simple style demonstrations. Since the only difference between them is the analysis, we only introduce the CoT style here:

- Input — 1–3 sentences of text from the FDM corpus or generated by LLMs.
- Output — The output copies the input text and the desired entities related to FDM are surrounded by the tokens defined in the task description.
- Analysis — During experiments, LLMs sometimes mislabeled entities of the needed classes that are unrelated to FDM. Thus, we added analysis to the training samples to guide the LLMs generation and decrease the task complexity, including one for positive inputs: "The context explicitly mentions entities of desired classes as related to FDM", one for negative inputs: "The context does not mention entities of desired classes, so no label will be added", and one for mixed inputs: "The context mentions FDM and entities of desired classes, but some of them are not explicitly related to FDM, so I won't label them". The analysis were put between each input and output to form a CoT prompt style, decomposing the task into two steps: (1) analyze the classifications of entities and identify their relationships with FDM. (2) output the entire context and label the entities with defined symbol pairs.

For retrieving k-shot demonstrations, we propose two retrieval strategies to obtain training samples for each test input. (1) Random retrieval: We randomly selected k demonstrations from the training set. (2) Context-level KNN retrieval: Providing demonstrations closer to the input in the embedding space can improve the ICL performance. Thus, we used a popular pretrained text similarity model SimCSE to obtain context-level representations for each training sample and the input paragraph. Then we calculated the cosine similarity to find k demonstrations most close to the input.

Test Input — At the end of an ICL prompt, an entire input paragraph from the test set is appended to indicate the text needs to be labeled.

Feeding the LLM with the prompt and setting the model hyperparameters as shown in Table 1, it is expected to generate the "Analysis" and "Output" for the test input based on the demonstrations provided. The LLMs sometimes generated nonexistent line breaks within the output paragraph compared to the original input, which were eliminated to avoid the wrong calculation of entity positions. Finally, entities surrounded by the defined symbols were extracted and recorded in the corresponding classes with their start position and end position to achieve process-level NER.

3.6. Fine-tuning based domain-specific NER

Compared to ICL, fine-tuning enables the development of a customized NER model for a manufacturing domain like FDM. The fine-tuned models can recognize both the category of an entity and its related manufacturing process from the test input without any demonstration, reducing the time and token consumption when the number of paragraphs to be processed is large.

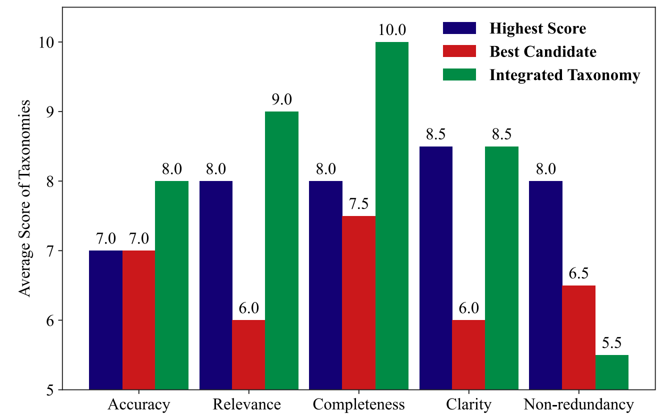

Fig. 4. The comparison between candidate taxonomy and integrated taxonomy.

The format of training data followed the specification of LLM used. For example, each training data for fine-tuning an OpenAI's conversational chat completion model should be a message containing a system prompt, a user prompt, and an assistant prompt, sharing the same contents with the task description, input, and output of ICL approach, respectively. Only the last sentence of ICL task description — "Here are some examples" was removed in the system prompt of fine-tuning. These data were uploaded to the API for training preparation.

In this case, we fine-tuned GPT-3.5-turbo as an example since more advanced models were not accessible for fine-tuning. The hyperparameters of fine-tuning mainly include number of epochs, batch size, and learning rate multiplier, as shown in Table 2. To avoid overfitting problem, we used smaller dataset and fewer training epochs in the beginning stage and increased these numbers gradually while observing the training performance. The deployment method of the fine-tuning model was the same with the original LLM, sharing the model hyperparameters illustrated in Table 1. The prompt containing task description and the test input was fed into the fine-tuned GPT-3.5 to classify entities related to FDM into the taxonomy without any demonstration.

4. Results and evaluation

4.1. Taxonomy customization

We used GPT-4 to generate 10 candidate FDM taxonomies through RAG to mitigate hallucination, and integrated them to improve the stability. The temperature of GPT-4 was kept at 0.0 during the entire process. As shown in Table 3, the integrated taxonomy contains 18 categories with subcategories and examples of each.

To evaluate the effectiveness of integration, we asked 2 FDM experts to rank the candidates and the integrated taxonomy in accuracy, relevance, completeness, clarity, and non-redundancy. For comparison, the rankings were converted into scores, with the first place corresponding to 10 and the eleventh corresponding to 0, and the scores of each metric given by two experts were averaged. The evaluation results are illustrated in Fig. 4, including the highest score achieved in each metric among all the candidate taxonomies, the performance of the best candidate which obtained the highest total points in 5 metrics, and the scores achieved by the integrated taxonomy. The ranking

Table 3
The integrated taxonomy of FDM.

No	Categories	Subcategories and examples
1	Printing Equipment	Printer: brands, models and types; Components: extruder, print bed, nozzle, cooling system
2	Material Types	Thermoplastic: PLA, ABS, PETG, TPU; Composites
3	Printing Parameters	Layer Height and resolution, print speed, extrusion and bed temperature, infill pattern
4	Software Concepts	Software: CAD, slicing, G-code generation and simulation; File formats: STL, OBJ
5	Product Metrics	Surface finish; Dimensional accuracy; Mechanical properties: strength
6	Common Issues	Warping, stringing, layer misalignment
7	Post-processing Techniques	Sanding, painting, gluing, assembly, chemical treatments
8	Applications	Prototyping; Education; Artistic creation; Industry: automotive, aerospace, medical
9	Standards and Regulations	Safety and quality standards: ISO, ASTM; Testing standards; Environmental regulations
10	3D Design and Modeling	Geometry terms; Design principles
11	Maintenance and Troubleshooting	Calibration, bed Leveling, nozzle cleaning
12	Technical Specifications	Printer Dimensions; Filament specifications
13	Environment and Sustainability	Recyclability, energy consumption, biodegradability, emission
14	Safety and Health	Health hazards: fumes, UV exposure; Safety equipment: fume extractors
15	Community and Education	Forums, communities, user groups, educational resources, tutorials
16	Costs and Pricing	Costs: printer, material; Expenses: operating, maintenance
17	Market and Industry	Manufacturers; Distributors; Consumers
18	Innovation and Trends	Emerging technologies; New materials; Market Trends; Research Topics

results showcase the integrated FDM taxonomy has better performance than all the candidates in each metric except non-redundancy, and its overall performance far surpasses the best candidate taxonomy. To demonstrate the effectiveness of our domain-specific NER approaches, we selected the first 10 classes of integrated taxonomy for recognition, from FDM equipment to 3D design and modeling, since they were more widely discussed in the FDM academic materials while the other classes were not frequently mentioned.

4.2. NER results

We labeled 300 input texts for training set, with 235 texts from the corpus and 65 generated by GPT-4. Meanwhile, the task description was added in each training sample of fine-tuning, while the analysis was inserted between corresponding input and output of each ICL sample. For the test input, 100 entire paragraphs were randomly selected from the rest part of the corpus.

The ICL approach was deployed with Llama-2, GPT-3.5, GPT-4, and Claude-3-opus, respectively. The hyperparameters of Llama-2 were different from other models, with temperature of 0.01, top_k of 0. From the entire test set, Claude-3-opus + ICL found 1264 FDM-related entities of predefined 10 classes. Fig. 5 illustrates an output of Claude-3-opus + ICL (context-level KNN, k=20, simple style), in which only the entities related to FDM are labeled with symbols corresponding to their classes, while the entities related to other processes such as SLA and SLM are not labeled, though these entities belong to the classes defined in the taxonomy. The performances of GPT-4 + ICL and the fine-tuned model in this example text were the same, demonstrating the effectiveness of our framework in process-level NER. Recognition of these process-specific entities is key knowledge for getting started with the FDM technique.

Stereolithography (SLA), Selective Laser Melting (SLM), and Fused Deposition Modeling (FDM) are three distinct 3D printing technologies. SLA uses UV light to cure liquid resin into solid objects, ideal for detailed prototypes and models, utilizing photopolymeric resins. Key parameters include layer thickness and curing time, influencing resolution and surface finish. SLM, meanwhile, employs a high-power laser to fuse metallic powders, commonly used in aerospace, automotive, and medical industries for its ability to create strong, complex metal parts from materials like steel, titanium, and aluminum. Its key parameters include laser power, scanning speed, and layer thickness, impacting part density and mechanical properties. FDM, the most widely accessible, builds objects by extruding @@thermoplastic materials^^ like @@ABS^^ and @@PLA^^ layer by layer, popular in @@prototyping&&, @@educational models&&, and @@hobbyist projects&&. @@Nozzle temperature~~, @@print speed~~, and @@layer height~~, which affect the @@strength++, @@accuracy++, and @@surface quality++ of the final print are important parameters of this technique.

FDM-related entity categories:
@@Material^^ @@Application&& @@Printing Parameter~~ @@Quality Metric++

Fig. 5. An example of process-level manufacturing NER by our framework.

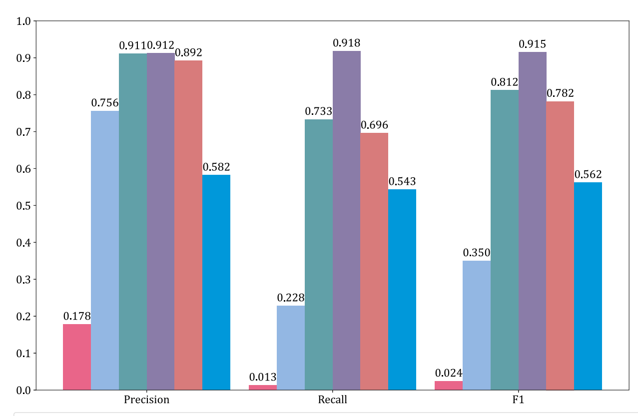

Fig. 6. F1 score comparison between ICL approaches, fine-tuned model and BERT-MRC.

The comparison between the performances of fine-tuned model, ICL approach (context-level KNN, k=20, simple style), and baseline model BERT-MRC is shown in Fig. 6. Even the amount of training data was very limited, the advanced LLMs such as GPT series, Claude-3, and fine-tuned model based on GPT-3.5 were still impressive. Among them, GPT-4 + ICL, Claude-3-opus + ICL, and fine-tuned GPT-3.5 outperformed the baseline model BERT-MRC, illustrating the effectiveness of LLM-based NER approaches in data-limited scenarios. According to the previous research, the performance of the baseline model would increase with more training data provided. Claude-3-opus + ICL achieved the best result with F1 value of 0.915, owing to its remarkable recall rate of 0.918, which is 0.185 above the second place, GPT-4. These results was reasonable as Claude-3-opus and GPT-4 were advanced in most generation tasks. The fine-tuned model significantly outperformed its original model GPT-3.5 + ICL with the F1 value of 0.782 vs 0.350, which was also close to the ICL approach with more advanced original model GPT-4, with only 0.03 lower in F1 value, revealing fine-tuning approach is another promising way for domain-specific NER, and may have better performance than the ICL approach with the same original model. The limited demonstrations in ICL prompt and the full utilization of all the training data in fine-tuning could explain this phenomenon.

Fig. 7 illustrates the enhancement in NER performance as the k value increases in the ICL approach and compares the performances of Claude-3-opus + ICL (simple style) with different retrieval modules, demonstrating that the context-level KNN retrieval is always better than the random retrieval. When k = 2, the KNN retrieval module outperformed the random retrieval module by about 0.0783 in F1

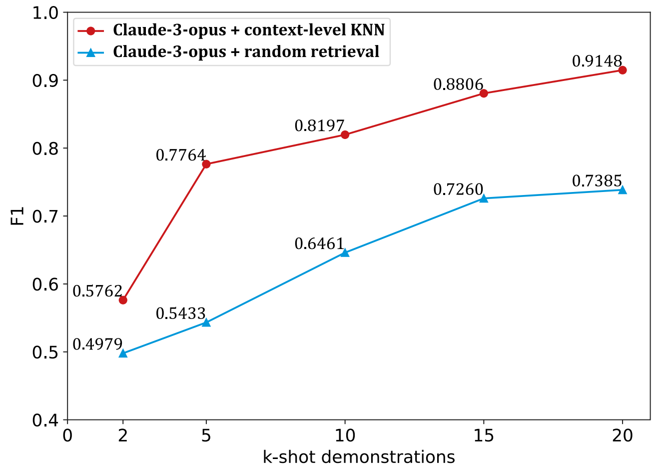

Fig. 7. The comparison on retrieval modules with Claude-3-opus + ICL approach (simple style).

Table 4
The comparison on different prompt styles with Claude-3-opus + ICL approach (context-level KNN, K=20).

Prompt Style	Precision	Recall	F1
Simple Style	0.9119	0.9177	0.9148
CoT Style	0.9206	0.9177	0.9192

value, which grew to 0.1763 when k = 20. It suggested using demonstrations closer to the input in the embedding space notably improved the performance of LLM + ICL approach. We also tested the k = 0 scenario which involves directly using Claude-3-opus without ICL for process-level NER. However, the performance was quite poor since the LLM could not understand the annotation style described in the task description without any demonstration, leading to several issues. For example, most input texts were not correctly reproduced in the outputs, many symbol pairs were added to the same side of their corresponding entities, and the entities in the middle of the input text were often missed in recognition. Thus, calculating the F1 value for k = 0 scenario was impossible.

We also explored the influence of different prompt styles on the performance. As shown in Table 4, with KNN retrieval configuration and K = 20, the Claude-3-opus + ICL approach with CoT prompt increases the precision by 0.0087 than the simple prompt. This performance indicated the effectiveness of the analysis component in mitigating wrong labeling by assisting reasoning of LLMs.

4.3. NER approaches comparison and selection

Based on our experiment, the ICL approach and the fine-tuning approach are both suitable for process-level NER, but they perform differently under various configurations. Firstly, the ICL approach is more suit to small-scale since it does not require training a new model, while fine-tuning approach is more suitable for batch processing since its prompt requires no demonstration, reducing time and token consumption. Meanwhile, the number of demonstrations used in the ICL approach is limited by the LLM's character limit, which is usually less than 10,000. In contrast, the fine-tuned model leverage a larger amount of training data, often leading to better performance when ample data are available for training.

4.4. Knowledge absorption

These performances demonstrated the effectiveness of our framework in recognizing domain-specific manufacturing entities, which contain plenty of knowledge about the specific manufacturing process. For example, as illustrated in Table 5, we rank FDM-related entities of each class by the number of times they were mentioned in the test set, revealing important information provided by the texts, including the major printing materials, the key printing parameters, the most common issues, etc. According to the texts, the most used FDM materials were thermoplastics, the major applications of FDM were involved in prototyping and biomedical field, while poor surface quality and warping were the most mentioned issues which should be concerned. By keeping the corpus up to date, this information can promote upskilling of FDM engineer, assist in parsing additive manufacturing report, and serve as a checklist for FDM process design and troubleshooting.

Besides FDM, this NER framework can be adopted in other manufacturing processes by building corpus of other domains and adjusting the taxonomy generation prompt, providing essential knowledge to improve the production process and employee ability.

5. Discussion and conclusion

This work achieves the recognition of process-level manufacturing entities with minimal training data by an LLM-based domain-specific NER framework, enabling manufacturing knowledge extraction from natural language texts. By leveraging the internalized manufacturing knowledge of LLM and the expert domain knowledge in academic materials through RAG, our work begins with the automated customization of process-level taxonomy instead of designing taxonomy by expert. We introduce two NER approaches based on ICL and fine-tuning, respectively. Both of them can accurately recognize the entities related to a specific manufacturing domain and divide them into categories defined by the taxonomy with very scarce training data. Since the training set is extremely small, their performances maybe further increased with more samples added. We also compare their advantages and limitations

Table 5
The most frequently mentioned FDM-related entities of each class recognized by Claude-3-opus + ICL approach (context-level KNN, K=20, CoT style).

Material type	Application	Product metric	Printing parameter	Post-processing technique
PLA	prototype	mechanical properties	layer thickness	painting
ABS	biomedicine	tensile strength	temperature	coating
Polymer	scaffold	surface roughness	raster angle	heating
PC	tissue engineering	surface finish	layer height	vapor deposition
thermalplastics	biomedical application	build time	nozzle diameter	chemical treatment
Printing equipment	**Common issue**	**Software concept**	**Design**	[a]**Standard**
FDM machine	poor surface quality	STL file	shape	ISO 178
nozzle	warping	.gcode	size	ASTM D638
printer	defects	CAD	geometry	ASTM D2344M
liquefier head	weak interlayer bonding	SML	internal structure	ASTM F2792
hot-end	printing lines	Slic3r	support	ISO/ASTM 52901:2017

[a] All the entities of the standard class only occurs once in the NER result of test set.

to provide suggestions for method selection in different scenarios. At last, we give an example of leveraging the NER framework to acquire process-level manufacturing knowledge from natural language corpus. Some key knowledge of a certain manufacturing process is implied in the extracted entities, which is important for forming an overview of a specific manufacturing technique and assisting manufacturing process design and optimization.

For the future work, the application of RAG in corpus construction is worth exploring. A possible direction is building a vector database of the whole papers and books and using LLMs to generate texts based on the vector database through RAG. These texts may have higher quality than the paragraphs directly extracted from academic materials. Moreover, this LLM-based NER framework is a cornerstone of process-level manufacturing knowledge graph construction. After linking the extracted entities by their relationships, the knowledge graph can serve as a process-level manufacturing knowledge base for employee upskilling, decision making, and process optimization. By applying LLMs in both process-level NER and relation extraction, it is possible to develop an end-to-end process-specific manufacturing knowledge graph construction framework.

CRediT authorship contribution statement

Xuan Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **John Ahmet Erkoyuncu:** Writing – review & editing, Validation. **Jerry Ying Hsi Fuh:** Writing – review & editing, Supervision, Funding acquisition. **Wen Feng Lu:** Writing – review & editing, Supervision. **Bingbing Li:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bingbing Li reports financial support was provided by U.S. Department of Energy and National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is funded by the U.S. Department of Energy (DOE) Office of Manufacturing and Energy Supply Chains (DE-EE0009726). We acknowledge Prof. Larry Smarr and Prof. Thomas DeFanti from University of California San Diego for HyperCluster computing support of San Diego Supercomputer Center (SDSC) National Research Platform (NRP) Nautilus sponsored by the U.S. National Science Foundation (NSF) (2100237, 2120019).

Data availability

Data will be made available on request.

References

[1] M. Fakhar Manesh, M.M. Pellegrini, G. Marzi, M. Dabic, Knowledge management in the fourth industrial revolution: Mapping the literature and scoping future avenues, IEEE Trans. Eng. Manage. 68 (1) (2021) 289–300, Available from: http://dx.doi.org/10.1109/TEM.2019.2963489.

[2] P. Tamla, F. Freund, M.L. Hemmje, Supporting named entity recognition and document classification in a knowledge management system for applied gaming, in: Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020), 2020, pp. 108–121, Available from: http://dx.doi.org/10.5220/0010145001080121.

[3] A. Kumar, B. Starly, "Fabner": information extraction from manufacturing process science domain literature using named entity recognition, J. Intell. Manuf. 33 (8) (2021) 2393–2407, Available from: http://dx.doi.org/10.1007/s10845-021-01807-x.

[4] M. Shrivastava, K. Seri, H. Wagatsuma, A named entity recognition model for manufacturing process based on the BERT language model scheme, in: Lecture Notes in Computer Science, Springer Nature Switzerland, 2022, pp. 576–587, Available from: http://dx.doi.org/10.1007/978-3-031-24667-8_50.

[5] S. Iloga, O. Romain, M. Tchuenté, An efficient generic approach for automatic taxonomy generation using HMMs, Pattern Anal. Appl. 24 (1) (2020) 243–262, Available from: http://dx.doi.org/10.1007/s10044-020-00918-0.

[6] R. Sujatha, R. Bandaru, R. Rao, Taxonomy construction techniques–issues and challenges, Indian J. Comput. Sci. Eng. 2 (5) (2011) 661–671, Available from: https://ijcse.com/docs/INDJCSE11-02-05-006.pdf.

[7] K. Shuster, S. Poff, M. Chen, D. Kiela, J. Weston, Retrieval augmentation reduces hallucination in conversation, in: M.-F. Moens, X. Huang, L. Specia, S.W.-t. Yih (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 3784–3803, Available from: http://dx.doi.org/10.18653/v1/2021.findings-emnlp.320.

[8] J. He, L. Wang, Y. Hu, N. Liu, H. Liu, X. Xu, H.T. Shen, ICL-D3IE: In-context learning with diverse demonstrations updating for document information extraction, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 19485–19494, Available from: https://doi.org/10.1109/ICCV51070.2023.01785.

[9] 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, W. Fedus, Emergent abilities of large language models, Trans. Mach. Learn. Res. (2022) Survey Certification. Available from: https://openreview.net/forum?id=yzkSU5zdwD.

[10] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, 2023, http://arxiv.org/abs/2201.11903.

[11] H. Fan, J. Fuh, W.F. Lu, A.S. Kumar, B. Li, Unleashing the potential of large language models for knowledge augmentation: A practical experiment on incremental sheet forming, Procedia Comput. Sci. 232 (2024) 1269–1278, Available from: https://doi.org/10.1016/j.procs.2024.01.125.

[12] S. Wang, X. Sun, X. Li, R. Ouyang, F. Wu, T. Zhang, J. Li, G. Wang, Gpt-ner: Named entity recognition via large language models, 2023, Available from: https://doi.org/10.48550/arXiv.2304.10428.

[13] F. Yan, P. Yu, X. Chen, LTNER: Large language model tagging for named entity recognition with contextualized entity marking, 2024, Available from: https://doi.org/10.48550/arXiv.2404.05624.

[14] X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, J. Li, A unified MRC framework for named entity recognition, 2022, Available from: https://arxiv.org/abs/1910.11476.

[15] M. Alavi, D.E. Leidner, Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues, MIS Q. 25 (1) (2001) 107, Available from: http://dx.doi.org/10.2307/3250961.

[16] G. Hackbarth, The impact of organizational memory on IT systems, AMCIS 1998 Proceedings (1998) 197, Available from: https://aisel.aisnet.org/amcis1998/197.

[17] M. Jawad, C. Dhawale, A.A.B. Ramli, H. Mahdin, Adoption of knowledge-graph best development practices for scalable and optimized manufacturing processes, MethodsX 10 (2023) 102124, Available from: http://dx.doi.org/10.1016/j.mex.2023.102124.

[18] A. de Bem Machado, S. Secinaro, D. Calandra, F. Lanzalonga, Knowledge management and digital transformation for industry 4.0: a structured literature review, Knowl. Manag. Res. Pract. 20 (2) (2021) 320–338, Available from: http://dx.doi.org/10.1080/14778238.2021.2015261.

[19] M. Capestro, S. Kinkel, Industry 4.0 and knowledge management: A review of empirical studies, in: Knowledge Management and Industry 4.0, Springer International Publishing, 2020, pp. 19–52, Available from: http://dx.doi.org/10.1007/978-3-030-43589-9_2.

[20] J. Li, A. Sun, J. Han, C. Li, A survey on deep learning for named entity recognition, IEEE Trans. Knowl. Data Eng. 34 (1) (2022) 50–70, Available from: https://doi/org/10.1109/TKDE.2020.2981314.

[21] L. Weston, V. Tshitoyan, J. Dagdelen, O. Kononova, A. Trewartha, K.A. Persson, G. Ceder, A. Jain, Named entity recognition and normalization applied to large-scale information extraction from the materials science literature, J. Chem. Inf. Modeling 59 (9) (2019) 3692–3702, Available from: https://doi/org/10.26434/chemrxiv.8226068.v1.

[22] D.Q. Nguyen, Z. Zhai, H. Yoshikawa, B. Fang, C. Druckenbrodt, C. Thorne, R. Hoessel, S.A. Akhondi, T. Cohn, T. Baldwin, et al., Chemu: named entity recognition and event extraction of chemical reactions from patents, in: Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42, Springer, 2020, pp. 572–579, Available from: https://doi.org/10.1007/978-3-030-45442-5_74.

[23] H. Cho, H. Lee, Biomedical named entity recognition using deep neural networks with contextual information, BMC Bioinf. 20 (2019) 1–11, Available from: https://doi.org/10.1186/s12859-019-3321-4.

[24] L. Yao, H. Huang, K.-W. Wang, S.-H. Chen, Q. Xiong, Fine-grained mechanical Chinese named entity recognition based on ALBERT-AttBiLSTM-CRF and transfer learning, Symmetry 12 (12) (2020) 1986, Available from: https://doi.org/10.3390/sym12121986.

[25] S. Addepalli, T. Weyde, B. Namoano, O.A. Oyedeji, T. Wang, J.A. Erkoyuncu, R. Roy, Automation of knowledge extraction for degradation analysis, CIRP Ann 72 (1) (2023) 33–36, Available from: https://doi.org/10.1016/j.cirp.2023.03.013.

[26] T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, in: NIPS '20, Curran Associates Inc., Red Hook, NY, USA, 2020, Available from: https://doi.org/10.5555/3495724.3495883.

[27] T. Kojima, S.S. Gu, M. Reid, Y. Matsuo, Y. Iwasawa, Large language models are zero-shot reasoners, in: Proceedings of the 36th International Conference on Neural Information Processing Systems, in: NIPS '22, Curran Associates Inc., Red Hook, NY, USA, 2024, Available from: https://doi.org/10.5555/3600270.3601883.

[28] O. Rubin, J. Herzig, J. Berant, Learning to retrieve prompts for in-context learning, in: M. Carpuat, M.-C. de Marneffe, I.V. Meza Ruiz (Eds.), Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Seattle, United States, 2022, pp. 2655–2671, Available from: https://doi.org/10.18653/v1/2022.naacl-main.191.

[29] J. Chen, Y. Lu, H. Lin, J. Lou, W. Jia, D. Dai, H. Wu, B. Cao, X. Han, L. Sun, Learning in-context learning for named entity recognition, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 13661–13675, Available from: https://aclanthology.org/2023.acl-long.764.

[30] C. Cai, Q. Wang, B. Liang, B. Qin, M. Yang, K.F. Wong, R. Xu, In-context learning for few-shot multimodal named entity recognition, in: The 2023 Conference on Empirical Methods in Natural Language Processing, 2023, Available from:.

[31] J. Choi, B. Lee, Accelerating materials language processing with large language models, Commun. Mater. 5 (1) (2024) 13, Available from: https://doi.org/10.1038/s43246-024-00449-9.

[32] M. Monajatipoor, J. Yang, J. Stremmel, M. Emami, F. Mohaghegh, M. Rouhsedaghat, K.W. Chang, LLMs in biomedicine: A study on clinical named entity recognition, 2024, Available from: https://doi.org/10.48550/arXiv.2404.07376.

[33] Y. Hu, Q. Chen, J. Du, X. Peng, V.K. Keloth, X. Zuo, Y. Zhou, Z. Li, X. Jiang, Z. Lu, et al., Improving large language models for clinical named entity recognition via prompt engineering, J. Am. Med. Inform. Assoc. (2024) ocad259, Available from: https://doi.org/10.1093/jamia/ocad259.

[34] Z. Li, X. Li, Y. Liu, H. Xie, J. Li, F.l. Wang, Q. Li, X. Zhong, Label supervised llama finetuning, 2023, Available from: https://doi.org/10.48550/arXiv.2310.01208.

[35] Z. Wan, F. Cheng, Z. Mao, Q. Liu, H. Song, J. Li, S. Kurohashi, Gpt-re: In-context learning for relation extraction using large language models, 2023, Available from: https://doi.org/10.48550/arXiv.2305.02105.

[36] H. Fan, X. Liu, J.Y.H. Fuh, W.F. Lu, B. Li, Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics, J. Intell. Manuf. (2024) 1–17, Available from: https://doi.org/10.1007/s10845-023-02294-y.

[37] X. Wang, N. Anwer, Y. Dai, A. Liu, ChatGPT for design, manufacturing, and education, Procedia CIRP 119 (2023) 7–14, Available from: https://doi.org/10.1016/j.procir.2023.04.001.

[38] S. Badini, S. Regondi, E. Frontoni, R. Pugliese, Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting, Adv. Ind. Eng. Polym. Res. 6 (3) (2023) 278–287, Available from: https://doi.org/10.1016/j.aiepr.2023.03.003.