Supporting information for

Accelerating materials language processing with large language models

Jaewoong Choi¹, Byungju Lee^{1,*}

¹Computational Science Research Center, Korea Institute of Science and Technology, Seoul, Republic of Korea

*Corresponding author: Dr. Byungju Lee (blee89@kist.re.kr)

Tables

Supplementary Table 1. Summary of comparing GPT-enabled text classification	models
against the SOTA models	

Dataset	Approach	Model	Precision	Recall	Accuracy
Battery- related papers	SOTA BatteryBERT- cased		96.6%	99.5%	97.5%
	Zero-shot learning	GPT-3.5 ('text- embedding- ada-002') with Original labels	63.2%	100%	63.2%
		GPT-3.5 ('text- embedding- ada-002') with Crude labels	84.5%	97.9%	87.3%
		GPT-3.5 ('text- embedding- ada-002') with Designated labels	88.6%	98.3%	91.0%
		GPT-3.5 ('text- embedding- ada-002') with Verbose labels	90.8%	98.9%	93.0%
		2-way 1-shot learning with GPT-3.5 ('text-davinci- 003')	95.0%	98.6%	95.7%
	Few-shot learning	2-way 5-shot learning with GPT-3.5 ('text-davinci- 003')	95.0%	99.1%	96.1%
		2-way 5-shot learning with GPT-4 ('gpt- 4-0613')	95.4%	98.9%	96.3%
	Fine-tuning of GPT	GPT-3 ('davinci')	95.8%	98.9%	96.6%

Supplementary Table 2. Summary of comparing GPT-enabled NER models against the SOTA models

Dataset	Approach	Model	Category	Precision	Recall	F1-score
	SOTA		APL	83.7%	77.8%	80.6%
			СМТ	83.3%	88.9%	86.0%
		MotREDT	DSC	90.3%	94.0%	92.1%
		uncased	MAT	87.3%	93.5%	90.3%
			PRO	85.5%	81.0%	83.2%
			SMT	79.1%	83.8%	81.4%
Solid-state			SPL	87.7%	77.1%	82.1%
materials			APL	96.4%	71.8%	82.3%
		GPT-3 ('davinci')	СМТ	94.0%	90.3%	92.1%
	Fine-		DSC	97.3%	93.0%	95.1%
	tuning of		MAT	97.5%	91.7%	94.6%
	GPT		PRO	95.3%	80.2%	87.1%
			SMT	93.9%	84.7%	89.1%
			SPL	99.8%	80.8%	89.3%
	SOTA	MatBERT- uncased	BASEMAT	-	-	72.0%
			DOPANT	-	-	82.0%
Dopant			DOPMODQ	-	-	62.0%
materials	Fine- tuning of GPT	GPT-3 ('davinci')	BASEMAT	93.4%	62.0%	74.6%
			DOPANT	95.6%	64.4%	77.0%
			DOPMODQ	92.7%	59.4%	72.4%
AuNPs	SOTA	MatBERT-	DES	70%	52%	56%
		uncased	MOR	83%	64%	70%
	Few-shot learning	Random	DES	63.2%	68.6%	65.8%
		retrieval				
		with GPT-		97.4%	61.7%	
		3.5 ('text-	MOR			75.6%
		davinci-				
		003')				
		Task	DES	65.1%	80.0%	71.8%

informed random retrieval with GPT- 3.5 ('text- davinci- 003')	MOR	97.9%	68.0%	80.2%
kNN	DES	63.6%	99.8%	77.7%
retrieval with GPT- 3.5 ('text- davinci- 003')	MOR	97.7%	83.0%	89.8%
kNN	DES	69.7%	99.6%	82.0%
retrieval with GPT- 4 ('gpt-4- 0613')	MOR	99.8%	87.7%	93.3%

Supplementary Table 3. Summary of comparing GPT-enabled extractive QA models against the SOTA models

Dataset	Approach	Model	Precision	Recall	F1-score
	SOTA	BatteryBERT-	77.49%	71.69%	74.48%
		cased			
	Zero-shot	GPT-3.5	60.92%	79.96%	69.15%
	learning	('text-davinci-			
Battery- device QA		003')			
	Task-	GPT-3.5	72.89%	80.65%	76.57%
	informed	('text-davinci-			
	Zero-shot	003')			
	learning				
	Fine-tuning	GPT-3	88.07%	88.35%	88.21%
		('davinci')			

Label	Publication title	Journal
False	Quality of health care with regard to detection and treatment of mental disorders in patients with coronary heart disease (MenDis-CHD): study protocol	BMC Psychology
	Handwriting capacity in children newly diagnosed with Attention Deficit Hyperactivity Disorder	Research in Developmental Disabilities
	Effects of compensatory cognitive training intervention for breast cancer patients undergoing chemotherapy: a pilot study	Supportive Care in Cancer
	Architecture of fluid intelligence and working memory revealed by lesion mapping	Brain Structure and Function
	Neural correlates of RDoC-specific cognitive processes in a high-functional autistic patient: a statistically validated case report	Journal of Neural Transmission
True	In-operando temperature measurement across the interfaces of a lithium-ion battery cell	Electrochimica Acta
	An all-solid state NASICON sodium battery operating at 200 °C	Journal of Power Sources
	Surface reactions and performance of non-aqueous electrolytes with lithium metal anodes	Journal of Power Sources
	Metal tetrabromophthalocyanines mediate the structure and electrochemical performance of lithium iron phosphate as cathode materials for lithium-ion batteries	Journal of Electroanalytical Chemistry
	Electrochemical characterization of electrolytes for lithium-ion batteries based on lithium difluoromono(oxalato)borate	Journal of Power Sources

Supplementary Table 4. Parts of text classification dataset for each label. Papers in the False label are related to medical or psychological field applications, while papers in the True label include materials science studies such as battery synthesis, reaction or characterization of materials.

Figures

(a)

Abstract: Flexible Li-O2 batteries have attracted worldwide research interests and been considered to be potential alternatives for the next-generation flexible devices. Nitrogen-doped carbon nanofibers (N-CNFs) prepared by electrospinning are used as flexible substrate and an amorphous TiO2 layer is coated by atomic layer deposition (ALD) and then decorated with Ru nanoparticles. The Ru/N-CNFs@TiO2 composite is directly used as a free-standing electrode for Li-O2 batteries and the electrode delivers a high specific capacity, improved round-trip efficiency and good cycling ability. The superior electrochemical performance can be attributed to the amorphous TiO2 protecting layer and superior catalytic activity of Ru nanoparticles. Based on density functional theory (DFT) calculations from first principles, the carbon electrode after coating with TiO2 is more stable during discharge/charge process. The analysis of Li2O2 on three different interfaces (Li2O2/N-CNFs, Li2O2/TiO2, and Li2O2/Ru) indicates that the electron transport capacity was higher on Ru and TiO2 compared with N-CNFs, therefore, Li2O2 could be formed and decomposed more easily on the Ru/N-CNFs@TiO2 cathode. This work paves a way to develop the free-standing cathode materials for the future development of high-performance flexible energy storage systems.



(d) (b) (c) MAT DSC MOR Monoclinic zirconia nanowires were synthesized by Sm doping favors the formation of hexagonal phase in Fig 4a corresponds to the TEM images of AuNRs chemical vapor decomposition using ZrCl4 powder as the host lattice of ZnS: Sm dispersion in the water having length 53-60nm and a starting material. Based on x-ray diffraction and The self-activated luminescence intensity increases width 20-25nm (aspect ratio =2.5) along with some Raman spectroscopy data, the resulting crystal square or rectangular shaped particles with the increase in the amount of Sm doping up to a maximum of 0.2 mol% of dopant. structure was found to be single crystalline monoclinic zirconia. An Au thin film was pre-depositing on a graphite substrate as a catalyst (e) Context Answer Question The blended slurry was then cast onto a clean current collector (Al foil for the cathode and Cu What is cathode? foil for the anode) and dried at 90 °C under vacuum overnight. Finally, the obtained Answer: Al foil What is anode? electrodes were cut into desired shapes on demand. It should be noted that the electrode What is electrolyte? mass ratio of cathode/anode is set to about 4. thus achieving the battery balance.

Supplementary Figure 1. Examples of annotation dataset. (A) Battery-related paper classification dataset (https://huggingface.co/batterydata). This data includes 46,663 labelled papers for which the labels are battery or non-battery are annotated based on their publication sources. The authors defined the battery-related journals and selected the papers in those journals as battery-related papers (ground-truth). (B) solid-state materials NER dataset. This dataset includes 800 annotated abstracts. This data set is intended to provide a comprehensive range of relevant information without focusing on any specific aspect of solid materials. Due to the broad definition of entities, solid state data sets usually contain more entities per paragraph. Accordingly, sufficient entity-centric entity tagging prompts could be obtained. (C) doped materials NER dataset. This dataset includes 455 annotated abstracts. The process involved annotating tokens on a sentence-by-sentence basis, where each sample corresponds to a single sentence. Sentences were only annotated if they contained precise and direct information about doping in solid-state materials. Accordingly, we tried to apply few-shot learning of GPT 3.5 model rather than promptcompletion fine-tuning module, which requires sufficient number of datasets. (D) AuNP synthesis NER dataset. The objective of this dataset is to collect information about the morphologies and descriptions of AuNP from specific sections of full-text literature on AuNP synthesis. A sole annotator annotated a collection of 85 characterization paragraphs extracted from 73 articles related to AuNP synthesis. Similar to doped materials set, we tried to apply few-shot learning of GPT 3.5 model considering the amount of available data. (E) battery-device QA dataset. The particular Q&A dataset follows the format of the

SQuAD dataset and includes three types of questions: 'what is the anode?', 'what is the cathode?', and 'what is the electrolyte?'.



Supplementary Figure 2. Performance of AuNPs entity recognition of few-shot learning using GPT-3.5 ('text-davinci-003') (A–B) AuNPs entity recognition recall and precision of random retrieval, task-informed random retrieval, and kNN retrieval methods. Here, the precision and recall are calculated, after boundary relaxation is conducted on the prediction results. As the entity-wise performances of prior models are not disclosed, we reported the performance of our methods only. In case of DES, random retrieval, taskinformed random retrieval and kNN retrieval showed recall values of 63.2%, 65.1% and 63.6%, while precision values are 68.6%, 80.0% and 99.8%. By providing similar examples in the prompt, we were able to obtain high entity recognition performance. In case of MOR, three methods provided high recall of 97.4%, 97.9% and 97.7%, while the precision values are relatively low, 61.7%, 68.0% and 83.0%. Here, we can observe that the effect of adding task-informing sentences is positive on performance, while kNN retrieval strategy is effective in increasing precision performance. This indicates that providing similar examples can help the identification of MOR entities more precisely. Task: the task is to extract the descriptive entities in the given documents. below are some examples.

prompt: ... From the TEM image of Au@PS core – shell particle (Fig. 6), it is evident that the size of spherical Au particle is ~115 nm, and the outside PS shell (formed by emulsifier - free polymerization) has a thickness of 2–3 nm ...

==>\n completion: ... From the TEM image of Au@PS core – shell particle (Fig. 6), it is evident that the size of @@spherical## Au particle is ~115 nm, and the outside PS shell (formed by emulsifier - free polymerization) has a thickness of 2–3 nm ... \n\n###\n\n

prompt: ... Rhombic dodecahedral gold nanocrystals bounded exclusively by { 110 } surfaces were synthesized following our reported procedure . Edge- and corner - truncated octahedral nanocrystals contain significant { 110 } faces from the truncated edges

==>\n completion: ... @@Rhombic dodecahedral## gold nanocrystals bounded exclusively by { 110 } surfaces were synthesized following our reported procedure . Edge- and corner - truncated @@octahedral## nanocrystals contain significant { 110 } faces from the truncated edges \n\n###\n\n

prompt: ... Strong interactions of rod – rod , and rod – plane here may favor the multiple - ring formation for the gold nanorod suspension , differing from the interactions of sphere – sphere and sphere – plane for the spherical gold nanoparticles

==>\n completion: ... Strong interactions of rod - rod , and rod - plane here may favor the multiple - ring formation for the gold nanorod suspension , differing from the interactions of sphere - sphere and sphere - plane for the @@spherical## gold nanoparticles \n/n###(n\n

prompt: ... Fig . 1(c) clearly indicates that the SiO2@Au core – shell NCs synthesized were speckled particles i.e., they retain the gold particle size as was used in the formation of core – shell particles ==>\n

Supplementary Figure 3. Example of few-shot learning prompt based on kNN similar examples; The prompt consists of task information, 3-shot (prompt-completion pairs), and new input.