Extracting a Database of Challenges and Mitigation Strategies for Sodium-ion Battery Development

Mrigi Munjal[1,*], Thorben Prein[2,*], Vineeth Venugopal[1], Kevin J. Huang[1], Elsa Olivetti[1]

[1] Massachusetts Institute of Technology, [2] Technische Universität München
t.prein@tum.de, {mrigi, vineethv, kjhuang, elsao}@mit.edu
[*]Equal contribution

Abstract

Sodium-ion batteries (SIBs) have been gaining attention for applications like grid-scale energy storage, largely owing to the abundance of sodium and an expected favorable $ kWh^{-1} figure. SIBs can leverage the well-established manufacturing knowledge of Lithium-ion Batteries (LIBs), but several materials synthesis and performance challenges for electrode materials need to be addressed. This work extracts a large database of challenges restricting the performance and synthesis of SIB cathode active materials (CAMs) and pairs them with corresponding mitigation strategies from the SIB literature by employing custom natural language processing (NLP) tools. The derived insights enable scientists in research and industry to navigate a large number of proposed strategies and focus on impactful scalability-informed mitigation strategies to accelerate the transition from lab to commercialization.

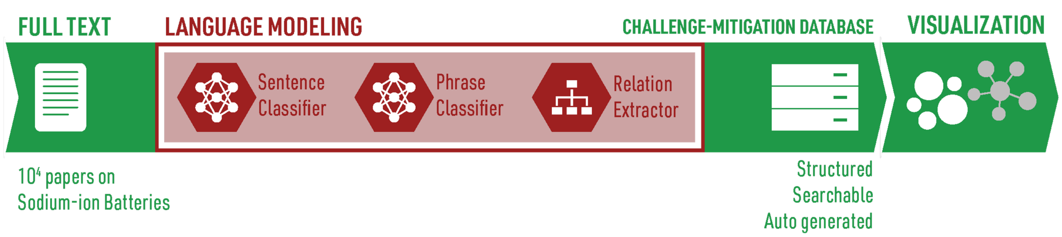

Figure 1: **NLP Pipeline**: We implemented a sequential filtering and visualization pipeline, employing sentence classification [1], phrase-level classification, and relationship extraction [2, 3]. The outcomes are visually represented through BERT-based topic modeling [4] in Figure 3 and knowledge graphs in Figure 5.

1 Introduction

Clean energy transition is crucial to mitigating climate change [5]. Energy storage devices, particularly compact chemical energy formats like batteries, are essential for managing the intermittent nature of renewable sources like solar and wind energy [6]. Lithium-ion batteries (LIBs) are the most common of all batteries today since they offer the highest energy density and output voltage compared to alternatives [7]. However, concerns have been raised regarding the skewed geographic impact of lithium extraction and the price impact of rapid growth [8]. The development of alternative battery chemistries, such as those based on sodium, could offer diversification opportunities [9, 10, 11]. Sodium-ion batteries (SIBs) can be good replacements for LIBs for grid storage. While SIB fabrication can parallel that of LIBs in terms of cell manufacturing and assembly [12, 13], the commercialization

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

of SIBs is lagging behind. This limitation predominantly stems from the performance and synthesis routes of cathode active materials for SIBs [14, 15].

Keeping track of developments can be complex as insights across material types are buried in an enormous corpus of >10,000 publications [16]. Therefore, it is vital to offer a coherent overview and an ability to efficiently query for researchers to develop these batteries.

NLP-based studies in materials science in the last few years have focused on the extraction of quantitative synthesis related data and materials properties, enabling training of regressors for property related models [17, 18, 19, 20]. In contrast to these studies that often omit the authors' rationale and compel researchers to depend on model-generated interpretations of the data, our methodology focuses on capturing author rationale regarding cause-effect relationships for challenges. These are then aligned with the relevant mitigation strategies. This approach maintains a high degree of adaptability in the array of suggested solutions. We believe that the intentional focus of our method on textual references to materials engineering methods and mechanisms, rather than on quantitative data can enhance the understanding of existing approaches. Our primary contributions are as follows:

- Detailed database on SIBs materials challenges and mitigation strategies
- Interactive search tool for scientists to find SIB-related mitigation strategies and linked mechanistic causes corresponding to observed performance characteristics
- Classifiers and training data for efficient battery literature screening, extendable to lithium-ion battery research

Our novel approach utilizes a systematic extraction of challenges and mitigation strategies from the literature using a two-stage process that increases the accuracy and relevance of the information gathered. Moreover, this specialized focus on SIBs, which are a critical and emerging area in battery technology, fills a specific knowledge gap in the field. The interactive search tool not only aids in research but also in practical problem-solving, allowing for a more dynamic and user-friendly way to access complex information. Our methodology not only enhances the efficiency of literature review but also provides a scalable model that can be adapted to other areas of battery research.

2 Methods and Framework

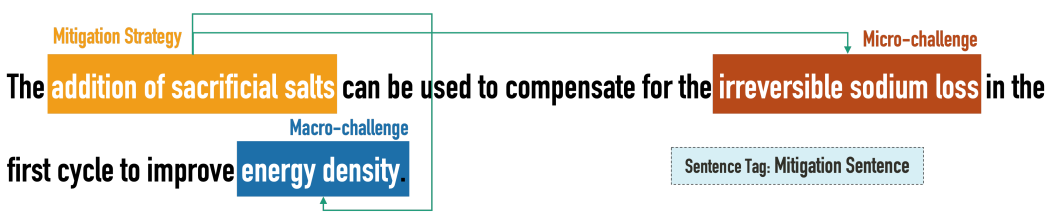

Figure 2: **Senetence, Phrase and Relationship Extraction:** The Sentence classification tag of this example sentence is "Mitigation". The phrases and relationships between phrases are identified in this annotated example.

In recent years, there has been a significant surge in published research papers on SIBs to the order of 10^4 [16]. It would take over 20,000 uninterrupted hours to manually read and comprehend every single paper, assuming an average of two hours per paper. However, our pipeline can efficiently process the entire corpus of scientific literature in just 6 hours, without requiring active human intervention. Recently, several fields have been trying to use visual maps to organize and analyze topics in text [21]. Our methodology drew inspiration from seminal works in sentence-oriented sentiment analysis [22, 23, 24], sentence-based search mechanisms [1, 25], open information extraction [3, 2], and topic modeling [4]. However, to the best of our knowledge, application of this scale of processing has not been developed for battery literature.

We extracted a structured database, which enables us to identify prominent topics from an extensive breadth of topics and infer generalizable implications about scalability-informed lab research. The relevant papers were downloaded using a custom download pipeline developed by the group [18]. We organized extracted information for 18 cathode material types across layered metal oxides, Prussian

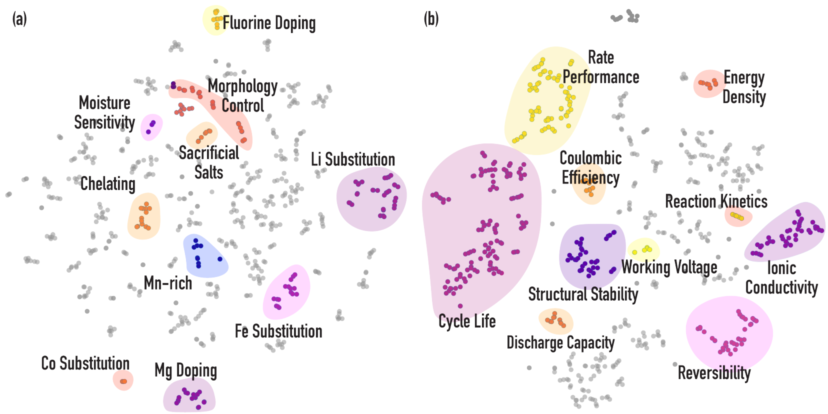

Figure 3: **Exploratory view of extracted mitigation sentences** (a) Visualized space of embeddings of mitigation strategy phrases. Marked regions showcase various mitigation strategy topics (b) Visualized space of embeddings of challenges. Marked regions showcase various micro-challenges. The candidate material here is Sodium Iron Manganese Oxide.

blue analogs, and polyanionics. Our framework allows for enhanced accessibility for non-experts in the domain and interpolating scalability strategies across material types. Figure 1 illustrates the NLP pipeline tasks. Our adaptable pipeline can be used for several similar tasks. Our sentence classification, phrase classification and relationship extraction methodology is described below.

Sentence Classification: Our sentences of interest are defined as:

- **Challenge Sentences** encapsulate discussions about all performance or materials-related flaws, their mechanistic origins, and shortcomings in synthesis procedures. (e.g., "*Irreversible sodium loss in sodium-ion batteries results in low specific capacity.*")
- **Mitigation Sentences** involve references to enhancing the material's key performance indices or associated properties and methods. (e.g. in Figure 2)

Phrase Classification and Relationship Extraction: To extract relevant context spans from the sentences, we developed a phrase-level classification scheme. We assessed challenges at two scales for phrase-level classification: macro-challenges and micro-challenges. Macro-challenges are key performance challenges directly linked to resultant performance like "*low specific capacity*," "*poor rate capability*.", etc. Whereas, micro-challenges are mechanistic causes of these macro-challenges that indicate the underlying phenomena that contribute to macro-challenges like "*low redox activity*", "*irreversible Na loss*," etc. Besides these two types of challenges, we extracted phrases related to mitigation strategies (e.g., "*addition of sacrificial salts*"). We also extracted the relationship among those phrases. Figure 3 visualizes the mitigation strategies space for NFMO (Sodium Iron Manganese Oxide) clustered using phrases. Extracted challenges and mitigation strategies were can be to constitute challenge-mitigation pairs.

Model Evaluation: For the development of our sentence and phrase classifiers, we benchmarked with a variety of approaches on our dataset, using stratified data splits and hyperparameter optimization. As seen in Table 1, the performance of BERT-based models [26] was commendable. The best results were attained with SciBERT [27] and MatSciBERT [28], which were both pretrained in the domain of scientific publications. We also discovered that recent autoregressive Large Language Models (LLMs), such as GPT-3 [29], yielded promising results, even when only presented with 10 in-context examples [29]. We utilized a widely employed model, namely, the Dynamic Graph Based Information Extraction (DyGIE) for phrase and relationship extraction [3].

Our approach enabled us to create a comprehensive database of materials-specific challenges and their mechanistic sources that impact material performance, along with corresponding mitigation strategies. This valuable overview allows us to scrutinize lab-probed strategies and hone in on compelling candidates for further analysis. In total, we obtained a database of approximately 31,000 challenge and mitigation sentences. Out of these, our classifiers identified 9,000 relations. Analyzing the diversity of papers in the final relational database, we note that 91% of the initial papers are

	F1	P	R
MATSCIBERT	83.1 (1.2)	83.4 (1.1)	83.8 (1.3)
SCIBERT	**84.1 (1.7)**	**84.4 (1.5)**	**84.2 (1.4)**
GPT3 @10 SHOTS	73.2	75.9	72.5

	Sentences	Phrases	Relations
CHALLENGE	84.1	67.5	39.4
MITIGATION	83.1	67.9	50.8

Table 1: (a) Model comparison on the sentence classification task for challenge sentences; (b) F1 scores for sentence classification, phrase classification and relationship extraction

represented in the mitigation sentences database and 84% in the challenge sentences database, underscoring the comprehensive coverage of our source material. Domain experts evaluated the informativeness and correctness of the extracted relationships which yielded a correctness score of 85%. The database can be queried for a large variety of challenge-mitigation pairs using challenge topics to aid in expediting the development of commercial-scale sodium-ion batteries.

3 Results and Discussion

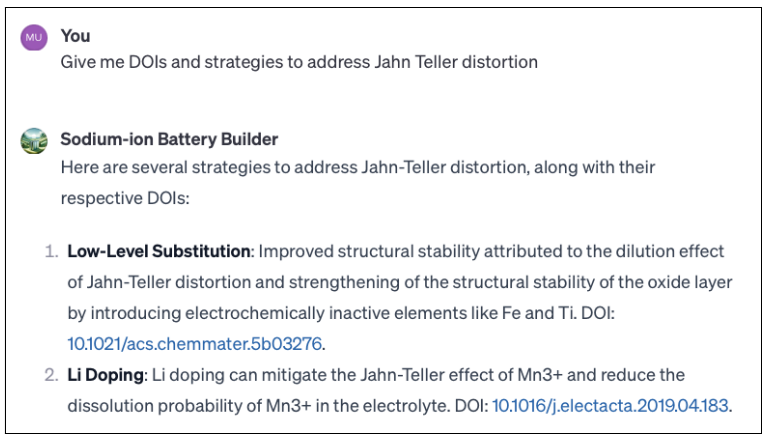

Figure 4: **Sodium-ion Battery Builder GPT** Snapshot of our chatbot to interact with our database using OpenAI custom GPTs. The chatbot can be queries to understand underlying mechanistic causes of challenges encountered in battery development as well as strategies to address them from reliable sources.

Our NLP pipeline enables a broader understanding of SIBs. Traditional literature reviews often depend on specialized review articles or focus on particular materials or methods, limiting the scope of research insights. Conducting effective searches typically necessitates familiarity with domain-specific terminology, such as types of materials, key performance indicators, current research challenges, and techniques for processing and analysis. While the initial training of our machine learning models involved domain experts for data labeling, the use of the pipeline does not necessitate prior knowledge in the field. This strategy helps in mitigating biases associated with the selection of work and topics. Our models have also shown effective adaptability to related fields like LIB data without the need for further expert input. Furthermore, this method promotes diversity in the database by encompassing a wider range of research works, reducing the risk of overlooking significant studies due to narrow search criteria or unconscious biases.

Our extracted databases were integrated with ChatGPT's custom GPT builder [30] to create a Sodium-ion Battery Builder chatbot (Figure 4) that can be queried to find relevant mitigation strategies for a given macro- or micro-challenge. The chatbot can also handle queries for retrieving linked micro-challenges for a given macro-challenge. The chatbot is designed to provide the user with the DOI that is linked to retrieved scientific information to increase reliability and minimize hallucination. Incorporating entire research papers to construct a specialized GPT for battery development is impractical, given the vast quantity and extensive size of these documents, which exceed the context window limitations of generative models. Conversely, our database adopts a selective approach, concentrating exclusively on segments of the literature that address specific challenges and their respective mitigation strategies in battery technology. This targeted filtration enables efficient querying for this Retrieval Augmented Generation chatbot.

We clustered our mitigation database (Figure 3) using sentence BERT [31] computed embeddings of the strategies, revealing a holistic map of the diverse mitigation strategies space. We also created knowledge graphs as shown in Figure 5 to identify "challenge-mitigation pairs" of interest.

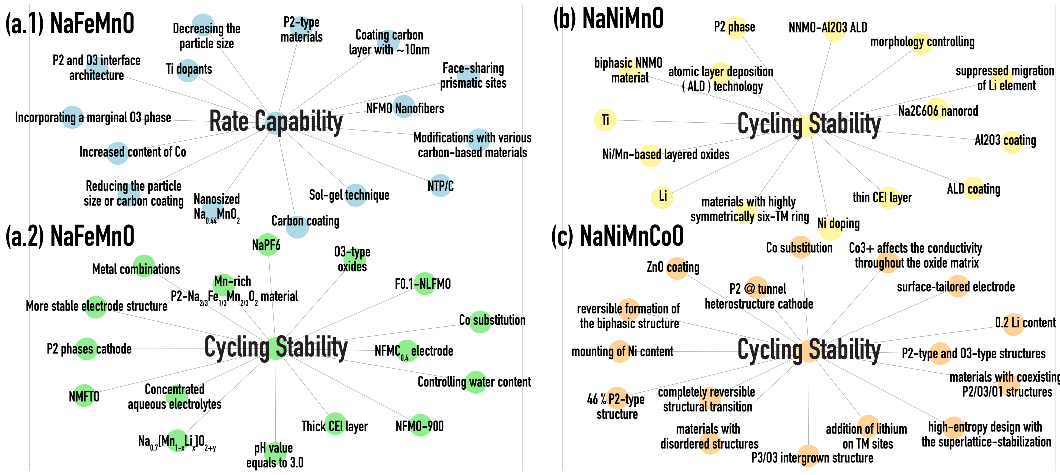

Figure 5: **Knowledge-graph representation** of some mitigation strategies linked to the macro-challenges of 'Rate Capability' and 'Cycling Stability' which are the two most commonly reported macro-challenges. These graphs were constructed by utilizing our relationship extractor to get "challenge-mitigation pairs". (a.1) and (a.2) are for NaFeMnO type materials, (b) is for NaNiMnO type materials and (c) is for NaNiMnCoO type materials.

4 Limitations and Future Outlook

Sodium-ion batteries are a promising energy storage option due to abundant and affordable sodium. Utilizing NLP, we extract crucial data to aid in the development of SIBs. Our framework propels efficient SIB exploration by pulling information from a large corpus. NLP has the potential to steer future research toward fruitful avenues, bringing us closer to realizing the full potential of commercial SIBs. However, transitioning from identifying 'challenge-mitigation' patterns to selecting the optimal strategy remains a complex task necessitating human expertise. This NLP framework allows us to select case studies of importance that can be evaluated further with a technoeconomic lens for a scalability analysis coupled with extracted performance metrics.

5 Data Availability

The extracted database and classifiers are available at github.com/olivettigroup/NLP4SIB.

Sodium Ion Battery Builder GPT built on OpenAI's ChatGPT platform as a custom GPT is available here: chat.openai.com/g/g-2gOTffBeL-sodium-ion-battery-builder

Acknowledgments and Disclosure of Funding

The authors thank the support from our sponsors, Shell USA and Dr Ryan Stephens (Team Lead, Energy Storage at Shell, USA. This work is also supported by the National Science Foundation DMREF Awards 1922090, 1922311, 1922372, the National Science Foundation FMRG award, and the Office of Naval Research (ONR) under contract N00014-20-1-2280.

References

[1] Dan Lahav, Jon Saad Falcon, Bailey Kuehl, Sophie Johnson, Sravanthi Parasa, Noam Shomron, Duen Horng Chau, Diyi Yang, Eric Horvitz, Daniel S Weld, et al. A search engine for discovery of scientific challenges and directions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11982–11990, 2022.

[2] Zexuan Zhong and Danqi Chen. A frustratingly easy approach for entity and relation extraction. *arXiv preprint arXiv:2010.12812*, 2020.

[3] David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. Entity, relation, and event extraction with contextualized span representations. *arXiv preprint arXiv:1909.03546*, 2019.

[4] Nicole Peinelt, Dong Nguyen, and Maria Liakata. tbert: Topic models and bert joining forces for semantic similarity detection. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 7047–7055, 2020.

[5] IPCC. *Summary for policymakers*, pages 7–22. Cambridge University Press, Cambridge, UK, 2007.

[6] John B. Goodenough and Youngsik Kim. Challenges for rechargeable li batteries. *Chemistry of Materials*, 22(3):587–603, 2010.

[7] Verónica Palomares, Paula Serras, Irune Villaluenga, Karina B. Hueso, Javier Carretero-González, and Teófilo Rojo. Na-ion batteries, recent advances and present challenges to become low cost energy storage systems. *Energy Environ. Sci.*

[8] Peter Greim, A. A. Solomon, and Christian Breyer. Assessment of lithium criticality in the global energy transition and addressing policy gaps in transportation. *Nature Communications*, 11(1):4570, Sep 2020.

[9] Shyue Ping Ong, Vincent L. Chevrier, Geoffroy Hautier, Anubhav Jain, Charles Moore, Sangtae Kim, Xiaohua Ma, and Gerbrand Ceder. Voltage, stability and diffusion barrier differences between sodium-ion and lithium-ion intercalation materials. *Energy Environ. Sci.*, 4:3680–3688, 2011.

[10] Naoaki Yabuuchi, Kei Kubota, Mouad Dahbi, and Shinichi Komaba. Research development on sodium-ion batteries. *Chemical Reviews*, 114(23):11636–11682, 2014. PMID: 25390643.

[11] Partha Saha, Moni Kanchan Datta, Oleg I. Velikokhatnyi, Ayyakkannu Manivannan, David Alman, and Prashant N. Kumta. Rechargeable magnesium battery: Current status and key challenges for the future. *Progress in Materials Science*, 66:1–86, 2014.

[12] Lina Zhao, Teng Zhang, Wei Li, Tao Li, Long Zhang, Xiaoguang Zhang, and Zhiyi Wang. Engineering of sodium-ion batteries: Opportunities and challenges. *Engineering*, 2021.

[13] Nuria Tapia-Ruiz, A Robert Armstrong, Hande Alptekin, Marco A Amores, Heather Au, Jerry Barker, Rebecca Boston, William R Brant, Jake M Brittain, Yue Chen, et al. 2021 roadmap for sodium-ion batteries. *Journal of Physics: Energy*, 3(3):031503, 2021.

[14] Robert Usiskin, Yaxiang Lu, Jelena Popovic, Markas Law, Palani Balaya, Yong-Sheng Hu, and Joachim Maier. Fundamentals, status and promise of sodium-based batteries. *Nature Reviews Materials*, 6(11):1020–1035, June 2021.

[15] Jang-Yeon Hwang, Seung-Taek Myung, and Yang-Kook Sun. Sodium-ion batteries: present and future. *Chemical Society Reviews*, 46:3529–3614, 2017.

[16] The lens - free & open patent and scholarly search. https://www.lens.org/.

[17] Olga Kononova, Haoyan Huo, Tanjin He, Ziqin Rong, Tiago Botari, Wenhao Sun, Vahe Tshitoyan, and Gerbrand Ceder. Text-mined dataset of inorganic materials synthesis recipes. *Scientific data*, 6(1):203, 2019.

[18] Edward Kim, Kevin Huang, Adam Saunders, Andrew McCallum, Gerbrand Ceder, and Elsa Olivetti. Materials synthesis insights from scientific literature via text extraction and machine learning. *Chemistry of Materials*, 29(21):9436–9444, 2017.

[19] Christopher Karpovich, Elton Pan, Zach Jensen, and Elsa Olivetti. Interpretable machine learning enabled inorganic reaction classification and synthesis condition prediction. *Chemistry of Materials*, 35(3):1062–1079, 2023.

[20] Thorben Prein, Elton Pan, Tom Doerr, Elsa Olivetti, and Jennifer Rupp. Mtencoder: A multi-task pretrained transformer encoder for materials representation learning. In *AI for Accelerated Materials Design-NeurIPS 2023 Workshop*, 2023.

[21] R. Maskat, S. M. Shaharudin, Deden Witarsyah, and H. Mahdin. A survey on forms of visualization and tools used in topic modelling. *Journal of Information and Visualization*, 7(2), 2023.

[22] Manish Munikar, Sushil Shakya, and Aakash Shrestha. Fine-grained sentiment classification using bert. In *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, volume 1, pages 1–5. IEEE, 2019.

[23] Zabit Hameed and Begonya Garcia-Zapirain. Sentiment classification using a single-layered bilstm model. *Ieee Access*, 8:73992–74001, 2020.

[24] Andrea Chiorrini, Claudia Diamantini, Alex Mircoli, and Domenico Potena. Emotion and sentiment analysis of tweets using bert. In *EDBT/ICDT Workshops*, volume 3, 2021.

[25] Johannes Daxenberger, Benjamin Schiller, Chris Stahlhut, Erik Kaiser, and Iryna Gurevych. Argumentext: argument classification and clustering in a generalized search scenario. *Datenbank-Spektrum*, 20:115–121, 2020.

[26] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[27] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

[28] Tanishq Gupta, Mohd Zaki, NM Anoop Krishnan, and Mausam. Matscibert: A materials domain language model for text mining and information extraction. *npj Computational Materials*, 8(1):102, 2022.

[29] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[30] Natalie. Creating a gpt. `https://help.openai.com/en/articles/8554397-creating-a-gpt`, 2023. Accessed: [insert date here].

[31] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[32] Cohere api. Retrieved from `[cohere.com]`, [2023]. Accessed 1.3.2023.

[33] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

A Appendix

A.1 Methods: Automated Literature Extraction

A.1.1 Corpus

A comprehensive corpus spanning the SIB-literature was assembled by querying the Lens API [16] using relevant keywords. The acquisition of the papers was made possible through licensing agreements with multiple publishers. Subsequent subdivision into paragraphs was carried out leveraging the group's prior work [18]. This initial corpus was then filtered for cathode materials of interest using several text-based matching rules.

A.1.2 Classifiers

Our pipeline harnessed sentence and phrase-level classifiers to screen and filter papers for pertinent information and underscore key mechanisms as cited by the authors. It embodied two types of classification models:

Sentence Level The initial filtering stage operates at the sentence level where we utilized fine-tuned BERT models to discriminate among three sentence categories: challenge sentences, mitigation sentences, and sentences of neither class. This methodology proved effective in efficiently screening the millions of sentences present in our corpus, subsequently filtering out information of distinct value. Given a sentence s_i, our classifier predicted the probability distribution over the classes c_i, eventually assigning the most suitable label. Our three classes were defined as Challenge Sentences, Mitigation Sentences and Non-Target Sentences..

For the development of our sentence-level screening classifier, we annotated a dataset comprising approximately 2,500 sentences. To expedite the labeling endeavor while securing sufficient quantities of infrequently occurring improvement and challenge sentences, we exploited an in-context learned GPT-esque large-language model for a cycle of weak labeling [32]. Ultimately, selected weak-labeled sentences were supplemented by an equal number of randomly chosen sentences and presented to human annotators, thus assuring expert-level quality. Owing to the presence of highly domain-specific vocabulary, we employed the expertise of battery domain experts for labeling. The main merit of this procedure lies in the cutting of labeling time by augmenting the proportion of scarce challenge and improvement sentences. Moreover, this methodology ensures the inclusion of hard-to-predict sentences that had been incorrectly selected by the weak labeller. Overall sentence diversity is ensured by overcoming potential biases of the weak-labeller by introducing a substantial volume of randomly sampled sentences. We computed the inter-annotator rating to be high with agreements of 80% Cohen's κ among the trio of annotators.

For the development of our classifiers, we commenced by benchmarking a variety of approaches on our dataset, using stratified data splits and hyperparameter optimisation.

Phrase Level The phrases were categorized as mitigation strategies (e.g., "doping with Li"), undesirable material-related outcomes (e.g., "low Mn dissolution"), or performance metrics (e.g., "energy density"). Formally, our task was: given a set of all candidate spans s_i, assign each span to the correct entry in the set of defined entity classes $\mathcal{E}$ using $y_e(s_i) \in \mathcal{E}$. This set of classes encompassed valid and invalid phrases. In the second stage the same spans are then investigated for causal relationships, scrutinizing all potential combinations using $y_r(s_i, s_j) \in \mathcal{R}$.

To assemble a dataset for our task, we curated annotation guidelines and utilize our sentence classifiers to evaluate the sodium-ion battery corpus. We randomly sampled 600 sentences and around $2,000$ phrases, with equal shares of challenge and improvement types for expert annotation. Comparing the inter-annotator agreement scores amongst a subset of 60 sentences, we calculated the average inter-annotator agreement by F1 scores for pairwise comparison. Our examination revealed that the discrepancy is primarily caused by the high complexity leading to divergences in annotators' span delimitation. We found the annotations to remain accurate and preserved the key messages conveyed by the sentences in our dataset. This underscored the high degree of flexibility inherent in our task.

The comparison of inter-annotator agreement and classifier performance demonstrated the outstanding level of capability attained by our model. It effectively exploited sentence context to discern lengthy phrases within the sentences, providing insights into mitigation strategies as well as challenges. Our model successfully discerns even non-trivial relationships.

A.1.3 Database Creation and Accessibility

The core of our methodology was formed by a sequential application of the developed methods. The extraction and presentation of information were facilitated through four main steps. Initially, the acquired publications were scrutinized for the studied cathode active material discussed, which includes layered metal oxides, polyanionic compounds, and Prussian blue analogues. We utilized text-guided rules to match elemental formulas and

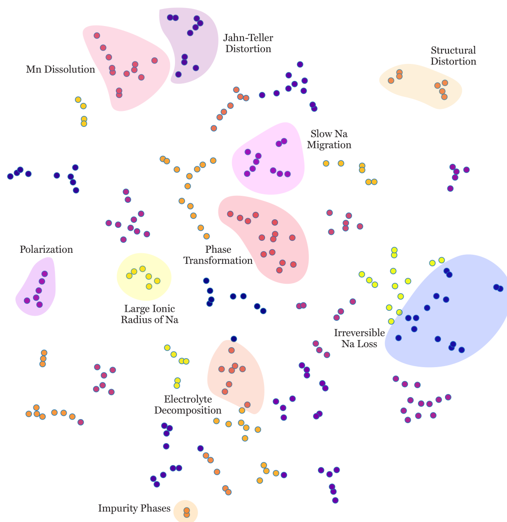

Figure 6: Challenge sentences space extracted using our NLP framework from the sodium-ion batteries literature on NFMO (sodium iron manganese oxide) cathodes with labelled approximate regions of prominent micro-challenges.

	F1	P	R
MATSCIBERT	83.1 (1.2)	83.4 (1.1)	83.8 (1.3)
SCIBERT	**84.1 (1.7)**	**84.4 (1.5)**	**84.2 (1.4)**
SENTENCE BERT	79.2 (1.8)	79.3 (2.0)	79.2 (1.8)
GPT3 @10 SHOTS	73.2	75.9	72.5
TF-IDF	70.2	70.1	70.5
RANDOM	51.4	51.2	51.8

	Sentences	Entities	Relations
CHALLENGE	84.1	67.5	39.4
IMPROVEMENT	83.1	67.9	50.8

Table 2: Model Comparison on the Sentence Dataset (Left) and Results for the Open Information Extraction Task (Right).

delineate materials by composition. Subsequently, the papers were introduced to our sentence classifier to discern sentences related to challenges and mitigation strategies. Following this, in the phrase and relation identification phase, we highlighted sections in the literature of particular interest. Lastly, we visualized the database of identified challenges and mitigation strategies by encoding each sentence using sentence BERT models [31, 4].

By applying this methodology to our corpus of approximately 2200 papers on selected cathode chemistries, we obtained a database of 31,000 challenge and mitigation sentences. Out of these, our classifiers identified 9,000 relations. Analyzing the diversity of papers in the final relational database, we noted that 91% of the papers are represented in the final improvement, and 82% in the challenge database, underscoring the comprehensive coverage of our source material. To evaluate the accuracy of our database, we randomly selected 200 entries, equally distributed between challenge and mitigation strategy entries. Two domain experts then assessed these entries for their accuracy and completeness. Our findings indicated very good overall correctness exceeding 90% in our database. Enabling usage of our data for further analysis.

To categorize the identified mitigation and challenge mentions, we utilized BERT-based clustering methods to visualize the extracted phenomena. We collated a dataset of 1100 root causes related to materials, which led to key performance degradation. These root cause or "mirco-challenges" were carried and of several types: elemental phenomena (e.g., electronic conductivity), structural phenomena (e.g., secondary phase formation), morphological phenomena (e.g., surface impurities), and key performance metrics (e.g., energy density). Subsequently, we assessed various clustering methods based on their cluster purity, which evaluated the model's ability to group text mentions of related phenomena closely together, modeling topics of interest. The most successful results were attained when phrases were embedded using Sentence-BERT models [31], which yielded the highest cluster purity values. We thus incorporated this approach into our methodology and introduced an additional visualization step. Utilizing the UMAP [33] algorithm, we condensed the dimensionality into 2D space, generating an interactive map of the curated database. Here, relevant strategies were grouped together, providing a conducive platform for comprehensive exploration of pairs of challenges and mitigation strategies. It further facilitated their selection for evaluation of large-scale manufacturing feasibility. Beyond its application in the realm of SIBs, our methodology was successfully transposed to the field of LIBs. Assessing database accuracy demonstrates a tolerable reduction in performance, with the correctness of database entries dropping to 90%. This successful domain shift paves the way for us to extend the array of investigated mitigation strategies to those strategies reported for LIBs, which have already seen industrial-scale adoption.

A.2 Case Study Identification and Evaluation

We can use our developed methodology for downstream tasks to identify and evaluate case studies and combine it with a process-based cost modeling for scalability inspection as shown in Figure 7.

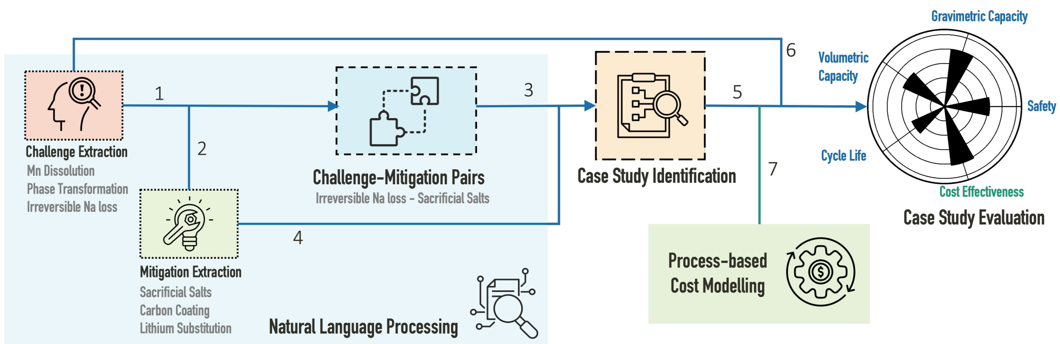

Figure 7: **Combining NLP techniques with Process-based Cost Modeling**: Extracted challenges and mitigation strategies were mapped to constitute challenge-mitigation pairs (Arrows 1 and 2). These challenge-mitigation pairs represent individual cases that can be further evaluated (Arrow 3). To get a broader overview of all represented strategies, we utilized the visualized space of mitigation strategies (Figure 3 (a)) to downselect challenge-mitigation case studies (Arrow 4) and create a database of sentences of interest and corresponding DOIs. After the identification of the case studies (Arrow 5), we used the extracted challenges database of sentences and corresponding DOIs (Arrow 6) along with a cost model built in-house for sodium-ion battery cathode materials (Arrow 7) to quantitatively assess the scalability barriers. The blue parameters were obtained via NLP and the green through cost modeling.