LLMs and MI Bring Innovation to Material Development Platforms

OBUCHI Kiichi, FUNAYA Koichi, TOYAMA Kiyohiko, TANAKA Shukichi, ONORO RUBIO Daniel, MIN, Martin Renqiang, MELVIN Iain

Abstract

In this paper, we introduce efforts to apply large language models (LLMs) to the field of material development. NEC is advancing the development of a material development platform. By applying core technologies corresponding to two material development steps, namely investigation activities (Read paper/patent) and experimental planning (Design Experiment Plan), the platform organizes documents such as papers and reports as well as data such as experimental results and then presents in an interactive way to users. In addition, with techniques that reflect physical and chemical principles into machine learning models, AI can learn even with limited data and accurately predict material properties. Through this platform, we aim to achieve the seamless integration of materials informatics (MI) with a vast body of industry literature and knowledge, thereby bringing innovation to the material development process.

Keywords
material area-specific LLM, knowledge graph, ontology, information extraction technology, Materials Informatics (MI), PIML (Physics-Informed Machine Learning), AI model for generating molecular structures

1. Introduction

Amidst the demand for enhanced performance and diversification of needs in industrial materials as well as environmental considerations, efficient material development is required. Currently, material developers primarily engage in two tasks: (1) investigations involving literature, patents, internal technical reports, and customer needs, and (2) prototyping involving repeated experiments based on experimental planning. However, it is challenging to extract the desired information from vast amounts of data and to find the desired material designs within a vast search space. This inevitably leads to wasting time.

For example, in typical material design cases, vast amounts of historical internal documents remain unused, hindering effective investigation. In addition, in prototyping, the lack of large-scale open databases and the high hurdles associated with simulations and high-throughput experiments (experiment automation) pose challenges to obtaining sufficient data. NEC is working to develop a materials development platform that addresses these issues by applying generative AI technology such as large language models (LLMs) to the field of materials development. In this article, we report on the application of our material development platform to the development of polymeric materials, which are used in various aspects of our daily lives, among others,

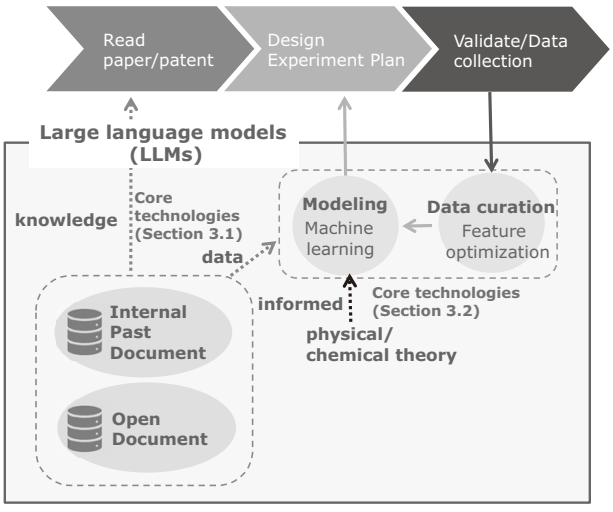

Fig. 1 Material development platform.

and there is a need to rapidly develop materials that fulfill various functions.

First, as shown in **Fig. 1**, by using this platform, it becomes possible to utilize previously unused internal data as well as external documents such as papers, patents, and open databases. An LLM uses proprietary technology to efficiently learn about materials from the documents, and this knowledge is then used to support investigations or applied to predict material properties and experimental planning using NEC's proprietary modeling techniques.

In addition, through the chat-based approach of LLMs, the barrier to utilizing AI is significantly lowered. More specifically, by introducing an easy-to-use method like chats, we eliminate the need for programming skills and proficiency in tools that have been barriers in previous AI applications. The aims are to enable more material developers, including those who may have reservations about digital technologies, to focus their time on their main goal of material development and to further improve the sophistication of methods with AI integrated into the material development platform.

In this paper, we first provide an overview of this material development platform, followed by an explanation of the core AI technologies integrated into it. Finally, as an actual application case, we introduce a case involving bioplastics.

2. What is the Material Development Platform?

Fig. 2 outlines the steps involved in a material development platform.

In our material development platform, the aim is to improve efficiency by reducing the labor and time required to develop materials in the materials industry. To achieve this, advanced technologies such as generative AI are used to streamline the tasks at each stage of the workflow shown in Fig. 1 as follows.

First, at the stage of investigation, papers and reports as well as data such as experimental results are organized and presented in an interactive way to users so that it is easier to search for and narrow down molecular structure candidates for materials to be developed from a vast amount of information sources.

Next, at the stage of experimental planning, the platform provides a mechanism that allows for learning even with minimal data and for accurate prediction of material properties. As mentioned later, to achieve this mechanism, we adopted a method that reflects physical and chemical principles into machine learning models. By utilizing machine learning models developed in this way, it becomes possible to identify molecular structures that have a high potential of realizing the required material properties. This helps streamline the traditionally labor-intensive stage of experimental evaluation by narrowing down the molecular structures for experimental validation.

In section 3, we will introduce the core technologies that are essential for realizing such a material development platform.

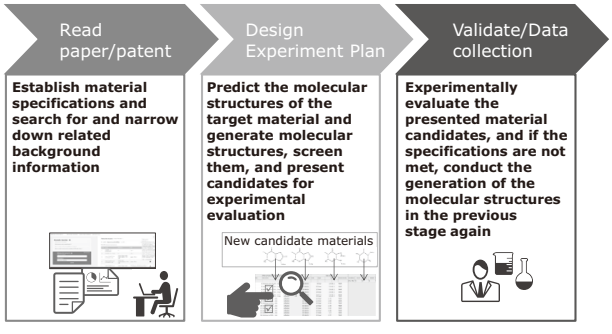

Fig. 2 Outline of the material development platform.

3. Core Technologies in the Material Development Platform

3.1 Searching for and narrowing down background information on material candidates

The initial task in the material development platform involves gathering a wide range of background information related to candidate materials and narrowing it down while establishing the policies for development work and assessing the feasibility of material candidates. This corresponds to the investigation activities in Fig. 2.

In the material development platform, this task is achieved through the following steps:

1) Automatically importing material information from information sources
2) Searching for and narrowing down related information using a dialogue interface.

In step 1, information regarding material characteristics and manufacturing methods is extracted from accessible information sources such as papers, internal reports, and experimental data. This information is then stored in a database in the format of a so-called knowledge graph. A knowledge graph is a data structure that consists of tuples (= data items) composed of a subject, a predicate, and an object[1)2)]. To make it easy to use in the materials industry, the material development platform defines each element (subject, predicate, and object) using a standard ontology of materials science. By defining the relationships between elements in the knowledge graph schema,

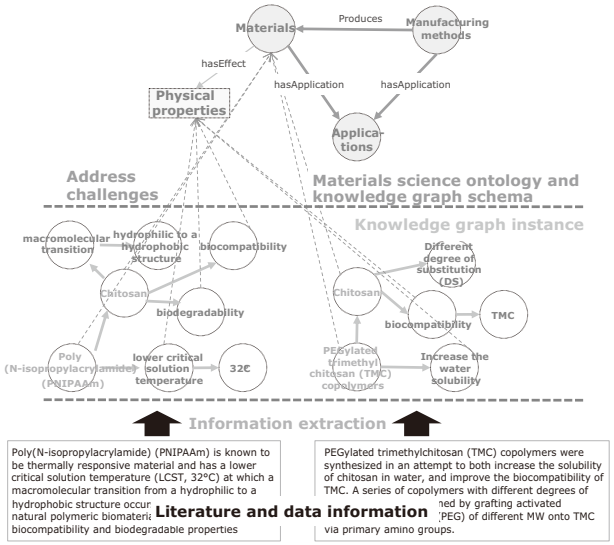

Fig. 3 Details of Step 1.

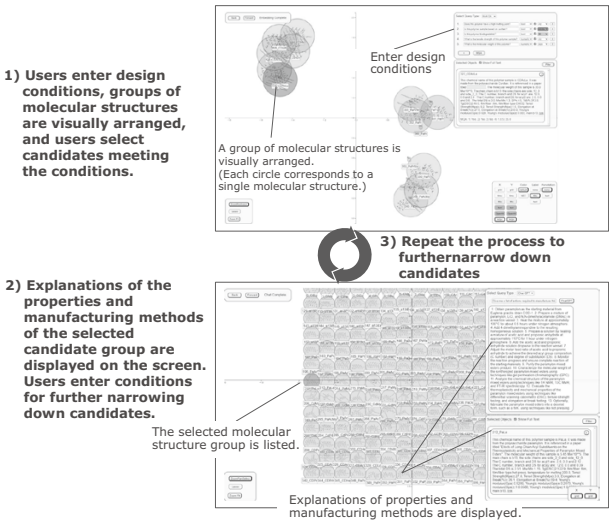

Fig. 4 Details of Step 2.

the platform ensures the clarity and reproducibility required for material development tasks.

Furthermore, by performing deep learning on the above-mentioned ontology and knowledge graph schema to train the LLMs, we have developed a method to directly extract knowledge graphs from literature without requiring complex language processing pipelines. This allows for the accurate and fast extraction of information from vast amounts of literature with simple procedures.

By utilizing this extracted knowledge graph, it is not only possible to perform detailed searches for background information but also to greatly reduce the effort required to prepare training data for generative AI. This reduces the effort required to train predictive AI to predict material properties or to train generative AI to generate molecular structures that achieve the desired material properties (**Fig. 3**).

In step 2, a mechanism for interactively narrowing down candidate molecular structures is realized by training the inference engine on LLMs with a combination of the knowledge graph acquired from step 1 and the literature. The interactive interface enables users to easily enter multiple design conditions crucial for material design to narrow down the candidate group of molecular structures through a simple operation of selection and elimination. For example, users can use natural language or other means to set the conditions related to the mechanical, thermal, and electrical properties as well as the manufacturing methods of the polymer they want to develop. The interface then arranges each candidate molecular structure on the screen based on its compatibility with the conditions, enabling users to visually select molecular structures that seem suitable for their objectives. As users make selections, explanations of the candidate's properties and manufacturing methods are displayed, so users can confirm the appropriateness of their selection (**Fig. 4**).

3.2 Polymer-based machine learning models and molecular generation technology

Materials informatics (MI) has rapidly developed in recent years as a field that utilizes information science such as AI to support material development. However, organic polymers have been said to be one of the most difficult materials to apply machine learning to compared to inorganic materials, where crystal structures have a deep relationship with physical properties, and organic small molecules, whose chemical structures are easily representable. This is because there are technological challenges such as representing long entangled polymer chains, understanding the interactions between fillers or additives in mixtures, and the cross-linking in rubber materials that cannot be addressed simply by applying advanced machine learning architects.

Therefore, NEC is developing technology to solve these challenges with the physics-informed machine learning (PIML) approach, which incorporates the physical laws of polymers into machine learning models. Specifically, we are developing techniques that use values obtained from the principles of physics and chemistry as the feature values input to machine learning and technologies that incorporate theoretical formulas that express

polymer-specific characteristics into machine learning architects. Efforts are also being made to generate the molecules themselves and are advancing our technological development while pursuing the strengths of both rule-based models based on specialized knowledge of polymers and model-based approaches that apply generative AI.

In this way, by using PIML and molecular generation technology, we are developing innovative technologies that can break through the current situation where machine learning models do not function due to the lack of data in vast search spaces.

4. Case of Application to Bioplastic Development

Plastic is indispensable in our daily lives at present, but there are challenges such as reducing CO_2 emissions and addressing the issue of marine plastic waste. NEC has been proactively addressing these challenges and has been developing bioplastics since the early 2000s with a record of commercialization (**Photo**)[3]. With increasing awareness of environmental issues, further acceleration of development is expected in the future.

Thus, we are proceeding with demonstrations by applying the material development platform. When we attempted to explore new materials for bioplastics in a demonstration experiment, it was not possible to predict the material properties when conventional materials informatics (MI) was applied. In contrast, by applying the core technologies mentioned in section 3.2, a level of property prediction accuracy that is sufficient for practical use was achieved. Also, as mentioned in section 3.1, using LLMs and papers on bioplastics, we obtained examples of identifying candidate molecular structures based on raw material names, strength, biodegradability, and other factors, and also obtained detailed descriptions of their manufacturing methods. By effectively utilizing these core technologies in this way, we aim to efficiently develop new bioplastics and contribute to the creation of a sustainable society.

Photo New bioplastic products made by NEC.

5. Conclusion

In this paper, we introduced a material development platform as an approach to applying LLMs in the field of material development. In the future, for the materials industry, which will needs to respond even faster to customers' demands in the future, NEC aims to bring innovation to material development using AI consisting of these two core technologies.

References

1) Bhushan Kotnis, Kiril Gashteovski, Daniel Rubio, Ammar Shaker, Vanesa Rodriguez-Tembras, Makoto Takamoto, Mathias Niepert, and Carolin Lawrence: MILIE: Modular & Iterative Multilingual Open Information Extraction, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp.6939-6950, 2022. https://aclanthology.org/2022.acl-long.478/
2) Gorjan Radevski, Kiril Gashteovski, Chia-Chien Hung, Carolin Lawrence, and Goran Glavas: Linking Surface Facts to Large-Scale Knowledge Graphs, CoRR 2023. https://arxiv.org/abs/2310.14909
3) NEC: NeCycle®, the bioplastics developed by NEC (Japanese) https://www.necplatforms.co.jp/product/necycle/index.html

Authors' Profiles

OBUCHI Kiichi
Assistant Manager
Secure System Platform Research Laboratories

FUNAYA Koichi
Chief Technology Architect
NEC Laboratories Europe

TOYAMA Kiyohiko
Lead Research Engineer
Secure System Platform Research Laboratories

TANAKA Shukichi
Director
Secure System Platform Research Laboratories

ONORO RUBIO Daniel
Senior Research Engineer
NEC Laboratories Europe

MIN, Martin Renqiang
Department Head
NEC Laboratories America

MELVIN Iain
Researcher
NEC Laboratories America

The details about this paper can be seen at the following.

Related URL:

Discussion on future R&D demands of research companies learning from pioneers at the edge of forefront of data utlization (Japanese)
https://events.nikkeibp.co.jp/event/2023/bpi230706/

Information about the NEC Technical Journal

Thank you for reading the paper.
If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese

English

Vol.17 No.2 Special Issue on Revolutionizing Business Practices with Generative AI
– Advancing the Societal Adoption of AI with the Support of Generative AI Technologies

Remarks for Special Issue on Revolutionizing Business Practices with Generative AI
Approaches to Generative AI Technology: From Foundational Technologies to Application Development and Guideline Creation

Vol.17 No.2
June 2024

Special Issue TOP

Papers for Special Issue

Market Application of Rapidly Spreading Generative AI
NEC Innovation Day 2023: NEC's Generative AI Initiatives
Streamlining Doctors' Work by Assisting with Medical Recording and Documentation
Using Video Recognition AI x LLM to Automate the Creation of Reports
Understanding of Behaviors in Real World through Video Analysis and Generative AI
Automated Generation of Cyber Threat Intelligence
NEC Generative AI Service (NGS) Promoting Internal Use of Generative AI
Utilization of Generative AI for Software and System Development
LLMs and MI Bring Innovation to Material Development Platforms
Disaster Damage Assessment Using LLMs and Image Analysis

Fundamental Technologies that Enhance the Potential of Generative AI
NEC's LLM with Superior Japanese Language Proficiency
NEC's AI Supercomputer: One of the Largest in Japan to Support Generative AI
Towards Safer Large Language Models (LLMs)
Federated Learning Technology that Enables Collaboration While Keeping Data Confidential and its Applicability to LLMs
Large Language Models (LLMs) Enable Few-Shot Clustering
Knowledge-enhanced Prompt Learning for Open-domain Commonsense Reasoning
Foundational Vision-LLM for AI Linkage and Orchestration
Optimizing LLM API usage costs with novel query-aware reduction of relevant enterprise data

For AI Technology to Penetrate Society
Movements in AI Standardization and Rule Making and NEC Initiatives
NEC's Initiatives on AI Governance toward Respecting Human Rights
Case Study of Human Resources Development for AI Risk Management Using RCModel

NEC Information

2023 C&C Prize Ceremony