RESEARCH ARTICLE

ADVANCED INTELLIGENT SYSTEMS Open Access

www.advintellsyst.com

Forecasting Research Trends Using Knowledge Graphs and Large Language Models

Maciej Tomczak, Yang Jeong Park, Chia-Wei Hsu, Payden Brown, Dario Massa, Piotr Sankowski, Ju Li,* and Stefanos Papanikolaou**

Since ancient times, oracles (e.g., Delphi) has the ability to provide useful visions of where the society is headed, based on key event correlations and educated guesses. Currently, foundation models are able to distill and analyze enormous text-based data that can be used to understand where societal components are headed in the future. This work investigates the use of three large language models (LLM) and their ability to aid the research of nuclear materials. Using a large dataset of *Journal of Nuclear Materials* papers spanning from 2001 to 2021, models are evaluated and compared with perplexity, similarity of output, and knowledge graph metrics such as shortest path length. Models are compared to the highest performer, OpenAI's GPT-3.5. LLM-generated knowledge graphs with more than 2×10^5 nodes and 3.3×10^5 links are analyzed per publication year, and temporal tracking leads to the identification of criteria for publication innovation, controversy, influence, and future research trends.

1. Introduction

The key to making successful predictions has always been the accuracy and rigidity of the required input data and a well-defined mechanism for producing predictions. Input data, such as microscopy images or tabular data, have been the major component of recent successful applications in machine learning. In materials science, it has been known for decades that structure-process-property relationships represent the key to any predictions of material properties. More generally, science in all of its forms, advances in time through empirical, in-context relations ("gentle" vs. "strong" pot stirring), which eventually evolve into actual mathematical relationships after many reincarnations in the scientific literature. Thus, it is natural to hypothesize that novel concepts and relationships exist, in nascent form, in current publications, in fuzzy, unstructured manners. The importance of identifying these emerging concepts and relationships cannot be overstated. In this work, we demonstrate how large language models (LLMs) can be used to develop a predictive tool for such identification, in the context of nuclear materials research, as it is expressed in the publications of the *Journal of Nuclear Materials* across the last two decades.

LLMs have recently revolutionized the understanding and generation of text that is produced by the use of naturally utilized languages, such as human-spoken or/and coding.[1–3] A key focus is the application of LLMs on scientific literature analysis.[4] A number of benchmarks and criteria have been developed for the assessment of LLM proficiency in a particular knowledge domain,[5,6] such as memorization, comprehension, and reasoning in scientific concepts. However, the understanding of higher-level proficiency of LLMs in scientific domains requires the development of tools that characterize the whole concept structure of the model.

Measures for the quality of LLMs are promoted through benchmarks for assessing their capabilities. General benchmarks can be used for models' world knowledge across diverse domains.[5] Additionally, one could use benchmarks for assessing the alignment with factual information.[7] However, in science subjects, such benchmarks are clearly insufficient for the understanding and reasoning of scientific literature. In this context, recent models have focused on text-mining and integrating

M. Tomczak, D. Massa, S. Papanikolaou
NOMATEN Centre of Excellence
National Center for Nuclear Research
A. Sołtana 7, Swierk/Otwock 05-400, Poland
E-mail: stefanos.papanikolaou@ncbj.gov.pl

Y. J. Park, C.-W. Hsu, P. Brown, J. Li, S. Papanikolaou
Department of Nuclear Science and Engineering
Massachusetts Institute of Technology
7 Massachusetts Avenue, Cambridge 02139, MA, USA
E-mail: yangjeongpark@kist.re.kr; liju@mit.edu

Y. J. Park
Computational Science Research Center
Korea Institute of Science and Technology
Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea

D. Massa
IDEAS-NCBR
Chmielna 69, Warsaw 00-801, Poland

P. Sankowski
University of Warsaw
Stefana Banacha 2, Warsaw 02-097, Poland

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202401124.

DOI: 10.1002/aisy.202401124

information across multiple documents.[8,9] Here, we focus on an alternative approach, in terms of creating benchmarks of complete knowledge graphs (KGs) constructed by LLMs on a large sequence of scientific publications, that may outperform traditional benchmarks in terms of multimodal content and overall contextual interpretation.

KGs, as opposed to word clouds (**Figure 1**), are concrete, structured, directional networks that represent data in a graphical form, especially events and entities as nodes, and the relationships between them as edges. While the importance of such a network is critical in domain knowledge understanding, its construction has been costly and time-consuming. LLMs have accelerated such KG constructions significantly,[10,11] providing insights to the context of scientifically written paragraphs. In this work, we develop KGs of more than 10 000 publications in *Journal of Nuclear Materials* (JNM) and then analyze such complex network properties and features.

The focus of this work is the investigation of domain knowledge KGs, viewed in physics as complex networks, and their statistical properties. The investigation of complex networks using statistical physics has led to major discoveries in the last two decades[12,13] on the content and dynamics of networks such as cities or the internet. In this article, we investigate in detail the temporal evolution of KG networks as they form by gradually including text of a particular journal, JNM. We use three LLMs, pretrained in dialog tasks, for the construction of the KGs, and then investigate the network structure with time. Using GPT performance as a benchmark, we first compare three different LLMs using both standard measures and statistical differences of produced KGs features, such as the degree and rank, number of nodes and internodal shortest path (SP) length, and numerical distance in the word embeddings space (Word2Vec).[14] Then, we identify the semantic similarity distance distributions between the most central KG nodes and all other nodes, and we capture the exponential form, both for mutually "common" and "irrelevant" concept pairs. These two distributions appear to be universal across concepts. We also develop an index formed by the number of new nodes introduced by a newly added manuscript and the citations it receives in the future that can identify in a unique way its innovative or controversial character. Finally, we investigate the SP between distinct popular nodes/concepts in the network, and we investigate its temporal evolution. We identify a constitutive exponential law that characterizes the temporal evolution of internodal SPs, and through this pathway, we develop a predictive tool for future interweaving of interdisciplinary research topics.

2. Experimental Section

2.1. Knowledge-Graph Construction

2.1.1. GPT-3.5-Turbo

GPT-3.5-turbo[1] represents a significant evolution in the Generative Pretrained Transformer series.[15,16] This model was distinguished by its enhanced efficiency and scalability in generating human-like text. The training process of GPT-3.5-turbo involves unsupervised learning, where the model predicts the next token in a sequence, followed by supervised fine-tuning on specific tasks. Using reinforcement learning from human feedback,[17] the model was fine-tuned to follow instructions.

2.1.2. Llama-2-7b-Chat-Hf

Llama-2-7b-chat-hf,[18] a product of Meta AI's research, was another advanced language model that excels in conversational AI tasks. The architectural foundation of Llama-2-7b-chat-hf was also based on the transformer model, benefiting from extensive pretraining on conversational datasets.

2.1.3. Mixtral-8×7B-v0.1

Mixtral-8×7B-v0.1[19] was a state-of-the-art LLM developed by Mistral AI. It was designed as a Sparse Mixture of Experts (SMoE), utilizing an architecture where each layer contains multiple experts, specifically eight feedforward blocks. During inference, a router network selects two experts per layer to process the input, making the model efficient and powerful.

We use abbreviations GPT, LLaMA2, and Mixtral to note the above models. Additionally, we tested a fine-tuned version of Mixtral on our data, which we label as Mixtral-FT. We developed all the code using Python libraries, including OpenAI API and open-source NLP libraries such as HuggingFace.[20]

We focus on the journal JNM, which includes in a self-contained manner the evolution of research in the field of nuclear materials over the two preceding decades. After downloading all the literature published in JNM from 2003 to 2022, each article was split into chunks with 1,000 words for training. The open-source models are fine-tuned for text reproduction task

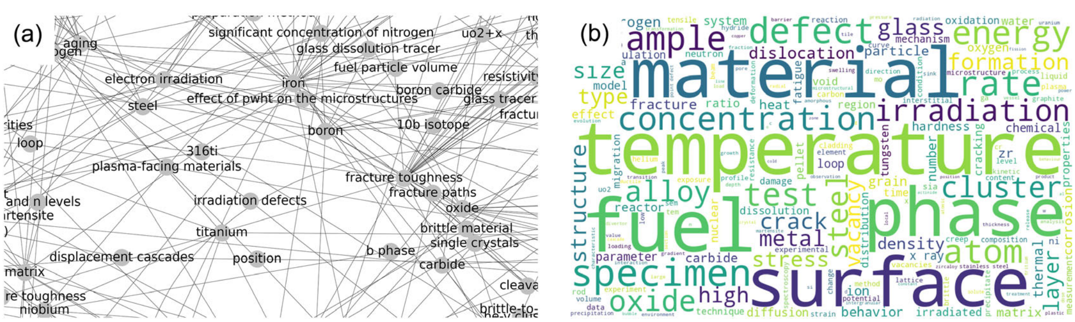

Figure 1. An example of a) a generated knowledge graph and b) a word cloud from literature published in the Journal of Nuclear Materials.

on the divided text. It is important to stress that the dialog tasks for which the models have been pretrained include, for example, text reproduction and question-answer pairs. Here we test a simple fine-tuning solution of such models on JNM content, which does not cover question-answer. We then test transferrability improvements to zero-shot KGs generation, a capability inherited by their pretraining on instruction following but for which the models are not specifically fine-tuned on in the context of this work. After a first evaluation of performance metrics on the pretrained and fine-tuned LLaMA and Mixtral models, selected models are implemented in KG generation from the provided paragraph chunks. In the prompt (Figure A9), models are asked to extract concepts in the text and describe connections between those concepts. Detailed analysis of the resulting KGs allows us to assess the usefulness of perplexity as a measure to estimate both the model capability to reproduce the text and to extract its essential components. For the GPT case, the perplexity could not be computed due to the limited access to the model through the API.

The text data was produced by using the Elsevier API. Elsevier API is available under license for noncommercial and research purposes (https://dev.elsevier.com/). Full texts can be accessed from the Science Direct database upon a call through the API. A variable number of papers was accessed in JNM from 2003 to 2022 (see Figure A1 of the Appendix), and they are sequentially extracted into KGs per year.

2.2. Evaluation Metrics and Statistical Procedures

To assess both language-model fidelity and the faithfulness of extracted KGs, we employ four complementary groups of metrics, outlined below.

2.2.1. Language-Model Quality

Perplexity: Given a token sequence $x_{1..T}$, the perplexity is defined by

$$\text{perplexity} \equiv \left(\prod_{t=1}^{T} P_{\text{LM}}(x_t | x_{<t})^{-1} \right)^{1/T} \quad (1)$$

where T is the total number of words, P_{LM} is the probability function of the language model. A lower value indicates higher likelihood (and thus better fit) of the held-out JNM text. Perplexity for GPT-3.5-turbo was unavailable because the OpenAI API does not expose token-wise log-probs; for LLaMA-2 and Mixtral variants, we compute it with the HuggingFace perplexity script on the same 1,000-word chunks used during fine-tuning.

2.2.2. Knowledge-Graph Topology

Degree Distribution and Rank Plot: We treat each concept node as unweighted and directed and compute the out-degree k. Rank–degree plots were fitted with the discrete power-law $P(k) \propto k^{-\gamma}$ using Clauset's MLE routine, revealing the expected scale-free behavior for GPT-3.5.

Table 1. Values of fitted parameters for GPT-3.5-turbo.

#	A	μ	σ
1	156.21	300.30	0.97
2	135.66	320.55	22.01

SP Length: For every yearly KG, we calculate the unweighted, directed SP between all reachable node pairs via Dijkstra's algorithm as implemented in NetworkX.[21]

Graph Spectrum: Real eigenvalues of the adjacency matrix were obtained with SciPy.[22] The multiplicity of low-modulus eigenvalues acts as a proxy for redundancy—pronounced spikes denote simpler, more repetitive KGs.

2.2.3. Embedding-Space Semantics

We train a 200-dimensional word2vec model (skip-gram, window = 5, min-count = 3) on the full JNM corpus. Euclidean distances between the top-5 central nodes and every other node produce bimodal histograms, each fitted with the sum of two Gaussians (Equation (1) and **Table 1**). These fits separate mutually common from mutually irrelevant concept pairs.

2.2.4. Temporal Impact Indices

Innovation–Controversy Quadrant: For each manuscript we log (i) the number of new nodes it adds to the cumulative KG and (ii) citations accrued up to the 2024-12 Scopus snapshot. The scatter is partitioned into four empirically chosen quadrants, enabling rapid classification of a paper's community impact.

Shortest-Path Decay Model: Let t be years since the two nodes first become connected. We fit

$$\text{SP}(t) = a \exp(-bt) + c \quad (2)$$

with 'curve_fit' (SciPy) over all node pairs whose manuscripts jointly exceed 200 citations.

3. Results

3.1. LLM-Derived KGs

First, we investigate the perplexity of LLMs (see Equation (1)), a standard measure for the characterization of their associated uncertainty in the predictions. As seen in **Figure 2**, both LLaMA and Mixtral models show lower perplexity values in their fine-tuned states. However, the pretrained version of Mixtral model seems already very close to its optimal fine-tuned state.

We investigate basic structural features of the final KGs, being the cumulative product of years 2003–2022 (Figure 4). During the production of KGs, the LLaMA-FT model was found not capable of reproducing meaningful KGs. The reason for such an outcome should be matter of further investigation: while one could argue it to be found in the fine-tuning procedure, which, as explained above, does not focus on KGs generation, it was interesting to notice how the fine-tuned Mixtral model succeeds in the

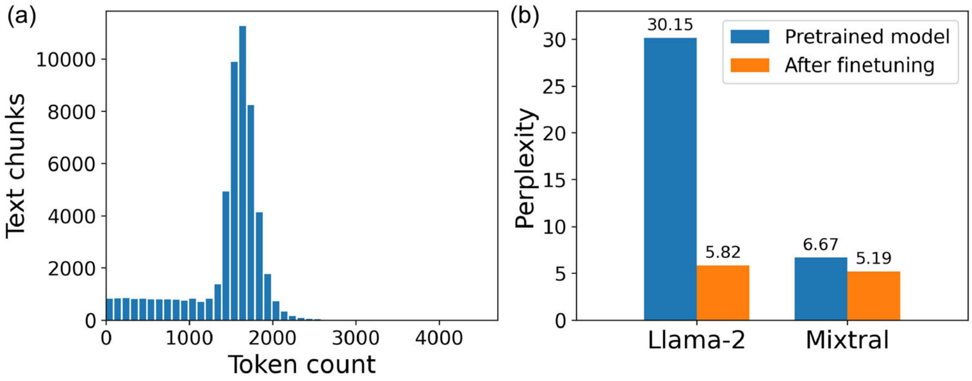

Figure 2. Article length in the dataset and perplexity of the utilized LLMs in this work: a) Token count distribution in processed text chunks and b) Perplexity values of pretrained (left bars) and fine-tuned(right bars) models.

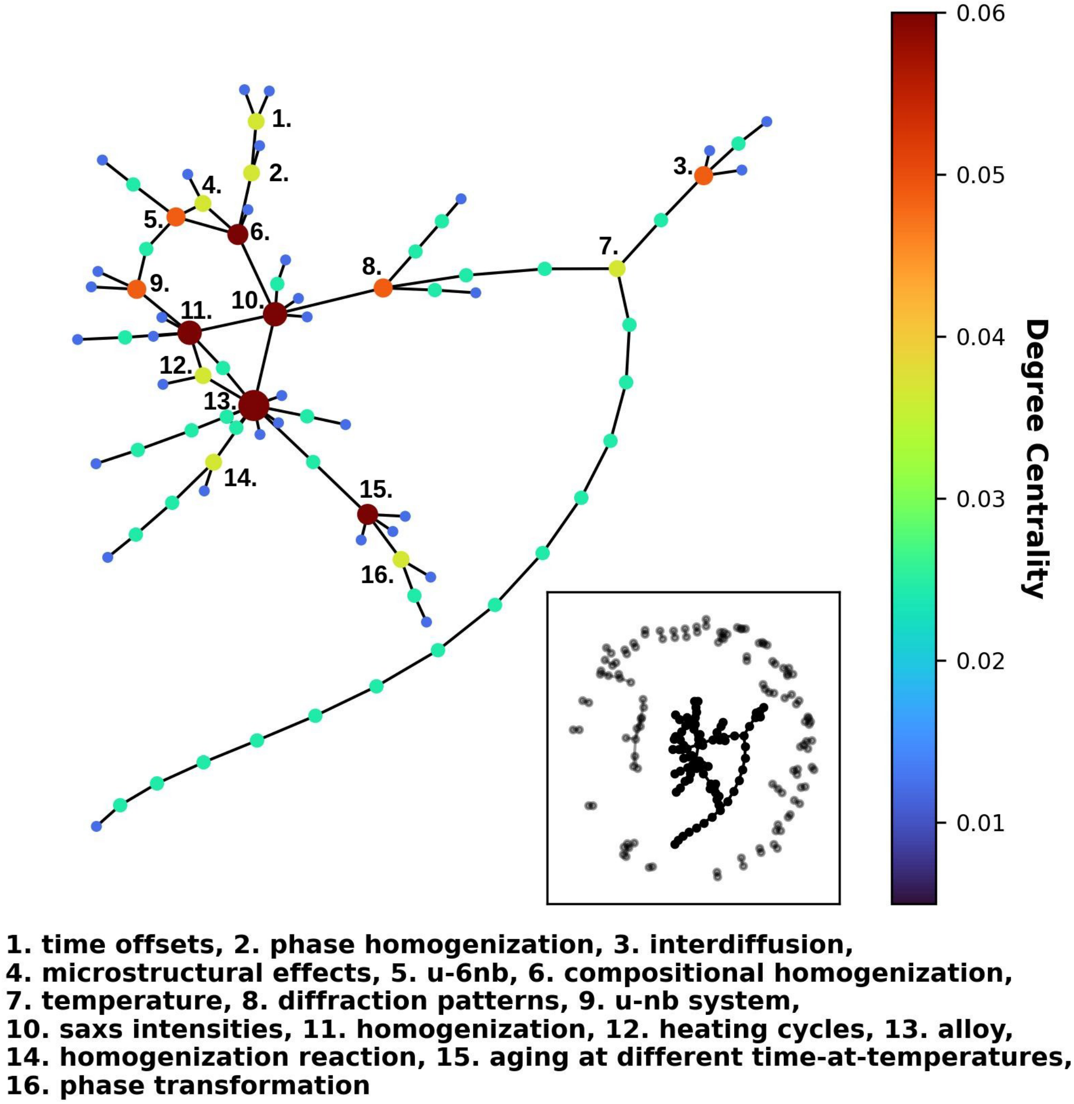

Figure 3. Biggest component of a GPT graph made from a single paper and degree of centrality of the main nodes. The whole article graph is shown in the lower-right inset.

task. An example of generated KG with GPT model for a single article from the database is shown in **Figure 3**, focusing on the biggest graph component and the centrality degree of its nodes. It was possible to notice how the model highlights and connects physically meaningful concepts: the highest centrality belongs to the most general concept of *alloy*, which then connects to more specific experimental characterization techniques and materials science concepts. Moreover, the smaller components of the graph were distributed as small disconnected satellites, implying that the model was capable of identifying a core structure in the article.

In the Appendix Figure A2–A6, we also report a detailed example of progressive building of KGs from text. In **Figure 4**, the degree distributions of the KGs display a power-law structure, with GPT3.5 signifying a scale-free property. While Mixtral significantly approaches the GPT baseline in both its pretrained and fine-tuned versions, the LLaMA model stands as an outlier, showing higher degree and rank, which implies higher number of nodes and shorter paths in the generated KGs, signatures of possible worse building of connections. A similar shape in the path lengths (Figure 4c) and Word2Vec space distances (Figure 4d) histograms suggests that the constructed KGs are encoding meaningful information.

3.2. Spectral Signatures of Redundancy in Extracted Graphs

We also investigate the graphs spectra from the different KGs, another standard measure of network structure characterization **Figure 5**. The presence of very pronounced peaks at a specific

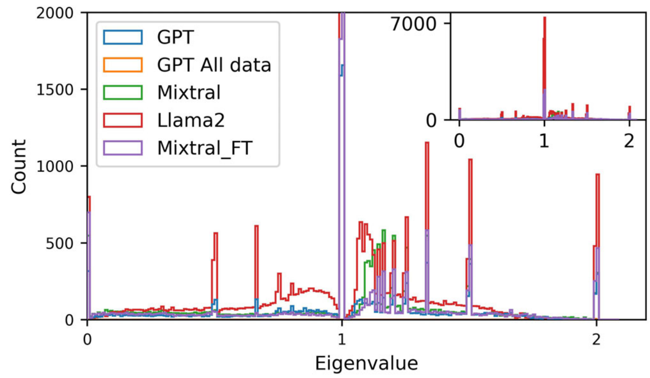

Figure 5. Graph Spectrum: histogram of real eigenvalues for GPT, LlaMA2, Mixtral, and Mixtral-FT.

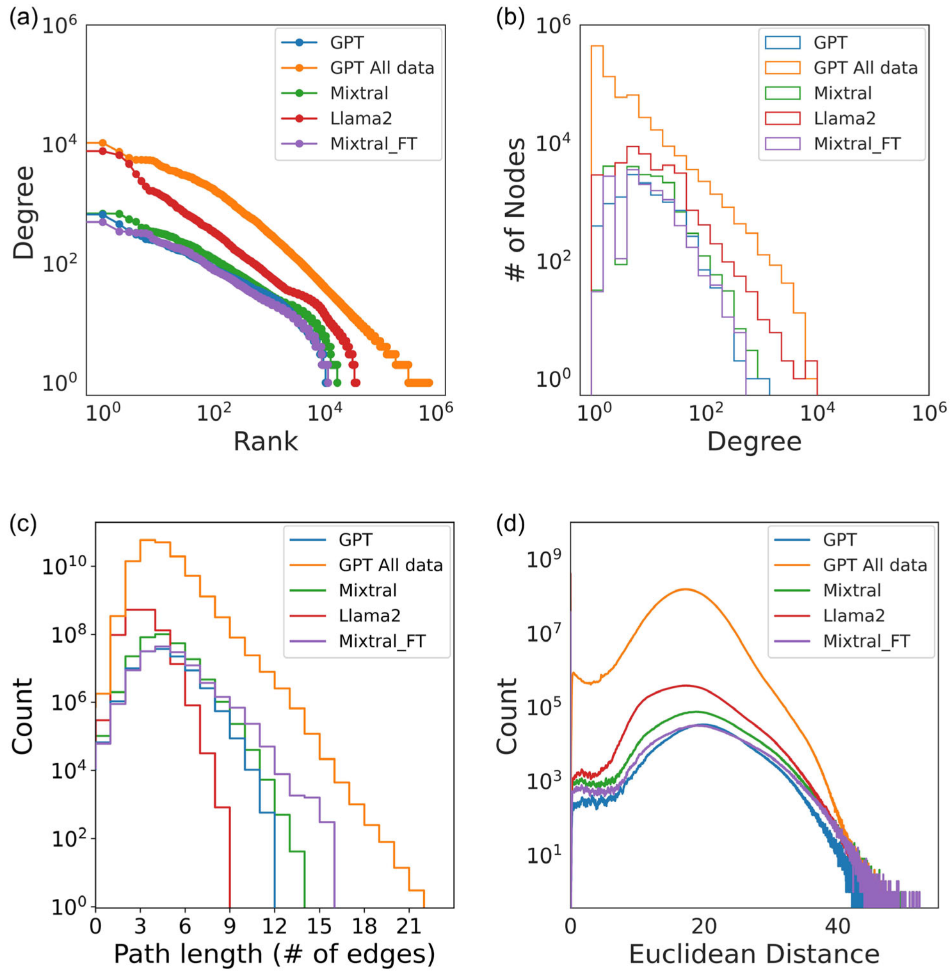

Figure 4. Comparison of graphs from different LLMs: a) Degree rank plot, b) Degree histogram, c) Path lengths histogram, and d) Euclidean distances in Word2Vec space histogram.

small set of eigenvalues represents a signature of more redundant and simple extracted graphs' structures. The majority of counts is found for the central eigenvalue 1, with LlaMA displaying the largest differences among the models, but a pattern similarity is present in the comparison.

3.3. Estimation of the Publication Impact

Then, we proceed to identify predictive tools, based on these KGs. For this purpose, we first focus on the most central nodes of the graph and investigate the semantic similarity to all other nodes. In **Figure 6**, we report an example of similarity analysis in word2vec space with GPT, involving concepts which include: temperature (a state parameter), irradiation (a process), hydrogen and tungsten (chemistry-related terms), samples (a grouping concept, meaning a more generic term). We do observe the expected vicinity of temperature and irradiation, two strongly related though distinct concepts. We also identify a highly universal structure close to the node location, with exponential fits applicable to both small and large Euclidean distances in the word2vec space from the node. We fit the histograms with a functional form of the kind

$$f(x) = A_1 \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + A_2 \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) \quad (3)$$

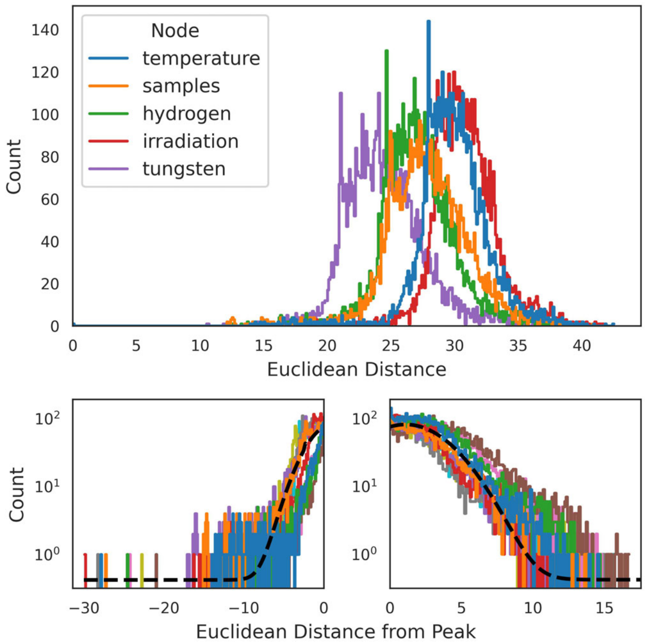

Figure 6. Histogram of Euclidean distances between 5 nodes with the highest degree centrality and every other node in the graph from GPT in Word2Vec space. The curves can be modeled by a sum of 2 Gaussians (black dashed line) according to the fitting function of Equation (3).

and report the fit parameters in Table 1. Based on this characterization, one could identify "mutually common" and "mutually irrelevant" concepts to describe the nodes that are either too close or too distant from the reference node.

Forecasting research trends also requires the investigation of the temporal evolution of the publication process, which is very computationally expensive analysis. We here propose a detailed example for GPT model only.

First, we investigate the effect of a new publication on the KG structure by tracking the number of new nodes added in the KG. By inspecting the number of new nodes as a function of the number of citations received by the manuscript postpublication, a criterion to distinguish innovation from controversy in the publication process can be proposed. In **Figure 7**, we identify four naturally defined quadrants that characterize the publications' effect on the research literature: "Innovative (C1)","Controversial (C2)","Standard (C3)","Incremental (C4)". It is clear that an increase of citations can turn C2 to C1, as well as turn C3 to C4. Based on this analysis, one can identify publications with particular characteristics: 1) *C1*: "Early studies on Cr-Coated Zircaloy-4 as enhanced accident-tolerant nuclear fuel claddings for light water reactors";[23] 2) *C2*: "Understanding of electrochemical behaviors of niobium in molten LiCl–KCl eutectic for pyrochemical decontamination process";[24] 3) *C3*: "Investigation of Zircaloy-fuel interaction in failed spent PWR fuel using EPMA".[25] 4) *C4*: "An insight into radiation resistance of D019 Ti3Al intermetallics".[26]

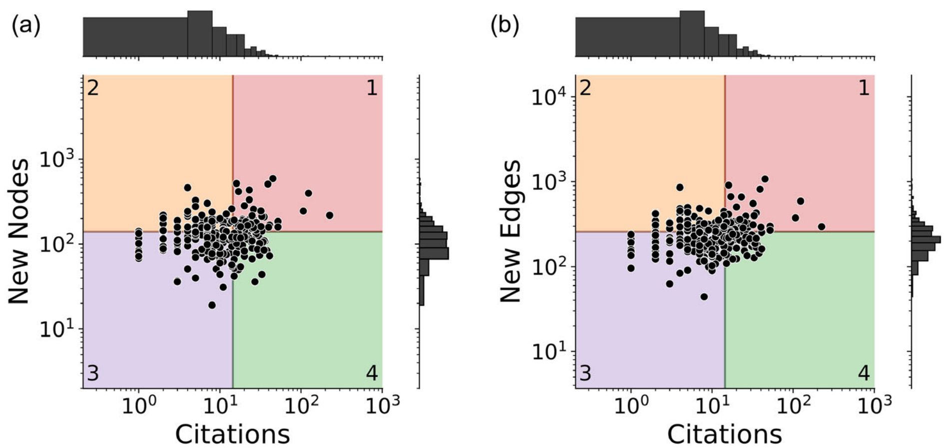

Figure 7. Proposed categorization of articles based on the comparison between the number of citations and effect on the size of the graph. "Innovative (C1)","Controversial (C2)","Standard (C3)","Incremental (C4)".

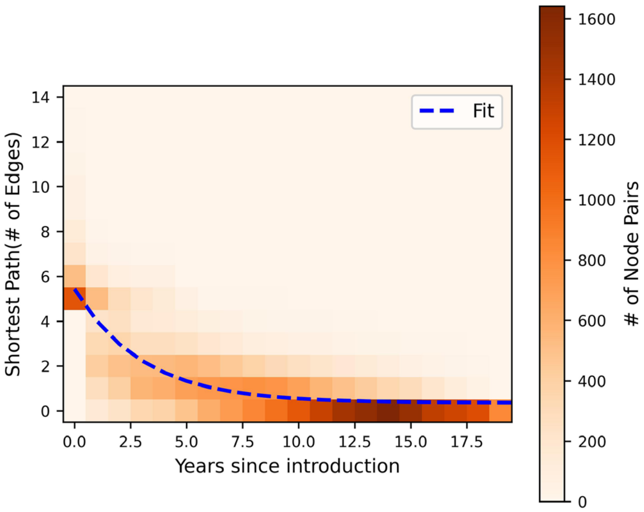

Figure 8. Decay of the SP between pairs of nodes with time progression, SP(t), and the corresponding fitting function (red line) of Eqn. 2. The pixel color in the heatmap is proportional to the number of examples with the same path length after the same number of years. Nodes were selected by summing the number of citations of papers that published them and rejecting those under 200.

In Appedix Figure A7, we report further analysis of the graphs' properties with respect to the number of citations of the article they are based on.

3.4. Temporal Convergence of Research Themes

Second, we investigate the distance of nodes in the KG network space, as a function of time. Naturally, concepts that are mutually distant in the KG are characterized by a large SP, as defined in graph theory.[27] Such SPs decrease in time, due to the fact that concepts are used together and intertwined through research investigations. By investigating the nodes that are mentioned by manuscripts whose sum of citations exceeds 200, the SP between the nodes can be computed, if it exists, and its yearly decrease (**Figure 8**) tracked. We find that a constitutive description of this decrease can be defined as the Equation (2) being t the time measured in years since a path connecting two nodes is created. The values of fitted parameters for the GPT case are $(a,b,c)=(5.07,\ 0.33,\ 0.36)$. In **Figure 9**, we report the decaying behavior for a set of nodes couples that show the largest changes in their SPs with time: these specific nodes are found to be of crucial interest in the nuclear materials community.

By focusing on individual node pairs (**Figure 10**), it is clear that this constitutive description can be used to predict the future commonality of mutually nontrivial concepts, thus predicting the "arrow of time" in this self-contained scientific field, as seen in **Table 2**.

Based on this prototype approach, we can identify the time when the research focus of the community will be on specific topics, such as the investigation of residual stress in large He bubbles (Table 2).

4. Discussion

In this work, we leverage the generalization ability of LLMs to construct high-fidelity KGs from unstructured text. The quality of the LLM is paramount; as shown in our comparative analysis, superior models like GPT-3.5 generate KGs with a power-law structure, suggesting the capture of meaningful, nonrandom relationships between concepts. This structured representation of domain knowledge is the essential substrate upon which any meaningful analysis of novelty or future trends can be built. The similarity in shape between the path length and word2vec distance histograms further suggests that the constructed KGs successfully encode meaningful information from the literature.

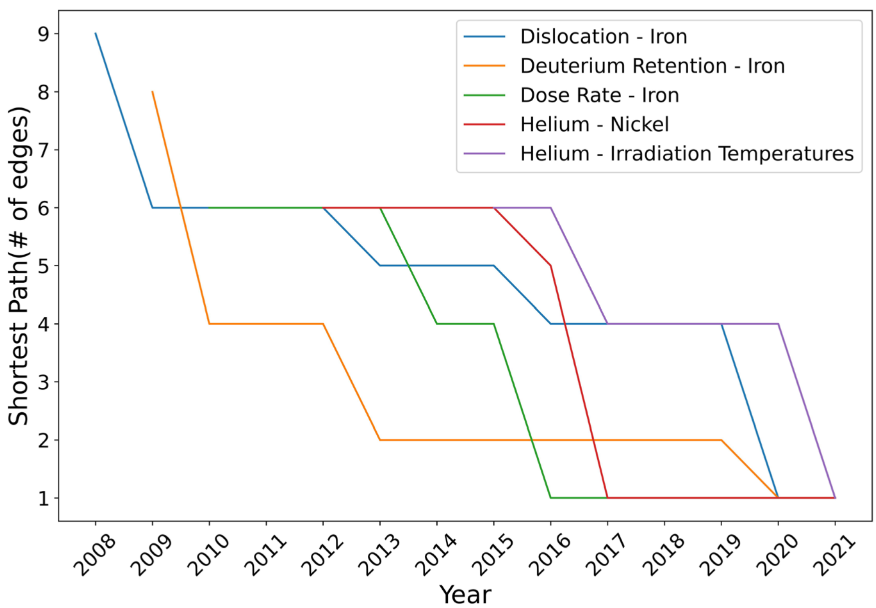

Figure 9. Biggest changes in the SP between pairs of nodes for GPT model.

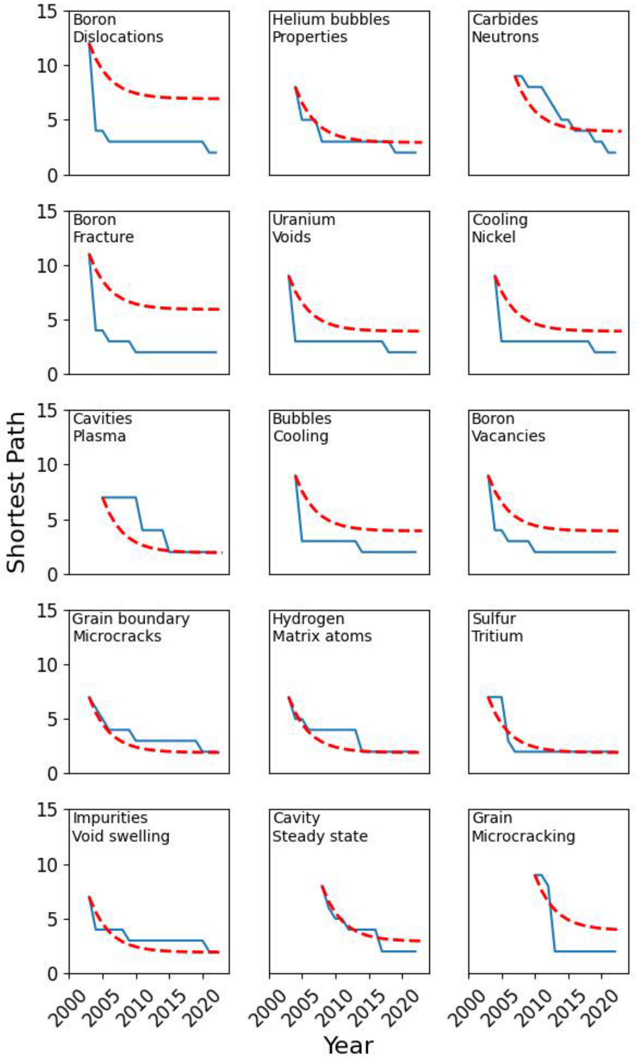

Figure 10. Approximation of shortening distance between pairs of nodes with time progression. Approximation, in red, starts at the point in time the pair is first connected in the graph and at the same distance. From left to right and top to bottom: sample; thermal expansion; cladding; scanning; electron microscopy; dislocations; oxide layer; corrosion; oxide; cladding; tests; error; steel; cracks; heat treatment; toughness; zirconium alloys; irradiated; and SEM observations.

The framework provides a quantitative method for identifying new concepts and assessing the nature of a publication's impact. By plotting the number of new nodes a paper introduces into the KG against the citations it receives, we can classify publications into distinct categories: "Innovative," "Controversial," "Standard," and "Incremental". An "Innovative" publication, for example, is defined by introducing many new nodes and receiving high citations, as shown in several examples such as Cr-Coated Zircaloy-4 fuel claddings. In contrast, a "Controversial" article might introduce many new concepts but initially receive fewer citations. This classification scheme offers a scalable method to automatically flag potentially transformative or debatable research upon its publication.

Our work provides the ability to forecast the evolution of the research field itself. The analysis of the SP between concepts in

Table 2. Projected year of combined focus for pairs of nodes and their SP in 2020.

Node 1	Node 2	SP	Projected year
Large bubbles	Residual stress	4	2023
Residual stress	Uranium	3	2023
Displacement dose	Residual stress	5	2025
Residual stress	Solution	3	2023
Annihilation	Residual stress	3	2023
Fuel rods	Residual stress	3	2023
Formation	Residual stress	4	2023
Grains	Residual stress	2	2023
Radiation damage	Residual stress	3	2023
History	Residual stress	4	2023
Oxide layer	Residual stress	3	2023
Residual stress	SIA clusters	4	2023
Database	Residual stress	3	2023
Point defects	Residual stress	2	2023
Oxide	Residual stress	3	2023

the KG network over time reveals the process of scientific integration, where disparate ideas are gradually connected through new research. We discovered that this convergence is not random but follows a constitutive exponential law, where the SP between two connected concepts predictably decreases over time. This empirical observation was modeled as $\mathrm{SP}(t) = ae^{-bt} + c$ as a predictive mathematical tool. By applying this model, our framework can forecast the year in which distinct research topics will merge into a combined focus. As demonstrated in our results, we can project that concepts like "residual stress" will become closely integrated with "large bubbles," "Uranium," within a few years. This transforms knowledge mining from a retrospective activity into a prospective one. Instead of simply summarizing what is known, this methodology allows researchers and funding agencies to anticipate and prioritize future interdisciplinary research areas, fulfilling the role of a data-driven forecast for research trends.

By explicitly linking LLM-derived KG metrics to empirical measures of future impact, our discussion clarifies how the experimental results empower new-knowledge mining. The structured indices and dynamic graph analysis we present offer both descriptive insight and actionable foresight, marking a significant advance in the automated discovery of emergent scientific concepts (Figure A8).

5. Conclusions

In conclusion, this work presents a framework for understanding conceptual development in scientific literature, as well as making predictions of future research interests. By using the embedding space-defined concepts of mutually "common" and "unrelated" concepts in LLM-constructed KGs, we identify a constitutive law for the relational aspect of distinct concepts in time, as their SPs decrease. We believe that refinements of this approach can lead to groundbreaking predictions of future research, analogous to long-term weather forecasts that are routine nowadays, but in the semantic space.

Appendix

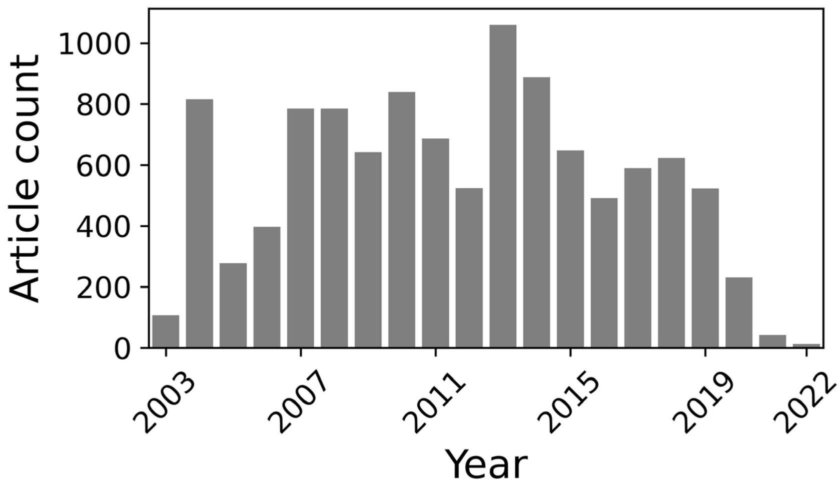

Figure A1. Number of articles per year considered from the *Journal of Nuclear Materials*.

... **Nuclear power**[10] has proven to be a **reliable**[14], **environmentally sustainable**[11], and **cost-effective**[13] source of large scale electricity. In the interest of continued **technological improvement**[9], further improvements in **operational reliability**[12], **economics**[5], and **safety**[4] under normal and transient conditions are being pursued worldwide. In regard to safety during anticipated transients and postulated accidents, the safety of the **fuel**[1] can be affected by numerous **fuel system phenomena**[7], including **cladding oxidation/hydriding**[2], **pellet-cladding interaction**[8], **pellet relocation and dispersal**[6], and **cladding embrittlement and fragmentation**[3] [1]. ...

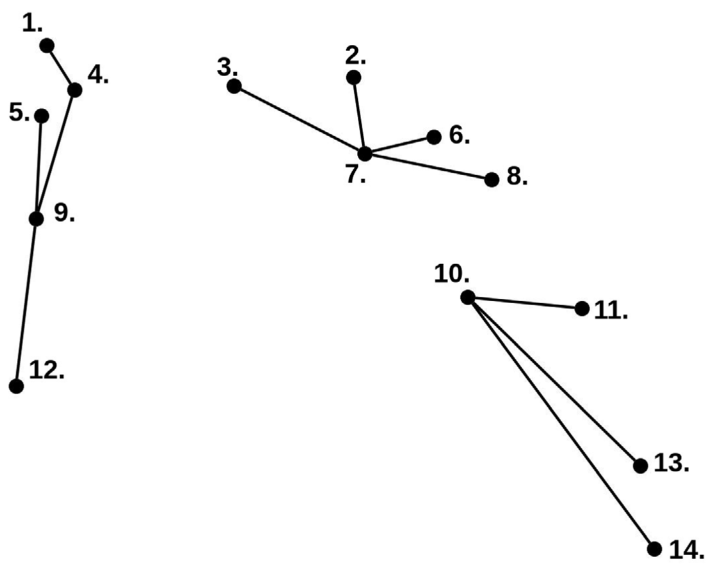

Figure A2. First step of building a graph.

... The high **power density**[16] that makes nuclear power **economical**[19] also makes the system susceptible to severe **accidents**[20]. Typical **power densities**[17] in **light water reactor**[18] (LWR) cores are 50-75 MWth m-3, which is about a **factor**[15] of 100 greater than the average power density in fossil energy plant boilers [21. In a typical **loss of coolant accident**[29] (**LOCA**[27]) scenario, a **reactor scram**[30] dramatically reduces the **power generation**[31] in the core. However, substantial amounts of **heat**[28] continue to be generated following the scram: 7% of **full power**[24] immediately after the scram, 1% of full power 4 h after the scram, and 0.2% of full power 10 days after the scram [3]. Considering that the **thermal full power levels** [25] in commercial **LWRs**[26] often exceed 3000 MW, the post-scram decay heat deposited in a **reactor core**[23] without circulating **coolant**[21] (i.e., approaching **adiabatic conditions**[22]) can lead to rapia temperature increases and subsequent core degradation. ...

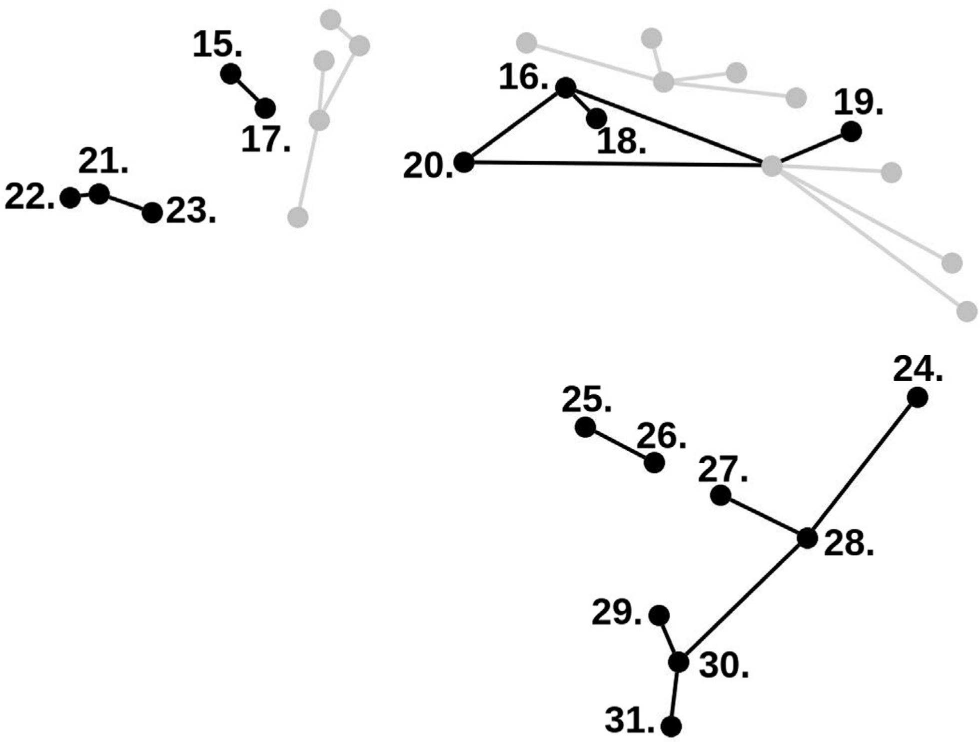

Figure A3. Second step of building a graph.

... Also due to the **high power density**[41] in **nuclear reactor cores**[37], the **system**[35] has relatively limited margin for accommodating additional energy generation under normal operating conditions. Unlike fossil **energy**[43] **electricity generation**[44] sources, fuel, **operation**[38], and **maintenance**[39] of nuclear reactors constitute only a minority fraction (one quarter) of the electricity generation cost (the rest being the capital costs). Therefore, the economics of nuclear reactor operation improves as the **power rating**[36] of these **units**[32] increases, which has led to many of the currently operating units undergoing power uprates. While plant power uprates involve a number of considerations, the **thermal limits**[33] of the current **fuels**[34] are one factor that is limiting future **power level increases**[42] since further local power increases in the core could result in fuel **damage**[45] under **anticipated operational occurrences**[40][1]. ...

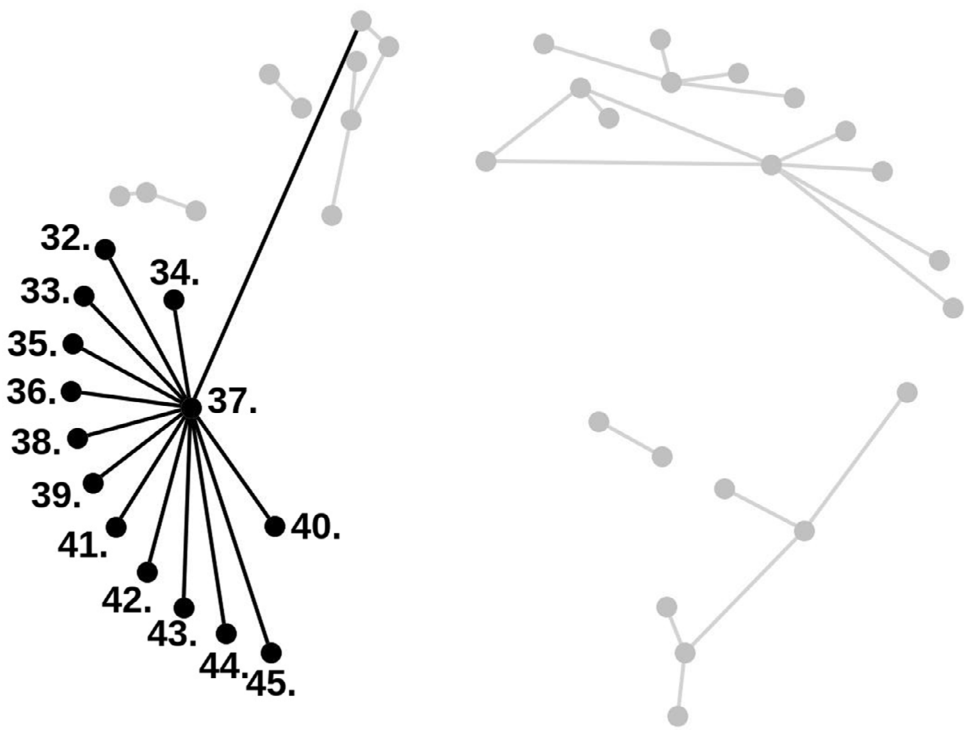

Figure A4. Third step of building a graph.

... In addition to **LOCA events**[52] as discussed above, another **safety design basis**[56] is related to the **potentially rapid power excursions**[54] that can occur from **reactivity insertion events**[53], such as control rod ejection in an LWR. Energy deposition resulting from **reactivity insertion**[48] in the **core**[49] can be limited by appropriate **reactor physics**[50] design of the core, balancing the design reactivity worth of the **control rods**[46] with the **fuel temperature feedback**[47]. However, the specific **fuel system response**[55] to any given magnitude of **energy deposition**[51] varies during its lifetime because of the changing **thermal-mechanical properties**[57] of the fuel/cladding system. ...

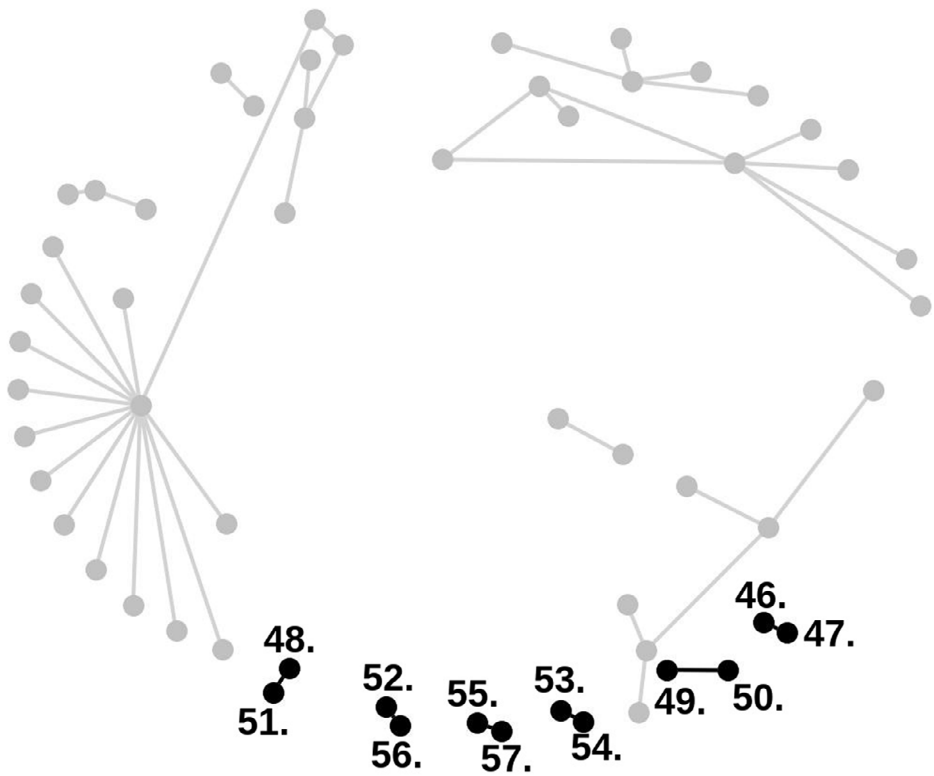

Figure A5. Fourth step of building a graph.

ADVANCED INTELLIGENT SYSTEMS
Open Access
www.advintellsyst.com

... As discussed in Section 2, numerous operational and retrofit **design changes**[67] were instituted in LWRs in the 1980s following the LOCA at Three Mile Island, and these changes have substantially reduced the probability of **core degradation**[68] during a severe accident. These specific design features and upgrades were primarily associated with maintaining **adequate core cooling**[64] in the event that the **primary cooling system**[65] is not functional. The recent **station blackout (SBO) accidents**[62] at three of the Japanese **Fukushima Dai-ichi reactors**[63] following the devastating 2011 earthquake and tsunami have sparked renewed interest in exploring the possibility of further design and fuel system improvements that could improve the safety of LWRs under **design-basis (DB)**[59] and **beyond-design-basis (BDB) accident scenarios**[58]. Section 3 provides a brief overview of **core degradation phenomena**[60] under DB and BDB **accident conditions**[61] and possible mitigation by design changes and utilization of new fuel systems. Section 4 reviews the overarching motivation for investigating **accident tolerant fuel (ATF) systems**[66], and Section 5 briefly summarizes some of the ATF concepts currently being explored. Several accompanying papers in this special issue provide more detailed information on the current status of some specific ATF systems. ...

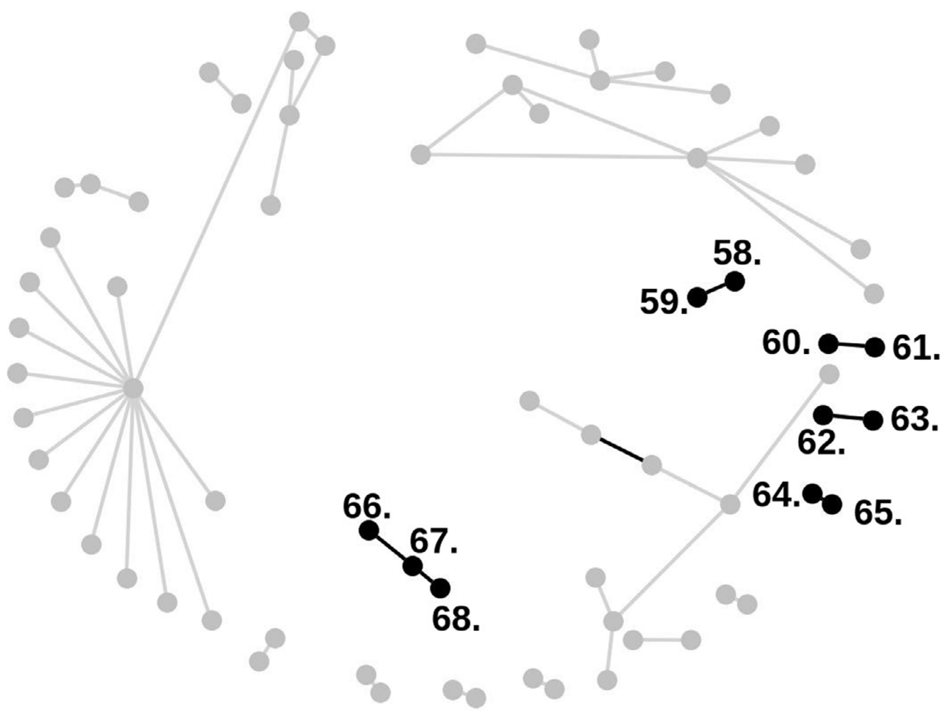

Figure A6. Fifth step of building a graph.

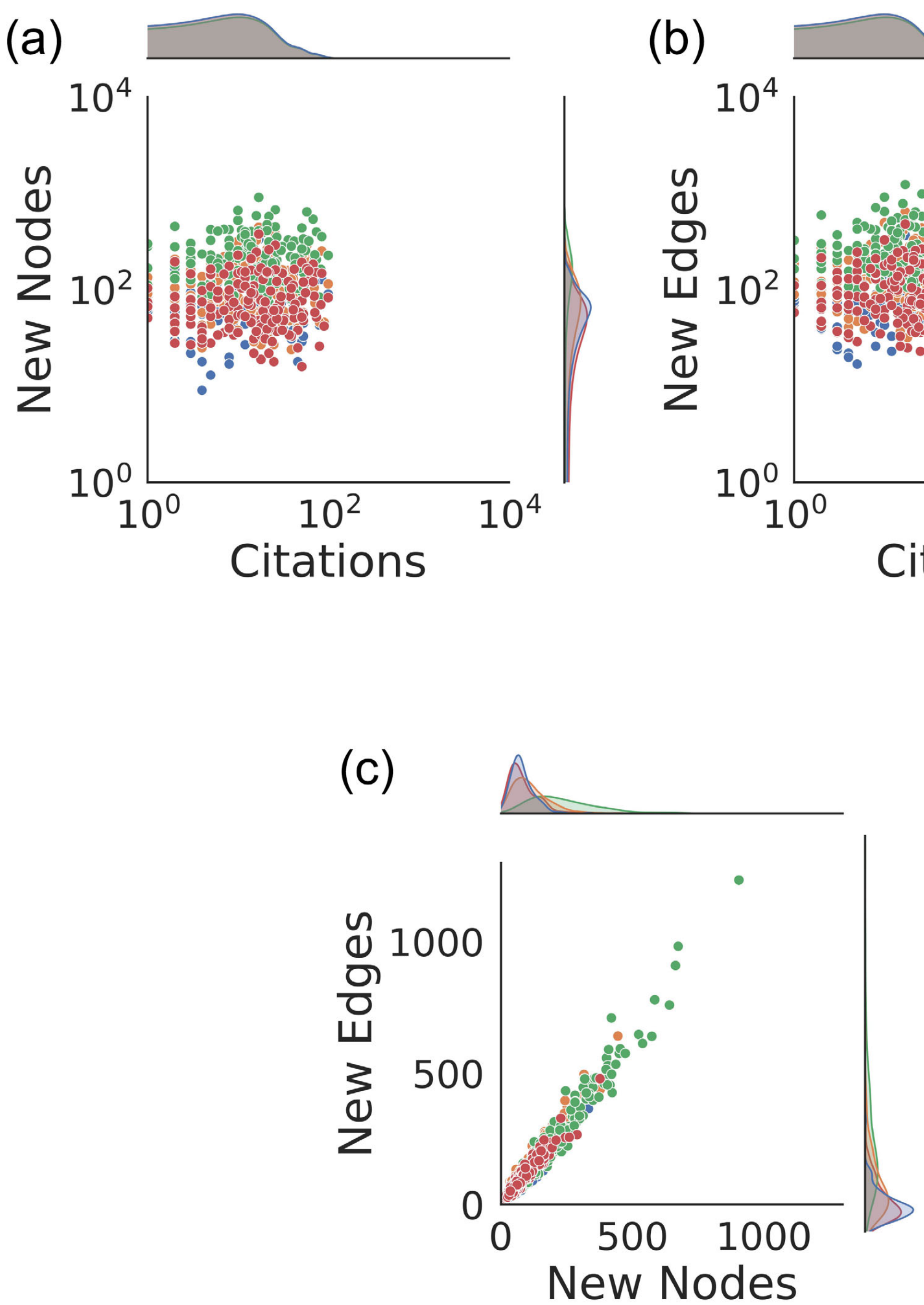

Figure A7. Impact of the graph vs citations of papers from different LLMs. a) Number of new nodes by the paper versus number of its citation. b) Number of new edges introduction by the paper versus number of its citation. c) Number of new nodes versus new edges introduction by the article.

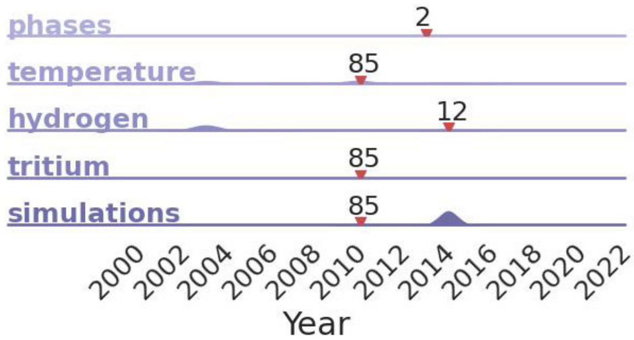

Figure A8. Node appearance frequency in literature by year. Red triangles mark the year of publication of the most cited article mentioning each node with the number equal to the number of citations of that article to date.

Acknowledgements

This work was supported by a grant from the National Research Foundation of Korea (NRF) funded by the Korean government, Ministry of Science and ICT (MSIT) (grant no. RS-2024-00356670). The author(s) D. Massa and S. Papanikolaou, were supported by the European Union Horizon 2020 research and innovation program, under grant agreement no. 857470 and from the European Regional Development Fund under the program of the Foundation for Polish Science International Research Agenda PLUS, grant no. MAB PLUS/2018/8, and the initiative of the Ministry of Science and Higher Education 'Support for the activities of Centers of Excellence established in Poland under the Horizon 2020 program' under agreement no. MEiN/2023/DIR/3795. P.S. was partially supported by ERC consolidator grant TUgbOAT no. 772346 and NCN no. c. YJP, CWH, PB, and JL acknowledge support by DTRA (award no. HDTRA1-20-2-0002) Interaction of Ionizing Radiation with Matter (IIRM) University Research Alliance (URA), and the Defense Advanced Research Projects Agency (DARPA) under agreement no. HR00112490369.

Open Access funding enabled and organized by MIT gold 2025.

```
You are a scientific network graph maker who extracts terms and their
relations from a given context. You are provided with a context chunk
(delimited by ```).
Your task is to extract the ontology of terms related to nuclear
materials, material sciences, underlying physics, and history in general
mentioned in the given context.
These terms should represent the key concepts as per the context.
Thought 1: While traversing through each sentence, think about the key
terms mentioned in it.
    -  Terms may include object, entity, location, organization, person,
       condition, acronym, documents, service, concept, etc.
    -  Terms should be as atomistic as possible.
    -  Avoid duplicated relationship as much as possible.
    -  If there are unicode symbols, be sure to decode them rather than
       encode them. (Negative example: \u03b2-z, Positive example: %β-z).
Thought 2: Think about how these terms can have one on one relation with
other terms.
    -  Terms that are mentioned in the same sentence or the same paragraph
       are typically related to each other.
    -  Terms can be related to many other terms.
Thought 3: Find out the relation between each such related pair of terms.

Format your output as a list of dictionary. Each element of the list
contains a pair of terms and the relation between them, like the
following:
{"ontology":
    [
        {
            "node_1": "A concept from extracted ontology",
            "node_2": "A related concept from extracted ontology",
            "edge": "relationship between the two concepts, node_1 and
node_2
                    in one or two sentences"
      }, {...}
   ]
}
```

Figure A9. The prompt used to generate KGs from literature.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Maciej Tomczak: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); validation (lead); visualization (equal); writing—original draft (supporting); writing—review & editing (equal). **Yang Jeong Park**: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); writing—original draft (supporting); writing—review & editing (equal). **Chia-Wei Hsu**: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); writing—original draft (supporting); writing—review and editing (equal). **Payden Brown**: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); writing—original draft (supporting); writing—review and editing (equal). **Dario Massa**: conceptualization (equal); resources (equal); supervision (equal); visualization (equal); writing—original draft (supporting); writing—review and editing (equal). **Piotr Sankowski**: conceptualization (equal); funding acquisition (equal); investigation (supporting); methodology (equal); project administration (equal); resources (equal); supervision (supporting); validation (equal); writing—review and editing (supporting). **Ju Li**: conceptualization (equal); formal analysis (equal); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (equal); supervision (lead); writing—original draft (supporting); writing—review and editing (equal). **Stefanos Papanikolaou**: conceptualization (lead); formal analysis (equal); funding acquisition (lead); investigation (lead); methodology (equal); project administration (lead); resources (equal); software (equal); supervision (lead); validation (equal); visualization (equal); writing—original draft (lead); writing—review and editing (equal). **Maciej Tomczak** and **Yang Jeong Park** contributed equally to this work.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

knowledge graphs, large language models, materials informatics, nuclear materials

Received: January 16, 2025
Revised: July 22, 2025
Published online:

[1] J. Achiam, S. Adler, S. Agrawal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, et al. (Preprint), arXiv:2303.08774, v1, Submitted: Mar. **2023**.
[2] G. Team, R. Anil, S. Borgeaud, Y. Wu, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, et al. (Preprint), arXiv:2312.11805, v1, Submitted: Dec. **2023**.
[3] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriquez, A. Joulin, E. Grave, G. Lample (Preprint), arXiv:2302.13971, v1, Submitted: Feb. **2023**.
[4] Microsoft Research AI4Science, Microsoft Azure Quantum (Preprint), arXiv:2311.07361, v1, Submitted: Nov. **2023**.
[5] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, J. Steinhardt (Preprint), arXiv:2009.03300, v1, Submitted: Sep. **2020**.
[6] H. Cai, X. Cai, J. Chang, S. Li, L. Yao, C. Wang, Z. Gao, Y. Li, M. Lin, S. Yang, J. Wang, M. Xu, J. Huang, X. Fang, J. Zhuang, Y. Yin, Y. Li, C. Chen, Z. Cheng, Z. Zhao, L. Zhang, G. Ke (Preprint), arXiv:2403.01976, v1, Submitted: Mar. **2024**.
[7] J. Li, X. Cheng, W. X. Zhao, J.-Y. Nie, J.-R. Wen (Preprint), arXiv:2305.11747, v1, Submitted: May **2023**.
[8] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, *ACM Trans. Comput. Healthcare* **2021**, *3*, 1.
[9] M. Parmar, S. Mishra, M. Purohit, M. Luo, M. H. Murad, C. Baral (Preprint), arXiv:2204.07600, v1, Submitted: Apr. **2022**.
[10] L. Yang, H. Chen, Z. Li, X. Ding, X. Wu, *IEEE Trans. Knowledge Data Eng.* **2024**, *36*, 3091.
[11] S. Hao, B. Tan, K. Tang, B. Ni, X. Shao, H. Zhang, E. P. Xing, Z. Hu (Preprint), arXiv:2206.14268, v1, Submitted: Jun. **2022**.
[12] R. Albert, A.-L. Barabási, *Rev. Modern Phys.* **2002**, *74*, 47.
[13] A.-L. Barabási, J. Frangos, *Linked: How Everything is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*, Basic books, New York **2014**.
[14] T. Mikolov, K. Chen, G. Corrado, J. Dean (Preprint), arXiv:1301.3781, v3, Submitted: Sep. **2013**.
[15] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *OpenAI Technical Report* **2018**, https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035 (accessed: May 30, 2025).
[16] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *OpenAI blog* **2019**, *1*, 9.
[17] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. F. Christiano, J. Leike, R. Lowe, *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730.
[18] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., (Preprint), arXiv:2307.09288, v1, Submitted: Jul. **2023**.
[19] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, D. d. L. Casas, E. B. Hanna, F. Bressand, G. Lengyel, G. Bour, G. Lample, L. R. Lavaud, L. Saulnier, M.-A. Lachaux, P. Stock, S. Subramanian, S. Yang, S. Antoniak, T. L. Scao, T. Gervet, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed (Preprint), arXiv:2401.04088, v1, Submitted: Jan. **2024**.
[20] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. V. Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. M. Rush (Preprint), arXiv:1910.03771, v5, Submitted: Jul. **2020**.
[21] A. A. Hagberg, D. A. Schult, P. J. Swart, in *Proc. of the Python in Science Conf.*, SciPy, Pasadena, CA **2008**, pp. 11–15, https://doi.org/10.25080/tcwv9851 (accessed: June 14, 2025).
[22] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, . Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, et al., *Nat. Methods* **2020**, *17*, 261.

[23] J.-C. Brachet, I. Idarraga-Trujillo, M. L. Flem, M. L. Saux, V. Vandenberghe, S. Urvoy, E. Rouesne, T. Guilbert, C. Toffolon-Masclet, M. Tupin, C. Phalippou, F. Lomello, F. Schuster, A. Billard, G. Velisa, C. Ducros, F. Sanchette, *J. Nuclear Mater.* **2019**, *517*, 268.
[24] G. Y. Jeong, S. Sohn, Y. Jeon, J. Park, *J. Nuclear Mater.* **2019**, *524*, 39.
[25] Y. H. Jung, S. J. Baik, S. B. Ahn, *J. Nuclear Mater.* **2019**, *517*, 349.
[26] R. Voskoboinikov, *J. Nuclear Mater.* **2019**, *519*, 239.
[27] E.W. Dijkstra, *Edsger Wybe Dijkstra: His Life, Work, and Legacy*, Vol. *45*, Association for Computing Machinery, New York, NY **2022**, pp. 287–290.

