Toward Human-AI Co-creation to Accelerate Material Discovery

Anonymous Author(s)
Affiliation
Address
email

Abstract

Urgency of scientific discovery in chemistry carries with it the extra burden of assessing risks of the proposed novel solutions. Despite the several advances in Machine Learning and AI, there is still a gap in technologies to support the creation of end-to-end discovery applications, integrating the myriad of available technologies into a coherent, orchestrated, yet flexible discovery process. Such applications need to handle knowledge management at scale, enabling knowledge consumption and production at the right time and in the right form. Furthermore, the discovery of novel functional materials strongly relies on the development of exploration strategies in chemical space. For instance, generative models have gained attention within the scientific community due to their ability to generate enormous volumes of novel molecules across material domains. These models exhibit extreme creativity that often translates in low viability of the generated candidates. In the context of materials discovery, viability is a complex metric evaluated by subject matter experts from complementary domains, such as synthetic organic chemistry, process scale-up, development of intellectual property, regulatory compliance, and such. In this scenario, we observe an excellent opportunity to incorporate AI techniques to support the experts, but it is also clear the need for a platform to exploit the human-ai interaction focusing on reducing the time of the first discovery and the cost of opportunity. In this work, we propose a platform for human-AI Co-creation to accelerate material discovery, which has four main components: (1) generative models, (2) dataset triage, (3) molecule adjudication, and (4) risk assessment.

1 Introduction

In the technological era, faster scientific discovery grants higher resilience at societal and individual level to the ever-emerging challenges, such as the latest pandemic. Even under normal circumstances, the entities facing discovery tasks, from the level of individual researcher to an academic or industrial organization, face multiple inefficiencies and bottlenecks that hinder overall progress. In this contribution, we focus on challenges pertinent to the discovery of polymeric materials and mitigation strategies rooted in human-AI co-creation. Many of the raised points pertain to other domains of chemical and material sciences [10, 14].

Any chemical product, from a small-molecule drug to a piece of plastic wrapping, has a non-trivial footprint over its lifecycle. During discovery phase, assessment of the utility of new materials cannot be restricted to the functional characteristics anymore. It must include a comprehensive set of factors ranging from the nature of the precursors (is it petrochemistry or bioengineering?) and catalysts (is it a rare-earth metal or enzyme?) to the degradation pathways, patterns of bioaccumulation, and toxicity. Historically, it occurred more than once, that the technological attractiveness came with such a fine print that rendered the technology undesirable after severe consequences have been encountered.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

The old pattern of discovery that disregards technological risks of the new materials is not tenable anymore [4]. A separate category of risks is associated with discovery activities as such. After all, discovery requires investment of appreciable resources with the goal of achieving such level of success that offsets the costs. Here, "resources", "costs", and "success" should be understood in the broadest manner. For example, academic discovery measures success primarily in the volume of published papers as a proof of work, whereas industrial discovery relies on development of intellectual property and reduction to practice. In any case, the discovery process is associated with risks such as inability to solve the task with given budget and/or given time, unacceptable opportunity costs, inadequate selection of discovery targets, and so on.

In the contemporary scientific discovery, subject matter experts (SMEs) are the owners and the stakeholders. Regardless of comparative performance of SMEs and AI systems in various specialized tasks, these roles will remain uniquely human in the foreseeable future. These roles fundamentally dictate that selection of goals and targets, project scoping, development of the technical specifications for the discovery components, establishing constraints, and making actionable decisions fall under SMEs responsibility and require adequate level of expertise and knowledge. Expert knowledge is conceptual and supports reasoning within the specific scientific domain. In a sense, expert knowledge bears similarity to "common sense" knowledge, considering that it must be shared among domain experts as a common denominator enabling scientific discourse. Expert knowledge is combined with tacit knowledge, which is highly personalized as a reflection of the unique scientific trajectories of the experts. A part of the tacit knowledge is attributable to the institutional knowledge encapsulating concepts and practices of institutions (cf., scientific school of thought). The nature of SMEs knowledge is such that it cannot be efficiently captured by contemporary AI systems because it is not captured in structured or unstructured sources. Simultaneously, SMEs knowledge and expertise have glaring deficiencies. They are incomplete, they are inseparable from humans' cognitive biases, and they are coupled with the fundamental irrationality of human decision-making. Therefore, the role of the SMEs in the discovery process is to contribute complementary cognitive capabilities and conceptual knowledge as the primary agent in the discovery process, while knowingly mitigating counterproductive aspects of human intelligence via interactions with AI components [1, 3].

In summary, the investigation of new AI-based approaches and their supporting technologies to accelerate materials discovery is critical. Such approaches must empower SMEs with an integrated platform to exploit the human-AI interaction. We advocate that such platform must focus on reducing the time of the first discovery and the cost of opportunity, levering the human-AI Co-creation symbiosis to accelerate material discovery. In this work, we will discuss a knowledge-driven discovery workbench that integrates four main components: (1) generative models, (2) dataset triage, (3) molecule adjudication, and (4) risk assessment. In the next section, we present an initial use case to guide the rest of this text. In this use case, we integrated the four components to support SMEs in discovering new Photoacid generators (PAG), a critical class of materials used in semiconductor manufacturing evaluated more closely for environmental risks [4]. After, we discuss the four main components of the discovery workbench.

2 The PAG use case

Sustained growth of computing platforms is contingent on sustainable development of semiconductor technology, which cannot be taken for granted, as the pandemic-induced chip shortages demonstrated. Technology of chemical amplification is the foundation of high performance of semiconductor lithography. Chemical amplification requires specialized molecular components such as photo-acid generators (PAGs) which attract progressively higher levels of regulatory scrutiny. Industrial stakeholders recognize that the healthy and meaningful growth of semiconductor manufacturing is demands discovery of high-performing PAGs with acceptable environmental characteristics over their lifecycle. One specific recent constraint, an expanded regulatory definition of the class of per- and polyfluoroalkyl substances (PFAS). Many broadly used PAGs now fall under its scope because fluorination is a robust design principle, allowing to stabilize strong acids formed during PAGs transformation in the lithographic process.

We implemented a process for PAG discovery where AI and SMEs arrive to the new PAG candidates in a collaborative manner. This process starts integrating historical data with generative models, the idea is to extract examples of molecules from documents, e.g., patents, to feed the AI-based molecule generators [7, 12]. Both the inputs to the generative models and outputs (chemical structures of PAG

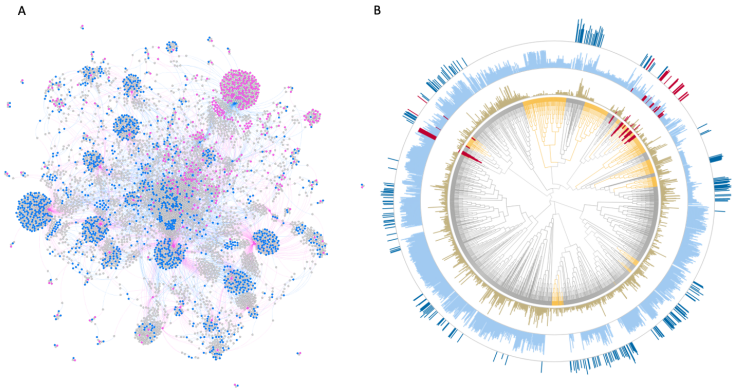


Figure 1: Panel A shows constitutional similarity network of ionic PAGs found in patents. Grey nodes are PAG salts (ionic pairs), pink nodes are anions, and blue nodes are cations. Nodes of the cation and anion comprising the salt are connected to the node representing the salt. Panel B shows phylogenetic tree of PAG cations extracted from patents (grey branches) and produced by generative model (orange and red branches), where red branches represent generated cations accepted by SMEs. Radial bar charts represent computed properties: LD50 (tan), Biodegradability (light blue), and λ_{max} (blue).

cations) are computationally characterized to produce labels including wavelength of the highest photon absorption λ_{max}, biodegradability, and 50% lethal dose LD50. Figure 1 shows the historical data along with the result of generative models for the PAGs. In practice, we processed approximately 5K patents, extracted approximately 4K structures of ionic PAGs comprised of 1.3K cations and 0.5K anions. The sample produced by the generative models comprised approximately 10K PAG cations. Since the volume of molecules generated is much more than the capacity of synthesizing them, we introduce peer-reviewing mechanisms where SMEs interact with the generative models. First, SMEs assign viability labels to a subsample of the generated structures in *adjudication phase* by answering the question "Does this candidate merit further exploration?". Here, we do not resolve the factors dictating the SME's decision and focus on learning the general pattern of rejections. Typically, they are associated with issues of synthetic costs, scale-up possibility of synthesis, possibility of undesirable secondary chemical processes, etc. Adjudication phase enables us to model the rejection priorities of the SMEs and prioritize the entire generated sample. Second, SMEs define specialized cutoffs based on the application and according SMEs' specialization, for example, functional application, generation of intellectual property, and environmental regulations. Along with the adjudication labels, these explicit filters enable *triage phase* where we reduce the number of molecules as much as by three orders of magnitude and rank them according to their viability as discovery targets. Finally, the set of molecule candidates goes through a risk assessment methodology where a multidisciplinary group of SMEs evaluates the molecule to decide whether it will be synthesized or not based on the known-unknowns and knowns-characteristics (Section 3.3).

3 Discovery Workbench

To tackle the PAG problem, we created the *Discovery Workbench* (DWb) framework, which aims at accelerating the construction of end-to-end discovery applications. DWb offers mechanisms to integrate multiple discovery tools and to support common discovery workflows by means of an integrated and user-friendly web-based interface. DWb provides systematization to support experts with the discovery process and tracks the knowledge that is created along that process throughout its entire life cycle, from hypothesis generation and collection (*i.e.* material candidates), to hypothesis testing (*i.e.* automated simulations and lab experiments), and then to drawing conclusions (*i.e.* visualizations and comparisons).

At its core is the idea of capturing and organizing the knowledge created and obtained during a discovery process executed by domain experts, so it can be reused in future iterations to accelerate

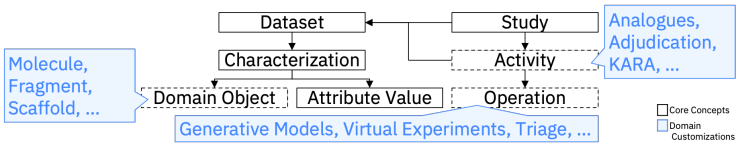


Figure 2: Core concepts of the DWb (in black) and the domain customizations (in blue).

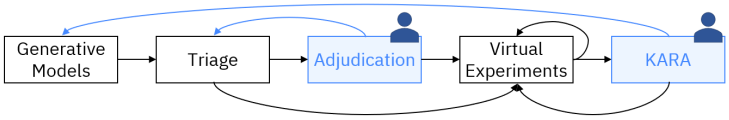


Figure 3: Example of workflows supported by DWb in the PAG use case. Blue boxes show expert-in-the-loop activities and blue arrows highlight machine teaching opportunities.

other instances of that same process. To achieve this, DWb deals with few core concepts, some of them being customized per domain (figure 2):

Dataset A collection of *Characterizations*.

Characterization A collection of *Attribute Values* for a *Domain Object*, representing a possible representation of that object.

Domain Object The finest granularity of instances that the system will handle, which is customized per domain and can be anything (*e.g.*, molecules, complex materials, oil reservoirs).

Study A focused investigation, created by domain experts, with associated *Datasets* of interest in which we can perform *Activities*.

Activity Customized per domain, transforms a *Dataset*, generating more information for the *Characterizations*, filtering the *Dataset*, etc. It may trigger automated *Operations* and/or involve expert-in-the-loop processes (*e.g.*, adjudication, KARA).

Operation Customized per domain, it is the execution of an algorithm that typically produces a new *Dataset* (*e.g.*, generative models, triage).

It provides multiple ways for the domain experts to interact with the system, empowering them to perform their daily activities. For example, in the PAG use case, we can follow the execution of a "triage" operation with an "adjudication" activity or a "virtual experiment" one (shown in figure 3.

From a technical standpoint, DWb is a combination of several components. Its architecture is focused on the extensibility in multiple levels: adaptation to new domains through extensions to its core ontology, integration of new plugins, and customization of new deployments following a Software Product Line approach. An overview of its architecture can be seen in figure 4.

3.1 Generative Models

Let's compare discovery of materials and small-molecules. Primary difficulties of small-molecule discovery are large cardinality of the set of molecules (often estimated as 10^60) and severe limitations

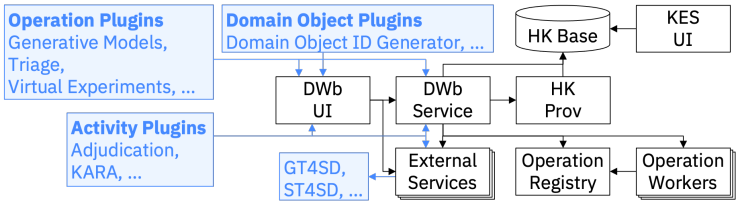


Figure 4: Overview of DWb architecture with the components in black and the extensibility points in blue.

of synthetic accessibility of large fraction of this set. Discovery of polymer materials includes synthesis of precursors as the initial step, so the large but finite cardinality of the set of molecular precursors, catalysts, and molecular additives is a contributing factor. However, polymerization itself, polymer formulation, and material processing require choosing values of continuous degrees of freedom with infinite cardinality. Moreover, there's a separate combinatorial explosion of the experimental protocols transforming a specific precursor into the material suitable for a technological application.

Design of experiments solving search and optimization tasks under such circumstances must be as efficient as possible to meet practical constraints on time and cost of discovery. Generative models (GMs) offer an attractive route of finding discovery targets, or at least plausible candidates, by expanding sets of known solutions to include new, generated candidates according to learnt joint probability distribution of the features over the initial set. Technical details of various generative modeling strategies can be found in many works like *Hoffman et al.* [4] and *Maziarka et al.* [7].

Operationally, the main question that needs to be addressed in application of generative models to materials discovery is how to reconcile the volume of the generated structures with finite throughput of chemical processes, regardless of the level of automation and autonomy. GMs in practice require a some auxiliary screening and constraining via discriminative models and filters. Tacit expert knowledge plays enormous role here – in lieu of GMs that can blend conceptual knowledge with factoids, SMEs assume discriminative role either directly or indirectly, via models of the SMEs decisions. Level of acceptance of GM candidates by SMEs can vary appreciably among different GM models and depending on the nature of SMEs task. For example, materials scientists are being more stringent in evaluation of synthetic practicality of new candidates and more accepting in evaluation of their functional potential.

In the end-to-end materials discovery the AI component based on a GM is expected to generate hypotheses that pass peer-review by the SME who is the owner of the discovery task. This leads to the first metric of GM quality assessment, the fraction of candidates that pass SME-driven peer-review. We would like to emphasize, that this is an observed and factually relevant metric, and it does not imply that SME decisions are flawless. This metric equally applies "autonomous lab" discovery - any autonomous scientific equipment has an SME owner and stakeholder who is directly responsible for the efficiency and cost of platform utilization. The secondary metrics include comparison of the generated and initial samples in terms of the geometry (topological data analysis on point clouds, manifold learning, and such), and statistics of the constitutional features and properties as proposed by *Tadesse et al.* [15].

3.2 Dataset Curation

After the generative model execution, the materials' dataset can be very extensive with numerous entries that are unrelated with the current task at hand (in our experience, about 10^5 candidates). While this is the generative model strength (quickly generate an enormous number of candidates), it is also its downsize – it is impractical to investigate each individual candidate. Thus, we follow the generative model with a *dataset curation* process to select/prioritize the materials.

We based this process in two strategies: one based on the implicit restrictions specific to the user and the study and another based on the computable properties. There are numerous restrictions that will never become explicit, as it has to do with (i) the uniqueness of the structure and (ii) the knowledge access. Due to the nature of Generative Model, it can pick up such patterns of atom and functional group arrangements that would not come to the SME's mind – so they cannot say upfront that some of those arrangements would be unfavorable, therefore not being able to create a restriction for the model. Moreover, many restrictions are institutional and/or tacit knowledge, not being able to extracted them from literature as they are specific to a group and may even be trade secrets.

Since we have this implicit knowledge, the DWb provides tools to empower the SMEs to express this knowledge. DWb has many data exploration features, such as sort, filter, visual analytics, statistical analysis, cluster analysis, etc. Additionally, we developed the adjudication plugin to support an expert-in-the-loop activity that captures a SME initial evaluation of a given material. The SME should go through every material and answer a single question with few answer choices (for example: "Does this candidate merit further exploration?", with answers "No", "Uncertain", and "Yes") by evaluating only its constitution – substructures, functional groups, ionic core.

The triage and adjudication process are intertwined, such as the expert input during adjudication feeds back to the triage model (as highlighted by the blue arrow in figure 3). By performing the adjudication, the SME is informing the triage AI models about the boundaries of the acceptable creative space.

After a relatively small amount of labeling by SME (about 1000 can be evaluated in under an hour), the triage model is trained to identify the patterns of SME decisions and can rank the structures on its own. After the triage step, we expect that 10% of the initial generative model dataset will be available for the next step, the risk assessment.

3.3 Risk Assessment in Materials Discovery

In the discovery process of a novel material, evaluating the multiple dimensions that may prevent the material from succeeding for a purpose is essential. For example, some molecule candidates generated by AI models may not be successfully synthesized or represent a great danger to the environment due to their toxicity. In this work, we refer to risk assessment as assessing the Probability-of-Success (POS) of a novel material obtained from synthesizing a molecule candidate to fulfill its purpose. Evaluating the multiple dimensions (or risk factors) heavily depends on experts' knowledge from different areas. Computational tools, such as molecular dynamics simulators, may support the experts in this task. Despite the complexity and the heavy computational demand of those tools, ultimately, the SMEs have the final word due to the potential limitation and uncertainty of simulation results. These resource requirements (computational and human) limit the assessment of numerous candidates. So, the previous phases of triage and adjudication are fundamental to filtering the most promising ones.

The strong dependency on the tacit knowledge of multiple experts and the inherent uncertainty in the available supporting data characterizes the risk assessment in materials discovery as a Knowledge-intensive process [2]. Assessing data availability, quantity, and quality in this context is essential. We represent this assessment through the so-called Level-of-Knowledge (LOK) metric (see [9]). Methods that don't correctly handle all this complexity often lead to biased and inconsistent assessments. We apply the Knowledge-augmented Risk Assessment (KaRA) framework to tackle this challenge.

3.3.1 Knowledge-augmented Risk Assessment (KaRA) Framework

The KaRA framework combines multiple AI techniques that consider SMEs' feedback on top of a structured domain knowledge base to support the risk assessment processes of candidates in knowledge-intensive contexts. This framework is the extension of the geological success of prospects (see [13], and [16]). The recent advances in AI and Knowledge Engineering provided the basis for the development of KaRA. Knowledge engineering practices and technologies are applied to represent and integrate the domain knowledge from multiple data sources and stakeholders and provide easy access to this knowledge through the DWb platform. At the same time, AI provides the appropriate tools for inference, prediction, uncertainty reduction, consensus reaching, etc.

The methodology implemented by KaRA divides the assessment into multiple risk factors. Individual risk factor evaluation encompasses three phases: risk factor characterization, LOK assessment, and POS assessment. Different AI agents support expert decisions at each one of them. As mentioned, the domain knowledge is structured and integrated by the KB managed by the DWb. The KB also handles all the data used by KaRA components. KaRA aims to enable a co-creation environment for SMEs and AI agents to support the risk assessment process. Figure 5 summarizes the architecture, the methodology workflow, and some of the main features of KaRA. We discuss those features in the sequence according to each phase.

Risk Factor Characterization We adopted a similar approach to [8] and [5] in the risk factor characterization phase. For each risk factor, experts should answer a standard questionnaire showcasing the data analysis workflow results that could influence LOK or POS assessments. The characterization questionnaires structure the candidates in the KB, providing an easy way to represent them for supervised and unsupervised machine learning algorithms. One representative example used in the following phases is retrieving similar candidates in terms of characterization for comparison purposes. The differentiator of KaRA will be the ability to query and reason through the KB to retrieve and rank relevant evidence that may help the experts answer the characterization questions. The expert will curate the evidence recovered from the KB, keeping only the ones relevant to the answer.

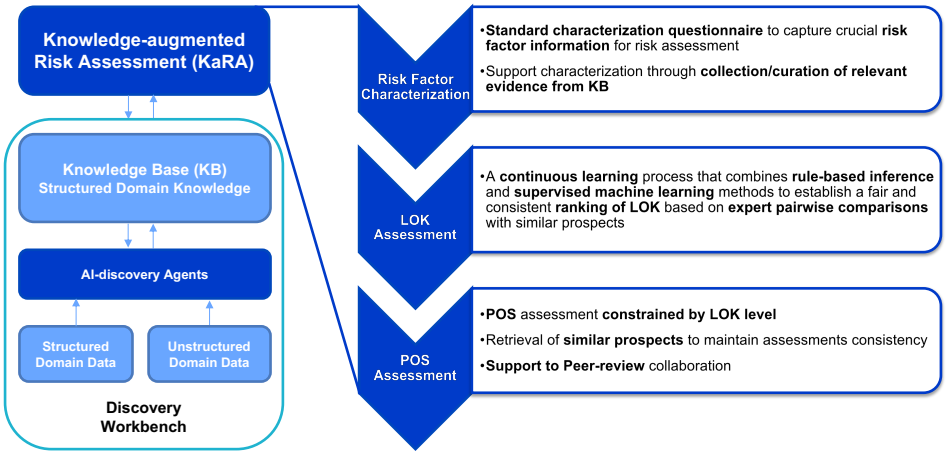


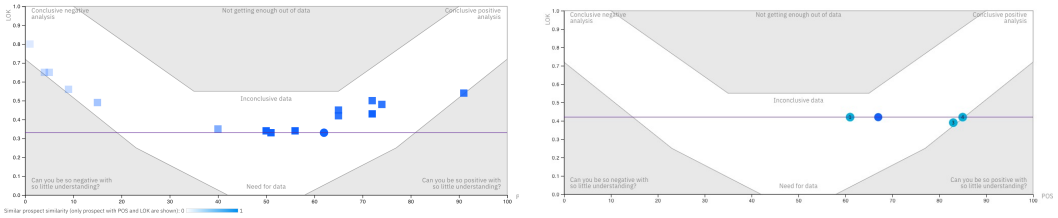
Figure 5: Overview of KaRA – methodology workflow and features.

This feedback feeds the ranking algorithms used in the evidence retrieval, which are continuously improved. This mechanism establishes a co-creation process where AI and experts collaborate to accurately characterize the candidate risk factors taking advantage of available evidence stored in the KB and the expert's tacit knowledge.

LOK Assessment The role of the *Level-of-Knowledge* (LOK) assessment phase is to establish fair and consistent LOK score scales dependent on the risk factor characterizations produced in the previous step to support the POS assessment process, such as done in [6] and [11]. The LOK scales are created based on pairwise comparisons made by the experts and previously assessed, and peer-reviewed candidates present in KB. These candidates serve as training examples for a supervised machine learning algorithm that gives an initial LOK estimate, the so-called reference LOK. Experts' pairwise LOK comparisons calibrate this estimate by solving a Linear Programming (LP) model that determines the smallest overall adjustments in the LOK values needed to respect the comparisons. A rule-based inference strategy maintains the consistency of comparisons for a particular individual. The comparisons at the individual level capture the ranking opinions of each expert. The reference LOK estimates calibrated by these comparisons create the expert LOK scale. The system also determines the subset of consistent LOK comparisons that best represent experts' consensus by solving an Integer Programming (IP) model that minimizes the conflicts between the output comparisons and the overall expert opinions. We obtain the so-called global LOK scale when these comparisons calibrate the reference LOK estimates with the same LP model previously described. Additional candidates and comparisons continuously adjust expert and global LOK scales. This approach is another example of a human-AI co-creation strategy applied to support an essential step in the discovery process.

POS Assessment In the last phase, experts' should collaborate to give a final POS value to the risk factor of the target candidate. The process collects individual POS assessments of multiple experts, and then a peer-review evaluation is conducted to reach a consensus value. Both steps constrain the experts' possible POS assessment values given a certain LOK level. The specification of how the LOK constrains the POS values is part of the specialization of the framework. KaRA's default method uses the so-called confidence-likelihood plot, a strategy devised in the oil and gas industry (see [11] and [6]). In this approach, POS values are constrained to the middle of the scale (around fifty percent) when the LOK is low, while at high LOK levels, the POS value should be either very low (around zero percent) or very high (around a hundred percent). The rationale is that in a low confidence situation, one should not be conclusive about the POS value, while in high confidence cases, one must be very assertive. The plots shown in Figure 6 specify this relationship by delimiting a region (in white) of allowed LOK (y-axis) and POS (x-axis) assessment pairs.

Figure 6a shows the supporting plot for the individual expert's POS assessment step, while Figure 6b presents the one for the peer-review POS assessment. In Figure 6a, we use the expert's LOK score to position the line on the y-axis. The expert is then constrained to give a POS value in the



(a) Expert POS assessment

(b) Peer-review POS assessment

Figure 6: Confidence-likelihood plots.

intersection between this line and the white region. The blue squares represent the assessments of similar candidates, where the blueness is proportional to the similarity with the target candidate. They serve as a reference to possibly maintain consistency with prior assessments. In Figure 6b, the global LOK score determines the position of the LOK line, and the circles represent the multiple assessments of the different experts. With both information, the peer-review step may reach a final consensus driven by the multiple experts' opinions, potentially decreasing the biases in the process or at least making it more explicit.

At the end of the process, the evaluated candidates will have a detailed assessment of each dimension of interest (risk factors). The generative models may use these results to calibrate their loss functions to increase the probability of generating new candidates with higher potential POS for each dimension. The complete workflow of the proposed materials discovery methodology establishes a continuous learning process through the collaboration and orchestration of AI agents and SMEs.

4 Final Remarks

In this paper we discussed some of the challenges and opportunities faced when we try to bring AI to support the end-to-end discovery process. We also introduced our ongoing work to create a method and its supporting technology that enable an efficient and effective collaboration between the SMEs and AI tools throughout the discovery process, applying a Human-AI co-creation approach. The proposed approach recognizes the role of the SMEs as the main owner and decision-maker of the discovery process, aiming at providing AI-based technology to assist the SMEs in their abductive reasoning and to allow the capture of the SME's knowledge for further reuse.

To develop and evaluate this approach, we have been working on a software framework called *Discovery Workbench* (DWb) that enables the construction of end-to-end discovery applications. DWb offers mechanisms to integrate multiple discovery tools and to support common discovery workflows by means of an integrated and user-friendly web-based interface. DWb provides systematization to support experts with the discovery process and tracks the knowledge that is created along that process throughout its entire life cycle, from hypothesis generation and collection (*i.e.* material candidates), to hypothesis testing (*i.e.* automated simulations and lab experiments), and then to drawing conclusions (*i.e.* visualizations and comparisons).

We have already created a PAG discovery application based on DWb, with promising initial results in terms of acceleration and systematization of the discovery process, mainly due to more flexible system integration and better knowledge management. We have also started to explore other use cases to test the generality of the approach, such as meta-organic frameworks and biomass materials.

References

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–18. Association for Computing Machinery, 2018.

[2] Claudio Di Ciccio, Andrea Marrella, and Alessandro Russo. Knowledge-intensive processes: Characteristics, requirements and analysis of contemporary approaches. *Journal on Data Semantics*, 4:29–57, 3 2015.

[3] Claudio Di Ciccio, Andrea Marrella, and Alessandro Russo. Knowledge-Intensive Processes: Characteristics, Requirements and Analysis of Contemporary Approaches. 4(1):29–57, 2015.

[4] Samuel C. Hoffman, Vijil Chenthamarakshan, Dmitry Yu Zubarev, Daniel P. Sanders, and Payel Das. Sample-Efficient Generation of Novel Photo-acid Generator Molecules using a Deep Generative Model. (arXiv:2112.01625), 2021.

[5] K Jan-Erik, C Bundiit, and PC Simplicil. The CCOP guidelines for risk assessment of petroleum prospects. *Professional Paper of Coordinating Committee for Offshore Prospecting in Asia*, pages 4–21, 2000.

[6] DC Lowry, RJ Suttill, and RJ Taylor. Advances in risking exploration prospects. *The APPEA Journal*, 45(1):143–158, 2005.

[7] Łukasz Maziarka, Agnieszka Pocha, Jan Kaczmarczyk, Krzysztof Rataj, Tomasz Danel, and Michał Warchoł. Mol-CycleGAN: A generative model for molecular optimization. 12(1):2, 2020.

[8] Alexei V Milkov. Risk tables for less biased and more consistent estimation of probability of geological success (pos) for segments with conventional oil and gas prospective resources. *Earth-Science Reviews*, 150:453–476, 2015.

[9] Nicola Paltrinieri, Louise Comfort, and Genserik Reniers. Learning about risk: Machine learning for risk assessment. *Safety science*, 118:475–486, 2019.

[10] Edward O. Pyzer-Knapp, Jed W. Pitera, Peter W. J. Staar, Seiji Takeda, Teodoro Laino, Daniel P. Sanders, James Sexton, John R. Smith, and Alessandro Curioni. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. 8(1):84, 2022.

[11] Peter R Rose. *Risk analysis and management of petroleum exploration ventures*, volume 12. American Association of Petroleum Geologists Tulsa, OK, 2001.

[12] Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. GraphAF: A Flow-based Autoregressive Model for Molecular Graph Generation. (arXiv:2001.09382), 2020.

[13] F. Silva, S. Fernandes, J. Casacão, C. Libório, J. Almeida, S. Cersósimo, C.R. Mendes, R. Brandão, and R. Cerqueira. Machine-learning in oil and gas exploration: A new approach to geological risk assessment. 2019(1):1–5, 2019.

[14] Changwon Suh, Clyde Fare, James A. Warren, and Edward O. Pyzer-Knapp. Evolving the materials genome: How machine learning is fueling the next generation of materials discovery. 50(1):1–25, 2020.

[15] Girmaw Abebe Tadesse, Jannis Born, Celia Cintas, Matteo Manica, and Kommy Weldemariam. Mpego: A toolkit for multi-level performance evaluation of generative models for material discovery domains. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022.

[16] E. Vital Brazil, R. Ferreira, V. Torres, L. Martins, C. Raoni, R. Cerqueira, A. Fernandez, J. Almeida, S. Fernandes, B. Carvalho, J. Casacão, C. Libório, M. Quintela, M. Ramos, D. Patrocinio, M. Ferraz, and S. Cersosimo. GSA – GeoScience Advisor. 2021(1):1–5, 2021.