

# Weekly Status Report

UGP1 | MSE 496



# Review of Goals

 **DONE**

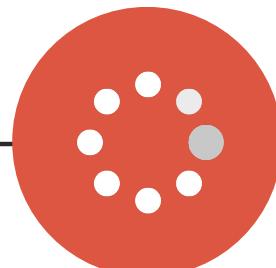
 **ONGOING**



**01**

## PAPER REVIEW

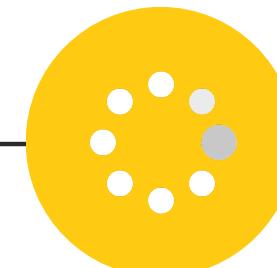
To find and document academic papers relevant to atomic layer deposition and the use of deep learning in the same.



**02**

## DATASET COLLATION

To find an appropriate dataset for the task and use information retrieval strategies to collate information from the dataset.



**03**

## PIPELINE SHORTLISTING

To shortlist a set of models that can be used to build the final product.

# Goal # 1

## PAPER REVIEW

### GENERAL STATUS

#### PROGRESS CHECK

- Read a review paper on ALD, to get an understanding of the scientific principles behind the entire process.
- Read a review paper on the use of Gen AI in materials science.
- Found few papers that have used DL/GenAI on related materials science topics before.
- Reviewed the efficacy of LLaMA-MAT-Chat on the task at hand and some general queries.
- Decided a tentative data ingestion pipeline for dataset creation and usage in downstream tasks.

#### ACTION ITEMS

- Finalize the data ingestion pipeline and test the nuances.
- Of the papers reviewed, finalize a pipeline for the final product and start working on it.

## Primer



# Atomic layer deposition

Erwin Kessels<sup>1</sup>✉, Anjana Devi<sup>2,3,4</sup>, Jin-Seong Park<sup>5</sup>, Mikko Ritala<sup>6</sup>, Angel Yanguas-Gil<sup>7</sup> & Claudia Wiemer<sup>1,8</sup>

## Abstract

Atomic layer deposition (ALD) is a surface-controlled chemical vapour deposition method, in which materials are prepared one atomic layer at the time. With ALD, film thickness can be controlled very precisely, and it allows the user to cover large areas and surfaces with a complex three-dimensional structure uniformly and conformally. ALD is used for the deposition of high-quality thin films and nanostructures, as well as for surface functionalization and interface engineering in a wide range of applications, both from a research and development perspective, as well as for high-volume manufacturing. This Primer outlines the method of ALD, describing the precursors, coreactants

## Sections

Introduction

Experimentation

Results

Applications

Reproducibility and  
data deposition

Limitations and optimizations

Outlook



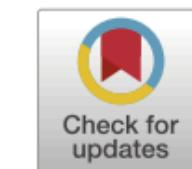
Contents lists available at [ScienceDirect](#)

## Journal of Materiomics

journal homepage: [www.journals.elsevier.com/journal-of-materiomics/](http://www.journals.elsevier.com/journal-of-materiomics/)



# Generative artificial intelligence and its applications in materials science: Current situation and future perspectives



Yue Liu <sup>a, d</sup>, Zhengwei Yang <sup>a</sup>, Zhenyao Yu <sup>a</sup>, Zitu Liu <sup>a</sup>, Dahui Liu <sup>a</sup>, Hailong Lin <sup>b</sup>,  
Mingqing Li <sup>b</sup>, Shuchang Ma <sup>a</sup>, Maxim Avdeev <sup>e, f</sup>, Siqi Shi <sup>b, c, \*</sup>

<sup>a</sup> School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

<sup>b</sup> State Key Laboratory of Advanced Special Steel, School of Materials Science and Engineering, Shanghai University, Shanghai, 200444, China

# Goal # 2

## DATASET COLLATION

### GENERAL STATUS

#### PROGRESS CHECK

- Found a relevant website for collecting data regarding ALD.
- The website contains extremely detailed ALD details of a large number of materials.
- Listed a few methods to ingest the data and make a relevant dataset.
- Investigated LLaMA-Mat-Chat for research paper summarization task.

#### ACTION ITEMS

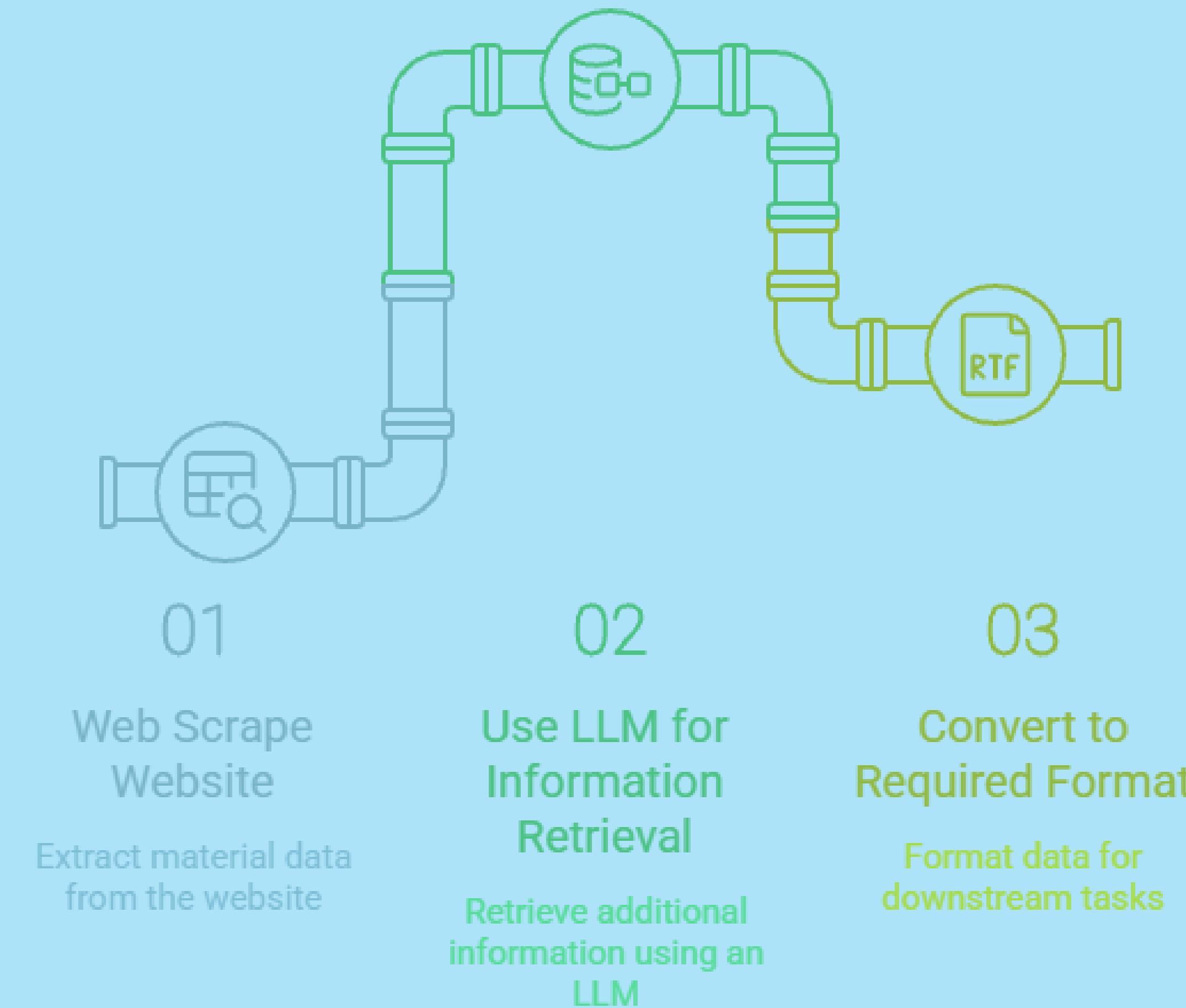
- Check the feasibility of the suggested methods and also take into considering the hardware constraints for the same.
- Investigate general purpose LLMs and other models for the summarization tasks.

Website:

[atomiclimits.com/alddatabase](http://atomiclimits.com/alddatabase)

List of processes						
Z	Material	Reactant A	Reactant B	Reactant C	Further reactants	References
3	Lithium					
	$\text{Li}_2\text{O}$	$\text{Li}(\text{O}^t\text{Bu})$	$\text{H}_2\text{O}$			<a href="#">Altonen et al.</a> <a href="#">Kozen et al.</a> ADD
		$\text{Li}(\text{O}^t\text{Bu})$	$\text{O}_2$ plasma			<a href="#">Hornsveld et al.</a> ADD
		$\text{LiN}(\text{SiMe}_3)_2$	$\text{O}_2$ plasma	$\text{H}_2$ plasma		<a href="#">Pieters et al.</a> ADD
	$\text{LiN}(\text{SiMe}_3)_2$		$\text{H}_2\text{O}$			<a href="#">Pieters et al.</a> ADD
	$\text{Li}_2\text{S}$	$\text{Li}(\text{O}^t\text{Bu})$	$\text{H}_2\text{S}$			<a href="#">Meng et al.</a> ADD
	$\text{Li}_3\text{N}$	$\text{LiN}(\text{SiMe}_3)_2$	$\text{NH}_3$			<a href="#">Østreng et al.</a> ADD
	$\text{LiF}$	$\text{Li}(\text{O}^t\text{Bu})$	$\text{TiF}_4$			<a href="#">Xie et al.</a> <a href="#">Tiurin et al.</a> ADD
		$\text{Li}(\text{O}^t\text{Bu})$	$\text{HF}$			<a href="#">Chen et al.</a> ADD
		$\text{Li}(\text{O}^t\text{Bu})$	$\text{NH}_4\text{F}$			<a href="#">Kvalvik et al.</a> ADD
		$\text{Li}(\text{thd})$	$\text{TiF}_4$	$\text{Mg}(\text{thd})_2$		<a href="#">Mäntymäki et al.</a> ADD
		$\text{Li}(\text{thd})$	$\text{TiF}_4$			<a href="#">Mäntymäki et al.</a> ADD

# Data Ingestion Pipeline



Aaltonen *et al.* Kozen *et al.*

a.processList--link 109.53 × 21

Hornsvedt *et al.* ADD

Pieters *et al.* ADD

Pieters *et al.* ADD

Meng *et al.* ADD

Østreng *et al.* ADD

Xie *et al.* Tiurin *et al.* ADD

Chen *et al.* ADD

Kvalvik *et al.* ADD

Mäntymäki *et al.* ADD

Mäntymäki *et al.* ADD

Lee *et al.* ADD

Li <sub>2</sub> O	Li(O <sup>t</sup> Bu)	H <sub>2</sub> O	
			Aaltonen <i>et al.</i> Kozen <i>et al.</i>
			a.processList--link 109.53 × 21
			Hornsvedt <i>et al.</i> ADD
			Pieters <i>et al.</i> ADD
			Pieters <i>et al.</i> ADD
			Meng <i>et al.</i> ADD
			Østreng <i>et al.</i> ADD
			Xie <i>et al.</i> Tiurin <i>et al.</i> ADD
			Chen <i>et al.</i> ADD
			Kvalvik <i>et al.</i> ADD
			Mäntymäki <i>et al.</i> ADD
			Mäntymäki <i>et al.</i> ADD
			Lee <i>et al.</i> ADD

Elements Console Sources Network Performance Memory Application Privacy and security Lighthouse >

```
<!DOCTYPE html>
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> scroll
  <head>@@</head>
  <body cz-shortcut-listen="true">
    <noscript>You need to enable JavaScript to run this app.</noscript>
    <div id="root">
      <div class="app">
        <div class="table">@@</div> grid
        <br>
        <div class="processList">
          <div class="processList--title">@@</div> grid
          <table class="processList--table">
            <tbody>
              <tr class="header">@@</tr>
              <tr class="processList--subtitle">@@</tr>
              <tr title class>@@</tr>
              <tr title class>
                <td></td>
                <td></td>
                <td>@@</td>
                <td>@@</td>
                <td></td>
                <td></td>
                <td></td>
                <td>@@</td>
                <td></td>
                <td></td>
                <td></td>
                <td></td>
                <td></td>
                <td>@@</td>
                <td><a class="processList--link" href="https://doi.org/https://doi.org/10.1039/C7RA07722J" target="_blank" rel="noopener noreferrer" style="margin-left: 0.2em;">@@</a>
                  <button class="jss92 jss129 jss131 jss132 jss134 jss135" tabindex="0" type="button"><@@</button> flex
                </td>
              </tr>
              <tr title class>@@</tr>
              <tr title class>@@</tr>
              <tr title class>@@</tr>
              <tr title class>@@</tr>
              <tr title class>
                <td></td>
              </tr>
            </tbody>
          </table>
        </div>
      </div>
    </div>
  </body>

```

## Possibility of HTML Scrapping

---

# AUTOMATED EXTRACTION OF MATERIAL PROPERTIES USING LLM-BASED AI AGENTS

---

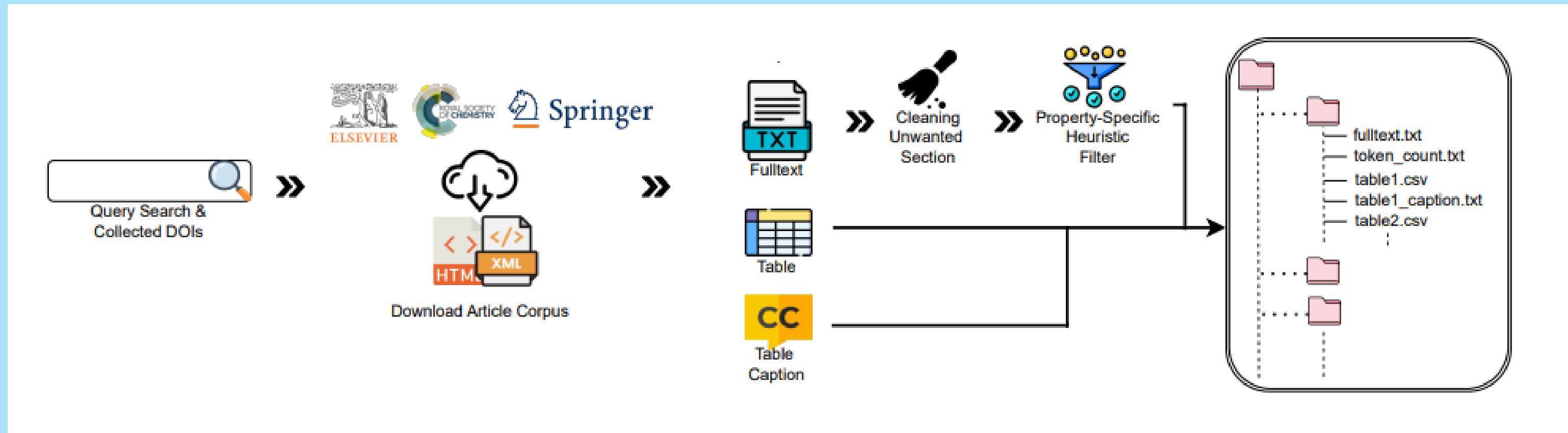
**Subham Ghosh**

Mehta Family School of Data Science and Artificial Intelligence  
Indian Institute of Technology Roorkee  
Uttarakhand, India, 247667  
subham\_g1@mfs.iitr.ac.in

**Abhishek Tewari**

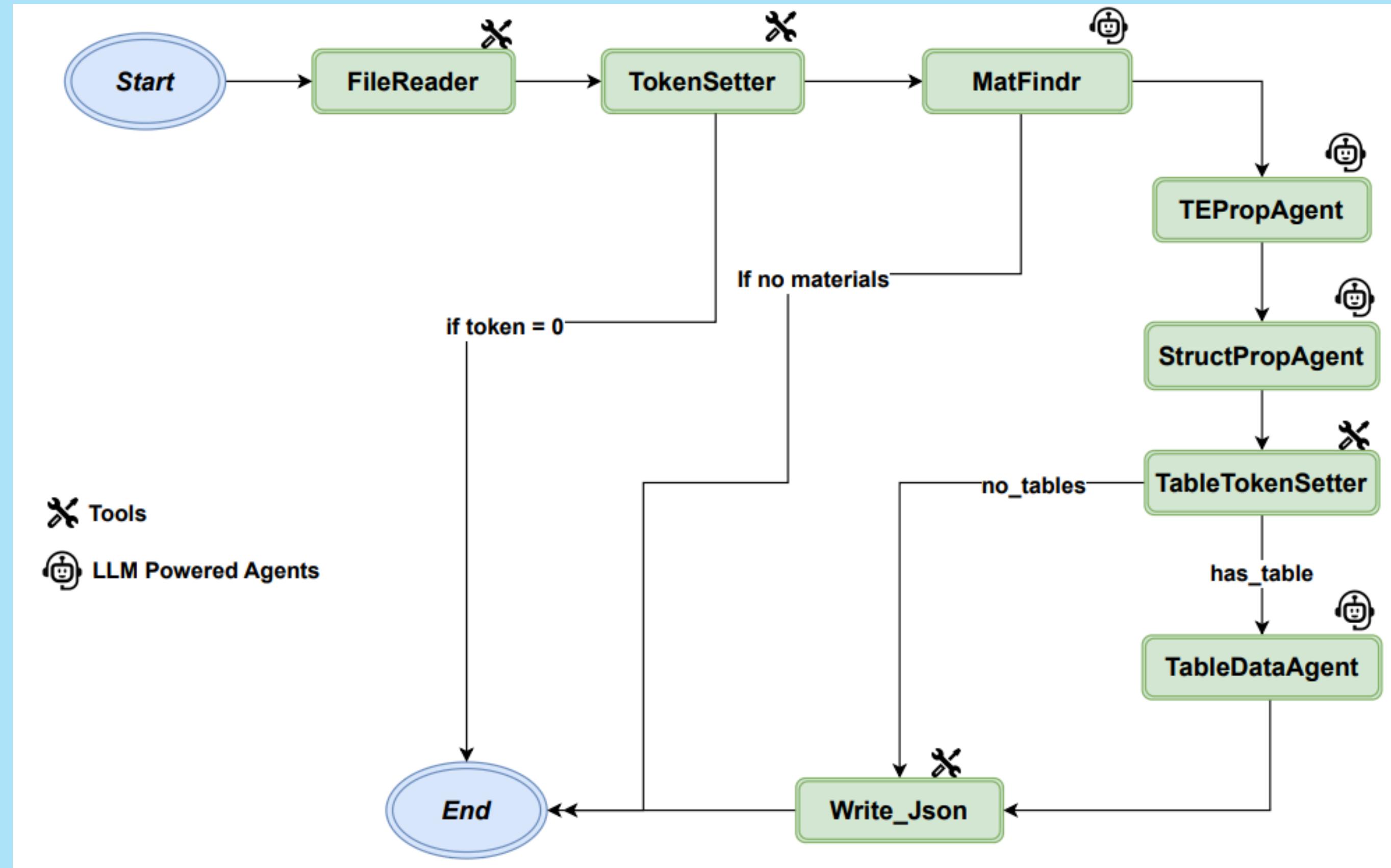
Mehta Family School of Data Science and Artificial Intelligence  
Department of Metallurgical and Materials Engineering  
Indian Institute of Technology Roorkee  
Uttarakhand, India, 247667  
abhishek@mt.iitr.ac.in

October 3, 2025



Code is available at:

[https://github.com/CMEG-IITR/Agentic\\_data\\_extraction](https://github.com/CMEG-IITR/Agentic_data_extraction)



[11]:

```
input_text = f"""
Instruction: Extract the most important information from the given text.
```

Rules:

- Only include essential details.
- Output as bullet points.
- Do not add new information.

Text:

```
{text}
```

Answer:

```
"""
```

Code

Markdown

```
In[3]: input_ids = tokenizer(input_text, return_tensors="pt").to("cuda")

outputs = model.generate(
    **input_ids,
    max_new_tokens=256
)

generated_ids = outputs[0][input_ids["input_ids"].shape[-1]:]
outs = tokenizer.decode(generated_ids, skip_special_tokens=True)

print(outs)
```

Setting `pad\_token\_id` to `eos\_token\_id`:2 for open-end generation.

- ALD is a subset of CVD methods
- ALD films possess excellent uniformity and unparalleled conformality
- ALD can be used for surface functionalization, interface engineering, nanoparticles fabrication and precise nanoengineering of materials
- ALD takes place by repeating cycles, each adding the same amount of material, which is typically less than a monolayer
- ALD can be represented as ABABAB or simply as AB-type
- The reactant in the reaction step A is typically called the 'precursor' , and it introduces the main element of the material to the film surface. The reactant in step B is often labelled as the 'coreactant' , and it can add a second element to the material, giving the possibility of transforming the main element of the film into an oxide, nitride, sulfide, fluoride or other compounds as being targeted.

```
▶ input_ids = tokenizer(input_text, return_tensors="pt").to("cuda")

outputs = model.generate(
    **input_ids,
    max_new_tokens=256
)

generated_ids = outputs[0][input_ids["input_ids"].shape[-1]:]
outs = tokenizer.decode(generated_ids, skip_special_tokens=True)

print(outs)
```

Setting `pad\_token\_id` to `eos\_token\_id`:2 for open-end generation.

- Lithium carbonate is considered a potential electrode passivating film in Li-ion batteries and electrolyte material or sensing layer in electrochemical devices like fuel cells or chemical sensors.
- Atomic layer deposition (ALD), which is based on sequential and self-limiting half-reactions between precursors (co-reactants) and surface, has emerged as a powerful tool since it shows potential towards exceptional conformality on high-aspect ratio structures, thickness control at sub-nanometer level, and tunable film properties.
- Plasma-assisted and thermal ALD were adopted to grow ultra-thin, conformal Li<sub>2</sub>O<sub>3</sub> films between 50

# Goal # 3

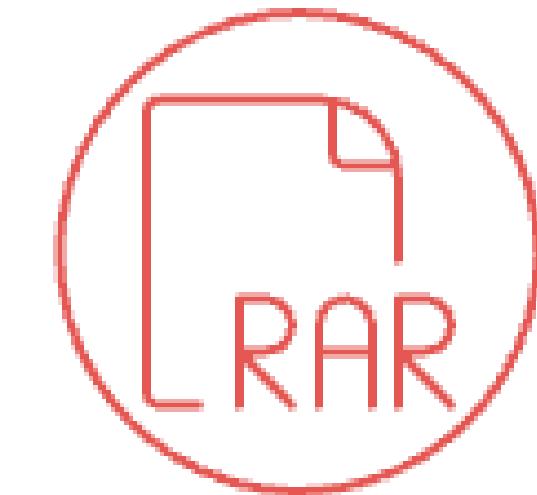
## FINAL PIPELINE DESIGN

GENERAL STATUS



RAG with LLaMA-Mat-Chat

Easier to implement



QLoRA/LoRA Adaptation

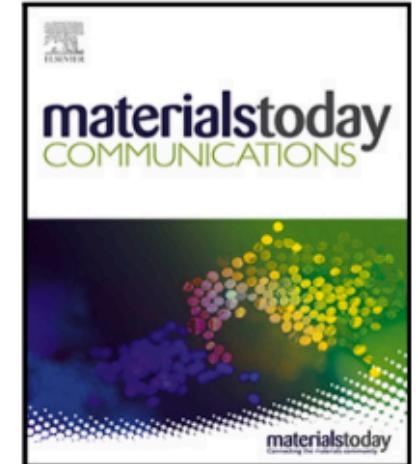
Resource intensive



Contents lists available at [ScienceDirect](#)

## Materials Today Communications

journal homepage: [www.elsevier.com/locate/mtcomm](http://www.elsevier.com/locate/mtcomm)



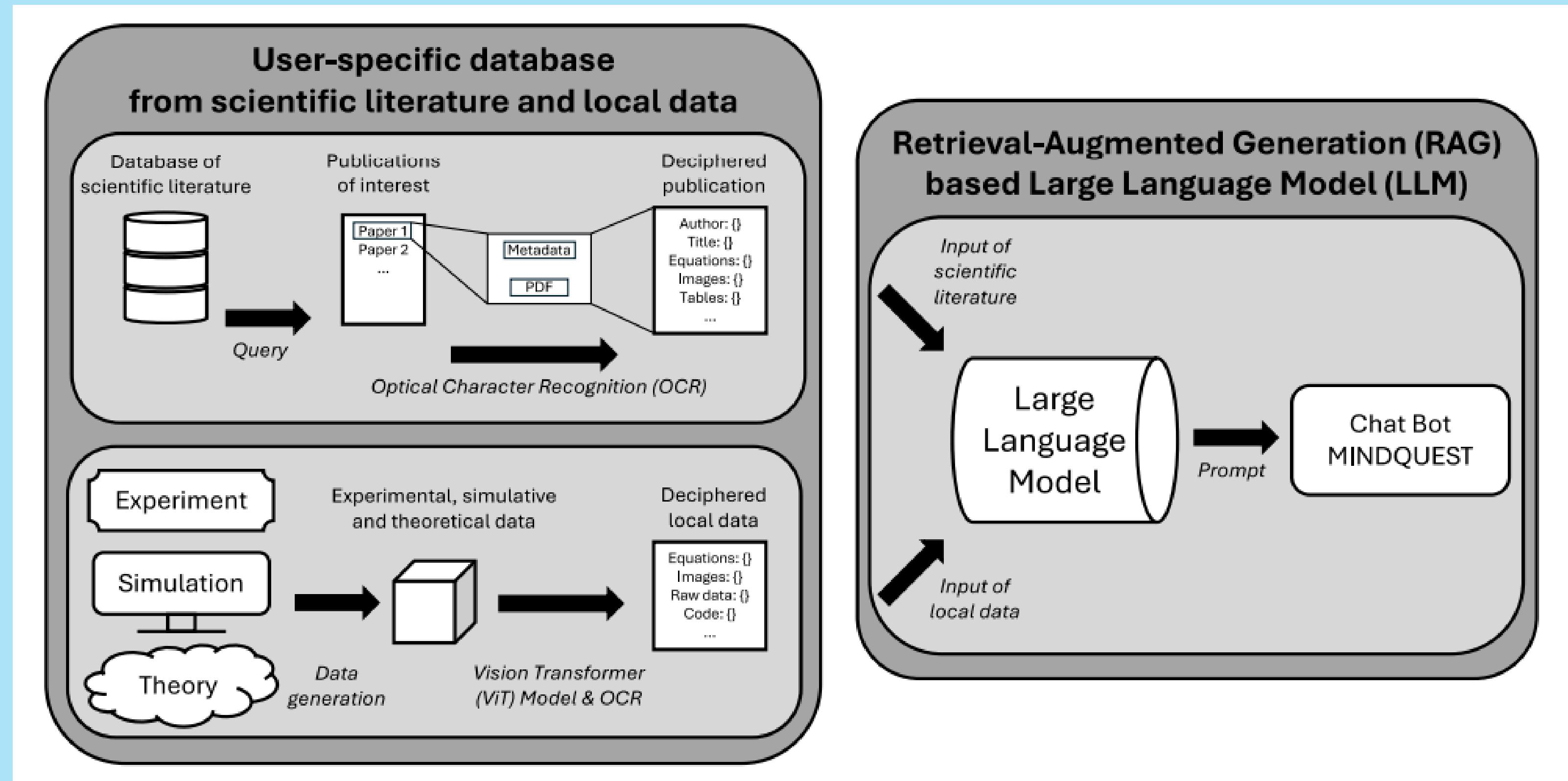
# Towards an automated workflow in materials science for combining multi-modal simulation and experimental information using data mining and large language models

Balduin Katzer <sup>a,b</sup>,\*, Steffen Klinder <sup>a</sup>, Katrin Schulz <sup>a,b</sup>,\*

<sup>a</sup> Karlsruhe Institute of Technology (KIT), Institute for Applied Materials (IAM), Kaiserstr. 12, Karlsruhe, 76131, Germany

<sup>b</sup> Karlsruhe University of Applied Sciences (HKA), Moltkestr. 30, Karlsruhe, 76133, Germany





# LLMATDESIGN: AUTONOMOUS MATERIALS DISCOVERY WITH LARGE LANGUAGE MODELS

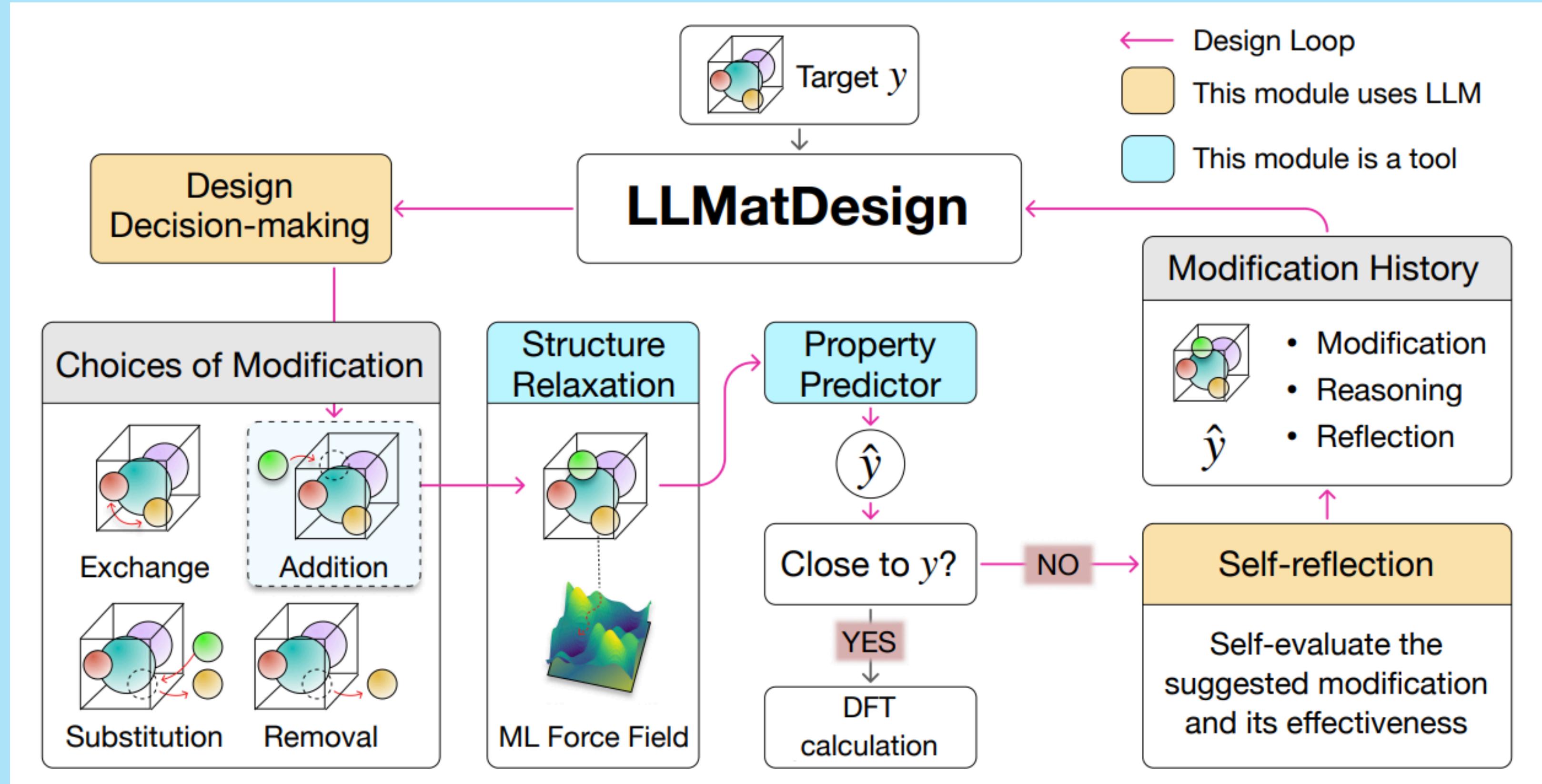
Shuyi Jia<sup>†</sup>, Chao Zhang<sup>†</sup>, Victor Fung<sup>†§</sup>

<sup>†</sup>Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

<sup>§</sup>Corresponding author: [victorfung@gatech.edu](mailto:victorfung@gatech.edu)

## Abstract

Discovering new materials can have significant scientific and technological implications but remains a challenging problem today due to the enormity of the chemical space. Recent advances in machine learning have enabled data-driven methods to rapidly screen or generate promising materials, but these methods still depend heavily on very large quantities of training data and often lack the flexibility and chemical understanding often desired in materials discovery. We introduce LLMatDesign, a novel language-based framework for interpretable materials design powered by large language models (LLMs). LLMatDesign utilizes LLM agents to translate human instructions, apply modifications to materials, and evaluate outcomes using provided tools. By incorporating self-reflection on its previous decisions, LLMatDesign adapts rapidly to new tasks and conditions in a zero-shot manner. A systematic evaluation of LLMatDesign on several



**Thank you**