

# Class 1: Materials Informatics: Introduction to Materials Data Science

**Pawan K. Tripathi**

Research Assistant Professor Case Western Reserve University, Cleveland, Ohio, USA

---

## Today's Agenda

1. **Welcome & Course Overview**
  2. **What are “Materials”?** (Conceptual View)
  3. **What is Data Science & AI?** (Review & Context)
  4. **Why Combine Them? Introduction to Materials Informatics (MI)**
  5. **MI Workflow Overview**
  6. **Q&A and Next Steps**
- 

## Course Goals

- Understand **what** Materials Informatics (MI) is and **why** it's important.
- Learn about the **types of data** used in MI.
- Explore how **data science and machine learning** are applied to materials problems.
- Get hands-on experience with basic **MI tools and workflows** (using Python).
- Appreciate the **challenges and future potential** of the field.

**No prior Materials Science background is required!** We focus on the data perspective.

---

## Course Overview

This course provides an introduction to the emerging field of Materials Informatics, which applies data science and machine learning techniques to accelerate the discovery, design, and development of new materials. Students will learn how to acquire, analyze, and model materials data to solve real-world materials science problems.

## Course Objectives

Upon successful completion of this course, students will be able to:

- Understand the fundamental concepts of Materials Informatics.
- Apply data mining and machine learning techniques to materials data.
- Utilize Python libraries for materials data analysis and modeling.
- Design and execute materials discovery workflows.
- Evaluate the performance of machine learning models in materials science.
- Communicate materials informatics findings effectively.
- Finish a project leveraging deep learning, graph learning for the materials image or tabular data ## Course Logistics
- **Format:** 6 remote sessions (90 mins each) via [Zoom]
- **Tools:** Google Colab (free, web-based Python environment)
- **Communication:** [Platform Name, e.g., Slack, Email List] for questions & announcements
- **Materials:** Slides & Colab notebooks provided before class
- **Expectations:** Active participation, willingness to learn basic Python/data concepts.

---

## Part 1: What are “Materials”? (The Conceptual View)

- Think beyond just “stuff”. Materials are substances engineered for specific **properties** and **applications**.
- **Examples:**
  - **Strong but light metals:** Aerospace (Aluminum alloys)
  - **Efficient energy converters:** Solar cells (Silicon, Perovskites)
  - **Flexible conductors:** Wearable electronics (Conducting polymers)
  - **Biocompatible implants:** Medical devices (Titanium, Polymers)
  - **Hard coatings:** Cutting tools (Titanium Nitride)
- **Traditional Goal of Materials Science:** Design and discover new materials with desired properties through experimentation and physical understanding.



##### (Source:Hitachi)

---

### Categories of Materials (High Level)

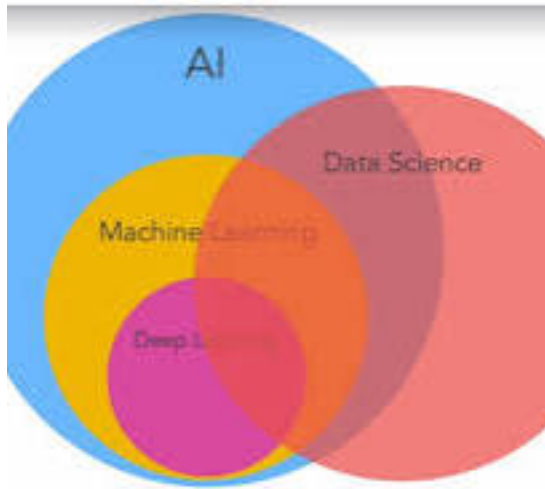
- **Metals:** Strong, conduct electricity/heat (e.g., Steel, Copper, Aluminum)
- **Ceramics:** Hard, heat-resistant, insulators (e.g., Glass, Concrete, Alumina)
- **Polymers (Plastics):** Lightweight, flexible, insulators (e.g., Polyethylene, PVC, Nylon)
- **Composites:** Combinations for unique properties (e.g., Carbon fiber reinforced polymer)

**Key Idea:** The arrangement of atoms and molecules dictates a material's properties!

---

### Part 2: What is Data Science & AI? (Review & Context)

- **Data Science:** The practice of extracting knowledge and insights from data.
- **Artificial Intelligence (AI):** Creating systems that can perform tasks typically requiring human intelligence.
- **Machine Learning (ML):** A subset of AI where systems learn patterns from data without being explicitly programmed for each task.



---

### Key Data Science Activities

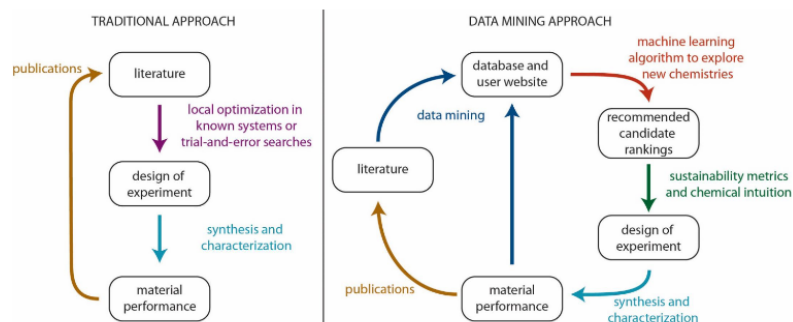
- **Data Collection:** Gathering raw information.
- **Data Cleaning/Preprocessing:** Fixing errors, handling missing values, formatting data. (Often 80% of the work!)
- **Exploratory Data Analysis (EDA):** Understanding data through statistics and visualization.
- **Modeling (Machine Learning):** Building predictive models (e.g., predicting a property).
- **Visualization:** Communicating insights effectively through charts and graphs.
- **Deployment:** Putting models into practice.

**Key Idea:** Data science provides the **tools** to analyze complex information and make predictions.

---

### Part 3: Why Combine Materials Science and Data Science?

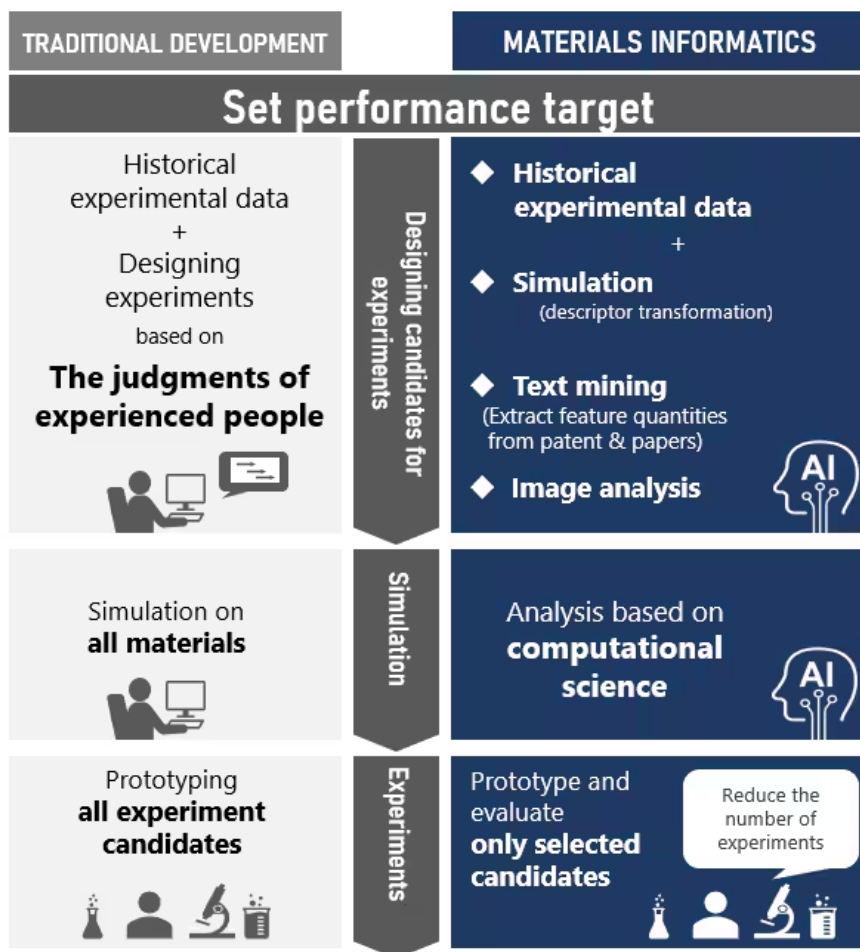
- **The Challenge:** Traditional materials discovery is often:
  - **Slow:** Relies on intuition and many experiments.
  - **Expensive:** Lab equipment, materials, time.
  - **Complex:** Huge number of possible material combinations. (Imag-



ine trying all possible recipes!)

## Enter: Materials Informatics (MI)

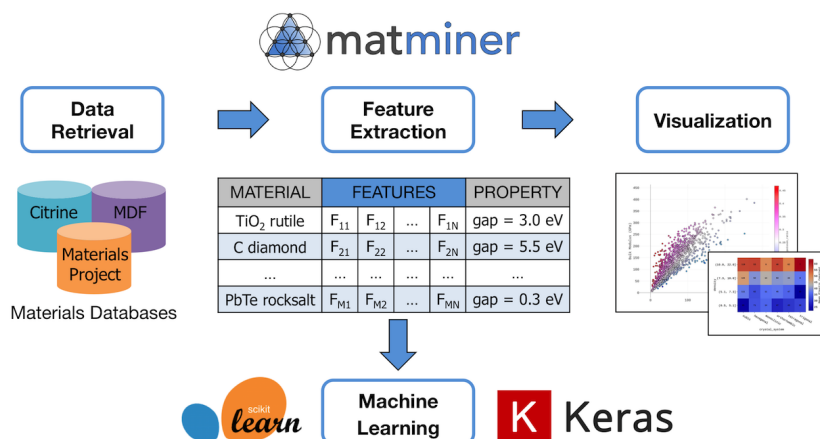
- **Definition:** An emerging field that applies the principles of **data science** and **AI** to solve problems in **materials science and engineering**.
- **Goal:** Accelerate the discovery, design, and deployment of new materials.
- Inspired by concepts like the **Materials Genome Initiative (MGI)**: Aiming to halve the time and cost of bringing new materials to market.



##### (Source: <https://www.hitachi-hightech.com/us/en/products/ict-solution/randd/mi/>)

### What Can MI Do? Examples:

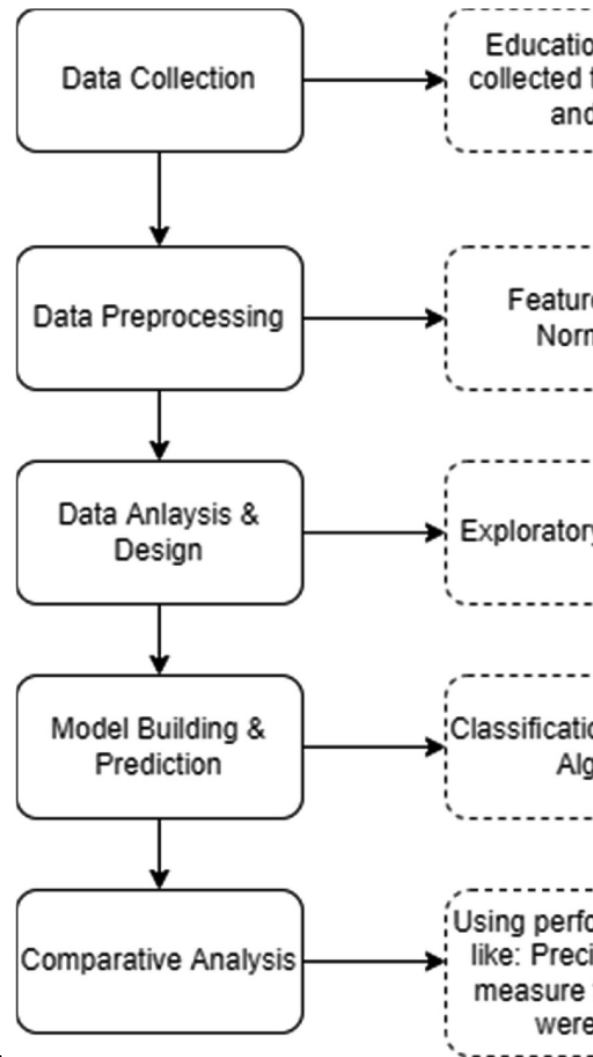
- **Predict Properties:** Estimate a material's hardness, melting point, conductivity, or stability *before* making it in the lab, based only on its composition or structure.
- **Discover New Candidates:** Search vast databases of potential materials to find promising candidates for a specific application (e.g., a better battery material).
- **Optimize Processes:** Use data from manufacturing to improve material quality or reduce waste.
- **Extract Knowledge:** Analyze scientific literature automatically to find



trends or hidden correlations.  
##### (Source: Matminer) —

## Part 4: A Simple MI Workflow

- Data Acquisition:**
  - Get data from experiments, simulations (computational chemistry/physics), databases (like Materials Project), or literature.
  - Challenge: Data can be sparse, noisy, heterogeneous.*
- Data Representation / Featurization:**
  - Convert material information (composition, structure) into numerical features computers can understand.
  - Challenge: How to best represent complex material characteristics?*
- Model Building (Machine Learning):**
  - Train ML models (regression, classification) to learn relationships between features and properties.
  - Challenge: Choosing the right model, avoiding overfitting.*
- Prediction / Discovery:**
  - Use the model to predict properties of new/unseen materials or screen candidates.
  - Challenge: How reliable are the predictions?*
- Validation:**
  - Test the most promising predictions through targeted experiments or simulations.



- *Challenge: Closing the loop back to the real world.*  
 \*\*\* (Source:[https://link.springer.com/chapter/10.1007/978-981-99-7456-6\\_4](https://link.springer.com/chapter/10.1007/978-981-99-7456-6_4))

---

## Part 5: Q&A and Next Steps

- **Any Questions?**
- **Preparation for Next Class:**
  - Ensure you can access Google Colab (<https://colab.research.google.com/>).
  - (Optional) If you're new to Python, browse a basic tutorial (links can



be provided).

- Think about: How would you represent a simple material like water ( $\text{H}_2\text{O}$ ) or salt ( $\text{NaCl}$ ) as data for a computer?
- **Next Time:** Diving deeper into representing materials data and features.

---

**Thank You!**