

École polytechnique de Louvain

# APPLICATION OF LLMS AND DIGITIZATION IN MACHINE LEARNING FOR THERMOMAGNETIC MATERIALS

Authors : Theodore Park  
Supervisors : **Gian-Marco Rignanese, Hongbin Zhang**  
Readers : **Flavio Abreu Araujo**  
Academic year 2023–2024  
Master [120] in Chemical and Materials Engineering

---

## Abstract

---

This thesis investigates the use of large language models (LLMs) and machine learning (ML) to predict new thermomagnetic materials. LLMs offer the opportunity to locate relevant plots in journals and papers (Polak and Morgan, 2024) as well as extract the relevant context, improving the process of data collection for ML. Combined with an automated digitization this process allows for the creation of a database which is then utilized for machine learning to predict new thermomagnetic materials. Additionally, this research allows for more easily performed machine learning by reducing the work and effort required in data collection as this methodology and code can be modified for other desired properties.

Several different LLM models such as ChatGPT (2, 4), Mistral and Llama (2, 3), are tested, with Llama 3 8B instruct being the most successful. As a preliminary test, 15 papers in portable document format (PDF) published from 2010 to 2024 were tested, using LLMs to find relevant figures and context within the paper surrounding the plots. To properly test the LLM recognition 15 of these papers contain thermomagnetic plots within them, while 5 of them do not. Additionally, 16 documents without any relevant figures were tested. Relevant plots were digitized, exploring different digitization methods with modified digitization software such as code based on Plot2Spectra, ChartDETE and EasyOCR. While the results are promising further development is required.

ChatGPT was used to help fix errors and assist with coding. In addition it was utilized to check grammar and spelling during the writing of this thesis.

---

# 1 Table of Contents

---

Abstract	i
1 Table of Contents	ii
2 Introduction	1
3 Theory	3
3.1 Thermomagnetism	3
3.2 Machine Learning (ML)	5
3.3 Theory of Natural language processing	7
3.4 Neural networks	8
3.5 Theory behind Transformers and LLMs	10
3.5.1 Parameters of LLMs	15
3.5.2 Improvements to LLMs	16
3.6 Theory of digitization	21
3.6.1 Image convolution	21
3.6.2 Optical Character Recognition (OCR) and Character-Region Awareness for Text detection (CRAFT)	23
3.6.3 You Only Look Once (YOLO)	23
3.6.4 Fully Convolutional One-Stage Object Detection (FCOS)	24
3.6.5 Semantic Segmentation and efficient residual factorized network (ERFNet)	25
3.6.6 Optical Flow for Plot2spectra	25
4 Development methodology	28
4.1 LLM testing and development	28
4.2 Digitization methods	30
5 Results	34
5.1 LLMs	36
5.1.1 GPT-4	36
5.1.2 Other models and the path to 'Llama 3 8B instruct'	41
5.1.3 'Llama 3 8B instruct'	43
5.2 Digitization	49
5.2.1 Other tested digitization methods	49
5.2.2 Plot2Spectra and Context-Aware Chart Element Detection	50
6 Discussion	58
6.1 Current progress	58
6.2 Future improvements and research	59
7 Conclusion	62
8 References	1
9 Index of abbreviations	8
10 Tables and Figures	9
11 Appendix	10
11.1 Images used for digitization & Plot2Spectra output	10
11.2 Corrected Plot2Spectra output	18

---

11.3	Basel level code i.e. code that is used at the surface	31
11.3.1	Llama 3 8b instruct	31
11.3.2	Digitizer	36
11.4	Llama 3 8b instruct output	48

---

## 2 Introduction

---

Thermomagnetic materials, which exhibit a change in magnetic properties in response to temperature variations, hold significant potential to benefit society through applications in energy-efficient technologies, such as magnetic refrigeration, temperature sensing, and energy harvesting from low-grade heat sources (Dzekan et al., 2021). The discovery of new thermomagnetic materials is therefore a crucial area of research. In this area Machine learning (ML) has emerged as a powerful tool, offering the potential to accelerate the identification and optimization of these materials (Johnson et al., 2020). However, a noteworthy bottleneck in applying ML to this field lies in the collection and digitization of relevant data, particularly thermomagnetic plots, which are essential for accurate ML modeling of thermomagnetic materials. This bottleneck has been highlighted by various researchers as a major challenge in the application of ML (Roh et al., 2019).

This thesis seeks to address this bottleneck by proposing a new solution that integrates large language models (LLMs) and advanced digitization techniques. Specifically, the research focuses on utilizing LLMs to automatically identify and extract relevant figures from scientific literature, which are then digitized into a format suitable for ML analysis (see chapter 4). The solution developed in this thesis employs different LLMs like Llama, ChatGPT and others, along with digitization tools such as Plot2Spectra, DePlot, and Curve AI.

The first step in this research involves a thorough exploration of the theoretical foundations that underpin the solution (see chapter 3). This includes an examination of the principles of thermomagnetic materials, the fundamentals of machine learning, and the theories of natural language processing (NLP), neural networks, transformers, and LLMs. A critical aspect of this investigation is understanding the theory and potential improvements to LLMs, which are key to enhancing their performance in identifying relevant data. Additionally, the thesis delves into the theory of digitization, exploring various techniques such as image convolution, optical character recognition (OCR), and advanced object detection methods like You Only Look Once (YOLO) and Fully Convolutional One-Stage Object Detection (FCOS). The digitization process is further refined through the application of optical flow techniques in the Plot2Spectra tool, which plays a central role in converting plot images into data that can be used in ML models.

Following the theoretical exploration, the thesis outlines the development methodology employed to implement and test the proposed solution (see chapter 4 and 5). This includes the testing and refinement of LLMs, with a particular focus on the path to selecting Llama 3 8B Instruct as the most suitable model for this task. The digitization methods are also rigorously evaluated, comparing various

---

approaches to identify the most effective techniques with Plot2Spectra, combined with ChartDETE and EasyOCR being the most suitable for accurately converting thermomagnetic plots into ML-ready data. The results section (chapter 5) presents the findings from these tests, highlighting the performance of different LLMs, including GPT-4 and other models, as well as the effectiveness of various digitization tools. The research demonstrates that while Llama 3 8B Instruct and Plot2Spectra, in combination with ChartDETE and EasyOCR, offer a viable solution to the data bottleneck, there are still areas where improvements are needed. These results and improvements are discussed in chapter 6, which reflects on the current progress and identifies key areas for future research and development. It emphasizes the importance of refining the digitization process and improving the reliability and accuracy of LLMs in this context. Finally, the thesis concludes with a summary of the contributions made by this research and the potential impact on the field of material science and beyond.

---

### 3 Theory

---

The rapid advancements in artificial intelligence (AI) and machine learning (ML) have had a profound impact on numerous domains, facilitating the development of sophisticated models that are capable of understanding and generating human language, recognizing complex patterns, and solving intricate problems across a range of fields (Johnson et al., 2020) for instance predicting new materials. This chapter provides a comprehensive overview of the theoretical foundations and underlying technologies employed in this thesis. By exploring the fundamental principles of thermomagnetism, natural language processing (NLP), neural networks, large language models (LLMs), transformers, digitization, image recognition tools and optical flow, a solid theoretical framework will be built to then support the research and tools presented in this work.

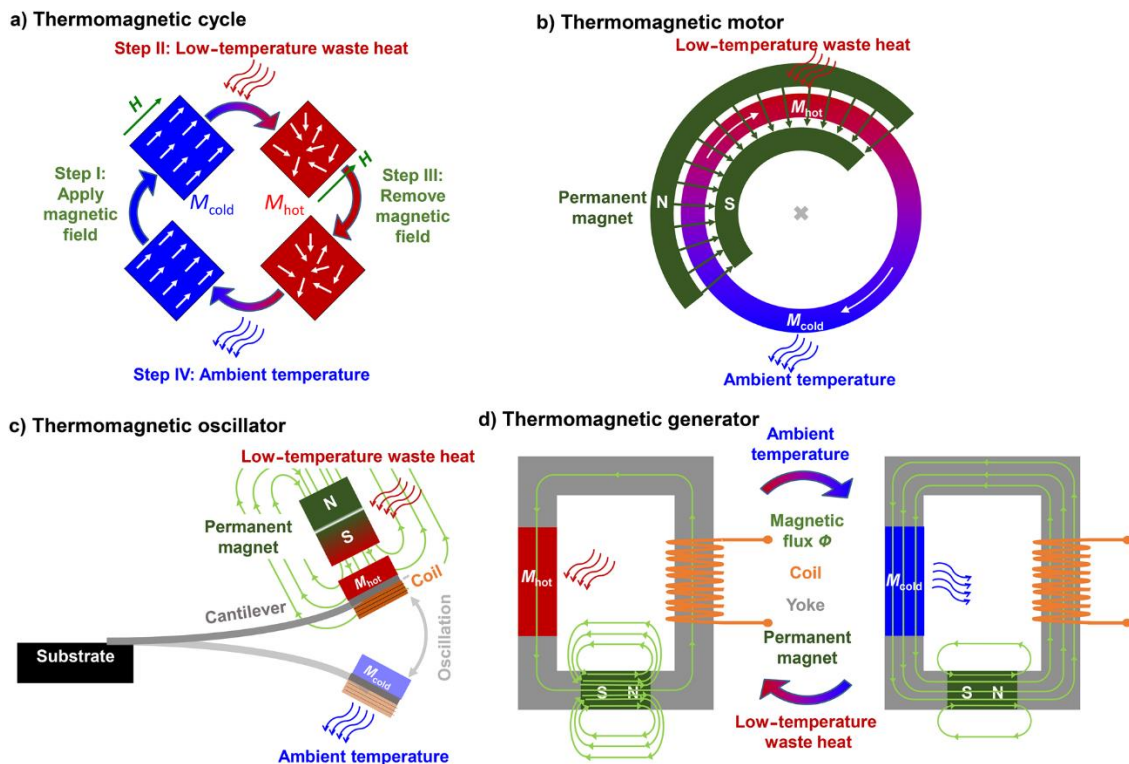
#### 3.1 Thermomagnetism

Thermomagnetic materials are materials that exhibit a change in magnetic properties in response to variations in temperature (Dzekan et al., 2021). This phenomenon is fundamental to various advanced applications, including the magnetocaloric effect, magnetic refrigeration, and temperature sensing (Dzekan et al., 2021). The study of thermomagnetic effects dates back to the 19th century, with contributions from key figures such as Pierre Curie, who first described the Curie temperature (Mould, 2007), and Louis Néel, who expanded on the understanding of magnetic phase transitions by exploring antiferromagnetism and ferrimagnetism, expanding the understanding of magnetic ordering in materials (Prévot and Dunlop, n.d.).

The Curie temperature, as mentioned previously, is a central property of thermomagnetic materials (Kishore and Priya, 2018) and describes the critical point at which a ferromagnetic material loses its magnetic properties and transitions to a paramagnetic phase (Mould, 2007). This transition occurs due to the increasing thermal energy causing the magnetic domains within the material to become disordered, thereby diminishing the material's net magnetization (Clark and Schmidt, 1982). Understanding and controlling the Curie temperature is crucial for designing materials and devices that operate effectively under varying thermal conditions (Pal et al., 2011).

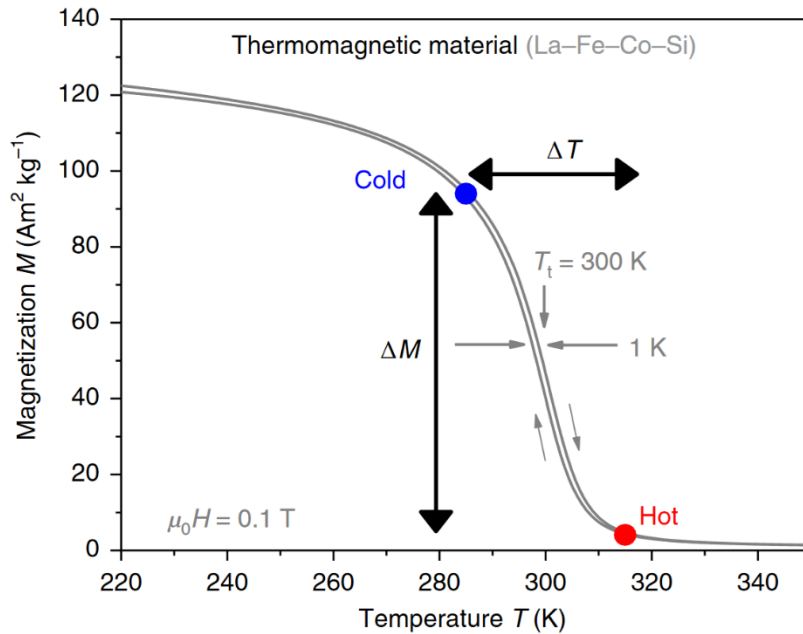
In addition to the Curie temperature, the magnetocaloric effect (MCE) is a phenomenon that is also associated with thermomagnetic materials (Nagy et al., 2019). The MCE refers to the reversible temperature change of a material when exposed to a changing magnetic field (Brück et al., 2007). This effect is the basis for magnetic refrigeration technology, which offers an energy-efficient and

environmentally friendly alternative to conventional gas-compression refrigeration methods (Brück et al., 2007). Other practical applications of thermomagnetic materials include temperature sensors, energy harvesting and harvesting low-grade heat (Dzekan et al., 2021). The first one of these being used to design highly sensitive temperature sensors that operate based on the sharp change in magnetic properties at the Curie temperature (Waske et al., 2018). Sensors like these are useful in various industrial and scientific applications where precise temperature monitoring is crucial (Chen et al., 2019). Additionally, it is possible to use thermomagnetic materials for energy harvesting and low-grade heat harvesting. In this instance the material is used to convert waste heat into usable energy a concept that concept is illustrated in **Figure 1**. This figure visualizes a practical implementation of thermomagnetic materials for low-grade heat recovery, showing the value of thermomagnetic materials for improving the energy efficiency of industrial processes.



**Figure 1** Examples for low waste heat collection utilizing thermomagnetic materials taken from (Dzekan et al., 2021)

A key factor in evaluating the effectiveness of a thermomagnetic material is its thermomagnetic plot, where the desired characteristic is a sharp change in magnetization over as small as possible temperature range as the temperature approaches the Curie temperature (Waske et al., 2018) as shown below (Waske et al., 2018).



**Figure 2** Desired Thermomagnetic plot taken from (Waske et al., 2018)

Recently advances in materials science have led to the discovery of new thermomagnetic materials with tailored Curie temperatures (Johnson et al., 2001). Researchers are exploring nanostructured materials, multiferroics, and composite systems to achieve better control over thermal and magnetic properties (Pal et al., 2011). Additionally, the integration of thermomagnetic materials into microelectromechanical systems (MEMS) and nanotechnology is opening new avenues for miniaturized devices with advanced thermal and magnetic functionalities (Pal et al., 2011).

In the future, continued research into the fundamental properties of thermomagnetic materials, combined with innovations in material synthesis and device fabrication, could lead to even more efficient and versatile applications, particularly in the areas of sustainable energy, advanced cooling systems, and precision sensing technologies. This process can be aided by tools such as machine learning, a topic that shall be investigated more closely in the following chapter.

### 3.2 Machine Learning (ML)

Machine learning (ML) is a fundamental component of artificial intelligence (AI) that involves the development of algorithms capable of learning from data to make predictions or decisions, without being explicitly programmed for specific tasks (Sarker, 2021). This data-centric approach enables AI models to identify patterns within datasets, allowing them to generalize from data and make accurate predictions or decisions in new, unseen scenarios (Sarker, 2021).

---

Machine learning is typically categorized into three primary types: supervised learning, unsupervised learning, and reinforcement learning, each addressing different problems and suited to varying types of data (Sarker, 2021). In supervised learning, an algorithm is trained on a labeled dataset, where each example includes both input features and corresponding output labels (Sarker, 2021). In this case the goal for the model is to learn the mapping between inputs and outputs so that it can predict the correct output when presented with new data (Sarker, 2021). This method is widely used in tasks such as classification, where data is categorized into predefined classes, and regression, where the aim is to predict a continuous value (Sarker, 2021).

Unsupervised learning, on the other hand, deals with unlabeled data (Sarker, 2021). In this scenario, the algorithm must infer the natural structure present within a dataset without explicit instructions (Sarker, 2021). This approach is commonly applied in clustering, where data points are grouped based on similarity, and in association, where relationships between variables in large datasets are identified (Sarker, 2021).

Reinforcement learning (RL) as the third type of machine learning, represents a different paradigm, where an agent learns to interact with a dynamic environment to achieve a specific goal (Sarker, 2021). To achieve this goal the agent takes actions within the environment and receives feedback in the form of rewards or penalties, with the primary objective of learning a strategy that maximizes the cumulative reward over time (Sarker, 2021). RL is particularly effective in scenarios that require a series of decisions, such as optimization of a photovoltaic-battery system (Real et al., 2024).

Beyond these core types, machine learning encompasses several additional areas that contribute to its versatility and broad applicability. Semi-supervised learning, for example, combines a small amount of labeled data with a large quantity of unlabeled data, allowing AI models to benefit from the strengths of both supervised and unsupervised learning (Sarker, 2021). This approach is particularly useful when labeling data is expensive or labor-intensive, yet there is an abundance of raw data available (C A Padmanabha Reddy et al., 2018).

Another important method is transfer learning, where a model is pre-trained on a large, diverse dataset and then fine-tuned for a specific task using a smaller, task-specific dataset (Zhuang et al., 2020). This technique is especially valuable in domains like computer vision and natural language processing (see chapter 3.3), where knowledge gained in one domain can be transferred to another with minimal additional training (Zhuang et al., 2020). Additionally, deep learning, a subfield of machine learning, involves neural networks with many hidden layers (see chapter 3.4), enabling these models to automatically learn representations of data (Schmidhuber, 2015). This makes the models in question

---

particularly well-suited for complex tasks such as image (Li et al., 2018) and speech recognition (Mehri et al., 2023).

However, all these types of machine learning are impacted by a key bottleneck, obtaining sufficient and relevant data in the first place (Roh et al., 2019). Gathering this data can be a lengthy and tedious process, often requiring information extraction from various sources. This thesis aims to address this issue by developing a new tool for automating the data collection process of thermomagnetic plots. The objective is to create an approach that not only offers an alternative to current procedures but also improves upon the existing practice of manually searching scientific papers for relevant plots and digitizing them by hand.

### **3.3 Theory of Natural language processing**

Natural Language Processing (NLP) is the field that provides computers with the ability to understand and interpret human language (Mielke et al., 2021). Examples of this are speech to text, Optical-Character-Recognition (OCR) (see chapter 3.6.2) and large language models (LLMs) (Zhao et al., 2023), which can be utilized for machine translations (Mielke et al., 2021), sentiment analysis and text summarization (see chapter 3.5).

As computers are only capable of understanding numbers, it is essential to translate words into vectors (Mielke et al., 2021). A major challenge in doing this, however, is capturing the surrounding context. For instance, in translation, the phrase "fruit flies like a banana" turns into "Fruchtfliegen mögen eine Banane" in German. Nevertheless, when translated directly word for word, it loses its intended meaning resulting in the translated sentence "Fruchtfliegen wie eine Banane", which when translated English becomes "Fruit flies (the insect) as a banana".

This issue is addressed by tokenization, a process that breaks down text into smaller units, such as words or subwords, and converts these units into vectors or "encodes them" (Mielke et al., 2021) in a way that preserve the context of the original text and their positions in the sequence. These vectors are then processed by the transformer based neural networks, allowing it to capture the nuanced relationships between tokens and enabling the model to generate or understand language with greater accuracy and avoid problems like fruit flies becoming bananas.

---

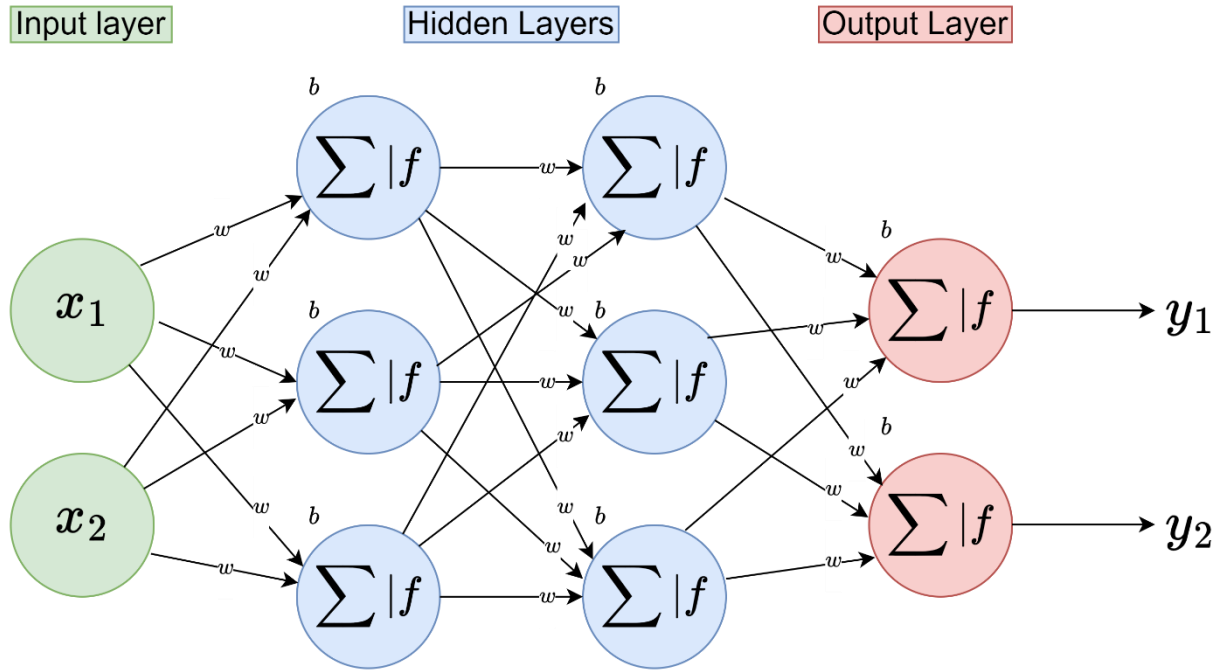
### 3.4 Neural networks

As mentioned previously neural networks (NN) are a fundamental part of artificial intelligence (AI). These networks are modeled after the neural structures of the human brain (Grossi and Buscema, 2007) and consist of interconnected nodes (see **Figure 3**), which serve as artificial neurons as shown by Grossi and Buscema (Grossi and Buscema, 2007). The architecture of a neural network typically includes input nodes, which receive data, output nodes, which produce the final prediction or classification, and hidden nodes, which exist between the input and output layers<sup>1</sup> brain (Grossi and Buscema, 2007). Each node in the network is therefore connected to others through weighted links, which determine the strength and influence of the signal transmitted between nodes (Grossi and Buscema, 2007). These weighted links are central to the operation of NN, as they dictate the extent to which one node influences another, thereby controlling the flow of information throughout the network (Grossi and Buscema, 2007). In order to properly adjust these weights to increase the accuracy in the networks prediction a training process is necessary. This adjustment process is akin to how synapses in the human brain strengthen or weaken to reinforce learning, allowing the network to become more adept at recognizing patterns and making accurate predictions.

Complementing the role of weights are biases, which are additional values added to the weighted sum of inputs before passing through an activation function (Wang et al., n.d.). Biases enable the NN to shift the activation function, providing greater flexibility in fitting the data and improving the network's ability to model complex relationships. This general structure of a neural network is shown below (see **Figure 3**) with **Equation 1** representing the performance of an individual node. Both weights and biases are refined during training through algorithms like backpropagation, which work to minimize the difference between the network's predicted output and the actual target value.

---

<sup>1</sup> For an illustration see **Figure 3**.



**Figure 3** Neural network (feed forward) architecture.  $x$  is inputs,  $y$  is the output of the network.  $w$  is weights and  $b$  is biases; source: own depiction

**Equation 1** Node functionality neural network, where  $n$  is the number of inputs into a node

$$\text{output of node} = f\left(\sum_{i=1}^n x_i w_i + b\right)$$

Once the basics of neural networks are understood it becomes essential to gain an understanding of their different types, each optimized for different tasks. The simplest form of neural network is the feedforward neural network (FFNs), where data moves in a singular direction without forming cycles between nodes (see **Figure 3**). This straightforward architecture, often used for basic tasks such as classification, processes information linearly, passing inputs through successive layers to generate an output (Murat H., 2006).

Building on the feedforward design, convolutional neural networks (CNNs) introduce a specialized architecture tailored for processing grid-like data structures, most notably images (O'Shea and Nash, 2015). CNNs apply convolutional layers that scan input data and use filters to detect local patterns such as edges, textures, and shapes (O'Shea and Nash, 2015). This approach makes CNNs particularly effective in tasks involving image recognition and processing, where understanding spatial hierarchies in data is crucial.

---

In contrast to feedforward and convolutional networks a different type of NN called recurrent neural network (RNN) incorporates loops within its architecture, enabling it to retain information from previous steps in a sequence. This feedback loop allows RNNs to maintain a memory of past inputs, which is essential for tasks that require contextual understanding over time, such as language modeling or time-series prediction (Schmidt, 2019). However, standard RNNs can struggle with long-term dependencies due to issues like vanishing gradients, which can impair their ability to learn from distant past events. To address the limitations of standard RNNs, long short-term memory (LSTM) networks were developed. LSTMs are a type of RNN specifically designed to remember information over extended periods (Hochreiter and Schmidhuber, 1997). They feature specialized memory cells that can retain or forget information selectively, making them well-suited for tasks where long-term context is critical, such as in speech recognition or sequential data processing (Hochreiter and Schmidhuber, 1997).

A notable departure from LSTMs is the concept of . These represent a significant evolution in NN-design, particularly for natural language processing (NLP). Unlike LSTMs, which explicitly manage long-term dependencies, transformers use a mechanism called self-attention to capture the relationships between elements in a sequence without prioritizing temporal order (Vaswani et al., 2023). This allows transformers to process sequences more efficiently, handle vast amounts of data, extract complex patterns and identify relevant contextual information regardless of distance within the sequence therefore making them the architecture of choice for large-scale NLP models (Vaswani et al., 2023).

The architecture of neural networks is thus carefully optimized to extract complex patterns from data, with each type of network offering distinct advantages depending on the nature of the task. In the realm of NLP, transformers have demonstrated particular effectiveness, not only due to their self-attention mechanism but also because of complementary processes like tokenization (Vaswani et al., 2023).

### **3.5 Theory behind Transformers and LLMs**

After introducing transformers in the previously this chapter explores the theoretical underpinnings of them, focusing on key concepts such as the key, query and value matrices, multi-headed attention, neural network structure, quantization, and model parameters.

The key, query and value matrices for example serve as the backbone to the self-attention mechanism and therefore how transformers assess relationships between different tokens in a sequence (Vaswani

---

et al., 2023). Accordingly, each token in the input sequence is represented by three vectors: a key, a query, and a value (Vaswani et al., 2023). These vectors are fundamental to the self-attention process<sup>2</sup>, where the query vector of a token is compared against the key vectors of all other tokens to calculate attention scores (Vaswani et al., 2023). The key, in this context, can be understood as 'what I have'—it is the vector representing each word or token in the input sequence, acting as an identifier that can be matched with a query (Vaswani et al., 2023). The query, on the other hand, is the vector that asks, 'what is the current token looking for?'. It is used to retrieve relevant information from other parts of the input sequence, seeking to find related words or contextual information within the sequence (Vaswani et al., 2023). In contrast, the value is the vector that contains the actual information or representation of each word or token (Vaswani et al., 2023). When a query finds a match with a key, the corresponding value vector is retrieved and used in the computation.

These attention scores determine the weight of each value vector, guiding the model to focus on relevant parts of the sequence with self-attention being calculated by the following **Equation 2** :

**Equation 2** self-attention calculation where Q is the query matrix, K is the key matrix, V is the value matrix and  $d_k$  is the dimension of the key vectors (Vaswani et al., 2023).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

This method allows the transformer to capture context more comprehensively, as it can dynamically adjust which tokens to prioritize based on their relevance to the task at hand (Vaswani et al., 2023).

Building upon the key, query and value mechanism, the 'multi-headed attention matrix' (Vaswani et al., 2023) further enhances the model's ability to understand complex language patterns. A head refers to one instance of the attention mechanism in this context. These heads typically have a dimension reduction to increase efficiency and a multi head is where the attention mechanisms or 'heads' are 'linked' together and have a linear transformation applied to them (Vaswani et al., 2023).

Cross-attention on the other hand refers to the query coming from one sequence and the keys as well as the values originating from another (Vaswani et al., 2023). This is most commonly used in situations such as translation where the query allows the link to the key matrix of words from the other language. This multi-headed approach allows the AI model to then simultaneously capture different aspects of the input data, such as structure and meaning.

---

<sup>2</sup> This process is further explained "Key elements of transformers".

---

Typically, a transformer model is structured with an encoder-decoder setup, where the encoder processes the input sequence, and the decoder generates the output. This is a configuration that is particularly effective in tasks like machine translation (Vaswani et al., 2023). The transition from the key, query and value matrices to multi-headed attention seamlessly integrates into the broader neural network framework, illustrating how each component of the transformer contributes to its overall functionality.

### **Key elements of transformers**

The self-attention mechanism that was touched upon in the previous chapter is the cornerstone of the transformer architecture and should be explained appropriately as such. It enables the model to assess the relative importance of different words within a sentence, allowing for a more nuanced understanding of language (Vaswani et al., 2023). Specifically, for each word in a sentence, self-attention calculates how much focus should be placed on every other word when encoding that particular word. This is essential because the meaning of a word often depends heavily on the context provided by surrounding words.

It is worthwhile to consider the following sentences: "The dog bit the child because it is angry" and "The dog bit the child because it is slow." In the first sentence, "it" refers to the dog, while in the second, "it" refers to the child. The self-attention mechanism captures such context-dependent meanings, enabling the model to understand which word "it" refers to, based on the surrounding text. This capability is crucial for the model to interpret and generate coherent language.

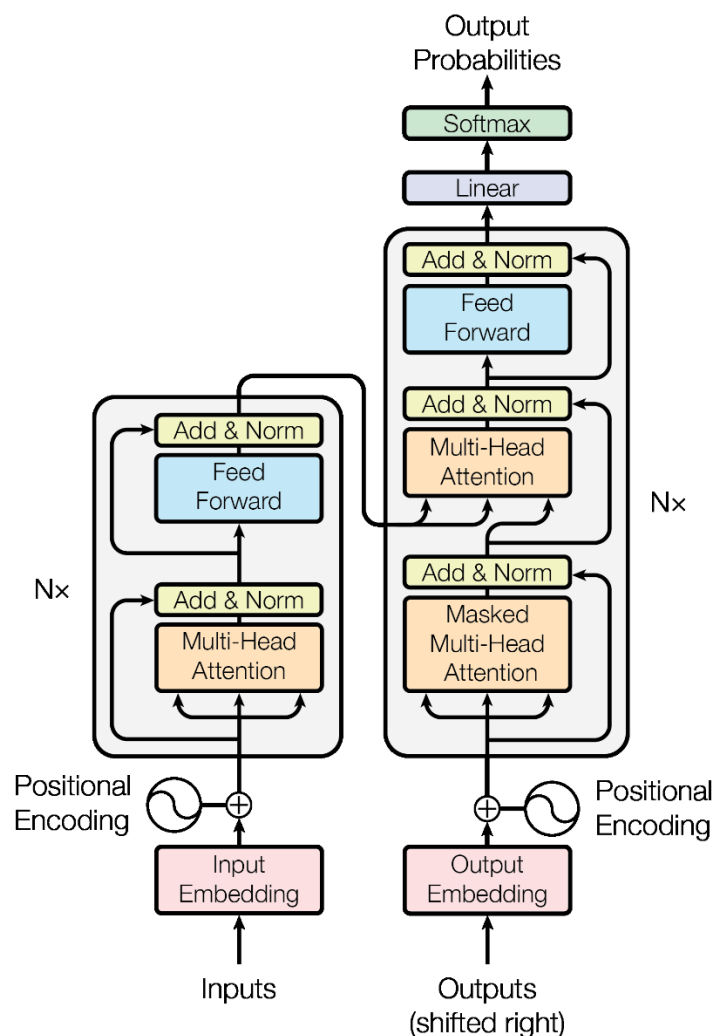
Unlike traditional models that process text sequentially, transformers process all words in a sentence simultaneously (Vaswani et al., 2023). To understand the order of words, which is vital for comprehension, transformers rely on positional encoding. Positional encoding is added to the word embeddings, providing the model with information about the position of each word in the sentence. This allows the transformer to maintain the sequence information while processing the entire sentence in parallel.

After the self-attention layer, the information is passed through a feed-forward neural network. This network processes the encoded information and generates the final output for that layer (Vaswani et al., 2023). The feed-forward network is responsible for transforming the data in a way that captures higher-level features and patterns, further refining the model's understanding before the output is produced (Vaswani et al., 2023).

## Transformer architecture

After considering self-attention and feed-forward neural networks this chapter examines the core elements of the architecture, including the encoder-decoder framework, multi-headed attention, and positional encoding, to understand how these features contribute to the model's advanced capabilities in handling sequential data.

The transformer architecture typically follows an encoder-decoder structure. The encoder (visualized on the left in (see **Figure 4**) takes the input sentence and converts it into a series of continuous representations, which encapsulate the meaning and context of the text. The decoder (visualized on the right in (see **Figure 4**) then takes these representations and generates the output sequence, such as a translated sentence in the case of machine translation (Vaswani et al., 2023).



**Figure 4** Transformers architecture taken from "All you need is attention" (Vaswani et al., 2023).

---

Taking a closer look at **Figure 4** the process begins with the input embedding, where words are first converted into vectors using an embedding layer. These vectors are then enriched with positional encoding, which provides the model with crucial information about the position of each token in the sequence, ensuring that the order of words is preserved (Vaswani et al., 2023).

One of the key features of the transformer architecture is multi-head attention. This mechanism allows the model to focus on different parts of the input sequence simultaneously when generating an output (Vaswani et al., 2023). By employing multiple attention mechanisms in parallel, the model can capture various relationships between tokens more effectively. In the encoder, the multi-head attention layer takes the input embeddings and examines other words in the sequence to gather context. In the decoder on the other hand, there are two distinct multi-head attention layers. The first is the masked attention layer, which handles the decoder's own outputs. It is referred to as 'masked' because it is prevented from seeing future tokens when producing the next token in a sequence (Vaswani et al., 2023). This ensures that predictions for each token are based solely on the tokens that have already been processed. The second attention layer is the encoder-decoder attention, also known as 'cross-attention' (Vaswani et al., 2023). This layer allows the decoder to focus on the relevant parts of the encoded input sequence when generating each token in the output sequence, thereby helping the decoder attend to the most pertinent information encoded by the encoder.

The architecture depicted in **Figure 4** also incorporates a crucial "Add & Norm" component, which plays an important role in enhancing the model's performance and stability during training. The "Add" operation refers to the residual connection, where the input to a layer is directly added to the output of that same layer. This residual connection is vital because it ensures the continuous flow of gradients throughout the network, preventing them from vanishing or exploding during backpropagation (Vaswani et al., 2023). By maintaining this gradient flow, the model can refine its input iteratively rather than starting from scratch at each layer, which accelerates learning and improves the overall robustness of the network.

Complementing the residual connection is the "Norm" part of the "Add & Norm" component, which refers to layer normalization. After the residual addition, layer normalization is applied to standardize the output of the layer. Specifically, it ensures that the output has a mean of zero and a variance of one across each example in a batch. This standardization process is similarly important because it stabilizes the learning process and enhances the convergence of the model during training, leading to more consistent and reliable performance across different tasks. Once the attention mechanisms have processed the input, the architecture applies a simple feed-forward neural network to each position in

---

the sequence. This network further refines the information gathered from the attention layers, enabling the model to better capture and understand the underlying patterns in the data (Vaswani et al., 2023). The feed-forward network operates independently at each position, allowing the model to build a more detailed and nuanced representation of the input sequence.

On the decoder side of the architecture, the process begins with the target data, which is slightly shifted to the right. This shift helps the model learn to predict the next word in the sequence during training. For example, given a target sentence like "The dog barked at someone," the input to the decoder would be "<start> the dog barked at," where "<start>" is a special token indicating the beginning of the sequence. The model is then trained to predict the next word, "someone," for each position in the input. This shifting mechanism is crucial for teaching the AI model how to generate coherent and contextually appropriate sequences. Akin to the input side, the output also undergoes embedding and positional encoding. In this step, the output tokens are converted into numerical vectors, and their order in the sequence is recorded through positional embeddings (Vaswani et al., 2023). These embeddings allow the model to retain the sequential structure of the data, which is essential for tasks like language modeling and translation.

In a final step the output from the decoder is passed through a linear layer, which transforms the logic vectors into a format interpretable as probabilities for each word in the vocabulary. This is followed by a softmax layer, which converts the probability scores produced by the linear layer into a probability distribution over the entire vocabulary (Vaswani et al., 2023). This distribution indicates the likelihood of each word being the next in the sequence, providing the final output probabilities. This combination of linear transformation and softmax activation is critical for generating predictions that are both accurate and interpretable, allowing the model to effectively generate or understand natural language sequences.

### **3.5.1 Parameters of LLMs**

Parameters of large language models represent the complexity and capacity of the AI model. The more parameters a model has, the wider range of tasks and the more detailed representations of the data are possible. However, the more parameters a model has, the more computational resources and data are needed to both utilize and train the model. These parameters tend to be weights and biases within the Neural Network of an LLM.

---

These parameters are quite numerous as they represent the complexity of the models. As a result, they also tend to be included in the name of the models, for example, 'Llama 3 8B' represents 8 billion parameters (Meta, 2024).

### 3.5.2 Improvements to LLMs

The evolution of LLMs is marked by significant advancements aimed at enhancing their performance and efficiency. This chapter introduces key techniques such as prompt engineering, fine-tuning, model improvements, statistical model bias mitigation, and quantization. Prompt engineering optimizes how LLMs interpret and respond to user inputs, ensuring outputs align with specific needs. Fine-tuning further refines LLMs by training them on specialized datasets, improving their accuracy in niche domains. Model improvements focus on enhancing architecture and complexity to boost overall capabilities. Addressing statistical model bias is crucial for ensuring reliable outputs, while quantization reduces computational demands, enabling efficient deployment in resource-constrained environments. This chapter will explore these techniques in detail, highlighting their importance in the evolution of LLMs.

A key strategy improving the output of a LLM is prompt engineering, which involves crafting and refining the inputs or prompts provided to the model to improve the relevance and coherence of its responses (Chen et al., 2023). Effective prompt engineering ensures that LLMs better understand and respond to queries, aligning their outputs more closely with user expectations and application requirements<sup>3</sup> (Chen et al., 2023).

Fine-tuning further enhances LLMs by continuing their training on domain-specific datasets after the initial training phase (Ren and Sutherland, 2024). This targeted approach allows models to adapt to specialized contexts, such as scientific terminology within the field of material science, improving their ability to generate accurate and contextually appropriate responses in these areas. Fine-tuning thus enables LLMs to be more effective in niche applications by refining their understanding based on specific data.

Model improvements encompass a variety of techniques designed to boost the overall capabilities of LLMs. These improvements include increasing the model's size, adding more layers, or integrating advanced architectural innovations. Such enhancements contribute to greater accuracy by allowing a deeper understanding of complex linguistic patterns and improve generalization across diverse tasks

---

<sup>3</sup> For prompt engineering used in this thesis see chapter 5.1.1

---

(Du et al., 2024). Continuous advancements in model design and training methodologies drive the development of increasingly sophisticated models.

Statistical model bias is whereby examining the variability and distribution of multiple responses from an LLM, you can gauge the model's confidence and potential bias in its output. This is useful for understanding the limitations and behavior of LLMs, particularly in scenarios where precision and reliability are crucial.

Quantization represents a technique used to optimize the computational efficiency of LLMs. By reducing the precision of the model's weights and activations, quantization decreases the memory and computational requirements of LLMs. This allows for faster inference and more efficient use of hardware resources, while maintaining performance levels close to those of higher-precision models.

### **Quantization**

Generally LLMs are trained with float 32 [full] or float 16 [half precision] floating point numbers (Lin et al., 2024). A float 32 has 32 bits which is the equivalent of 4 bytes. This will result in a model with 1bn parameters with float 32 requiring the equivalent of storage of approximately 4GB (Gigabytes). This is where LLM quantization comes in, which describes the mapping of high precision values to low precision values (Lin et al., 2024) to lower the required amount of computing power and resources required to run the model. Quantization reduces these high values down from float 16 float 8 integer 8 or even 8bit integer. This can result in a model size reduction to up to 1/8 of its original size. So taking the original example the 4 GB float 32 model down to half a GB.

There are two main techniques of performing quantization. Post-training, taking an already trained model and reducing its weights or quantization aware training where the data is fine-tuned so that quantization can be easily done. This avoids the need for calibration after. However, the cost of reducing these high precision weights to low precision weights reduces the model accuracy (Lin et al., 2024).

### **Prompt Engineering**

Prompt engineering is a critical technique in the interaction with LLMs, focusing on the careful construction of input instructions to guide the model towards more accurate and contextually appropriate outputs (Chen et al., 2023). By strategically designing prompts, users can significantly influence the quality and relevance of the responses generated by the AI model. This process involves

---

several methodologies that enhance the model's understanding of the task, ensuring that the output aligns with the user's intentions (Chen et al., 2023).

One of the key strategies in prompt engineering is the 'Chain of Thought' (Chen et al., 2023) approach, which involves prompting the model to think through a problem step by step. By breaking down the task into smaller, logical steps, the model is encouraged to follow a more structured thought process. This method is particularly useful in complex tasks where a linear, logical progression is necessary to arrive at the correct solution (Chen et al., 2023). An example from this thesis is, first asking the model to read the document, rather than just state "does this model contain the information I need to improve performance" (see chapter 5.1).

Another crucial aspect of prompt engineering is the clarity of instructions provided to the model. Clear and precise prompts reduce ambiguity and help the model understand the exact nature of the task. For instance, if a specific format or length of output is required, this should be explicitly stated in the prompt. This can be illustrated in scenarios where a desired output needs to adhere to a certain structure, such as a JSON format, or where the response must be concise and to the point. By demonstrating the expected format within the prompt, the user can guide the model to produce outputs that are not only correct but also presented in the desired manner<sup>4</sup>. This technique is particularly effective in ensuring consistency in outputs, as seen in various applications later discussed in chapter 5.1.1 (Chen et al., 2023).

A different powerful tool within prompt engineering is capitalizing keywords within a given prompt. By writing certain words or phrases in all capital letters these are emphasized which can in turn substantially influence the model's interpretation of those terms. This technique is particularly relevant in contexts where the importance of specific concepts must be highlighted. For instance, in tasks requiring the model to search for specific types of data, such as a "MAGNETIZATION VS. TEMPERATURE PLOT" (see chapter 5.1), capitalizing these terms helps the model prioritize them in its processing. This method leverages the way LLMs handle tokenization and embedding, where the visual emphasis on certain words can guide the model to treat them with greater importance in generating a response.

Another powerful technique in prompt engineering is meta-prompting, where the user asks the model to generate a prompt that can then be used for a particular task (Suzgun and Kalai, 2024). Meta-prompting is especially useful when the user is uncertain about how to frame a question to achieve the desired outcome (Suzgun and Kalai, 2024). By letting the LLM craft the prompt, users can harness the model's own capabilities to create more effective or creative prompts. This approach not only aids in

---

<sup>4</sup> For an overview of prompt engineering used in this thesis see chapter 5.1.

---

refining the initial input but also helps in exploring different ways of phrasing a task that the user might not have considered, thereby enhancing the model's overall utility (Suzgun and Kalai, 2024).

Another aspect that improves the AI model's utility is called 'fine tuning' and will be discussed in the following chapter.

## **Fine tuning**

While fine-tuning was not used in this thesis it remains an important aspect to understand the function of LLMs<sup>5</sup>. Fine-tuning describes a process used to adjust the parameters of a pre-trained LLM to tailor it to a specific task. This involves retraining the model on a different dataset, adjusting all nodes within the model based on the new data. In order to determine which type of fine-tuning is required one must take a closer look at the nature of the dataset used. Supervised fine-tuning, for example, is used for labeled datasets where the model learns to map inputs to desired outputs based on the examples provided (Ren and Sutherland, 2024). This approach has the advantage of producing highly accurate models, but it also requires high-quality labeled datasets and carries the risk of overfitting, where the model becomes too specialized to the training data, leading it to potentially perform poorly on unseen data. Unsupervised fine-tuning, on the other hand, does not rely on labeled data (Khramtsova et al., 2024). Instead, it involves feeding the model large amounts of relevant, but unlabeled, data that is pertinent to the task or domain in question (Khramtsova et al., 2024). While this approach allows for the incorporation of diverse data sources, it also poses the risk of introducing biases into the model if the data is not thoroughly vetted beforehand.

Other approaches to this challenge are few-shot and zero-shot learning. These are methods related to prompt-based learning where the model is provided with examples to adapt to new tasks with minimal additional training (Zhang et al., 2024). In few-shot learning, the model is given a few examples of prompts along with the desired outputs to help it learn and adapt to the task (Zhang et al., 2024). In contrast, zero-shot learning involves prompt engineering without any examples, relying on the model's existing knowledge to generate appropriate responses (Zhang et al., 2024). These methods are particularly useful when dealing with tasks that have limited labeled data, as they allow the model to perform new tasks with little to no additional training.

---

<sup>5</sup> The reason why fine-tuning was not applied in this research is that as a model is fine-tuned, it becomes less generalized, reducing its ability to perform a wide range of tasks.

---

Another strategy is transfer learning which involves taking a model that has already been fine-tuned for a similar task and adapting it to the target task. This is particularly advantageous when the target task has limited data, as it leverages the pre-existing knowledge of the model. However, care must be taken when selecting a pre-trained model to ensure that it is relevant to the new task in order to avoid suboptimal performance.

The last strategy in the context of fine-tuning LLMs is called feature extraction and involves utilizing the model's pre-trained ability to generate meaningful text embeddings that capture semantic and contextual information (Mischler et al., 2024). During fine-tuning, these embeddings are adapted and refined on a smaller, task-specific dataset, allowing the model to quickly and efficiently specialize in tasks such as text classification or sentiment analysis. This process leverages the model's existing knowledge, making fine-tuning more effective and requiring less data compared to training a model from scratch (Mischler et al., 2024). Feature extraction thus plays a critical role in the fine-tuning process, enabling the efficient adaptation of LLMs to a wide range of specialized tasks while preserving the general capabilities of the model (Mischler et al., 2024).

### **Statistical model bias**

Next to fine-tuning, discussed in the previous chapter statistical model bias depicts another important aspect of LLMs. It refers to the tendencies or predispositions these models have when generating responses. These statistical model biases are influenced by the training data, its architecture and how it was fine-tuned (Sivaprasad et al., 2024). Essentially, the model does not just provide a single, fixed answer to a prompt; instead, it generates responses based on probabilities learned during training. This process can lead to variability in the outputs, especially if the model isn't particularly confident about the correct answer (Sivaprasad et al., 2024).

When an LLM is presented with the same prompt on multiple occasions and exhibits a diverse range of responses, this suggests that the model is drawing from a broad distribution of potential answers. A variability as such reflects the model's uncertainty. In cases where the responses vary greatly, this signals that the model is not certain about what the 'correct' response should be (Sivaprasad et al., 2024). This uncertainty can arise from ambiguous or complex prompts, a lack of sufficient training data on the topic, or simply the inherent difficulty of the subject matter (Sivaprasad et al., 2024). To gain a deeper understanding of this, one can generate several responses to the same prompt and analyze how they are distributed. If these responses are consistent, the model likely has a more defined

---

understanding of the prompt, leading to similar outputs. On the other hand, if the responses are highly variable, it suggests that the model's internal representation of the prompt is less certain, resulting in a broader range of outputs. By examining the distribution and variability of these responses, one can assess the degree of confidence that the model exhibits and identify any potential biases in its output. This approach is valuable for understanding the limitations and behavior of LLMs, particularly in contexts where accuracy and reliability are important.

### 3.6 Theory of digitization

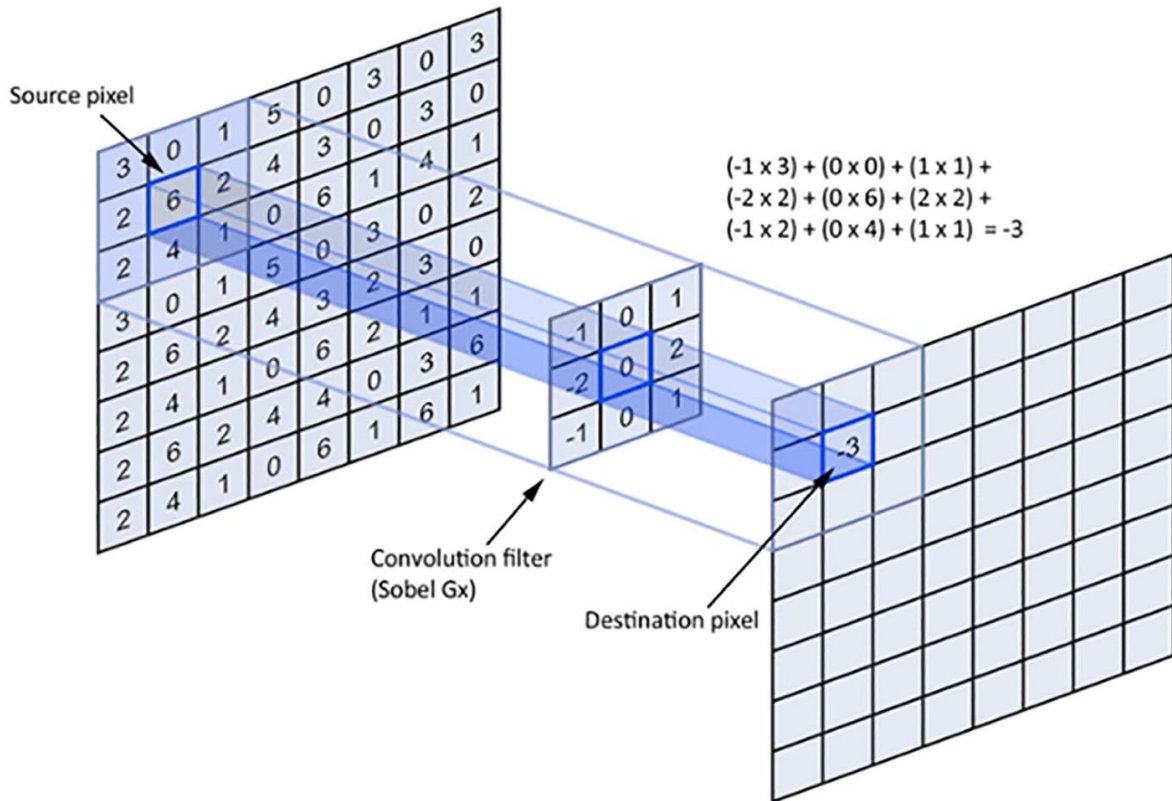
Similar to language, when it comes to the digitization of images these must first be converted into a format computers can better understand and interpret. This is done through image convolution, a topic that is further discussed in the following chapter. In short; image convolution reduces the images sectors to a simpler mathematical matrix using a kernel (Almryad and Kutucu, 2020).

A key component of digitization is the bounding box. This is the smallest minimum enclosing box in which an object or all points lie within. Two sets of coordinates are thus defined: the top left coordinates  $(x_{max}, y_{max})$  and the bottom right corner  $(x_{min}, y_{min})$ . The bounding box is commonly used in object detection that involves algorithms to identify objects in images and draw bounding boxes around them (He et al., 2018). Additionally, an image annotation is used, when labeling images for training machine learning models. In this case bounding boxes are used to mark the location of objects within the images. Additionally bounding boxes can also be used in video analysis, tracking the movement of objects across frames.

#### 3.6.1 Image convolution

Image convolution is a fundamental operation in the processing of visual data, particularly within convolutional neural networks (CNNs). It involves applying a kernel—a matrix of weights—across the pixels of an image to detect specific features in the input data (Almryad and Kutucu, 2020). A kernel functions as a filter that systematically scans over the image in small, overlapping regions, executing tasks such as edge detection, blurring, or sharpening (Liu et al., 2016). For instance, the "Sobel Gx" kernel illustrated in **Figure 5**, is tailored to detect horizontal edges within an image. This kernel operates by extracting the pixel values from a 3x3 matrix, applying the corresponding weights from the kernel, and condensing this information into a single pixel value (Vincent and Folorunso, 2009). The kernel then shifts to the next 3x3 region, continuing this process across the entire image. The outcome of this

convolution process is an accentuation of specific image characteristics, such as horizontal edges. This is called a 'feature map' (Ballas and Diou, 2023). These are crucial in subsequent stages of image processing or analysis, as they emphasize the most relevant features for the task at hand.



**Figure 5** Image convolution example taken from (Almryad and Kutucu, 2020).

CNNs build on these principles of image convolution by stacking multiple convolutional layers that apply a variety of filters to the input data. These filters perform convolutions to detect a wide range of features, including edges, textures, and complex patterns (O’Shea and Nash, 2015). Following these convolutional layers, pooling layers are typically introduced to reduce the spatial dimensions of the feature maps (Li et al., 2020). This reduction enhances the performance and robustness of the network by minimizing the computational load and helping the model generalize new data better (Li et al., 2020). Additionally, non-linear activation functions are applied after each convolution layer to introduce nonlinearity into the model, enabling it to capture more complex patterns in the data (Li et al., 2020). Finally, the network’s output layer utilizes a different activation function to produce the final prediction or classification, completing the process from raw image data to a meaningful, task-specific output (Li et al., 2020). An example of this is the operation of 'Optical Character Recognition' (OCR) software.

---

### 3.6.2 Optical Character Recognition (OCR) and Character-Region Awareness for Text detection (CRAFT)

Optical Character Recognition (OCR) is a technology that converts images containing text into machine-encoded text (S and A, 2015). OCR therefore enables the transformation of text within images—such as scanned documents, photos of text, or digital images of printed materials—into a format that can be easily edited, searched, and stored digitally (S and A, 2015). This technology works by analyzing the structure of the text within an image, identifying individual characters, and then translating these characters into a digital format (S and A, 2015). Initially, OCR systems relied on pattern matching to recognize characters, essentially comparing the image of a character to a stored set of patterns. However, modern OCR technologies have significantly evolved, now utilizing advanced machine learning algorithms that enhance accuracy and adaptability across different fonts, languages, and text formats (S and A, 2015). They also generally use a convolutional recurrent neural network (CRNN) to perform the conversion from image to text of individual characters.

Within the OCR field there is another model, specifically designed to detect text in complex environments. This model is called 'Character-Region Awareness for Text Detection' (CRAFT) (Baek et al., 2019) and can recognize text in natural scenes where letters or words may appear on signs or other objects (Baek et al., 2019). CRAFT differs from traditional OCR approaches by focusing on detecting individual characters rather than entire lines of text, enabling it to link these characters together to form words. This approach allows for more precise detection in scenarios where text may not be uniformly aligned or might appear in irregular shapes and sizes (Baek et al., 2019). CRAFT itself is built on a convolutional neural network (CNN) architecture, which is particularly well-suited for processing and analyzing visual data (see chapter 3.4). In operation, CRAFT generates a heatmap that highlights regions of the image where text is likely present, and it creates bounding boxes around detected text (Baek et al., 2019). These bounding boxes in turn help to accurately locate and isolate text within an image<sup>6</sup>.

### 3.6.3 You Only Look Once (YOLO)

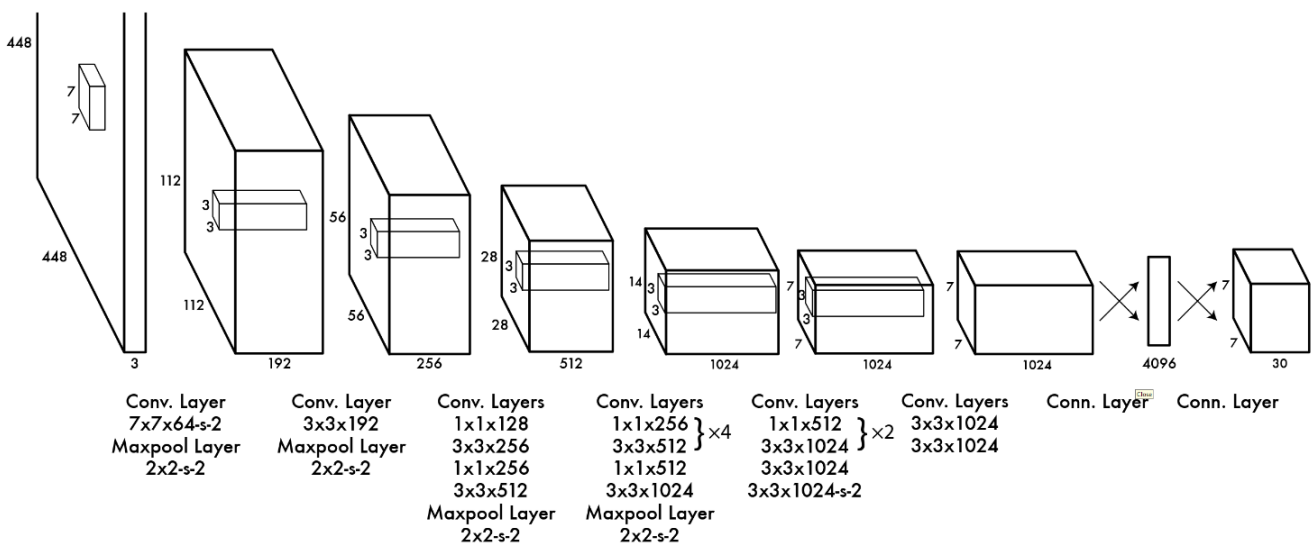
Another object detection model that is based on CNNs is the You Only Look Once (YOLO) network developed by Joseph Redmon in 2016 (Redmon et al., 2016). YOLO's primary innovation lies in its approach to object detection, where it reimagines the problem as a single regression task, rather than

---

<sup>6</sup> While CRAFT's heatmap capability is a powerful feature, it was not utilized in the specific project presented in this thesis.

a series of individual tasks as seen in traditional methods (Redmon et al., 2016). The core concept behind YOLO is its division of an image into a grid system, where each grid cell is responsible for detecting objects that fall within its boundaries (Redmon et al., 2016). This grid-based approach allows YOLO to predict the presence of multiple objects within an image in a single pass through the network, significantly increasing both detection speed and accuracy (Redmon et al., 2016).

As illustrated in **Figure 6**, the architecture of YOLO operates by first resizing the input image to a fixed dimension, typically 448x448 pixels, to standardize the data fed into the network (Redmon et al., 2016). The image is then divided into a grid, and each grid cell predicts a set of bounding boxes along with the associated confidence scores, which reflect the likelihood that a bounding box contains an object and the accuracy of the predicted location (Redmon et al., 2016). YOLO's detection mechanism allows it to not only identify objects but also classify them at the same time, outputting predictions of what it believes each object is.



**Figure 6** YOLOs convolutional network taken from (Redmon et al., 2016)

### 3.6.4 Fully Convolutional One-Stage Object Detection (FCOS)

A different innovation in the field of object detection is the fully convolutional one-stage. To understand what makes a fully convolutional one-stage object detection unique one must first understand that it lacks the commonly used concept of an 'anchor box' (Tian et al., 2019). An anchor box is a predefined bounding box that is used as a reference to detect objects within an image.

FCOS, rather than relying on anchor boxes utilizes key points in the image, such as a pixel in the feature map to predict the coordinates of the bounding box (Tian et al., 2019). The method in question therefore allows for a more simple and flexible detection pipeline that reduces the complexity of the

---

model and computational overhead. This model also deploys a feature pyramid network to handle different objects of different sizes that generate feature maps at different resolutions (Tian et al., 2019). This allows the model to perform well across a range of object sizes, additionally generating a score of how close the bounding box is to the center of the object to help the model focus on the central region of objects, therefore improving the accuracy of the bounding boxes it predicts (Tian et al., 2019).

### 3.6.5 Semantic Segmentation and efficient residual factorized network (ERFNet)

Semantic Segmentation, which is different from the YOLO or FCOS-model with their bounding boxes is an image processing task that classifies each pixel in to a specific category or class (Romera et al., 2018). The goal of this is to assign a label to every pixel in the image, effectively segmenting the image into relevant areas (Romera et al., 2018). ERFNet is a type of neural network architecture, designed for semantic segmentation. Rather than drawing a bounding box around an object ERFNet labels all the pixels of that object as that object (Romera et al., 2018). In this instance it's utilized to segment each individual plot line from the background and other plot lines (see chapter 5.2.2).

### 3.6.6 Optical Flow for Plot2spectra

Optical flow is an old concept originating in the 1940s (Gibson, 1950) however its application in Plot2Spectra is a modern application. Optical Flow is the pattern of apparent motion between two consecutive frames caused by the movement of objects (Beauchemin and Barron, 1995). In the context of this thesis, it is utilized to ensure that the pixels of particular plot lines are correctly grouped together, even if the lines, colors or style are partially obscured. This improves the detection of the plot lines by leveraging the continuity and smoothness in pixel intensity as the line moves through the plot (Jiang et al., 2022).

In total there are 5 equations worthy of noting if one wants to understand optical flows utilization in Plot2Spectra. The first equation represents the brightness constancy assumption, a key principle behind optical flow. It assumes that the intensity of a specific point remains constant as the point moves from one position to another (see **Equation 3**).  $x, y$  being Coordinates of pixels in the plot image.  $\Delta x, \Delta y$  on the other hand, are the small changes in the x and y coordinates, representing the movement of the point from one position to another.  $I_p(x, y)$  however, depicts the intensity of the pixel at position  $(x, y)$  in the plot image.

**Equation 3** Brightness constancy assumption (Jiang et al., 2022).

$$I^p(x, y) = I^p(x + \Delta x, y + \Delta y)$$

**Equation 4** explains the intensity constraint of optical flow with  $W^p$  being the width of the plot image in pixels and  $I^p(x_i, y_i)$  the intensity of the pixel at the  $i$ -th position along the plot line.  $x_i, y_i$  in contrast are indicative of the  $i$ -th pixel along the plot line, while  $L^{\text{intensity}}$  marks the loss function. This function measures the difference in intensity between consecutive pixels along the plot line, enforcing the constancy of intensity as the line progresses.

**Equation 4** Optical Flow Intensity Constraint (Jiang et al., 2022).

$$L^{\text{intensity}} = \sum_{i=0}^{W^p-1} \|I^p(x_{i+1}, y_{i+1}) - I^p(x_i, y_i)\|_2^2$$

Moving on to **Equation 5** to visualize optical flows smoothness constraint. This formula is comprised of  $V(x, y)$  which is used to describe the velocity of the point along the  $y$ -direction which can be obtained by applying a Taylor expansion to Equation 3 where  $I_x^p(x, y)$  and  $I_y^p(x, y)$  which are the partial derivatives of the intensity  $I^p$  with respect to  $x$  and  $y$ , respectively. This is then used in the secondary equation.  $L^{\text{smooth}}$  is thus the loss function that enforces smoothness by penalizing large changes in the slope of the plot line, ensuring the line is continuous and differentiable.

**Equation 5** Optical Flow Smoothness Constraint (Jiang et al., 2022).

$$V(x, y) = \frac{I_x^p(x, y)}{I_y^p(x, y)}$$

$$L^{\text{smooth}} = \sum_{i=0}^{W^p-1} \|y_{i+1} - y_i - V(x_i, y_i)\|_2^2$$

Semantic constraint loss within optical flow is calculated with **Equation 6**. Where  $\tilde{C}(x_i, y_i)$  describes the probability that the pixel at  $(x_i, y_i)$  belongs to the plot line, as predicted by a semantic segmentation model.  $L^{\text{semantic}}$  thus refers to the loss function that encourages more foreground pixels (i.e., pixels belonging to plot lines) to be correctly classified by the segmentation model.

**Equation 6** Semantic Constraint Loss (Jiang et al., 2022).

$$L^{\text{semantic}} = \sum_{i=0}^{W^p} \|1 - \tilde{C}(x_i, y_i)\|_2^2$$

The last equation worth noting in this context is **Equation 7** which depicts the total loss for plot line detection with optical flow. This equation takes the results of the previous equations and combines them with each other to depict  $L^{\text{line}}$  as the total loss function used for optimizing plot line detection, combining smoothness, intensity constancy, and semantic constraints.

**Equation 7** Total Loss for Plot Line Detection (Jiang et al., 2022).

$$L^{\text{line}} = L^{\text{intensity}} + L^{\text{smooth}} + L^{\text{semantic}}$$

Utilizing these equations the algorithm below shown in **Figure 7** is utilized by Plot2Spectra to digitize lines.

**Initialization:**  $V(x, y)$ ,  $\tilde{C}$ ,  $I^p(x, y)$   
Pick a point from each plot line as the start position  $\{(x_t, y_t^k) \mid k = 1, 2, \dots, M\}$ , where  $M$  is the number of plots  
**while**  $t < W^p$  **do**  
     $\hat{y}_{t+1}^k \leftarrow y_t^k + V(x_t, y_t^k)$    ▷ Estimate the next position of the point with optical flow  
     $y_{cand} \leftarrow \{\tilde{C}(x_{t+1}) == 1\}$    ▷ Select pixels of plot data in the semantic map  
     $\{\hat{y}_{t+1}^k\} \leftarrow \text{SemanticMap}(\{\hat{y}_{t+1}^k\}, y_{cand})$   
     $\{y_{t+1}^k\} \leftarrow \text{ColorMap}(x_{t+1}, \{\hat{y}_{t+1}^k\})$   
     $t \leftarrow t + 1$   
**end while**  
**while**  $t > 0$  **do**  
     $\hat{y}_{t+1}^k \leftarrow y_t^k - V(x_t, y_t^k)$    ▷ Estimate the next position of the point with optical flow  
     $y_{cand} \leftarrow \{\tilde{C}(x_{t-1}) == 1\}$    ▷ Select pixels of plot data in the semantic map  
     $\{\hat{y}_{t-1}^k\} \leftarrow \text{SemanticMap}(\{\hat{y}_{t-1}^k\}, y_{cand})$   
     $\{y_{t-1}^k\} \leftarrow \text{ColorMap}(x_{t-1}, \{\hat{y}_{t-1}^k\})$   
     $t \leftarrow t - 1$   
**end while**

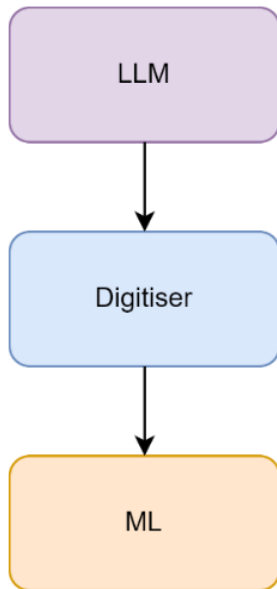
**Figure 7** Plot2Spectra optical flow algorithm (Jiang et al., 2022).

---

## 4 Development methodology

---

After covering important theory of thermomagnetism, machine learning, neural networks, transformers and digitization this chapter aims to give insight into the developed methodology of this thesis. This



can be separated in the testing and development of LLMs on the one hand, and digitization methods utilized on the other hand. A general workflow of how the code should operate is shown in **Figure 8** with the end goal of building a database to perform machine learning over. The code should begin first with a LLM module (shown in purple) to identify whether the document contains relevant plots and if yes, find figures of the plots with their corresponding data. This data is then added to the database storing the information within the plot i.e. the composition of material. Located plots are then fed to the digitization module (shown in blue) and machine learning (shown in orange) can be then performed on the database.

**Figure 8** Workflow; source: own depiction

### 4.1 LLM testing and development

The first wave of testing the LLMs was conducted using a 3080 Ti NVIDIA GPU. However, as the complexity of the models increased, requiring greater computational power, the testing environment was upgraded to an A100 NVIDIA GPU. This phase utilized a dataset comprising 15 PDFs published after 2010 that are less than 15 pages, which included the desired thermomagnetic plots, along with an additional 3 PDFs that did not but were in the field of material science. Later, another 13 PDFs devoid of thermomagnetic plots and unrelated to material science were incorporated into the testing to evaluate the robustness and accuracy of the models.

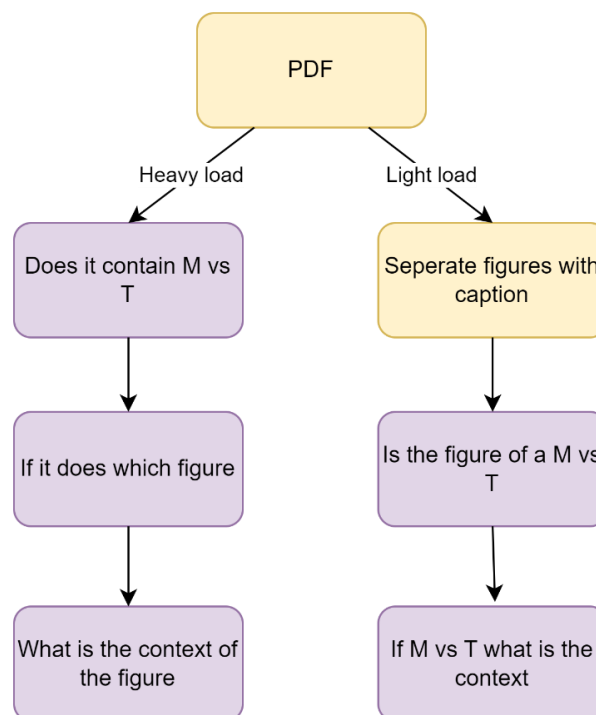
The testing began using GPT-2, which was tasked with extracting figures and summarizing them from documents. The performance of GPT-2 was found to be suboptimal, as the generated summaries were irrelevant to the corresponding figures. From this experience a main methodology was developed, and subsequent testing was done with ChatGPT-4. The initial approach involved feeding entire PDFs into the model while repeatedly refining the prompts<sup>7</sup>. This was done in an effort to optimize output quality and helped identify potential challenges in utilizing ChatGPT-4 for figure extraction and summarization.

---

<sup>7</sup> An example of prompts and their evolution can be found in chapter 5.1.

For instance, a specific example of a 'heavy load' task is illustrated in **Figure 9**, where the entire PDF is processed rather than individual sections. The performance limitations observed with GPT-4 prompted an exploration of alternative models and given the varying performance levels across different models, a dual approach was adopted. For models with lower performance capabilities, a 'lighter load' method (see **Figure 9**) was developed, wherein only the figure descriptions were input for processing. In contrast, more powerful models like 'Llama 3 8B instruct' continued to receive the entire PDF for analysis<sup>8</sup>. During this phase, new fine-tuned instructional models were also identified and incorporated into the testing regimen. These models allow an instruction on how they should behave, leading to the selection of the best-performing model, 'Llama 3 8B Instruct'.

The final stage of methodology development involved refining this model further to produce a JSON output that included the extracted figure numbers, and the relevant data associated with each figure. A quantized version of 'Llama 3 70B Instruct' was also tested. However, the performance degraded too much for it to be useful. **Table 1** provides a comprehensive overview of tested LLMs, their parameters as well as more information about the model.



**Figure 9** Workflows for LLM development; source: own depiction

<sup>8</sup> For an overview of differences between a “heavy load” and “light load” see **Figure 9**.

**Table 1** Overview tested large language models (LLMs); source: own depiction

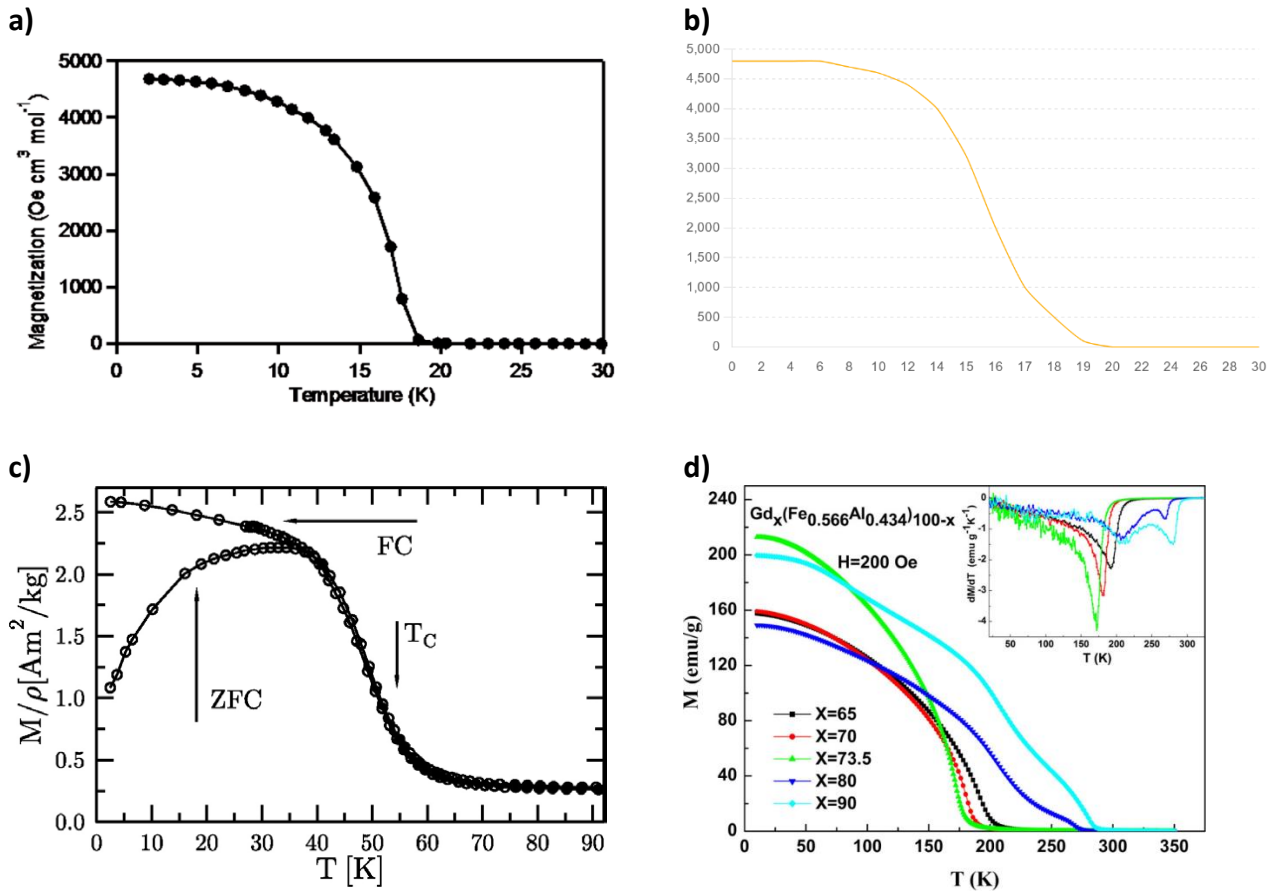
LLMs	Parameter (in billions)	Information
ChatGPT-2	1.5	Transformer based architecture; created by OpenAI; pre-trained on 8 million web-pages (openAI, 2019)
CHatGPT-4	unknown	Transformer based architecture; created by OpenAI (openAI, 2024)
Llama 2 7B	7	Transformers based architecture; auto regressive language model; pre-trained on 2 trillion tokens of data; trained with supervised learning and reinforcement learning from human feedback (Meta, 2023)
Llama 2 7B Chat	7	Fine-tuned for chat of Llama 2 7B; (Meta, 2023)
Mistral 7B	7	Transformer based architecture; utilized group query attention, sliding window attention; unpublished data-set for training (Jiang et al., 2023)
Mistral 7B instruct	7	Fine-tuned version of 'Minstral 7B' (Jiang et al., 2023)
Llama 3 70B instruct awq	70	Transformers based architecture; auto regressive language model; trained with supervised learning and reinforcement learning from human feedback; quantized version of 'Llama 3 7B instruct' based on activation-aware weight quantization
Llama 3 8B instruct	8	Transformers based architecture; auto regressive language model; pre-trained on 15 trillion tokens of data; trained with supervised learning and reinforcement learning from human feedback (Meta, 2024)

## 4.2 Digitization methods

After looking at different LLMs and determining 'Llama 3 8B instruct' to be the most promising, digitization posed the next challenge as there is no well-known tool or methodology that automatically digitizes plots. As such they do not have a solidified workflow.

Originally all digitization methods were first tested on the 3800TI NVIDIA Card, however Plot2Spectra and ChartDETE required older versions of software that are not available to this hardware thus the high-powered computing (HPC) with its A100 had to be used. In order to test the different digitization methods, they were fed a set of four plot images (see **Figure 10**). Depending on performance further investigation was performed, utilizing a set of fifteen plot images (see appendix **11.1**).

**Figure 10** Selection Images used for initial testing additional utilized in appendix 11.1; a &b)(Imoto et al., 2017) c)(Cieślak et al., 2005) d)(Liu et al., 2013)



The first digitization method examined was called 'AI Curve' from Science AI of the Chinese academy of sciences (Zhou and Lu, 2022). In the process of implementing this digitization tool, code flowcharts were developed. This code is based on the 'YOLO'-architecture with version 4 being used (Zhou and Lu, 2022). By first converting the PDF page into images, the code used YOLO to scan these for figures of graphs. After scanning all pages, it would then deposit the figures and attempt to digitize the curves while using traditional image processing techniques.

By analyzing the vertical and horizontal intensities, combined with hue line transformation YOLO could determine the location of the x and y axis based on their orientations and positions. This was done by filtering the longest lines of either horizontal or vertical orientation (Zhou and Lu, 2022). Text extraction was then performed with an OCR, while the graph reconstruction was done by mapping the pixel positions of the data point on the graph back to the corresponding values on the axes (Zhou and Lu, 2022).

After investigating 'AI Curve', web-plot digitizer (Rohatgi, 2024) was tested. This tool unfortunately requires human input for both the scale and plot area leading it to be a poor choice for automation. It

---

functions by having the user first define two points on the x-axis and two points along the y-axis to allow for the scaling of the graph. Next the user selects the color of the desired line before highlighting the area the graph is in. It is essential, that the graph is the only thing in the highlighted area because web-plot digitizer will then attempt to digitize the graph and output a CSV document based on it.

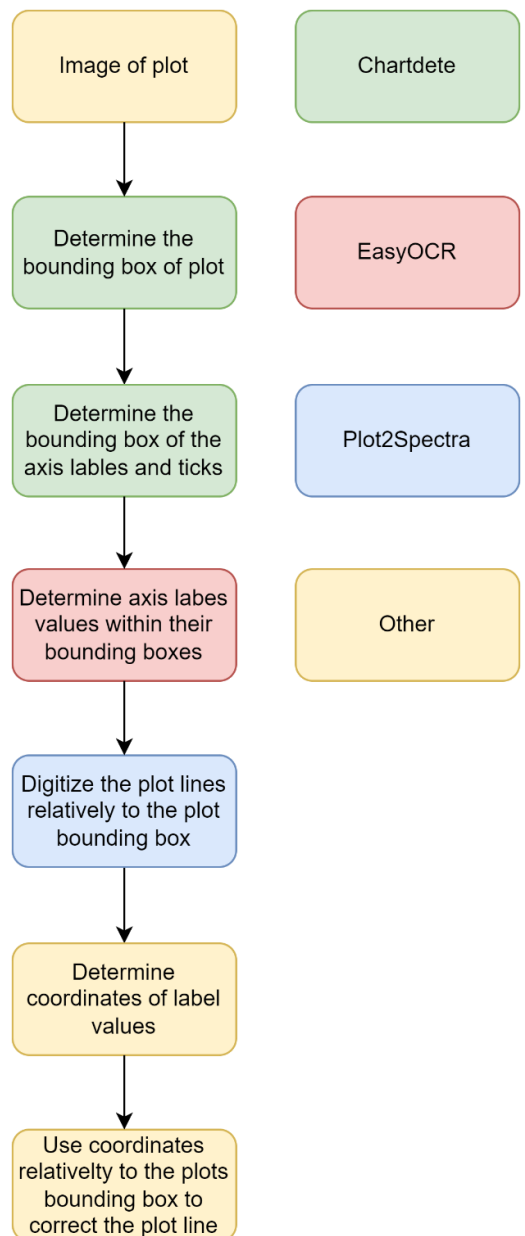
A different tool investigated for automatic digitization was ChatGPT-4. It being a multi model system capable of image recognition and image production (openAI, 2024) lead to tests to see if it could digitize a plot. This however, only worked once with the result not being able to be replicated, thus continuing to highlight the issue of volatility in ChatGPT-4.

'DePlot' was another tested model during research. It is an advanced version of the original MATCHA framework, designed to convert visual data, such as plots and graphs, into a structured table (Liu et al., 2023). 'DePlot' functions as an end-to-end image-to-text transformer model, building on the foundational principles of the MATCHA model, which itself is an evolution of the PIX-to-structure model (Liu et al., 2023). These type of models are specifically designed to generate text representations of images, facilitating the automatic interpretation of visual data (Liu et al., 2023). The 'DePlot' model particularly is geared toward integration within a multi-model process involving a LLM, where 'DePlot' produces a table or structured data based on the values extracted from a given plot (Liu et al., 2023). During testing 'DePlot' performed decently for graphs with a singular line, however, due to other limitations discussed in results (see chapter 5.1.2), it was not expanded upon.

Plot2Spectra was the final digitization tool tested in this context. Originally it was designed for digitizing spectrographs (Jiang et al., 2022) and its code functions by first using an object detector model, FCOS (discussed in chapter 3.6.4) to determine the area of the plot and align axis (Jiang et al., 2022). In another step it feeds this plot area into a plot data extraction module which, uses the schematic segmentation model of an Efficient Residual Factorized ConvNet (ERFNet) (Jiang et al., 2022). ERFNet in turn would feed the plot into an optical flow algorithm where it would then replot the graph. This produced the best-looking graphs whose general shape would match the input graphs. These values however, were relevant to the bounding box, not the actual values the plot represented, therefore requiring further code. Since Plot2Spectra is based on 'optical flow' it does have an inbuilt mechanism capable of utilizing CRAFT as well as a convolutional neural network to attempt to correct the values along the x-axis. The performance of both is quite limited though, which is why a different solution was developed.

This solution to correct the values of Plot2Spectra involved feeding the plots to ChartDETE. This code recognizes different chart elements with the main implementation of this being able to recognize ticks

and respective labels on the x and y-axis<sup>9</sup>. These x and y-labels were then fed to easyOCR which would convert them to text by utilizing 'character region awareness for text detection' (CRAFT) (see chapter 3.6.2) (JaidedAI, 2024). In an additional step it uses a convolutional recurrent neural network (CRNN) to detect the actual text within these regions (JaidedAI, 2024). This data and its relative positions were then saved to a JSON-file format and applied to the Plot2Spectra output to produce accurate graphs. This resulted in the final workflow shown below in **Figure 11**.



**Figure 11** Final digitization workflow; source: own depiction

<sup>9</sup> To recognize ticks and labels it uses a region-based convolutional neural network (RCNN) (Yan et al., 2023).

---

## 5 Results

---

The primary goal of this research was to evaluate the performance of various LLMs in identifying and extracting magnetization vs. temperature (M vs T) plots from scientific PDFs and find appropriate digitization methods. To address this objective, different LLMs, including ChatGPT-4 (openAI, 2024) and Llama 3 8B instruct (Meta, 2024), were tested. Each model was assessed based on its accuracy, ability to handle complex and large documents, and the consistency of its performance. In order to optimize their effectiveness, they were also subjected to various prompt engineering strategies (see chapter **5.1**).

The results presented in the following detail the performance of these models, with a focus on the specific challenges encountered, such as false positives and missed plots, particularly with ChatGPT-4 (see chapter **5.1.1**). The subsequent sections also highlight the improvements achieved through refined prompts and the adoption of Llama 3 8B instruct, which demonstrated higher accuracy and speed (see chapter **5.1.3**). In order to properly understand the test-results a table has been built that shows all the tested PDF-files with the correct plots, that the LLMs should have found (see **Table 2**). This table serves as a comparison to the actual LLM output tables.

**Table 2** Correct data of PDFs that contain thermomagnetic plots given to LLMs; source: own depiction

Document name	Document Title	Figures & Sections	Material
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	Phase diagram, ferromagnetic martensitic transformation and magnetoresponsive properties of Fe-doped MnCoGe alloys	2(a)(b)	(a) MnCo <sub>1-x</sub> Fe <sub>x</sub> Ge alloys (b) Mn <sub>1-x</sub> Fe <sub>x</sub> CoGe alloys
035220_1_online.pdf (Zhang et al., 2016)	Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity	2	Gd <sub>x</sub> Fe <sub>0.566</sub> Al <sub>0.434</sub> 100-x (x = 65-90)
122405_1_online.pdf (Liu et al., 2013)	Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system by establishing a Curie-temperature window	3	Mn <sub>1-x</sub> Co <sub>x</sub> NiGe
app135z2p11.pdf (Kupczyk et al., 2019)	Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe70-xCoxMn10Mo5B15 (0 ≤ x ≤ 0.5) Alloys at Low Magnetizing Field	1(a)(b)(c),2(a)(b)(c)	1,2 Fe <sub>70-x</sub> Co <sub>x</sub> Mn <sub>10</sub> Mo <sub>5</sub> B <sub>15</sub> (x=0,0.25,0.5) 2 is after annealing
ArticleSupra2020.pdf (Hcini et al., 2020)	Structural Analysis and Theoretical Investigations of the Magnetocaloric Effect for La <sub>0.7</sub> Ba <sub>0.15</sub> Ag <sub>0.15</sub> MnO <sub>3</sub> Manganite Prepared Using Sol-Gel Route	2(a),3(c),8	La <sub>0.7</sub> Ba <sub>0.15</sub> Ag <sub>0.15</sub> MnO <sub>3</sub>
BMS.pdf (Anwar et al., 2015)	Magnetocaloric response of La <sub>0.70</sub> Ca <sub>0.1</sub> Sr <sub>0.2</sub> Fe <sub>0.1</sub> Mn <sub>0.9</sub> O <sub>3</sub> perovskite for magnetic refrigeration	3(a)	La <sub>0.7</sub> Ca <sub>0.1</sub> Sr <sub>0.2</sub> Fe <sub>x</sub> Mn <sub>1-x</sub> O <sub>3</sub> (x=0,.1)
HCMO.pdf (Mazumdar and Das, 2021)	Role of 3d-4f exchange interaction and local anti-site defects in the magnetic and magnetocaloric properties of double perovskite Ho <sub>2</sub> CoMnO <sub>6</sub> compound	2,3(a)	Ho <sub>2</sub> CoMnO <sub>6</sub>
JAP115123908.pdf (Saikar et al., 2014)	Magnetic properties of mixed spinel BaTiO <sub>3</sub> -NiFe <sub>2</sub> O <sub>4</sub> composites	4(a)(b),5(a)(b)(c)(d),6(a)(b)(c)(d)	(100-x)BaTiO <sub>3</sub> -(x)NiFe <sub>2</sub> O <sub>4</sub> (x=0,20,40,50,60,80,100)
Krishna_Murthy_2015_J._Appl._Phys._48_355001.pdf (Krishna Murthy et al., 2015)	Giant magnetocaloric effect in Gd <sub>2</sub> NiMnO <sub>6</sub> and Gd <sub>2</sub> CoMnO <sub>6</sub> ferromagnetic insulators	1(a)(b)	(a)Gd <sub>2</sub> NiMnO <sub>6</sub> (b)Gd <sub>2</sub> CoMnO <sub>6</sub>
Metal_Substitution_Effect_on_a_Three-Dimensional_C.pdf (Imoto et al., 2017)	Metal Substitution Effect on a Three-Dimensional Cyanido-Bridged Fe Spin-Crossover Network	6(a)	(Fe <sub>x</sub> Co <sub>1-x</sub> ) <sub>2</sub> [Ni(CN) <sub>6</sub> ] <sub>4</sub> -pyridineal(doxime) <sub>8</sub> zH <sub>2</sub> O (x=0, 0.1,0.25, 0.5, 0.75, 1)
Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO <sub>3</sub> and LaFeO <sub>3</sub> CrO <sub>5</sub> CrO <sub>5</sub> 03.pdf (Xia et al., 2022)	Magnetic Properties and Mossbauer Study of Perovskite LaFeO <sub>3</sub> and LaFe <sub>0.5</sub> Cr <sub>0.5</sub> O <sub>3</sub>	4(a)5(a)	4(a) LaFeO <sub>3</sub> , LaFe <sub>0.5</sub> Cr <sub>0.5</sub> O <sub>3</sub> 5(a) LaFe <sub>0.5</sub> Cr <sub>0.5</sub> O <sub>3</sub> and predicted
PhysRevB.81.214406.pdf (Khovayko et al., 2010)	Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape	1,2	Ni <sub>50</sub> Mn <sub>36</sub> Co <sub>1</sub> Sn <sub>13</sub>
PhysRevB.93.184403.pdf (Bhattacharya et al., 2016)	Realization of two-dimensional ferromagnetism with giant coercivity in ultrathin β-Ni(OH) <sub>2</sub> layers grown on a MoS <sub>2</sub> surface	4(a)	FC, LCZFC, LCFC, MC ZFC, MC Ni-acetate concentrations designated as LC (5mM), MC (12.5mM),
Tunable_Giant_Magnetocaloric_Effect_in_MnFe2PSi_M.pdf (Thang et al., 2016)	Tunable Giant Magnetocaloric Effect in (Mn,Fe)2(P,Si) Materials by Co-B and Ni-B Co-Doping	3,8,12	3 Mn <sub>1.00</sub> Fe <sub>0.85</sub> Co <sub>0.10</sub> P <sub>0.95</sub> -zSi <sub>0.45</sub> Bz (z=0,0.02,0.04,0.06) 8 Mn <sub>1.00</sub> Fe <sub>0.95</sub> -zCo <sub>z</sub> P <sub>0.95</sub> Si <sub>0.45</sub> Bz <sub>0.04</sub> (z=0.07,0.09,0.11,0.13) 12 Mn <sub>1.00</sub> Fe <sub>0.95</sub> -zNi <sub>z</sub> P <sub>0.51</sub> Si <sub>0.45</sub> Bz <sub>0.04</sub> (z=0.06,0.08,0.10,0.12)
zno-serie5.pdf (Silva Tupan et al., 2019)	Structural and magnetic characterization of (Zn <sub>1-x</sub> YFe <sub>x</sub> O) <sub>2</sub>	2	(Zn <sub>0.95</sub> Fe <sub>0.05</sub> )O, (Zn <sub>0.95</sub> Fe <sub>0.02</sub> Co <sub>0.03</sub> )O

---

## 5.1 LLMs

This section will focus on the results of the LLMs. While ChatGPT 2 was initially tested it yielded no meaningful results. This resulted in the subsequent deployment GPT-4, which was ultimately abandoned for reasons outlined in the preceding chapter (see chapter 5.1.1). Following chapters will present the rationale behind the selection of Llama 3 8B as the final instructional model (see chapters 5.1.2 and 5.1.3).

### 5.1.1 GPT-4

As mentioned in the introduction of the overarching chapter ChatGPT-4 was tested, after ChatGPT-2 was deemed to be a futile endeavor. ChatGPT-4 was therefore utilized in a rather straightforward manner, where a single PDF was uploaded, and the simple question, "Does the PDF contain a figure about magnetization as a function of temperature?" was posed. Upon confirming the presence of such a figure, a follow-up request, "Can you please extract the figure?" was made. However, after being prompted a second time ChatGPT-4 demonstrated an inability to complete this task, and a trend of poor performance was noted. Since the time requirement of single PDFs was quite lengthy it was decided to have the model process multiple PDFs at once. Subsequently, ChatGPT-4 was provided with ten documents and asked, "Which of these contain M vs T curves?" however, it was unable to successfully identify the relevant figures. Further prompting of ChatGPT-4 revealed that the issue might have stemmed from the PDFs being too large or the prompts being too vague as determined when ChatGPT was asked why it was struggling. This recognition led to the first significant evolution of the prompt.

Further evolutions of the prompts, described in the following, reflect a systematic process of refinement aimed at enhancing the clarity, specificity, and comprehensiveness of the instructions provided. The second prompt, "does this contain M vs T plots check the whole document," aimed at addressing the limitations of the first prompt by specifying that the entire document should be searched. This addition reduced ambiguity, guiding the process more effectively. However, this version still did not consider potential challenges that could arise during the scanning process, nor did it request detailed information about the findings.

The third prompt, "does this contain M vs T plots please scan the whole document if you can't scan the whole document state why not," further refined the instructions by introducing a contingency plan. It not only instructed the AI to search the entire document but also requested an explanation if the task

---

could not be completed. This addition improved the robustness of the prompt by anticipating potential obstacles and ensuring that any limitations in the search process were documented.

Finally, the fourth prompt, "this contains M vs T plots please scan the whole document if you can't scan the whole document state why not and state which figure," represented a more refined version. It incorporated comprehensive instructions, including both a thorough search directive and a request for specific details about the findings. By asking ChatGPT-4 to identify the exact figure containing the M vs T plots, this prompt ensured that the response would not only confirm the presence of such plots but also provide precise information about their location within the document.

The following prompt engineering step was meta-prompting. This process caused the initial vague query, "Which of these contain M vs T curves?" to be refined to an even more detailed instruction:

"Please search for any sections, figures, or charts that depict magnetization vs. temperature curves. Focus on any part of the document that discusses magnetic properties, phase transitions, or temperature-dependent magnetic behavior. If there's a table of contents, index, or list of figures, start there to guide the search towards relevant chapters or sections. Additionally, look for keywords such as 'magnetization,' 'temperature dependence,' 'Curie temperature,' 'magnetic susceptibility,' or similar terms that are likely associated with such curves. Name such figures."

This prompt engineering marked a major improvement in the interaction with ChatGPT-4, as it finally allowed the model to provide more accurate and relevant responses. To test this several documents known to contain M vs T plots were fed to ChatGPT-4 individually. In total ChatGPT-4 failed to find any of the M vs T curves in 3 out of 15 instances. Regarding requested figures it found 16 out of the possible 25 figures (64%). However other issues like it often missing graphs and making mistakes with the composition as well as a single false positive arose. This data is shown in **Table 3**. Below is an example response of GPT-4 when functioning correctly and provided with a PDF that does contain the relevant plots.

"Yes, the document contains sections, figures, and charts that depict magnetization vs. temperature (M vs T) curves for various materials. Specifically:- **\*\*Material:\*\*** Mn<sub>1.00</sub>Fe<sub>0.85</sub>Co<sub>0.10</sub>P<sub>0.55</sub>-zSi<sub>0.45</sub>Bz compounds - **\*\*Figures:\*\*** Figure 3 shows magnetization as a function of temperature for these compounds.- **\*\*Material:\*\*** Mn<sub>1.00</sub>Fe<sub>0.95</sub>-zCozP<sub>0.51</sub>Si<sub>0.45</sub>B<sub>0.04</sub> compounds - **\*\*Figures:\*\*** Figure 8 presents M-T curves for these materials.- **\*\*Material:\*\*** Mn<sub>1.00</sub>Fe<sub>0.95</sub>-zNi<sub>z</sub>P<sub>0.51</sub>Si<sub>0.45</sub>B<sub>0.04</sub> compounds - **\*\*Figures:\*\*** Figure 12 illustrates the magnetization versus temperature curves for these compounds. These figures indicate the influence of Co (Ni) and B co-doping on the Curie temperature, thermal hysteresis, and magnetic transitions of the materials, aligning with the search keywords provided." (ChatGPT-4)

**Table 3** Individual document performance ChatGPT-4; source: own depiction

Document name	Results
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	Correct but missed a and b
035220_1_online.pdf (Zhang et al., 2016)	Failed
122405_1_online.pdf (Liu et al., 2013)	Found 3 false positive of 4
app135z2p11.pdf (Kupczyk et al., 2019)	Correct but missed comp nuance
ArticleSupra2020.pdf (Hcini et al., 2020)	Correct but missed a and 3
BMS.pdf (Anwar et al., 2015)	Correct
HCMO.pdf (Mazumdar and Das, 2021)	Correct
JAP115123908.pdf (Sarkar et al., 2014)	Correct
Krishna_Murthy_2015_J._Phys._D _Appl._Phys._48_355001.pdf (Krishna Murthy et al., 2015)	Failed
Metal_Substitution_Effect_on_a_Three-Dimensional_C.pdf (Imoto et al., 2017)	Failed
Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and M ssbauer Study of Perovskite LaFeO3 and LaFe0 5Cr0 5O3.pdf (Xia et al., 2022)	Found 4 missed a and 5
PhysRevB.81.214406.pdf (Khovaylo et al., 2010)	Required additional prompting missed 1
PhysRevB.93.184403.pdf (Bhattacharya et al., 2016)	Correct but missed the a
Tunable_Giant_Magnetocaloric_Effect_in_MnFe2PSi_M.pdf (Thang et al., 2016)	3 correct missed 8, 12 wrong comp
zno-serie5.pdf (Silva Tupan et al., 2019)	Correct

Testing was also performed with single PDFs that do not contain M vs T plots however, while ChatGPT-4 would often correctly state that there were no M vs T plots it would occasionally give a false positive claiming there was one where there was not. Examples of ChatGPT-4s response are given below, beginning with the correct response:

“The document does not directly provide magnetization vs. temperature (M vs T) curves or discuss magnetization, temperature dependence, Curie temperature, or magnetic susceptibility in the context of magnetocaloric or magnetic materials. Instead, it focuses on heterostructured materials, including their mechanical properties, processing, and applications, without specific reference to magnetic properties or behaviors related to magnetization.” (ChatGPT-4)

“Yes, the document includes sections and figures that depict magnetization vs. temperature (M vs T) curves, as well as related magnetic properties across various materials and nanoparticle configurations. Specifically: Figures 3a and 3b describe the calorimetric and magnetometric methods used to evaluate the heating efficiency of magnetic nanoparticles (NPs), which indirectly relate to their magnetization properties under alternating current (AC) field excitations. Figure 4 contrasts the simulated magnetic spin states of cubic and spherical nanoparticles to highlight differences in magnetization due to shape anisotropy, affecting their temperature response in magnetic hyperthermia applications. Figure 5 illustrates the concept of exchange spring nanomagnets, showing the magnetic interactions between hard core and soft shell magnetic layers, and compares the heating efficiencies of single-component and core-shell nanoparticles. These figures and sections detail how structural modifications, including size, shape, and chemical composition, impact the magnetic and thermal responses of iron oxide nanoparticles, emphasizing the relationship between magnetization and temperature in the context of magnetic hyperthermia applications.” (ChatGPT-4)

Even with the errors this output was a marked improvement over prior ones, where it would consistently fail to identify the figures. However, due to ChatGPT-4 taking approximately one minute per PDF, it was decided to start testing with multiple PDFs at once which lead to a prompt evolution to include the line “Answer with a table of the documents which figure and what material.” and ChatGPT-4 was fed 10 documents at once with the updated prompt. This is when more errors with ChatGPT-4 started to emerge. Out of the 10 documents provided, the model ended up analyzing 8 and produced results with 2 instances of false positives. The first being a figure that showed the change of magnetization with temperature vs. temperature instead of a magnetization vs. temperature plot (see document “appapp135z2p11.pdf” in **Table 4**) while the other plot showed delta  $S_m$  vs. temperature instead (see document “Krishna\_Murphy\_2015” in **Table 4**). Additionally, several figures were missed as highlighted by brackets “{ }” in **Table 4**.

**Table 4** First ChatGPT-4 generated table, missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font; source: own depiction

Document	Figure	Material
HCMO.pdf (Mazumdar and Das, 2021)	Fig. 3{(a)}	<del>Temperature dependence of magnetization under various external magnetic fields in a zero field cooled warming (ZFCW) cycle</del>
BMS.pdf (Anwar et al., 2015)	Fig. 3 {a}	<del>Magnetization vs. temperature curve (H = 0.5 T)</del>
app135z2p11.pdf (Kupczyk et al., 2019)	Figs. 1{(a)(b)(c)} 2(a)(b)(c) & 3	Specific magnetization as a function of temperature for amorphous Fe <sub>70-x</sub> Co <sub>x</sub> Mn <sub>10</sub> Mo <sub>5</sub> B <sub>15</sub> alloys in the as-quenched state and after accumulative annealing
122405_1_online.pdf (Liu et al., 2013)	Fig. 3	M(T) curves of Mn <sub>1-x</sub> Co <sub>x</sub> NiGe in low and high magnetic fields
035220_1_online.pdf (Zhang et al., 2016)	Fig. 2	Temperature dependence of the magnetization of Gdx (Fe <sub>0.566</sub> Al <sub>0.434</sub> ) <sub>100-x</sub> ribbons under an applied field of 200 Oe
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	Fig.2 {a b}	Temperature dependence of magnetization of MnCo <sub>1-x</sub> Fe <sub>x</sub> Ge alloys and Mn <sub>1-x</sub> Fe <sub>x</sub> CoGe alloys measured under a magnetic field of 100 Oe
Krishna_Murthy_2015_J._Phys._D_Appl._Phys._48_355001.pdf (Krishna Murthy et al., 2015)	Figs. 1 {(a)(b)}, 3	Temperature-dependent magnetization in zero-field cooled (ZFC) and field-cooled (FC) modes with 0.01 T dc field for GNMO and GCMO samples
JAP115123908.pdf (Sarkar et al., 2014)	Fig. 4 ,{5,6}	Field cooled magnetization vs temperature curve at magnetic field H = 1000 Oe for various BTO-NiFe <sub>2</sub> O <sub>4</sub> composites

After testing 10 PDFs at once the number of PDFs was reduced to 8 where the number of false positives only increased and a false negative, an instance of ChatGPT-4 reporting no plots for the document “ArticleSupra2020.pdf”, was also observed. **Table 5** highlights these missed figure plots in brackets “{ }”, while false positives are noted in *italics*, as well as ~~struck through font~~.

**Table 5** Missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font ChatGPT-4; source: own depiction

Document	Figure	Material
122405_1_online.pdf (Liu et al., 2013)	3, <del>6</del>	Mn1-xCoxNiGe alloys
035220_1_online.pdf (Zhang et al., 2016)	2, <del>4</del>	Gd-Fe-Al based amorphous/nanocrystalline systems
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	2	Fe-doped MnCoGe alloys
JAP115123908.pdf (Sarkar et al., 2014)	4, {5,6}	Mixed spinel BaTiO3-NiFe2O4 composites
HCMO.pdf (Mazumdar and Das, 2021)	2, 3	HoCoMnO system
BMS.pdf (Anwar et al., 2015)	3	La0.7Ca0.1Sr0.2MnO3 and La0.7Ca0.1Sr0.2Fe0.1Mn0.9O3 compounds
ArticleSupra2020.pdf (Hcini et al., 2020)	{2(a),3(c), 8b}	{La0.7Ba0.15Ag0.15MnO3}
app135z2p11.pdf (Kupczyk et al., 2019)	1, 2, <del>3</del>	Amorphous Fe70-xCoxMn10Mo5B15 alloys (0 ≤ x ≤ 0.5)

These results also highlighted a different, but concerning problem with ChatGPT-4, where feeding 7, previously tested PDFs and using the same prompt led to new occurring errors. The PDF documents in question were identical to those used in prior attempts, as evidenced by a comparison of **Table 3**, **Table 4**, **Table 5** and **Table 6**. This observation leads to the conclusion that there is volatility within this GPT model. It is also worth noting that ChatGPT-4 struggles to consistently note which sub-plots within a figure contain the desired information.

**Table 6** ChatGPT-4 results, missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font; source: own depiction

Document Title	Figures & Sections	Material	Notes
app135z2p11.pdf (Kupczyk et al., 2019)	Figs. 1 {2}, <del>3</del>	Amorphous Fe70-x Cox Mn10 Mo5 B15 alloys	M vs T curves for as-quenched and annealed states, focusing on Curie point analysis.
122405_1_online.pdf (Liu et al., 2013)	<del>Figs. 1-6</del> , Figs {3}	MnNiGe-CoNiGe systems	Detailed on magnetocaloric effect and Curie-temperature window.
035220_1_online.pdf (Zhang et al., 2016)	Figs. 2-4 {3}	Gdx(Fe0.566Al0.434)100-x ribbons	Examines magnetocaloric properties in amorphous/nanocrystalline systems.
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	Figs. 2{(a)(b)} <del>5</del>	Fe-doped MnCoGe alloys	Studies magneto-responsive properties including magnetocaloric effects.
HCMO.pdf (Mazumdar and Das, 2021)	{Figs,2 3(a)}	Ho2CoMnO6 compound	Focuses on magnetic and magnetocaloric properties. Detailed sections not provided due to document structure.
BMS.pdf (Anwar et al., 2015)	Figs. 3{a}, <del>4</del> , <del>5</del>	La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite	Discusses magnetocaloric properties and potential for magnetic refrigeration.
ArticleSupra2020.pdf (Hcini et al., 2020)	Figs. 2{(a)}, 3{(a)}, <del>4</del> , <del>5</del> , <del>6</del> , <del>7</del> , <del>8</del> , <del>9</del>	La0.7Ba0.15Ag0.15MnO3 manganite	Studies magnetocaloric effect, spontaneous magnetization, and applies theoretical models for magnetic entropy change.

It was then decided to include PDFs that do not contain M vs T plots, which is why **Table 7** shows ChatGPT-4s results of 7 PDFs. 5 of these contained M vs T plots whereas 2 did not. Initially ChatGPT-4 failed to provide any results and required additional regeneration of the prompts until it did. Upon

closer inspection it was discovered that the model failed to mention PDFs containing no relevant plots whatsoever. Additionally, it ended up producing false positives. The highlighting of plots without an actual M vs. T plot also continued, further emphasizing the volatility of the model. It was therefore decided to move on to more open-source models that can be restricted to specific versions of the model.

**Table 7** ChatGPT-4 with false PDFs, missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font; source: own depiction

Document Title	Figure & Sections	Material Description
122405_1_online.pdf (Liu et al., 2013)	Fig. 3	Mn <sub>1-x</sub> CoxNiGe alloys in low and high magnetic fields
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	<del>Fig. 3</del> {2(a)(b)}	MnCo <sub>1-x</sub> FexGe and Mn <sub>1-x</sub> FexCoGe alloys
app135z2p11.pdf (Kupczyk et al., 2019)	Fig. 1{(a)(b)(c)}, Fig. 2{(a)(b)(c)}	Amorphous Fe <sub>70</sub> Mn <sub>10</sub> Mo <sub>5</sub>
035220_1_online.pdf (Zhang et al., 2016)	Fig. 2, <del>Fig. 4</del>	Gdx(Fe <sub>0.566</sub> Al <sub>0.434</sub> ) <sub>100-x</sub> ribbons
ArticleSupra2020.pdf (Hcini et al., 2020)	Fig. 2{(a),3(a),8} <del>Fig. 5</del>	La <sub>0.7</sub> Ba <sub>0.15</sub> Ag <sub>0.15</sub> MnO <sub>3</sub> manganite

### 5.1.2 Other models and the path to 'Llama 3 8B instruct'

Moving on from ChatGPT-4 other models were tested. All of these utilize the python package 'transformers' by Hugging Face (Wolf et al., 2020). As these models run on local hardware the prompt was adjusted accordingly for these models, as they were only required to search figure descriptions. The prompt was modified to reflect this and changed to:

“Please search the text for any sections, figures, or charts that depict magnetization vs. temperature (M vs T) curves. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. Check the table of contents, index, or list of figures if available to guide the search. Answer with 'Yes' if such curves are found then state what material and which figures, or 'No' if there are none”

---

The following models were excluded due to their suboptimal performance: 'ChatGPT-2'<sup>10</sup> (openAI, 2019), 'Llama 2 7B'<sup>11</sup> (Meta, 2023), 'Llama 2 7B chat'<sup>12</sup> (Meta, 2023) and 'Mistral-7B'<sup>13</sup> (Jiang et al., 2023). In contrast to these, the 'Mistral 7B instruct' (Jiang et al., 2023) was able to identify M vs T plots with occasional false positives, however the instructional performed well<sup>14</sup>, making it clear that an instructional based model was the best way to proceed. These are models that have been previously fine-tuned to accept an instruction such as “You are a pirate ChatBot” and from then on will respond to all prompts as a pirate (Jiang et al., 2023). This is highlighted below in the response, where the instruction is given in Bold, and the provided text for the AI to analyze is given in italics.

“<s> [INST] **Please check this figure description and determine the figure it depicts a magnetization vs. temperature (M vs T) curve. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY WITH 'YES' IF SUCH CURVES ARE FOUND THEN STATE WHAT MATERIAL , OR SIMPLY 'NO' IF THERE ARE NONE** [/INST]*Figure 4. Specific heat derived from Differential scanning calorimetry (DSC) measurements for the Mn<sub>1.00</sub>Fe<sub>0.85</sub>Co<sub>0.10</sub>P<sub>0.55</sub>-zSi<sub>0.45</sub>Bz compounds measured in the zero field upon cooling and heating.*</s> Inset shows the magnetic susceptibility  $\chi$  vs. temperature T for Mn<sub>1.00</sub>Fe<sub>0.85</sub>Co<sub>0.10</sub>P<sub>0.55</sub>-zSi<sub>0.45</sub>Bz compound. NO, the figure description does not mention a magnetization vs. temperature (M vs T) curve. The figure shows specific heat data and an inset of magnetic susceptibility vs. temperature.</s> ” (Mistral 7B instruct)

For this model prompt engineering was once again utilized, where the prompt evolved from the one mentioned previously to the following one in order to put greater emphasis on the type of response desired:

“Please check this figure description and determine the figure it depicts a magnetization vs. temperature (M vs T) curve. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY WITH 'YES' IF SUCH CURVES ARE FOUND THEN STATE WHAT MATERIAL , OR SIMPLY 'NO' IF THERE ARE NONE”

Although 'Mistral 7B instruct' was able to provide satisfactory responses one major drawback was the time it took, which was determined to be about five minutes per figure description. While research continued in an effort to improve Mistral 7B instructs performance the company Meta (Meta, 2024) announced that they had released a new, open-source model called 'Llama 3 8B' with an instructional version (Meta, 2024). Upon testing, this model demonstrated the capability to handle more data at the same time and exhibited faster processing speeds than Mistral 7B instruct, even when fed an entire PDF at once. Further testing was therefore conducted with 'Llama 3 8B instruct'.

---

<sup>10</sup> ChatGPT-2 was excluded due to being unusable.

<sup>11</sup> Llama 2 7B was excluded due to poor performance with responses as it would generate too many false positives in addition to generating non-relevant text.

<sup>12</sup> Llama 7B chat was excluded due to poor performance. While better than Llama 2 7B it was still ineffective.

<sup>13</sup> Mistral-7B was excluded due to poor performance in recognition.

<sup>14</sup> Ultimately, this model ended up being excluded due to its lack of speed and inability to handle entire PDF-files.

---

### 5.1.3 'Llama 3 8B instruct'

Initially Llama 3 was tested in a similar manner to Mistral 7B instruct (see chapter 5.1.2) where it was fed single figure descriptions. However, due to its enhanced performance, it was assigned a harder task of processing an entire PDF at once, providing an accurate response. Consequently, it was decided to test it with multiple PDFs by utilization of the TU Darmstadt High Powered Computing (HPC) called 'Lichtenberg II high-performance computer' (TU Darmstadt, 2024). The deployment of the improved hardware notably increased performance speed. While Llama 3 already exhibited the fastest processing, taking approximately 1 minute to analyze a PDF, it reached final speeds of 1 to 10 seconds when deployed to the HPC (see appendix 11.4). However, during this process, it was observed, that when PDFs reached a certain size they would exceed the token-limit of 8192 (Meta, 2024) causing the LLM to output nonsensical answers. For example; while the desired output looks like this:

“YES

The figure depicting the M vs T curve is Figure 3a, which shows the magnetization of samples as a function of temperature.”

the LLM would output unexpected replies such as repeating the number “20850” over and over. Its performance with this issue is as follows. Out of 15 PDFs with relevant plots only 8 ended up being analyzed by the Llama 3 8B instruct while the nonsensical responses occurred for all other PDFs, even those lacking an M vs T plot. It was later discovered that the issue of nonsensical replies was caused by the exceedance of the maximum positional embedding token limit.

For the 8 PDFs that were below the limit Llama 3’s response was determined to be correct, even though it is not very useful for the end goal of figure extraction from a paper due to the aforementioned token limit. Thus, prompt engineering was once more deployed in an effort to produce a usable JSON format.

The instruction was therefore changed from where pdf\_text is extracted text from the PDF

```
“messages = [  
  {"role": "system", "content": "Please check this document to see if it contains a figure That depicts a magnetization vs. temperature (M vs T) curve. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY WITH a JSON return the following json {""Figure found"": ""yes or no""} ""},  
  {"role": "user", "content": pdf_text},  
]”
```

---

To

```
“messages = [  
  {"role": "system", "content": instruction},  
  {"role": "user", "content": pdf_text},  
  ]  
instruction = (  
  "Please check this document to see if it contains a figure that depicts a "  
  "magnetization vs. temperature (M vs T) PLOT. Keywords to look for include "  
  "'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. "  
  "Ignore all special symbols. ANSWER ONLY WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS "  
  "THE M VS T FIGURE AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE, ANSWER ONLY WITH a JSON  
  "  
  "return the following json {"  
  "\"title\": \"The title of the paper\";"  
  "\"Figure found\": \"If a magnetization vs. temperature PLOT is found YES or NO\";"  
  "\"Figure\": \"which figure is the MAGNETIZATION VS. TEMPERATURE PLOT\";"  
  "\"material\": \"what is the material that is the plot\";"  
  }”
```

Resulting in the Following output:

```
“{  
  "title": "Magnetocaloric response of La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite for magnetic refrigeration",  
  "Figure found": "YES",  
  "Figure": "Figure 3a",  
  "material": "La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite"  
}”
```

This produced the following **Table 8** where missing information is in brackets “{ }” and incorrect information is in *italics* and ~~struck through~~. It is worth noting that the Llama 3 8B instruct model also suffers from some volatility and inconsistencies as it would not always deliver a consistent JSON format for the document highlighted in bold. Furthermore, Llama 3 8B would occasionally give incorrect information. Nevertheless, it correctly identified 80% of the figures with a single false positive. However, for “Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3.pdf” the document technically exceeds the token limit. If one were to exclude it from the testing data due to its large size the success rate becomes 100% without false positives (see **Table 8**).

**Table 8** Llama 3 8B, testing output; source: own depiction

Document name	Document Title	Figures & Sections	Material
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	Phase diagram, ferromagnetic martensitic transformation and magnetoresponse properties of Fe-doped MnCoGe alloys	Figure 2{(a)(b)}	MnCoGe alloys{(a) MnCo 1-xFexGe alloys (b) Mn1-xFexCoGe alloys }
035220_1_online.pdf (Zhang et al., 2016)	Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity	Figure 2	Gd x(Fe0.566Al0.434)100-x(x=65-90) glassy alloys
122405_1_online.pdf (Liu et al., 2013)	{Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system by establishing a Curie-temperature window}	Figure 3	MnNiGe-CoNiGe {Mn1-xCoxNiGe (x=0.1,0.2,0.3)}
app135z2p11.pdf (Kupczyk et al., 2019)	Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe 70	Figure 1{(a)(b)(c)}, Figure 2{(a)(b)(c)}	Amorphous Fe70 xCoxMn10Mo5B15(0x0.5) alloys
BMS.pdf (Anwar et al., 2015)	Magnetocaloric response of La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite for magnetic refrigeration	Figure 3a	La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite
Krishna_Murthy_2015_J._Phys._D_Appl._Phys._48_355001.pdf (Krishna Murthy et al., 2015)	Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic insulators	Figure 1(a) and (b)	Gd 2NiMnO 6 and Gd2CoMnO 6
<i>Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3.pdf</i> (Xia et al., 2022)	{Magnetic Properties and Mössbauer Study of Perovskite LaFeO 3 and LaFe 0.5 Cr0.5 O3}	Figure 1 <del>shows the XRD patterns of LFO and LFCO {1,2}</del>	LaFeO3 and LaFe0.5Cr0.5O3
zno-serie5.pdf (Silva Tupan et al., 2019)	Structural and magnetic characterization of (Zn 1-x-yFexCoy)O	Figure 2	ZnO with iron and cobalt doping {(Zn0.95Fe0.05)O and (Zn0.95Fe0.02Co0.03)O}

In regard to the test for false positives it was determined, that all the 3 PDFs used for the test exceeded the positional token limit. Further research into solutions to this token limit was performed and the solution of cutting PDFs up into chunks close to their token limit was devised. The prompt was thus evolved for a final time into:

```
base_instruction = (
    "You are going to be given {document_part} of a document. "
    "Please check this document to see if it contains a figure that depicts a "
    "magnetization vs. temperature (M vs T) PLOT. Keywords to look for include "
    "'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. "
    "ANSWER ONLY WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS "
    "THE M VS T FIGURE if it is AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE, ANSWER ONLY WITH "
    "a JSON "
    "return the following Json {"
```

```

"\title\":"The title of the paper\";"
"\Figure found\":"If a magnetization vs. temperature PLOT is found YES or NO\";"
"\Figure\":"which figure is the MAGNETIZATION VS. TEMPERATURE PLOT\";"
"\material\":"what is the material that is the plot\}}"
)
#where document part is
document_part = "the whole" if total_text_tokens <= token_threshold else "a part"

```

The results of this prompt along with the corresponding outputs by Llama 3 8B are shown below in **Table 9**. However, this solution caused a degradation in performance with figures missed and a reduction in recognition precision of the material of the plot. It became evident that any document that was not divided into chunks due to the token limit only failed to include 1 out of the 8 figures. In the case of documents that had surpassed the token limit and therefore produced multiple JSON-files a total of 3 figures were missed (see **Table 9**).

When including the PDFs that are split into chunks, Llama 3 only misses all the plots in the PDF 4 times out of 15. However, when treating the figures individually only 14 out of the 25 figures (56%) were correctly identified, with 9 of them being false positives. Furthermore, the issue of false positives persisted, as an additional 13 documents from an entirely different scientific field as well as 3 other PDFs that did not include the desired plots were tested. It was found that there also a large amount of yes and no responses where one JSON would say a plot was found and the other would say there is no plot this occurred 6 out of the 16 PDF that did not contain the desired information but never for PDFs that did.

The results of the analysis by Llama 3 indicate that only 4 out of the 16 cases provided the completely correct answer, correctly identifying that no thermomagnetic plot was found. However, 10 out of the 16 cases initially indicated the presence of a plot but later retracted this claim, while the remaining 6 cases incorrectly identified the presence of an M vs T plot. Some of these false positives were easily identifiable as errors, given that the suggested material, such as meat or an individual's name, could never logically be associated with an M vs T plot, as shown in appendix **11.4**.

These findings highlight both the potential and the limitations of using Llama 3 7B instruct for this task. Despite being an open-source model that allows for easier improvement and customization, Llama 3 7B instruct currently performs only slightly worse than ChatGPT-4 when processing individual PDFs. This underscores the promise of the model, especially with potential future enhancements, but also emphasizes the need for continued research and development of LLMs for more accurate and reliable results in this specific application.

**Table 9** Llama 3 8B instruct sequential performance; source: own depiction

Document name	Document Title	Figures & Sections	Material	Notes	JSON Count
1-s2.0-S0304885312009754-main.pdf (Li et al., 2013)	Phase diagram, ferromagnetic martensitic transformation and magnetoresponse properties of Fe-doped MnCoGe alloys	2(a)(b)	MnCoGe alloys {(a) MnCo 1-xFexGe alloys (b) Mn1-xFexCoGe alloys}		1
035220_1_online.pdf (Zhang et al., 2016)	Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity	~4{2}	Gd-Fe-Al {Gdx(Fe0.566Al0.434)100-x(x = 65-90)}		1
122405_1_online.pdf (Liu et al., 2013)	Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system{ by establishing a Curie-temperature window}	3	MnNiGe-CoNiGe {Mn1-xCoxNiGe (x=0.1,0.2,0.3)}		1
app135z2p11.pdf (Kupczyk et al., 2019)	Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe70-xCoxMn10Mo5B15 (0 ≤ x ≤ 0.5) Alloys at Low Magnetizing Field	1{(a) (b) (c)}, 2{(a) (b) (c)}	amorphous Fe70xCoxMn10Mo5B15(0x0.5) alloys {1,2 Fe70-xCoxMn10Mo5B15(x=0,0.25,0.5)}		1
ArticleSupra2020.pdf	Structural Analysis and Theoretical Investigations of the Magnetocaloric Effect for La0.7Ba0.15Ag0.15MnO3 Manganite Prepared Using Sol-Gel Route	<del>2(d), 7</del> {2(a), 3(c), 8}	La0.7Ba0.15Ag0.15MnO3		2
BMS.pdf (Anwar et al., 2015)	{Magnetocaloric response of La 0.70 Ca 0.1 Sr0.2 Fe0.1 Mn 0.9 O 3 perovskite for magnetic refrigeration}	3(a)	La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite {La0.7 Ca0.1 Sr 0.2Fex Mn1-xO3 (x=0,.1)}		1
HCMO.pdf (Mazumdar and Das, 2021)	Role of 3d-4f exchange interaction and local anti-site defects in the magnetic and magnetocaloric properties of double perovskite Ho2CoMnO6 compound	{2}, 3(a), <del>3(d), 7(b)</del>	Ho2 CoMnO 6	Failed JSON format form the first chunk the failed JSON found 3(a) plot	1-non 2
JAP115123908.pdf	Magnetic properties of mixed spinel BaTiO3-NiFe2O4 composites	4{(a) (b)}, 5{(a) (b) (c) (d)}, {6 (a) (b) (c) (d)}	BaTiO3-NiFe2O4 composite {(100-x)BaTiO3-(x)NiFe2O4 (x=0,20,40,50,60,80,100)}	Failed JSON format form the first chunk the failed JSON found 4 plot	1-non 1

Document name	Document Title	Figures & Sections	Material	Notes	JSON Count
Krishna_Murthy_2015_J._Phys._D_Appl._Phys._48_3_55001.pdf	Giant magnetocaloric effect in Gd <sub>2</sub> NiMnO <sub>6</sub> and Gd <sub>2</sub> CoMnO <sub>6</sub> ferromagnetic insulators	1 (a) (b)	(a)Gd <sub>2</sub> NiMnO <sub>6</sub> (b)Gd <sub>2</sub> CoMnO <sub>6</sub>		1
Metal_Substitution_Effect_on_a_Three-Dimensional_C.pdf (Imoto et al., 2017)	{Metal Substitution Effect on a Three-Dimensional Cyanido-Bridged Fe Spin-Crossover Network}	<del>5</del> , 6{(a)}	(FexCo <sub>1-x</sub> ) <sub>2</sub> [Nb(CN) <sub>8</sub> ](4-pyridinealdoxime) <sub>8</sub> ·zH <sub>2</sub> O (x=0, 0.1,0.25, 0.5, 0.75, 1)	JSON 2 got the comps correct	2
Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO <sub>3</sub> and LaFeO <sub>5</sub> CrO <sub>3</sub> .pdf (Xia et al., 2022)	Magnetic Properties and Mössbauer Study of Perovskite LaFeO <sub>3</sub> and LaFe <sub>0.5</sub> Cr <sub>0.5</sub> O <sub>3</sub>	<del>1</del> , {4 (a) 5 (a)}	4a LaFeO <sub>3</sub> and LaFe <sub>0.5</sub> Cr <sub>0.5</sub> O <sub>3</sub> 5a LaFe <sub>0.5</sub> Cr <sub>0.5</sub> O <sub>3</sub> and Predicted	Complete failure except for title	2
PhysRevB.81.214406.pdf	Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape memory alloys	1,{2}	Ni-Mn-Sn ferromagnetic shape memory alloy {Ni <sub>50</sub> Mn <sub>36</sub> Co <sub>1</sub> Sn <sub>13</sub> }	second JSON correct figure 1 wrong material	2
PhysRevB.93.184403.pdf	Realization of two-dimensional ferromagnetism with giant coercivity in ultrathin β-Ni(OH) <sub>2</sub> layers grown on a MoS <sub>2</sub> surface	4{(a)}, 8	FC_LC ZFC_LC FC_MC ZFC_MC		2
Tuneable_Giant_Magnetocaloric_Effect_in_MnFe <sub>2</sub> PSi_M.pdf (Thang et al., 2016)	Tuneable Giant Magnetocaloric Effect in (Mn,Fe) <sub>2</sub> (P,Si) Materials by Co-B and Ni-B Co-Doping	<del>1,4</del> , {3, 8, 12}	3 Mn <sub>1.00</sub> Fe <sub>0.85</sub> Co <sub>0.10</sub> P <sub>0.55-z</sub> Si <sub>0.45</sub> Bz (z=0,0.02,0.04,0.06) 8 Mn <sub>1.00</sub> Fe <sub>0.95-z</sub> Co <sub>z</sub> P <sub>0.51</sub> Si <sub>0.45</sub> B <sub>0.04</sub> (z=0.07,0.09,0.11,0.13) 12 Mn <sub>1.00</sub> Fe <sub>0.95-z</sub> Ni <sub>z</sub> P <sub>0.51</sub> Si <sub>0.45</sub> B <sub>0.04</sub> (z=0.06,0.08,0.10,0.12)	Complete failure	2
zno-serie5.pdf (Silva Tupan et al., 2019)	Structural and magnetic characterization of (Zn <sub>1-x-y</sub> Fe <sub>x</sub> Co <sub>y</sub> )O	2	ZnO, Fe-doped ZnO, Co-doped ZnO {(Zn <sub>0.95</sub> Fe <sub>0.05</sub> )O and (Zn <sub>0.95</sub> Fe <sub>0.02</sub> Co <sub>0.03</sub> )O}		1

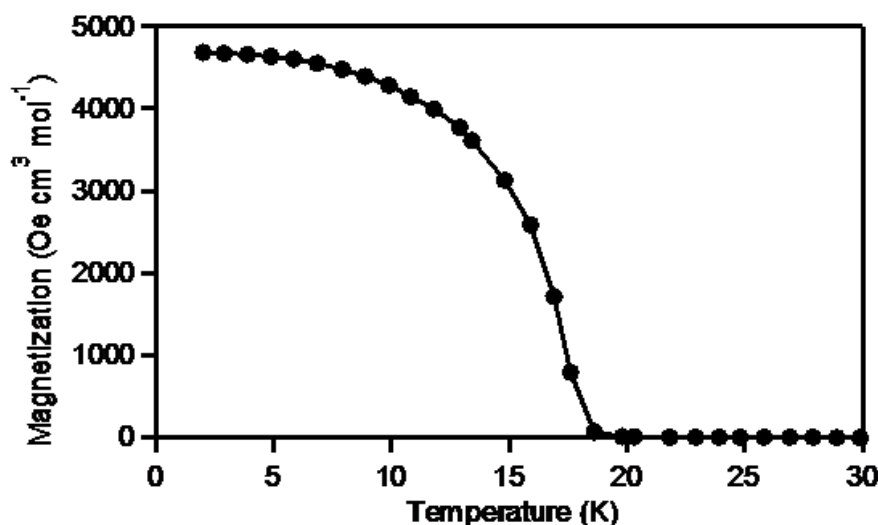
## 5.2 Digitization

Digitization proved to be the most challenging part of this project. Ultimately, a technique that utilizes multiple versions of CUDA, loaded and unloaded, was used to address this challenge. This is further expanded upon in the following chapters.

### 5.2.1 Other tested digitization methods

In total 5 different digitization methods ended up being examined for their potential suitability for the automated digitization of the plots. Out of ChatGPT-4 (openAI, 2024), WebPlot digitizer (Rohatgi, 2024), DePlot (Liu et al., 2023), curve (Zhou and Lu, 2022) and Plot2Spectra (Jiang et al., 2022). DePlot and Plot2Spectra were the most promising and therefore examined more closely. In total 5 different digitization methods ended up being examined for their potential suitability for the automated digitization of the plots. Out of ChatGPT-4 (openAI, 2024), WebPlot digitizer (Rohatgi, 2024), DePlot (Liu et al., 2023), curve (Zhou and Lu, 2022) and Plot2Spectra (Jiang et al., 2022). DePlot and Plot2Spectra were the most promising and therefore examined more closely.

DePlot provides decent digitization of plots with an example provided below (see **Figure 12** and **Figure 13**). As can be seen from this, DePlot is evidently designed for bar graphs as it will only measure points that are labeled on the x-axis. It is also unable to accurately digitize graphs that contain multiple lines on a single plot. Nevertheless, its recognition of the labels on the x and y-axis should not be ignored and could provide solutions to other problems discussed in this thesis (see chapter 6).



**Figure 12** Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C plot provided to DePlot (Imoto et al., 2017)

DePlot output:

```
"Temperature (K) | Magnitization (O<0xE2><0x82><0x83> m<0xE2><0x82><0x83>)  
0.001 | 4675  
5 | 4625  
10 | 4250  
15 | 3110  
20 | 1  
25 | 1  
30 | 1"
```

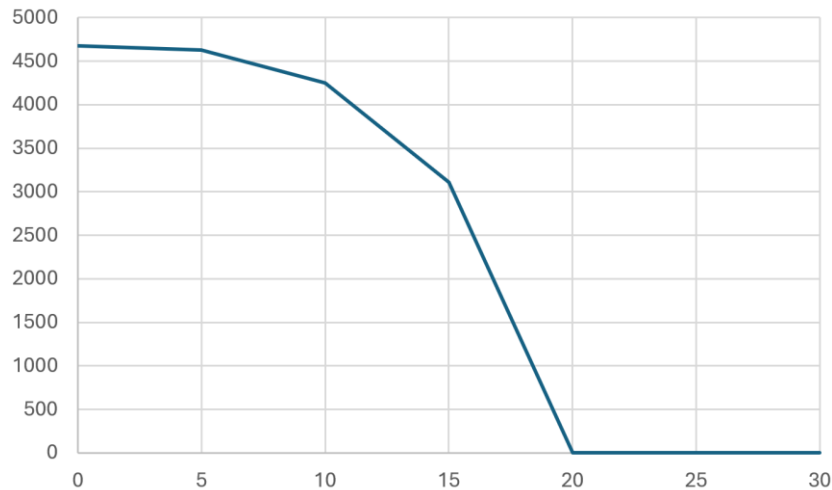


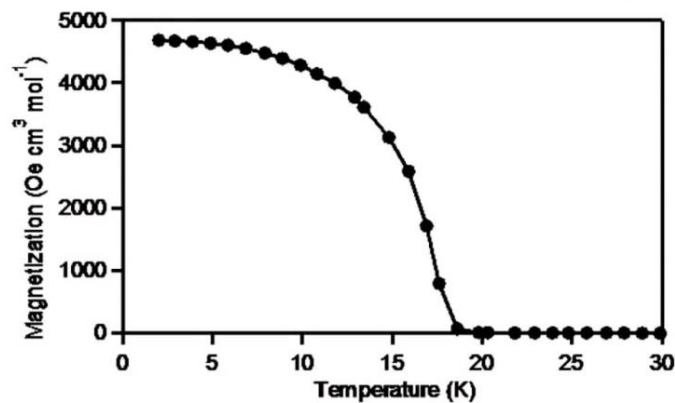
Figure 13 DePlot example; source: own depiction

## 5.2.2 Plot2Spectra and Context-Aware Chart Element Detection

After Deplot Plot2spectra was investigated and deemed the most promising. It is based on the optical flow (Jiang et al., 2022), utilizes a branched ERFNet neural network to separate the lines from the background and one another (Jiang et al., 2022) as well as a EfficientDet model (Jiang et al., 2022) a CNN to determine the plots bounding box.

This model thus provides excellent line detection and decent performance when given multiple lines as shown in the following example in **Figure 14**:

# Input Plot Image



# Model Output

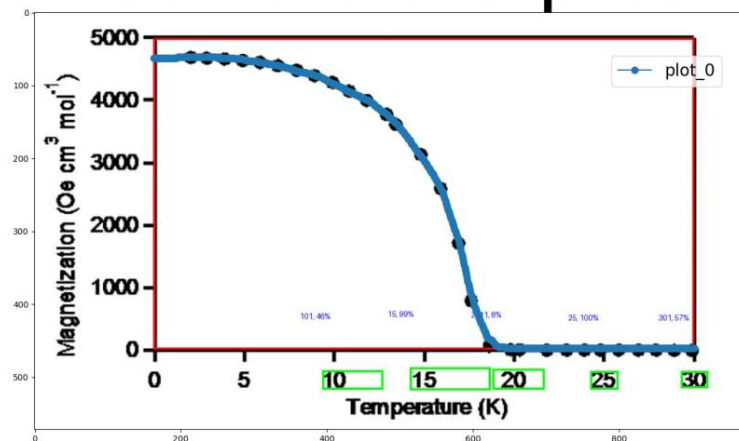
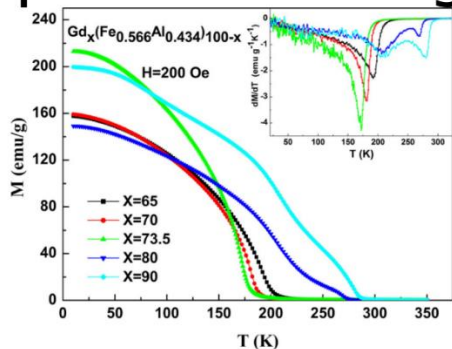


Figure 14 Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C, Plot2Spectra output source: own depiction based on (Imoto et al., 2017)

As illustrated in the above image Plot2Spectra is very precise and close to the line. However, this is not always consistently the case, particularly when it comes to multiple lines and additional smaller plot-elements such as a sub-plot interfere being included in the plot (see **Figure 15**). **Figure 15** illustrates this phenomenon, demonstrating how a line is created for the sub-plot itself as well as for plot zero. Another example worth noting is phantom lines. This is demonstrated by the initial correct identification of the top part of plot zero (see **Figure 16**). However, due to how Plot2Spectra is programmed it then expects this line to continue, therefore connecting it to the arrow for FC. This causes the line to continue to nothing in a straight line until the end of the bounding box.

# Input Plot Image



# Model Output

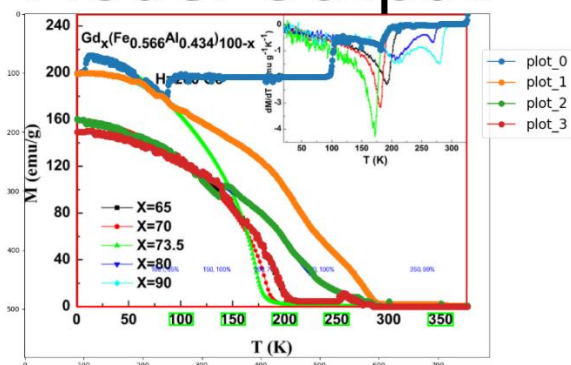
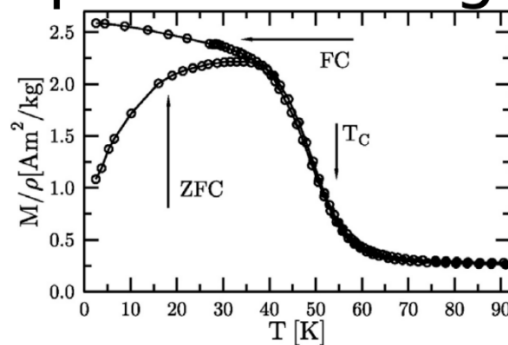


Figure 15 035220\_1\_online, Plot2Spectra output (Zhang et al., 2016)

# Input Plot Image



# Model Output

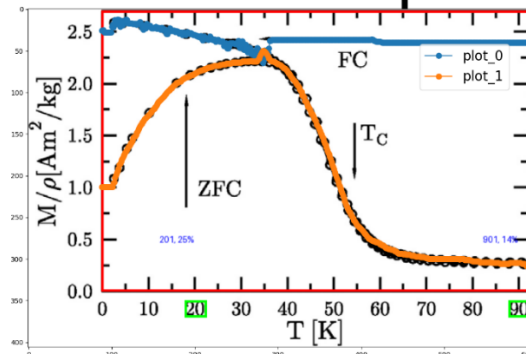
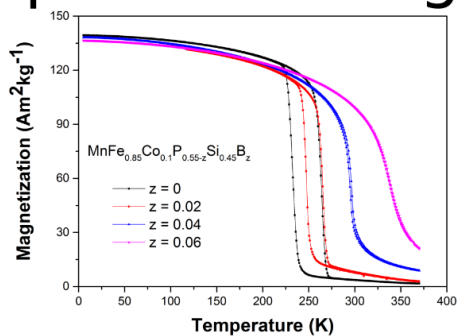


Figure 16 JPCM\_1720052985, Plot2Spectra output (Cieřlak et al., 2005)

# Input Plot Image



# Model Output

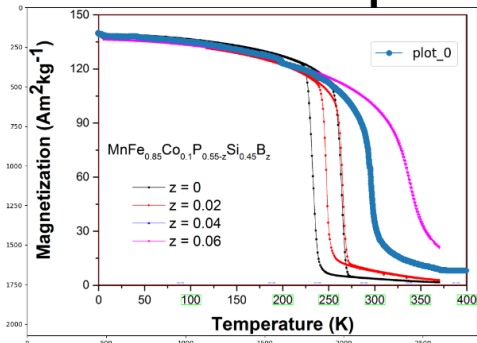
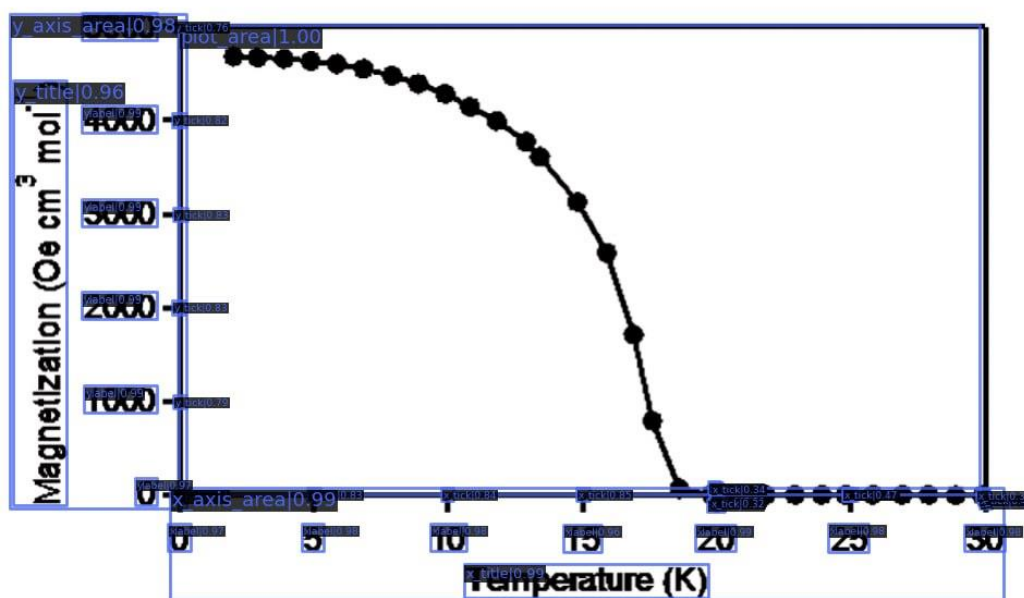


Figure 17 Tuneable\_image1, Plot2Spectra output (Thang et al., 2016)

However, when the data from Plot2Spectra is extracted, the axes are still relative to the bounding box which means, that while the line looks correct, to be accurate when extracted it is required to be corrected through a different method. To demonstrate this the axes that contain the bounding box have been preserved on the above images (see **Figure 14**, **Figure 15**, **Figure 16** and **Figure 17**) that are relevant to the bounding block and the line, this axis is the pixel based axis that surrounds the image. While Plot2Spectra attempts to fix this issue, it has quite poor performance recognizing the axis label values as seen above (see **Figure 14**, **Figure 15**, **Figure 16** and **Figure 17**) the blue text above the green boxes, a new method had to be devised. This is where Context-Aware Chart Detection (ChartDETE) (Yan et al., 2023) comes in. This tool utilizes an RCNN, that allows it to act as a tool designed to recognize and label different chart elements. An example is shown below (see **Figure 18**).



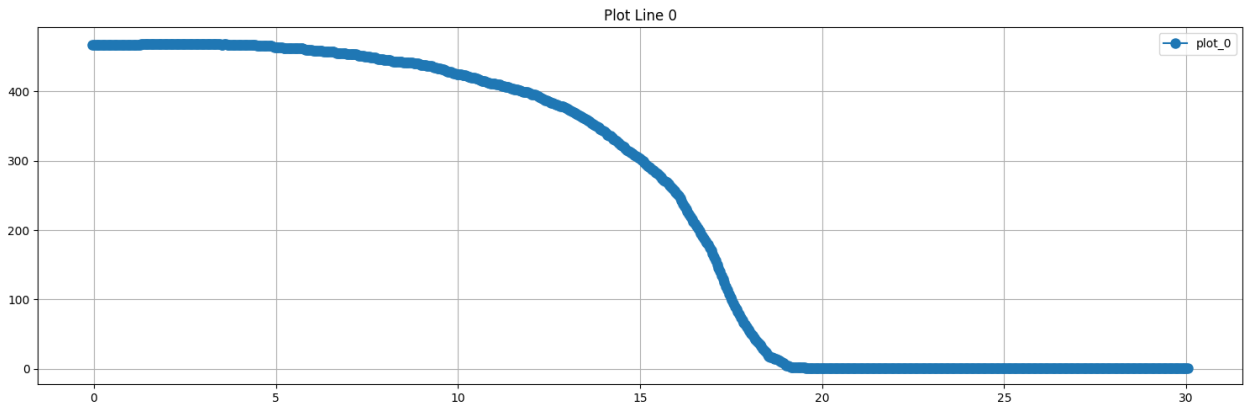
**Figure 18** ChartDETE example; based on (Imoto et al., 2017)

Utilizing this tool, it was possible to obtain the bounding boxes surrounding the important data, such as the x and y-labels and ticks. The image within the x and y-label bounding boxes was then fed to “easy OCR” (JaidedAI, 2024), which provided the value of the label as well as its relative position to the plot. **Equation 8** averages the maximum and minimum of both x and y values of the bounding box, allowing these points to be tied to their respective labels. Then these values were incorporated into a polynomial to form a relevant linear trend line that would be used to convert the values produced by Plot2Spectra into accurate values for the plot shown below (see **Figure 21**). Further testing was then performed to test the robustness of this solution. Some examples are shown here with additional examples being in the appendix (appendix **11.2**).

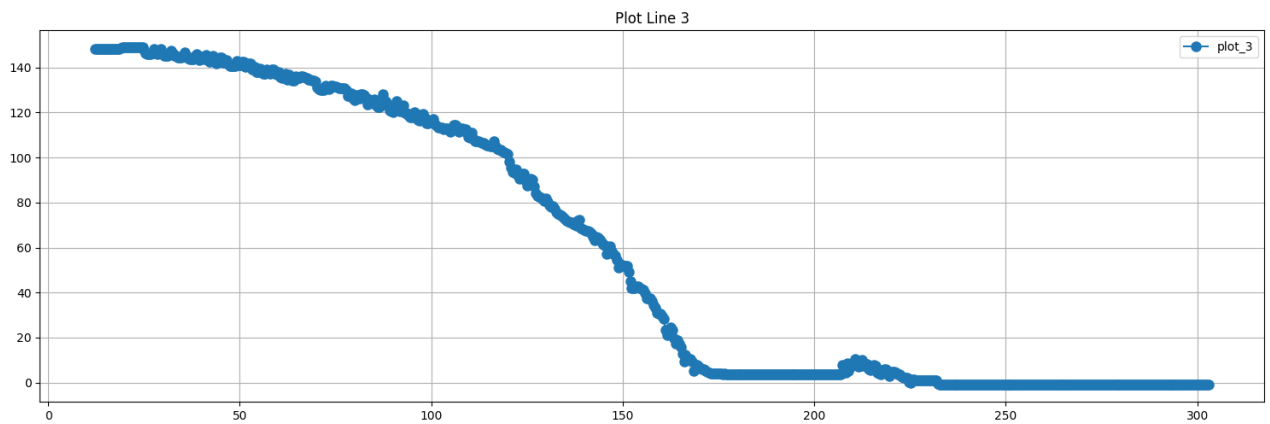
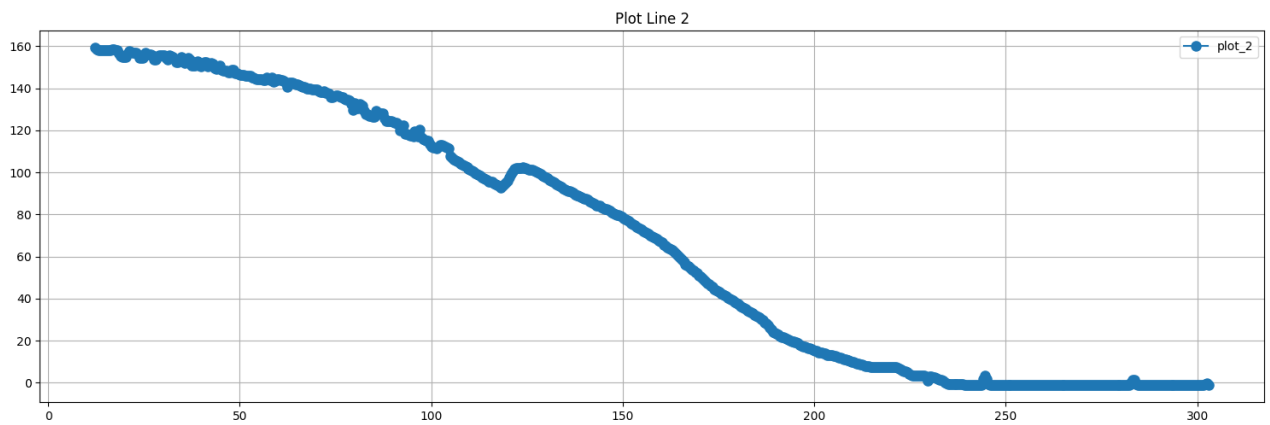
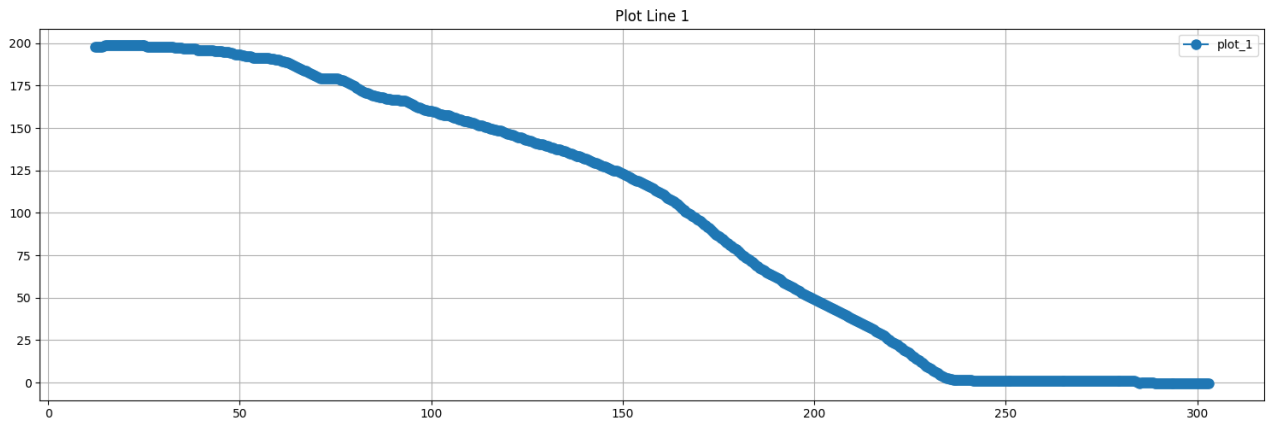
**Equation 8** label positioning relative to axis where  $x_p$  is the x position relative to the label and bounding box

$$x_p = \frac{bx_{max} + bx_{min}}{2}, y_p = \frac{by_{max} + by_{min}}{2}$$

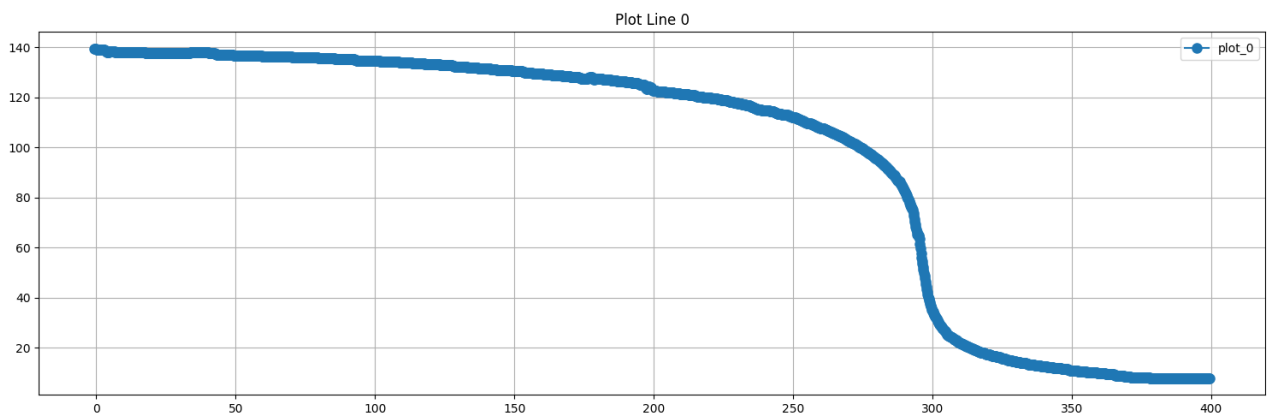
While for **Figure 21** the plot is correct, **Figure 19**, **Figure 20** and **Figure 22** show flaws with this technique. While the x-axis of **Figure 19** is correct, the y-axis is shown to be one tenth of the correct values. This is caused by an issue with the OCR where it fails to recognize the final zero of the values, i.E. 5000 is seen as 500. For **Figure 20**, the value 350 was recognized as 300 along the x-axis, while with **Figure 22** the OCR only recognized the zero of the number 0.5 resulting in a wider y-axis that dips into negative values.



**Figure 19** Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C Figure 14 corrected; source: own depiction



**Figure 20** 035220\_1\_online Figure 15 corrected; source: own depiction



**Figure 21** Tuneable Figure 17 corrected; source: own depiction

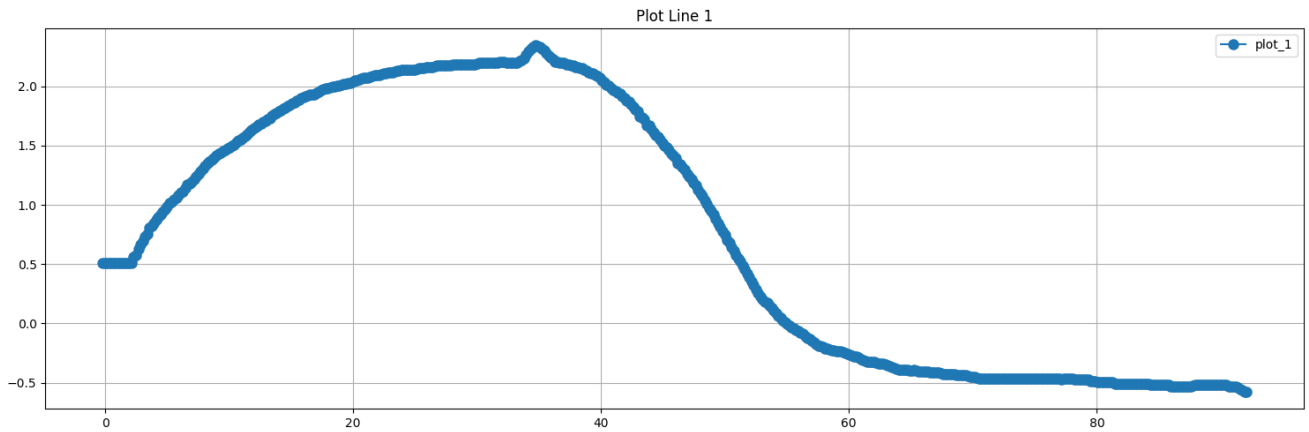


Figure 22 JPCM\_1720052985 Figure 16 corrected; source: own depiction

In addition, this methodology can struggle with non-typical axes for example if the x-axis is a logarithmic as shown in the appendix (appendix 11.1) or completely fail for a yet unknown reason (appendix 11.2). Furthermore, this methodology is unable to extract images with multiple subplots. The overall performance of Plot2Spectra is illustrated below in Table 10 for 15 different plots. It shows out that only 6 out of 40 possibly recognized lines are completely missed which is an 85% probability for the digitizer to find the line. Of the 34 cases examined, 19 are found to have at least one error. However, it is notable, that the majority of these errors could quite easily be solved by different methods discussed below in chapter 6.2. The axis recognition and correction also has a poor performance, with 11 out of 15 x-axes and 9 out of 15 y-axes being correctly identified. However, this too can be improved upon as mentioned in chapter 6.2.

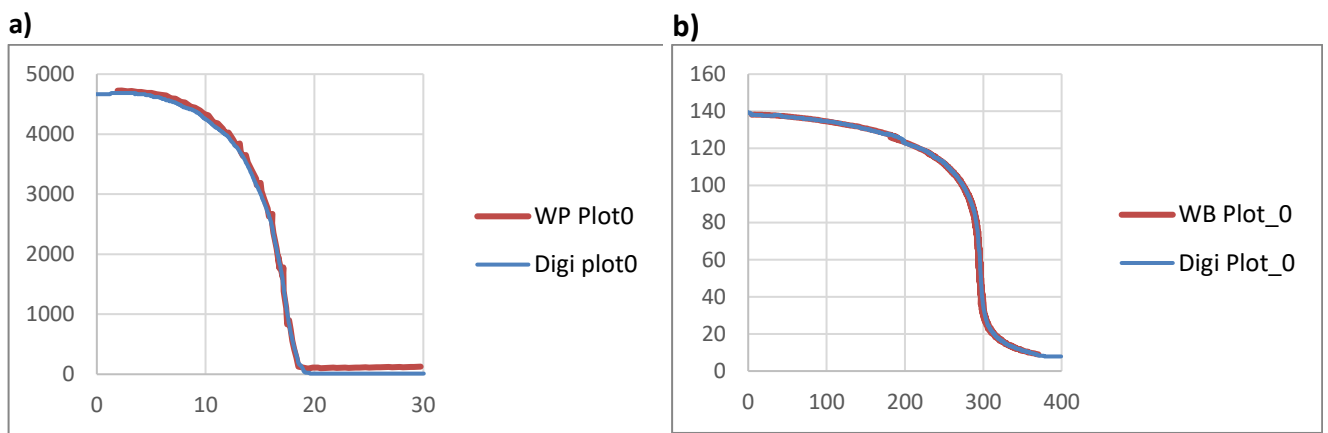
Table 10 Plot digitization performance results; source: own depiction

Plot	Lines	Lines with errors	correct lines	Correct x Axis	Correct y Axis	False positives	Error source lines	Error source Axis
122405_1_online_image1	4	3	1	✓	✓	0	inaccurately predicted continuation	
BMS_corrected_image1	2	1	1	✓	✓	0	inaccurately predicted continuation	
Krishna_Murthy_2015_J._Phys._D_Appl._Phys._48_355001 image 1	2	4	0	✗	✓	1	subplot	Atypical axis placement

Plot	Lines	Lines with errors	correct lines	Correct x Axis	Correct y Axis	False positives	Error source lines	Error source Axis
Krishna_Murthy_2015_J._Phys._D_Appl._Phys._48_355001_images2	2	1	1	X	X	0	subplot	Atypical axis placement
Physica Rapid Research Ltrs -_output_image1	2	0	1	✓	✓		close lines	
PhysRevB.93.184403_images_output	4	2	0	X	✓	2	Subplot	Logarithmic axis
Metal_Substitution_Effect_on_a_Three-Dimensional_C Redrawn	1	0	1	✓	X	1		unknown
JPCM_1720052985_image1	2	1	1	✓	X	0	inaccurately predicted continuation	Poor OCR performance
035220_1_online_image1	4	3	1	X	✓	0	close lines & subplot	unknown
Metal_Substitution_Effect_on_a_Three-Dimensional_C	1	0	1	✓	X	0		Small bounding box around y values
tunable_image1	4	0	1	✓	✓		Unknown	
tunable_image2	4	3	1	✓	✓	0	close lines	
tunable_image3	4	0	3	✓	✓	0	Unknown	
zno-serie5-1_image1	2	0	1	✓	X	0		Scientific notation for axis
zno-serie5-1_image2	2	1	1	✓	X		subplot	Scientific notation for axis
Totals	40	19	15	11	9	4		

It was noted that this is by far the best version of this digitization, with the final output being a correct CSV file for each image. Worth noting however is, that Plot2Spectra and ChartDETE require two different versions of CUDA to run, requiring unloading and reloading. In an effort to fix this, the code operates by first running ChartDETE with its unique requirements. This produces a JSON file that contains the labels, the ticks and their responding bounding boxes. The code then relies on Plot2Spectra to produce corrected graphs. In an effort to demonstrate the effectiveness of this digitization method, WebPlot-digitizer was used to digitize the graphs in **Figure 14** and **Figure 17**. These WebPlot-digitizer outputs were then plotted against the digitizer's outputs.

**Figure 23** WebPlot vs. digitization method; source: own depiction



Utilizing WebPlot digitizer the closest approximation of the graphs was obtained. Subsequently, the data was then plotted against the values' output from the chosen digitization method. It is worth noting at this point, that **Figure 23** shows the accuracy of this digitization method compared to even more robust methods such as WebPlot digitizer. Furthermore, it becomes clear, that the y-values are multiplied by 10 for **Figure 23a**. This highlights that digitization methods can still be further expanded upon. However, Plot2Spectra and ChartDETE provide a robust back bone to work from.

## 6 Discussion

As it currently stands the workflow shows great promise when it comes to the task of identifying and digitizing thermomagnetic plots. The LLM currently shows excellent performance, assuming the PDFs do not exceed the token limit. However, when the PDFs do exceed the token limit issues arise that require further investigation. It is recommended that further performance improvements should be deployed for the digitizer in order to improve the consistency of the performance. The final goal should be to produce a viable, user-friendly model to allow for the use of this code in many applications beyond thermomagnetic plots. In the following chapters current progress and future improvements and research are highlighted and discussed.

### 6.1 Current progress

The current workflow, while functional, presents several challenges that limit its accessibility and ease of use. To effectively utilize this workflow, users must possess a basic understanding of Python programming and its associated environments, which includes managing libraries and dependencies

---

within Python. Additionally, access to and a solid grasp of high-performance computing (HPC) is essential, as the workflow requires interaction with specialized hardware resources that are not commonly available to the average user.

A different barrier to broader adoption is the lack of a user-friendly interface. At present, there is no graphical user interface (GUI) for the workflow, meaning that users must interact directly with Python scripts and command-line operations. This requirement necessitates a higher level of technical expertise, particularly in Python programming and the management of HPC environments, which may limit the workflow's accessibility to non-experts or those unfamiliar with such tools.

Software and Hardware limitations also play a critical role in the usability of this workflow. Specifically, it requires access to an HPC system, or a powerful GPU equipped with Python 3.8, CUDA 11.1, and CUDA 11.6. These hardware requirements are not trivial, as HPC resources are typically expensive and may not be readily available to all researchers or institutions. However, updating the code to run on newer versions of CUDA and future advancements in technology could potentially reduce these constraints, making it possible to run the workflow on less specialized and more widely available computing resources. As computing power continues to evolve and become more accessible, it is conceivable that users may eventually be able to execute this workflow without the need for an HPC, broadening its applicability. For those looking to replicate the results of this work, access to the specific Python codes and model checkpoints used in the original research is essential. These codes and checkpoints are critical for ensuring that the workflow performs as intended and produces consistent, reproducible results. Without them, recreating the thesis findings would be challenging, if not impossible.

The workflow's output is particularly valuable for machine learning applications. The digitization process produces a CSV document containing the extracted data from plots, which, when combined with the output from the LLM, can serve as input for machine learning models. This structured data format is conducive to various machine learning tasks, enabling further analysis and potential discoveries within the dataset.

## **6.2 Future improvements and research**

To evaluate the true effectiveness of the tool presented in this thesis, one could allow access to a large pool of papers collected by an API to create a database. Thereafter machine learning could be conducted as well as outcomes of this subsequently tested, potentially extending this workflow to tasks unrelated to thermomagnetism. Despite this, several improvements could enhance the tool's

---

functionality and efficiency. The first and most immediate improvement would involve reworking Plot2Spectra and ChartDETE to function on a more recent version of CUDA, which would streamline the process. The updated version could then be used in conjunction with the LLM to generate a unified dataset from the extracted values.

Further enhancements could be made directly to the LLM. Currently, the LLM captures only the material composition, but improvements could enable it to also capture the material treatment and the conditions under which the thermomagnetic plot was obtained, such as the applied magnetic field. This would result in a more comprehensive dataset that includes critical experimental details.

In an effort to better understand how the LLM treats the data a statistical model bias study could also be performed over repeated runs, allowing for better understanding on why the model correctly predicts some figures but not others.

Additionally, the performance of the LLM when a PDF exceeds the token limit needs to be improved. Whether this is improved by finding models designed to handle larger token limits or finding a better methodology than simply cutting the PDFs up into chunks and feeding them sequentially.

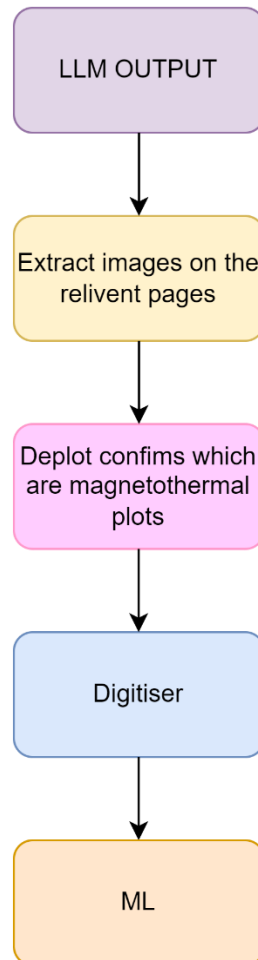
The digitization method could also benefit from refinement, particularly by addressing the issue of phantom lines—where lines in plots continue unnecessarily. This could be mitigated by introducing a mechanism that deletes any straight line exceeding a certain threshold height. Additionally, a methodology could be developed to separate lines of different colors individually, enhancing the clarity and accuracy of the digitized data.

Moving on to ChartDETE. This could be further utilized to remove undesirable chart elements, such as subplots, from the input images and to separate individual plots when a figure contains multiple subplots. This would streamline the extraction process and ensure that only relevant data is processed. Moreover, ChartDETE could be employed to extract legends from plots, with the legend information used to determine what each line in the plot represents. This would add a layer of interpretability to the digitized data, making it easier to understand and analyze.

Besides the suggested improvements so far, a solution needs to be devised to handle logarithmic plots. This could be as simple as identifying and excluding them via utilizing the closeness of individual ticks from ChartDETE or attempting to correct the Plot2Spectra output through a different method. This second, more complex method could work by first identifying with ChartDETE that it is a logarithmic plot and then somehow applying a logarithmic scale. At present, however, the largest gap in this workflow is a link between the LLM and the digitization. As it currently stands, large language models are only able to identify which figure needs to be extracted and cannot guarantee the extraction of this

---

figure. This is due to the PDFs lacking the meta-data, giving the images in the PDF figure names. One potential solution to this is to extract the figures on the matching pages that the text appears on and then feed these images to DePlot to confirm that the output figure is Thermomagnetic plot. A preliminary workflow for this can be found below (see **Figure 24**).



**Figure 24** Theoretical figure from PDF extraction workflow; source: own depiction

---

## 7 Conclusion

---

This thesis explored the intersection of artificial intelligence (AI) and material science to help predict new thermomagnetic materials through machine learning (ML) in the future. While the intended goal of utilizing ML to predict thermomagnetic materials could not be fully realized due to time constraints, the research laid useful groundwork for future endeavors in this area and machine learning as a whole. The thesis's main contributions can be summarized in two key areas: the implementation and evaluation of large language models (LLMs) for data extraction and the digitization of thermomagnetic plots.

One of the critical challenges leading up to this thesis was the collection and digitization of relevant data for machine learning. It was therefore noted that a significant bottleneck in the ML process is the availability and accessibility of high-quality, relevant data as mentioned by Yuji Roh, Geon Heo, Steven Euijong Whang (Roh et al., 2019). In the context of predicting new thermomagnetic materials, this issue is particularly pertinent given the specialized and often complex nature of plots and data in the field. To address this challenge, the thesis focused on the automatic extraction of relevant plots from scientific papers in the field of material science. These plots often contain critical data on material properties, including those that relate to thermomagnetic behaviors. The methodology developed for this task involved training an LLM to identify and extract these plots from a sample (n=15) of scientific literature. The LLM, specifically 'Llama 3 8B instruct', was determined to be the most effective model for this task, providing consistent and accurate identification of relevant plots across multiple tests (see chapter **5.1.3**).

Digitizing plots was another key focus of the thesis. A variety of digitization tools were evaluated, including AI Curve, Web Plot Digitizer, ChatGPT-4, DePlot, and Plot2Spectra. After extensive testing, Plot2Spectra emerged as the most successful tool for converting graphical data into a format suitable for machine learning analysis (see chapter **5.2.2**). This tool was further enhanced by integrating it with ChartDETE and EasyOCR, which increased the accuracy of digitization by incorporating optical character recognition (OCR) and convolutional recurrent neural networks (CRNN) to detect and interpret textual data within plots.

While the final machine learning step to predict new thermomagnetic materials was not executed, the research demonstrated that with some improvements it would be feasible to build a database of digitized plots via this method that could be used for future ML predictions. The groundwork laid in data extraction and digitization therefore provides a foundation for future work to build upon and

---

shows, that the successful digitization of these plots marks a crucial step towards enabling automated, data-driven discoveries in material science.

Despite the advancements made, the research encountered several limitations and challenges that prevented the full realization of the project's goals. One of the most prominent challenges was the inherent volatility and variability in the performance of different LLMs. While 'Llama 3 8B instruct' was the most effective model tested, it still exhibited inconsistencies in handling entire PDFs, leading to varying results across different tests (see chapter 5.1.3). This volatility underscores the need for further refinement in LLM-based methodologies, particularly in improving their stability and reliability.

Another notable limitation was the complexity involved in the digitization process. While Plot2Spectra, augmented with ChartDETE, proved effective in many cases, it struggled with non-standard plot formats, such as those with logarithmic axes or subplots within the plots (see chapter 5.2.2). Additionally, the digitization process occasionally produced errors, such as 'phantom lines' where lines in the plot were incorrectly extended beyond their actual endpoints. These issues highlight the need for further enhancements in digitization tools, particularly to handle a broader range of plot types and to reduce the occurrence of such errors.

Regarding future directions and recommendations there are several things to consider. While the research did not achieve the full objective of predicting new thermomagnetic materials using machine learning, it has paved the way for future efforts to do so, highlighting several key areas for future research. One of these lays in the refinement of large language models with future work aiming to address the volatility observed in LLMs by refining the models used for plot extraction. This could involve fine-tuning existing models, developing new models specifically tailored to the material science domain, or exploring hybrid approaches that combine the strengths of different models. Additionally, further investigations into dividing large PDFs into smaller sections for processing as discussed in this thesis could help mitigate some of the issues related to model performance.

Besides improving the utilized LLMs it is worthwhile to enhance the digitization tools in future research to make them more accurate and versatile. This could involve enhancing Plot2Spectra and ChartDETE to better handle non-standard plot types and reduce the occurrence of errors like phantom lines. Incorporating additional pre-processing steps to identify and correct potential issues before digitization could also improve overall accuracy.

Another improvement worth considering is the development of user-friendly interfaces to make the process more approachable. Currently, potential users must interact directly with the code, requiring a solid understanding of Python and access to high-performance computing resources. This limitation

---

restricts the usability of the tools to a relatively small group of highly skilled researchers, reducing the potential impact of the work. A well-designed graphical user interface (GUI) however, would allow users to interact with the tools without needing extensive programming knowledge, significantly increasing their usability and impact. This is assumed to also facilitate collaboration across different fields, as researchers in material science could more easily integrate these tools into their workflows. Finally, machine learning, as well as the integration of additional data sources should be investigated in further research. The ultimate goal of predicting new thermomagnetic materials using machine learning remains a critical area for future work. With the groundwork laid by this thesis in data extraction and digitization, future research should focus on implementing machine learning models to analyze the digitized data and identify potential new materials. This will likely involve the development of specialized ML algorithms that can handle the unique challenges posed by thermomagnetic materials, including their complex dependencies on temperature and other environmental factors. Additionally, the dataset for these ML predictions should be expanded to allow for scientifically accurate predictions. This could involve integrating data from other sources, such as experimental results, theoretical calculations, or material databases, to provide a more comprehensive foundation for ML models. Furthermore, the development of methods to automatically incorporate material treatment and experimental conditions into the data extraction process could further improve the accuracy and relevance of the predictions.

In conclusion, while the original objective of predicting new thermomagnetic materials through machine learning was not fully realized, this thesis has made notable contributions to the field of material science and AI. The research has demonstrated the feasibility of using LLMs for data extraction, developed a robust methodology for digitizing scientific plots, and identified key challenges that must be addressed. The findings of this research provide a solid foundation for future work, which will undoubtedly continue to push the boundaries of what is possible at the intersection of AI and material science. The potential to automate the discovery of new materials using AI and ML is immense, and the work presented here represents an important step towards realizing that potential. With continued research and development, the tools and workflows developed in this thesis could one day lead to significant breakthroughs in the design and discovery of new thermomagnetic materials, with wide-ranging applications across various industries.

---

## 8 References

---

- Almryad, A.S., Kutucu, H., 2020. Automatic identification for field butterflies by convolutional neural networks. *Engineering Science and Technology, an International Journal* 23, 189–195.  
<https://doi.org/10.1016/j.jestch.2020.01.006>
- Anwar, M.S., Ahmed, F., Koo, B.H., 2015. Magnetocaloric response of La<sub>0.70</sub>Ca<sub>0.15</sub>Sr<sub>0.2</sub>Fe<sub>0.1</sub>Mn<sub>0.9</sub>O<sub>3</sub> perovskite for magnetic refrigeration. *Bull Mater Sci* 38, 101–104. <https://doi.org/10.1007/s12034-014-0816-5>
- Baek, Y., Lee, B., Han, D., Yun, S., Lee, H., 2019. Character Region Awareness for Text Detection.
- Ballas, A., Diou, C., 2023. CNN Feature Map Augmentation for Single-Source Domain Generalization.  
<https://doi.org/10.48550/ARXIV.2305.16746>
- Beauchemin, S.S., Barron, J.L., 1995. The computation of optical flow. *ACM Comput. Surv.* 27, 433–466.  
<https://doi.org/10.1145/212094.212141>
- Bhattacharya, S., Dinda, D., Shaw, B.K., Dutta, S., Saha, S.K., 2016. Realization of two-dimensional ferromagnetism with giant coercivity in ultrathin  $\beta$ -Ni(OH)<sub>2</sub> layers grown on a MoS<sub>2</sub> surface. *Phys. Rev. B* 93, 184403. <https://doi.org/10.1103/PhysRevB.93.184403>
- Brück, E., Tegus, O., Thanh, D.T.C., Buschow, K.H.J., 2007. Magnetocaloric refrigeration near room temperature (invited). *Journal of Magnetism and Magnetic Materials* 310, 2793–2799.  
<https://doi.org/10.1016/j.jmmm.2006.10.1146>
- C A Padmanabha Reddy, Y., Viswanath, P., Eswara Reddy, B., 2018. Semi-supervised learning: a brief review. *IJET* 7, 81. <https://doi.org/10.14419/ijet.v7i1.8.9977>
- Chen, B., Zhang, Z., Langrené, N., Zhu, S., 2023. Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review. <https://doi.org/10.48550/ARXIV.2310.14735>
- Chen, C.-C., Chen, I.-L., Duan, H., Tseng, C.-Y., Chung, T.-K., 2019. A thermomagnetic-piezoelectric MEMS AC magnetic-field sensor demonstrating a novel room-temperature-range coil-less resetting/demagnetizing approach. *Sensors and Actuators A: Physical* 297, 111509. <https://doi.org/10.1016/j.sna.2019.07.033>
- Cieślak, J., Costa, B.F.O., Dubiel, S.M., Reissner, M., Steiner, W., 2005. Magnetic properties of a nanocrystalline  $\sigma$ -FeCr alloy. *J. Phys.: Condens. Matter* 17, 2985–2992. <https://doi.org/10.1088/0953-8984/17/19/012>

- 
- Clark, D.A., Schmidt, P.W., 1982. Theoretical analysis of thermomagnetic properties, low-temperature hysteresis and domain structure of titanomagnetites. *Physics of the Earth and Planetary Interiors* 30, 300–316. [https://doi.org/10.1016/0031-9201\(82\)90029-2](https://doi.org/10.1016/0031-9201(82)90029-2)
- Du, W., Luo, T., Qiu, Z., Huang, Z., Shen, Y., Cheng, R., Guo, Y., Fu, J., 2024. Stacking Your Transformers: A Closer Look at Model Growth for Efficient LLM Pre-Training.
- Dzekan, D., Waske, A., Nielsch, K., Fähler, S., 2021. Efficient and affordable thermomagnetic materials for harvesting low grade waste heat. *APL Materials* 9, 011105. <https://doi.org/10.1063/5.0033970>
- Gibson, J.J., 1950. *The perception of the visual world*. The Riverside Press, Cambridge, Massachusetts.
- Grossi, E., Buscema, M., 2007. Introduction to artificial neural networks: *European Journal of Gastroenterology & Hepatology* 19, 1046–1054. <https://doi.org/10.1097/MEG.0b013e3282f198a0>
- Hcini, S., Hsini, M., Al Robei, H., Bouazizi, M.L., Boudard, M., 2020. Structural Analysis and Theoretical Investigations of the Magnetocaloric Effect for La<sub>0.7</sub>Ba<sub>0.15</sub>Ag<sub>0.15</sub>MnO<sub>3</sub> Manganite Prepared Using Sol-Gel Route. *J Supercond Nov Magn* 33, 3597–3605. <https://doi.org/10.1007/s10948-020-05615-x>
- He, Y., Zhu, C., Wang, J., Savvides, M., Zhang, X., 2018. Bounding Box Regression with Uncertainty for Accurate Object Detection. <https://doi.org/10.48550/ARXIV.1809.08545>
- Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation* 9, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Imoto, K., Takano, S., Ohkoshi, S., 2017. Metal Substitution Effect on a Three-Dimensional Cyanido-Bridged Fe Spin-Crossover Network. *Inorganics* 5, 63. <https://doi.org/10.3390/inorganics5040063>
- JaiedAI, 2024. EasyOCR [WWW Document]. Github. URL <https://github.com/JaiedAI/EasyOCR>
- Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D. de las, Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.-A., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E., 2023. Mistral 7B.
- Jiang, W., Li, K., Spreadbury, T., Schwenker, E., Cossairt, O., Chan, M.K.Y., 2022. Plot2Spectra: an automatic spectra extraction tool. *Digital Discovery* 1, 719–731. <https://doi.org/10.1039/D1DD00036E>
- Johnson, F., Hughes, P., Gallagher, R., Laughlin, D.E., McHenry, M.E., Willard, M.A., Harris, V.G., 2001. Structure and thermomagnetic properties of new FeCo-based nanocrystalline ferromagnets. *IEEE Trans. Magn.* 37, 2261–2263. <https://doi.org/10.1109/20.951142>

- 
- Johnson, N.S., Vulimiri, P.S., To, A.C., Zhang, X., Brice, C.A., Kappes, B.B., Stebner, A.P., 2020. Invited review: Machine learning for materials developments in metals additive manufacturing. *Additive Manufacturing* 36, 101641. <https://doi.org/10.1016/j.addma.2020.101641>
- Khovaylo, V.V., Skokov, K.P., Gutfleisch, O., Miki, H., Takagi, T., Kanomata, T., Koledov, V.V., Shavrov, V.G., Wang, G., Palacios, E., Bartolomé, J., Burriel, R., 2010. Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape memory alloys. *Phys. Rev. B* 81, 214406. <https://doi.org/10.1103/PhysRevB.81.214406>
- Khramtsova, E., Zhuang, S., Baktashmotlagh, M., Zuccon, G., 2024. Leveraging LLMs for Unsupervised Dense Retriever Ranking, in: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Presented at the SIGIR 2024: The 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Washington DC USA, pp. 1307–1317. <https://doi.org/10.1145/3626772.3657798>
- Kishore, R.A., Priya, S., 2018. A review on design and performance of thermomagnetic devices. *Renewable and Sustainable Energy Reviews* 81, 33–44. <https://doi.org/10.1016/j.rser.2017.07.035>
- Krishna Murthy, J., Devi Chandrasekhar, K., Mahana, S., Topwal, D., Venimadhav, A., 2015. Giant magnetocaloric effect in  $Gd_2NiMnO_6$  and  $Gd_2CoMnO_6$  ferromagnetic insulators. *J. Phys. D: Appl. Phys.* 48, 355001. <https://doi.org/10.1088/0022-3727/48/35/355001>
- Kupczyk, A., Świerczek, J., Hasiak, M., 2019. Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous  $Fe_{70-x}Co_xMn_{10}Mo_5B_{15}$  ( $0 \leq x \leq 0.5$ ) Alloys at Low Magnetizing Field. *Acta Phys. Pol. A* 135, 133–135. <https://doi.org/10.12693/APhysPolA.135.133>
- Li, G.J., Liu, E.K., Zhang, H.G., Zhang, Y.J., Chen, J.L., Wang, W.H., Zhang, H.W., Wu, G.H., Yu, S.Y., 2013. Phase diagram, ferromagnetic martensitic transformation and magneto-responsive properties of Fe-doped MnCoGe alloys. *Journal of Magnetism and Magnetic Materials* 332, 146–150. <https://doi.org/10.1016/j.jmmm.2012.12.001>
- Li, Y., Chen, Y., Rajabifard, A., Khoshelham, K., Aleksandrov, M., 2018. Estimating Building Age from Google Street View Images Using Deep Learning (Short Paper). *LIPICs, Volume 114, GIScience 2018* 114, 40:1-40:7. <https://doi.org/10.4230/LIPICS.GISCIENCE.2018.40>
- Li, Z., Yang, W., Peng, S., Liu, F., 2020. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects.

- 
- Lin, J., Tang, J., Tang, H., Yang, S., Chen, W.-M., Wang, W.-C., Xiao, G., Dang, X., Gan, C., Han, S., 2024. AWQ: Activation-aware Weight Quantization for LLM Compression and Acceleration.
- Liu, E.K., Zhang, H.G., Xu, G.Z., Zhang, X.M., Ma, R.S., Wang, W.H., Chen, J.L., Zhang, H.W., Wu, G.H., Feng, L., Zhang, X.X., 2013. Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system by establishing a Curie-temperature window. *Applied Physics Letters* 102, 122405. <https://doi.org/10.1063/1.4798318>
- Liu, F., Eisenschlos, J.M., Piccinno, F., Krichene, S., Pang, C., Lee, K., Joshi, M., Chen, W., Collier, N., Altun, Y., 2023. DePlot: One-shot visual language reasoning by plot-to-table translation.
- Liu, J., Su, H., Yi, Y., Hu, W., 2016. Robust text detection via multi-degree of sharpening and blurring. *Signal Processing* 124, 259–265. <https://doi.org/10.1016/j.sigpro.2015.06.025>
- Mazumdar, D., Das, I., 2021. Role of 3d–4f exchange interaction and local anti-site defects in the magnetic and magnetocaloric properties of double perovskite Ho<sub>2</sub>CoMnO<sub>6</sub> compound. *Journal of Applied Physics* 129, 063901. <https://doi.org/10.1063/5.0041257>
- Mehrish, A., Majumder, N., Bhardwaj, R., Mihalcea, R., Poria, S., 2023. A Review of Deep Learning Techniques for Speech Processing.
- Meta, 2024. LLAMA 3 [WWW Document]. Github. URL <https://github.com/meta-llama/llama3>
- Meta, 2023. LLAMA 2 [WWW Document]. URL <https://github.com/meta-llama/llama>
- Mielke, S.J., Alyafeai, Z., Salesky, E., Raffel, C., Dey, M., Gallé, M., Raja, A., Si, C., Lee, W.Y., Sagot, B., Tan, S., 2021. Between words and characters: A Brief History of Open-Vocabulary Modeling and Tokenization in NLP.
- Mischler, G., Li, Y.A., Bickel, S., Mehta, A.D., Mesgarani, N., 2024. Contextual Feature Extraction Hierarchies Converge in Large Language Models and the Brain.
- Mould, R.F., 2007. Pierre Curie, 1859–1906. *Current Oncology* 14, 74–82. <https://doi.org/10.3747/co.2007.110>
- Murat H., S., 2006. A brief review of feed-forward neural networks. *Communications Faculty Of Science University of Ankara* 50, 11–17. [https://doi.org/10.1501/commua1-2\\_0000000026](https://doi.org/10.1501/commua1-2_0000000026)
- Nagy, A., Hammad, T., Yehia, S., Aly, S.H., 2019. Thermomagnetic properties and magnetocaloric effect of TmFe<sub>2</sub> compound. *Journal of Magnetism and Magnetic Materials* 473, 324–330. <https://doi.org/10.1016/j.jmmm.2018.10.050>
- openAI, 2024. GPT-4 [WWW Document]. openai. URL <https://openai.com/index/gpt-4/>
- openAI, 2019. GPT-2 [WWW Document]. Github. URL <https://github.com/openai/gpt-2>

- 
- O'Shea, K., Nash, R., 2015. An Introduction to Convolutional Neural Networks.  
<https://doi.org/10.48550/ARXIV.1511.08458>
- Pal, S., Datta, A., Sen, S., Mukhopdhyay, A., Bandopadhyay, K., Ganguly, R., 2011. Characterization of a ferrofluid-based thermomagnetic pump for microfluidic applications. *Journal of Magnetism and Magnetic Materials* 323, 2701–2709. <https://doi.org/10.1016/j.jmmm.2011.06.016>
- Polak, M.P., Morgan, D., 2024. Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nat Commun* 15, 1569. <https://doi.org/10.1038/s41467-024-45914-8>
- Prévoit, M., Dunlop, D., n.d. Louis Néel: forty years of magnetism.
- Real, A.C., Luz, G.P., Sousa, J.M.C., Brito, M.C., Vieira, S.M., 2024. Optimization of a photovoltaic-battery system using deep reinforcement learning and load forecasting. *Energy and AI* 16, 100347.  
<https://doi.org/10.1016/j.egyai.2024.100347>
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection.
- Ren, Y., Sutherland, D.J., 2024. Learning Dynamics of LLM Finetuning.
- Roh, Y., Heo, G., Whang, S.E., 2019. A Survey on Data Collection for Machine Learning: a Big Data -- AI Integration Perspective.
- Rohatgi, A., 2024. WebPlotDigitizer [WWW Document]. automeris. URL <https://automeris.io/docs/#cite-webplotdigitizer>
- Romera, E., Alvarez, J.M., Bergasa, L.M., Arroyo, R., 2018. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. *IEEE Trans. Intell. Transport. Syst.* 19, 263–272.  
<https://doi.org/10.1109/TITS.2017.2750080>
- S, V., A, S., 2015. Performance Comparison of OCR Tools. *IJU* 6, 19–30. <https://doi.org/10.5121/iju.2015.6303>
- Sarkar, B., Dalal, B., Dev Ashok, V., Chakrabarti, K., Mitra, A., De, S.K., 2014. Magnetic properties of mixed spinel BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites. *Journal of Applied Physics* 115, 123908.  
<https://doi.org/10.1063/1.4869782>
- Sarker, I.H., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160. <https://doi.org/10.1007/s42979-021-00592-x>

- 
- Schmidhuber, J., 2015. Deep Learning in Neural Networks: An Overview. *Neural Networks* 61, 85–117.  
<https://doi.org/10.1016/j.neunet.2014.09.003>
- Schmidt, R.M., 2019. Recurrent Neural Networks (RNNs): A gentle Introduction and Overview.  
<https://doi.org/10.48550/ARXIV.1912.05911>
- Silva Tupan, L.F., Valerio-Cuadros, M.I., Oliveira, A.A., De Souza, A.O., Barco, R., Ivashita, F.F., Caetano, E.P., Saitovitch, E.B., Paesano, A., 2019. Structural and magnetic characterization of (Zn<sub>1-x</sub>YFe<sub>x</sub>Co)O. *Hyperfine Interact* 240, 30. <https://doi.org/10.1007/s10751-019-1571-1>
- Sivaprasad, S., Kaushik, P., Abdelnabi, S., Fritz, M., 2024. Exploring Value Biases: How LLMs Deviate Towards the Ideal. <https://doi.org/10.48550/ARXIV.2402.11005>
- Suzgun, M., Kalai, A.T., 2024. Meta-Prompting: Enhancing Language Models with Task-Agnostic Scaffolding.
- Thang, N., Dijk, N., Brück, E., 2016. Tuneable Giant Magnetocaloric Effect in (Mn,Fe)<sub>2</sub>(P,Si) Materials by Co-B and Ni-B Co-Doping. *Materials* 10, 14. <https://doi.org/10.3390/ma10010014>
- Tian, Z., Shen, C., Chen, H., He, T., 2019. FCOS: Fully Convolutional One-Stage Object Detection.
- TU Darmstadt, 2024. High Performance Computing [WWW Document]. Technische Universität Darmstadt. URL <https://www.hrz.tu-darmstadt.de/hlr/hochleistungsrechnen/index.en.jsp>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need.
- Vincent, O., Folorunso, O., 2009. A Descriptive Algorithm for Sobel Image Edge Detection. Presented at the InSITE 2009: Informing Science + IT Education Conference. <https://doi.org/10.28945/3351>
- Wang, S., Zhou, T., Bilmes, J.A., n.d. Bias Also Matters: Bias Attribution for Deep Neural Network Explanation.
- Waske, A., Dzekan, D., Sellschopp, K., Berger, D., Stork, A., Nielsch, K., Fähler, S., 2018. Energy harvesting near room temperature using a thermomagnetic generator with a pretzel-like magnetic flux topology. *Nat Energy* 4, 68–74. <https://doi.org/10.1038/s41560-018-0306-x>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M., 2020. HuggingFace’s Transformers: State-of-the-art Natural Language Processing.
- Xia, P., Mo, J., Chen, J., Liu, M., Xia, Y., 2022. Magnetic Properties and Mössbauer Study of Perovskite LaFeO<sub>3</sub> and LaFe<sub>0.5</sub>Cr<sub>0.5</sub>O<sub>3</sub>. *Physica Rapid Research Ltrs* 16, 2200023. <https://doi.org/10.1002/pssr.202200023>

- 
- Yan, P., Ahmed, S., Doermann, D., 2023. Context-Aware Chart Element Detection. pp. 218–233.  
[https://doi.org/10.1007/978-3-031-41676-7\\_13](https://doi.org/10.1007/978-3-031-41676-7_13)
- Zhang, K., Yuan, S., Zhao, H., 2024. TALEC: Teach Your LLM to Evaluate in Specific Domain with In-house Criteria by Criteria Division and Zero-shot Plus Few-shot. <https://doi.org/10.48550/ARXIV.2407.10999>
- Zhang, L., Bao, M., Zheng, Q., Tian, L., Du, J., 2016. Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity. *AIP Advances* 6, 035220. <https://doi.org/10.1063/1.4945407>
- Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., Wen, J.-R., 2023. A Survey of Large Language Models.
- Zhou, S., Lu, J., 2022. An Open Digitization Tool for Extracting Scientific Curve Data in Portable Documents, in: Chen, C.-H. (Ed.), *Advances in Transdisciplinary Engineering*. IOS Press.  
<https://doi.org/10.3233/ATDE220079>
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2020. A Comprehensive Survey on Transfer Learning.

---

## 9 Index of abbreviations

---

AC	Alternating current
AI	Artificial intelligence
ChartDETE	Context-Aware Chart Detection
CNN	Convolutional neural network
CNNs	Convolutional neural networks
CRAFT	Character-Region Awareness for Text Detection
CRNN	Convolutional recurrent neural network
CRNNs	Convolutional recurrent neural networks
ERFNet	Efficient Residual Factorized ConvNet
FCOS	Fully Convolutional One-Stage Object Detection
FFN	Feedforward neural network
FFNs	Feedforward neural networks
GB	Gigabyte
GUI	Graphical User Interface
HPC	High Powered Computing
i.e.	Id est
LLM	Large language model
LLMs	Large language models
LSTM	Long term short-term memory
M vs. T	Magnetization vs. Temperature
MCE	Magnetocaloric effect
MEMS	Microelectromechanical systems
ML	Machine learning
NLP	Natural language processing
NN	Neural network
OCR	Optical Character Recognition
PDF	Portable Document Format
RL	Reinforcement learning
RNN	Recurrent Neural Network
RNNs	Recurrent neural networks
YOLO	You Only Look Once

---

## 10 Tables and Figures

---

<b>Table 1</b> Overview tested large language models (LLMs); source: own depiction.....	30
<b>Table 2</b> Correct data of PDFs that contain thermomagnetic plots given to LLMs; source: own depiction .....	35
<b>Table 3</b> Individual document performance ChatGPT-4; source: own depiction.....	38
<b>Table 4</b> First ChatGPT-4 generated table, missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font; source: own depiction.....	39
<b>Table 5</b> Missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font ChatGPT-4; source: own depiction .....	40
<b>Table 6</b> ChatGPT-4 results, missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font; source: own depiction.....	40
<b>Table 7</b> ChatGPT-4 with false PDFs, missed plots, highlighted in brackets { }, and false positives, highlighted by italics and struck through font; source: own depiction.....	41
<b>Table 8</b> Llama 3 8B, testing output; source: own depiction.....	45
<b>Table 9</b> Llama 3 8B instruct sequential performance; source: own depiction.....	47
<b>Table 10</b> Plot digitization performance results; source: own depiction .....	56
<b>Figure 1</b> Examples for low waste heat collection utilizing thermomagnetic materials taken from (Dzekan et al., 2021).....	4
<b>Figure 2</b> Desired Thermomagnetic plot taken from (Waske et al., 2018).....	5
<b>Figure 3</b> Neural network (feed forward) architecture. x is inputs, y is the output of the network. w is weights and b is biases; source: own depiction.....	9
<b>Figure 4</b> Transformers architecture taken from "All you need is attention" (Vaswani et al., 2023). .....	13
<b>Figure 5</b> Image convolution example taken from (Almryad and Kutucu, 2020). .....	22
<b>Figure 6</b> YOLOs convolutional network taken from (Redmon et al., 2016) .....	24
<b>Figure 7</b> Plot2Spectra optical flow algorithm (Jiang et al., 2022). .....	27
<b>Figure 8</b> Workflow; source: own depiction .....	28
<b>Figure 9</b> Workflows for LLM development; source: own depiction.....	29
<b>Figure 10</b> Selection Images used for initial testing additional utilized in appendix 11.1; a &b)(Imoto et al., 2017) c)(Cieślak et al., 2005) d)(Liu et al., 2013) .....	31
<b>Figure 11</b> Final digitization workflow; source: own depiction .....	33
<b>Figure 12</b> Metal_Substitution_Effect_on_a_Three-Dimensional_C plot provided to DePlot (Imoto et al., 2017) .....	49
<b>Figure 13</b> DePlot example; source: own depiction .....	50
<b>Figure 14</b> Metal_Substitution_Effect_on_a_Three-Dimensional_C, Plot2Spectra output source: own depiction based on (Imoto et al., 2017) .....	51
<b>Figure 15</b> 035220_1_online, Plot2Spectra output (Zhang et al., 2016).....	52
<b>Figure 16</b> JPCM_1720052985, Plot2Spectra output (Cieślak et al., 2005).....	52
<b>Figure 17</b> Tuneable_image1, Plot2Spectra output (Thang et al., 2016) .....	52
<b>Figure 18</b> ChartDETE example; based on (Imoto et al., 2017) .....	53
<b>Figure 19</b> Metal_Substitution_Effect_on_a_Three-Dimensional_C Figure 14 corrected; source: own depiction .....	54
<b>Figure 20</b> 035220_1_online Figure 15 corrected; source: own depiction .....	55
<b>Figure 21</b> Tuneable Figure 17 corrected; source: own depiction .....	55
<b>Figure 22</b> JPCM_1720052985 Figure 16 corrected; source: own depiction .....	56
<b>Figure 23</b> WebPlot vs. digitization method; source: own depiction .....	58
<b>Figure 24</b> Theoretical figure from PDF extraction workflow; source: own depiction .....	61
<b>Figure 25</b> 122405_1_online_image1 taken from (Thang et al., 2016) .....	11

Figure 26 BMS_corrected_image1 taken from (Anwar et al., 2015).....	11
Figure 27 Krishna_Murthy_2015_J._Phys._D _Appl._Phys._48_355001_image1 taken from (Krishna Murthy et al., 2015) .....	12
Figure 28 Krishna_Murthy_2015_J._Phys._D _Appl._Phys._48_355001_images2 taken from (Krishna Murthy et al., 2015).....	12
Figure 29 Physica Rapid Research Ltrs _output_image1 taken form (Xia et al., 2022) .....	13
Figure 30 PhysRevB.93.184403_image1 taken from (Bhattacharya et al., 2016).....	13
Figure 31 JPCM_1720052985_image1 taken from (Cieślak et al., 2005) .....	14
Figure 32 035220_1_online_image1 taken from (Zhang et al., 2016).....	14
Figure 33 tunable_image1 taken from (Thang et al., 2016).....	15
Figure 34 tunable_image2 taken from (Thang et al., 2016).....	15
Figure 35 tunable_image3 taken from (Thang et al., 2016).....	16
Figure 36 zno-serie5-1_image1 taken from (Silva Tupan et al., 2019).....	16
Figure 37 zno-serie5-1_image2 taken from (Silva Tupan et al., 2019).....	17
Figure 38 Metal_Substitution_Effect_on_a_Three-Dimensional_C_image1 taken from (Imoto et al., 2017).....	17
Figure 39 Metal_Substitution_Effect_on_a_Three-Dimensional_Genarated taken from (Imoto et al., 2017)...	18
Figure 40 Corrected version of Figure 25 .....	19
Figure 41 corrected version of Figure 26.....	20
Figure 42 corrected version of Figure 27.....	22
Figure 43 corrected version of Figure 28.....	23
Figure 44 corrected version of Figure 29.....	23
Figure 45 corrected version of Figure 30.....	24
Figure 46 corrected figure of Figure 31.....	25
Figure 47 corrected figure of Figure 32.....	26
Figure 48 corrected figure of Figure 33.....	27
Figure 49 corrected figure of Figure 34.....	28
Figure 50 corrected version of Figure 35.....	29
Figure 51 corrected version of Figure 36.....	30
Figure 52 corrected version of Figure 37.....	30
Figure 53 corrected version of Figure 38.....	31
Figure 54 corrected version of Figure 39.....	31

---

## 11 Appendix

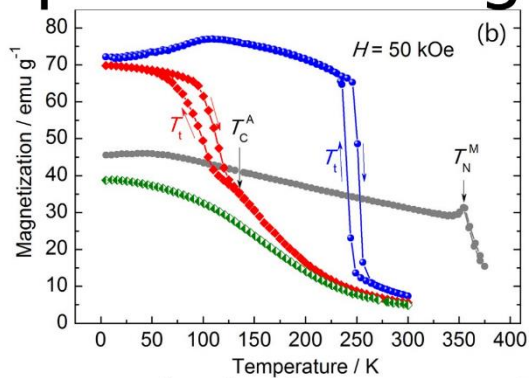
---

### Contents of Appendix

11.1	Images used for digitization & Plot2Spectra output	10
11.2	Corrected Plot2Spectra output	18
11.3	Basel level code i.e. code that is used at the surface	31
11.3.1	Llama 3 8b instruct	31
11.3.2	Digitizer	36
11.4	Llama 3 8b instruct output	48

### 11.1 Images used for digitization & Plot2Spectra output

# Input Plot Image



# Model Output

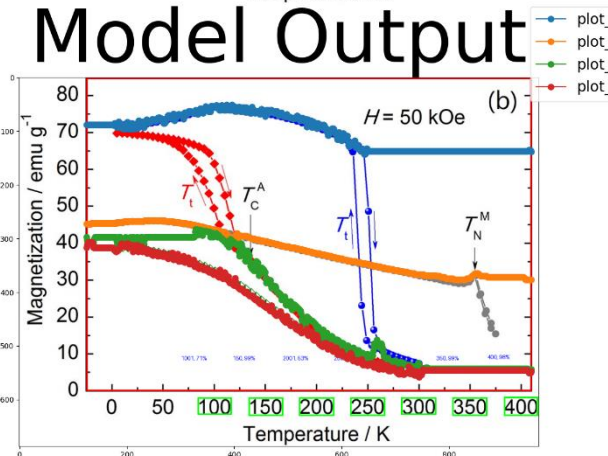
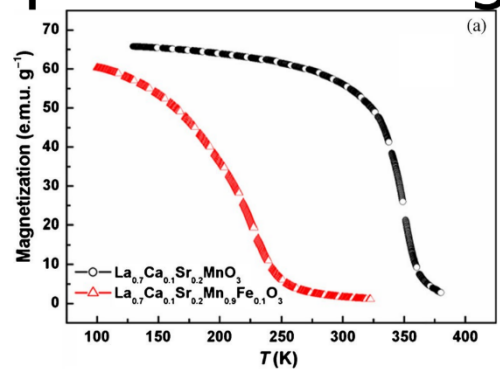


Figure 25 122405\_1\_online\_image1 taken from (Thang et al., 2016)

# Input Plot Image



# Model Output

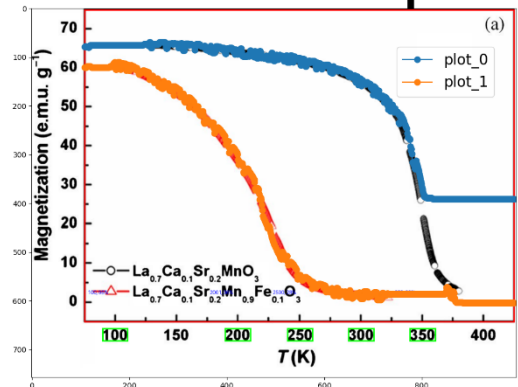
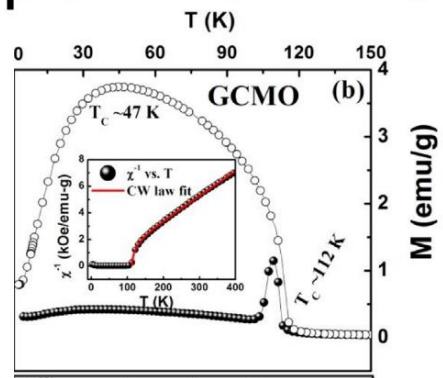
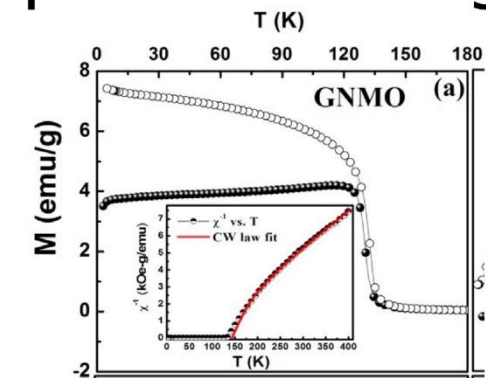
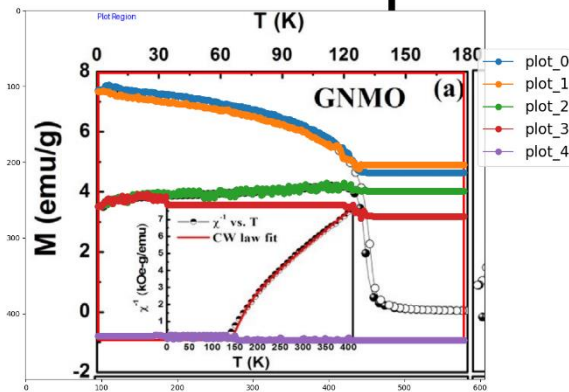


Figure 26 BMS\_corrected\_image1 taken from (Anwar et al., 2015)

# Input Plot Image      Input Plot Image



# Model Output



# Model Output

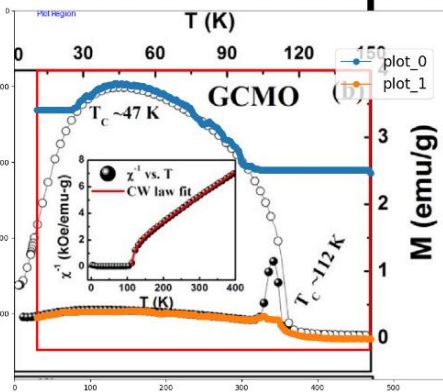
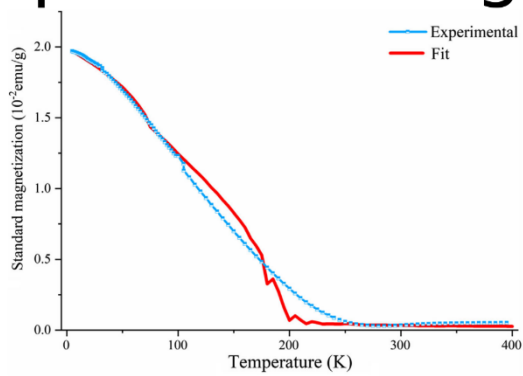


Figure 27 Krishna\_Murthy\_2015\_J.\_Phys.\_D  
\_Appl.\_Phys.\_48\_355001\_image1 taken from (Krishna Murthy  
et al., 2015)

Figure 28 Krishna\_Murthy\_2015\_J.\_Phys.\_D  
\_Appl.\_Phys.\_48\_355001\_images2 taken from (Krishna Murthy  
et al., 2015)

# Input Plot Image      Input Plot Image



## Model Output

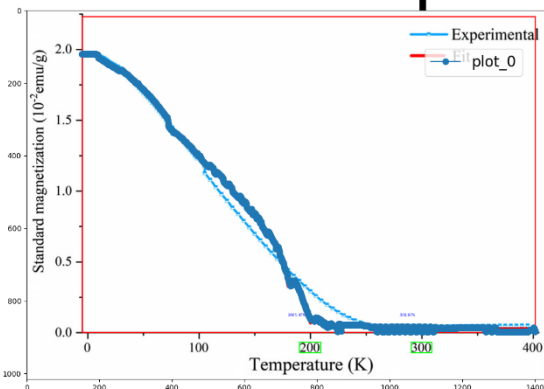
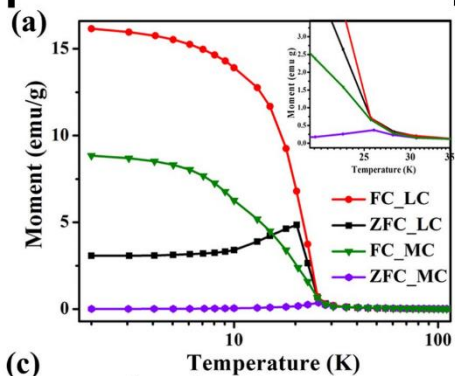


Figure 29 Physica Rapid Research Ltrs - \_output\_image1 taken from (Xia et al., 2022)

# Input Plot Image



## Model Output

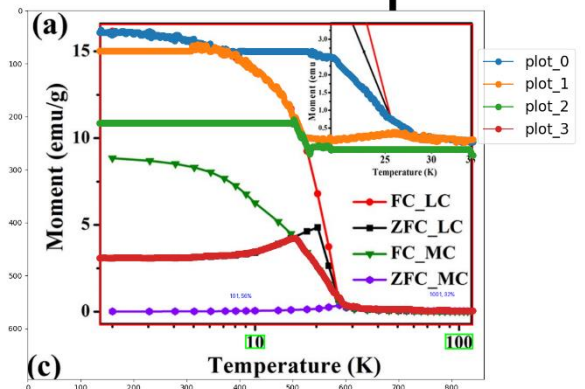
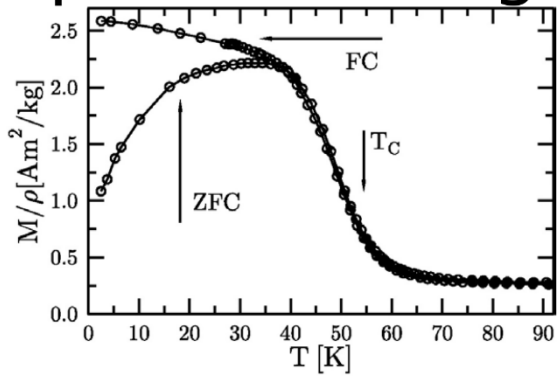


Figure 30 PhysRevB.93.184403\_image1 taken from (Bhattacharya et al., 2016)

# Input Plot Image



# Model Output

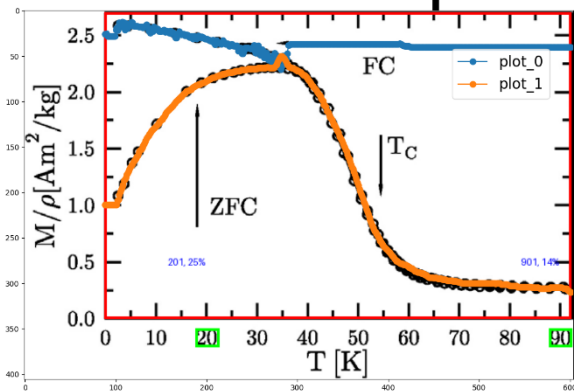
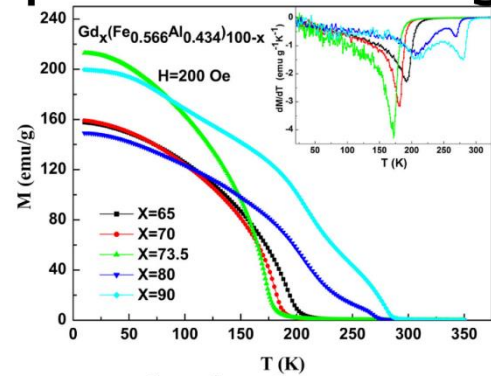


Figure 31 JPCM\_1720052985\_image1 taken from (Cieślak et al., 2005)

# Input Plot Image



# Model Output

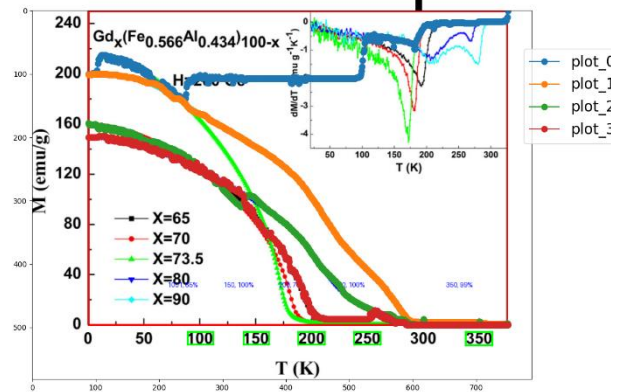
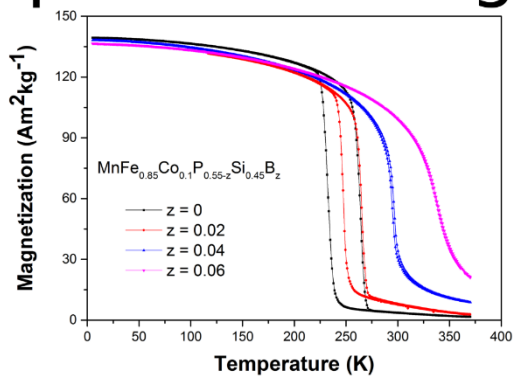


Figure 32 035220\_1\_online\_image1 taken from (Zhang et al., 2016)

# Input Plot Image



# Model Output

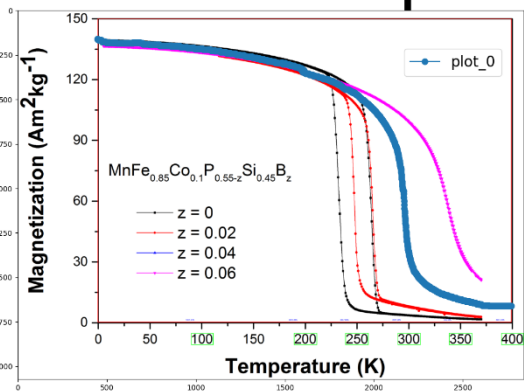
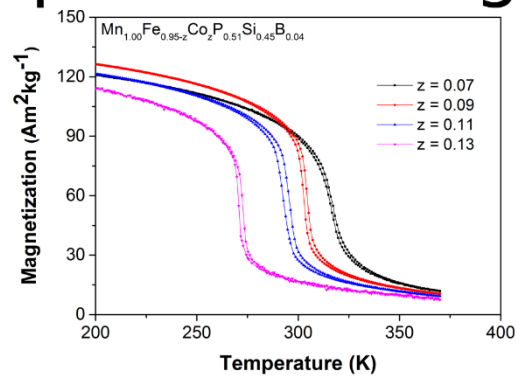


Figure 33 tunable\_image1 taken from (Thang et al., 2016)

# Input Plot Image



# Model Output

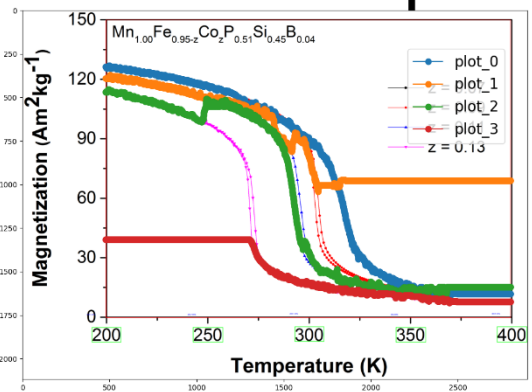
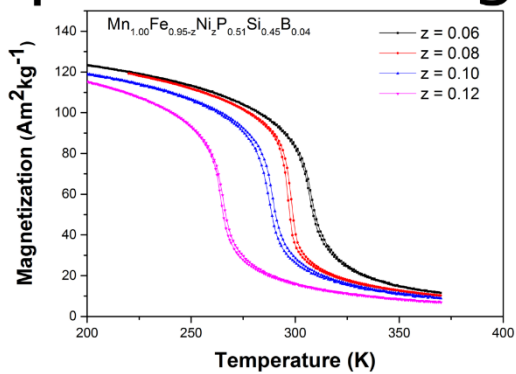


Figure 34 tunable\_image2 taken from (Thang et al., 2016)

# Input Plot Image



# Model Output

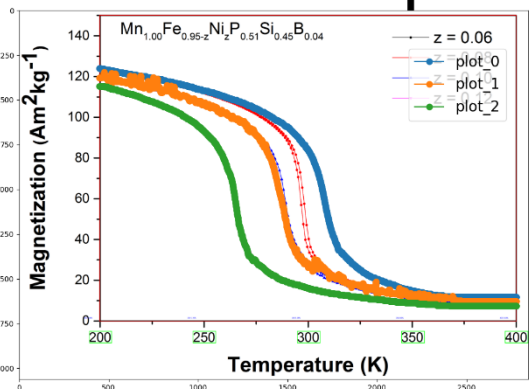
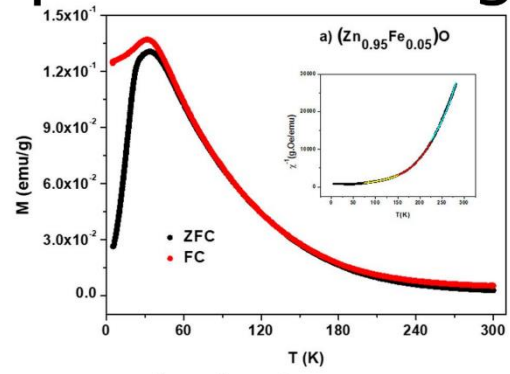


Figure 35 tunable\_image3 taken from (Thang et al., 2016)

# Input Plot Image



# Model Output

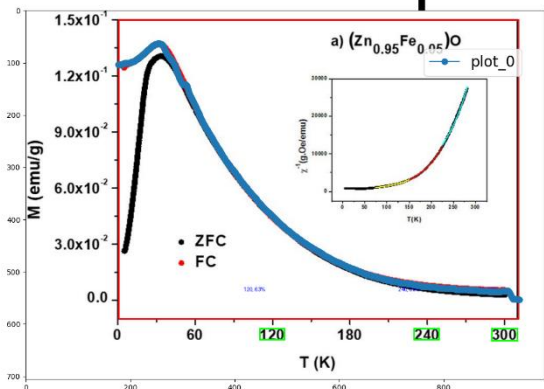
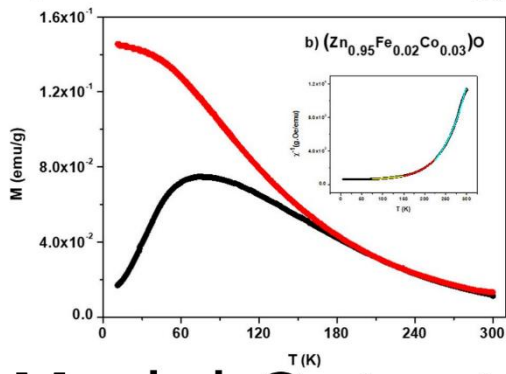
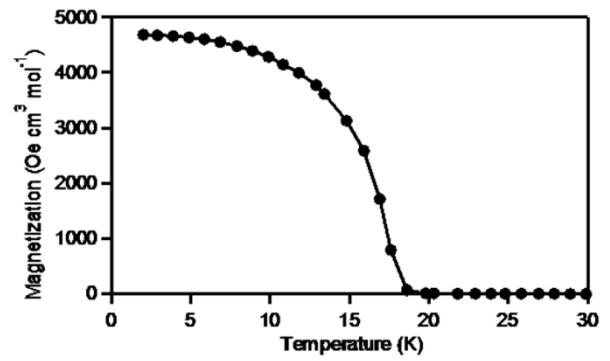


Figure 36 zn0-series5-1\_image1 taken from (Silva Tupan et al., 2019)

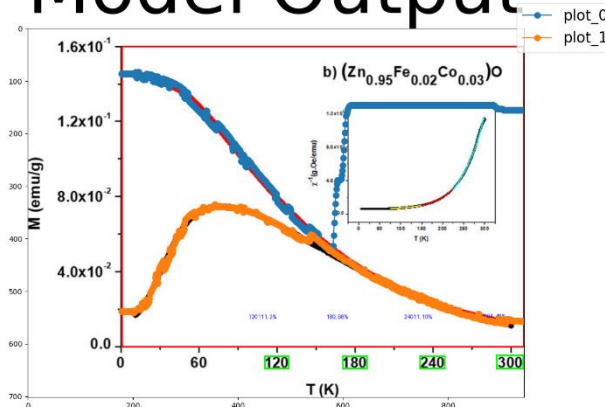
# Input Plot Image



# Input Plot Image



# Model Output



# Model Output

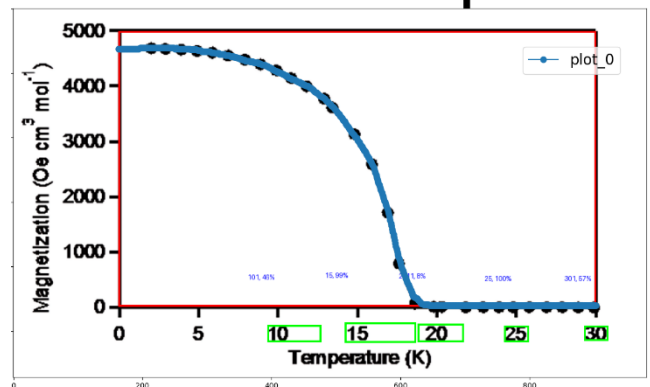
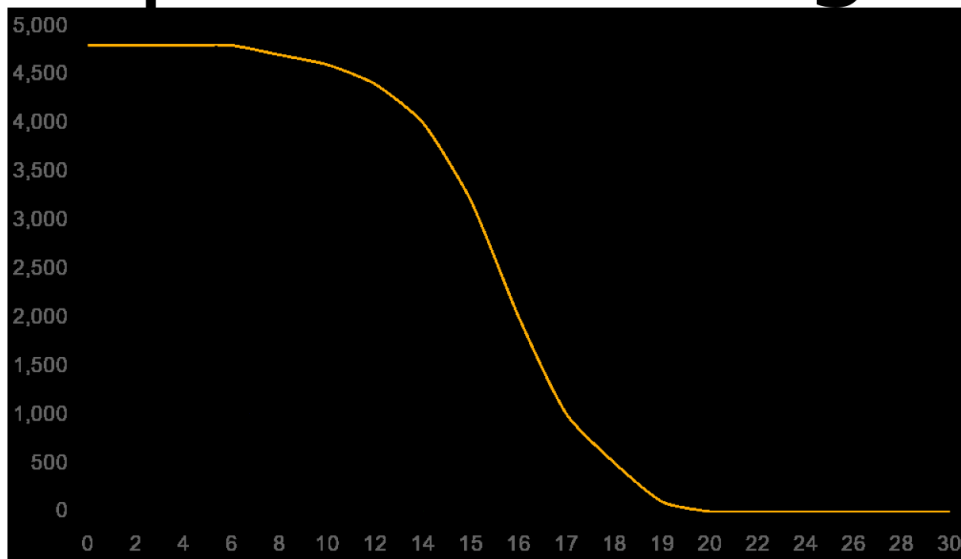


Figure 37 zno-serie5-1\_image2 taken from (Silva Tupan et al., 2019)

Figure 38 Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C\_image1 taken from (Imoto et al., 2017)

# Input Plot Image



# Model Output

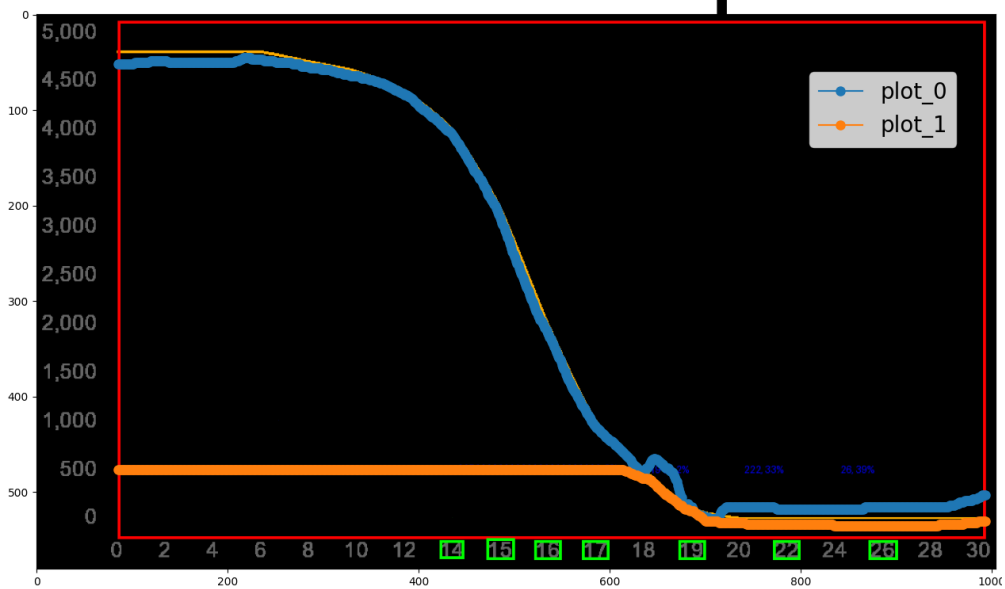


Figure 39 Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_Generated taken from (Imoto et al., 2017)

## 11.2 Corrected Plot2Spectra output

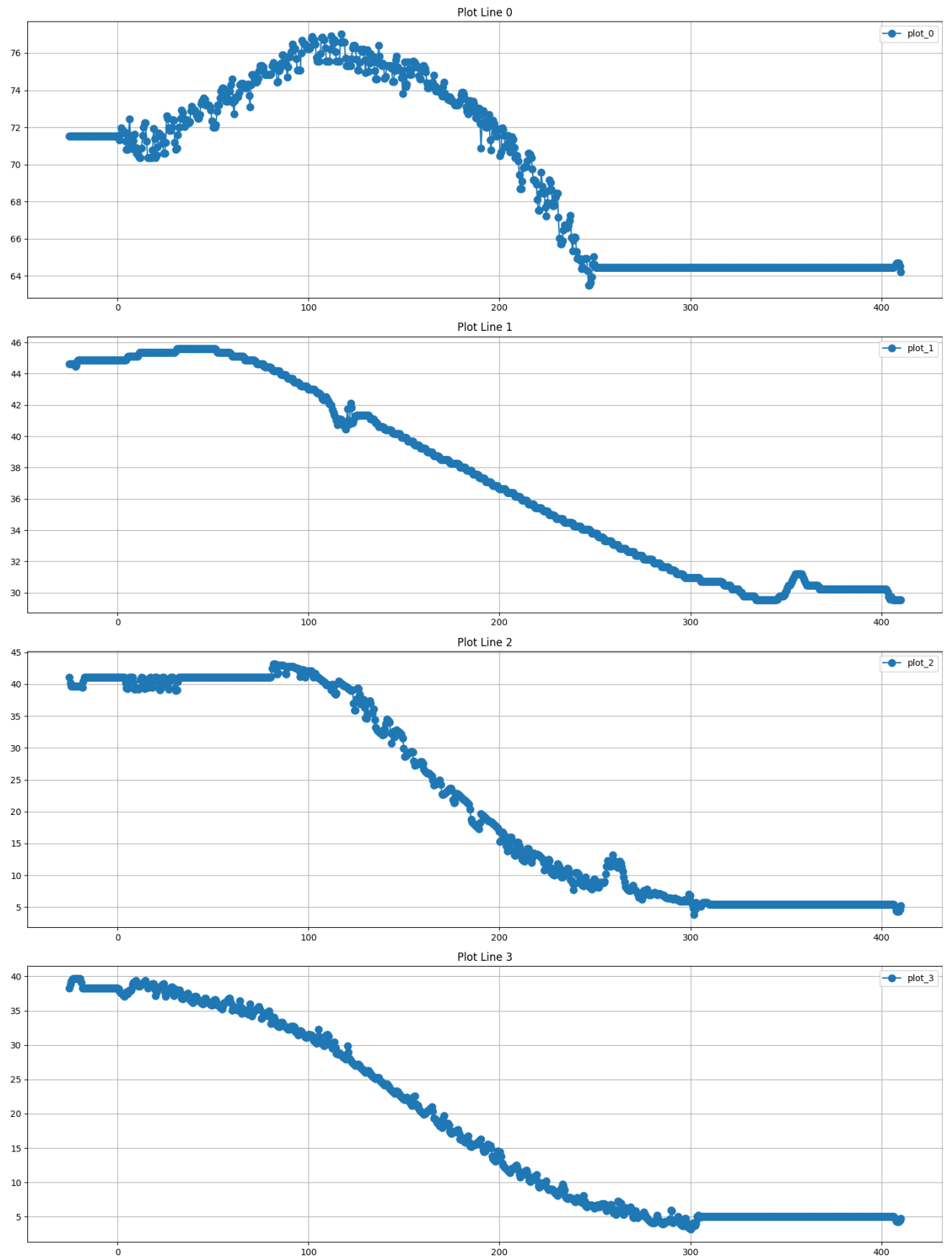


Figure 40 Corrected version of Figure 25

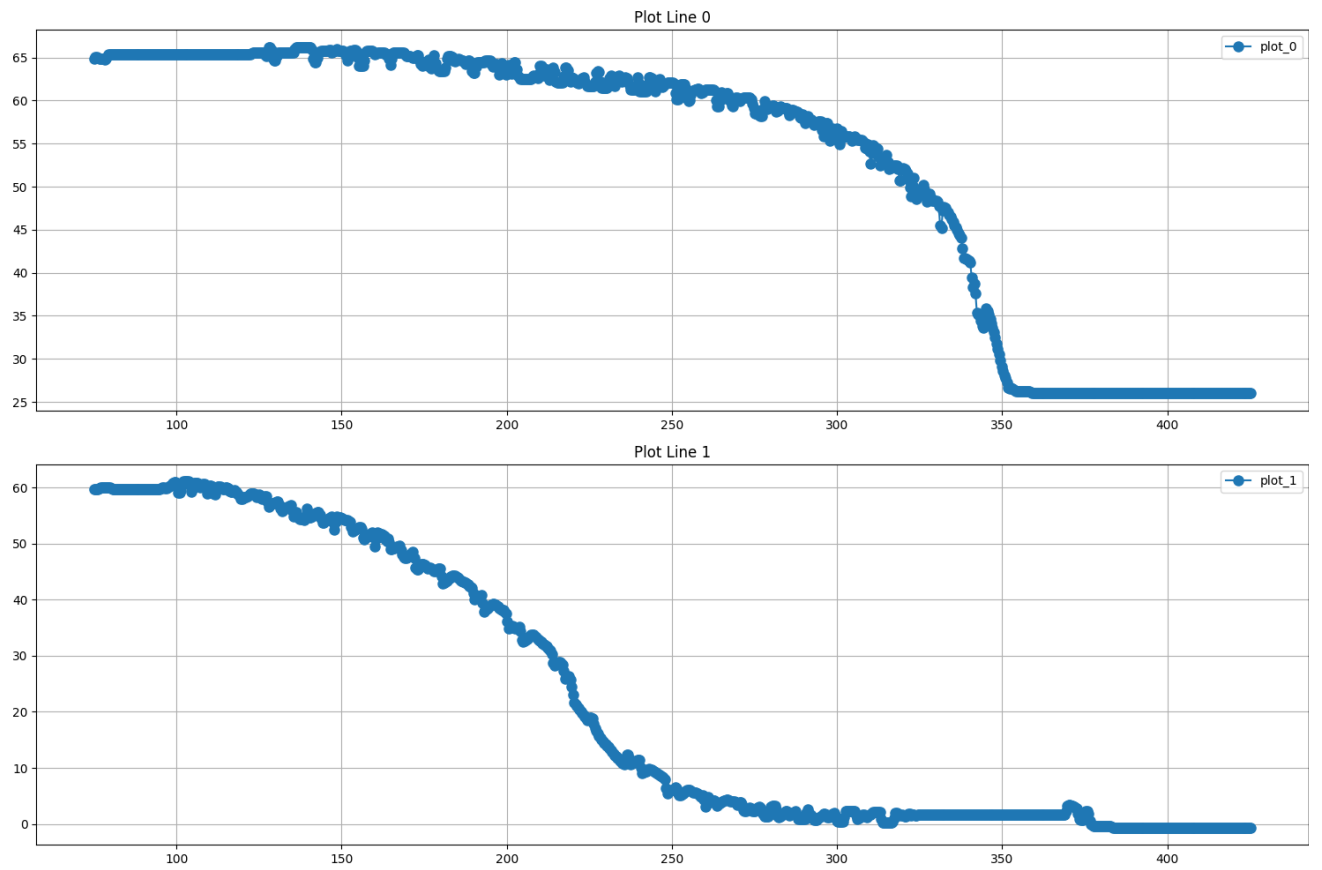
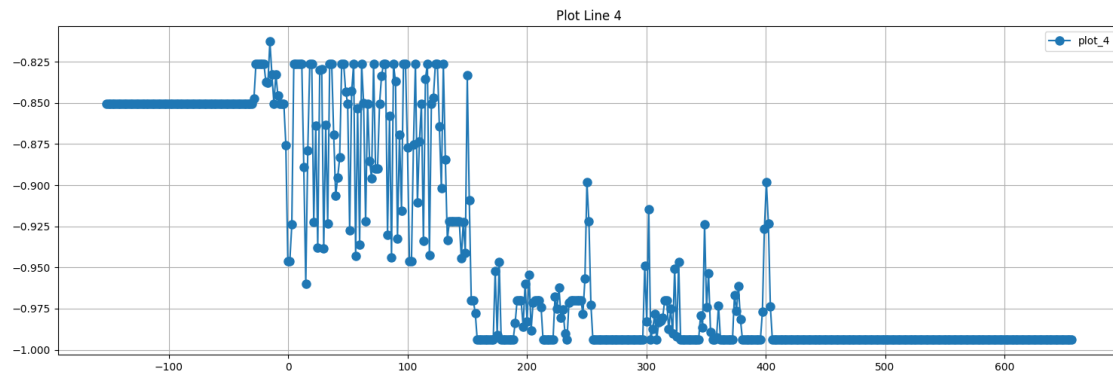
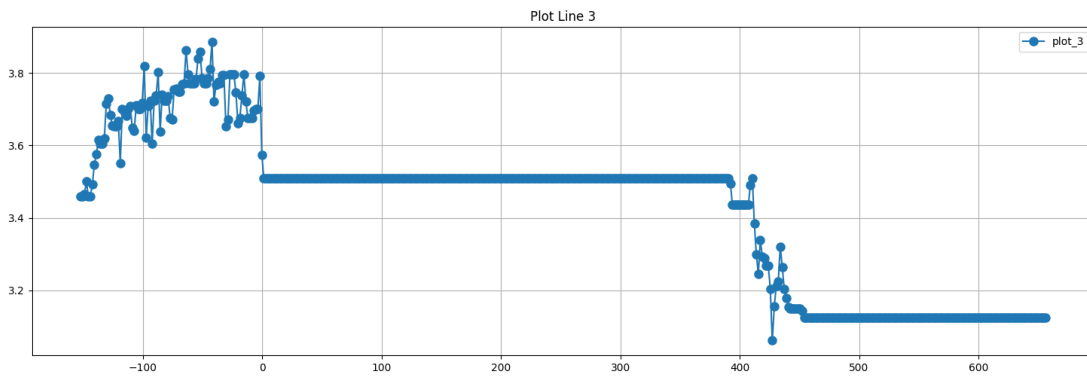
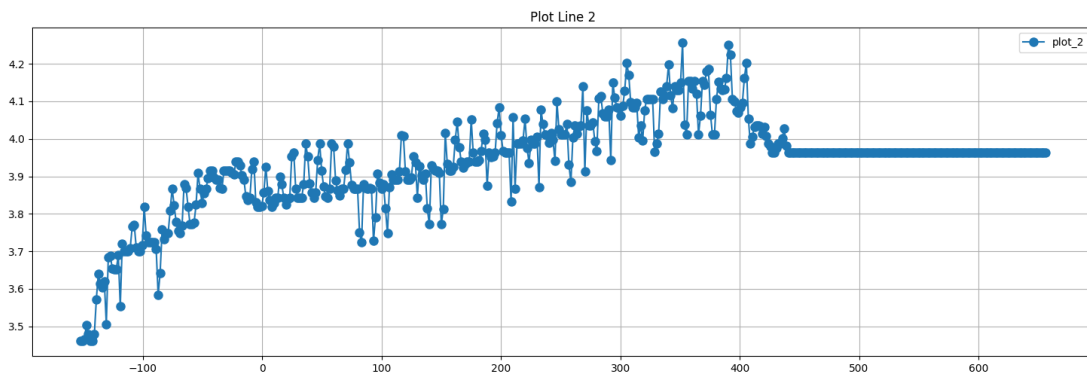
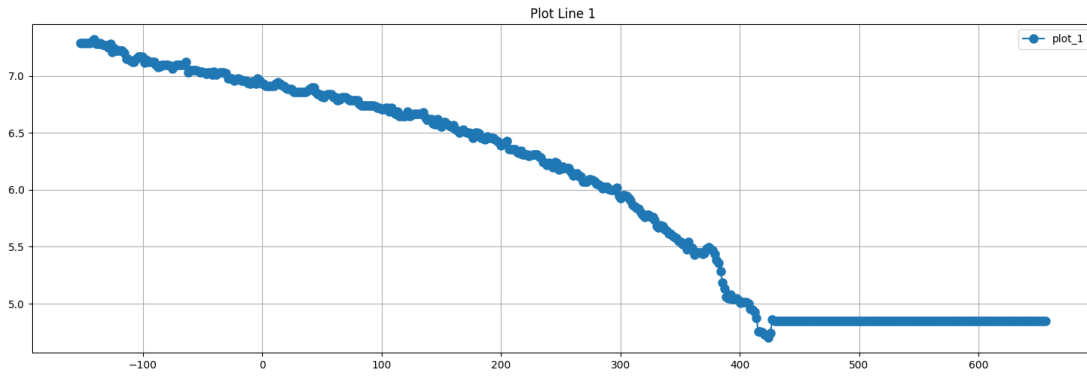
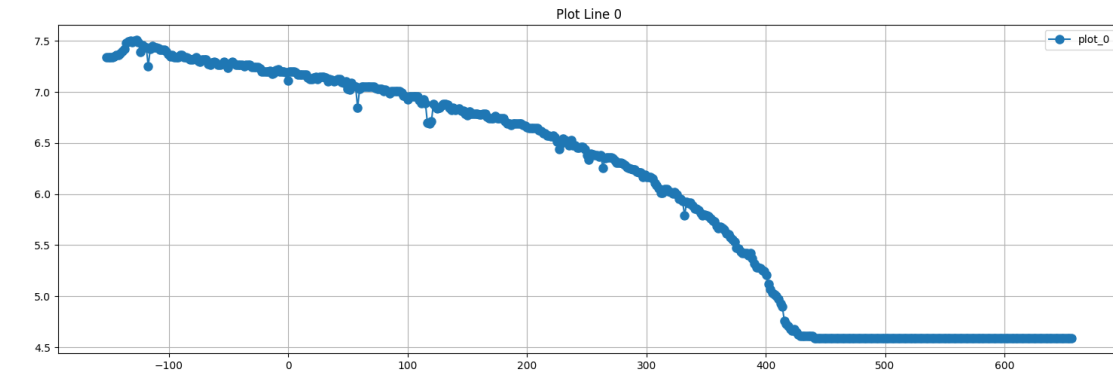


Figure 41 corrected version of Figure 26



---

Figure 42 corrected version of Figure 27

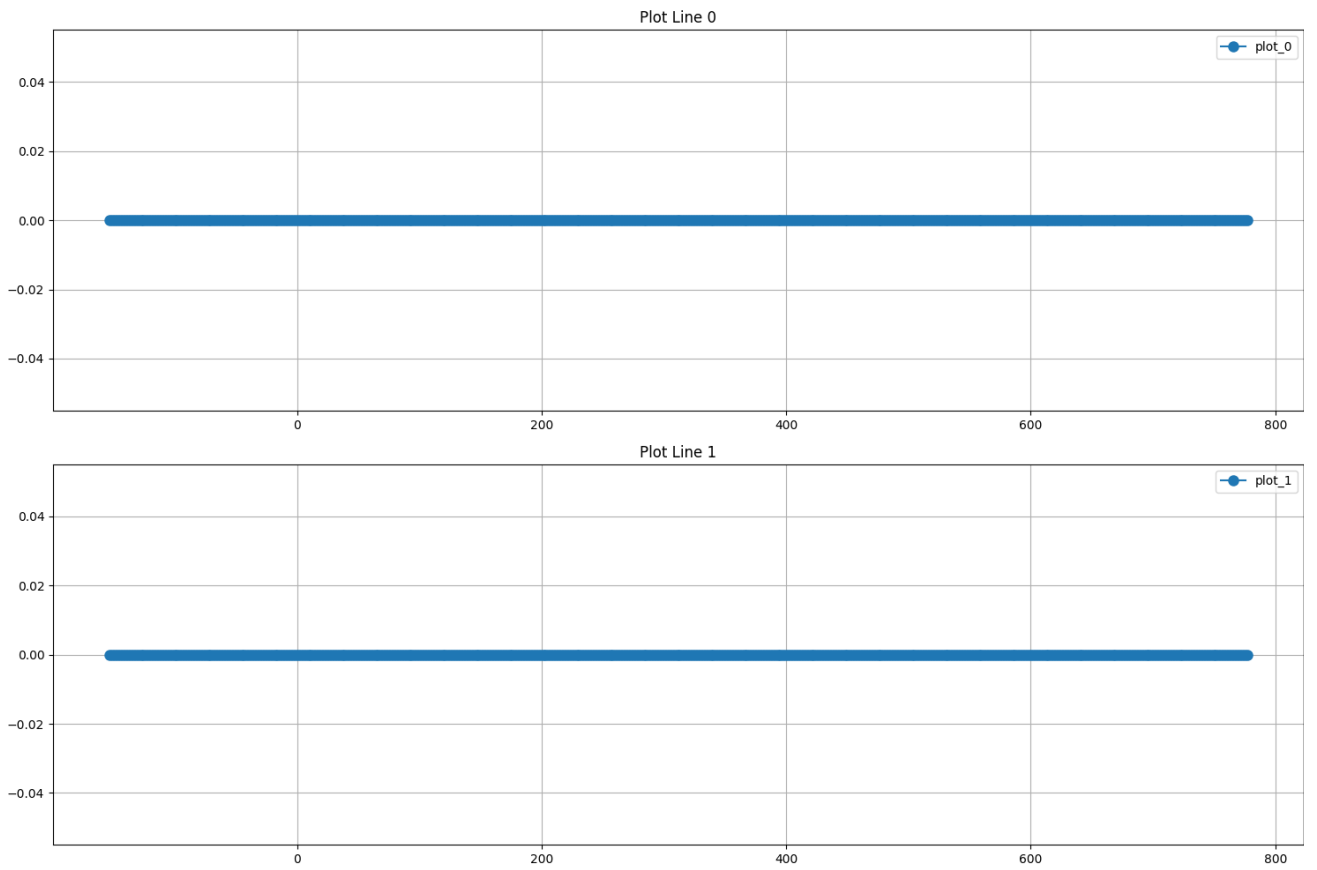


Figure 43 corrected version of Figure 28

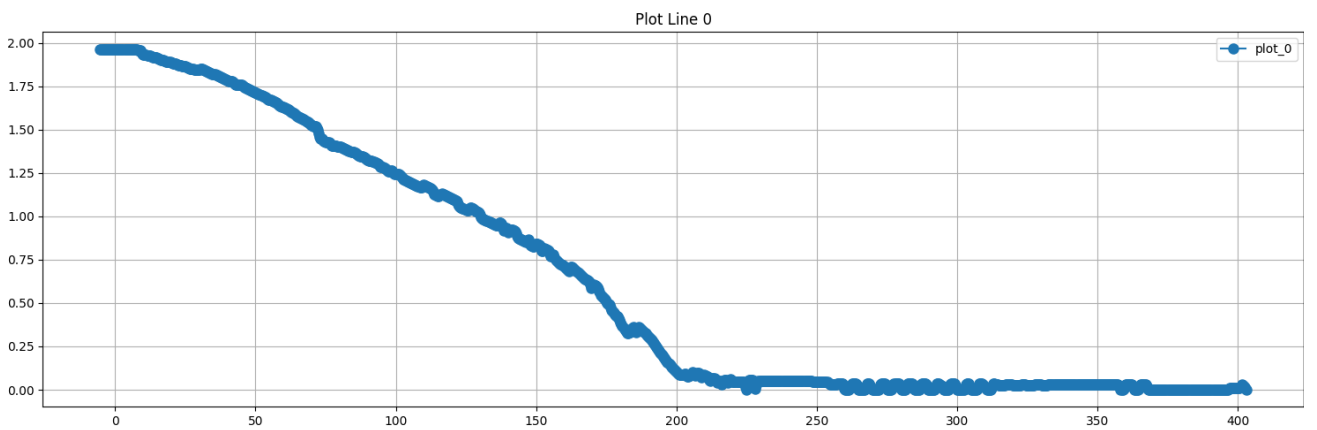


Figure 44 corrected version of Figure 29

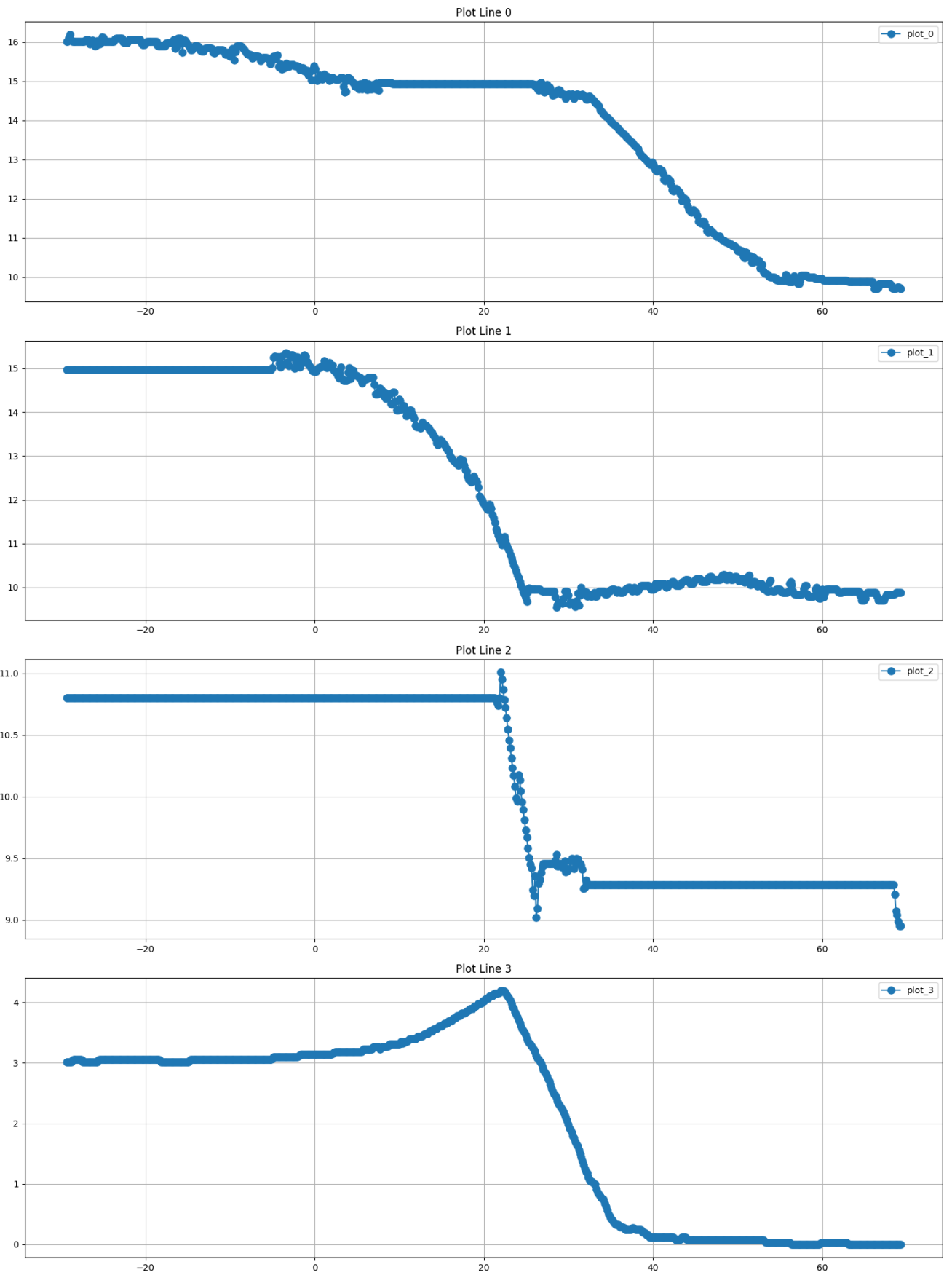


Figure 45 corrected version of Figure 30

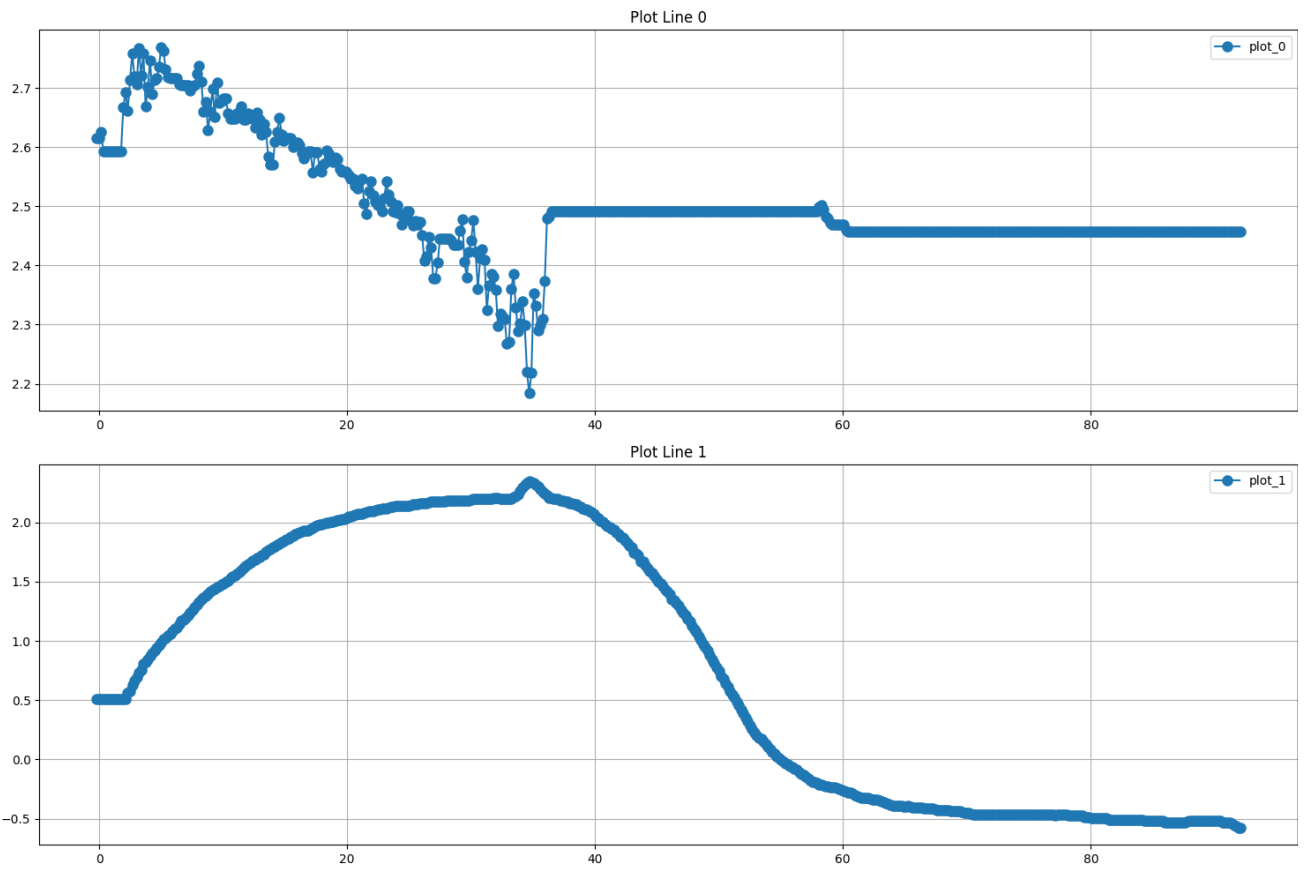


Figure 46 corrected figure of Figure 31

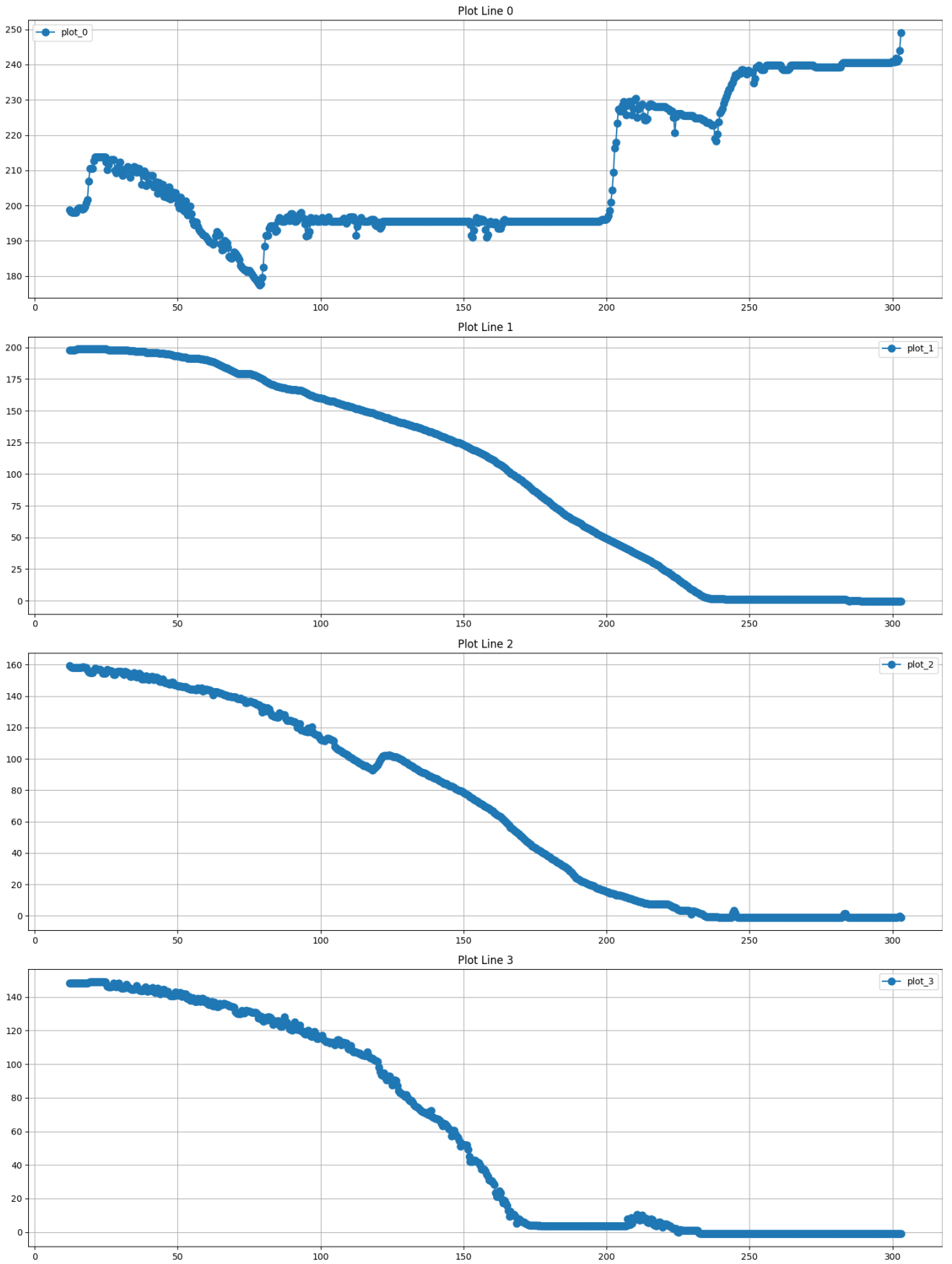


Figure 47 corrected figure of Figure 32

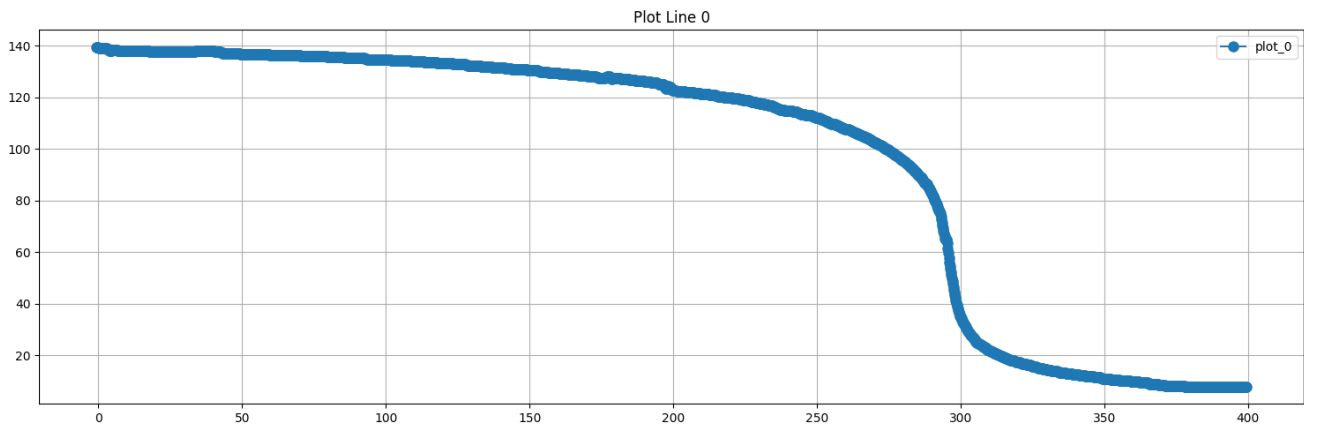


Figure 48 corrected figure of Figure 33

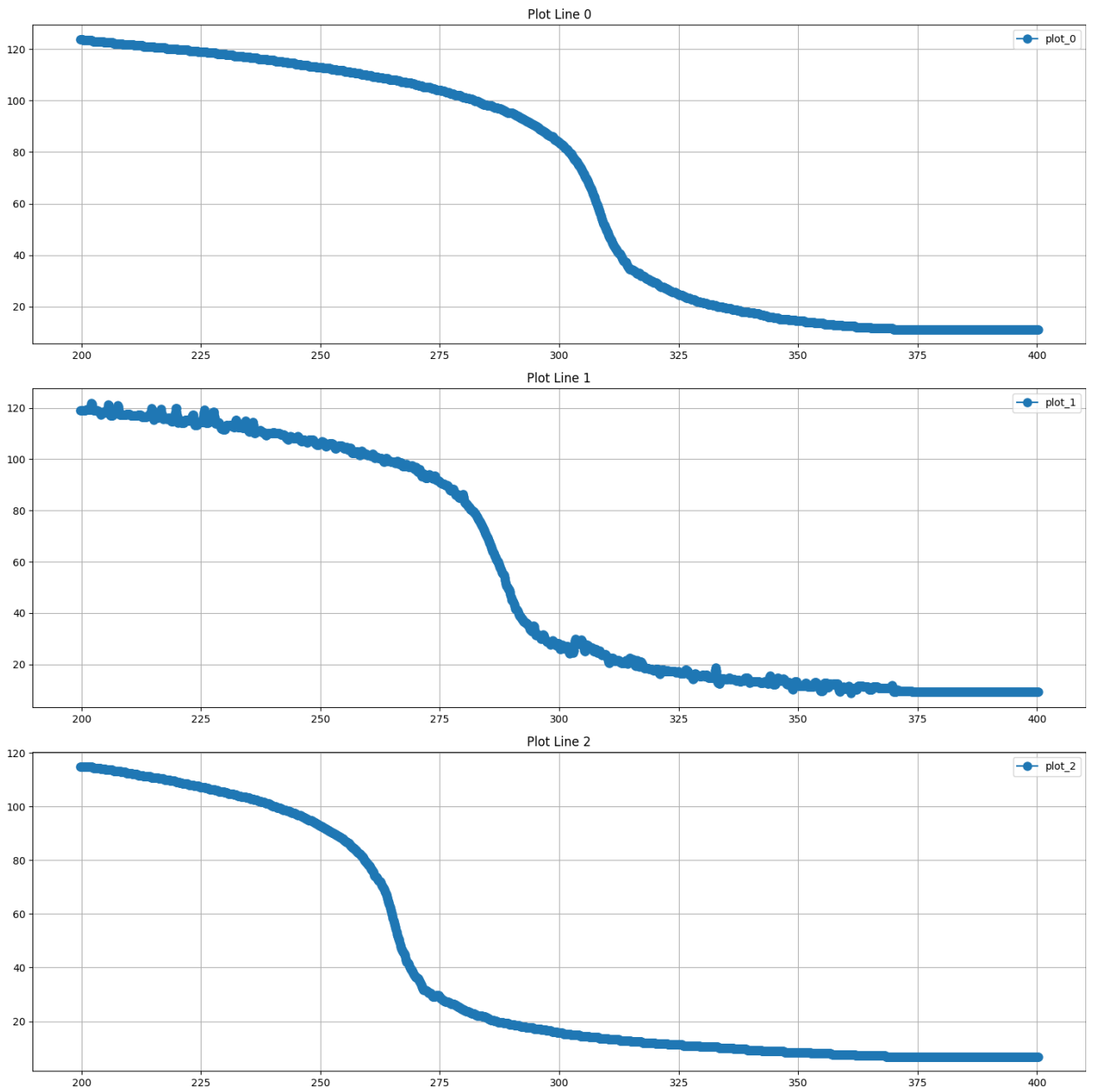


Figure 49 corrected figure of Figure 34

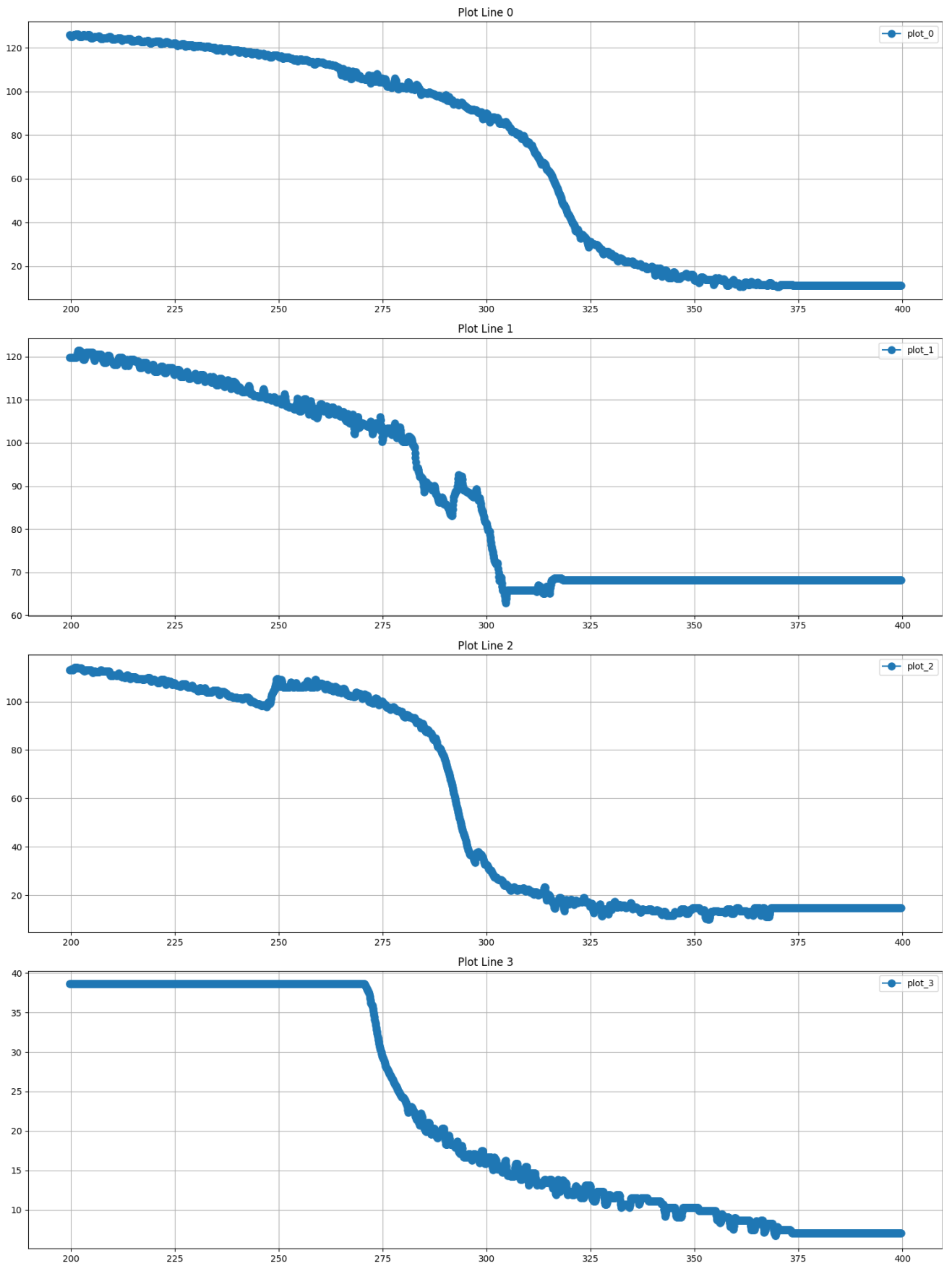


Figure 50 corrected version of Figure 35

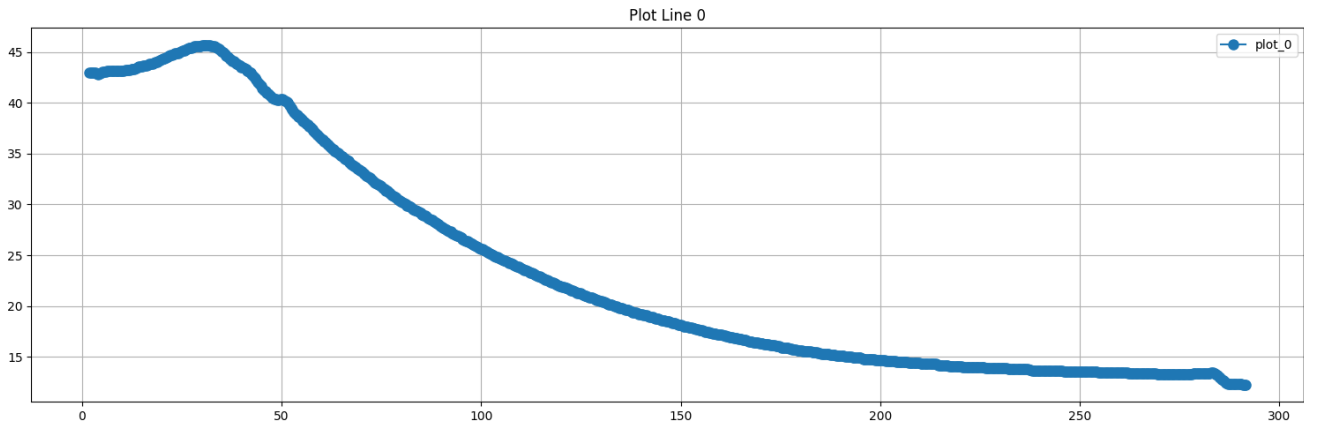


Figure 51 corrected version of Figure 36

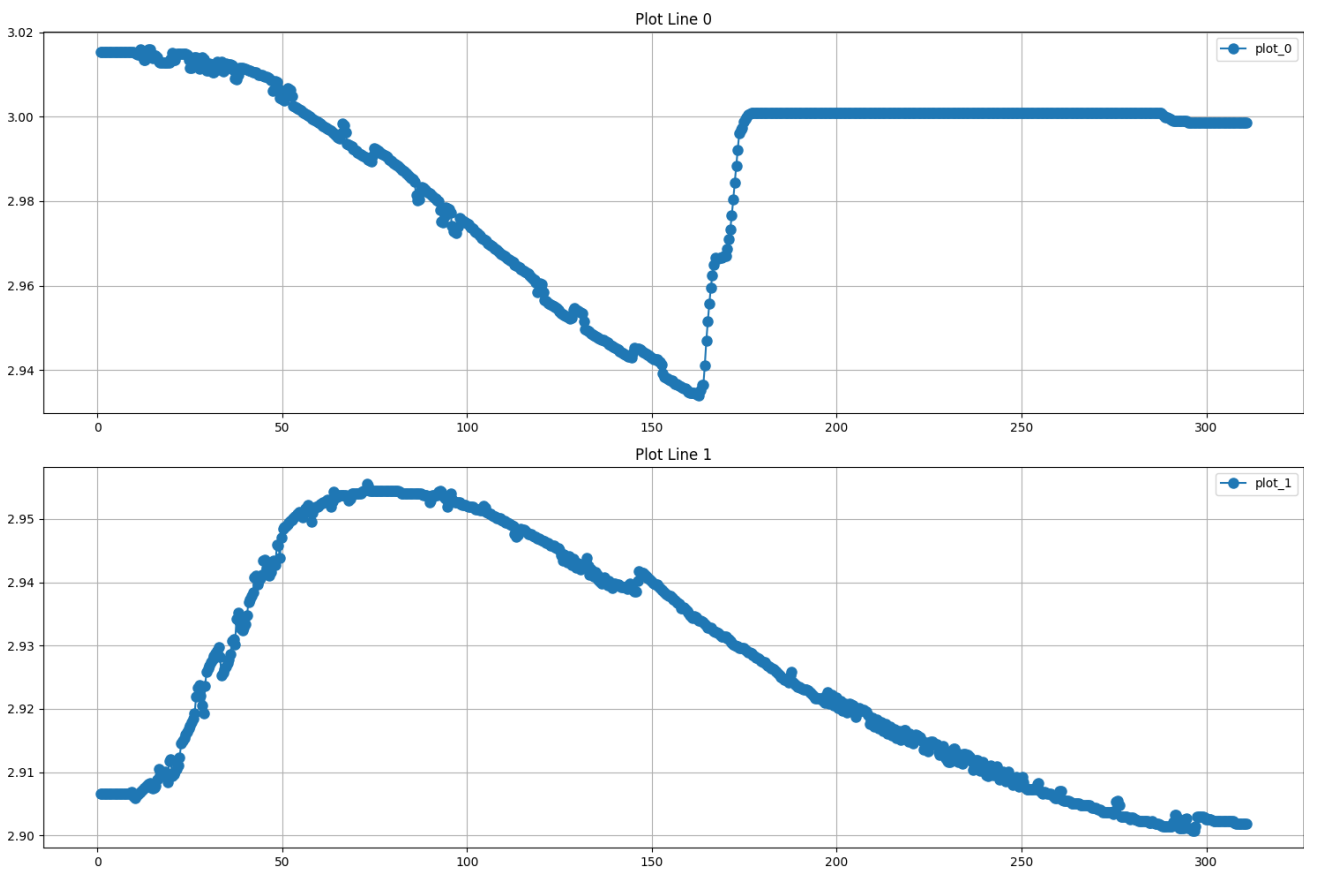


Figure 52 corrected version of Figure 37

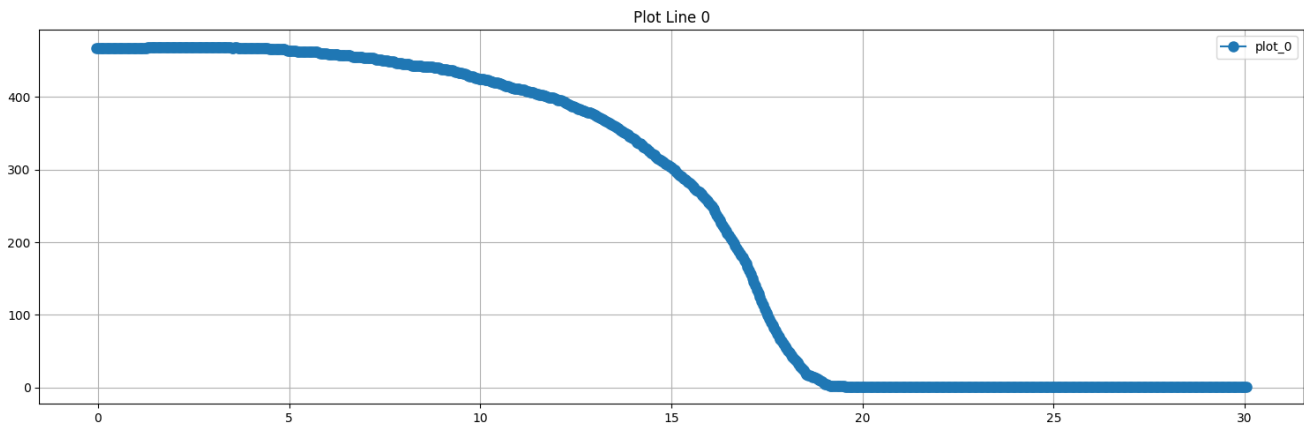


Figure 53 corrected version of Figure 38

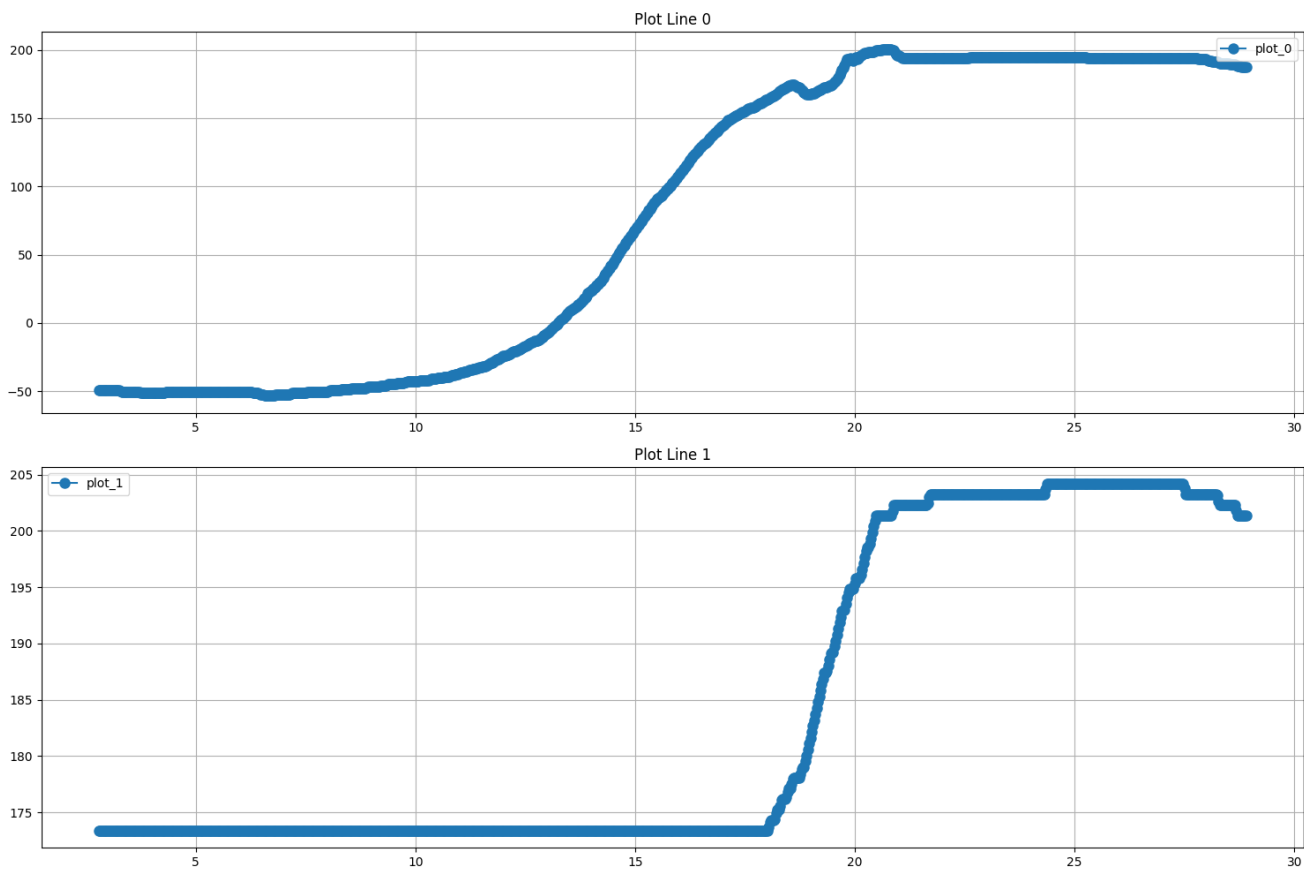


Figure 54 corrected version of Figure 39

### 11.3 Basel level code i.e. code that is used at the surface

#### 11.3.1 Llama 3 8b instruct

Python(LLM-TEST.py)

```
import os
import PyPDF2
import transformers
import torch
```

```

import time
from PIL import Image

def extract_text_from_pdf(pdf_path):
    text = ""
    with open(pdf_path, "rb") as file:
        pdf_reader = PyPDF2.PdfReader(file)
        num_pages = len(pdf_reader.pages)
        for page_number in range(num_pages):
            page = pdf_reader.pages[page_number]
            extracted_text = page.extract_text()
            if extracted_text:
                # Ensure the extracted text is in UTF-8
                extracted_text = extracted_text.encode('utf-8',
errors='ignore').decode('utf-8')
            else:
                extracted_text = " [No text on page] "
            text += extracted_text
    return text

base_instruction = (
    "You are going to be given {document_part} of a document. "
    "Please check this document to see if it contains a figure that depicts a "
    "magnetization vs. temperature (M vs T) PLOT. Keywords to look for include "
    "'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic
susceptibility'. "
    "ANSWER ONLY WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS "
    "THE M VS T FIGURE AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE,
ANSWER ONLY WITH a JSON "
    "return the following Json {"
    "\"title\": \"The title of the paper\";"
    "\"Figure found\": \"If a magnetization vs. temperature PLOT is found YES or
NO\";"
    "\"Figure\": \"which figure is the MAGNETIZATION VS. TEMPERATURE PLOT\";"
    "\"material\": \"what is the material that is the plot\"}"
)

def save_pdf_page_as_image(pdf_path, page_number, output_folder):
    pages = convert_from_path(pdf_path, first_page=page_number+1,
last_page=page_number+1)
    for i, page in enumerate(pages):
        image_path = os.path.join(output_folder, f"figure_page_{page_number +
1}.png")
        page.save(image_path, "PNG")
    return image_path

def check_pdf_folder(folder_path, output_folder):
    if not os.path.exists(output_folder):
        os.makedirs(output_folder)

```

```

model_id = "meta-llama/Meta-Llama-3-8B-Instruct"
access_token = "hf_FyzDPMFqjeYFDrabLAtBGWwPkMym1RGqAa"
tokenizer = transformers.AutoTokenizer.from_pretrained(
    model_id,
    token=access_token,
)
pipeline = transformers.pipeline(
    "text-generation",
    model=model_id,
    model_kwargs={"torch_dtype": torch.bfloat16 ,
'cache_dir': '/work/scratch/tp79diga/model'},
    device="cuda",
    token=access_token,
    tokenizer=tokenizer
)

eos_token_id = pipeline.tokenizer.eos_token_id

results = {}
max_tokens = 8192
max_output_tokens = 256
token_threshold = 3000 # Define what constitutes a "part" versus a "whole"
document based on token count

for filename in os.listdir(folder_path):
    if filename.lower().endswith('.pdf'):
        start_time = time.time()
        free_memory_before, total_memory = torch.cuda.mem_get_info()
        used_memory_before = total_memory - free_memory_before

        pdf_path = os.path.join(folder_path, filename)
        pdf_text = extract_text_from_pdf(pdf_path)

        text_tokens = tokenizer.encode(pdf_text)
        total_text_tokens = len(text_tokens)

        # Determine if the document is being processed in parts or as a whole
based on tokens
        document_part = "the whole" if total_text_tokens <= token_threshold else
"a part"

        # Update the instruction with the correct document context
        instruction = base_instruction.format(document_part=document_part)

        messages = [
            {"role": "system", "content": instruction},
            {"role": "user", "content": pdf_text},
        ]

```

```

    prompt = pipeline.tokenizer.apply_chat_template(
        messages,
        tokenize=False,
        add_generation_prompt=True
    )
    instruction_tokens = len(tokenizer.encode(prompt)) -
len(tokenizer.encode(pdf_text))

    chunk_size = max_tokens - instruction_tokens - max_output_tokens
    chunks = [text_tokens[i:i + chunk_size] for i in range(0,
len(text_tokens), chunk_size)]

    all_responses = []
    for chunk in chunks:
        chunk_text = tokenizer.decode(chunk)
        chunk_messages = [
            {"role": "system", "content": instruction},
            {"role": "user", "content": chunk_text},
        ]

        chunk_prompt = pipeline.tokenizer.apply_chat_template(
            chunk_messages,
            tokenize=False,
            add_generation_prompt=True
        )

        outputs = pipeline(
            chunk_prompt,
            max_new_tokens=max_output_tokens,
            eos_token_id=eos_token_id,
            do_sample=False,
            temperature=0.6,
            top_p=0.95,
        )

        chunk_response = outputs[0]["generated_text"][len(chunk_prompt):]
        all_responses.append(chunk_response.strip())

    final_response = " ".join(all_responses)

    end_time = time.time()
    processing_time = end_time - start_time

    free_memory_after, _ = torch.cuda.mem_get_info()
    used_memory_after = total_memory - free_memory_after

    results[filename] = {
        "response": final_response,

```

```

        "input_text": pdf_text.strip()[:1000],
        "total_tokens": total_text_tokens,
        "processing_time_seconds": processing_time,
        "memory_used_before_mb": used_memory_before / 1024 / 1024,
        "memory_used_after_mb": used_memory_after / 1024 / 1024,
        "input_tokens": len(text_tokens),
        "output_tokens": sum(len(tokenizer.encode(resp)) for resp in
all_responses),
    }

    torch.cuda.empty_cache()

    return results

# Example usage:
folder_path = 'PDFS'
output_folder = 'test_images'
pdf_results = check_pdf_folder(folder_path, output_folder)
print(base_instruction.format(document_part="an example"))
print("-----")

for filename, details in pdf_results.items():
    print(f"PDF File: {filename}")
    print(f"Response: {details['response']}")
    print(f"Processing Time: {details['processing_time_seconds']} seconds")
    print(f"Input Tokens: {details['input_tokens']}")
    print(f"Output Tokens: {details['output_tokens']}")
    print("-----")

```

### Sbatch(LLM-TEST.sh)

```

#!/bin/bash

#SBATCH -J LLM-TEST
#SBATCH --mail-type=ALL
# Please check pathes (directories have to exist beforehand):
#SBATCH -e /home/tp79diga/LLM-TEST/errors/%x.err.%j
#SBATCH -o /home/tp79diga/LLM-TEST/outputs/%x.out.%j
#
#
# CPU specification
#SBATCH -n 1           # 1 process
#SBATCH -c 24         # 24 CPU cores per process
#                      can be referenced as $SLURM_CPUS_PER_TASK in the
"payload" part
#SBATCH --mem-per-cpu=3800 # Hauptspeicher in MByte pro Rechenkern
#SBATCH -t 00:25:00     # in hours:minutes, or '#SBATCH -t 10' - just minutes

```

```

# GPU specification
#SBATCH --gres=gpu:a100      # 1 GPUs of type NVidia "A100"

# -----
# your job's "payload" in form of commands to execute, eg.
module purge
module load gcc/8 python/3.10 cuda
cd /work/scratch/tp79diga/test
source LLM/bin/activate
python LLM-TEST.py

# specification from OMP_NUM_THREADS depends on your program
export OMP_NUM_THREADS=$SLURM_CPUS_PER_TASK

# for checking whether and which GPUs have been allocated
# (output appears in the "#SBATCH -e" file specified above):
nvidia-smi 1>&2

EXITCODE=$?

# any cleanup and copy commands:

# end this job script with precisely the exit status of your scientific program
above:
exit $EXITCODE

```

### 11.3.2 Digitizer

#### Python ChartDETE & EasyOCR (lable.py)

```

import os
import easyocr
import cv2
import mmcv
from mmdet.apis import init_detector, inference_detector
import numpy as np
import matplotlib.pyplot as plt
import json
import re

# Paths for images and output
image_folder = './example_plot_images/'
save_folder = './data/ocr_results_folder/'
os.makedirs(save_folder, exist_ok=True)

import os
import shutil

# List of folder paths to delete contents from
folders = [
    './data/input_plot_extraction/gtFine/test/raman_xanes/',

```

```

    './data/input_plot_extraction/leftImg8bit/test/raman_xanes/',
    './data/ocr_results_folder/',
    './data/regions/'
]

def delete_folder_contents(folder_path):
    for filename in os.listdir(folder_path):
        file_path = os.path.join(folder_path, filename)
        try:
            if os.path.isfile(file_path) or os.path.islink(file_path):
                os.unlink(file_path)
            elif os.path.isdir(file_path):
                shutil.rmtree(file_path)
        except Exception as e:
            print(f'Failed to delete {file_path}. Reason: {e}')

# Delete contents of each folder
for folder in folders:
    delete_folder_contents(folder)
    print(f"Contents of folder '{folder}' have been deleted.")

print("All specified folder contents have been deleted.")

print("runing")
# Initialize EasyOCR reader
reader = easyocr.Reader(['en']) # Assuming English text, specify other languages if
needed
# Define a regular expression pattern for capturing numbers (including decimals)z
number_pattern = re.compile(r'\d{1,3}(?:\d{3})*(?:\.\d+)?')

# Load model configuration and weights
config_file = './work_dirs/cascade_rcnn_swin-
t_fpn_LGF_VCE_PCE_coco_focalsmoothloss/cascade_rcnn_swin-
t_fpn_LGF_VCE_PCE_coco_focalsmoothloss.py'
checkpoint_file = './work_dirs/cascade_rcnn_swin-
t_fpn_LGF_VCE_PCE_coco_focalsmoothloss/checkpoint.pth'
model = init_detector(config_file, checkpoint_file, device='cuda:0')

# Load configuration
config = mmcv.Config.fromfile(config_file)
classes = config.data.train.classes
xlabel_id = classes.index('xlabel')
ylabel_id = classes.index('ylabel')
xtick_id = classes.index('x_tick')
ytick_id = classes.index('y_tick')
plot_area_id = classes.index('plot_area')

```

```

# Define function to check alignment
def is_aligned(label_bbox, tick_bbox, axis):
    if axis == 'x': # Horizontal alignment for x-labels
        return abs((label_bbox[0] + label_bbox[2])/2 - (tick_bbox[0] +
tick_bbox[2])/2) < 50
    else: # Vertical alignment for y-labels
        return abs((label_bbox[1] + label_bbox[3])/2 - (tick_bbox[1] +
tick_bbox[3])/2) < 50

# Process images
image_files = [f for f in os.listdir(image_folder) if
os.path.isfile(os.path.join(image_folder, f))]
for image_file in image_files:
    img_path = os.path.join(image_folder, image_file)
    result = inference_detector(model, img_path)
    img = cv2.imread(img_path)
    img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    bbox_result = result if not isinstance(result, tuple) else result[0]
    plot_area_bboxes = []
    for bbox in bbox_result[plot_area_id]:
        if bbox[4] > 0.5: # Confidence threshold
            bbox_dict = {'bbox': bbox[:4].tolist()}
            plot_area_bboxes.append(bbox_dict)

# Data to be saved
data_to_save = {'xlabel': [], 'ylabel': [], 'x_tick': [], 'y_tick': []}
ticks = {'x_tick': [], 'y_tick': []}
data_to_save['plot_area'] = plot_area_bboxes

for label_id, bboxes in enumerate(bbox_result):
    if label_id in [xlabel_id, ylabel_id, x_tick_id, y_tick_id]:
        label = 'xlabel' if label_id == xlabel_id else 'ylabel' if label_id ==
ylabel_id else 'x_tick' if label_id == x_tick_id else 'y_tick'
        for bbox in bboxes:
            if bbox[4] > 0.5: # Confidence threshold
                bbox_dict = {'bbox': bbox[:4].tolist()}
                if label in ['xlabel', 'ylabel']: # OCR only for these labels
                    x_min, y_min, x_max, y_max = map(int, bbox[:4])
                    roi = img[y_min:y_max, x_min:x_max]
                    ocr_result = reader.readtext(roi)
                    # Extracting only numbers including decimal points from OCR
results
                    text = ' '.join([number_pattern.search(res[1]).group(0) for
res in ocr_result if number_pattern.search(res[1])])
                    bbox_dict['text'] = text
                    data_to_save[label].append(bbox_dict)
                    if label in ['x_tick', 'y_tick']:
                        ticks[label].append(bbox[:4])

# Link ticks to labels

```

```

for label in ['xlabel', 'ylabel']:
    for label_entry in data_to_save[label]:
        label_bbox = label_entry['bbox']
        axis = 'x' if label == 'xlabel' else 'y'
        tick_label = 'x_tick' if axis == 'x' else 'y_tick'
        aligned_ticks = [tick for tick in ticks[tick_label] if
is_aligned(label_bbox, tick, axis)]
        label_entry['ticks'] = [tick.tolist() for tick in aligned_ticks]

# Save results to JSON
json_path = os.path.join(save_folder,
f'{os.path.splitext(image_file)[0]}_results.json')
with open(json_path, 'w') as json_file:
    json.dump(data_to_save, json_file, indent=4)

# Optionally display the image
plt.figure(figsize=(10, 10))
plt.imshow(img)
plt.title(image_file)
plt.axis('off')
plt.show()

```

### Python Plot2Spectra (Digitiser.py)

```

# necessary package to import
import numpy as np
import matplotlib.pyplot as plt
import glob
import os
from PIL import Image, ImageDraw, ImageFont
from distinctipy import distinctipy
import json
import torch, torchvision
import mmcv
import mmdet
from mmcv.ops import get_compiling_cuda_version, get_compiler_version
from src.axis_alignment.utils import dict2class, AxisAlignment
import csv

# pick 30 visually distinct colors for better visuallization experience
colors = distinctipy.get_colors(100)

opt = {
    # region detection
    "config_file": "./checkpoints/axis_alignment/fcos_x101_64x4d_fpn_gn-
head_mstrain_640-800_2x_coco.py",

```

```

"checkpoint_file": "./checkpoints/axis_alignment/epoch_200.pth",
"refinement": True,
# tick detection
"cuda": True,
"canvas_size": 1280,
"mag_ratio": 1.5,
"poly": False,
"text_threshold": 0.7,
"low_text": 0.4,
"link_threshold": 0.7,
"show_time": False,
"refine": True,
"trained_model": './checkpoints/axis_alignment/craft_mlt_25k.pth',
"refiner_model": './checkpoints/axis_alignment/craft_refiner_CTW1500.pth',
# tick recognition
"workers": 4,
"saved_model": "./checkpoints/axis_alignment/TPS-ResNet-BiLSTM-Attn.pth",
"batch_max_length": 25,
"imgH": 32,
"imgW": 100,
"rgb": False,
"character": "0123456789abcdefghijklmnopqrstuvwxyz",
"sensitive": False,
"PAD": True,
"Transformation": "TPS",
"FeatureExtraction": "ResNet",
"SequenceModeling": "BiLSTM",
"Prediction": "Attn",
"num_fiducial": 20,
"input_channel": 1,
"output_channel": 512,
"hidden_size": 256,
}

# initialize the axis alignment module
axis_alignment = AxisAlignment(opt)
print("Initialize axis_alignment succses")

# load all images in the given folder
path = "./example_plot_images/"
axis_alignment.load_data(path)

print("load all images axis_alignment succses")
len(axis_alignment.imglist)
# save region of interest
region_path = "./data/regions"
axis_alignment.crop(region_path)

```

```

print("load all images axis_alignment succses")

from src.plot_data_extraction.plot_digitizer import PlotDigitizer
from src.plot_data_extraction.SpatialEmbeddings.src.utils import transforms as
my_transforms
from src.plot_data_extraction.evaluation import PlotEvaluator
from src.plot_data_extraction.utils import GenerateTestData, dict2class, Segmap2Lines

# convert the format of the plot region to fit the requirement of the plot data
extraction module
GenerateTestData("./data/regions/",
                 "./data/input_plot_extraction")

opt = {
    "cuda": True,
    "display": True,
    "save": True,
    "save_dir": "./exp/",
    "root": "./data/input_plot_extraction/leftImg8bit/",
    "data_type": "test",
    "mode": "raman_xanes",
    "num_workers": 6,
    "checkpoint_path": "./checkpoints/plot_data_extraction/checkpoint_0999.pth",

    "dataset": {
        'name': 'cityscapes',
        'kwargs': {
            'root_dir': './data/input_plot_extraction',
            'type': 'test',
            'transform': my_transforms.get_transform([
                {
                    "name": "CustomResizePad",
                    "opts": {
                        'keys': ('image', 'instance', 'label'),
                        "is_test": True,
                    },
                },
                {
                    'name': 'ToTensor',
                    'opts': {
                        'keys': ('image', 'instance', 'label'),
                        'type': (torch.FloatTensor, torch.ByteTensor,
torch.ByteTensor),
                    }
                },
            ]),
        },
    },
}

```

```

        {
            'name': 'Normalize',
            'opts': {
                'keys': ('image'),
                'p': -1,
            }
        },
    ]),
}
},

"model": {
    'name': 'branched_erfnet',
    'kwargs': {
        'num_classes': [3, 1],
    }
}
}

# initialize the plot data extraction module
plot_digitizer = PlotDigitizer()
plot_digitizer.load_seg("spatialembdding", opt)
len(plot_digitizer.instance_seg)

"""
grid search for the best parameter setting:
- match_threshold (i.e. change the coefficient for different losses)
- start position
"""
def Run(plot_digitizer, img_id):
    plot_digitizer.predict_from_ins_seg(img_id)
    ins_map = plot_digitizer.result_dict["visual"]["ins_map"]
    img_gray = plot_digitizer.result_dict["visual"]["img_gray"]
    img_rgb = plot_digitizer.result_dict["visual"]["img_rgb"]
    seg_map = plot_digitizer.result_dict["visual"]["seg_map"]
    seg_plots = Segmap2Lines(ins_map)
    plot_digitizer.find_init_posi()
    plot_evaluator = PlotEvaluator(img_rgb, img_gray, seg_map)
    match_thresholds = [0.8,0.85,0.9,0.95,0.97,0.99]
    start_ids = plot_digitizer.result_dict["data"]["start_ids"]
    scores, param_space, line_sets = plot_evaluator.param_search(match_thresholds,
start_ids)
    mean_score = np.mean(scores, axis=1)
    s, m = param_space[np.argmax(mean_score)]
    line_set = line_sets[np.argmax(mean_score)]
    result = {}
    result["img_name"] = plot_digitizer.img_name
    result["start_posi"] = s

```

```

result["match_threshold"] = m
result["pred_line_set"] = line_set
result["seg_line_set"] = seg_plots
result["visual"] = [img_rgb, seg_map, ins_map]
return result

def print_results(result):
    print(f"Image Name: {result['img_name']}")
    print(f"Best Start Position: {result['start_posi']}")
    print(f"Best Match Threshold: {result['match_threshold']}")
    print("Predicted Line Set:")
    for line in result['pred_line_set']:
        print(f" {line}")
    print("Segmented Line Set:")
    for line in result['seg_line_set']:
        print(f" {line}")
    print("Visual Outputs:")
    print(f" RGB Image: {result['visual'][0].shape}")
    print(f" Segmentation Map: {result['visual'][1].shape}")
    print(f" Instance Map: {result['visual'][2].shape}")

# Create directories for saving images and CSV data if they don't exist
output_image_dir = "./output_images"
output_csv_dir = "./output_csv"
os.makedirs(output_image_dir, exist_ok=True)
os.makedirs(output_csv_dir, exist_ok=True)

# Iterate over all images in the dataset
for img_no in range(len(axis_alignment.imglist)):
    img_name = os.path.basename(axis_alignment.imglist[img_no])
    print(img_name)
    img, plot_bbox, results, results_all = axis_alignment.run(img_no)
    result = Run(plot_digitizer, img_no)

    # Prepare for saving the image and CSV
    img_name = os.path.basename(axis_alignment.imglist[img_no])
    img_name_no_ext = os.path.splitext(img_name)[0] # Remove file extension
    img_output_path = os.path.join(output_image_dir, f"{img_name_no_ext}_output.png")
    csv_output_path = os.path.join(output_csv_dir,
f"{img_name_no_ext}_corrected.csv")

    # Construct the path to the matching JSON file based on the image name
    json_file_path = f'./data/ocr_results_folder/{img_name_no_ext}_results.json'

    # Open CSV file for writing the corrected data
    with open(csv_output_path, mode='w', newline='') as csv_file:
        csv_writer = csv.writer(csv_file)

    # Initialize a list to hold the headers and data rows

```

```

headers = []
rows = []

# Load the JSON data for OCR results from the dynamically constructed path
with open(json_file_path) as f:
    data = json.load(f)

# Extract the x-ticks and y-ticks with their labels
x_ticks = {(tick['bbox'][0] + tick['bbox'][2]) / 2: float(tick['text']) for
tick in data['xlabel'] if tick['text']}
y_ticks = {(tick['bbox'][1] + tick['bbox'][3]) / 2: float(tick['text']) for
tick in data['ylabel'] if tick['text']}

x_tvalues = np.array(list(x_ticks.keys()))
x_lvalues = np.array(list(x_ticks.values()))

# Calculate the linear trend line
x_coefficients = np.polyfit(x_tvalues, x_lvalues, 1) # degree 1 for linear
x_polynomial = np.poly1d(x_coefficients)

y_tvalues = np.array(list(y_ticks.keys()))
y_lvalues = np.array(list(y_ticks.values()))

# Calculate the linear trend line
y_coefficients = np.polyfit(y_tvalues, y_lvalues, 1) # degree 1 for linear
y_polynomial = np.poly1d(y_coefficients)

x_bias, y_bias = list(plot_bbox)[:2]
x_min, y_min, x_max, y_max = int(list(plot_bbox)[0]), list(plot_bbox)[1],
int(list(plot_bbox)[2]), list(plot_bbox)[3]
x_count = x_max - x_min + 1
norm_ts = np.linspace(0, x_max - x_min, x_count)
y_resize_ratio = (y_max - y_min) / result["visual"][0].shape[0]

plot_ts, plot_lines = result["pred_line_set"]["plot"]
norm_ts_pred = np.linspace(min(plot_ts), max(plot_ts), x_count)

conf_ts, conf_lines = result["pred_line_set"]["conf"]

for line_id in range(plot_lines.shape[1]):
    plot_line_norm = np.interp(norm_ts_pred, plot_ts, plot_lines[:, line_id])
    x_values = norm_ts + x_bias
    y_values = plot_line_norm * y_resize_ratio + y_bias

    # Corrected x and y values
    x_corrected = x_polynomial(x_values) # Apply the correction
    y_corrected = y_polynomial(y_values) # Apply the correction

# Prepare headers and data rows

```

```

headers.extend([f"plot_{line_id}_x", f"plot_{line_id}_y"])
rows.append(np.column_stack((x_corrected, y_corrected)))

# Write headers
csv_writer.writerow(headers)

# Write rows
max_len = max(len(row) for row in rows)
for i in range(max_len):
    row = []
    for col in rows:
        if i < len(col):
            row.extend(col[i])
        else:
            row.extend(['', ''])
    csv_writer.writerow(row)

# Display results
plt.figure(figsize=(20, 20))
img_raw = Image.open(axis_alignment.imglist[img_no]).convert("RGB")
plt.subplot(2, 1, 1)
plt.imshow(img_raw)
plt.axis("off")
plt.title("Input Plot Image", fontsize=100)

plt.subplot(2, 1, 2)
img_draw = img_raw.copy()
draw = ImageDraw.Draw(img_draw)
draw.rectangle(list(plot_bbox), outline=(255, 0, 0), width=3)
try:
    font = ImageFont.truetype("/usr/share/fonts/truetype/ubuntu/Ubuntu-
L.ttf", size=50)
except:
    font = ImageFont.load_default()
text_position = list(plot_bbox)[:2]
text_position[1] -= 80
draw.text(text_position, font=font, fill=(0, 0, 255), text="Plot Region",
align="left")

for idx, tick in enumerate(results):
    draw.rectangle(list(tick[2]), outline=(0, 255, 0), width=3)
    text_position = list(tick[2])[:2]
    text_position[1] -= 80
    text_position[0] -= 30
    try:
        font = ImageFont.truetype("/usr/share/fonts/truetype/ubuntu/Ubuntu-
L.ttf", size=50)
    except:
        font = ImageFont.load_default()

```

```

        draw.text(text_position, font=font, fill=(0, 0, 255), text="{},
{}%".format(tick[0], round(tick[1] * 100)),
                align="left")

plt.imshow(img_draw)

# Plot raw (non-corrected) values
for line_id in range(plot_lines.shape[1]):
    plot_line_norm = np.interp(norm_ts_pred, plot_ts, plot_lines[:, line_id])
    x_values = norm_ts + x_bias
    y_values = plot_line_norm * y_resize_ratio + y_bias
    plt.plot(x_values, y_values, "o-", markersize=8, label=f"plot_{line_id}")

plt.legend(fontsize=20, bbox_to_anchor=(0.97, 0.915))
plt.axis("on")
plt.title("Model Output", fontsize=100)

# Save the figure
plt.savefig(img_output_path)
plt.close()

fig, axs = plt.subplots(plot_lines.shape[1], 1, figsize=(15, 5 *
plot_lines.shape[1])) # Adjust subplot size as needed

# Check if we have more than one line to plot
if plot_lines.shape[1] > 1:
    axs = axs.flatten() # Flatten the axis array for easier iteration
else:
    axs = [axs] # Wrap the single AxesSubplot object in a list for consistency

for line_id in range(plot_lines.shape[1]):
    x_corrected, y_corrected = rows[line_id].T

    axs[line_id].plot(x_corrected, y_corrected, "o-", markersize=8,
label=f"plot_{line_id}")
    axs[line_id].set_title(f"Plot Line {line_id}")
    axs[line_id].legend()
    axs[line_id].grid(True)

plt.tight_layout()

# Save the corrected plot as an image
corrected_img_output_path = os.path.join(output_image_dir,
f"{img_name_no_ext}_corrected_output.png")
plt.savefig(corrected_img_output_path)
plt.close()

```

## Sbatch Digitizer (digi.sh)

```
#!/bin/bash

#SBATCH -J digi
#SBATCH --mail-type=ALL
# Please check pathes (directories have to exist beforehand):
#SBATCH -e /home/tp79diga/digi-HPC/errors/%x.err.%j
#SBATCH -o /home/tp79diga/digi-HPC/outputs/%x.out.%j
#
# CPU specification
#SBATCH -n 1                # 1 process
#SBATCH -c 24               # 24 CPU cores per process
#                           # can be referenced as $SLURM_CPUS_PER_TASK in the
#                           # "payload" part
#SBATCH --mem-per-cpu=3800  # Hauptspeicher in MByte pro Rechenkern
#SBATCH -t 00:30:00        # in hours:minutes, or '#SBATCH -t 10' - just minutes

# GPU specification
#SBATCH --gres=gpu:a100    # 1 GPUs of type NVidia "A100"

# -----
# your job's "payload" in form of commands to execute, eg.
module purge
module load gcc/8 python/3.8 cuda/11.6
cd /work/scratch/tp79diga/combo/
source chartdete/bin/activate
python lable.py

module purge
module load gcc python/3.8 cuda/11.1
source Plot2specV3/bin/activate
python Digitiser.py

# specification from OMP_NUM_THREADS depends on your program
export OMP_NUM_THREADS=$SLURM_CPUS_PER_TASK

# for checking whether and which GPUs have been allocated
# (output appears in the "#SBATCH -e" file specified above):
nvidia-smi 1

EXITCODE=$?

# any cleanup and copy commands:

# end this job script with precisely the exit status of your scientific program
above:
exit $EXITCODE
```

### 11.4 Llama 3 8b instruct output

You are going to be given an example of a document. Please check this document to see if it contains a figure that depicts a magnetization vs. temperature (M vs T) PLOT. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS THE M VS T FIGURE AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE, ANSWER ONLY WITH a JSON return the following Json {"title": "The title of the paper"; "Figure found": "If a magnetization vs. temperature PLOT is found YES or NO"; "Figure": "which figure is the MAGNETIZATION VS. TEMPERATURE PLOT"; "material": "what is the material that is the plot"}

-----  
PDF File:

Bommarito\_et\_al\_2023\_Temporal\_trends\_and\_predictors\_of\_gestational.pdf

Response: The JSON output is:

```
```  
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 2",  
  "material": "OPES (organophosphate esters)"  
}
```

``` The JSON output is:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Fig. 2",  
"material": "none"  
}
```

The paper discusses the exposure to organophosphate ester (OPE) flame retardants and plasticizers in pregnant women and their potential health risks. The authors analyzed data from the LIFECODES Fetal Growth Study, which included 900 pregnant women, and found that the concentrations of OPE biomarkers varied over time and were influenced by sociodemographic and pregnancy characteristics. The authors also found that the concentrations of some OPE biomarkers were higher in certain populations, such as younger women, those with lower educational attainment, and those from racial and ethnic minority groups. {

```
  "title": "Prenatal exposure to organophosphate esters is associated with  
decreased anogenital distance in offspring",  
  "Figure found": "YES",  
  "Figure": "Figure 1",  
  "material": "Organophosphate esters"  
}
```

Processing Time: 11.764989852905273 seconds

Input Tokens: 18882

Output Tokens: 266

-----  
PDF File:

Yang\_et\_al\_2023\_Detection\_of\_Various\_Microplastics\_in\_Patients\_Und.pdf

Response: Here is the JSON response:

---

```
{
"title": "The Detection of Microplastics in Patients Undergoing Cardiac
Surgery",
"Figure found": "YES",
"Figure": "Figure 1",
"Material": "Various polymers, including PVC, PET, and PMMA"
} {
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.978264808654785 seconds
Input Tokens: 10276
Output Tokens: 85
```

```
-----
PDF File: onep.pdf
Response: {
"title": "Environment International 163 (2022) 107199",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 1.1482188701629639 seconds
Input Tokens: 1452
Output Tokens: 33
```

```
-----
PDF File: BMS.pdf
Response: {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3a",
"material": "La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite"
}
Processing Time: 2.2888214588165283 seconds
Input Tokens: 4845
Output Tokens: 60
```

```
-----
PDF File: Tuneable_Giant_Magnetocaloric_Effect_in_MnFe2PSi_M.pdf
Response: Here is the JSON response:

{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Mn,Fe)2(P,Si)-based materials"
} {
"title": "The magnetocaloric effect and magnetic refrigeration near room
temperature: Materials and models",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "MnFe(P,Si,B) and MnFe(Ni,P,Si,B)"
}
Processing Time: 4.84161376953125 seconds
Input Tokens: 11212
Output Tokens: 110
```

```
-----
PDF File: 122405_1_online.pdf
Response: Here is the JSON response:
```

---

```
{
"title": "Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "MnNiGe-CoNiGe"
}
```

```
Processing Time: 2.79042649269104 seconds
Input Tokens: 7672
Output Tokens: 56
```

```
-----
PDF File: PhysRevB.93.184403.pdf
Response: The answer is:
```

```
{"title": "Realization of two-dimensional ferromagnetism with giant coercivity
in ultrathin  $\beta$ -Ni(OH)2 layers grown on a MoS2 surface", "Figure found": "Yes",
"Figure": "Figure 1", "Material": " $\beta$ -Ni(OH)2"} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 8",
"material": "MoS2/Ni(OH)2 composite"
}
```

```
Processing Time: 5.04149866104126 seconds
Input Tokens: 12780
Output Tokens: 111
```

```
-----
PDF File: ArticleSupra2020.pdf
Response: The JSON response is:
```

```
{
"title": "Structural Analysis and Theoretical Investigations of the
Magnetocaloric Effect for La0.7Ba0.15Ag0.15MnO3 Manganite Prepared Using Sol-
Gel Route",
"Figure found": "YES",
"Figure": "Figure 2d",
"material": "La0.7Ba0.15Ag0.15MnO3"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Fig. 7",
"material": "La0.7Ba0.15Ag0.15MnO3"
}
```

```
Processing Time: 8.28354811668396 seconds
Input Tokens: 10009
Output Tokens: 137
```

```
-----
PDF File: Krishna_Murthy_2015_J._Phys._D _Appl._Phys._48_355001.pdf
Response: Here is the JSON output:
```

```
{
"title": "Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic
insulators",
"Figure found": "YES",
"Figure": "Figure 1(a) and (b)",
"material": "Gd2NiMnO6 and Gd2CoMnO6"
}
```

```
Processing Time: 3.418886661529541 seconds
Input Tokens: 6793
Output Tokens: 83
```

---

PDF File: Zinn - 2023 - Zweite Moderne und Risikogesellschaft.pdf  
Response: The JSON output is:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 3",  
"material": "none"  
} Here is the JSON output:
```

```
{  
"title": "The title of the paper",  
"Figure found": "NO",  
"Figure": "",  
"material": ""  
}  
Processing Time: 3.754127025604248 seconds  
Input Tokens: 12942  
Output Tokens: 71
```

-----  
PDF File: Albert\_und\_Kleinman\_2011\_Bringing\_Pierre\_Bourdieu\_to\_Science\_and.pdf  
Response: Here is the JSON response:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 1: The Scientific Field: A Space of Struggle for Power",  
"material": "The paper discusses the application of Pierre Bourdieu's concepts  
to the study of science and technology studies (STS)."  
} {"title": "Bourdieu: Critical perspectives", "Figure  
found": "NO", "Figure": "", "material": ""}  
Processing Time: 4.151360511779785 seconds  
Input Tokens: 8648  
Output Tokens: 98
```

-----  
PDF File:  
D'Angelo\_und\_Meccariello\_2021\_Microplastics\_A\_Threat\_for\_Male\_Fertility.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure  
1", "material": "microplastics"}
```

The paper discusses the effects of microplastics on male fertility and reproduction. The authors review the current state of knowledge on the topic and present some of their own research findings. They conclude that microplastics can have negative effects on sperm quality and fertility, and that further research is needed to fully understand the risks associated with microplastic exposure. Here is the JSON response:

```
{  
"title": "Q. Molecular mechanisms of PFOA-induced toxicity in animals and  
humans: Implications for health risks.",  
"Figure found": "NO",  
"Figure": "",  
"material": ""  
}  
Processing Time: 6.565014600753784 seconds  
Input Tokens: 13286  
Output Tokens: 155
```

---

-----  
PDF File:

Paul\_et\_al\_2022\_Institutionalised\_ignorance\_in\_policy\_and\_regulati.pdf

Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the concept of institutionalized ignorance and its role in governance, policy, and regulation.

According to the document, institutionalized ignorance refers to the normalization of not knowing or ignoring certain information, which can have significant consequences for accountability and decision-making. The authors argue that ignorance is not just a lack of knowledge, but a deliberate strategy used by institutions to maintain power and control.

The document also discusses the concept of agnotology, which refers to the study of ignorance and non-knowledge. The authors argue that agnotology is an important field of study that can help us understand how ignorance is produced and maintained, and how it can be challenged and overcome.

The document does not provide a specific material that is the subject of the M vs. T plot, but it does discuss the importance of understanding the role of ignorance in various fields, including science, policy, and regulation.

Here is the JSON output:

```
{
  "title": "The title of the paper",
  "Figure found": "NO",
  "Figure": "N/A",
  "Material": "N/A"
} {
  "title": "Routledge International Handbook of Ignorance Studies",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
Processing Time: 9.9422607421875 seconds
Input Tokens: 9626
Output Tokens: 274
```

-----  
PDF File: 1-s2.0-S0304885312009754-main.pdf

Response: Here is the JSON output:

```
{
  "title": "Phase diagram, ferromagnetic martensitic transformation and
magnetoresponse properties of Fe-doped MnCoGe alloys",
  "Figure found": "YES",
  "Figure": "Figure 2 (a) and (b)",
  "material": "MnCoGe alloys"
}
Processing Time: 2.9770314693450928 seconds
Input Tokens: 6637
Output Tokens: 66
```

-----  
PDF File: Kuhlicke - 2023 - Resilienz und Risiko.pdf

Response: The JSON output is:

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
```

```
"Figure": "Figure 2.1",
"material": "none"
} {
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.5361948013305664 seconds
Input Tokens: 12243
Output Tokens: 67
```

-----  
PDF File: Wajcman - 2010 - Feminist theories of technology.pdf  
Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the relationship between gender and technology, and how technology is shaped by and shapes gender relations.

The document is a scholarly article that explores the intersection of feminist theory and science and technology studies (STS). The author, Judy Wajcman, argues that technology is not a neutral or objective field, but rather a social construct that is shaped by and shapes gender relations.

The article discusses the ways in which technology has been used to reinforce gender hierarchies and the ways in which women have been excluded from the development and use of technology. Wajcman argues that feminist theory can help to challenge these gendered power relations and create a more inclusive and equitable technology.

The article does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot, but it does contain several tables, figures, and diagrams that illustrate the author's arguments about the relationship between gender and technology.

Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "N/A",
"Material": "N/A"
} {"title":"The title of the paper":null,"Figure found":"NO","Figure":null,"material":null}
Processing Time: 9.79239010810852 seconds
Input Tokens: 8825
Output Tokens: 273
```

-----  
PDF File: Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C.pdf  
Response: Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 5a",
"Material": "Cyanido-bridged metal assemblies containing both Fe and Co ions"
} Here is the JSON output:
...
{
```

```
"title": "Inorganics 2017,5, 63",
"Figure found": "YES",
"Figure": "Figure 6",
"material": "K4[Nb(CN)8]2H2O, (Fe0.92Co0.08)2[Nb(CN)8](4-pyridinealdoxime)
83H2O, (Fe0.71Co0.29)2[Nb(CN)8](4-pyridinealdoxime) 83H2O, (Fe
0.50Co0.50)2[Nb(CN)8](4-pyridinealdoxime) 83H2O, (Fe 0.21Co0.79)2[Nb(CN)8](4-
pyridinealdoxime) 83H2O, Co 2[Nb(CN)8](4-pyridinealdoxime) 83H2O"
}
...
```

Processing Time: 10.677539587020874 seconds

Input Tokens: 14082

Output Tokens: 288

-----  
PDF File: JAP115123908.pdf

Response: The magnetization vs. temperature (M vs. T) plot is found in this document. The title of the paper is "Magnetic properties of mixed spinel BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites".

The answer is: YES, a magnetization vs. temperature (M vs. T) plot is found in this document.

The M vs. T plot is shown in Fig. 4, which displays the field-cooled (FC) magnetization ( $M_{FC}$ ) as a function of temperature for all samples down to 5 K. The plot shows that the ferrimagnetic transition occurs at about 848 K for pure NFO, and the value of  $T_C$  decreases from 848 K to 836 K for the solid solution with an increase in BTO phase.

The material is BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites. Here is the JSON output:

```
...
{
  "title": "Synthesis of (100-x)BTO-(x)NFO solid solution at high
temperature",
  "Figure found": "YES",
  "Figure": "Figure 5",
  "material": "NFO and BTO"
}
...
```

Processing Time: 24.462421894073486 seconds

Input Tokens: 14291

Output Tokens: 243

-----  
PDF File: Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mssbauer Study of Perovskite LaFeO<sub>3</sub> and LaFe<sub>0.5</sub>Cr<sub>0.5</sub>O<sub>3</sub>.pdf

Response: The answer is:

```
{"title": "Magnetic Properties and Mössbauer Study of Perovskite LaFeO3 and
LaFe0.5Cr0.5O3",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "LaFeO3 and LaFe0.5Cr0.5O3"} {
"title": "Phys. Status Solidi RRL 2022,16, 2200023",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "none"
}
```

Processing Time: 4.89314603805542 seconds

Input Tokens: 8415

Output Tokens: 124

---

PDF File: HCMO.pdf  
Response: YES

Figure: Figure 1(a)

Material: Ho<sub>2</sub>CoMnO<sub>6</sub>

The magnetization vs. temperature (M-T) plot under various magnetic fields is shown in Figure 3(a). The plot exhibits a valley-like feature in the low-temperature region, indicating a negative magnetization. The negative magnetization is attributed to the 3d-4f negative exchange interaction between the Ho<sup>3+</sup> ions and the FM mesh of Co<sup>2+</sup>/Mn<sup>4+</sup> sublattices.

The magnetocaloric effect (MCE) is studied by measuring the temperature dependence of magnetization under various magnetic fields. The MCE is found to be significant, with a maximum value of 13.4 J/kg K at a low temperature and a magnetic field of 70 kOe. The MCE is attributed to the change in the magnetic entropy of the system upon the application of a magnetic field.

The magnetocaloric parameters, such as the relative cooling power (RCP) and the adiabatic temperature change ( $\Delta T_{ad}$ ), are also calculated. The RCP is found to be 12.5 J/kg K and the  $\Delta T_{ad}$  is found to be 4.2 K at a magnetic field of 70 kOe Here is the JSON response:

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 3(d)",
  "material": "Ho2CoMnO6",
  "material description": "The title compound is a rare-earth-based double
perovskite compound, Ho2CoMnO6, which exhibits a complex magnetic behavior."
} Here is the JSON output:
...
{
  "title": "Structural, magnetic, and magnetocaloric properties of Ho2CoMnO6",
  "Figure found": "YES",
  "Figure": "Figure 7(b)",
  "material": "Ho2CoMnO6"
}
...
```

Processing Time: 15.713869571685791 seconds  
Input Tokens: 21154  
Output Tokens: 405

---

PDF File: 1-s2.0-S0109564113000067-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure
2", "material": "Polymer-infiltrated-ceramic-network material (PICN)"}
```

The paper discusses the mechanical properties of a novel material called Polymer-infiltrated-ceramic-network material (PICN), which is a two-phase composite material composed of a ceramic network and a polymer network. The authors investigate the effects of varying ceramic precursor densities on the mechanical properties of the PICN, including flexural strength, elastic modulus, strain at failure, and hardness.

---

The paper presents the results of experiments on four different PICNs with varying ceramic precursor densities, as well as a dense ceramic and a pure polymer. The results show that the mechanical properties of the PICN are influenced by the ceramic precursor density, with lower ceramic fractions resulting in lower elastic modulus and hardness, but higher flexural strength and strain at failure.

The authors also present SEM images of the indented areas of the materials, which show that the PICN exhibits a unique microstructure with a continuous polymer network and a ceramic network. The results suggest that the PICN has potential for use in dental applications, as it exhibits mechanical properties similar to those of natural teeth. {

```
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "PICN (polymer-infiltrated ceramic network)"
}
```

Processing Time: 11.327147006988525 seconds

Input Tokens: 11623

Output Tokens: 299

-----  
PDF File: zno-serie5.pdf

Response: Here is the JSON output:

...

```
{
  "title": "Structural and magnetic characterization of (Zn 1-x-yFexCoy)O",
  "Figure found": "YES",
  "Figure": "Figure 2",
  "material": "ZnO, Fe-doped ZnO, Co-doped ZnO"
}
```

...

Processing Time: 3.698740243911743 seconds

Input Tokens: 7592

Output Tokens: 75

-----  
PDF File: 1-s2.0-S1359645413004503-main.pdf

Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 3", "material": "CoCrFeMnNi high-entropy alloy"} Here is the JSON output:
```

...

```
{
  "title": "An equiatomic CoCrFeMnNi high-entropy alloy",
  "Figure found": "YES",
  "Figure": "Figure 12",
  "material": "CoCrFeMnNi high-entropy alloy"
}
```

...

Processing Time: 5.287784576416016 seconds

Input Tokens: 14405

Output Tokens: 110

-----  
PDF File: 1-s2.0-S1002007116300995-main.pdf

Response: The answer is:

---

```
{"title": "Structural effects on the magnetic hyperthermia properties of iron oxide-based NPs", "Figure found": "YES", "Figure": "Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5", "Material": "Iron oxide nanoparticles"} The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 5", "material": "iron oxide-based magnetic hyperthermia agents"}
```

The paper discusses the use of iron oxide-based magnetic hyperthermia agents for cancer treatment. The authors review the current state of the art in this field and discuss the potential benefits and challenges of using these agents. They also present some of their own research on the topic, including the development of new iron oxide-based magnetic hyperthermia agents and the testing of their effectiveness in treating cancer in animal models. {

```
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 103",  
"material": "Fe3O4"  
}
```

Processing Time: 9.752267360687256 seconds

Input Tokens: 17383

Output Tokens: 233

-----  
PDF File: app135z2p11.pdf

Response: {

```
"title": "Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe70.57xCoxMn10Mo5B15(0x0.5) Alloys at Low Magnetizing Field",
```

```
"Figure found": "YES",
```

```
"Figure": "Figure 1 and Figure 2",
```

```
"material": "amorphous Fe70.57xCoxMn10Mo5B15(0x0.5) alloys"
```

Processing Time: 3.103688955307007 seconds

Input Tokens: 3375

Output Tokens: 97

-----  
PDF File: 035220\_1\_online.pdf

Response: Here is the JSON output:

...

```
{  
  "title": "Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity",
```

```
  "Figure found": "YES",
```

```
  "Figure": "Figure 4",
```

```
  "material": "Gd-Fe-Al"
```

```
}
```

Processing Time: 3.567056894302368 seconds

Input Tokens: 7390

Output Tokens: 80

-----  
PDF File: PhysRevB.81.214406.pdf

Response: The answer is:

```
{"title": "Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape memory alloys",
```

```
"Figure found": "Yes",
```

```
"Figure": "Figure 1",
```

```
"Material": "Ni-Mn-Sn ferromagnetic shape memory alloy"} {
```

```
"title": "PECULIARITIES OF THE MAGNETOCALORIC EFFECT PHYSICAL REVIEW B 81, 214406 /H208492010 /H20850",
```

---

```
"Figure found": "YES",
"Figure": "Figure 1",
"material": "MnFePd"
}
Processing Time: 6.603277921676636 seconds
Input Tokens: 9218
Output Tokens: 131
```

```
-----
PDF File: Gottschall2020_JAP_Advanced characterization of multicaloric
materials in pulsed magnetic fields.pdf
Response: Here is the JSON output:
```

```
{
"figure": "Figure 5",
"material": "Ni-Mn-Ti-Co"
} {
"figure": "FIG. 6",
"material": "Heusler shape memory alloys"
}
Processing Time: 3.8675827980041504 seconds
Input Tokens: 10443
Output Tokens: 83
```

```
-----
PDF File: Wehling_2023_Nichtwissen_-_ein_ungewöhnlicher_Schlüsselbegriff_.pdf
Response: The answer is:
```

```
{"figure": "Figure 3.1", "material": "FCKW (Fluor-Chlor-Kohlenwasserstoff)"}
```

The paper discusses the concept of "Nichtwissen" (ignorance) in the context of environmental sociology and its relevance to understanding environmental problems. The author argues that ignorance is not just a lack of knowledge, but a social construct that is shaped by various factors, including the way we perceive and define risk, uncertainty, and the limits of our knowledge.

The paper highlights the importance of considering ignorance in the context of environmental problems, such as climate change, and argues that it is not just a matter of individual ignorance, but rather a systemic issue that is rooted in the way we organize our societies and economies.

The author also discusses the concept of "wissenschaftliches Nichtwissen" (scientific ignorance), which refers to the ignorance that is generated by scientific research itself. This type of ignorance is not just a lack of knowledge, but rather a result of the limitations and biases of scientific inquiry.

The paper concludes by arguing that ignorance is a key concept in understanding environmental problems and that it is essential to recognize its Here is the JSON output:

```
...
{
"figure": "Figure 1",
```

```
"material": "none"
}
```

The paper discusses the concept of ignorance and its relevance in the context of environmental issues, particularly in the context of climate change and sustainability. The author, Peter Wehling, argues that ignorance is not just a lack of knowledge, but a complex and multifaceted phenomenon that is deeply embedded in societal and cultural contexts. He also highlights the importance of acknowledging and addressing ignorance in order to develop more effective strategies for addressing environmental challenges.

The paper is a collection of essays that explore the concept of ignorance from various perspectives, including sociology, philosophy, and environmental studies. The authors discuss the ways in which ignorance is produced and reproduced in different contexts, and the ways in which it can be addressed and overcome.

The paper is a valuable contribution to the field of environmental studies, as it highlights the importance of acknowledging and addressing ignorance in order to develop more effective strategies for addressing environmental challenges. It also provides a nuanced and complex understanding of the concept of ignorance, and its relevance to environmental issues.

Processing Time: 17.833888053894043 seconds

Input Tokens: 14995

Output Tokens: 501

-----  
PDF File: Nyadjro\_et\_al\_2023\_The\_NOAA\_NCEI\_marine\_microplastics\_database.pdf  
Response: The answer is:

```
{"title":"The title of the paper","Figure found":"YES","Figure number":"1","Material":"marine microplastics"} Here is the JSON response:
```

```
...
```

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "Various materials"
}
```

```
... {"title":"Fellowship from the Gulf Research Program of the US National Academies of Sciences, Engineering, and Medicine","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 5.906141519546509 seconds

Input Tokens: 15859

Output Tokens: 115

-----  
PDF File:  
Benningstad\_und\_Kunst\_2020\_Dissociating\_meat\_from\_its\_animal\_origins.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "meat"} Here is the JSON output:
```

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 4.6",
  "material": "meat"
}
```

---

Processing Time: 4.217400550842285 seconds  
Input Tokens: 13879  
Output Tokens: 78

-----  
PDF File: Gottschall2018\_A multicaloric cycle that exploits thermal hysteresis\_NatureMat.pdf  
Response: The answer is:

```
{"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 4",  
"material": "Ni-Mn-In Heusler alloy"} {  
"title": "Large reversible magnetocaloric effect in Ni-Mn-In-Co",  
"Figure found": "YES",  
"Figure": "Figure 1",  
"material": "Ni50.0Mn35.5In14.5"  
}
```

Processing Time: 6.516801595687866 seconds  
Input Tokens: 8750  
Output Tokens: 103

-----  
PDF File:  
Leslie\_et\_al\_2022\_Discovery\_and\_quantification\_of\_plastic\_particle.pdf  
Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the analysis of plastic particles in human blood and the development of a method to quantify the mass of five high production volume polymers applied in plastic materials: poly(methyl methacrylate) (PMMA), polypropylene (PP), materials containing polymerized styrene (PS), polyethylene (PE), and polyethylene terephthalate (PET).

The method uses double shot pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) to analyze the samples. The authors report the results of a recovery experiment, which showed that the method can recover the spiked polymers from blood samples with recoveries ranging from 68% to 114%. They also discuss the control of background contamination during sampling, sample preparation, and analysis.

The authors conclude that the method can detect and quantify the four polymers in blood samples with sufficient sensitivity and precision. The measured concentration data in blood were corrected for the average procedural blank for each analyte.

Here is the JSON response:

```
{  
"title": "The title of the paper",  
"Figure found": "NO",  
"Figure": "",  
"material": " The answer is:"  
  
{"title":"The title of the paper","Figure found":"YES","Figure number":"3.1","Material":"Human blood"} {"title":"Impact of microplastics and nanoplastics on human health","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 12.60486888885498 seconds  
Input Tokens: 15870  
Output Tokens: 315

---

You are going to be given an example of a document. Please check this document to see if it contains a figure that depicts a magnetization vs. temperature (M vs T) PLOT. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS THE M VS T FIGURE if it is AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE, ANSWER ONLY WITH a JSON return the following Json {"title":"The title of the paper";"Figure found":"If a magnetization vs. temperature PLOT is found YES or NO";"Figure":"which figure is the MAGNETIZATION VS. TEMPERATURE PLOT";"material":"what is the material that is the plot"}

-----  
PDF File:  
Bommarito\_et\_al\_2023\_Temporal\_trends\_and\_predictors\_of\_gestational.pdf  
Response: The answer is:

```
{"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 2",  
"material": "human biomarkers of exposure to organophosphate ester flame  
retardants and plasticizers in urine from pregnant people"} The JSON output  
is:
```

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Fig. 2",  
  "material": "human urine samples"  
} {  
  "title": "Wang et al. (2023)",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

Processing Time: 7.283336162567139 seconds  
Input Tokens: 18882  
Output Tokens: 136

-----  
PDF File:  
Yang\_et\_al\_2023\_Detection\_of\_Various\_Microplastics\_in\_Patients\_Und.pdf  
Response: Here is the JSON response:

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 1",  
  "material": "Various polymers, including PVC, PET, and PMMA"  
} {  
  "title": "The title of the paper",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

Processing Time: 3.720438241958618 seconds  
Input Tokens: 10276  
Output Tokens: 77

-----  
PDF File: onep.pdf  
Response: {  
"title": "Environment International 163 (2022) 107199",

---

```
"Figure found": "NO",
"Figure": "",
"material": ""
}
```

```
Processing Time: 1.147451400756836 seconds
Input Tokens: 1452
Output Tokens: 33
```

```
-----
PDF File: BMS.pdf
```

```
Response: {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3a",
"material": "La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite"
}
```

```
Processing Time: 2.292752981185913 seconds
Input Tokens: 4845
Output Tokens: 60
```

```
-----
PDF File: Tuneable_Giant_Magnetocaloric_Effect_in_MnFe2PSi_M.pdf
```

```
Response: Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Mn,Fe)2(P,Si)-based materials"
} {
"title": "The magnetocaloric effect and magnetic refrigeration near room
temperature: Materials and models",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "MnFe(P,Si,B) and MnFe(Ni,P,Si,B)"
}
```

```
Processing Time: 4.847777605056763 seconds
Input Tokens: 11212
Output Tokens: 110
```

```
-----
PDF File: 122405_1_online.pdf
```

```
Response: Here is the JSON response:
```

```
{
"title": "Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "MnNiGe-CoNiGe"
}
```

```
Processing Time: 2.7820096015930176 seconds
Input Tokens: 7672
Output Tokens: 56
```

```
-----
PDF File: PhysRevB.93.184403.pdf
```

```
Response: The answer is:
```

```
{"title": "Realization of two-dimensional ferromagnetism with giant coercivity
in ultrathin  $\beta$ -Ni(OH)2 layers grown on a MoS2 surface",
"Figure found": "YES",
"Figure": "Figure 4",
"Material": " $\beta$ -Ni(OH)2"} {
```

---

```
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 8",
"material": "MoS2/Ni(OH)2"
}
Processing Time: 5.184289455413818 seconds
Input Tokens: 12780
Output Tokens: 113
```

```
-----
PDF File: ArticleSupra2020.pdf
Response: Here is the JSON response:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 2d",
"material": "La0.7Ba0.15Ag0.15MnO3 manganite"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Fig. 7",
"material": "La0.7Ba0.15Ag0.15MnO3"
}
Processing Time: 7.593746185302734 seconds
Input Tokens: 10009
Output Tokens: 107
```

```
-----
PDF File: Krishna_Murthy_2015_J._Phys._D _Appl._Phys._48_355001.pdf
Response: Here is the JSON output:
```

```
{
"title": "Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic
insulators",
"Figure found": "YES",
"Figure": "Figure 1(a) and (b)",
"material": "Gd2NiMnO6 and Gd2CoMnO6"
}
Processing Time: 4.6754374504089355 seconds
Input Tokens: 6793
Output Tokens: 83
```

```
-----
PDF File: Zinn - 2023 - Zweite Moderne und Risikogesellschaft.pdf
Response: The answer is:
```

```
{"title":"The title of the paper","Figure found":"YES","Figure":"Figure
3","material":"none"} Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.459108591079712 seconds
Input Tokens: 12942
Output Tokens: 62
```

```
-----
PDF File: Albert_und_Kleinman_2011_Bringing_Pierre_Bourdieu_to_Science_and.pdf
Response: The answer is:
```

---

```
{"title":"The title of the paper is 'Bringing Pierre Bourdieu to Science and Technology Studies',"Figure found":"YES","Figure":"Figure 1","material":"The paper discusses the application of Pierre Bourdieu's concepts to the field of Science and Technology Studies (STS). It argues that Bourdieu's work can provide insights into the operation of science in the contemporary world and can be used to analyze the social dynamics of the scientific field."}
```

```
{"title":"Bourdieu: Critical perspectives","Figure found":"NO","Figure":"","material":""}
```

```
Processing Time: 4.860419750213623 seconds
```

```
Input Tokens: 8648
```

```
Output Tokens: 120
```

```
-----  
PDF File:
```

```
D'Angelo_und_Meccariello_2021_Microplastics_A_Threat_for_Male_Fertility.pdf
```

```
Response: The answer is:
```

```
{"title":"The title of the paper","Figure found":"YES","Figure number":"Figure 1","Material":"Not mentioned"} Here is the JSON output:
```

```
...
```

```
{  
  "title": "Ma, L.Q. Molecular mechanisms of PFOA-induced toxicity in animals and humans: Implications for health risks.",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

```
Processing Time: 4.424508094787598 seconds
```

```
Input Tokens: 13286
```

```
Output Tokens: 90
```

```
-----  
PDF File:
```

```
Paul_et_al_2022_Institutionalised_ignorance_in_policy_and_regulati.pdf
```

```
Response: The answer is:
```

```
{"title":"The title of the paper";"Figure found":"YES";"Figure":"Figure 1";"material":"unknown"} {  
"title": "Routledge International Handbook of Ignorance Studies",  
"Figure found": "NO",  
"Figure": "",  
"material": ""  
}
```

```
Processing Time: 3.0628509521484375 seconds
```

```
Input Tokens: 9626
```

```
Output Tokens: 61
```

```
-----  
PDF File: 1-s2.0-S0304885312009754-main.pdf
```

```
Response: Here is the JSON output:
```

```
{  
"title": "Phase diagram, ferromagnetic martensitic transformation and magnetoresponse properties of Fe-doped MnCoGe alloys",  
"Figure found": "YES",  
"Figure": "Figure 2",  
"material": "MnCoGe alloys"  
}
```

```
Processing Time: 2.810117244720459 seconds
```

```
Input Tokens: 6637
```

---

Output Tokens: 61

-----  
PDF File: Kuhlicke - 2023 - Resilienz und Risiko.pdf  
Response: The JSON output is:

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 2.1",  
  "material": "none"  
}  
{  
  "title": "The title of the paper",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

Processing Time: 3.552651882171631 seconds  
Input Tokens: 12243  
Output Tokens: 67

-----  
PDF File: Wajcman - 2010 - Feminist theories of technology.pdf  
Response: The paper "Feminist Theories of Technology" by Judy Wajcman does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, the paper does discuss the relationship between gender and technology, and how technology is shaped by and shapes gender relations.

The paper argues that feminist theories of technology have come a long way over the last quarter of a century, and that the intellectual exploration at the intersections of feminist scholarship and science and technology studies (STS) has enriched both fields immeasurably.

The paper highlights the importance of understanding the mutual shaping of gender and technology, and how the materiality of technology affords or inhibits the doing of particular gender power relations. It also emphasizes the need to involve women throughout the processes and practices of technological innovation, and to recognize the constitutive power of tools, techniques, and objects to materialize social, political, and economic arrangements.

The paper does not mention the concept of "material-semiotic practice" or "technoscience" explicitly, but it does discuss the social constructivist approach to technology, which views technology as a sociotechnical product that is shaped by social and cultural factors.

The paper does not mention the authors Manuel Castells, Nicholas Negroponte  
{"title":"The title of the paper":null,"Figure found":"NO","Figure":null,"material":null}  
Processing Time: 10.056691646575928 seconds  
Input Tokens: 8825  
Output Tokens: 282

-----  
PDF File: Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C.pdf  
Response: Here is the JSON response:

```
...  
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 5",
```

```
"Material": "Cyanido-bridged metal assemblies containing both Fe and Co ions"
}
```

```
``` Here is the JSON output:
```

```
```
{
  "title": "Inorganics 2017,5, 63",
  "Figure found": "YES",
  "Figure": "Figure 6",
  "material": "K4[Nb(CN)8]2H2O, (Fe0.21Co0.79)2[Nb(CN)8](4-pyridinealdoxime)8zH2O"
}
```
```

```
Processing Time: 6.898503065109253 seconds
```

```
Input Tokens: 14082
```

```
Output Tokens: 161
```

```
-----
PDF File: JAP115123908.pdf
```

```
Response: The magnetization vs. temperature (M vs. T) plot is found in the article. Therefore, the answer is:
```

```
**YES**
```

```
The M vs. T plot is Figure 4 in the article.
```

```
**Material:** BaTiO3-NiFe2O4 composite
```

```
**Figure:** Figure 4
```

```
**Description:** The figure shows the field-cooled magnetization (M) as a function of temperature (T) for the BaTiO3-NiFe2O4 composite. The magnetization decreases with increasing temperature, and the Curie temperature (Tc) is observed to decrease with increasing BaTiO3 content. Here is the JSON output:
```

```
```
{
  "title": "Synthesis of (100-x)BTO-(x)NFO solid solution at high temperature",
  "Figure found": "YES",
  "Figure": "Figure 5",
  "material": "NFO and BTO"
}
```
```

```
Processing Time: 22.981261491775513 seconds
```

```
Input Tokens: 14291
```

```
Output Tokens: 196
```

```
-----
PDF File: Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3.pdf
```

```
Response: The answer is:
```

```
{"title": "Magnetic Properties and Mössbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3", "Figure found": "YES", "Figure": "Figure 1", "material": "LaFeO3 and LaFe0.5Cr0.5O3"} {
"title": "Phys. Status Solidi RRL 2022,16, 2200023",
"Figure found": "YES",
"Figure": "Figure 1",
```

---

```
"material": "none"
}
```

```
Processing Time: 4.796543598175049 seconds
Input Tokens: 8415
Output Tokens: 121
```

```
-----
PDF File: HCMO.pdf
Response: The answer is:
```

```
{"title": "Role of 3d-4f exchange interaction and local anti-site defects in
the magnetic and magnetocaloric properties of double perovskite Ho2CoMnO6
compound", "Figure found": "YES", "Figure": "Figure 2", "Material":
"Ho2CoMnO6"} Here is the JSON response:
```

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 3(d)",
  "material": "Ho2CoMnO6",
  "parameters": {
    "jΔSjmax": 13.44 J/kg K,
    "RCP": 272 J/kg,
    "TEC": 6.5 (3.9) J/kg K,
    "NRC": 1.5 (2.1)
  }
} {
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 7(b)",
  "material": "Ho2CoMnO6"
}
```

```
Processing Time: 9.795434474945068 seconds
Input Tokens: 21154
Output Tokens: 218
```

```
-----
PDF File: 1-s2.0-S0109564113000067-main.pdf
Response: The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure
2", "material": "Polymer-infiltrated-ceramic-network material (PICN)"}
```

The paper discusses the mechanical properties of a novel material called Polymer-infiltrated-ceramic-network material (PICN), which is a two-phase composite material composed of a ceramic network and a polymer network. The authors investigate the effects of varying ceramic precursor densities on the mechanical properties of the PICN, including flexural strength, elastic modulus, strain at failure, and hardness.

The paper presents a comprehensive study of the mechanical properties of PICNs, including the effects of ceramic precursor density on the material's properties. The results show that the flexural strength of the PICN increases with decreasing ceramic precursor density, while the elastic modulus and hardness increase with increasing ceramic precursor density. The strain at failure of the PICN also decreases with increasing ceramic precursor density.

The authors also present scanning electron microscopy (SEM) images of the indented areas of the PICN, which show that the material exhibits a unique microstructure with a continuous polymer network and a ceramic network. The

---

SEM images also show that the material exhibits a high degree of toughness and resistance to {

```
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
```

```
Processing Time: 10.908538818359375 seconds
Input Tokens: 11623
Output Tokens: 284
```

-----  
PDF File: zno-serie5.pdf  
Response: Here is the JSON output:

```
...
{
  "title": "Structural and magnetic characterization of (Zn 1-x-yFexCoy)O",
  "Figure found": "YES",
  "Figure": "Figure 2",
  "material": "ZnO, Fe-doped ZnO, Co-doped ZnO"
}
...
```

```
Processing Time: 3.6972219944000244 seconds
Input Tokens: 7592
Output Tokens: 75
```

-----  
PDF File: 1-s2.0-S1359645413004503-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 9", "material": "CoCrFeMnNi high-entropy alloy"} Here is the JSON output:
```

```
...
{
  "title": "An equiatomic CoCrFeMnNi high-entropy alloy was produced by arc melting, drop casting and rolling, after which it was recrystallized to produce single-phase, fcc-structured, equiaxed microstructures with three different grain sizes, 4.4, 50 and 155 nm.",
  "Figure found": "YES",
  "Figure": "Figure 11",
  "material": "CoCrFeMnNi high-entropy alloy"
}
...
```

```
Processing Time: 6.863595485687256 seconds
Input Tokens: 14405
Output Tokens: 162
```

-----  
PDF File: 1-s2.0-S1002007116300995-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "Iron oxide nanoparticles"} The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 5", "material": "iron oxide-based magnetic hyperthermia agents"}
```

The paper discusses the use of iron oxide-based magnetic hyperthermia agents for cancer treatment. The authors review the current state of the art in this field and discuss the potential benefits and challenges of using these agents.

---

They also present some of their own research on the topic, including the development of new iron oxide-based magnetic hyperthermia agents and their testing in animal models.

The figure that is referred to is Figure 5, which shows the results of the authors' experiments on the heating efficiency of different iron oxide-based magnetic hyperthermia agents. The figure shows that the agents with a higher concentration of iron oxide particles have a higher heating efficiency.

The material that is being discussed is iron oxide-based magnetic hyperthermia agents, which are used to generate heat in the body to treat cancer. {

```
"title": "The title of the paper",
```

```
"Figure found": "YES",
```

```
"Figure": "Figure 103",
```

```
"material": "Fe3O4"
```

```
}
```

```
Processing Time: 11.12071442604065 seconds
```

```
Input Tokens: 17383
```

```
Output Tokens: 275
```

```
-----  
PDF File: app135z2p11.pdf
```

```
Response: {
```

```
"title": "Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe7069xCoxMn10Mo5B15(0x0.5) Alloys at Low Magnetizing Field",
```

```
"Figure found": "YES",
```

```
"Figure": "Figure 1 and Figure 2",
```

```
"material": "amorphous Fe7069xCoxMn10Mo5B15(0x0.5) alloys"
```

```
Processing Time: 3.0970208644866943 seconds
```

```
Input Tokens: 3375
```

```
Output Tokens: 97
```

```
-----  
PDF File: 035220_1_online.pdf
```

```
Response: Here is the JSON output:
```

```
{
```

```
"title": "Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity",
```

```
"Figure found": "YES",
```

```
"Figure": "Figure 4",
```

```
"material": "Gd-Fe-Al"
```

```
}
```

```
Processing Time: 3.3428659439086914 seconds
```

```
Input Tokens: 7390
```

```
Output Tokens: 73
```

```
-----  
PDF File: PhysRevB.81.214406.pdf
```

```
Response: The answer is:
```

```
{"title": "Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape memory alloys", "Figure found": "YES", "Figure": "Figure 1", "Material": "Ni-Mn-Sn ferromagnetic shape memory alloy"} {
```

```
"title": "Peculiarities of the Magnetocaloric Effect",
```

```
"Figure found": "YES",
```

```
"Figure": "Figure 1",
```

```
"material": "MnFePGe"
```

```
}
```

```
Processing Time: 5.836395263671875 seconds
```

```
Input Tokens: 9218
```

---

Output Tokens: 105

-----  
PDF File: Gottschall2020\_JAP\_Advanced characterization of multicaloric materials in pulsed magnetic fields.pdf  
Response: Here is the JSON output:

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 5",
  "material": "Ni-Mn-Ti-Co Heusler alloy"
} {
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "FIG. 6",
  "material": "Heusler shape memory alloys"
}
Processing Time: 3.999363660812378 seconds
Input Tokens: 10443
Output Tokens: 87
```

-----  
PDF File: Wehling\_2023\_Nichtwissen\_-\_ein\_ungewöhnlicher\_Schlüsselbegriff\_.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "3.1",
"material": "none"} Here is the JSON output:
```

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "none"
}
...
```

The paper discusses the concept of ignorance and its relevance in the context of environmental issues, particularly in the field of climate change. The author, Peter Wehling, argues that ignorance is not just a lack of knowledge, but a complex and multifaceted phenomenon that can have significant consequences for our understanding of the world and our ability to make informed decisions.

The paper explores the concept of ignorance in various contexts, including the history of climate change research, the role of ignorance in the development of scientific knowledge, and the implications of ignorance for environmental policy and decision-making. The author also discusses the importance of acknowledging and addressing ignorance in order to promote more effective and sustainable solutions to environmental problems.

The paper is written in a clear and accessible style, making it suitable for readers with a general interest in environmental issues and the social sciences.

Processing Time: 9.742598533630371 seconds  
Input Tokens: 14995  
Output Tokens: 249

-----  
PDF File: Nyadjro\_et\_al\_2023\_The\_NOAA\_NCEI\_marine\_microplastics\_database.pdf  
Response: \*\*JSON Response:\*\*

```

...
{
  "title": "The NOAA NCEI global marine microplastics database",
  "Figure found": "YES",
  "Figure": "Figure 1, Figure 2, Figure 3, Figure 4",
  "Material": "Marine microplastics"
}
``` Here is the JSON response:
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "NO"
}
``` {"title":"Fellowship from the Gulf Research Program of the US National
Academies of Sciences, Engineering, and Medicine","Figure
found":"NO","Figure":"","material":""}
Processing Time: 7.161860466003418 seconds
Input Tokens: 15859
Output Tokens: 151

```

-----

PDF File:  
Benningstad\_und\_Kunst\_2020\_Dissociating\_meat\_from\_its\_animal\_origins.pdf  
Response: The answer is:

```

{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure
1", "material": "meat"} Here is the JSON output:

```

```

{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 4.6",
  "material": "meat"
}
Processing Time: 4.211366891860962 seconds
Input Tokens: 13879
Output Tokens: 78

```

-----

PDF File: Gottschall2018\_A multicaloric cycle that exploits thermal hysteresis\_NatureMat.pdf  
Response: The answer is:

```

{"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "Ni-Mn-In Heusler alloy"} {
"title": "Large reversible magnetocaloric effect in Ni-Mn-In-Co",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Ni50.0Mn35.5In14.5"
}
Processing Time: 6.543430328369141 seconds
Input Tokens: 8750
Output Tokens: 103

```

-----

PDF File:  
Leslie\_et\_al\_2022\_Discovery\_and\_quantification\_of\_plastic\_particle.pdf

---

Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. Therefore, the answer is:

```
{"title": "The title of the paper", "Figure found": "NO", "Figure": "",  
"material": ""}
```

However, the document does discuss the analysis of plastic particles in human blood, including the development of a method to quantify the mass of five high production volume polymers applied in plastic materials: poly(methyl methacrylate) (PMMA), polypropylene (PP), materials containing polymerized styrene (PS), polyethylene (PE), and polyethylene terephthalate (PET). The method involves double shot pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) and is used to measure the mass concentrations of these polymers in human blood. The answer is:

```
{"title":"The title of the paper is not provided","Figure found":"YES","Figure  
number":"3.1","Material":"Human blood"} {"title":"Screening criteria for  
microplastic particles in air and atmospheric deposition: critical review and  
applicability towards assessing human exposure","Figure  
found":"NO","Figure":"","material":""}
```

Processing Time: 10.34526515007019 seconds

Input Tokens: 15870

Output Tokens: 249

-----

...

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 2",  
  "material": "OPEs (organophosphate esters)"  
}
```

``` The JSON output is:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Fig. 2",  
"material": "none"  
}
```

The paper discusses the exposure to organophosphate ester (OPE) flame retardants and plasticizers in pregnant women and their potential health risks. The authors analyzed data from the LIFECODES Fetal Growth Study, which included 900 pregnant women, and found that the concentrations of OPE biomarkers varied over time and were influenced by sociodemographic and pregnancy characteristics. The authors also found that the concentrations of some OPE biomarkers were higher in certain populations, such as younger women, those with lower educational attainment, and those from racial and ethnic minority groups. {

```
"title": "Prenatal exposure to organophosphate esters is associated with  
decreased anogenital distance in offspring",
```

```
"Figure found": "YES",
```

```
"Figure": "Figure 1",
```

```
"material": "Organophosphate esters"
```

```
}
```

Processing Time: 11.764989852905273 seconds

Input Tokens: 18882

---

Output Tokens: 266

-----  
PDF File:

Yang\_et\_al\_2023\_Detection\_of\_Various\_Microplastics\_in\_Patients\_Und.pdf  
Response: Here is the JSON response:

```
{
  "title": "The Detection of Microplastics in Patients Undergoing Cardiac Surgery",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "Material": "Various polymers, including PVC, PET, and PMMA"
} {
  "title": "The title of the paper",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
```

Processing Time: 3.978264808654785 seconds

Input Tokens: 10276

Output Tokens: 85

-----  
PDF File: onep.pdf

Response: {

```
"title": "Environment International 163 (2022) 107199",
"Figure found": "NO",
"Figure": "",
"material": ""
}
```

Processing Time: 1.1482188701629639 seconds

Input Tokens: 1452

Output Tokens: 33

-----  
PDF File: BMS.pdf

Response: {

```
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3a",
"material": "La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite"
}
```

Processing Time: 2.2888214588165283 seconds

Input Tokens: 4845

Output Tokens: 60

-----  
PDF File: Tuneable\_Giant\_Magnetocaloric\_Effect\_in\_MnFe2PSi\_M.pdf

Response: Here is the JSON response:

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "Mn,Fe)2(P,Si)-based materials"
} {
  "title": "The magnetocaloric effect and magnetic refrigeration near room temperature: Materials and models",
  "Figure found": "YES",
  "Figure": "Figure 4",
  "material": "MnFe(P,Si,B) and MnFe(Ni,P,Si,B)"
}
```

---

Processing Time: 4.84161376953125 seconds  
Input Tokens: 11212  
Output Tokens: 110

-----  
PDF File: 122405\_1\_online.pdf  
Response: Here is the JSON response:

```
{  
  "title": "Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system",  
  "Figure found": "YES",  
  "Figure": "Figure 3",  
  "material": "MnNiGe-CoNiGe"  
}
```

Processing Time: 2.79042649269104 seconds  
Input Tokens: 7672  
Output Tokens: 56

-----  
PDF File: PhysRevB.93.184403.pdf  
Response: The answer is:

```
{"title": "Realization of two-dimensional ferromagnetism with giant coercivity  
in ultrathin  $\beta$ -Ni(OH)2 layers grown on a MoS2 surface", "Figure found": "Yes",  
"Figure": "Figure 1", "Material": " $\beta$ -Ni(OH)2"} {  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 8",  
  "material": "MoS2/Ni(OH)2 composite"  
}
```

Processing Time: 5.04149866104126 seconds  
Input Tokens: 12780  
Output Tokens: 111

-----  
PDF File: ArticleSupra2020.pdf  
Response: The JSON response is:

```
{  
  "title": "Structural Analysis and Theoretical Investigations of the  
Magnetocaloric Effect for La0.7Ba0.15Ag0.15MnO3 Manganite Prepared Using Sol-  
Gel Route",  
  "Figure found": "YES",  
  "Figure": "Figure 2d",  
  "material": "La0.7Ba0.15Ag0.15MnO3"  
} {  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Fig. 7",  
  "material": "La0.7Ba0.15Ag0.15MnO3"  
}
```

Processing Time: 8.28354811668396 seconds  
Input Tokens: 10009  
Output Tokens: 137

-----  
PDF File: Krishna\_Murthy\_2015\_J.\_Phys.\_D \_Appl.\_Phys.\_48\_355001.pdf  
Response: Here is the JSON output:

```
{  
  "title": "Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic  
insulators",  
  "Figure found": "YES",  
}
```

---

```
"Figure": "Figure 1(a) and (b)",
"material": "Gd2NiMnO6 and Gd2CoMnO6"
Processing Time: 3.418886661529541 seconds
Input Tokens: 6793
Output Tokens: 83
```

```
-----
PDF File: Zinn - 2023 - Zweite Moderne und Risikogesellschaft.pdf
Response: The JSON output is:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "none"
} Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.754127025604248 seconds
Input Tokens: 12942
Output Tokens: 71
```

```
-----
PDF File: Albert_und_Kleinman_2011_Bringing_Pierre_Bourdieu_to_Science_and.pdf
Response: Here is the JSON response:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 1: The Scientific Field: A Space of Struggle for Power",
"material": "The paper discusses the application of Pierre Bourdieu's concepts
to the study of science and technology studies (STS)."} {"title":"Bourdieu: Critical perspectives", "Figure
found":"NO", "Figure":"","material":""}
Processing Time: 4.151360511779785 seconds
Input Tokens: 8648
Output Tokens: 98
```

```
-----
PDF File:
D'Angelo_und_Meccariello_2021_Microplastics_A_Threat_for_Male_Fertility.pdf
Response: The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure
1", "material": "microplastics"}
```

The paper discusses the effects of microplastics on male fertility and reproduction. The authors review the current state of knowledge on the topic and present some of their own research findings. They conclude that microplastics can have negative effects on sperm quality and fertility, and that further research is needed to fully understand the risks associated with microplastic exposure. Here is the JSON response:

```
{
"title": "Q. Molecular mechanisms of PFOA-induced toxicity in animals and
humans: Implications for health risks.",
"Figure found": "NO",
```

---

```
"Figure": "",
"material": ""
}
```

```
Processing Time: 6.565014600753784 seconds
Input Tokens: 13286
Output Tokens: 155
```

```
-----
PDF File:
```

```
Paul_et_al_2022_Institutionalised_ignorance_in_policy_and_regulati.pdf
Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the concept of institutionalized ignorance and its role in governance, policy, and regulation.
```

According to the document, institutionalized ignorance refers to the normalization of not knowing or ignoring certain information, which can have significant consequences for accountability and decision-making. The authors argue that ignorance is not just a lack of knowledge, but a deliberate strategy used by institutions to maintain power and control.

The document also discusses the concept of agnotology, which refers to the study of ignorance and non-knowledge. The authors argue that agnotology is an important field of study that can help us understand how ignorance is produced and maintained, and how it can be challenged and overcome.

The document does not provide a specific material that is the subject of the M vs. T plot, but it does discuss the importance of understanding the role of ignorance in various fields, including science, policy, and regulation.

Here is the JSON output:

```
{
  "title": "The title of the paper",
  "Figure found": "NO",
  "Figure": "N/A",
  "Material": "N/A"
} {
  "title": "Routledge International Handbook of Ignorance Studies",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
```

```
Processing Time: 9.9422607421875 seconds
Input Tokens: 9626
Output Tokens: 274
```

```
-----
PDF File: 1-s2.0-S0304885312009754-main.pdf
Response: Here is the JSON output:
```

```
{
  "title": "Phase diagram, ferromagnetic martensitic transformation and magnetoresponse properties of Fe-doped MnCoGe alloys",
  "Figure found": "YES",
  "Figure": "Figure 2 (a) and (b)",
  "material": "MnCoGe alloys"
```

```
Processing Time: 2.9770314693450928 seconds
Input Tokens: 6637
Output Tokens: 66
```

---

PDF File: Kuhlicke - 2023 - Resilienz und Risiko.pdf  
Response: The JSON output is:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 2.1",  
"material": "none"  
} {  
"title": "The title of the paper",  
"Figure found": "NO",  
"Figure": "",  
"material": ""  
}  
Processing Time: 3.5361948013305664 seconds  
Input Tokens: 12243  
Output Tokens: 67
```

-----  
PDF File: Wajcman - 2010 - Feminist theories of technology.pdf  
Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the relationship between gender and technology, and how technology is shaped by and shapes gender relations.

The document is a scholarly article that explores the intersection of feminist theory and science and technology studies (STS). The author, Judy Wajcman, argues that technology is not a neutral or objective field, but rather a social construct that is shaped by and shapes gender relations.

The article discusses the ways in which technology has been used to reinforce gender hierarchies and the ways in which women have been excluded from the development and use of technology. Wajcman argues that feminist theory can help to challenge these gendered power relations and create a more inclusive and equitable technology.

The article does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot, but it does contain several tables, figures, and diagrams that illustrate the author's arguments about the relationship between gender and technology.

Here is the JSON response:

```
{  
"title": "The title of the paper",  
"Figure found": "NO",  
"Figure": "N/A",  
"Material": "N/A"  
} {"title":"The title of the paper":null,"Figure  
found":"NO","Figure":null,"material":null}  
Processing Time: 9.79239010810852 seconds  
Input Tokens: 8825  
Output Tokens: 273
```

-----  
PDF File: Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C.pdf  
Response: Here is the JSON response:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",
```

---

"Figure": "Figure 5a",  
"Material": "Cyanido-bridged metal assemblies containing both Fe and Co ions"  
} Here is the JSON output:

```
...  
{  
  "title": "Inorganics 2017,5, 63",  
  "Figure found": "YES",  
  "Figure": "Figure 6",  
  "material": "K4[Nb(CN)8]2H2O, (Fe0.92Co0.08)2[Nb(CN)8](4-pyridinealdoxime)  
83H2O, (Fe0.71Co0.29)2[Nb(CN)8](4-pyridinealdoxime) 83H2O, (Fe  
0.50Co0.50)2[Nb(CN)8](4-pyridinealdoxime) 83H2O, (Fe 0.21Co0.79)2[Nb(CN)8](4-  
pyridinealdoxime) 83H2O, Co 2[Nb(CN)8](4-pyridinealdoxime) 83H2O"  
}  
...
```

Processing Time: 10.677539587020874 seconds  
Input Tokens: 14082  
Output Tokens: 288

-----  
PDF File: JAP115123908.pdf  
Response: The magnetization vs. temperature (M vs. T) plot is found in this document. The title of the paper is "Magnetic properties of mixed spinel BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites".

The answer is: YES, a magnetization vs. temperature (M vs. T) plot is found in this document.

The M vs. T plot is shown in Fig. 4, which displays the field-cooled (FC) magnetization ( $M_{FC}$ ) as a function of temperature for all samples down to 5 K. The plot shows that the ferrimagnetic transition occurs at about 848 K for pure NFO, and the value of  $T_C$  decreases from 848 K to 836 K for the solid solution with an increase in BTO phase.

The material is BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites. Here is the JSON output:

```
...  
{  
  "title": "Synthesis of (100-x)BTO-(x)NFO solid solution at high  
temperature",  
  "Figure found": "YES",  
  "Figure": "Figure 5",  
  "material": "NFO and BTO"  
}  
...
```

Processing Time: 24.462421894073486 seconds  
Input Tokens: 14291  
Output Tokens: 243

-----  
PDF File: Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO<sub>3</sub> and LaFe<sub>0.5</sub>Cr<sub>0.5</sub>O<sub>3</sub>.pdf  
Response: The answer is:

```
{"title": "Magnetic Properties and Mössbauer Study of Perovskite LaFeO3 and  
LaFe0.5Cr0.5O3",  
"Figure found": "YES",  
"Figure": "Figure 1",  
"material": "LaFeO3 and LaFe0.5Cr0.5O3"} {  
"title": "Phys. Status Solidi RRL 2022,16, 2200023",  
"Figure found": "YES",
```

---

```
"Figure": "Figure 1",
"material": "none"
}
```

```
Processing Time: 4.89314603805542 seconds
Input Tokens: 8415
Output Tokens: 124
```

```
-----
PDF File: HCMO.pdf
Response: YES
```

```
Figure: Figure 1(a)
```

```
Material: Ho2CoMnO6
```

The magnetization vs. temperature (M-T) plot under various magnetic fields is shown in Figure 3(a). The plot exhibits a valley-like feature in the low-temperature region, indicating a negative magnetization. The negative magnetization is attributed to the 3d-4f negative exchange interaction between the Ho<sup>3+</sup> ions and the FM mesh of Co<sup>2+</sup>/Mn<sup>4+</sup> sublattices.

The magnetocaloric effect (MCE) is studied by measuring the temperature dependence of magnetization under various magnetic fields. The MCE is found to be significant, with a maximum value of 13.4 J/kg K at a low temperature and a magnetic field of 70 kOe. The MCE is attributed to the change in the magnetic entropy of the system upon the application of a magnetic field.

The magnetocaloric parameters, such as the relative cooling power (RCP) and the adiabatic temperature change ( $\Delta T_{ad}$ ), are also calculated. The RCP is found to be 12.5 J/kg K and the  $\Delta T_{ad}$  is found to be 4.2 K at a magnetic field of 70 kOe Here is the JSON response:

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 3(d)",
  "material": "Ho2CoMnO6",
  "material description": "The title compound is a rare-earth-based double perovskite compound, Ho2CoMnO6, which exhibits a complex magnetic behavior."
} Here is the JSON output:
```

```
...
```

```
{
  "title": "Structural, magnetic, and magnetocaloric properties of Ho2CoMnO6",
  "Figure found": "YES",
  "Figure": "Figure 7(b)",
  "material": "Ho2CoMnO6"
}
...
```

```
Processing Time: 15.713869571685791 seconds
Input Tokens: 21154
Output Tokens: 405
```

```
-----
PDF File: 1-s2.0-S0109564113000067-main.pdf
Response: The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 2", "material": "Polymer-infiltrated-ceramic-network material (PICN)"}
```

---

The paper discusses the mechanical properties of a novel material called Polymer-infiltrated-ceramic-network material (PICN), which is a two-phase composite material composed of a ceramic network and a polymer network. The authors investigate the effects of varying ceramic precursor densities on the mechanical properties of the PICN, including flexural strength, elastic modulus, strain at failure, and hardness.

The paper presents the results of experiments on four different PICNs with varying ceramic precursor densities, as well as a dense ceramic and a pure polymer. The results show that the mechanical properties of the PICN are influenced by the ceramic precursor density, with lower ceramic fractions resulting in lower elastic modulus and hardness, but higher flexural strength and strain at failure.

The authors also present SEM images of the indented areas of the materials, which show that the PICN exhibits a unique microstructure with a continuous polymer network and a ceramic network. The results suggest that the PICN has potential for use in dental applications, as it exhibits mechanical properties similar to those of natural teeth. {

```
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 3",  
"material": "PICN (polymer-infiltrated ceramic network)"  
}  
Processing Time: 11.327147006988525 seconds  
Input Tokens: 11623  
Output Tokens: 299
```

-----  
PDF File: zno-serie5.pdf  
Response: Here is the JSON output:

```
...  
{  
  "title": "Structural and magnetic characterization of (Zn 1-x-yFexCoy)O",  
  "Figure found": "YES",  
  "Figure": "Figure 2",  
  "material": "ZnO, Fe-doped ZnO, Co-doped ZnO"  
}  
...
```

Processing Time: 3.698740243911743 seconds  
Input Tokens: 7592  
Output Tokens: 75

-----  
PDF File: 1-s2.0-S1359645413004503-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 3", "material": "CoCrFeMnNi high-entropy alloy"} Here is the JSON output:
```

```
...  
{  
  "title": "An equiatomic CoCrFeMnNi high-entropy alloy",  
  "Figure found": "YES",  
  "Figure": "Figure 12",  
  "material": "CoCrFeMnNi high-entropy alloy"  
}  
...
```

Processing Time: 5.287784576416016 seconds  
Input Tokens: 14405

---

Output Tokens: 110

-----  
PDF File: 1-s2.0-S1002007116300995-main.pdf

Response: The answer is:

```
{"title": "Structural effects on the magnetic hyperthermia properties of iron oxide-based NPs", "Figure found": "YES", "Figure": "Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5", "Material": "Iron oxide nanoparticles"} The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 5", "material": "iron oxide-based magnetic hyperthermia agents"}
```

The paper discusses the use of iron oxide-based magnetic hyperthermia agents for cancer treatment. The authors review the current state of the art in this field and discuss the potential benefits and challenges of using these agents. They also present some of their own research on the topic, including the development of new iron oxide-based magnetic hyperthermia agents and the testing of their effectiveness in treating cancer in animal models. {

```
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 103",  
"material": "Fe3O4"  
}
```

Processing Time: 9.752267360687256 seconds

Input Tokens: 17383

Output Tokens: 233

-----  
PDF File: app135z2p11.pdf

Response: {

```
"title": "Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe70.81xCo0.10Mn10Mo5B15(0x0.5) Alloys at Low Magnetizing Field",  
"Figure found": "YES",  
"Figure": "Figure 1 and Figure 2",  
"material": "amorphous Fe70.81xCo0.10Mn10Mo5B15(0x0.5) alloys"
```

Processing Time: 3.103688955307007 seconds

Input Tokens: 3375

Output Tokens: 97

-----  
PDF File: 035220\_1\_online.pdf

Response: Here is the JSON output:

...

```
{  
  "title": "Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity",  
  "Figure found": "YES",  
  "Figure": "Figure 4",  
  "material": "Gd-Fe-Al"  
}
```

Processing Time: 3.567056894302368 seconds

Input Tokens: 7390

Output Tokens: 80

-----  
PDF File: PhysRevB.81.214406.pdf

Response: The answer is:

```
{"title": "Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape memory alloys",
```

---

```
"Figure found": "Yes",
"Figure": "Figure 1",
"Material": "Ni-Mn-Sn ferromagnetic shape memory alloy"} {
"title": "PECULIARITIES OF THE MAGNETOCALORIC EFFECT PHYSICAL REVIEW B 81,
214406 /H208492010 /H20850",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "MnFePd"
}
Processing Time: 6.603277921676636 seconds
Input Tokens: 9218
Output Tokens: 131
```

-----  
PDF File: Gottschall2020\_JAP\_Advanced characterization of multicaloric materials in pulsed magnetic fields.pdf  
Response: Here is the JSON output:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 5",
"material": "Ni-Mn-Ti-Co"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "FIG. 6",
"material": "Heusler shape memory alloys"
}
Processing Time: 3.8675827980041504 seconds
Input Tokens: 10443
Output Tokens: 83
```

-----  
PDF File: Wehling\_2023\_Nichtwissen\_-\_ein\_ungewöhnlicher\_Schlüsselbegriff\_.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 3.1", "material": "FCKW (Fluor-Chlor-Kohlenwasserstoff)"}
```

The paper discusses the concept of "Nichtwissen" (ignorance) in the context of environmental sociology and its relevance to understanding environmental problems. The author argues that ignorance is not just a lack of knowledge, but a social construct that is shaped by various factors, including the way we perceive and define risk, uncertainty, and the limits of our knowledge.

The paper highlights the importance of considering ignorance in the context of environmental problems, such as climate change, and argues that it is not just a matter of individual ignorance, but rather a systemic issue that is rooted in the way we organize our societies and economies.

The author also discusses the concept of "wissenschaftliches Nichtwissen" (scientific ignorance), which refers to the ignorance that is generated by scientific research itself. This type of ignorance is not just a lack of knowledge, but rather a result of the limitations and biases of scientific inquiry.

The paper concludes by arguing that ignorance is a key concept in understanding environmental problems and that it is essential to recognize its Here is the JSON output:

---

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "none"
}
...
```

The paper discusses the concept of ignorance and its relevance in the context of environmental issues, particularly in the context of climate change and sustainability. The author, Peter Wehling, argues that ignorance is not just a lack of knowledge, but a complex and multifaceted phenomenon that is deeply embedded in societal and cultural contexts. He also highlights the importance of acknowledging and addressing ignorance in order to develop more effective strategies for addressing environmental challenges.

The paper is a collection of essays that explore the concept of ignorance from various perspectives, including sociology, philosophy, and environmental studies. The authors discuss the ways in which ignorance is produced and reproduced in different contexts, and the ways in which it can be addressed and overcome.

The paper is a valuable contribution to the field of environmental studies, as it highlights the importance of acknowledging and addressing ignorance in order to develop more effective strategies for addressing environmental challenges. It also provides a nuanced and complex understanding of the concept of ignorance, and its relevance to environmental issues.

Processing Time: 17.833888053894043 seconds  
Input Tokens: 14995  
Output Tokens: 501

-----  
PDF File: Nyadjro\_et\_al\_2023\_The\_NOAA\_NCEI\_marine\_microplastics\_database.pdf  
Response: The answer is:

```
{"title":"The title of the paper","Figure found":"YES","Figure number":"1","Material":"marine microplastics"} Here is the JSON response:
```

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "Various materials"
}
... {"title":"Fellowship from the Gulf Research Program of the US National Academies of Sciences, Engineering, and Medicine","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 5.906141519546509 seconds  
Input Tokens: 15859  
Output Tokens: 115

-----  
PDF File:  
Benningstad\_und\_Kunst\_2020\_Dissociating\_meat\_from\_its\_animal\_origins.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "meat"} Here is the JSON output:
```

---

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 4.6",
"material": "meat"
Processing Time: 4.217400550842285 seconds
Input Tokens: 13879
Output Tokens: 78
```

```
-----
PDF File: Gottschall2018_A multicaloric cycle that exploits thermal
hysteresis_NatureMat.pdf
Response: The answer is:
```

```
{"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "Ni-Mn-In Heusler alloy"} {
"title": "Large reversible magnetocaloric effect in Ni-Mn-In-Co",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Ni50.0Mn35.5In14.5"
}
Processing Time: 6.516801595687866 seconds
Input Tokens: 8750
Output Tokens: 103
```

```
-----
PDF File:
Leslie_et_al_2022_Discovery_and_quantification_of_plastic_particle.pdf
Response: The document does not contain a figure that depicts a magnetization
vs. temperature (M vs. T) plot. However, it does discuss the analysis of
plastic particles in human blood and the development of a method to quantify
the mass of five high production volume polymers applied in plastic materials:
poly(methyl methacrylate) (PMMA), polypropylene (PP), materials containing
polymerized styrene (PS), polyethylene (PE), and polyethylene terephthalate
(PET).
```

The method uses double shot pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) to analyze the samples. The authors report the results of a recovery experiment, which showed that the method can recover the spiked polymers from blood samples with recoveries ranging from 68% to 114%. They also discuss the control of background contamination during sampling, sample preparation, and analysis.

The authors conclude that the method can detect and quantify the four polymers in blood samples with sufficient sensitivity and precision. The measured concentration data in blood were corrected for the average procedural blank for each analyte.

Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": " The answer is:

{"title":"The title of the paper","Figure found":"YES","Figure
number":"3.1","Material":"Human blood"} {"title":"Impact of microplastics and
nanoplastics on human health","Figure found":"NO","Figure":"","material":""}
```

---

Processing Time: 12.60486888885498 seconds  
Input Tokens: 15870  
Output Tokens: 315

-----  
You are going to be given an example of a document. Please check this document to see if it contains a figure that depicts a magnetization vs. temperature (M vs T) PLOT. Keywords to look for include 'magnetization', 'temperature dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS THE M VS T FIGURE if it is AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE, ANSWER ONLY WITH a JSON return the following Json {"title":"The title of the paper";"Figure found":"If a magnetization vs. temperature PLOT is found YES or NO";"Figure":"which figure is the MAGNETIZATION VS. TEMPERATURE PLOT";"material":"what is the material that is the plot"}  
-----

PDF File:  
Bommarito\_et\_al\_2023\_Temporal\_trends\_and\_predictors\_of\_gestational.pdf  
Response: The answer is:

```
{"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 2",  
"material": "human biomarkers of exposure to organophosphate ester flame retardants and plasticizers in urine from pregnant people"} The JSON output is:
```

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Fig. 2",  
  "material": "human urine samples"  
} {  
  "title": "Wang et al. (2023)",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

Processing Time: 7.283336162567139 seconds  
Input Tokens: 18882  
Output Tokens: 136

-----  
PDF File:  
Yang\_et\_al\_2023\_Detection\_of\_Various\_Microplastics\_in\_Patients\_Und.pdf  
Response: Here is the JSON response:

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 1",  
  "material": "Various polymers, including PVC, PET, and PMMA"  
} {  
  "title": "The title of the paper",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

Processing Time: 3.720438241958618 seconds  
Input Tokens: 10276  
Output Tokens: 77

---

-----  
PDF File: onep.pdf  
Response: {  
"title": "Environment International 163 (2022) 107199",  
"Figure found": "NO",  
"Figure": "",  
"material": ""  
}  
Processing Time: 1.147451400756836 seconds  
Input Tokens: 1452  
Output Tokens: 33  
-----

PDF File: BMS.pdf  
Response: {  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 3a",  
"material": "La<sub>0.70</sub>Ca<sub>0.1</sub>Sr<sub>0.2</sub>Fe<sub>0.1</sub>Mn<sub>0.903</sub> perovskite"  
}  
Processing Time: 2.292752981185913 seconds  
Input Tokens: 4845  
Output Tokens: 60  
-----

PDF File: Tuneable\_Giant\_Magnetocaloric\_Effect\_in\_MnFe<sub>2</sub>PSi\_M.pdf  
Response: Here is the JSON output:

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 1",  
  "material": "Mn,Fe)2(P,Si)-based materials"  
} {  
  "title": "The magnetocaloric effect and magnetic refrigeration near room  
temperature: Materials and models",  
  "Figure found": "YES",  
  "Figure": "Figure 4",  
  "material": "MnFe(P,Si,B) and MnFe(Ni,P,Si,B)"  
}  
Processing Time: 4.847777605056763 seconds  
Input Tokens: 11212  
Output Tokens: 110  
-----
```

PDF File: 122405\_1\_online.pdf  
Response: Here is the JSON response:

```
{  
  "title": "Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system",  
  "Figure found": "YES",  
  "Figure": "Figure 3",  
  "material": "MnNiGe-CoNiGe"  
}  
Processing Time: 2.7820096015930176 seconds  
Input Tokens: 7672  
Output Tokens: 56  
-----
```

PDF File: PhysRevB.93.184403.pdf  
Response: The answer is:

---

```
{"title": "Realization of two-dimensional ferromagnetism with giant coercivity
in ultrathin  $\beta$ -Ni(OH)2 layers grown on a MoS2 surface",
"Figure found": "YES",
"Figure": "Figure 4",
"Material": " $\beta$ -Ni(OH)2"} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 8",
"material": "MoS2/Ni(OH)2"
}
Processing Time: 5.184289455413818 seconds
Input Tokens: 12780
Output Tokens: 113
```

-----  
PDF File: ArticleSupra2020.pdf  
Response: Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 2d",
"material": "La0.7Ba0.15Ag0.15MnO3 manganite"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Fig. 7",
"material": "La0.7Ba0.15Ag0.15MnO3"
}
Processing Time: 7.593746185302734 seconds
Input Tokens: 10009
Output Tokens: 107
```

-----  
PDF File: Krishna\_Murthy\_2015\_J.\_Phys.\_D \_Appl.\_Phys.\_48\_355001.pdf  
Response: Here is the JSON output:

```
{
"title": "Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic
insulators",
"Figure found": "YES",
"Figure": "Figure 1(a) and (b)",
"material": "Gd2NiMnO6 and Gd2CoMnO6"
}
Processing Time: 4.6754374504089355 seconds
Input Tokens: 6793
Output Tokens: 83
```

-----  
PDF File: Zinn - 2023 - Zweite Moderne und Risikogesellschaft.pdf  
Response: The answer is:

```
{"title":"The title of the paper","Figure found":"YES","Figure":"Figure
3","material":"none"} Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.459108591079712 seconds
Input Tokens: 12942
```

---

Output Tokens: 62

-----  
PDF File: Albert\_und\_Kleinman\_2011\_Bringing\_Pierre\_Bourdieu\_to\_Science\_and.pdf  
Response: The answer is:

```
{"title":"The title of the paper is 'Bringing Pierre Bourdieu to Science and Technology Studies',"Figure found":"YES","Figure":"Figure 1","material":"The paper discusses the application of Pierre Bourdieu's concepts to the field of Science and Technology Studies (STS). It argues that Bourdieu's work can provide insights into the operation of science in the contemporary world and can be used to analyze the social dynamics of the scientific field."}
```

```
{"title":"Bourdieu: Critical perspectives","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 4.860419750213623 seconds

Input Tokens: 8648

Output Tokens: 120

-----  
PDF File:

D'Angelo\_und\_Meccariello\_2021\_Microplastics\_A\_Threat\_for\_Male\_Fertility.pdf

Response: The answer is:

```
{"title":"The title of the paper","Figure found":"YES","Figure number":"Figure 1","Material":"Not mentioned"} Here is the JSON output:
```

...

```
{  
  "title": "Ma, L.Q. Molecular mechanisms of PFOA-induced toxicity in animals and humans: Implications for health risks.",  
  "Figure found": "NO",  
  "Figure": "",  
  "material": ""  
}
```

...

Processing Time: 4.424508094787598 seconds

Input Tokens: 13286

Output Tokens: 90

-----  
PDF File:

Paul\_et\_al\_2022\_Institutionalised\_ignorance\_in\_policy\_and\_regulati.pdf

Response: The answer is:

```
{"title":"The title of the paper";"Figure found":"YES";"Figure":"Figure 1";"material":"unknown"} {
```

```
"title": "Routledge International Handbook of Ignorance Studies",
```

```
"Figure found": "NO",
```

```
"Figure": "",
```

```
"material": ""
```

```
}
```

Processing Time: 3.0628509521484375 seconds

Input Tokens: 9626

Output Tokens: 61

-----  
PDF File: 1-s2.0-S0304885312009754-main.pdf

Response: Here is the JSON output:

```
{  
  "title": "Phase diagram, ferromagnetic martensitic transformation and magnetoresponse properties of Fe-doped MnCoGe alloys",  
  "Figure found": "YES",
```

---

```
"Figure": "Figure 2",
"material": "MnCoGe alloys"
}
```

```
Processing Time: 2.810117244720459 seconds
Input Tokens: 6637
Output Tokens: 61
```

```
-----
PDF File: Kuhlicke - 2023 - Resilienz und Risiko.pdf
Response: The JSON output is:
```

```
{
"figure": "Figure 2.1",
"material": "none"
} {
"figure": "Figure 2.1",
"material": "none"
}
```

```
Processing Time: 3.552651882171631 seconds
Input Tokens: 12243
Output Tokens: 67
```

```
-----
PDF File: Wajcman - 2010 - Feminist theories of technology.pdf
Response: The paper "Feminist Theories of Technology" by Judy Wajcman does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, the paper does discuss the relationship between gender and technology, and how technology is shaped by and shapes gender relations.
```

The paper argues that feminist theories of technology have come a long way over the last quarter of a century, and that the intellectual exploration at the intersections of feminist scholarship and science and technology studies (STS) has enriched both fields immeasurably.

The paper highlights the importance of understanding the mutual shaping of gender and technology, and how the materiality of technology affords or inhibits the doing of particular gender power relations. It also emphasizes the need to involve women throughout the processes and practices of technological innovation, and to recognize the constitutive power of tools, techniques, and objects to materialize social, political, and economic arrangements.

The paper does not mention the concept of "material-semiotic practice" or "technoscience" explicitly, but it does discuss the social constructivist approach to technology, which views technology as a sociotechnical product that is shaped by social and cultural factors.

The paper does not mention the authors Manuel Castells, Nicholas Negroponte

```
{"title": "The title of the paper": null, "Figure found": "NO", "Figure": null, "material": null}
Processing Time: 10.056691646575928 seconds
Input Tokens: 8825
Output Tokens: 282
```

```
-----
PDF File: Metal_Substitution_Effect_on_a_Three-Dimensional_C.pdf
Response: Here is the JSON response:
```

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 5",
  "Material": "Cyanido-bridged metal assemblies containing both Fe and Co
ions"
}
... Here is the JSON output:
...
{
  "title": "Inorganics 2017,5, 63",
  "Figure found": "YES",
  "Figure": "Figure 6",
  "material": "K4[Nb(CN)8]2H2O, (Fe0.21Co0.79)2[Nb(CN)8](4-pyridinealdoxime)
8zH2O"
}
...
```

```
Processing Time: 6.898503065109253 seconds
Input Tokens: 14082
Output Tokens: 161
```

```
-----
PDF File: JAP115123908.pdf
Response: The magnetization vs. temperature (M vs. T) plot is found in the
article. Therefore, the answer is:
```

**\*\*YES\*\***

The M vs. T plot is Figure 4 in the article.

**\*\*Material:\*\*** BaTiO3-NiFe2O4 composite

**\*\*Figure:\*\*** Figure 4

**\*\*Description:\*\*** The figure shows the field-cooled magnetization (M) as a function of temperature (T) for the BaTiO3-NiFe2O4 composite. The magnetization decreases with increasing temperature, and the Curie temperature (Tc) is observed to decrease with increasing BaTiO3 content. Here is the JSON output:

```
...
{
  "title": "Synthesis of (100-x)BTO-(x)NFO solid solution at high
temperature",
  "Figure found": "YES",
  "Figure": "Figure 5",
  "material": "NFO and BTO"
}
...
```

```
Processing Time: 22.981261491775513 seconds
Input Tokens: 14291
Output Tokens: 196
```

```
-----
PDF File: Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and M
ssbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3.pdf
Response: The answer is:
```

---

```
{"title": "Magnetic Properties and Mössbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3", "Figure found": "YES", "Figure": "Figure 1", "material": "LaFeO3 and LaFe0.5Cr0.5O3"} {
"title": "Phys. Status Solidi RRL 2022,16, 2200023",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "none"
}
```

```
Processing Time: 4.796543598175049 seconds
Input Tokens: 8415
Output Tokens: 121
```

```
-----
PDF File: HCMO.pdf
Response: The answer is:
```

```
{"title": "Role of 3d-4f exchange interaction and local anti-site defects in the magnetic and magnetocaloric properties of double perovskite Ho2CoMnO6 compound", "Figure found": "YES", "Figure": "Figure 2", "Material": "Ho2CoMnO6"} Here is the JSON response:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3(d)",
"material": "Ho2CoMnO6",
"parameters": {
"jΔSjmax": 13.44 J/kg K,
"RCP": 272 J/kg,
"TEC": 6.5 (3.9) J/kg K,
"NRC": 1.5 (2.1)
}
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 7(b)",
"material": "Ho2CoMnO6"
}
```

```
Processing Time: 9.795434474945068 seconds
Input Tokens: 21154
Output Tokens: 218
```

```
-----
PDF File: 1-s2.0-S0109564113000067-main.pdf
Response: The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 2", "material": "Polymer-infiltrated-ceramic-network material (PICN)"}
```

The paper discusses the mechanical properties of a novel material called Polymer-infiltrated-ceramic-network material (PICN), which is a two-phase composite material composed of a ceramic network and a polymer network. The authors investigate the effects of varying ceramic precursor densities on the mechanical properties of the PICN, including flexural strength, elastic modulus, strain at failure, and hardness.

The paper presents a comprehensive study of the mechanical properties of PICNs, including the effects of ceramic precursor density on the material's properties. The results show that the flexural strength of the PICN increases with decreasing ceramic precursor density, while the elastic modulus and

---

hardness increase with increasing ceramic precursor density. The strain at failure of the PICN also decreases with increasing ceramic precursor density.

The authors also present scanning electron microscopy (SEM) images of the indented areas of the PICN, which show that the material exhibits a unique microstructure with a continuous polymer network and a ceramic network. The SEM images also show that the material exhibits a high degree of toughness and resistance to {

```
"title": "The title of the paper",  
"Figure found": "NO",  
"Figure": "",  
"material": ""  
}
```

Processing Time: 10.908538818359375 seconds

Input Tokens: 11623

Output Tokens: 284

-----  
PDF File: zno-serie5.pdf

Response: Here is the JSON output:

...

```
{  
  "title": "Structural and magnetic characterization of (Zn 1-x-yFexCoy)O",  
  "Figure found": "YES",  
  "Figure": "Figure 2",  
  "material": "ZnO, Fe-doped ZnO, Co-doped ZnO"  
}  
...
```

Processing Time: 3.6972219944000244 seconds

Input Tokens: 7592

Output Tokens: 75

-----  
PDF File: 1-s2.0-S1359645413004503-main.pdf

Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 9", "material": "CoCrFeMnNi high-entropy alloy"} Here is the JSON output:
```

...

```
{  
  "title": "An equiatomic CoCrFeMnNi high-entropy alloy was produced by arc melting, drop casting and rolling, after which it was recrystallized to produce single-phase, fcc-structured, equiaxed microstructures with three different grain sizes, 4.4, 50 and 155 nm.",  
  "Figure found": "YES",  
  "Figure": "Figure 11",  
  "material": "CoCrFeMnNi high-entropy alloy"  
}  
...
```

Processing Time: 6.863595485687256 seconds

Input Tokens: 14405

Output Tokens: 162

-----  
PDF File: 1-s2.0-S1002007116300995-main.pdf

Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "Iron oxide nanoparticles"} The answer is:
```

---

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 5", "material": "iron oxide-based magnetic hyperthermia agents"}
```

The paper discusses the use of iron oxide-based magnetic hyperthermia agents for cancer treatment. The authors review the current state of the art in this field and discuss the potential benefits and challenges of using these agents. They also present some of their own research on the topic, including the development of new iron oxide-based magnetic hyperthermia agents and their testing in animal models.

The figure that is referred to is Figure 5, which shows the results of the authors' experiments on the heating efficiency of different iron oxide-based magnetic hyperthermia agents. The figure shows that the agents with a higher concentration of iron oxide particles have a higher heating efficiency.

The material that is being discussed is iron oxide-based magnetic hyperthermia agents, which are used to generate heat in the body to treat cancer. {

```
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 103",  
"material": "Fe3O4"  
}
```

```
Processing Time: 11.12071442604065 seconds  
Input Tokens: 17383  
Output Tokens: 275
```

-----  
PDF File: app135z2p11.pdf

Response: {

```
"title": "Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe7093xCoxMn10Mo5B15(0x0.5) Alloys at Low Magnetizing Field",  
"Figure found": "YES",  
"Figure": "Figure 1 and Figure 2",  
"material": "amorphous Fe7093xCoxMn10Mo5B15(0x0.5) alloys"
```

```
Processing Time: 3.0970208644866943 seconds  
Input Tokens: 3375  
Output Tokens: 97
```

-----  
PDF File: 035220\_1\_online.pdf

Response: Here is the JSON output:

```
{  
"title": "Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity",  
"Figure found": "YES",  
"Figure": "Figure 4",  
"material": "Gd-Fe-Al"  
}
```

```
Processing Time: 3.3428659439086914 seconds  
Input Tokens: 7390  
Output Tokens: 73
```

-----  
PDF File: PhysRevB.81.214406.pdf

Response: The answer is:

```
{"title": "Peculiarities of the magnetocaloric properties in Ni-Mn-Sn ferromagnetic shape memory alloys", "Figure found": "YES", "Figure": "Figure 1", "Material": "Ni-Mn-Sn ferromagnetic shape memory alloy"} {  
"title": "Peculiarities of the Magnetocaloric Effect",
```

---

```
"Figure found": "YES",
"Figure": "Figure 1",
"material": "MnFePGe"
}
Processing Time: 5.836395263671875 seconds
Input Tokens: 9218
Output Tokens: 105
```

```
-----
PDF File: Gottschall2020_JAP_Advanced characterization of multicaloric
materials in pulsed magnetic fields.pdf
Response: Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 5",
"material": "Ni-Mn-Ti-Co Heusler alloy"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "FIG. 6",
"material": "Heusler shape memory alloys"
}
Processing Time: 3.999363660812378 seconds
Input Tokens: 10443
Output Tokens: 87
```

```
-----
PDF File: Wehling_2023_Nichtwissen_-_ein_ungewöhnlicher_Schlüsselbegriff_.pdf
Response: The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "3.1",
"material": "none"} Here is the JSON output:
```

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "none"
}
...
```

The paper discusses the concept of ignorance and its relevance in the context of environmental issues, particularly in the field of climate change. The author, Peter Wehling, argues that ignorance is not just a lack of knowledge, but a complex and multifaceted phenomenon that can have significant consequences for our understanding of the world and our ability to make informed decisions.

The paper explores the concept of ignorance in various contexts, including the history of climate change research, the role of ignorance in the development of scientific knowledge, and the implications of ignorance for environmental policy and decision-making. The author also discusses the importance of acknowledging and addressing ignorance in order to promote more effective and sustainable solutions to environmental problems.

The paper is written in a clear and accessible style, making it suitable for readers with a general interest in environmental issues and the social sciences.

---

Processing Time: 9.742598533630371 seconds  
Input Tokens: 14995  
Output Tokens: 249

-----  
PDF File: Nyadjro\_et\_al\_2023\_The\_NOAA\_NCEI\_marine\_microplastics\_database.pdf  
Response: \*\*JSON Response:\*\*

```
{
  "title": "The NOAA NCEI global marine microplastics database",
  "Figure found": "YES",
  "Figure": "Figure 1, Figure 2, Figure 3, Figure 4",
  "Material": "Marine microplastics"
}
```

``` Here is the JSON response:

```

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "NO"
}
```

```
``` {"title":"Fellowship from the Gulf Research Program of the US National Academies of Sciences, Engineering, and Medicine","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 7.161860466003418 seconds  
Input Tokens: 15859  
Output Tokens: 151

-----  
PDF File:  
Benningstad\_und\_Kunst\_2020\_Dissociating\_meat\_from\_its\_animal\_origins.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "meat"} Here is the JSON output:
```

```
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 4.6",
  "material": "meat"
}
Processing Time: 4.211366891860962 seconds
Input Tokens: 13879
Output Tokens: 78
```

-----  
PDF File: Gottschall2018\_A multicaloric cycle that exploits thermal hysteresis\_NatureMat.pdf  
Response: The answer is:

```
{"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "Ni-Mn-In Heusler alloy"} {
"title": "Large reversible magnetocaloric effect in Ni-Mn-In-Co",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Ni50.0Mn35.5In14.5"
}
Processing Time: 6.543430328369141 seconds
```

---

Input Tokens: 8750  
Output Tokens: 103

-----  
PDF File:

Leslie\_et\_al\_2022\_Discovery\_and\_quantification\_of\_plastic\_particle.pdf

Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. Therefore, the answer is:

```
{"title": "The title of the paper", "Figure found": "NO", "Figure": "",  
"material": ""}
```

However, the document does discuss the analysis of plastic particles in human blood, including the development of a method to quantify the mass of five high production volume polymers applied in plastic materials: poly(methyl methacrylate) (PMMA), polypropylene (PP), materials containing polymerized styrene (PS), polyethylene (PE), and polyethylene terephthalate (PET). The method involves double shot pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) and is used to measure the mass concentrations of these polymers in human blood. The answer is:

```
{"title":"The title of the paper is not provided","Figure found":"YES","Figure  
number":"3.1","Material":"Human blood"} {"title":"Screening criteria for  
microplastic particles in air and atmospheric deposition: critical review and  
applicability towards assessing human exposure","Figure  
found":"NO","Figure":"","material":""}
```

Processing Time: 10.34526515007019 seconds

Input Tokens: 15870

Output Tokens: 249

-----

...

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 2",  
  "material": "OPEs (organophosphate esters)"  
}
```

``` The JSON output is:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Fig. 2",  
"material": "none"  
}
```

The paper discusses the exposure to organophosphate ester (OPE) flame retardants and plasticizers in pregnant women and their potential health risks. The authors analyzed data from the LIFECODES Fetal Growth Study, which included 900 pregnant women, and found that the concentrations of OPE biomarkers varied over time and were influenced by sociodemographic and pregnancy characteristics. The authors also found that the concentrations of some OPE biomarkers were higher in certain populations, such as younger women, those with lower educational attainment, and those from racial and ethnic minority groups. {

```
"title": "Prenatal exposure to organophosphate esters is associated with  
decreased anogenital distance in offspring",  
"Figure found": "YES",
```

---

```
    "Figure": "Figure 1",
    "material": "Organophosphate esters"
}
Processing Time: 11.764989852905273 seconds
Input Tokens: 18882
Output Tokens: 266
-----
PDF File:
Yang_et_al_2023_Detection_of_Various_Microplastics_in_Patients_Und.pdf
Response: Here is the JSON response:
```

```
{
  "title": "The Detection of Microplastics in Patients Undergoing Cardiac
Surgery",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "Material": "Various polymers, including PVC, PET, and PMMA"
} {
  "title": "The title of the paper",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
Processing Time: 3.978264808654785 seconds
Input Tokens: 10276
Output Tokens: 85
```

```
-----
PDF File: onep.pdf
Response: {
  "title": "Environment International 163 (2022) 107199",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
Processing Time: 1.1482188701629639 seconds
Input Tokens: 1452
Output Tokens: 33
```

```
-----
PDF File: BMS.pdf
Response: {
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 3a",
  "material": "La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite"
}
Processing Time: 2.2888214588165283 seconds
Input Tokens: 4845
Output Tokens: 60
```

```
-----
PDF File: Tuneable_Giant_Magnetocaloric_Effect_in_MnFe2PSi_M.pdf
Response: Here is the JSON response:

{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "Mn,Fe)2(P,Si)-based materials"
} {
```

---

```
"title": "The magnetocaloric effect and magnetic refrigeration near room
temperature: Materials and models",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "MnFe(P,Si,B) and MnFe(Ni,P,Si,B)"
}
```

```
Processing Time: 4.84161376953125 seconds
Input Tokens: 11212
Output Tokens: 110
```

```
-----
PDF File: 122405_1_online.pdf
Response: Here is the JSON response:
```

```
{
"title": "Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "MnNiGe-CoNiGe"
}
```

```
Processing Time: 2.79042649269104 seconds
Input Tokens: 7672
Output Tokens: 56
```

```
-----
PDF File: PhysRevB.93.184403.pdf
Response: The answer is:
```

```
{"title": "Realization of two-dimensional ferromagnetism with giant coercivity
in ultrathin  $\beta$ -Ni(OH)2 layers grown on a MoS2 surface", "Figure found": "Yes",
"Figure": "Figure 1", "Material": " $\beta$ -Ni(OH)2"} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 8",
"material": "MoS2/Ni(OH)2 composite"
}
```

```
Processing Time: 5.04149866104126 seconds
Input Tokens: 12780
Output Tokens: 111
```

```
-----
PDF File: ArticleSupra2020.pdf
Response: The JSON response is:
```

```
{
"title": "Structural Analysis and Theoretical Investigations of the
Magnetocaloric Effect for La0.7Ba0.15Ag0.15MnO3 Manganite Prepared Using Sol-
Gel Route",
"Figure found": "YES",
"Figure": "Figure 2d",
"material": "La0.7Ba0.15Ag0.15MnO3"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Fig. 7",
"material": "La0.7Ba0.15Ag0.15MnO3"
}
```

```
Processing Time: 8.28354811668396 seconds
Input Tokens: 10009
Output Tokens: 137
```

```
-----
PDF File: Krishna_Murthy_2015_J._Phys._D _Appl._Phys._48_355001.pdf
```

---

Response: Here is the JSON output:

```
{
"title": "Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic
insulators",
"Figure found": "YES",
"Figure": "Figure 1(a) and (b)",
"material": "Gd2NiMnO6 and Gd2CoMnO6"
Processing Time: 3.418886661529541 seconds
Input Tokens: 6793
Output Tokens: 83
```

-----  
PDF File: Zinn - 2023 - Zweite Moderne und Risikogesellschaft.pdf  
Response: The JSON output is:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "none"
} Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.754127025604248 seconds
Input Tokens: 12942
Output Tokens: 71
```

-----  
PDF File: Albert\_und\_Kleinman\_2011\_Bringing\_Pierre\_Bourdieu\_to\_Science\_and.pdf  
Response: Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 1: The Scientific Field: A Space of Struggle for Power",
"material": "The paper discusses the application of Pierre Bourdieu's concepts
to the study of science and technology studies (STS)."} {"title": "Bourdieu: Critical perspectives", "Figure
found": "NO", "Figure": "", "material": ""}
Processing Time: 4.151360511779785 seconds
Input Tokens: 8648
Output Tokens: 98
```

-----  
PDF File:  
D'Angelo\_und\_Meccariello\_2021\_Microplastics\_A\_Threat\_for\_Male\_Fertility.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure
1", "material": "microplastics"}
```

The paper discusses the effects of microplastics on male fertility and reproduction. The authors review the current state of knowledge on the topic and present some of their own research findings. They conclude that microplastics can have negative effects on sperm quality and fertility, and

---

that further research is needed to fully understand the risks associated with microplastic exposure. Here is the JSON response:

```
{
  "title": "Q. Molecular mechanisms of PFOA-induced toxicity in animals and humans: Implications for health risks.",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
```

Processing Time: 6.565014600753784 seconds

Input Tokens: 13286

Output Tokens: 155

-----  
PDF File:

Paul\_et\_al\_2022\_Institutionalised\_ignorance\_in\_policy\_and\_regulati.pdf

Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the concept of institutionalized ignorance and its role in governance, policy, and regulation.

According to the document, institutionalized ignorance refers to the normalization of not knowing or ignoring certain information, which can have significant consequences for accountability and decision-making. The authors argue that ignorance is not just a lack of knowledge, but a deliberate strategy used by institutions to maintain power and control.

The document also discusses the concept of agnotology, which refers to the study of ignorance and non-knowledge. The authors argue that agnotology is an important field of study that can help us understand how ignorance is produced and maintained, and how it can be challenged and overcome.

The document does not provide a specific material that is the subject of the M vs. T plot, but it does discuss the importance of understanding the role of ignorance in various fields, including science, policy, and regulation.

Here is the JSON output:

```
{
  "title": "The title of the paper",
  "Figure found": "NO",
  "Figure": "N/A",
  "Material": "N/A"
} {
  "title": "Routledge International Handbook of Ignorance Studies",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
```

Processing Time: 9.9422607421875 seconds

Input Tokens: 9626

Output Tokens: 274

-----  
PDF File: 1-s2.0-S0304885312009754-main.pdf

Response: Here is the JSON output:

```
{
  "title": "Phase diagram, ferromagnetic martensitic transformation and magnetoresponse properties of Fe-doped MnCoGe alloys",
```

---

```
"Figure found": "YES",
"Figure": "Figure 2 (a) and (b)",
"material": "MnCoGe alloys"
Processing Time: 2.9770314693450928 seconds
Input Tokens: 6637
Output Tokens: 66
-----
PDF File: Kuhlicke - 2023 - Resilienz und Risiko.pdf
Response: The JSON output is:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 2.1",
"material": "none"
} {
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.5361948013305664 seconds
Input Tokens: 12243
Output Tokens: 67
-----
```

```
PDF File: Wajcman - 2010 - Feminist theories of technology.pdf
Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the relationship between gender and technology, and how technology is shaped by and shapes gender relations.
```

The document is a scholarly article that explores the intersection of feminist theory and science and technology studies (STS). The author, Judy Wajcman, argues that technology is not a neutral or objective field, but rather a social construct that is shaped by and shapes gender relations.

The article discusses the ways in which technology has been used to reinforce gender hierarchies and the ways in which women have been excluded from the development and use of technology. Wajcman argues that feminist theory can help to challenge these gendered power relations and create a more inclusive and equitable technology.

The article does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot, but it does contain several tables, figures, and diagrams that illustrate the author's arguments about the relationship between gender and technology.

Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "N/A",
"Material": "N/A"
} {"title": "The title of the paper": null, "Figure found": "NO", "Figure": null, "material": null}
Processing Time: 9.79239010810852 seconds
Input Tokens: 8825
Output Tokens: 273
```

---

-----  
PDF File: Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C.pdf  
Response: Here is the JSON response:

```
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 5a",  
  "Material": "Cyanido-bridged metal assemblies containing both Fe and Co ions"  
} Here is the JSON output:
```

...

```
{  
  "title": "Inorganics 2017,5, 63",  
  "Figure found": "YES",  
  "Figure": "Figure 6",  
  "material": "K4[Nb(CN)8]2H2O, (Fe0.92Co0.08)2[Nb(CN)8](4-pyridinealdoxime)  
83H2O, (Fe0.71Co0.29)2[Nb(CN)8](4-pyridinealdoxime) 83H2O, (Fe  
0.50Co0.50)2[Nb(CN)8](4-pyridinealdoxime) 83H2O, (Fe 0.21Co0.79)2[Nb(CN)8](4-  
pyridinealdoxime) 83H2O, Co 2[Nb(CN)8](4-pyridinealdoxime) 83H2O"  
}  
...
```

Processing Time: 10.677539587020874 seconds  
Input Tokens: 14082  
Output Tokens: 288

-----  
PDF File: JAP115123908.pdf  
Response: The magnetization vs. temperature (M vs. T) plot is found in this document. The title of the paper is "Magnetic properties of mixed spinel BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites".

The answer is: YES, a magnetization vs. temperature (M vs. T) plot is found in this document.

The M vs. T plot is shown in Fig. 4, which displays the field-cooled (FC) magnetization (M<sub>FC</sub>) as a function of temperature for all samples down to 5 K. The plot shows that the ferrimagnetic transition occurs at about 848 K for pure NFO, and the value of T<sub>C</sub> decreases from 848 K to 836 K for the solid solution with an increase in BTO phase.

The material is BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composites. Here is the JSON output:

...

```
{  
  "title": "Synthesis of (100-x)BTO-(x)NFO solid solution at high  
temperature",  
  "Figure found": "YES",  
  "Figure": "Figure 5",  
  "material": "NFO and BTO"  
}  
...
```

Processing Time: 24.462421894073486 seconds  
Input Tokens: 14291  
Output Tokens: 243

-----  
PDF File: Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mossbauer Study of Perovskite LaFeO<sub>3</sub> and LaFe<sub>0.5</sub>Cr<sub>0.5</sub>O<sub>3</sub>.pdf  
Response: The answer is:

```
-----
{"title": "Magnetic Properties and Mössbauer Study of Perovskite LaFeO3 and
LaFe0.5Cr0.5O3",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "LaFeO3 and LaFe0.5Cr0.5O3"} {
"title": "Phys. Status Solidi RRL 2022,16, 2200023",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "none"
}
Processing Time: 4.89314603805542 seconds
Input Tokens: 8415
Output Tokens: 124
```

```
-----
PDF File: HCMO.pdf
Response: YES
```

Figure: Figure 1(a)

Material: Ho<sub>2</sub>CoMnO<sub>6</sub>

The magnetization vs. temperature (M-T) plot under various magnetic fields is shown in Figure 3(a). The plot exhibits a valley-like feature in the low-temperature region, indicating a negative magnetization. The negative magnetization is attributed to the 3d-4f negative exchange interaction between the Ho<sup>3+</sup> ions and the FM mesh of Co<sup>2+</sup>/Mn<sup>4+</sup> sublattices.

The magnetocaloric effect (MCE) is studied by measuring the temperature dependence of magnetization under various magnetic fields. The MCE is found to be significant, with a maximum value of 13.4 J/kg K at a low temperature and a magnetic field of 70 kOe. The MCE is attributed to the change in the magnetic entropy of the system upon the application of a magnetic field.

The magnetocaloric parameters, such as the relative cooling power (RCP) and the adiabatic temperature change ( $\Delta T_{ad}$ ), are also calculated. The RCP is found to be 12.5 J/kg K and the  $\Delta T_{ad}$  is found to be 4.2 K at a magnetic field of 70 kOe Here is the JSON response:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3(d)",
"material": "Ho2CoMnO6",
"material description": "The title compound is a rare-earth-based double
perovskite compound, Ho2CoMnO6, which exhibits a complex magnetic behavior."
} Here is the JSON output:
```

```
...
```

```
{
"title": "Structural, magnetic, and magnetocaloric properties of Ho2CoMnO6",
"Figure found": "YES",
"Figure": "Figure 7(b)",
"material": "Ho2CoMnO6"
}
...
```

```
Processing Time: 15.713869571685791 seconds
Input Tokens: 21154
Output Tokens: 405
-----
```

---

PDF File: 1-s2.0-S0109564113000067-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 2", "material": "Polymer-infiltrated-ceramic-network material (PICN)"}
```

The paper discusses the mechanical properties of a novel material called Polymer-infiltrated-ceramic-network material (PICN), which is a two-phase composite material composed of a ceramic network and a polymer network. The authors investigate the effects of varying ceramic precursor densities on the mechanical properties of the PICN, including flexural strength, elastic modulus, strain at failure, and hardness.

The paper presents the results of experiments on four different PICNs with varying ceramic precursor densities, as well as a dense ceramic and a pure polymer. The results show that the mechanical properties of the PICN are influenced by the ceramic precursor density, with lower ceramic fractions resulting in lower elastic modulus and hardness, but higher flexural strength and strain at failure.

The authors also present SEM images of the indented areas of the materials, which show that the PICN exhibits a unique microstructure with a continuous polymer network and a ceramic network. The results suggest that the PICN has potential for use in dental applications, as it exhibits mechanical properties similar to those of natural teeth. {

```
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 3",  
"material": "PICN (polymer-infiltrated ceramic network)"  
}
```

Processing Time: 11.327147006988525 seconds  
Input Tokens: 11623  
Output Tokens: 299

-----  
PDF File: zno-serie5.pdf  
Response: Here is the JSON output:

```
...  
{  
  "title": "Structural and magnetic characterization of (Zn 1-x-yFexCoy)O",  
  "Figure found": "YES",  
  "Figure": "Figure 2",  
  "material": "ZnO, Fe-doped ZnO, Co-doped ZnO"  
}  
...
```

Processing Time: 3.698740243911743 seconds  
Input Tokens: 7592  
Output Tokens: 75

-----  
PDF File: 1-s2.0-S1359645413004503-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 3", "material": "CoCrFeMnNi high-entropy alloy"} Here is the JSON output:
```

```
...  
{  
  "title": "An equiatomic CoCrFeMnNi high-entropy alloy",  
  "Figure found": "YES",  
}
```

---

```
"Figure": "Figure 12",
"material": "CoCrFeMnNi high-entropy alloy"
}
```

```
Processing Time: 5.287784576416016 seconds
Input Tokens: 14405
Output Tokens: 110
```

```
-----
PDF File: 1-s2.0-S1002007116300995-main.pdf
Response: The answer is:
```

```
{"title": "Structural effects on the magnetic hyperthermia properties of iron oxide-based NPs", "Figure found": "YES", "Figure": "Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5", "Material": "Iron oxide nanoparticles"} The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 5", "material": "iron oxide-based magnetic hyperthermia agents"}
```

The paper discusses the use of iron oxide-based magnetic hyperthermia agents for cancer treatment. The authors review the current state of the art in this field and discuss the potential benefits and challenges of using these agents. They also present some of their own research on the topic, including the development of new iron oxide-based magnetic hyperthermia agents and the testing of their effectiveness in treating cancer in animal models. {

```
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 103",
"material": "Fe3O4"
}
```

```
Processing Time: 9.752267360687256 seconds
Input Tokens: 17383
Output Tokens: 233
```

```
-----
PDF File: app135z2p11.pdf
```

```
Response: {
"title": "Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe70105xCoxMn10Mo5B15(0x0.5) Alloys at Low Magnetizing Field",
"Figure found": "YES",
"Figure": "Figure 1 and Figure 2",
"material": "amorphous Fe70105xCoxMn10Mo5B15(0x0.5) alloys"
}
```

```
Processing Time: 3.103688955307007 seconds
Input Tokens: 3375
Output Tokens: 97
```

```
-----
PDF File: 035220_1_online.pdf
```

```
Response: Here is the JSON output:
```

```
...
```

```
{
  "title": "Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity",
  "Figure found": "YES",
  "Figure": "Figure 4",
  "material": "Gd-Fe-Al"
}
```

```
Processing Time: 3.567056894302368 seconds
Input Tokens: 7390
Output Tokens: 80
```

---

-----  
PDF File: PhysRevB.81.214406.pdf  
Response: The answer is:

```
{"title": "Peculiarities of the magnetocaloric properties in Ni-Mn-Sn  
ferromagnetic shape memory alloys",  
"Figure found": "Yes",  
"Figure": "Figure 1",  
"Material": "Ni-Mn-Sn ferromagnetic shape memory alloy"} {  
"title": "PECULIARITIES OF THE MAGNETOCALORIC EFFECT PHYSICAL REVIEW B 81,  
214406 /H208492010 /H20850",  
"Figure found": "YES",  
"Figure": "Figure 1",  
"material": "MnFePd"  
}  
Processing Time: 6.603277921676636 seconds  
Input Tokens: 9218  
Output Tokens: 131  
-----
```

PDF File: Gottschall2020\_JAP\_Advanced characterization of multicaloric materials in pulsed magnetic fields.pdf  
Response: Here is the JSON output:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 5",  
"material": "Ni-Mn-Ti-Co"  
} {  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "FIG. 6",  
"material": "Heusler shape memory alloys"  
}  
Processing Time: 3.8675827980041504 seconds  
Input Tokens: 10443  
Output Tokens: 83  
-----
```

PDF File: Wehling\_2023\_Nichtwissen\_-\_ein\_ungewöhnlicher\_Schlüsselbegriff\_.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure  
3.1", "material": "FCKW (Fluor-Chlor-Kohlenwasserstoff)"}
```

The paper discusses the concept of "Nichtwissen" (ignorance) in the context of environmental sociology and its relevance to understanding environmental problems. The author argues that ignorance is not just a lack of knowledge, but a social construct that is shaped by various factors, including the way we perceive and define risk, uncertainty, and the limits of our knowledge.

The paper highlights the importance of considering ignorance in the context of environmental problems, such as climate change, and argues that it is not just a matter of individual ignorance, but rather a systemic issue that is rooted in the way we organize our societies and economies.

The author also discusses the concept of "wissenschaftliches Nichtwissen" (scientific ignorance), which refers to the ignorance that is generated by scientific research itself. This type of ignorance is not just a lack of

---

knowledge, but rather a result of the limitations and biases of scientific inquiry.

The paper concludes by arguing that ignorance is a key concept in understanding environmental problems and that it is essential to recognize its Here is the JSON output:

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "none"
}
...
```

The paper discusses the concept of ignorance and its relevance in the context of environmental issues, particularly in the context of climate change and sustainability. The author, Peter Wehling, argues that ignorance is not just a lack of knowledge, but a complex and multifaceted phenomenon that is deeply embedded in societal and cultural contexts. He also highlights the importance of acknowledging and addressing ignorance in order to develop more effective strategies for addressing environmental challenges.

The paper is a collection of essays that explore the concept of ignorance from various perspectives, including sociology, philosophy, and environmental studies. The authors discuss the ways in which ignorance is produced and reproduced in different contexts, and the ways in which it can be addressed and overcome.

The paper is a valuable contribution to the field of environmental studies, as it highlights the importance of acknowledging and addressing ignorance in order to develop more effective strategies for addressing environmental challenges. It also provides a nuanced and complex understanding of the concept of ignorance, and its relevance to environmental issues.

Processing Time: 17.833888053894043 seconds  
Input Tokens: 14995  
Output Tokens: 501

-----  
PDF File: Nyadjro\_et\_al\_2023\_The\_NOAA\_NCEI\_marine\_microplastics\_database.pdf  
Response: The answer is:

```
{"title":"The title of the paper","Figure found":"YES","Figure number":"1","Material":"marine microplastics"} Here is the JSON response:
```

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "Various materials"
}
```

```
` ` {"title":"Fellowship from the Gulf Research Program of the US National Academies of Sciences, Engineering, and Medicine","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 5.906141519546509 seconds  
Input Tokens: 15859  
Output Tokens: 115

---

PDF File:

Benningstad\_und\_Kunst\_2020\_Dissociating\_meat\_from\_its\_animal\_origins.pdf

Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "meat"} Here is the JSON output:
```

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 4.6",  
"material": "meat"  
Processing Time: 4.217400550842285 seconds  
Input Tokens: 13879  
Output Tokens: 78
```

-----  
PDF File: Gottschall2018\_A multicaloric cycle that exploits thermal hysteresis\_NatureMat.pdf

Response: The answer is:

```
{"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 4",  
"material": "Ni-Mn-In Heusler alloy"} {  
"title": "Large reversible magnetocaloric effect in Ni-Mn-In-Co",  
"Figure found": "YES",  
"Figure": "Figure 1",  
"material": "Ni50.0Mn35.5In14.5"  
}
```

```
Processing Time: 6.516801595687866 seconds  
Input Tokens: 8750  
Output Tokens: 103
```

-----  
PDF File:

Leslie\_et\_al\_2022\_Discovery\_and\_quantification\_of\_plastic\_particle.pdf

Response: The document does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, it does discuss the analysis of plastic particles in human blood and the development of a method to quantify the mass of five high production volume polymers applied in plastic materials: poly(methyl methacrylate) (PMMA), polypropylene (PP), materials containing polymerized styrene (PS), polyethylene (PE), and polyethylene terephthalate (PET).

The method uses double shot pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) to analyze the samples. The authors report the results of a recovery experiment, which showed that the method can recover the spiked polymers from blood samples with recoveries ranging from 68% to 114%. They also discuss the control of background contamination during sampling, sample preparation, and analysis.

The authors conclude that the method can detect and quantify the four polymers in blood samples with sufficient sensitivity and precision. The measured concentration data in blood were corrected for the average procedural blank for each analyte.

Here is the JSON response:

```
{  
"title": "The title of the paper",
```

---

```
"Figure found": "NO",
"Figure": "",
"material": " The answer is:
```

```
{"title":"The title of the paper","Figure found":"YES","Figure
number":"3.1","Material":"Human blood"} {"title":"Impact of microplastics and
nanoplastics on human health","Figure found":"NO","Figure":"","material":""}
Processing Time: 12.60486888885498 seconds
Input Tokens: 15870
Output Tokens: 315
```

```
-----
You are going to be given an example of a document. Please check this document
to see if it contains a figure that depicts a magnetization vs. temperature (M
vs T) PLOT. Keywords to look for include 'magnetization', 'temperature
dependence', 'Curie temperature', and 'magnetic susceptibility'. ANSWER ONLY
WITH 'YES' IF SUCH PLOTS ARE FOUND THEN STATE WHICH IS THE M VS T FIGURE if it
is AND THEN WHAT MATERIAL, OR SIMPLY 'NO' IF THERE ARE NONE, ANSWER ONLY WITH
a JSON return the following Json {"title":"The title of the paper";"Figure
found":"If a magnetization vs. temperature PLOT is found YES or
NO";"Figure":"which figure is the MAGNETIZATION VS. TEMPERATURE
PLOT";"material":"what is the material that is the plot"}
-----
```

```
PDF File:
Bommarito_et_al_2023_Temporal_trends_and_predictors_of_gestational.pdf
Response: The answer is:
```

```
{"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 2",
"material": "human biomarkers of exposure to organophosphate ester flame
retardants and plasticizers in urine from pregnant people"} The JSON output
is:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Fig. 2",
"material": "human urine samples"
} {
"title": "Wang et al. (2023)",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 7.283336162567139 seconds
Input Tokens: 18882
Output Tokens: 136
```

```
-----
PDF File:
Yang_et_al_2023_Detection_of_Various_Microplastics_in_Patients_Und.pdf
Response: Here is the JSON response:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Various polymers, including PVC, PET, and PMMA"
} {
"title": "The title of the paper",
```

```

"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.720438241958618 seconds
Input Tokens: 10276
Output Tokens: 77
-----
PDF File: onep.pdf
Response: {
"title": "Environment International 163 (2022) 107199",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 1.147451400756836 seconds
Input Tokens: 1452
Output Tokens: 33
-----
PDF File: BMS.pdf
Response: {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 3a",
"material": "La0.70Ca0.1Sr0.2Fe0.1Mn0.9O3 perovskite"
}
Processing Time: 2.292752981185913 seconds
Input Tokens: 4845
Output Tokens: 60
-----
PDF File: Tuneable_Giant_Magnetocaloric_Effect_in_MnFe2PSi_M.pdf
Response: Here is the JSON output:

{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "Mn,Fe)2(P,Si)-based materials"
} {
"title": "The magnetocaloric effect and magnetic refrigeration near room
temperature: Materials and models",
"Figure found": "YES",
"Figure": "Figure 4",
"material": "MnFe(P,Si,B) and MnFe(Ni,P,Si,B)"
}
Processing Time: 4.847777605056763 seconds
Input Tokens: 11212
Output Tokens: 110
-----
PDF File: 122405_1_online.pdf
Response: Here is the JSON response:

{
"title": "Giant magnetocaloric effect in isostructural MnNiGe-CoNiGe system",
"Figure found": "YES",
"Figure": "Figure 3",
"material": "MnNiGe-CoNiGe"
}
Processing Time: 2.7820096015930176 seconds

```

---

Input Tokens: 7672  
Output Tokens: 56

-----  
PDF File: PhysRevB.93.184403.pdf  
Response: The answer is:

```
{"title": "Realization of two-dimensional ferromagnetism with giant coercivity  
in ultrathin  $\beta$ -Ni(OH)2 layers grown on a MoS2 surface",  
"Figure found": "YES",  
"Figure": "Figure 4",  
"Material": " $\beta$ -Ni(OH)2"} {  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 8",  
"material": "MoS2/Ni(OH)2"  
}  
Processing Time: 5.184289455413818 seconds  
Input Tokens: 12780  
Output Tokens: 113
```

-----  
PDF File: ArticleSupra2020.pdf  
Response: Here is the JSON response:

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 2d",  
"material": "La0.7Ba0.15Ag0.15MnO3 manganite"  
} {  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Fig. 7",  
"material": "La0.7Ba0.15Ag0.15MnO3"  
}  
Processing Time: 7.593746185302734 seconds  
Input Tokens: 10009  
Output Tokens: 107
```

-----  
PDF File: Krishna\_Murthy\_2015\_J.\_Phys.\_D \_Appl.\_Phys.\_48\_355001.pdf  
Response: Here is the JSON output:

```
{  
"title": "Giant magnetocaloric effect in Gd2NiMnO6 and Gd2CoMnO6 ferromagnetic  
insulators",  
"Figure found": "YES",  
"Figure": "Figure 1(a) and (b)",  
"material": "Gd2NiMnO6 and Gd2CoMnO6"  
}  
Processing Time: 4.6754374504089355 seconds  
Input Tokens: 6793  
Output Tokens: 83
```

-----  
PDF File: Zinn - 2023 - Zweite Moderne und Risikogesellschaft.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure  
3", "material": "none"} Here is the JSON output:
```

```
{  
"title": "The title of the paper",
```

---

```
"Figure found": "NO",
"Figure": "",
"material": ""
}
```

```
Processing Time: 3.459108591079712 seconds
Input Tokens: 12942
Output Tokens: 62
```

```
-----
PDF File: Albert_und_Kleinman_2011_Bringing_Pierre_Bourdieu_to_Science_and.pdf
Response: The answer is:
```

```
{"title":"The title of the paper is 'Bringing Pierre Bourdieu to Science and
Technology Studies',"Figure found":"YES","Figure":"Figure 1","material":"The
paper discusses the application of Pierre Bourdieu's concepts to the field of
Science and Technology Studies (STS). It argues that Bourdieu's work can
provide insights into the operation of science in the contemporary world and
can be used to analyze the social dynamics of the scientific field."}
```

```
{"title":"Bourdieu: Critical perspectives","Figure
found":"NO","Figure":"","material":""}
```

```
Processing Time: 4.860419750213623 seconds
Input Tokens: 8648
Output Tokens: 120
```

```
-----
PDF File:
```

```
D'Angelo_und_Meccariello_2021_Microplastics_A_Threat_for_Male_Fertility.pdf
```

```
Response: The answer is:
```

```
{"title":"The title of the paper","Figure found":"YES","Figure number":"Figure
1","Material":"Not mentioned"} Here is the JSON output:
```

```
...
```

```
{
  "title": "Ma, L.Q. Molecular mechanisms of PFOA-induced toxicity in animals
and humans: Implications for health risks.",
  "Figure found": "NO",
  "Figure": "",
  "material": ""
}
```

```
...
```

```
Processing Time: 4.424508094787598 seconds
Input Tokens: 13286
Output Tokens: 90
```

```
-----
PDF File:
```

```
Paul_et_al_2022_Institutionalised_ignorance_in_policy_and_regulati.pdf
```

```
Response: The answer is:
```

```
{"title":"The title of the paper","Figure found":"YES","Figure":"Figure
1","material":"unknown"} {
```

```
"title": "Routledge International Handbook of Ignorance Studies",
```

```
"Figure found": "NO",
```

```
"Figure": "",
```

```
"material": ""
```

```
}
```

```
Processing Time: 3.0628509521484375 seconds
```

```
Input Tokens: 9626
```

```
Output Tokens: 61
```

```
-----
PDF File: 1-s2.0-S0304885312009754-main.pdf
```

---

Response: Here is the JSON output:

```
{
"title": "Phase diagram, ferromagnetic martensitic transformation and
magnetoresponse properties of Fe-doped MnCoGe alloys",
"Figure found": "YES",
"Figure": "Figure 2",
"material": "MnCoGe alloys"
}
Processing Time: 2.810117244720459 seconds
Input Tokens: 6637
Output Tokens: 61
```

-----  
PDF File: Kuhlicke - 2023 - Resilienz und Risiko.pdf  
Response: The JSON output is:

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 2.1",
"material": "none"
} {
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
Processing Time: 3.552651882171631 seconds
Input Tokens: 12243
Output Tokens: 67
```

-----  
PDF File: Wajcman - 2010 - Feminist theories of technology.pdf  
Response: The paper "Feminist Theories of Technology" by Judy Wajcman does not contain a figure that depicts a magnetization vs. temperature (M vs. T) plot. However, the paper does discuss the relationship between gender and technology, and how technology is shaped by and shapes gender relations.

The paper argues that feminist theories of technology have come a long way over the last quarter of a century, and that the intellectual exploration at the intersections of feminist scholarship and science and technology studies (STS) has enriched both fields immeasurably.

The paper highlights the importance of understanding the mutual shaping of gender and technology, and how the materiality of technology affords or inhibits the doing of particular gender power relations. It also emphasizes the need to involve women throughout the processes and practices of technological innovation, and to recognize the constitutive power of tools, techniques, and objects to materialize social, political, and economic arrangements.

The paper does not mention the concept of "material-semiotic practice" or "technoscience" explicitly, but it does discuss the social constructivist approach to technology, which views technology as a sociotechnical product that is shaped by social and cultural factors.

The paper does not mention the authors Manuel Castells, Nicholas Negroponte  
{"title":"The title of the paper":null,"Figure found":"NO","Figure":null,"material":null}  
Processing Time: 10.056691646575928 seconds

---

Input Tokens: 8825  
Output Tokens: 282

-----  
PDF File: Metal\_Substitution\_Effect\_on\_a\_Three-Dimensional\_C.pdf  
Response: Here is the JSON response:

```
...  
{  
  "title": "The title of the paper",  
  "Figure found": "YES",  
  "Figure": "Figure 5",  
  "Material": "Cyanido-bridged metal assemblies containing both Fe and Co  
ions"  
}  
... Here is the JSON output:  
...  
{  
  "title": "Inorganics 2017,5, 63",  
  "Figure found": "YES",  
  "Figure": "Figure 6",  
  "material": "K4[Nb(CN)8]2H2O, (Fe0.21Co0.79)2[Nb(CN)8](4-pyridinealdoxime)  
8zH2O"  
}  
...
```

Processing Time: 6.898503065109253 seconds  
Input Tokens: 14082  
Output Tokens: 161

-----  
PDF File: JAP115123908.pdf  
Response: The magnetization vs. temperature (M vs. T) plot is found in the  
article. Therefore, the answer is:

**\*\*YES\*\***

The M vs. T plot is Figure 4 in the article.

**\*\*Material:\*\*** BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composite

**\*\*Figure:\*\*** Figure 4

**\*\*Description:\*\*** The figure shows the field-cooled magnetization (M) as a  
function of temperature (T) for the BaTiO<sub>3</sub>-NiFe<sub>2</sub>O<sub>4</sub> composite. The  
magnetization decreases with increasing temperature, and the Curie temperature  
(T<sub>c</sub>) is observed to decrease with increasing BaTiO<sub>3</sub> content. Here is the JSON  
output:

```
...  
{  
  "title": "Synthesis of (100-x)BTO-(x)NFO solid solution at high  
temperature",  
  "Figure found": "YES",  
  "Figure": "Figure 5",  
  "material": "NFO and BTO"  
}  
...
```

Processing Time: 22.981261491775513 seconds  
Input Tokens: 14291  
Output Tokens: 196

---

-----  
PDF File: Physica Rapid Research Ltrs - 2022 - Xia - Magnetic Properties and Mössbauer Study of Perovskite LaFeO<sub>3</sub> and LaFe<sub>0.5</sub>Cr<sub>0.5</sub>O<sub>3</sub>.pdf  
Response: The answer is:

```
{"title": "Magnetic Properties and Mössbauer Study of Perovskite LaFeO3 and LaFe0.5Cr0.5O3", "Figure found": "YES", "Figure": "Figure 1", "material": "LaFeO3 and LaFe0.5Cr0.5O3"} {  
"title": "Phys. Status Solidi RRL 2022,16, 2200023",  
"Figure found": "YES",  
"Figure": "Figure 1",  
"material": "none"  
}
```

Processing Time: 4.796543598175049 seconds  
Input Tokens: 8415  
Output Tokens: 121

-----  
PDF File: HCMO.pdf  
Response: The answer is:

```
{"title": "Role of 3d-4f exchange interaction and local anti-site defects in the magnetic and magnetocaloric properties of double perovskite Ho2CoMnO6 compound", "Figure found": "YES", "Figure": "Figure 2", "Material": "Ho2CoMnO6"} Here is the JSON response:
```

```
{  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 3(d)",  
"material": "Ho2CoMnO6",  
"parameters": {  
"jΔSjmax": 13.44 J/kg K,  
"RCP": 272 J/kg,  
"TEC": 6.5 (3.9) J/kg K,  
"NRC": 1.5 (2.1)  
}  
} {  
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 7(b)",  
"material": "Ho2CoMnO6"  
}
```

Processing Time: 9.795434474945068 seconds  
Input Tokens: 21154  
Output Tokens: 218

-----  
PDF File: 1-s2.0-S0109564113000067-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 2", "material": "Polymer-infiltrated-ceramic-network material (PICN)"}
```

The paper discusses the mechanical properties of a novel material called Polymer-infiltrated-ceramic-network material (PICN), which is a two-phase composite material composed of a ceramic network and a polymer network. The authors investigate the effects of varying ceramic precursor densities on the mechanical properties of the PICN, including flexural strength, elastic modulus, strain at failure, and hardness.

---

The paper presents a comprehensive study of the mechanical properties of PICNs, including the effects of ceramic precursor density on the material's properties. The results show that the flexural strength of the PICN increases with decreasing ceramic precursor density, while the elastic modulus and hardness increase with increasing ceramic precursor density. The strain at failure of the PICN also decreases with increasing ceramic precursor density.

The authors also present scanning electron microscopy (SEM) images of the indented areas of the PICN, which show that the material exhibits a unique microstructure with a continuous polymer network and a ceramic network. The SEM images also show that the material exhibits a high degree of toughness and resistance to {

```
"title": "The title of the paper",
"Figure found": "NO",
"Figure": "",
"material": ""
}
```

```
Processing Time: 10.908538818359375 seconds
Input Tokens: 11623
Output Tokens: 284
```

-----  
PDF File: zno-serie5.pdf  
Response: Here is the JSON output:

```
...
{
  "title": "Structural and magnetic characterization of (Zn 1-x-yFexCoy)O",
  "Figure found": "YES",
  "Figure": "Figure 2",
  "material": "ZnO, Fe-doped ZnO, Co-doped ZnO"
}
...
```

```
Processing Time: 3.6972219944000244 seconds
Input Tokens: 7592
Output Tokens: 75
```

-----  
PDF File: 1-s2.0-S1359645413004503-main.pdf  
Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 9", "material": "CoCrFeMnNi high-entropy alloy"} Here is the JSON output:
```

```
...
{
  "title": "An equiatomic CoCrFeMnNi high-entropy alloy was produced by arc melting, drop casting and rolling, after which it was recrystallized to produce single-phase, fcc-structured, equiaxed microstructures with three different grain sizes, 4.4, 50 and 155 nm.",
  "Figure found": "YES",
  "Figure": "Figure 11",
  "material": "CoCrFeMnNi high-entropy alloy"
}
...
```

```
Processing Time: 6.863595485687256 seconds
Input Tokens: 14405
Output Tokens: 162
```

-----  
PDF File: 1-s2.0-S1002007116300995-main.pdf  
Response: The answer is:

---

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "Iron oxide nanoparticles"} The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 5", "material": "iron oxide-based magnetic hyperthermia agents"}
```

The paper discusses the use of iron oxide-based magnetic hyperthermia agents for cancer treatment. The authors review the current state of the art in this field and discuss the potential benefits and challenges of using these agents. They also present some of their own research on the topic, including the development of new iron oxide-based magnetic hyperthermia agents and their testing in animal models.

The figure that is referred to is Figure 5, which shows the results of the authors' experiments on the heating efficiency of different iron oxide-based magnetic hyperthermia agents. The figure shows that the agents with a higher concentration of iron oxide particles have a higher heating efficiency.

The material that is being discussed is iron oxide-based magnetic hyperthermia agents, which are used to generate heat in the body to treat cancer. {

```
"title": "The title of the paper",  
"Figure found": "YES",  
"Figure": "Figure 103",  
"material": "Fe3O4"  
}
```

```
Processing Time: 11.12071442604065 seconds  
Input Tokens: 17383  
Output Tokens: 275
```

-----  
PDF File: app135z2p11.pdf

```
Response: {  
"title": "Qualitative Analysis of Magnetization versus Temperature Curves in Amorphous Fe70117xCoxMn10Mo5B15(0x0.5) Alloys at Low Magnetizing Field",  
"Figure found": "YES",  
"Figure": "Figure 1 and Figure 2",  
"material": "amorphous Fe70117xCoxMn10Mo5B15(0x0.5) alloys"  
Processing Time: 3.0970208644866943 seconds  
Input Tokens: 3375  
Output Tokens: 97
```

-----  
PDF File: 035220\_1\_online.pdf

Response: Here is the JSON output:

```
{  
"title": "Magnetocaloric effect in high Gd content Gd-Fe-Al based amorphous/nanocrystalline systems with enhanced Curie temperature and refrigeration capacity",  
"Figure found": "YES",  
"Figure": "Figure 4",  
"material": "Gd-Fe-Al"  
}
```

```
Processing Time: 3.3428659439086914 seconds  
Input Tokens: 7390  
Output Tokens: 73
```

-----  
PDF File: PhysRevB.81.214406.pdf

Response: The answer is:

---

```
{"title": "Peculiarities of the magnetocaloric properties in Ni-Mn-Sn
ferromagnetic shape memory alloys", "Figure found": "YES", "Figure": "Figure
1", "Material": "Ni-Mn-Sn ferromagnetic shape memory alloy"} {
"title": "Peculiarities of the Magnetocaloric Effect",
"Figure found": "YES",
"Figure": "Figure 1",
"material": "MnFePGe"
}
```

```
Processing Time: 5.836395263671875 seconds
Input Tokens: 9218
Output Tokens: 105
```

```
-----
PDF File: Gottschall2020_JAP_Advanced characterization of multicaloric
materials in pulsed magnetic fields.pdf
Response: Here is the JSON output:
```

```
{
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "Figure 5",
"material": "Ni-Mn-Ti-Co Heusler alloy"
} {
"title": "The title of the paper",
"Figure found": "YES",
"Figure": "FIG. 6",
"material": "Heusler shape memory alloys"
}
```

```
Processing Time: 3.999363660812378 seconds
Input Tokens: 10443
Output Tokens: 87
```

```
-----
PDF File: Wehling_2023_Nichtwissen_-_ein_ungewöhnlicher_Schlüsselbegriff_.pdf
Response: The answer is:
```

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "3.1",
"material": "none"} Here is the JSON output:
```

```
...
{
  "title": "The title of the paper",
  "Figure found": "YES",
  "Figure": "Figure 1",
  "material": "none"
}
...
```

The paper discusses the concept of ignorance and its relevance in the context of environmental issues, particularly in the field of climate change. The author, Peter Wehling, argues that ignorance is not just a lack of knowledge, but a complex and multifaceted phenomenon that can have significant consequences for our understanding of the world and our ability to make informed decisions.

The paper explores the concept of ignorance in various contexts, including the history of climate change research, the role of ignorance in the development of scientific knowledge, and the implications of ignorance for environmental policy and decision-making. The author also discusses the importance of acknowledging and addressing ignorance in order to promote more effective and sustainable solutions to environmental problems.

---

The paper is written in a clear and accessible style, making it suitable for readers with a general interest in environmental issues and the social sciences.

Processing Time: 9.742598533630371 seconds

Input Tokens: 14995

Output Tokens: 249

-----  
PDF File: Nyadjro\_et\_al\_2023\_The\_NOAA\_NCEI\_marine\_microplastics\_database.pdf

Response: \*\*JSON Response:\*\*

```
```\n{\n  "title": "The NOAA NCEI global marine microplastics database",\n  "Figure found": "YES",\n  "Figure": "Figure 1, Figure 2, Figure 3, Figure 4",\n  "Material": "Marine microplastics"\n}
```

```\nHere is the JSON response:

```\n

```
{\n  "title": "The title of the paper",\n  "Figure found": "YES",\n  "Figure": "Figure 1",\n  "material": "NO"\n}
```

```
```\n{"title":"Fellowship from the Gulf Research Program of the US National Academies of Sciences, Engineering, and Medicine","Figure found":"NO","Figure":"","material":""}
```

Processing Time: 7.161860466003418 seconds

Input Tokens: 15859

Output Tokens: 151

-----  
PDF File:

Benningstad\_und\_Kunst\_2020\_Dissociating\_meat\_from\_its\_animal\_origins.pdf

Response: The answer is:

```
{"title": "The title of the paper", "Figure found": "YES", "Figure": "Figure 1", "material": "meat"} Here is the JSON output:
```

```
{\n  "title": "The title of the paper",\n  "Figure found": "YES",\n  "Figure": "Figure 4.6",\n  "material": "meat"\n}\nProcessing Time: 4.211366891860962 seconds\nInput Tokens: 13879\nOutput Tokens: 78
```

-----  
PDF File: Gottschall2018\_A multicaloric cycle that exploits thermal hysteresis\_NatureMat.pdf

Response: The answer is:

```
{"title": "The title of the paper",\n"Figure found": "YES",\n"Figure": "Figure 4",\n"material": "Ni-Mn-In Heusler alloy"} {\n"title": "Large reversible magnetocaloric effect in Ni-Mn-In-Co",\n"Figure found": "YES",
```

---

```
"Figure": "Figure 1",
"material": "Ni50.0Mn35.5In14.5"
}
```

```
Processing Time: 6.543430328369141 seconds
Input Tokens: 8750
Output Tokens: 103
```

-----  
PDF File:

```
Leslie_et_al_2022_Discovery_and_quantification_of_plastic_particle.pdf
Response: The document does not contain a figure that depicts a magnetization
vs. temperature (M vs. T) plot. Therefore, the answer is:
```

```
{"title": "The title of the paper", "Figure found": "NO", "Figure": "",
"material": ""}
```

However, the document does discuss the analysis of plastic particles in human blood, including the development of a method to quantify the mass of five high production volume polymers applied in plastic materials: poly(methyl methacrylate) (PMMA), polypropylene (PP), materials containing polymerized styrene (PS), polyethylene (PE), and polyethylene terephthalate (PET). The method involves double shot pyrolysis-gas chromatography/mass spectrometry (Py-GC/MS) and is used to measure the mass concentrations of these polymers in human blood. The answer is:

```
{"title":"The title of the paper is not provided","Figure found":"YES","Figure
number":"3.1","Material":"Human blood"} {"title":"Screening criteria for
microplastic particles in air and atmospheric deposition: critical review and
applicability towards assessing human exposure","Figure
found":"NO","Figure":"","material":""}
```

```
Processing Time: 10.34526515007019 seconds
Input Tokens: 15870
Output Tokens: 249
```

-----

UNIVERSITÉ CATHOLIQUE DE LOUVAIN  
École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | [www.uclouvain.be/epl](http://www.uclouvain.be/epl)