Application of digitization and large language models for digital twins in permanent magnet materials

Amrita Khan[1], Aaron Dextre[1], Theodore Park[2], Imants Dirba[1], Hongbin Zhang[2], Oliver Gutfleisch[1]
[1]*Functional Materials, Institute of Materials Science, Technical University of Darmstadt, 64287 Darmstadt, Germany*

[2]*Theory of Magnetic Materials, Institute of Materials Science, Technical University of Darmstadt, Darmstadt, 64287, Germany*

Correspondence to: Imants Dirba imants.dirba@tu-darmstadt.de

1. Introduction

The rapid growth of permanent magnet applications, especially in electric vehicles and renewable energy technologies, has intensified the demand for high-performance permanent magnets, particularly Nd-Fe-B sintered magnets [1–3], requiring advanced digital tools for their characterization, properties prediction, and optimization. Large language models (LLMs) offer the opportunity to locate relevant plots in journals and papers as well as extract the relevant context, improving the process of data collection for machine learning (ML). Combined with an automated digitization, this process allows for the creation of a database which is then utilized for machine learning, creating a digital twin that accurately predicts the properties of permanent magnets under real-life conditions (the physical twin). This work investigates the use of plot digitization and LLMs in the development of a digital twin framework for Nd-Fe-B permanent magnets.

2. Methodology

Magnetization curves from literature, own experimental measurements, and commercial magnets manufacturers' datasheets were used in this study. Different digitization methods such as ChatGPT-4 [4], WebPlot digitizer [5], curve [6] and Plot2Spectra [7] were examined for their potential suitability for manual and automated digitization of the plots.

3. Results and Discussion

Accurate and automatized data digitization proved to be very challenging. In order to test the different digitization methods, they were fed a set of plot images from literature. An example of a thermomagnetic curve from [8] is shown in Fig. 1. After examining different digitization methods, DePlot and Plot2Spectra were the most promising for automatized and PlotDigitizer for manual data extraction. DePlot provides decent digitization of plots with an example provided in Fig. 2. As can be seen from this, DePlot is evidently designed for bar graphs as it will only measure points that are labelled 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 works well compared to the other methods.
After Deplot Plot2spectra was investigated and deemed the most promising. It is based on the optical flow, utilizes a branched ERFNet neural network to separate the lines from the background and one another, as well as a EfficientDet model, a CNN to determine the plot's bounding box. This model thus provides excellent line detection and decent performance when given multiple lines.

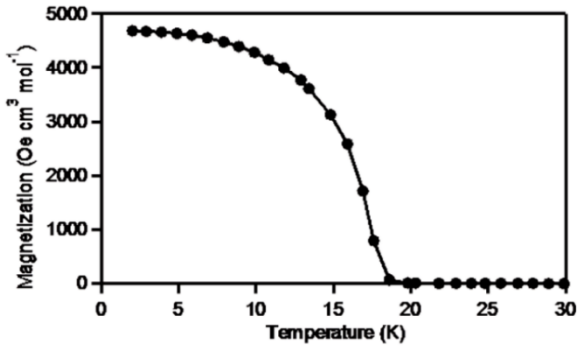

Fig. 1: Magnetization as a function of temperature plot used for testing different digitization methods.

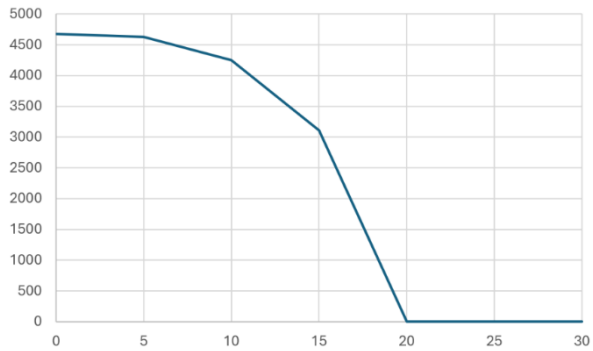

Fig. 2: Plot of the extracted data shows that the applied digitization approach using DePlot is satisfactory.

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. These challenges, as well as other practical examples on M(H) demagnetization curves and data collection, are explored and will be presented.

Acknowledgments

This work is financially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project ID No. 405553726, TRR 270

References

[1] O. Gutfleisch, M.A. Willard, E. Brück, C.H. Chen, S.G. Sankar, J.P. Liu, Magnetic Materials and Devices for the 21st Century: Stronger, Lighter, and More Energy Efficient, Advanced Materials 23 (2011) 821–842.

[2] A. Durgun, D. Ohmer, M. Katter, A. Thul, S. Steentjes, H.S. Amin, O. Gutfleisch, I. Dirba, Toward Maximum Utilization of Heavy Rare Earths in Sintered Nd–Fe–B Magnets by Grain Boundary Diffusion Source and Application Area Optimization, Adv Eng Mater (2025) e202501145.

[3] A. Durgun, I. Dirba, K. Opelt, C.-C. Lin, J. Gassmann, O. Gutfleisch, Combining grain boundary diffusion and segmentation: A novel production route for resource-efficient Nd–Fe–B magnets, Scripta Materialia 275 (2026) 117164.

[4] openAI, 2024, GPT-4, https://openai.com/index/gpt-4/

[5] A. Rohatgi, 2024, WebPlotDigitizer, https://automeris.io/docs/#cite-webplotdigitizer

[6] S. Zhou, J. Lu, An Open Digitization Tool for Extracting Scientific Curve Data in Portable Documents, in: C.-H. Chen (Ed.), Advances in Transdisciplinary Engineering, IOS Press, 2022.

[7] I. W. Jiang, K. Li, T. Spreadbury, E. Schwenker, O. Cossairt, M.K.Y. Chan, Plot2Spectra: an automatic spectra extraction tool, Digital Discovery 1 (2022) 719–731.

[8] K. Imoto, S. Takano, S. Ohkoshi, Metal Substitution Effect on a Three-Dimensional Cyanido-Bridged Fe Spin-Crossover Network, Inorganics 5 (2017) 63. https://doi.org/10.3390/inorganics5040063.