



## OPEN ACCESS

## EDITED BY

Alessandro Maffini,  
Polytechnic University of Milan, Italy

## REVIEWED BY

Riccardo Rossi,  
University of Rome Tor Vergata, Italy

## \*CORRESPONDENCE

R. Michael Churchill,  
✉ rchurchi@pppl.gov

RECEIVED 20 November 2024

ACCEPTED 12 December 2024

PUBLISHED 10 February 2025

## CITATION

Churchill RM (2025) AI foundation models for  
experimental fusion tasks.

*Front. Phys.* 12:1531334.

doi: 10.3389/fphy.2024.1531334

## COPYRIGHT

© 2025 Churchill. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC  
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# AI foundation models for experimental fusion tasks

R. Michael Churchill\*

Princeton Plasma Physics Laboratory, Princeton, NJ, United States

Artificial Intelligence (AI) foundation models, while successful in various domains of language, speech, and vision, have not been adopted in production for fusion energy experiments. This brief paper presents how AI foundation models can be used for fusion energy diagnostics, enabling, for example, visual automated logbooks to provide greater insights into chains of plasma events in a discharge, in time for between-shot analysis.

## KEYWORDS

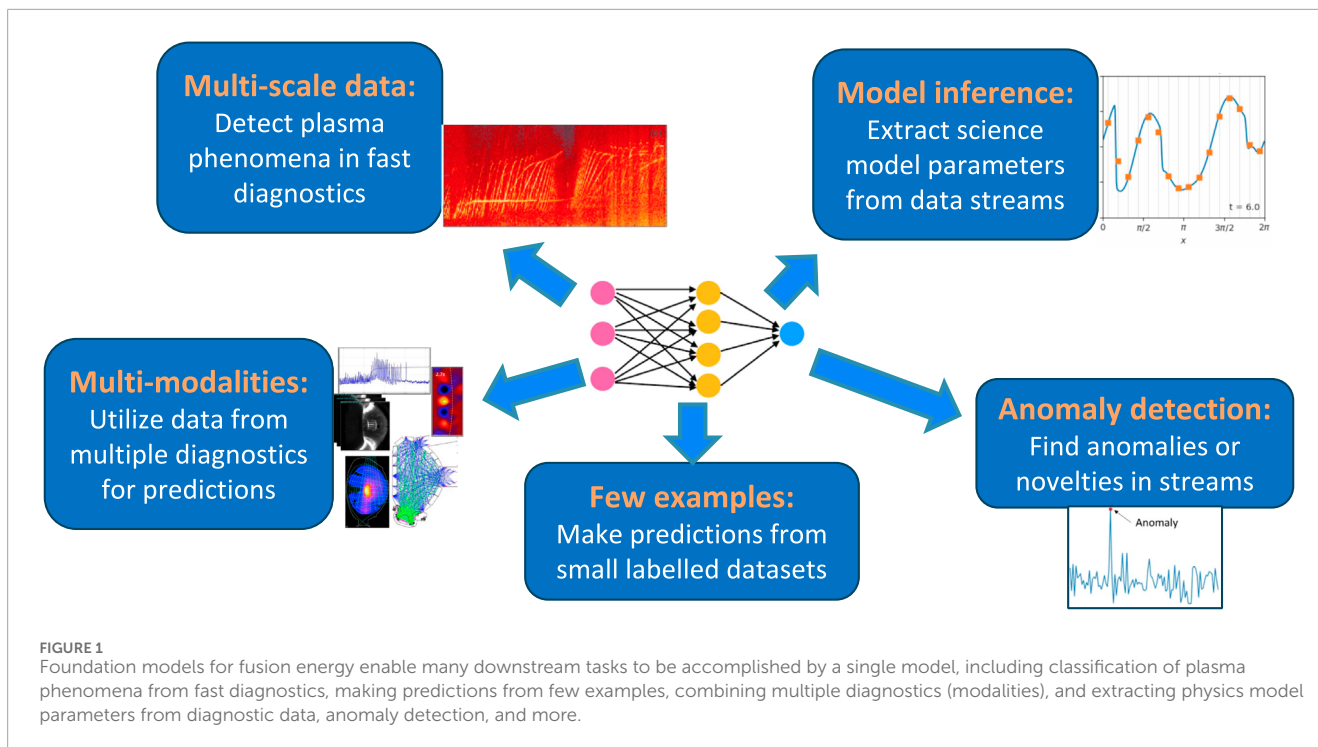
fusion energy, artificial intelligence, machine learning, foundation models, diagnostic

## 1 Introduction

AI foundation models [1] encapsulate a concept wherein an AI model is pre-trained in an unsupervised or self-supervised manner with a fundamental task, for example, predicting the next word in a sentence, on a wide range of data, and the trained model subsequently serves as a foundation to fine-tune the pre-trained foundation model for more detailed downstream tasks, for example, sentence generation, text summary, machine translation, etc. Essentially, instead of being a narrow expert, they are generalists. Although the concept of these models gained popularity with large language models (LLMs), such as those underlying ChatGPT [2], in principle, similar techniques can be utilized across a range of modalities, for example, images, audio, video, unstructured meshes, etc. Given the plethora of data on different modalities in experimental magnetic confinement fusion devices and the wide variety of tasks experimental fusion scientists need to perform, a natural question arises on whether AI foundation models can be created for experimental fusion data to enhance and accelerate fusion science. This paper seeks to explain at a conceptual level how these foundation models could be created and how they could effectively be used in experimental fusion settings.

## 2 Foundation models for fusion energy experiments

Currently, when AI/machine learning (ML) is used for tasks within fusion energy experiments, most often, the focus is on bespoke solutions for a particular task. These bespoke solutions require a lot of work from the practitioner, in gathering data, cleaning data, often performing data reductions (i.e., feature engineering), labeling data for classification problems, etc. The targeted tasks range widely, including models created specifically for anomaly detection [3], classification of plasma events [4–7], and time-series semantic search [8]. Figure 1 shows a representation of a foundation model that would instead serve the basis for these many tasks and more, reducing substantially the burden for repeating many of the steps for custom bespoke solutions. However, the question arises as to what this foundation model is and how this is achieved.



For LLMs, one of the more popular foundation models is the generative pre-trained transformer (GPT) [9]. This is a decoder-only transformer [10], pre-trained for next-token prediction (where tokens are created by splitting the text into a fixed vocabulary size of subwords, usually on order  $\sim 100k$ ). Mathematically, GPT model next-token prediction training maximizes the log-likelihood:

$$\max_{\theta} \sum_{i=1} \log p_{\theta}(s_i | s_{i-1}, s_{i-2}, \dots, s_{i-k+1}, s_{i-k}),$$

where  $p_{\theta}$  is the probability density represented by the transformer neural network with the parameter  $\theta$  and the token represented by  $s_j$ , with  $k$  being the context length of tokens provided to the model to make the next-token prediction of  $s_i$ . After training this model with gradient descent, the trained model can be used for various tasks, by, for example, adding a learnable layer at the end of the model and fine-tuning for supervised classification problems [9]. One of the most impactful findings with these trained models is “in-context learning” [11], in which a few examples can be input to the model and, without any additional training or fine-tuning, have the model complete a similar task (for example, provide example pairs of word translation from English -> Spanish and an empty word to translate, for example, provide as input to the model “dog -> perro, cat -> gato, bird ->”, and the model outputs “pajaro”). The increased in-context learning performance of GPT models with the size of the model (and size of data trained on) has enabled these models to be of general purpose (able to perform many different tasks) and led directly to the success of ChatGPT.

In experimental fusion energy sciences, the data are fundamentally different from text, in the first place being continuous instead of discrete but also consists of hundreds of different diagnostic data modalities, ranging from simple time series to more complex multi-channel, line-integrated 2d spatial videos. The time-series nature of the data maps well onto foundation models

created for audio or music [12], where their typical downstream tasks are speaker identification, automatic speech recognition, music generation, etc. Typically, to train these models, the self-supervised learning objective differs from the discrete language case since the continuous nature of the time series is a large space to attempt the next-token prediction. Instead, often, contrastive learning for self-supervised training is used, where a time series sequence is partially masked, and the model learns to predict this masked portion by discerning from a set, including the true sequence and many negative or false sequence samples:

$$\mathcal{L} = -\log \frac{\exp(\text{sim}(c_i, q_i^+)/\tau)}{\sum_{q \in Q} \exp(\text{sim}(c_i, q)/\tau)}$$

where  $\text{sim}(a, b) = a^T b / \|a\| \|b\|$  is the cosine similarity,  $c_i$  is the model output predicted sequence,  $q_i^+$  is the true sequence (quantized to ease learning),  $\tau$  is a modifiable temperature parameter, and  $Q = \{q_1^-, q_2^-, \dots, q_N^-, q_i^+\}$  is a set including the true sequence and a number of false sequences to discern between. Contrastive losses are more suitable for situations with continuous valued sequences. With the pre-trained model, a similar path can be followed for fine-tuning the models using a few specific labeled examples for supervised learning tasks like classification.

It should be noted here that while LLMs based on next-token prediction loss have been useful for both generative and discriminative downstream tasks, often, foundation models for audio or time-series have been focused on one set or the other (generative or discriminative downstream tasks). Figure 1 focuses on discriminative downstream tasks (e.g., classifying plasma modes in diagnostic data), but it should be noted that there are generative tasks that can be useful in fusion energy, such as scenario planning. Many foundation models for modalities like audio focused on

generative tasks use diffusion or flow matching models [13], although they are not studied here.

AI foundation models can be created for single diagnostics; however, AI model architectures exist to incorporate multiple modalities [14–16], thereby taking advantage of the correlations between modalities. For fusion experiments, this is particularly useful as information in, for example, the electron cyclotron emission imaging (ECEI) diagnostic and the beam emission spectroscopy (BES) diagnostic, measures different physical phenomena, and combining the data for predictions will potentially provide greater information than the sum of its parts.

Because AI foundation models are pre-trained to effectively learn the underlying data distribution, it is observed that large parameter models pre-trained on large amounts of unlabeled data perform better [11]. As a result, the consequence is that large high-performance computing (HPC) resources with many GPUs are needed to train these models. With the popularity of deep learning and foundation models, many good frameworks and tools are available to make this easier, including PyTorch, Hugging Face Accelerate, and MetaFAIR library.

### 3 Automated logbook

One relevant example of how to use such an AI foundation model for fusion energy experiment is shown in the automated logbook example in Figure 2. Fusion energy researchers have a deluge of data to process and understand from experiment, on short timescales between experimental discharges (usually 10–20 min) and longer timescales of months to years for understanding campaign-level data. Insights, if recorded, are normally formulated as text into personal or online logbooks. This manual analysis can be laborious. The AI foundation model could be used to automatically tag plasma events of interest in the diagnostic data, creating a metadata database and enabling fast visualization of plasma event sequences between plasma discharges.

As shown in Figure 2, first, a large dataset of raw diagnostic data from many plasma discharges is gathered, without having to label specific plasma events in the data. The AI foundation model is pre-trained on this data, passing in the sequences of data and using a contrastive loss to learn to predict masked portions of the sequence (the model shown is based on the wav2vec 2.0 model [17], with a convolutional neural network (CNN) encoder to reduce the data to a latent space representation, followed by a transformer model [10]). In the second step, a small dataset is gathered and labeled at time slices with a specific plasma event or mode, for example, neoclassical tearing modes (NTMs), Alfvén eigenmodes (AEs), edge harmonic oscillation (EHO), etc. The fine-tuning of the model can be to predict a single type of plasma event or different types of events. The size of this labeled dataset is smaller than would be required when training a model directly in a traditional supervised learning fashion since the pre-trained model has learnt good representations of the underlying data distribution. The size of this labeled dataset in principle can be as little as one or a few examples but, in practice, may require more and is problem-dependent. A decoder layer with learnable parameters is added onto the end of the pre-trained model, and with the labeled dataset, the model is fine-tuned to output predicted labels based on an input sequence. This fine-tuning

can involve only updating the decoder layer learnable parameters and retaining the rest of the pre-trained model parameters frozen, or unfreezing various layers of the pre-trained model and having those parameters also updated by the learning process. This fine-tuning needs to be done once, and then, the model is used for inference (in machine learning parlance prediction *versus* learning). As new plasma discharges are completed, the fine-tuned model takes in the new diagnostic data and predicts labels for the various plasma events. In the final step shown in Figure 2, these predictions can be visualized with the data in the automated logbook, for fast feedback to fusion researchers between plasma discharges and further investigation later. Detected modes can also trigger further analysis, for example, bandpass filtering on the detected mode frequencies, and visualizing the resulting spatial model structure in different diagnostics.

Although bespoke AI models could be created for each diagnostic or each plasma event, the traditional supervised learning route would almost surely require thousands of labeled examples gathered by researchers, a long tedious process often avoided. The AI foundation model offers a route where fewer labeled examples are needed. The foundation model can be fine-tuned for different plasma events. This enables identification of chains of events often important for understanding phenomena such as disruptions [18, 19].

Foundation models do require a large unlabeled dataset, and there are no well-defined rules for its size (this is dependent on the variety of the data and information content per sample, which may be hard to quantify). For many fusion energy experiments, substantial data can be available, depending on the device and diagnostic. An example of the largest diagnostic datasets on the DIII-D tokamak is shown in Table 1 (there are a total of 60 different diagnostic systems on DIII-D), showing a substantial amount of data available that can reasonably be expected to be sufficient for the purpose of training an AI foundation model.

In addition to the need for sizeable information-rich data to train on, out-of-distribution (OOD) data during inference need to be considered. Fusion experiments often push the boundaries to new areas, resulting in diagnostic data that may be far from that seen previously. Various works have approached this topic in bespoke AI models for fusion energy, seeking to enable models to adapt to new datasets [5, 20, 21]. In the context of AI foundation models, there are some indications in other fields, such as medical imaging, that foundation models are more robust to data distribution shift [22], even being useful to discriminate OOD data [23]. However, this needs to be researched in the specific context of AI foundation models for fusion energy diagnostic data.

### 4 Discussion

AI foundation models could serve to simplify and greatly expand the use of AI in experimental fusion energy. The ability to create good latent space representations of diagnostic data can aid in a number of downstream tasks for experimental fusion scientists, such as identification of plasma phenomena across multiple diagnostics, anomaly detection, extracting physics parameters from data, and use in control systems. The automation of these tasks leads to remarkable opportunities to gain further insights across many plasma discharges and uncover hidden relationships. Foundation models also ease

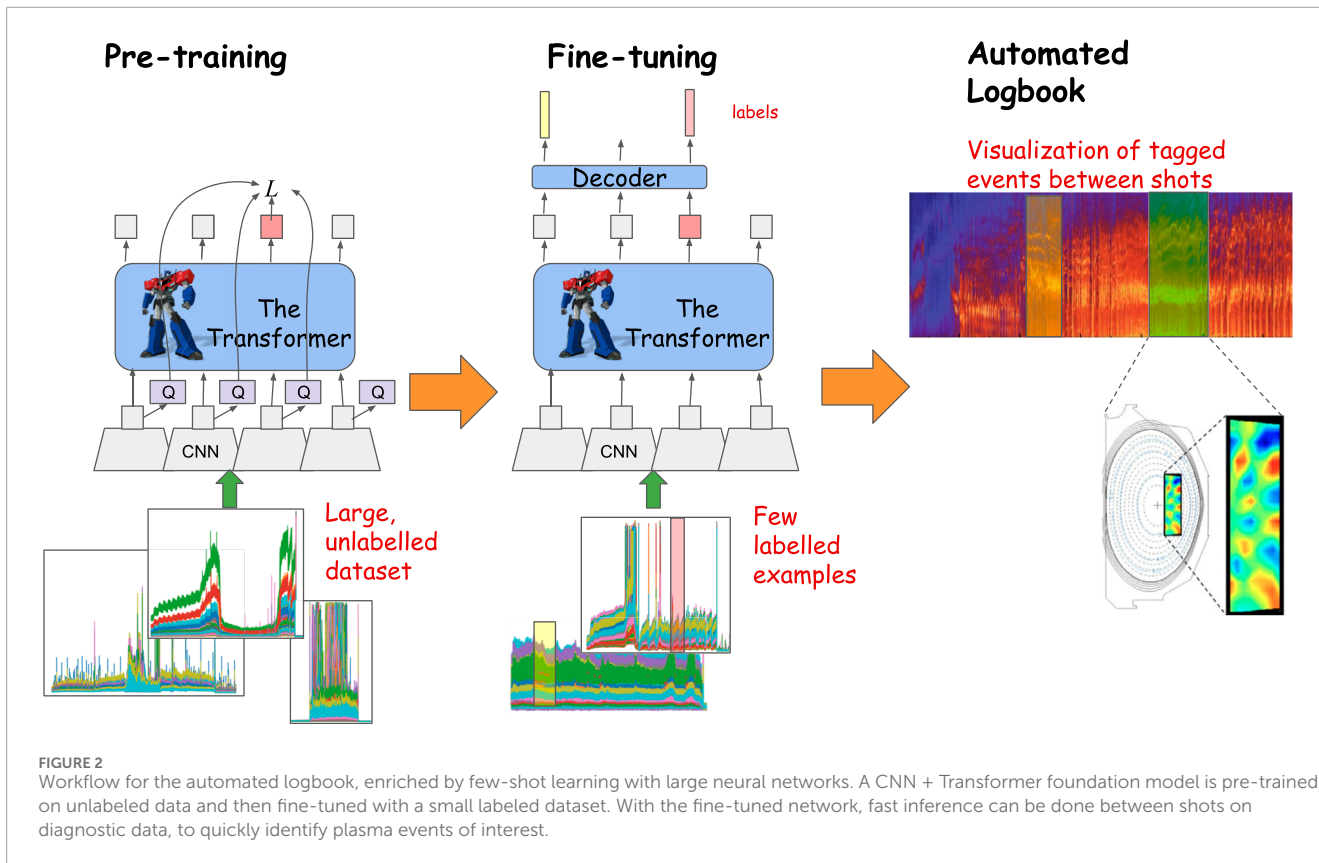


TABLE 1 Diagnostics on the DIII-D tokamak with the largest dataset sizes. Note that not all of these data are for overlapping plasma discharges (i.e., some plasma discharges will not have all of these diagnostics available).

Diagnostic	Spatial	Temporal	Total size [TB]
Fast camera	512 × 368	10 kHz	30.8
Helicon antenna cameras	-	-	8.6
IR TV camera	3 × (640 × 512; 464 × 4)	125 Hz; 12 kHz	10.1
Tangential viewing visible light camera	512 × 512	31 Hz	2.7
Beam emission spectroscopy	8 × 8	1 MHz	14.2
Electron cyclotron emission imaging	2 × 20 × 8	1 MHz	60.0
Ultra-fast charge exchange recombination spectroscopy (UF-CHERS)	16	1 MHz	1.7
Fast soft X-ray imaging	100	1 MHz	10.7

the burden on scientists from identifying and labeling thousands of examples for AI models, to a much more manageable level. Some work toward foundation models for fusion energy diagnostics has begun, for example, through the ExaLearn project, which was part of the Exascale Computing Project [Rodriguez et al., 2024 (unpublished study)], EUROfusion projects [24], and multi-modal bespoke models[25, 26], but until now, the full realization of AI foundation models as a production-ready tool in experimental fusion science has not been realized.

Although the focus of this paper has been foundation models for multi-modal time-series-based diagnostics, the advent of reasoning models such as the OpenAI o1 model [27] presents an opportunity to combine these in a hybrid system of AI agents, which can leverage these multi-modal time-series foundation models as tools to further automate discovery and utility of the investment in these experimental devices, including coupling with simulation. Creating these flexible building blocks of multi-modal time-series foundation models, to build these advanced workflows, could greatly aid fusion

energy scientists ultimately toward the realization of fusion energy as a clean and sustainable energy source.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

## Author contributions

RC: writing—original draft and writing—review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the US Department of Energy under DE-AC02-09CH11466.

## Acknowledgments

The author gratefully acknowledges stimulating conversations and collaboration with colleagues in the ExaLearn project

## References

- Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, Arx S, et al. On the opportunities and risks of foundation models. arXiv [Preprint]. *arXiv:2108.07258* (2021). Available from: <http://arxiv.org/abs/2108.07258> (Accessed August 27, 2021)
- OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 technical report. arXiv [Preprint]. *arXiv:2303.08774* (2024). Available from: <http://arxiv.org/abs/2303.08774> (Accessed November 19, 2024)
- Anand H, Sammulu BS, Olofsson KEJ, Humphreys DA. Real-time magnetic sensor anomaly detection using autoencoder neural networks on the DIII-D tokamak. *IEEE Trans. Plasma Sci.* (2022) 50:4126–30. doi:10.1109/TPS.2022.3181548
- Churchill RM, Tobias B, Zhu Y. Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data. *Phys. Plasmas* (2020) 27(6):062510. doi:10.1063/1.5144458
- Kates-Harbeck J, Svyatkovskiy A, Tang W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* (2019) 568:526–31. doi:10.1038/s41586-019-1116-4
- Rea C, Granetz RS, Montes K, Tinguely RA, Eidietis N, Hanson JM, et al. Disruption prediction investigations using Machine Learning tools on DIII-D and Alcator C-Mod. *Plasma Phys. Control. Fusion* (2018) 60: 8. doi:10.1088/1361-6587/aac7fe
- Škvára V, Šmídl V, Pevný T, Seidl J, Havránek A, Tskhakaya D, et al. Detection of alfvén Eigenmodes on COMPASS with generative neural networks. *Fusion Sci. Technol.* (2020) 76(8):962–71. doi:10.1080/15361055.2020.1820805
- Montes KJ, Rea C, Tinguely RA, Sweeney R, Zhu J, Granetz RS. A semi-supervised machine learning detector for physics events in tokamak discharges. *Nucl. Fusion* (2021) 61(2):026022. doi:10.1088/1741-4326/abcdb9
- Radford A, Wu J, Child R, Luan D, Dario A, Sutskever I Language models are unsupervised multitask learners. *Tech Rep OpenAI* (2019). Available from: <https://github.com/openai/gpt3> (Accessed July 17, 2019).
- Ashish V, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv [Preprint]. *arXiv:1706.03762*. (2017). Available from: <http://arxiv.org/abs/1706.03762> (Accessed July 17, 2019).
- Brown TB, Benjamin M, Nick R, Melanie S, Jared K, Prafulla D, et al. Language models are few-shot learners arXiv [Preprint]. *arXiv:2005.14165* (2020). Available from: <http://arxiv.org/abs/2005.14165> (Accessed June 17, 2020).
- Yinghao M, Anders Ø, Anton R, Bleiz MacSen Del S, Charalampos S, Chris D, et al. Foundation Models for Music: A Survey. arXiv [Preprint]. *arXiv:2408.14340* (2024). Available from: <http://arxiv.org/abs/2408.14340> (Accessed November 20, 2024)
- Lipman Y, Chen RTQ, Ben-Hamu H, Maximilian N, Matt L. Flow Matching for Generative Modeling. arXiv [Preprint]. *arXiv:2210.02747* (2023). Available from: <http://arxiv.org/abs/2210.02747> (Accessed October 2, 2024)
- Akbari H, Liangzhe Y, Rui Q, Wei-Hong C, Chang SF, Yin C, et al. VATT: Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text. arXiv [Preprint]. *arXiv:2104.11178* (2021). Available from: <http://arxiv.org/abs/2104.11178> (Accessed November 13, 2024)
- Jaegle A, Sebastian B, Jean-Baptiste A, Carl D, Catalin I, David D, et al. Perceiver IO: a general architecture for structured inputs and outputs. arXiv [Preprint]. *arXiv:2107.14795* (2021). Available from: <http://arxiv.org/abs/2107.14795> (Accessed January 18, 2022).
- Alayrac J-B, Jeff D, Pauline L, Antoine M, Iain B, Yana H, et al. Flamingo: a Visual Language Model for Few-Shot Learning. arXiv [Preprint]. *arXiv:2204.14198* (2022). Available from: <http://arxiv.org/abs/2204.14198> (Accessed November 20, 2024)
- Baevski A, Henry Z, Abdelrahman M, Michael A. wav2vec 2.0: a framework for self-supervised learning of speech representations. arXiv [Preprint]. *arXiv:2006.11477* (2020). Available from: <http://arxiv.org/abs/2006.11477> (Accessed June 25, 2020)
- de Vries PC, Johnson MF, Alper B, Buratti P, Hender TC, Koslowski HR, et al. Survey of disruption causes at JET. *Nucl. Fusion* (2011) 51(5):053018. doi:10.1088/0029-5515/51/5/053018
- Sabbagh SA, Berkery JW, Park YS, Ahn JH, Jiang Y, Rizques JD, et al. Disruption event characterization and forecasting in tokamaks. *Phys. Plasmas* (2023) 30:032506. doi:10.1063/5.0133825
- Murari A, Rossi R, Peluso E, Lungaroni M, Gaudio P, Gelfusa M, et al. On the transfer of adaptive predictors between different devices for both mitigation

and with attendees at the Visualizing Offline and Live Data with AI (VOLDA) workshop where the work was presented.

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



and prevention of disruptions. *Nucl. Fusion* (2020) 60(5):056003. doi:10.1088/1741-4326/ab77a6

21. Murari A, Rossi R, Craciunescu T, Vega J, Mailloux J, Abid N, et al. A control oriented strategy of disruption prediction to avoid the configuration collapse of tokamak reactors. *Nat Commun* (2024) 15(1):2424. doi:10.1038/s41467-024-46242-7
22. Duy MHN, Tan Ngoc P, Nghiem TD, Nghi QP, Quang P, Vinh T, et al. On the Out of Distribution Robustness of Foundation Models in Medical Image Segmentation. arXiv [Preprint]. *arXiv:2311.11096* (2023). Available from: <http://arxiv.org/abs/2311.11096> (Accessed December 10, 2024)
23. Liu J, Wen X, Zhao S, Chen Y, Qi X. Can OOD Object Detectors Learn from Foundation Models? arXiv [Preprint]. *arXiv:2409.05162* (2024). Available from: <http://arxiv.org/abs/2409.05162> (Accessed December 10, 2024)
24. de Vries G. EUROfusion spearheads advances in Artificial Intelligence and Machine Learning to unlock fusion energy. en-US (2024). Available from: <https://euro-fusion.org/eurofusion-news/eurofusion-spearheads-advances-in-artificial-intelligence-and-machine-learning-to-unlock-fusion-energy/> (Accessed December 10, 2024)
25. Zheng W, Fengming X, Zhongyong C, Dalong C, Bihao G, Chengshuo S, et al. Disruption prediction for future tokamaks using parameter-based transfer learning. *Commun. Phys.* (2023) 6.1:1–11. doi:10.1038/s42005-023-01296-9
26. Jalalvand A, SangKyeun K, Jaemin S, Qiming H, Max C, Peter S, et al. Multimodal Super-Resolution: Discovering hidden physics and its application to fusion plasmas. arXiv [Preprint]. *arXiv:2405.05908* (2024). Available from: <http://arxiv.org/abs/2405.05908> (Accessed November 19, 2024)
27. OpenAI. Learning to Reason with LLMs (2024). Available from: <https://openai.com/index/learning-to-reason-with-llms/> (Accessed November 20, 2024)