IMPERIAL

IMPERIAL COLLEGE LONDON

Dyson School of Design Engineering

DESIGN ENGINEERING MASTER'S PROJECT

(Early Stage Gateway Submission)

Solar Flare Prediction with Multimodal Transformer Architectures: A Novel Approach to Interpretable Space Weather Forecasting

---

Author:

Antanas Žilinskas

Supervised by:

Professor Robert N Shorten

This Early Stage Gateway submission is presented in partial fulfillment of the requirements for the Design Engineering Master's Project, part of the MEng degree in Design Engineering.

February 2, 2025

London, United Kingdom

Declaration of Originality

I hereby declare that the work presented in this thesis is my own, except where explicitly stated otherwise. To the best of my knowledge, it is original, and any collaborative contributions or referenced ideas have been duly acknowledged. Large Language Models (e.g., Claude and ChatGPT) were utilized to aid with project development, enhance the quality of writing, and as tools during research.

Copyright Declaration

Abstract

Solar flare forecasting is a critical challenge at the intersection of space weather prediction and advanced computational modeling. Current state-of-the-art approaches, including statistical methods, traditional machine learning models Zheng et al. (2023), and deep learning architectures based on convolutional and recurrent networks Jiao et al. (2020), provide incremental improvements but often struggle with the intrinsic complexity, multimodality, and nonlinearity of solar data. These shortcomings restrict both predictive accuracy and lead times, especially for the rare but highly impactful X-class flares.

The following research propose a plan for a novel transformer-based methodology inspired by Large Language Models (LLMs) Vaswani et al. (2023); Abduallah et al. (2023) to address these limitations. By integrating spatial magnetogram imagery, time-series flux measurements, and related solar observations into a unified, self-attention-driven framework Li et al. (n.d.), approach aims to set new performance benchmarks. Research will employ self-supervised pretraining to learn generalizable representations of solar phenomena Hoffmann et al. (n.d.), thereby overcoming data scarcity and improving model robustness. The resulting architecture seeks not only to enhance predictive skill but also to offer interpretability through attention mechanisms Devlin et al. (n.d.), granting insights into which features and temporal intervals are most indicative of impending events.

The anticipated contributions include more accurate, early detection of major solar flares, improved understanding of their underlying drivers, and guidance for space weather stakeholders to better protect infrastructure and human activity dependent on reliable technological systems. Ultimately, this work aspires to establish a new standard for solar flare predictions using transformer-based architecture.

Contents

Chapter 1

Introduction

Solar flares are colossal eruptions on the Sun's surface, releasing immense energy across the electromagnetic spectrum and causing severe disturbances in Earth's space weather environment (*ESA - What are solar flares?*, n.d.). Even moderate flares can disrupt aviation, maritime communications, and satellite operations, while intense X-class events threaten power grids and crewed space missions (Omatola and Okeme, n.d.). Table 1.1 summarizes flare classes and their terrestrial impacts. Although precautionary measures exist, timely and accurate forecasting remains critical for allowing operators and agencies to prepare effectively.

Table 1.1: Solar Flare Classifications and Potential Terrestrial Impacts

Class	Peak X-ray Flux (W/m^2)	Potential Impacts on Earth
A	$I < 10^{-8}$	Minimal to no observable impact.
B	$10^{-8} \leq I < 10^{-7}$	Generally negligible effects; minor disturbances in Earth's magnetosphere.
C	$10^{-7} \leq I < 10^{-6}$	Small; few noticeable consequences on Earth.
M	$10^{-6} \leq I < 10^{-5}$	Medium-sized; can cause brief radio blackouts affecting Earth's polar regions. Minor radiation storms sometimes follow.
X	$I \geq 10^{-5}$	Major events that can trigger planet-wide radio blackouts and long-lasting radiation storms.

Traditional statistical and early machine learning models (e.g., SVMs, random forests) achieved incremental improvements but failed to capture the complex temporal-spatial dynamics driving flares (Schrijver, 2009; Zheng et al., 2023). Recent deep learning methods, employing LSTMs and CNNs, have enhanced performance by integrating temporal and spatial features, while attention mechanisms have further boosted mid-level flare predictions (Jiao et al., 2020; Abduallah et al., 2023). Yet, accurately forecasting rare, high-impact X-class flares and generalizing across solar cycles remain unresolved

challenges (Hayes et al., 2021).

To overcome these limitations, this research introduces a multimodal transformer architecture inspired by Large Language Model (LLM) principles (Vaswani et al., 2023). Unlike traditional CNN-LSTM hybrids, it fuses diverse data—magnetograms, flux time series, and historical flare records—into one predictive pipeline (Abduallah et al., 2023; Schmude et al., 2024). By doing so, it aims to surpass current benchmarks, improve lead times, enhance interpretability, and establish a new standard for solar flare forecasting.

Chapter 2

Literature Review and Contextual Investigation

Solar flare prediction has a long history of methodological evolution, given the growing complexity of our understanding of solar physics and the growing importance of timely forecasts. Early efforts primarily used statistical approaches, grounded in empirically derived heuristics and morphological classifications of solar active regions (Schrijver, 2009). Though these statistical methods offered initial baselines, they were typically constrained by simplistic assumptions. As a result, they struggled to capture the inherently nonlinear and dynamic nature of the solar magnetic environment. Subsequent attempts to refine forecasts introduced traditional machine learning algorithms, such as Support Vector Machines (SVMs) and Random Forests (RFs), utilizing engineered features extracted from observed solar magnetic field parameters. While these techniques improved classification accuracy and provided more flexible decision boundaries than pure statistical models (Zheng et al., 2023; Abduallah et al., 2023), they still faced limitations. Notably, such models often failed to adequately capture the temporal dimension critical to flare onset and lacked mechanisms for assimilating the diverse sensor modalities available.

The advancement of deep learning techniques furthered the effectiveness of solar flare prediction. Convolutional Neural Networks (CNNs) enabled the automated extraction of spatial features from high-resolution solar imagery, whereas Long Short-Term Memory (LSTM) networks could model time dependencies more effectively. By combining CNNs and LSTMs, researchers created hybrid architectures that jointly processed spatial magnetogram data and temporal flux measurements. This fusion yielded improvements over traditional ML models, particularly in differentiating higher-class flares (e.g., M-class and X-class) from weaker events (Abduallah et al., 2023; Jiao et al., 2020). Moreover, the introduction of attention mechanisms into LSTM-based models began to address the challenge of focusing on the most relevant features at critical time steps (Zheng

et al., 2023). Given the complexity of the data available, research has also attempted to predict solar flare activity based on data of the Space-weather HMI Active Region Patches (SHARP) parameters (Abduallah et al., 2023). These measures, derived from observational data, describe the magnetic field configuration and complexity. However, such abstractions lose out on the resolution needed to deliver predictions more reliably. More widely, despite these advancements, even state-of-the-art deep learning models have struggled with data scarcity for rare X-class flares and have shown limited ability to generalize across multiple solar cycles (Hayes et al., 2021).

In parallel, the broader field of machine learning has advanced significantly with the introduction of transformer architectures (Vaswani et al., 2023). Originally designed for natural language processing tasks, transformers leverage the concept of self-attention to model long-range dependencies without relying on recurrent operations. By doing so, they offer flexible contextual modeling, scalability, and an ability to consider global relationships within a sequence. Large Language Models (LLMs), such as GPT and BERT, have demonstrated the raw power of transformers in capturing nuanced patterns from considerable text corpora (Devlin et al., n.d.). While these models were conceived for language, the underlying principles have proven remarkably transferable to other domains that benefit from long-context attention and rely on time-series data.

Outside of solar physics, transformers have shown promise in time-series forecasting tasks, climate modeling, and various scientific challenges where complex, nonlinear dynamics span multiple temporal and spatial scales. For instance, researchers have applied transformers to predict seasonal weather patterns, learn from satellite-derived Earth observation data, and analyze energy consumption trends—scenarios that, like solar activity, involve intricate relationships across time and space (Gettelman et al., 1997; Yıldız and Timoçin, 2023). In climate modeling, transformers have begun to tackle the complexity of Earth system processes, potentially offering improved long-horizon forecasts over previous recurrent or convolutional models (Liu et al., 2024; Blumenfeld, n.d.). In financial and economic time-series forecasting, transformers have demonstrated improved predictive performance and interpretability in understanding market shifts, which often require capturing long-term correlations (Schmude et al., 2024). These successes suggest that transformer architectures could similarly deliver improvements in solar flare prediction.

When examining the existing body of solar flare forecasting literature, a clear gap emerges: very few studies have seriously considered transformer-based models, let alone experimented with self-supervised learning or multimodal fusion strategies tailored for this domain. While recent attempts have begun to incorporate attention-like mechanisms or even basic transformer variants into the pipeline (Abduallah et al., 2023), they remain preliminary and have not fully leveraged the power of LLM-inspired architectures. There is sparse exploration into truly multimodal approaches that can fuse heterogeneous inputs—ranging

from image-based magnetograms, line-of-sight magnetic field measurements, X-ray flux time series, to derived solar features—within a single, unified transformer framework. Equally rare is any detailed investigation of self-supervised pretraining strategies, which could enable the model to learn robust, general representations of solar activity before downstream supervised forecasting tasks.

This avenue of research could benefit the solar flare predictions community significantly. Solar activity spans multiple timescales, from minutes to days, and its driving processes involve complex spatial structures and evolving magnetic fields. Transformers, with their global attention, can potentially model these long-term dependencies far better than classical recurrent or convolutional architectures (Kaplan et al., n.d.). Additionally, by treating input modalities as tokens within a transformer encoder, we can theoretically integrate images, scalar time series, and other data types in a more principled manner than ad-hoc model fusion strategies (Francisco et al., 2024).

Beyond the modeling capabilities, another challenge in solar flare forecasting is the interpretability of predictions. Existing black-box deep learning models often yield predictions without illuminating what aspects of solar activity led to a forecasted flare. Transformers offer a unique interpretability avenue: attention weights. By examining attention maps, it may be possible to identify specific spatial regions or temporal intervals that the model deems critical, thereby providing the solar physics community with valuable insights into the underlying solar processes that precede flares (Gallagher et al., n.d.).

Yet, to realize these benefits, several challenges must be addressed. Data complexity is significant: solar data come in multiple formats (e.g., magnetograms, UV/EUV imagery, spectral flux), each requiring different preprocessing steps. Moreover, rare but high-impact flare classes (X-class) have limited examples, calling for domain adaptation, transfer learning, or synthetic data generation. Ensuring that the model can capture long-term dependencies extending over several days of solar observations is another nontrivial endeavor. Transformers are well-equipped for long sequences, but efficiency and careful architectural design remain paramount (Fedus et al., 2022). Lastly, interpretability and uncertainty quantification are essential for operational trust—no forecasting system can be reliably used by mission operators or infrastructure managers without understanding the confidence and rationale behind its predictions.

The case for transformer-based methods in solar flare forecasting, therefore, rests firmly on these points. By using self-attention, transformers can fuse diverse signals that conventional architectures handle only in a piecemeal manner. The advanced representational power from self-supervised pretraining could help overcome data scarcity, learning general solar context before specializing in flare prediction. In doing so, these models can push beyond incremental gains, potentially offering a step-change in predictive skill. From earlier

detection of imminent X-class flares to increased model interpretability and more robust generalization across solar cycles, a transformer-centered approach is well positioned to reshape the field.

Chapter 3

Project Objectives and scope

3.0.1 Objectives

This project aims to set a new standard for solar flare forecasting by employing transformer-based, multimodal models. The primary objectives are to:

1. Develop a transformer architecture that integrates solar imagery, magnetograms, and X-ray flux time series into a unified predictive framework Abduallah et al. (2023); Zheng et al. (2023).

2. Explore self-supervised strategies to learn generalizable solar representations, addressing data scarcity and improving model robustness Fedus et al. (2022); Hoffmann et al. (n.d.).

3. Surpass current benchmarks on key solar forecasting metrics (TSS, POD, FAR), especially for rare, high-impact X-class flares Abduallah et al. (2023); Hayes et al. (2021).

In addition to these primary aims, secondary objectives include exploring interpretability tools, by visualizing attention maps to highlight key regions or temporal intervals Abduallah et al. (2023), implementing uncertainty quantification to inform operational risk assessments Abduallah et al. (2023), and incorporating domain-informed architectural modifications, such as tailored positional embeddings aligned with solar rotational periods Harra et al. (2016); Schrijver (2009).

3.0.2 Scope and Constraints

This work focuses on publicly available data from sources like SDO and GOES Yıldız and Timoçin (2023); Omatola and Okeme (n.d.). Emphasis will be placed on M- and X-class

flare prediction over at least one full solar cycle. While the methodology can extend to other datasets or incorporate additional physical processes, this study confines itself to well-established datasets and core predictive tasks Gettelman et al. (1997), deferring more complex scenarios for future investigations.

Chapter 4

Methodological Approach

4.0.1 Data Selection and Preprocessing

This project will leverage GOES X-ray flux time series for flare labeling, complemented by Solar Dynamics Observatory (SDO) imagery (HMI data) for spatial patterns and SHARP parameters representing magnetic field properties Abduallah et al. (2023); Zheng et al. (2023). Preprocessing involves normalization, resolution adjustments, temporal synchronization, and careful handling of missing values. Multiple years of data will be used, covering diverse solar conditions Yıldız and Timoçin (2023).

4.0.2 Modeling Approach

The core methodology centers on a multimodal transformer pipeline. Research will start by training a baseline time-series transformer on X-ray flux data Vaswani et al. (2023). Next, self-supervised pretraining tasks (masked token modeling, temporal patch prediction) will learn general solar feature representations prior to supervised fine-tuning Fedus et al. (2022); Hoffmann et al. (n.d.). A Vision Transformer (ViT)-inspired component will process imagery, while a separate transformer encoder handles flux or magnetic inputs. Cross-attention layers then fuse these data streams into a unified predictive head Abduallah et al. (2023).

4.0.3 Benchmarking and Validation Strategy

Performance will be compared to established baselines—such as CNN-LSTM hybrids and simpler transformers—using metrics like the True Skill Statistic (TSS), Probability of Detection (POD), False Alarm Rate (FAR), and Heidke Skill Score (HSS) Zheng et al. (2023). K-fold cross-validation across multiple solar cycles ensures robust generalization.

Early evaluations will guide iterative refinements Yıldız and Timoçin (2023).

4.0.4   Feasibility and Tools

The implementation will use PyTorch and run on high-performance local computing resources or GPU clusters. Adapting existing transformer codebases from climate and Earth observation domains will streamline development, ensuring both reliability and efficient experimentation Vaswani et al. (2023); Gettelman et al. (1997); Schmude et al. (2024).

Chapter 5

Project Management and Timeline

5.0.1 Work Plan

Term 1 established the project's theoretical foundation via an extensive literature review, clarifying research gaps and guiding the methodological approach Minaee et al. (n.d.). In parallel, initial technical setups and a preliminary experiment—training a small language model on a Shakespearean corpus—validated the computational workflow and transformer routines (see Listing 5.1), whilst providing a great breakdown of "Attention Is All You Need" paper Vaswani et al. (2023).

Listing 5.1: Sample Output from the Preliminary Language Model

```
MARCIUS:
God less unfound
A harm upon it!

Lord Marcius 'sut doth the twenty world,
And flover high gives, she'll hear shall mine mothes,
Having been most post were soul's a conspiror
And to Rome well thricke.
I would you drive you me to it.

SLEONTAS:
Take it to sign answer.

LADY ANNE:
Yea, bid me you upon that subject come to have hath
The prepartuative you.
```

GLOUCESTER:

GLOUCESTER:
You had not to stand of, when measure.
What Lady? this man! Catesby Is to think:
Yout the varlet and day not

5.0.2 Term 2 and Term 3

Term 2 will implement full-scale transformer models tailored to solar data Francisco et al. (2024), investigate self-supervised pretraining, and integrate multiple input modalities. Iterative refinement, benchmarking against baselines Schmude et al. (2024), and architectural improvements will be central activities. Term 3 will focus on robust validation, including cross-validation, interpretability analyses (via attention maps) Hoffmann et al. (n.d.), uncertainty quantification, and final performance reporting. Deliverables include a refined predictive model, comprehensive evaluations, and operational recommendations Li et al. (n.d.).

5.0.3 Week-by-week breakdown

Week	Tasks and Milestones
Term 2 (Spring Term)	
Week 1 (Jan 8 - Jan 14)	Finalize self-supervised pretraining tasks. Prepare datasets for multimodal input integration, ensuring alignment and preprocessing of magnetogram images, time-series flux, and SHARP parameters.
Week 2 (Jan 15 - Jan 21)	Implement the baseline transformer architecture for time-series flux data. Train and validate the initial model on a reduced dataset to debug computational workflow.
Week 3 (Jan 22 - Jan 28)	Introduce multimodal input handling by integrating spatial magnetograms and temporal flux measurements into the model pipeline. Perform initial tests with simplified inputs.
Week 4 (Jan 29 - Feb 4)	Develop and implement cross-attention layers for fusing diverse modalities. Conduct preliminary training on multimodal datasets.
Week 5 (Feb 5 - Feb 11)	Conduct benchmarking of the current model against CNN-LSTM and other baseline architectures using True Skill Statistics (TSS) and other metrics. Refine architecture based on performance gaps.

Week 6 (Feb 12 - Feb 18)	Focus on interpretability: implement attention map visualization and begin analyzing which features contribute most to flare predictions. Document insights.
Week 7 (Feb 19 - Feb 25)	Explore uncertainty quantification methods. Implement and test uncertainty-aware predictions to evaluate model robustness.
Week 8 (Feb 26 - Mar 3)	Conduct a full round of validation using cross-validation over multiple solar cycles. Ensure the model generalizes well across different data distributions.
Week 9 (Mar 4 - Mar 10)	Prepare for Late Stage Gateway: finalize a comprehensive presentation of model design, current progress, challenges, and schematic work plan for the remaining project duration.
Week 10 (Mar 11 - Mar 17)	Final rehearsal of the Late Stage Gateway presentation. Collect and address feedback from peers and supervisors. Prepare visual aids such as attention maps and model outputs for the presentation.
Week 11 (Mar 18 - Mar 21)	Deliver the Late Stage Gateway presentation. Document feedback and create an updated work plan for the final stages of the project leading to June 5 submission.
Term 3 (Summer Term)	
Week 1 (Mar 31 - Apr 6)	Incorporate feedback from the Late Stage Gateway into project plans. Begin final refinements of the multimodal transformer model based on validation insights.
Week 2 (Apr 7 - Apr 13)	Conduct additional experiments to validate improvements and robustness. Begin writing the introduction, background, and methodology sections of the final report.
Week 3 (Apr 14 - Apr 20)	Finalize interpretability tools (e.g., attention maps) and incorporate into analysis. Draft results section with a focus on benchmarking against baseline methods.
Week 4 (Apr 21 - Apr 27)	Integrate social, environmental, and technological impact analyses into the discussion section. Update diagrams and visualizations for the final report.
Week 5 (Apr 28 - May 4)	Review the full report draft with peers and supervisors. Incorporate feedback into final report revisions.
Week 6 (May 5 - May 11)	Conduct final testing and validation to ensure all criteria are met. Refine results and discussion sections based on testing insights.

Week 7 (May 12 - May 18)	Complete final report draft and submit for supervisor review. Prepare a slide deck for Demo Day and begin rehearsals.
Week 8 (May 19 - May 25)	Polish and finalize the project report. Conduct rehearsals for Viva and Demo Day presentations.
Week 9 (May 26 - Jun 1)	Submit the final report (Deadline: June 5). Finalize presentation materials for Demo Day. Prepare technical documentation and codebase for demonstration.
Week 10 (Jun 2 - Jun 8)	Deliver Demo Day presentation (Scheduled for the week of June 9). Prepare for VIVA sessions with practice Q and A sessions.
Week 11 (Jun 9 - Jun 15)	Deliver Viva presentation and Q and A session (Scheduled for the week of June 16). Document final reflections and insights for personal learning and growth.

5.0.4 Risks and Mitigation

Potential risks such as limited training data or computational bottlenecks have been considered from the outset Gao et al. (2020). Mitigation strategies—ranging from alternative datasets Liu et al. (2024) to more efficient transformer variants Fedus et al. (2022)—are in place, ensuring that the project's momentum will not be unduly hindered Rajbhandari et al. (n.d.).

Chapter 6

Expected Contributions and Impact

This research aims to establish a standard for solar flare forecasting by introducing transformer-based Vaswani et al. (2023); Hoffmann et al. (n.d.), multimodal Schmude et al. (2024), and self-supervised learning strategies specifically tailored to complex solar data. By integrating diverse data types—spatial magnetograms, flux time series, and potentially additional solar observations—into a unified framework Francisco et al. (2024); Abduallah et al. (2023), this project can yield richer, more accurate forecasts with longer lead times and enhanced interpretability Li et al. (n.d.).

Chapter 7

Conclusion and Next Steps

In conclusion, this project aligns latest transformer architectures with the pressing need for more accurate, interpretable solar flare forecasts. By systematically addressing the challenges of data complexity, rare events, and long-range temporal dependencies, it aims to push the boundaries of current methodologies and operational capabilities.

Moving forward, the immediate next steps involve finalizing the model design, initiating early experiments with multimodal transformers, and iteratively refining approaches based on preliminary results. Over subsequent terms, this iterative process will guide the transformation of theoretical concepts into a robust, high-impact predictive framework.

Bibliography

Abduallah, Y., Wang, J. T. L., Wang, H. and Xu, Y. (2023), 'Operational prediction of solar flares using a transformer-based framework', *Scientific Reports* **13**(1).

Blumenfeld, J. (n.d.), 'Nasa and ibm research apply ai to weather and climate | nasa earthdata'.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (n.d.), 'Bert: Pre-training of deep bidirectional transformers for language understanding'.

ESA - What are solar flares? (n.d.).

Fedus, W., Zoph, B., Shazeer, N. and Clark, A. (2022), 'Switch transformers: Scaling to trillion parameter models with simple and e cient sparsity'.

Francisco, G., Guastavino, S., Fernandes, J., Barata, T. and Moro, D. D. (2024), 'Multimodal flare forecasting with deep learning'.

Gallagher, P. T., Moon, Y. J. and Wang, H. (n.d.), 'I. data processing and first results'.

Gao, L., Thite, A., Nabeshima, N., Presser, S. and Leahy, C. (2020), 'The pile: An 800gb dataset of diverse text for language modeling'.

Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D. J., Stephens, G. L., Heever, S. C. V. D., Varble, A. C. and Zuidema, P. (1997), 'The future of earth system prediction: Advances in model-data fusion', *Science Advances* **8**(14).

Harra, L. K., Schrijver, C. J., Janvier, M., Toriumi, S., Hudson, H., Matthews, S., Woods, M. M., Hara, H., Guedel, M., Kowalski, A., Osten, R., Kusano, K. and Lueftinger, T. (2016), 'The characteristics of solar x-class flares and cmes: A paradigm for stellar superflares and eruptions?', *Solar Physics* **291**(6), 1761.

Hayes, L. A., O'hara, O. S. D., Murray, S. A. and Gallagher, P. T. (2021), 'Solar flare effects on the earth's lower ionosphere', *Solar Physics* **296**(11).

Hoffmann, J., x1d43f;, S. B., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., De, D., Casas, L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican,

K., Driessche, G. V. D., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J. W., Vinyals, O. and Sifre, L. (n.d.), 'Training compute-optimal large language models'.

Jiao, Z., Sun, H., Wang, X., Manchester, W., Gombosi, T., Hero, A. and Chen, Y. (2020), 'Solar flare intensity prediction with machine learning models'.

Kaplan, J., Mccandlish, S., Openai, Henighan, T., Brown, T. B., Gray, S., Radford, A., Wu, J. and Amodei, D. (n.d.), 'Scaling laws for neural language models'.

Li, H., Wang, Z., Wang, J., Kai, A., Lau, H. and Qu, H. (n.d.), 'Cllmate: A multimodal llm for weather and climate events forecasting'.

Liu, Y., Cao, J., Liu, C., Ding, K. and Jin, L. (2024), 'Datasets for large language models: A comprehensive survey'.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X. and Gao, J. (n.d.), 'Large language models: A survey'.

Omatola, . and Okeme, K. M. (n.d.), 'Impacts of solar storms on energy and communications technologies'.

Rajbhandari, S., Rasley, J., Ruwase, O. and He, Y. (n.d.), 'Zero: Memory optimizations toward training trillion parameter models'.

Schmude, J., Roy, S., Trojak, W., Jakubik, J., Civitarese, D. S., Singh, S., Kuehnert, J., Ankur, K., Gupta, A., Phillips, C. E., Kienzler, R., Szwarcman, D., Gaur, V., Shinde, R., Lal, R., Silva, A. D., Luis, J., Diaz, G., Jones, A., Pfreundschuh, S., Lin, A., Sheshadri, A., Nair, U., Anantharaj, V., Hamann, H., Watson, C., Maskey, M., Lee, T. J., Moreno, J. B. and Ramachandran, R. (2024), 'Prithvi wxc: Foundation model for weather and climate'.

Schrijver, C. J. (2009), 'Driving major solar flares and eruptions: A review', *Advances in Space Research* **43**(5), 739.

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Research, G., Uszkoreit, J., Jones, L. and Gomez, A. N. (2023), 'Provided proper attribution is provided, google hereby grants permission to reproduce the tables and figures in this paper solely for use in journalistic or scholarly works. attention is all you need'.

Yıldız, L. B. and Timoçin, E. (2023), 'The effect of solar flares on hf radio communications over turkey', *Geomagnetism and Aeronomy* **63**(5), 642.

Zheng, Y., Qin, W., Li, X., Ling, Y., Huang, X., Li, X., Yan, P., Yan, S. and Lou, H. (2023), 'Comparative analysis of machine learning models for solar flare prediction', *Astrophysics and Space Science* **368**(7).