Foundational Models in physics and its neighborhood

Peter Steinbach, Steve Schmerler

Helmholtz-Zentrum Dresden-Rossendorf, Department for Information Services and Computing,

MT Annual Meeting 2025, November 3, 2025

Table of Contents

Table of Contents

Some recent history: BERT

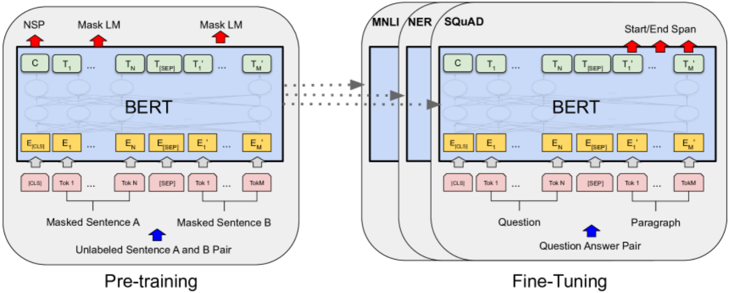


Figure 1 from [Devlin et al., 2018]

- name:
  Bidirectional
  Encoder
  Representations from
  Transformers

Some recent history: BERT

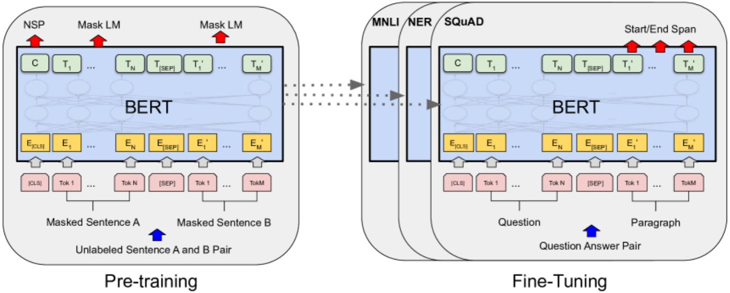


Figure 1 from [Devlin et al., 2018]

- name:
  Bidirectional
  Encoder
  Representations from
  Transformers
- task:
  string sequence to sequence translation

Some recent history: BERT

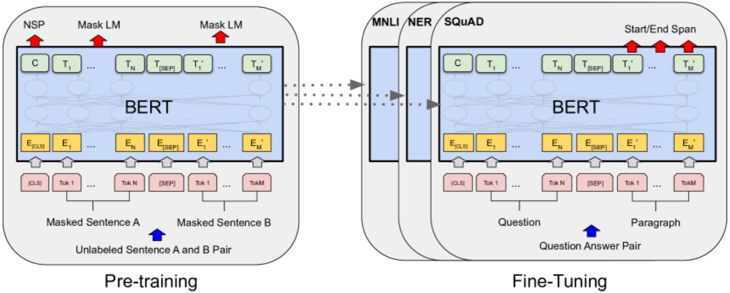


Figure 1 from [Devlin et al., 2018]

- name:
  Bidirectional
  Encoder
  Representations from
  Transformers
- task:
  string sequence to sequence translation
- data:
  (unlabelled) text pairs

Some recent history: BERT

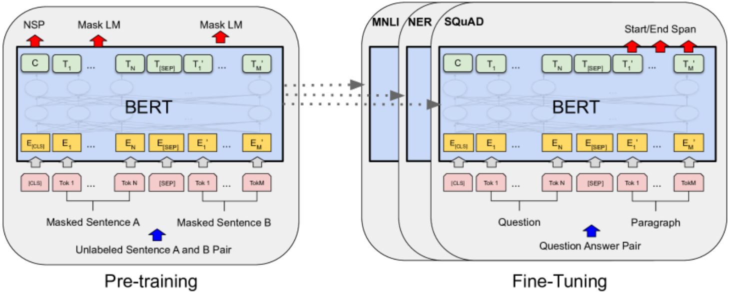


Figure 1 from [Devlin et al., 2018]

- name:
  Bidirectional
  Encoder
  Representations from
  Transformers
- task:
  string sequence to sequence translation
- data:
  (unlabelled) text pairs
- breakthrough:

Some recent history: BERT

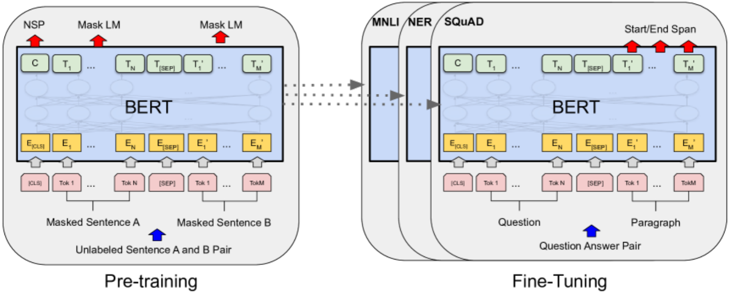


Figure 1 from [Devlin et al., 2018]

- name:
  Bidirectional
  Encoder
  Representations from
  Transformers
- task:
  string sequence to sequence translation
- data:
  (unlabelled) text pairs
- breakthrough:
  1. train once

Some recent history: BERT

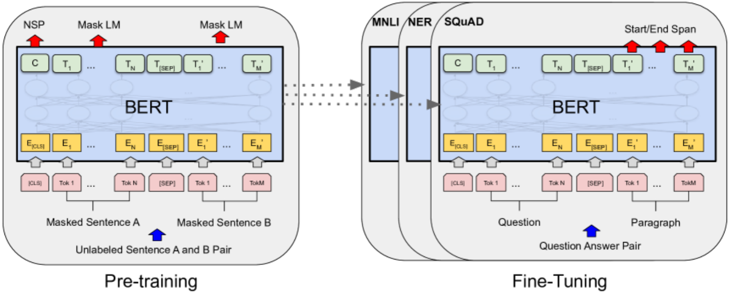


Figure 1 from [Devlin et al., 2018]

- name:
  Bidirectional
  Encoder
  Representations from
  Transformers
- task:
  string sequence to sequence translation
- data:
  (unlabelled) text pairs
- breakthrough:
  1. train once
  2. finetune and use on many unrelated tasks

  (MNLI, NER, SQuAD, . . .)

Some recent history: GPT-1

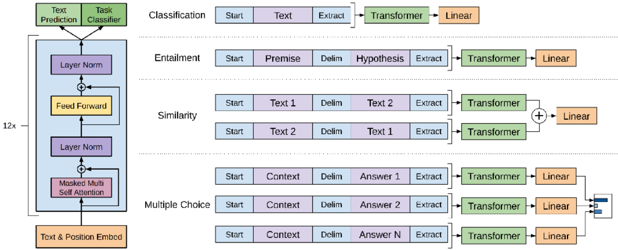


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer

Some recent history: GPT-1

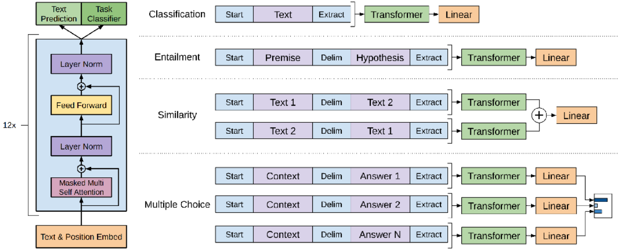


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding

Some recent history: GPT-1

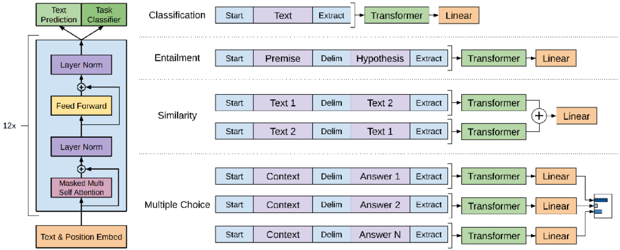


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding
- data:
  heaps of text (www)

Some recent history: GPT-1

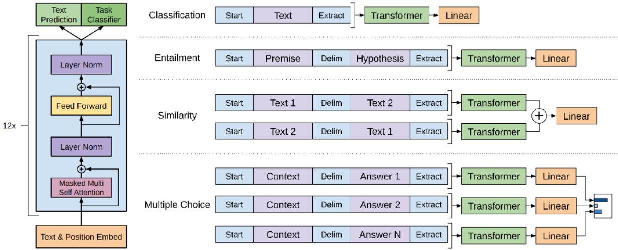


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding
- data:
  heaps of text (www)
- breakthrough:

Some recent history: GPT-1

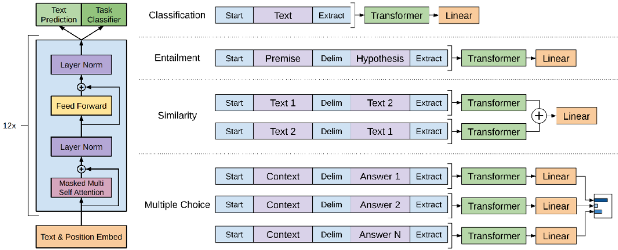


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding
- data:
  heaps of text (www)
- breakthrough:
  1. scalable

Some recent history: GPT-1

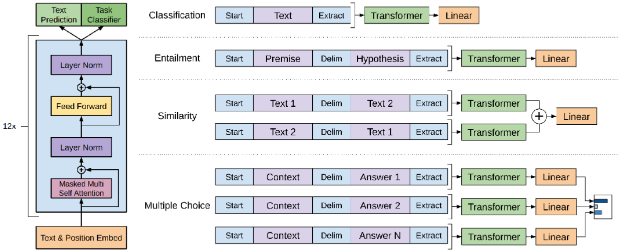


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding
- data:
  heaps of text (www)
- breakthrough:
  1. scalable
  2. better quality than BERT [Radford, Wu, et al., 2019]

Some recent history: GPT-1

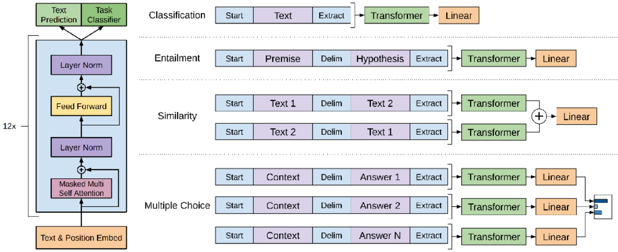


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding
- data:
  heaps of text (www)
- breakthrough:
  1. scalable
  2. better quality than BERT [Radford, Wu, et al., 2019]
  3. science behind closed doors

Some recent history: GPT-1

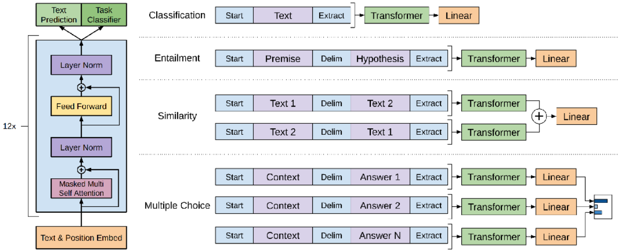
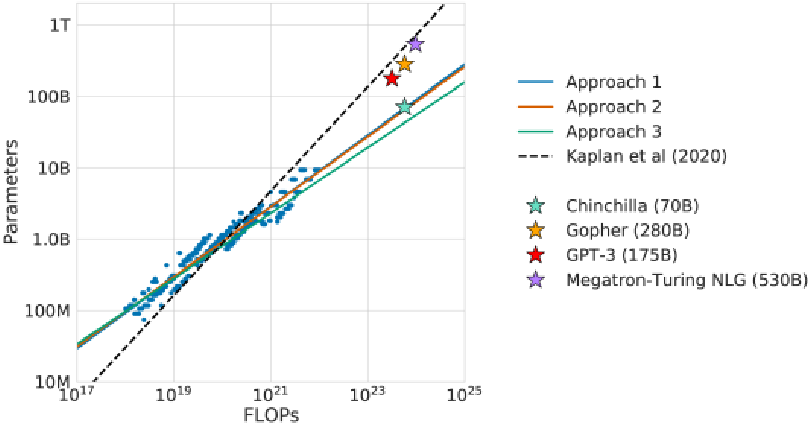


Figure 1 from [Radford, Narasimhan, et al., 2018]

- name:
  Generative **P**retrained **T**ransformer
- task:
  sequence to sequence decoding
- data:
  heaps of text (www)
- breakthrough:
  1. scalable
  2. better quality than BERT [Radford, Wu, et al., 2019]
  3. science behind closed doors
- pretraining + finetune

Scalability? [Hoffmann et al., 2022]



The bigger, the better!

(bigger models, more compute, more data result in better performance)

image-text model: OpenClip

arXiv > cs > arXiv:2212.07143v2

Search...

Help

Computer Science > Machine Learning

[Submitted on 14 Dec 2022 (v1), last revised 13 Jul 2024 (this version, v2)]

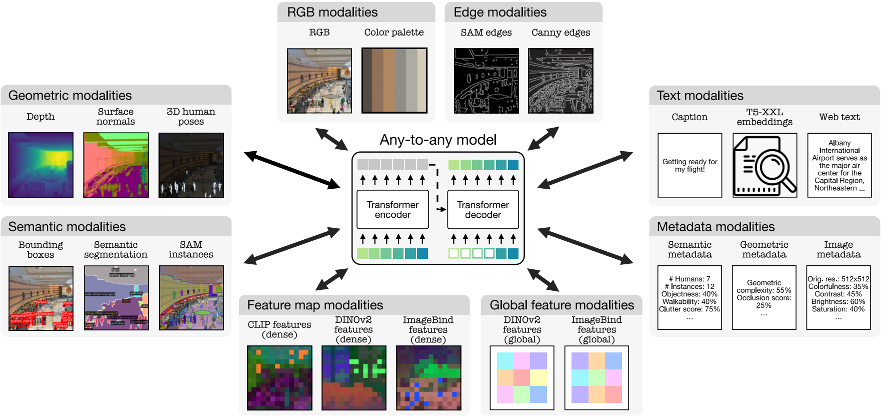
Reproducible scaling laws for contrastive language-image learning

Mehdi Cherti, Romain Beaumont, Ross Wightman, Mitchell Wortsman, Gabriel Ilharco, Cade Gordon, Christoph Schuhmann, Ludwig Schmidt, Jenia Jitsev

Scaling up neural networks has led to remarkable performance across a wide range of tasks. Moreover, performance often follows reliable scaling laws as a function of training set size, model size, and compute, which offers valuable guidance as large-scale experiments are becoming increasingly expensive. However, previous work on scaling laws has primarily used private data \& models or focused on uni-modal language or vision learning. To address these limitations, we investigate scaling laws for contrastive language-image pre-training (CLIP) with the public LAION dataset and the open-source OpenCLIP repository. Our large-scale experiments involve models trained on up to two billion image-text pairs and identify power law scaling for multiple downstream tasks including zero-shot classification, retrieval, linear probing, and end-to-end fine-tuning. We find that the training distribution plays a key role in scaling laws as the OpenAI and OpenCLIP models exhibit different scaling behavior despite identical model architectures and similar training recipes. We open-source our evaluation workflow and all models, including the largest public CLIP models, to ensure reproducibility and make scaling laws research more accessible. Source code and instructions to reproduce this study will be available at this https URL

github.com/mlfoundations/open_clip [Cherti et al., 2023]

image-to-any models: Imagebind, M4, ...



[Bachmann et al., 2024]

Table of Contents

Review Paper in HFMI: Motivation and Methods

Motivation

- Initiative in Helmholtz Foundational Model Initiative (HFMI)
- diverse set of disciplines (geoscience, life science, physics, materials, ...)
- goal: review literature for successful applications of FMs

Review Paper in HFMI: Motivation and Methods

Motivation

- Initiative in Helmholtz Foundational Model Initiative (HFMI)
- diverse set of disciplines (geoscience, life science, physics, materials, ...)
- goal: review literature for successful applications of FMs

Methods

1. search candidate papers (google scholar, science OS, consensus
2. download pdf files (or store URLs)
3. ingest to notebookLM
4. analyse and summarize
5. manually double check!

Classifying Physics: Physh [Smith, 2019]

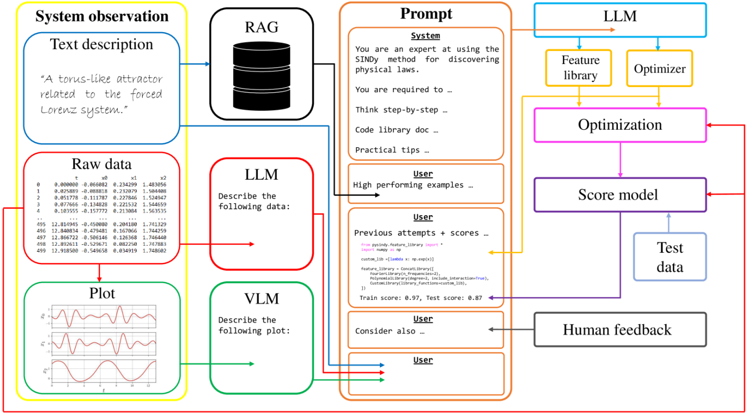
- Colliders
- Atomic, Molecular, Optics
- Condensed Matter
  separate
- Enery
- Fluid Dynamics
- Grav., Cosm. & Astro.
- interdisc. physics
- networks
- nonlinear dynamics
- nuclear physics
- particles and fields
- physics education
- physics of living systems
- plasma physics
- polymers & soft matter
- quantum information
- statistical physics

Classifying Physics: Physh [Smith, 2019]

Colliders	4
Atomic, Molecular, Optics	4
Condensed Matter	separate
Enery	4
Fluid Dynamics	20
Grav., Cosm. & Astro.	23
interdisc. physics	(14)
networks	17
nonlinear dynamics	9
nuclear physics	4
particles and fields	21
physics education	13
physics of living systems	2
plasma physics	3
polymers & soft matter	16
quantum information	10
statistical physics	4

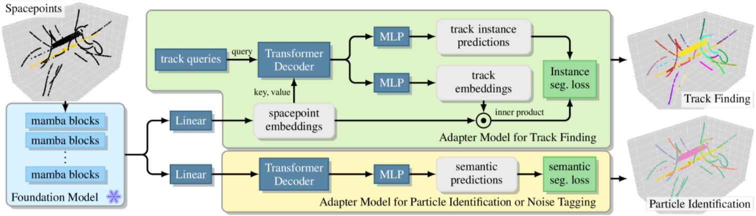
Table of Contents

Al-Khwarizmi: Discovering Physical Laws with Foundation Models

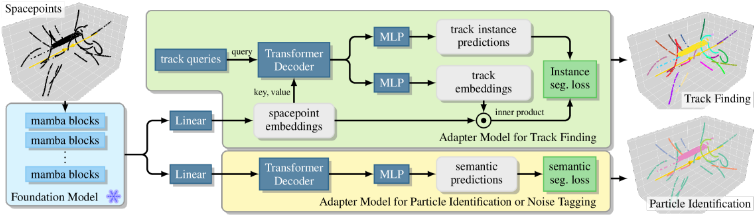


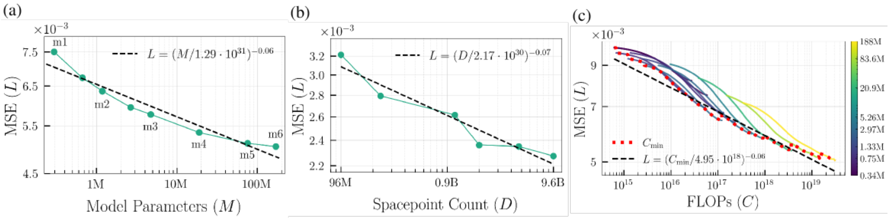
[Mower et al., 2025]

FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics



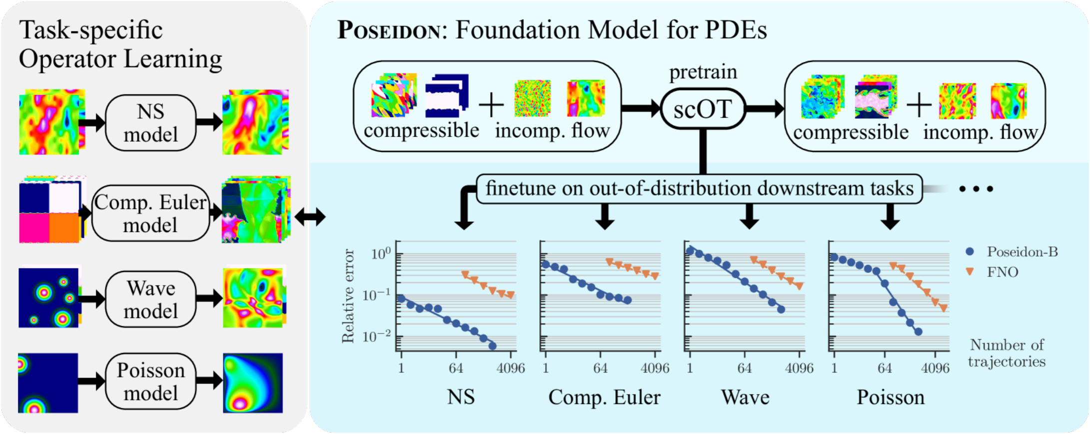
FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics





[Park et al., 2025]

Poseidon: Efficient Foundation Models for PDEs



[Herde et al., 2024]

Table of Contents

Main Takeaways

- Machine Learning Methods for Natural Language Processing nurtured idea of foundational models

Main Takeaways

- Machine Learning Methods for Natural Language Processing nurtured idea of foundational models
- key ingredients: **Parallelization of Models**$^{\mathrm{HPC}}$, **Availability of Data**$^{\mathrm{DM}}$, **Statistics** and **Machine Learning**

Main Takeaways

- Machine Learning Methods for Natural Language Processing nurtured idea of foundational models
- key ingredients: **Parallelization of Models**$^{\mathrm{HPC}}$, **Availability of Data**$^{\mathrm{DM}}$, **Statistics** and **Machine Learning**
- some parts of physics invest heavily in this (number of papers, assumed person hours)

Main Takeaways

- Machine Learning Methods for Natural Language Processing nurtured idea of foundational models
- key ingredients: **Parallelization of Models**$^{\mathrm{HPC}}$, **Availability of Data**$^{\mathrm{DM}}$, **Statistics** and **Machine Learning**
- some parts of physics invest heavily in this (number of papers, assumed person hours)
- stay tuned for our preprint!

Main Takeaways

- Machine Learning Methods for Natural Language Processing nurtured idea of foundational models
- key ingredients: **Parallelization of Models**$^{\text{HPC}}$, **Availability of Data**$^{\text{DM}}$, **Statistics** and **Machine Learning**
- some parts of physics invest heavily in this (number of papers, assumed person hours)
- stay tuned for our preprint!

Main Takeaways

- Machine Learning Methods for Natural Language Processing nurtured idea of foundational models
- key ingredients: **Parallelization of Models**$^{\mathrm{HPC}}$, **Availability of Data**$^{\mathrm{DM}}$, **Statistics** and **Machine Learning**
- some parts of physics invest heavily in this (number of papers, assumed person hours)
- stay tuned for our preprint!

Thank you for your attention!

Feel free to ask questions, provide feedback or comments.

Talk to us!

- helmholtz.ai/you-helmholtz-ai/ai-consulting/
- contact@helmholtz.ai for general questions
- consultant-helmholtz.ai@hzdr.de for questions regarding research field matter

Bibliography (I)

- Bachmann, Roman et al. (2024). *4M-21: An Any-to-Any Vision Model for Tens of Tasks and Modalities*. arXiv: 2406.09406 [cs.CV]. URL: https://arxiv.org/abs/2406.09406 (cit. on p. 20).
- Cherti, Mehdi et al. (June 2023). "Reproducible Scaling Laws for Contrastive Language-Image Learning". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 2818–2829. DOI: 10.1109/cvpr52729.2023.00276. URL: http://dx.doi.org/10.1109/CVPR52729.2023.00276 (cit. on p. 19).
- Devlin, Jacob et al. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *CoRR* abs/1810.04805. arXiv: 1810.04805. URL: http://arxiv.org/abs/1810.04805 (cit. on pp. 4–9).
- Herde, Maximilian et al. (Nov. 2024). *Poseidon: Efficient Foundation Models for PDEs*. arXiv:2405.19101 [cs]. DOI: 10.48550/arXiv.2405.19101. URL: http://arxiv.org/abs/2405.19101 (visited on 10/06/2025) (cit. on p. 30).

Bibliography (II)

- Hoffmann, Jordan et al. (2022). *Training Compute-Optimal Large Language Models*. arXiv: 2203.15556 [cs.CL]. URL: https://arxiv.org/abs/2203.15556 (cit. on p. 18).
- Mower, Christopher E. and Haitham Bou-Ammar (June 2025). *Al-Khwarizmi: Discovering Physical Laws with Foundation Models*. arXiv:2502.01702 [cs]. DOI: 10.48550/arXiv.2502.01702. URL: http://arxiv.org/abs/2502.01702 (visited on 10/07/2025) (cit. on p. 27).
- Park, David et al. (Aug. 2025). *FM4NPP: A Scaling Foundation Model for Nuclear and Particle Physics*. arXiv:2508.14087 [cs]. DOI: 10.48550/arXiv.2508.14087. URL: http://arxiv.org/abs/2508.14087 (visited on 10/09/2025) (cit. on pp. 28, 29).
- Radford, Alec, Karthik Narasimhan, et al. (2018). "Improving language understanding by generative pre-training". In: URL: https://web.archive.org/web/20210126024542/https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (cit. on pp. 10–17).

Bibliography (III)

- Radford, Alec, Jeffrey Wu, et al. (2019). "Language models are unsupervised multitask learners". In: *OpenAI blog* 1.8, p. 9. URL: https://web.archive.org/web/20210206183945/https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (cit. on pp. 10–17).
- Smith, Arthur (2019). "From PACS to PhySH". In: *Nature Reviews Physics* 1.1. for more details see https://physh.org/, pp. 8–11 (cit. on pp. 24, 25).