What Machine Learning Methods is Physics invested in?

Tobias Baier
Department of Physics
Technical University of Dresden
t.baier@hzdr.de

Steve Schmerler *
Helmholtz-Zentrum Dresden-Rossendorf / Helmholtz AI
Bautzner Landstrasse 400, 01328 Dresden, Germany
s.schmerler@hzdr.de

Peter Steinbach †
Helmholtz-Zentrum Dresden-Rossendorf / Helmholtz AI
Bautzner Landstrasse 400, 01328 Dresden, Germany
p.steinbach@hzdr.de

Abstract

The intersection of machine learning (ML) and physics is rapidly expanding, yet a comprehensive overview of which ML techniques are being adopted by physicists and within which sub-disciplines remains lacking. This work addresses this gap by presenting a large-scale bibliometric analysis of all accepted papers from the Machine Learning and the Physical Sciences (ML4PS) workshop at NeurIPS over the past years. We leverage Large Language Models (LLMs) to automatically classify each publication according to both its primary physics discipline and the employed ML task and methodology. This enables us to identify dominant trends in ML application across physics, and to track the evolution of these trends over time. Our results provide a quantitative snapshot of the current ML landscape within the physics community represented by ML4PS, highlighting prevalent approaches and emerging areas of research.

1 Introduction

The application of machine learning (ML) to physics problems has experienced strong growth in recent years, driven by the increasing availability of large datasets and the limitations of traditional analytical and computational approaches. While numerous successful applications have been demonstrated, a systematic, quantitative understanding of which ML methods are prevalent across different physics sub-disciplines remains largely absent. This knowledge gap hinders informed assessments of the fields progress, efficient resource allocation, and the identification of promising avenues for future research.

Current understanding relies heavily on anecdotal evidence and targeted reviews, lacking the breadth needed to capture the rapidly evolving landscape. A comprehensive analysis of ML method adoption is crucial for several reasons: to benchmark the effectiveness of various techniques against specific physics challenges, to identify potential transfer learning opportunities between disciplines and tasks, and to inform the development of novel algorithms specifically tailored to the unique characteristics of physical systems. Furthermore, understanding these trends is vital for educators seeking to prepare the next generation of physicists with the necessary skill set for this increasingly data-driven era. Our work addresses this critical need by presenting a large-scale, data-driven anal-

*corresponding authors
†corresponding authors

Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

ysis of machine learning methodologies employed within the physics research community. We use publicly available articles of [Har+25] as data source.

2 Related Work and Motivation

The relationship of physics and ML has been surveyed [LT24; Sur+24], discussed [SBO23] and reviewed [Men+25; Car+19; Kar+21]. While all of these formats of scientific publication yield merit within their own rights, they lack substantial insight into what is used in practice. We believe that a retrospective empirical analysis can add insights to the discussions mentioned above. In addition, our results can provide empirical grounding to prospective and sometimes speculating discussions.

The field of academic text classification is very broad and has been studied extensively. Early research restricted itself to classify articles by title and abstract only [RVA21; DLT19] or relied on extracted feature based predictions [McK+16]. More recently, full text classification methods with transformer based architectures have appeared [Arh+24]. For our study, we abstained from fine-tuning a LLM [RK24] due to the computational burden. In addition, context lengths of contemporary LLM models have surpassed the limit where full-text classification is impossible and fine-tuning is a necessity.

In our study, we would like to answer two scientific questions with the publications of the "Machine Learning and the Physical Sciences" workshop [Har+25] since its inception: 1. Which domains of physics do we believe to either contribute to the field of ML or profit from it? 2. What methods of ML are of interest to the physics community? We would like to add more facets to the discussion of ML in physics (and vice versa). Further, our results provide guidance to those with teaching duties, to those that provide and consume data as well as to those that can provide infrastructure to the physical science community. Finally, our results can surface those ML architectures or physics fields where more research might be required. We hope that our presented exploratory results give rise to a broader discussion on the junction of data science and physics.

3 Methods

3.1 Dataset

We collected all articles in PDF format from [Har+25] available in the publicly accessible GitHub repository. We used `marker-pdf` [Par24] to convert PDF documents to markdown files for further processing. The documents span the runtime of the workshop from 2017 to 2024 (except for 2018). We excluded any non-paper documents like posters and slides. As the markdown format was the target, we also excluded any form of images from the conversion, while tables are included in flattened form. Table 1 reports the number of papers we were able to convert to markdown format per year. Despite occasions where the conversion could not succeed, we converted 1067 out of 1070 articles. We also observed that the obtained markdown corpus exhibited a long-tailed distribution of number of lines. For this reason, we selected 5 % of the longest papers and reduced them to the first 506 lines of text. This affected 54 papers. By doing so, we restricted the number of tokens being processed by the LLM in addition. Based on samples from our validation set (Section A), we measured that 5500 tokens are required on average to encode a full paper. As the tokenizer is specific to the model family or even model version, we did not scrutinize this number to high precision.

Workshop Year	2017	2019	2020	2021	2022	2023	2024
Converted Papers	30	158	155	147	183	190	204

Table 1: Number of converted papers per year using `marker-pdf` [Par24].

3.2 Categorization into physics discipline

To study the distribution of workshop papers into physics disciplines, we employed the PhySH classification scheme [Smi24; Smi19]. We obtained the scheme from the publicly available API and limited ourselves to the top categories including examples of their sub-disciplines, which are listed in Section B.1.

To perform the classification, we employed few-shot prompting [Bro+20] of openly available large language models (LLM) APIs. The prompt (see Section C.1) was constructed from instructions about the task, three exemplars on what to reply given the title and a content place holder of a paper. Finally, we appended the full text of a given paper to the prompt. All models were used with `temperature` zero to restrict the model to an answers obtained by greedy search. The models in use were selected by availability and context length (ministral, Llama 3.1 , Gemma 3, all have $128k$ token context windows).

To assess the quality of our approach, we obtained 30 random papers of our corpus and manually classified them into PhySH categories. The Ministral model stands out from the other in all three metrics at hand (see Table 2). However, due to the limited size of the validation set, these numbers should not be taken at face value.

3.3 Categorization into ML methods

According to our literature review, there currently exists no accepted long-term classification scheme for scientific work in the ML or artificial intelligence community. Therefore, we opted for a pragmatic approach based on the table of contents of [Mur23]. Chapters serve as coarse top ML categories related to the task to solve. See the list of those 5 *ML task* categories in Section B.2. We further devised a subset of sections observed in these chapters, which gave rise to 21 *ML method* categories as reported in Section B.2. In line with Section 3.2, we performed an analysis of the quality of classification with the same models as above. Table 3 reports the quality of predictions of the ML tasks.

As we aspired to select only one top scoring model and valued accuracy over precision and recall, we selected the *Llama 3.1* model for the subsequent analysis. Our results from Table 2 and Table 3 not only underline the feasibility of our endeavor, but also the need for a larger validation set as well as improvements in prompting (see also [SH25]).

4 Results

We first present an overview of the obtained results in Figure 1. We show the frequency of articles in physics fields vs. ML model types and methods, aggregated over all years in our dataset. While we find that physics as a whole makes use of a wide variety of ML methods, some still stand out. For instance the usage of CNNs, as well as autoregressive models and transformers which hints at LLM applications. Further we find strong usage of MLPs, an architecture which is slightly out of fashion in other ML fields. One possible reason is that in physics (and other fields of science), many applications can be reduced to regression or classification problems for which MLPs are a flexible tool without the need for inductive bias such as in CNNs, for example.

Next, we like to unpack this aggregate view of the data and look at the adoption of ML methods over time. A time-resolved analysis of the data in Figure 1 can be found in Figure 2 which shows the adoption of ML methods in different fields in detail. We observe that from 2019 onwards, all investigated physics fields adopt ML techniques. The data also reflects trends from the wider ML community, for instance the decreased usage of GANs, possibly in favor of other generative approaches such as diffusion models, or the reduced usage of classical ML methods such as GPs over time. A more compact view into the time evolution is shown in Figure 3, which also includes the ML task classification. Here we analyze the article frequency w.r.t. our three classification targets (physics field, ML task, ML methods). We find that Gravitation, Cosmology & Astrophysics has the strongest publication record within the ML4PS workshop in recent years, followed by Particles & Fields and Condensed Matter, Materials & Applied Physics, which may hint at a preference of these fields to publish in ML4PS compared to, say, Physics of Living Systems (see Figure 4 for a plot containing all physics fields). In terms of ML tasks, Prediction, a classical ML application, is prevalent but shows a steady decrease in recent years, countered by an increasing usage of Generation methods, while other tasks remain on rather constant usage levels. When looking at methods, we note a decrease of CNN usage over time, possibly due to the success of newer architectures such as vision transformers. In addition we see a steady uptake in diffusion model usage. In summary, we find physics fields to adopt new ML methods over time, while also making continued use of more classical "tried and true" ML techniques such as predictive models (CNNs, MLPs).

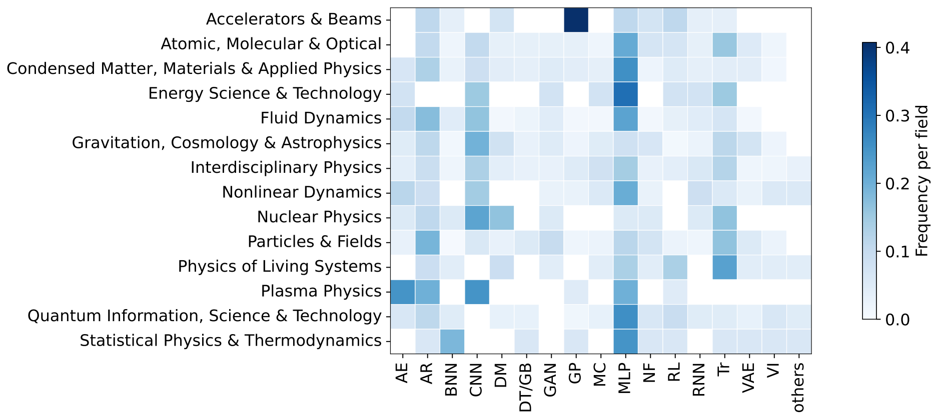

Figure 1: Articles in physics fields vs. ML model types and methods, aggregated over all years. We show data with least 10 articles per physics field / ML combination. To remove imbalance w.r.t. the number of articles per field, we do not produce a raw histogram but instead normalize row-wise by the total number of articles per field and report the article *frequency per field*. Please refer to Section B.2 for ML method abbreviations. Note that the strong usage of GPs in Accelerators & Beams is the result of high article counts in 2 years (see Figure 2) and could be regarded as an outlier.

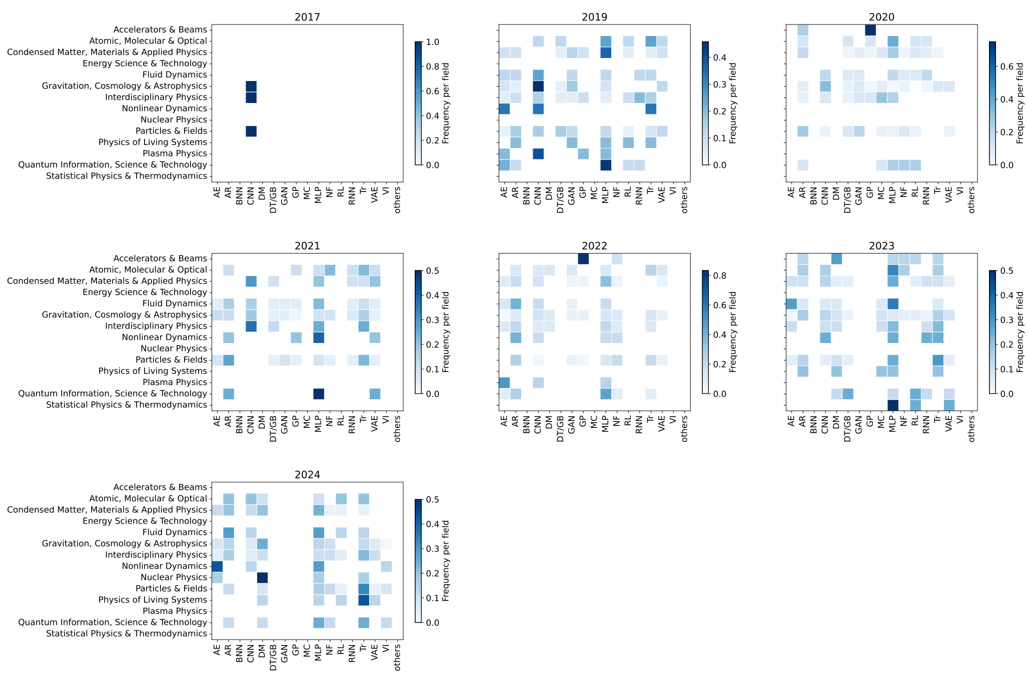

Figure 2: Articles in physics fields vs. ML model types and methods, per year (row-normalized separately, see Figure 1 for a time aggregation of the same data). A frequency of 1.0 represents cases where, in this year, articles assigned to one physics field report usage of one single ML method according to our LLM classification, hence all counts go into one bin. We show data with least 5 articles per physics field / ML combination per year.

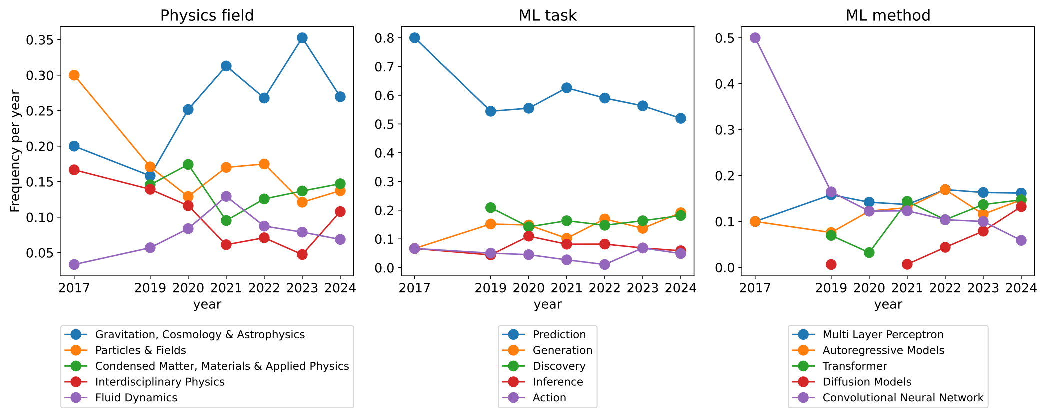

Figure 3: Article frequency over time for the 3 classification targets physics field (left), ML task (middle) and ML method (right). We are interested in the relative time evolution of categories and therefore normalize by the total number of articles per year across categories (colors) and report the article *frequency per year*. We show the top 5 (w.r.t. frequency) categories for better visibility. The full data including all categories is show in Figure 4.

5 Summary

In our study, we leveraged LLMs to classify scientific publications into their physics discipline, the ML task pursued as well as the ML methods in use. With this, we wish to instigate the bibliographic exploration of academic literature in which physics meets ML and vice verse. We observe that common trends of the ML community are reflected in this data. However, despite popular architectures (transformers, diffusion models) well working methods (MLPs) are found in our corpus throughout the years.

We like to emphasize the scope and limitations of this work by noting that by limiting our data source to ML4PS articles, our findings are biased towards the physics fields that tend to publish in this workshop. In addition to the produced results, this work may serve as a tooling template which can be applied to other and larger text corpora. On the methodological side, more research needs to be devoted to the technological challenges of analyzing large corpora of scientific articles as PDF files with LLMs regarding document conversion, prompting, reference ontologies in physics and machine learning or classification details, e.g. by relaxing the requirement of assigning a single prominent ML method per article.

Despite these limitations, we believe this avenue of research to be a promising direction to inform teaching, domain-specific operational decision making in the natural sciences and motivate research in ML methods.

Acknowledgements

We like to thank Philip Müller for reviewing the manuscript, valuable feedback on evaluating LLM replies and technical assistance with LLM setup and usage.

References

[Har+25] N. Hartmann et al. *Machine Learning and the Physical Sciences*. GitHub repository: https://github.com/ml4physicalsciences/ml4physicalsciences.github.io. 2017-2025. URL: https://ml4physicalsciences.github.io/2025/.

[LT24] H. Levine et al. "Machine learning meets physics: A two-way street". In: *Proceedings of the National Academy of Sciences of the United States of America* 121.27 (2024), e2403580121. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11228530/ (visited on 08/21/2025).

[Sur+24] R. Suresh et al. "Revolutionizing physics: a comprehensive survey of machine learning applications". In: *Frontiers in Physics* 12 (2024). URL: https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2024.1322162/full (visited on 08/21/2025).

[SBO23] A. Seyyedi et al. "Machine Learning and Physics: A Survey of Integrated Models". In: *ACM Comput. Surv.* 56.5 (2023), 115:1–115:33. URL: https://doi.org/10.1145/3611383 (visited on 08/21/2025).

[Men+25] C. Meng et al. "When physics meets machine learning: a survey of physics-informed machine learning". In: *Machine Learning for Computational Science and Engineering* 1.1 (2025), p. 20. URL: https://doi.org/10.1007/s44379-025-00016-0 (visited on 08/21/2025).

[Car+19] G. Carleo et al. "Machine learning and the physical sciences". In: *Reviews of Modern Physics* 91.4 (2019). arXiv:1903.10563 [physics], p. 045002. URL: http://arxiv.org/abs/1903.10563 (visited on 08/21/2025).

[Kar+21] G. Karagiorgi et al. *Machine Learning in the Search for New Fundamental Physics*. arXiv:2112.03769 [hep-ph]. 2021. URL: http://arxiv.org/abs/2112.03769 (visited on 08/21/2025).

[RVA21] M. Rivest et al. "Article level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling". In: *PloS one* 16.5 (2021), e0251493.

[DLT19] T. T. Dien et al. "Article classification using natural language processing and machine learning". In: *2019 international conference on advanced computing and applications (ACOMP)*. IEEE. 2019, pp. 78–84.

[McK+16] K. McKeown et al. "Predicting the impact of scientific concepts using full-text features". In: *Journal of the Association for Information Science and Technology* 67.11 (2016), pp. 2684–2696.

[Arh+24] C. Arhiliuc et al. "Journal article classification using abstracts: a comparison of classical and transformer-based machine learning methods". In: *Scientometrics* (2024), pp. 1–30.

[RK24] Z. R. K. Rostam et al. *Fine-Tuning Large Language Models for Scientific Text Classification: A Comparative Study*. 2024. URL: https://arxiv.org/abs/2412.00098.

[Par24] V. Paruchuri. *Marker - Convert PDF to markdown with high speed and accuracy.* 2024. URL: https://pypi.org/project/marker-pdf/0.1.3/.

[Smi24] A. Smith. *PhySH version 2.6.0*. 2024. URL: https://github.com/physh-org/PhySH/releases/tag/v2.6.0.

[Smi19] A. Smith. "From PACS to PhySH". In: *Nature Reviews Physics* 1.1 (2019). for more details see https://physh.org/, pp. 8–11.

[Bro+20] T. B. Brown et al. *Language Models are Few-Shot Learners*. 2020. URL: https://arxiv.org/abs/2005.14165.

[Mur23] K. P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL: https://probml.github.io/pml-book/book2.html.

[SH25] G. K. Shahi et al. *On the Effectiveness of Large Language Models in Automating Categorization of Scientific Texts*. 2025. URL: https://arxiv.org/abs/2502.15745.

Appendix

A Validation set

The validation set consists of these papers:

```
NeurIPS_ML4PS_2019_141.md
NeurIPS_ML4PS_2020_144.md
NeurIPS_ML4PS_2021_22.md
NeurIPS_ML4PS_2022_65.md
NeurIPS_ML4PS_2024_119.md
NeurIPS_ML4PS_2019_17.md
NeurIPS_ML4PS_2020_155.md
NeurIPS_ML4PS_2021_79.md
NeurIPS_ML4PS_2023_113.md
NeurIPS_ML4PS_2024_163.md
NeurIPS_ML4PS_2019_69.md
NeurIPS_ML4PS_2020_18.md
NeurIPS_ML4PS_2022_149.md
NeurIPS_ML4PS_2023_14.md
NeurIPS_ML4PS_2024_231.md
NeurIPS_ML4PS_2019_8.md
NeurIPS_ML4PS_2020_52.md
NeurIPS_ML4PS_2022_152.md
NeurIPS_ML4PS_2023_174.md
NeurIPS_ML4PS_2024_241.md
NeurIPS_ML4PS_2019_85.md
NeurIPS_ML4PS_2020_89.md
NeurIPS_ML4PS_2022_154.md
NeurIPS_ML4PS_2023_215.md
NeurIPS_ML4PS_2024_48.md
NeurIPS_ML4PS_2020_143.md
NeurIPS_ML4PS_2021_135.md
NeurIPS_ML4PS_2022_52.md
NeurIPS_ML4PS_2023_231.md
NeurIPS_ML4PS_2024_67.md
```

B Classification Categories

Below we list the categories which were used to classify each paper.

B.1 Physics categories

The PhySH classification scheme [Smi19] categories are:

1. Accelerators & Beams
2. Atomic, Molecular & Optical
3. Condensed Matter, Materials & Applied Physics
4. Energy Science & Technology
5. Fluid Dynamics
6. Gravitation, Cosmology & Astrophysics
7. Interdisciplinary Physics
8. Networks
9. Nonlinear Dynamics
10. Nuclear Physics
11. Particles & Fields

12. Physics Education Research
13. Physics of Living Systems
14. Plasma Physics
15. Polymers & Soft Matter
16. Quantum Information, Science & Technology
17. Statistical Physics & Thermodynamics

B.2 ML related categories

Regarding their ML content, we classified each paper into ML task and ML method. The *ML tasks* correspond to the main chapters of [Mur23]:

1. Action
2. Discovery
3. Generation
4. Inference
5. Prediction

Selected sections within the above chapters serve as *ML method* categories:

1. Autoencoders (AE)
2. Autoregressive Models (AR)
3. Bayesian Neural Network (BNN)
4. Convolutional Neural Network (CNN)
5. Diffusion Models (DM)
6. Decision Trees/Gradient Boosting (DT/GB)
7. Energy Based Model (EBM)
8. Generative Adversarial Networks (GAN)
9. Generalized Linear Models (GLM)
10. Gaussian processes (GP)
11. Latent Factor Models (LFM)
12. (Markov Chain) Monte Carlo (MC)
13. Multi Layer Perceptron (MLP)
14. Flows/Normalizing Flows (NF)
15. Reinforcement Learning (RL)
16. Recurrent Neural Network (RNN)
17. State-space Models (SSM)
18. Transformer (Tr)
19. Variational Autoencoders (VAE)
20. Variational Inference (VI)
21. others

C Prompting Details

We used few-shot prompting [Bro+20] with three exemplars. We also provided instructions on how the model should present a reply in order to extract the prediction.

C.1 Physics Discipline Classification

Instructions

```
You are a classifier. It is your job to sort papers into the disciplines of physics
    of APS Journals' PhySH classification scheme.
Things to consider:\n
- does the paper directly contribute to a discipline, by e.g. making an existing
    technique better -> choose that discipline\n
- does the paper introduce a technique and mention a possible usecase -> choose the
    discipline of the usecase\n
- does the paper introduce a technique and not mention a possible usecase -> choose
    the discipline, the technique belongs to
You are only allowed to pick A SINGLE CATEGORY! Have your answer formatting follow
    the formatting of the examples.
Choose from these categories:
```

The place-holders < rest of the paper > were kept as is.

Available Class Labels

During inference, these instruction were followed by the available class labels:

```
0: Accelerators & Beams
(concepts : Beam dynamics, Beam techniques, Cryogenics & vacuum technology,
    Electromagnetic field calculations, Radio frequency techniques, )
1: Atomic, Molecular & Optical
(concepts : Atom interferometry, Atomic & molecular collisions, Atomic & molecular
    processes in external fields, Atomic & molecular structure, Cold atoms & matter
     waves, Collective effects in atomic physics, Cooling & trapping, Laser systems
    , Nitrogen vacancy centers in diamond, Optical materials & elements, Optics &
    lasers, Photoemission, Schroedinger equation, Structured light, )
2: Condensed Matter, Materials & Applied Physics
(concepts : Artificial intelligence, Charge, Continuum mechanics, Devices,
    Electrical properties, Electronic structure, Electronically polarized systems,
    Functional materials, Junctions, Magnetism, Mechanical & acoustical properties,
     Mesoscopics, Optical & microwave phenomena, Phase transitions, Quantum master
    equation, Quasiparticles & collective excitations, Radio frequency techniques,
    Robotics, Structural properties, Superconductivity, Superfluidity, Surface &
    interfacial phenomena, Thermal properties, Topological phases of matter,
    Transport phenomena, )
3: Energy Science & Technology
(concepts : Energy sources, Energy storage, Energy utilization, Sustainability, )
4: Fluid Dynamics
(concepts : Acoustics, Aerodynamics, Artificial intelligence, Astrophysical fluid
    dynamics, Biological fluid dynamics, Boundary layers, Cavitation, Compressible
    flows, Convection, Drop & bubble phenomena, Electrokinetic flows, Flow control,
     Flow instability, Flows in porous media, Geophysical fluid dynamics, Granular
    flows, Interactions in fluids, Interfacial flows, Laminar flows, Low Reynolds
    number flows, Magnetohydrodynamics, Microfluidics, Multiphase flows,
    Nanofluidics, Nonlinear dynamics in fluids, Rarefied flows, Reacting flows,
    Rheology, Shear flows, Turbulence, Vortex flows, Waves and free surface flows,
    )
5: Gravitation, Cosmology & Astrophysics
(concepts : Cosmic rays & astroparticles, Cosmology, Electromagnetic radiation
    astronomy, Formation & evolution of stars & galaxies, Global Positioning System
    , Gravitation, Hydrostatic stellar nucleosynthesis, Large scale structure of
    the Universe, Transient & explosive astronomical phenomena, )
6: Interdisciplinary Physics
(concepts : Chemical Physics & Physical Chemistry, Complex systems, Engineering,
    Environmental research, Geophysics, Global Positioning System, History of
```

physics, Information & communication theory, Mathematical physics, Metrology, Physics & society, Social systems, Transportation research,)

7: Networks

(concepts : Artificial neural networks, Bayesian networks, Biological neural networks, Collective behavior in networks, Critical phenomena, Decision trees, Network evolution, Network formation & growth, Network optimization, Network phase transitions, Network searches, Network stability, Network structure, Renormalization, Self-organized criticality, Spreading, Support-vector machines , Synchronization, Traffic, Transport in networks,)

8: Nonlinear Dynamics

(concepts : Bifurcations, Chaos, Classical mechanics, Dynamics of networks, Pattern formation, Population dynamics, Synchronization,)

9: Nuclear Physics

(concepts : Lattice gauge theory, Nuclear astrophysics, Nuclear engineering, Nuclear reactions, Nuclear structure & decays, Nuclear tests of fundamental interactions, Relativistic heavy-ion collisions, Strong interaction, Synthesis of new superheavy elements,)

10: Particles & Fields

(concepts : Electroweak interaction, Hypothetical particle physics models, Lattice field theory, Naturalness, Particle astrophysics, Particle phenomena, Phase shift, Phenomenology, Quantum field theory, Strings & branes, Strong interaction, Sum rules, Total cross sections,)

11: Physics Education Research

(concepts : Assessment, Concepts & principles, Diversity & inclusion, Educational policy, Epistemology, attitudes, & beliefs, Instructional materials development , Instructional strategies, Learning environment, Learning theory, Research methodology, Scientific reasoning & problem solving, Student preparation, Technology,)

12: Physics of Living Systems

(concepts : Biomolecular & subcellular processes, Cellular organization, physiology & dynamics, Medical imaging, Medical physics & public health, Neuroscience, neural computation & artificial intelligence, Organismal, population, evolutionary & ecological systems, Renormalization group, Self-organized criticality,
)

13: Plasma Physics

(concepts : High-energy-density plasmas, Laboratory studies of space & astrophysical plasmas, Magnetohydrodynamics, Nonlinear phenomena in plasmas, Plasma acceleration & new acceleration techniques, Plasma discharges, Plasma fusion, Plasma interactions, Plasma ionization, Plasma medicine, Plasma opacity, Plasma optics, Plasma production & heating, Plasma thermodynamics, Plasma transport, Plasma waves, Radiation & particle generation in plasmas, Space weather,)

14: Polymers & Soft Matter

(concepts : Aging, Applications of soft matter, Chirality, Collective behavior, Confinement, Conformation & topology, Crystal phenomena, Electric field effects , Fluctuations & noise, Granular packing, Interactions & forces, Living matter & active matter, Mechanical deformation, Pattern formation, Patterning, Phase behavior, Phase transitions, Polymer behavior, Renormalization group, Rheology, Self-assembly, Surface & interfacial phenomena, Synthesis, Thermodynamics, Transport phenomena,)

15: Quantum Information, Science & Technology

(concepts : Quantum algorithms & computation, Quantum communication, protocols & technology, Quantum correlations, foundations & formalism,)

16: Statistical Physics & Thermodynamics

(concepts : Ballistic deposition, Classical statistical mechanics, Computational complexity, Conformal field theory, Equations of state, Fluctuations & noise, Fractals, Fractional dimension, Growth processes, Irreversible processes, Kinetic theory, Living matter & active matter, Nonequilibrium statistical mechanics, Optimization problems, Packing & jamming problems, Pattern formation , Percolation, Phase transitions, Population dynamics, Quantum statistical mechanics, Renormalization group, Self-avoiding walks, Self-organized systems, Stochastic processes, Synchronization, Thermodynamics, Tomography, Transport phenomena,)

Exemplars

This section was followed by few-shot examples:

```
Example 1:
Paper: # Inferring Parameters For Binary Black Hole Mergers Using Normalizing Flows

< rest of the paper >

Answer: The correct category is 5.
Reason: The paper uses Normalizing Flows to solve a Problem that lies in
    Astrophysics. Machine Learning is therefore just the tool. This makes the
    correct category Astrophysics. The correct category is 5.

Example 2:
Paper: # Score-Based Data Assimilation For A Two-Layer Quasi-Geostrophic Model

< rest of the paper >

Answer: The correct category is 6.
Reason: The paper presents advances in their calculations of geophysical models.
    This would put this paper under Geophysics. However there is no discipline
    called geophysics. Geophysics is present in the Interdisciplinary Physics
    category though. The correct category is 6.
```

Paper to Analyse

The prompt concluded by adding the entire paper:

```
Paper: #<title of paper>

<content of paper>

Answer:
```

C.2 ML Classification

For the *ML task* classification, we provided a prompt simimar to the one in Section C.1. We first instructed the model, then provided the available categories, added two examples and appended the paper text.

```
Read the paper provided to you below. You are a classifier. It is your job to sort
    papers into the given categories of General Machine Learning Tasks also listed
    below.
Things to consider:\n
- does the paper implement some Machine Learning technique, that falls directly into
     one of the tasks? -> choose that task
- does the paper introduce a new technique, that impacts performance of a task,
    hoever the technique belongs to a different task? -> choose the task, the
    technique belongs to
You are only allowed to pick A SINGLE CATEGORY! Have your answer formatting follow
    the formatting of the examples.
Choose from these categories:

1: Prediction; examples: generalized linear models, DNNs like MLPs, CNNs, RNNs, GNNs
     and Transformers, bayesian neural networks, gaussian processes, trees and
 ensembles,
2: Generation; examples: variational autoencoders, normalizing flows, diffusion
    models, GANs, autoregressive models (CNNs, Transformers, LLMs),
3: Discovery; examples: latent factor models, state-space models, graph learning,
    clustering, nonparametric bayesian models, representation learning with e.g.
 Autoencoders,
4: Action; examples: decision making under uncertainty, reinforcement learning,
    causality,
```

```
5: Inference; examples: Gaussian filtering and smoothing, Monte Carlo (general,
    Markov Chains, Sequential), Variational Inference, Message passing algorithms,


Example 1:
Paper: # Inferring Parameters For Binary Black Hole Mergers Using Normalizing Flows

< rest of the paper >

Answer: The correct category is 2.
Reason: The paper uses Normalizing Flows to solve a Problem that lies in
    Astrophysics. They do this by generating data, that feeds into a Predictor. The
     main
contribution is the data generation with the Normalizing Flow. The paper belongs
    into category 2 - Generation.

Example 2:
Paper: # Score-Based Data Assimilation For A Two-Layer Quasi-Geostrophic Model

< rest of the paper >

Answer: The correct category is 1.
Reason: The paper uses CNNs to simulate a model from Geophysics. Their proposed
    technique directly tries to solve the Prediction problem at hand and does not
contribute to a existing Prediction problem, using a model from a different category
    . The paper therefore belongs into Prediction.

Paper: <content of paper>

Answer:
```

We conducted the classification into *ML task* and *ML method* in two steps. An example for a *method* prompt is given below:

```
Read the paper provided to you below. You are a classifier. It is your job to sort
    papers into the given categories of Machine Learning techniques/architecturs
    also listed below.
Consider:
- if an architecture is used to solve a problem in a new way, choose the category
    the architecture belongs to
- if an architecture is used to improve the performance of a different architecture,
     than choose the first architecture (i.e. the one that lead to the improvement)
You are only allowed to pick A SINGLE CATEGORY! Have your answer formatting follow
    the formatting of the examples.
Choose from these categories:

1: Generative Adversarial Networks
2: Energy Based Model
3: Decision Trees/Gradient Boosting
4: Autoregressive Models (Causal CNNs, Transformers, LLMs)
5: Variational Autoencoders
6: Gaussian processes
7: Bayesian Neural Network
8: flows, normalizing flows
9: Diffusion models
10: Autoencoders
11: Generalized linear models
12: Transformer
13: Variational inference
14: (Markov Chain) Monte Carlo
15: Latent Factor Models
16: State-space models
17: Reinforcement Learning
18: Multi Layer Perceptron
19: Convolutional Neural Network
```

```
20: Recurrent Neural Network
21: others

Example 1:
Paper: # Generating Simulations data for Neutron Star Radiation Using A Diffusion
    Models

< rest of the paper >

Answer: The correct category is 5.
Reason: The paper uses Diffusion Models.

Example 2:
Paper: # Using Generative Adversarial Networks to model high energy jet formation at
     the LHC

< rest of the paper >

Answer: The correct category is 6.
Reason: The paper uses GANs.
```

D Classification Quality

As reported in Section 3.2 and 3.3, we used a validation corpus of 30 papers to assess the quality of the classification. Table 2 and 3 report these quality assessments.

Model	Accuracy / %	Precision / %	Recall / %
LLama 3.1	7.5	30.0	20.0
Ministral	10.0	34.3	23.7
Gemma 3	6.5	25.6	16.7

Table 2: Classification quality of three candidate LLMs given our few-shot prompt queries into all physics categories. The table was obtained from a validation set comprising 30 articles. Balanced Accuracy was used to account for class imbalances. Weighted Averaging was used for precision and recall.

Model	Accuracy / %	Precision / %	Recall / %
LLama 3.1	4.1	0.8	3.3
Ministral	2.7	3.6	6.67
Gemma 3	1.3	3.3	3.3

Table 3: Classification quality of three candidate LLMs given our few-shot prompt queries into all ML method categories. The table was obtained from a validation set comprising 30 articles. Balanced Accuracy was used to account for class imbalances. Weighted Averaging was used for precision and recall.

E Additional results

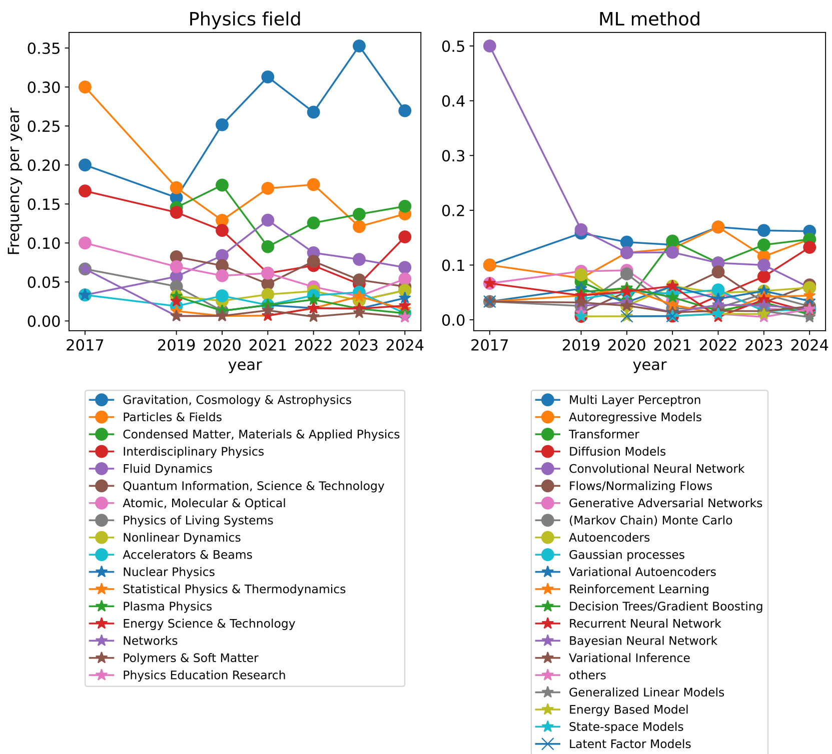

Figure 4: Article frequency over time for the 2 classification targets physics field (left) and ML method (right). We show all categories per classification target, while Figure 3 shows the top 5 categories.