Knowledge Graph in Astronomical Research with Large Language Models: Quantifying Driving Forces in Interdisciplinary Scientific Discovery

Abstract

Identifying and predicting the factors that contribute to the success of interdisciplinary research is crucial for advancing scientific discovery. However, there is a significant lack of methods to quantify the integration of new ideas and technological advancements within a field and how they trigger further scientific breakthroughs. Large language models, with their prowess in extracting key concepts from vast literature beyond keyword searches, provide a new tool to quantify such processes. In this study, we use astronomy as a case study to quantify this process. We extract concepts in astronomical research from 297,807 publications between 1993 and 2024 using large language models, resulting in a refined set of 24,939 concepts. These concepts are then adopted to form a knowledge graph, where the link strength between any two concepts is determined by their relevance based on the citation-reference relationships. By calculating this relevance across different time periods, we quantify the impact of numerical simulations and artificial intelligence on astronomical research, demonstrating the possibility of quantifying the gradual integration of interdisciplinary research and its further branching that leads to the flourishing of scientific domains.

1 Introduction

Interdisciplinary collaborations often drive innovation in research by introducing new theoretical, analytical, or computational tools to specific scientific domains. These new tools can revitalize and open up fields that might otherwise remain stagnant. For instance, the theoretical understanding of quantum physics and general relativity has driven much of modern cosmology [Weinberg, 2008], and each subsequent engineering breakthrough leads to new windows of observation. A prime example is the detection of gravitational waves with LIGO [Abbott *et al.*, 2016], which was made possible by the convergence of cutting-edge technologies in interferometry. Simultaneously, high-performance computing has paved the way for understanding complex systems in the cosmos, such as the evolution of galaxies [McAlpine *et al.*, 2016; Pillepich *et al.*, 2018] and the inner workings of stars and stellar atmospheres [Gudiksen *et al.*, 2011], through N-body or hydrodynamical simulations.

The advancement of astronomy also relies heavily on the revolution of statistical and analytical methods, which allow for proper inferences based on observations. The introduction of even well-known statistical techniques to astrophysics often leads to key turning points in the field. For example, a cornerstone of our understanding of cosmology comes from analyzing the power spectrum of the cosmic microwave background [Hu and Dodelson, 2002], while the detection of planetary systems outside the solar system has benefited from Gaussian Processes [Hara and Ford, 2023]. More recently, the advent of deep learning, with numerous successes in sciences such as AlphaFold [Jumper *et al.*, 2021], has propelled much of the field to rethink statistical inference in astronomy. This includes using generative models as surrogates for the likelihood or posterior [Cranmer *et al.*, 2020; Sun *et al.*, 2023a] and employing flow-based generative models to capture higher-order moment information in stochastic fields [Diaz Rivero and Dvorkin, 2020].

However, the underpinnings of these successful interdisciplinary results often stem from a rigorous process of debate and adaptation within the community. New thought processes are initially treated as disruptors, but a subset of these promising methods subsequently becomes integrated into the field's knowledge base. Over time, such integration gains significant traction and further creates branching of knowledge in the field, fostering its growth. Consider the example of numerical simulation, which was initially viewed as a "distraction" from pure mathematical interest in solving N-body problems and Navier-Stokes equations [Bertschinger, 1998]. However, astrophysics has gradually acknowledged that some aspects of the field are nonlinear and beyond analytical understanding. The integration of numerical simulations has subsequently led to the thriving study of galaxy evolution [McAlpine *et al.*, 2016], a widely researched topic, and has also gradually permeated into more specialized domains like solving the accretion physics of black holes and protoplanetary disks [Jiang *et al.*, 2014; Bai, 2016].

However, while such integration and branching off are intuitively clear, studying and quantifying them remains a challenge. Questions such as how long it might take for a field to adopt a new concept and what quantitative impact it has on the field still evades rigorous study. A key bottleneck is the difficulty in defining and extracting the various concepts described in a paper. The classical approach of classification using only keywords or the field [Xu *et al.*, 2018] of research might lack granularity. Other implicit methods that aim to extract vectorized semantic representations from papers [Meijer *et al.*, 2021] are hard to parse at the human level, let alone operate on such representations.

Recent breakthroughs in large language models (LLMs), particularly generalized pre-trained transformer techniques [Brown *et al.*, 2020; OpenAI *et al.*, 2023], have demonstrated exceptional zero-shot/few-shot capabilities across various downstream tasks and have shown broad domain knowledge coverage [Bubeck *et al.*, 2023]. The synergy between LLMs and knowledge graphs constitutes an active area of research. LLMs have shown reasonable performance in tasks such as entity identification for knowledge graph construction, and their capabilities can be significantly enhanced when coupled with knowledge graphs as external knowledge sources [Pan *et al.*, 2023; Zhu *et al.*, 2023].

Armed with this advancement, in this study, we explore the possibility of using LLMs as a bridging tool by distilling concepts from research papers in astronomy and astrophysics and constructing knowledge graphs to study their relationships and co-evolution over time. To the best of our knowledge, this is the first time an LLM-based knowledge graph has been constructed for astrophysics. The combination of the LLM-extracted concepts with our proposed citation-reference-based relevance allows us to quantitatively analyze cross-domain interactions over time and the co-evolution of subfields in astronomy.

This paper is organized as follows: In Section 2, we outline the dataset used for this study. Section 3 details the methodologies employed, including knowledge graph construction with large language model agents and the citation-reference-based relevance to quantify the interconnection between different concepts. We present our findings in Section 4, including a case study focusing on how numerical simulations were gradually adopted by the astronomical community, and by extension, quantifying the current impact of machine learning in astronomy. We discuss and conclude in Section 5.

2 Literature in Astronomical Research

This study employs a dataset of 297,807 arXiv papers in the fields of astronomy and astrophysics, collected from 1993 to 2024 and sourced from the NASA Astrophysics Data System (NASA/ADS) [Accomazzi, 2024]. Astrophysics is known to be a field where the vast majority of publications are on arXiv and easily searchable on ADS. Therefore, the number of arXiv papers here comprises a close-to-complete collection of literature that was published in the field.

We downloaded all PDFs from arXiv and performed OCR with Nougat [Blecher *et al.*, 2023]. Through human inspection, we found that Nougat did a great transcription of the data with minimal failure. The same set of data was currently used to train various specialized LLMs in astronomy (Pan et al., in prep., Arora et al,, in prep.), following AstroLLaMA and AstroLLaMA-Chat [Dung Nguyen *et al.*, 2023; Perkowski *et al.*, 2024], and auxiliary minor mistakes were identified and cleaned up during those iterations.

A key component of this paper is understanding the relation of concepts, as viewed by the research community, through the citation relation within the existing literature. The fact that NASA/ADS oversees a close to complete literature makes astronomy one of the well-curated fields to explore this study. We further extract the citation-reference relation for the entire corpus using the NASA/ADS API[1] to quantify the interaction among various scientific concepts during their co-evolution.

3 Constructing a Knowledge Graph for Astronomy

Constructing a knowledge graph between concepts in astrophysics requires two essential components: extracting the concepts in astronomical literature through large language model agents, and determining the strength of interconnectivity between concepts through the underlying relationships between paper citations. In this section, we explore these components in more detail.

3.1 Concept Extraction with Large Language Models

The key challenges in distilling concepts from publications using large language models are twofold. Firstly, LLM agents may generate hallucinations, producing lists of concepts that deviate from the expectations of human experts. Secondly, even when the concepts are accurately distilled, the models may yield concepts that are either too detailed, overly broad, or merely synonymous with each other, thereby diminishing the practical relevance of understanding their interrelationships. To address these challenges, we employ a multi-agent system in this study, as shown in Figure 1. This system consists of three parts: (a) extraction of concepts from astronomical publications; (b) nearest neighbor search of the concepts; and (c) merging of the concepts. This iterative approach enables control over the granularity of the knowledge graph, tailoring it to our purpose.

In this study, we focus on extracting key concepts from the titles and abstracts of astronomical publications to minimize computational cost. In astronomy, the abstract often encapsulates the essential information, including scientific motivation, methods, and data sources. The abstracts from the 300,000 papers amount to a total of approximately 2 billion tokens. To efficiently handle this large-scale data while maintaining cost-effectiveness, we leverage open-source large language models for concept extraction. Specifically, we employ MISTRAL-7B-INSTRUCT-V0.2[2] [Jiang *et al.*, 2023] as our inference model and JINA-EMBEDDINGS-V2-BASE-EN[3] [Günther *et al.*, 2023] for text embedding.

[1]https://ui.adsabs.harvard.edu/help/api/
[2]https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2
[3]https://huggingface.co/jinaai/jina-embeddings-v2-base-en

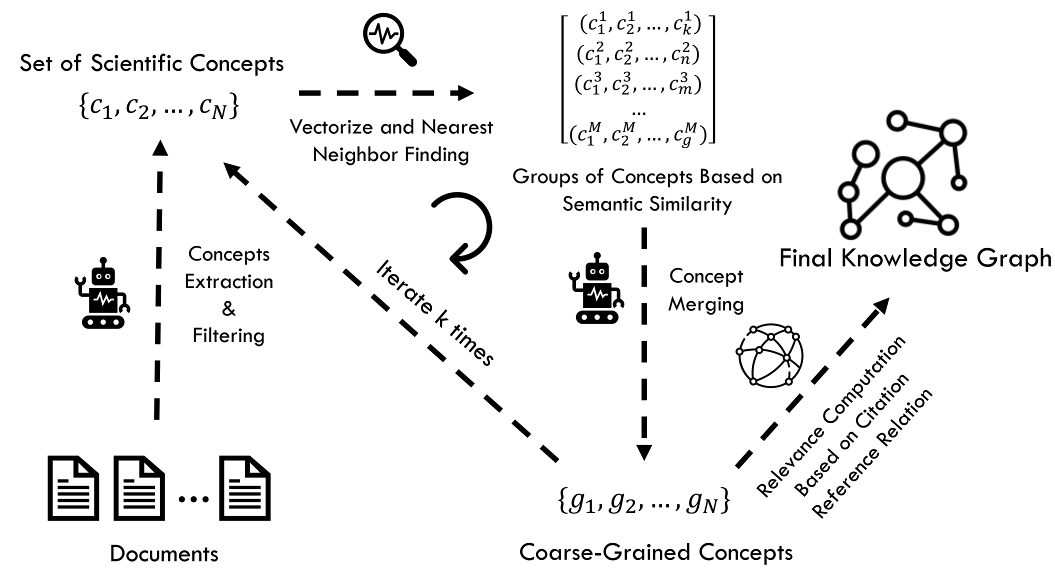

Figure 1: Schematic plot outlining the knowledge graph construction using large language model agents. The extraction of concepts comprises three main phases: (1) Concept Extraction, where agents construct scientific concepts from documents; (2) Vectorization and Nearest Neighbor Finding, in which concepts are vectorized and grouped by semantic similarity; (3) Concept Merging, where similar concepts are combined to form a more coarse-grained structures. The connections between concepts are then defined by citation-reference relevance as detailed in Section 3.2, with concepts involved in more citation-reference pairs assigned a higher relevance.

Concept Extraction: The first agent is prompted to extract a preliminary set of scientific concepts from the abstracts and titles[4]. While most of these concepts appear to be valid, some of them seem to be hallucinations that are not pertinent to astronomy, such as "misleading result" and "maternal entity in astronomy". To address this issue, a secondary LLM agent is deployed to explain and clarify each term, ensuring the removal of ambiguities and allowing only scientifically valid concepts to proceed. In this clarifying step, we utilize the entire document as an additional source enhanced by retrieval augmented generation to assist our agent in accurately understanding the meanings of various scientific terminologies. The validated scientific concepts are denoted as $\{c_1, c_2, \ldots, c_N\}$.

Vectorize and Nearest Neighbor Finding: Once the concepts are extracted and validated, they are transformed into vector representations using the text-embedding models, enabling the accurate computation of similarity measures. We group the concepts based on the cosine similarity of their corresponding vector representations into M clusters, represented as $\{\{c^i_j, j = 1, \ldots, k_i\}, i = 1, \ldots, M\}$. The number of elements in each cluster, k_i, is adaptively determined based on a predefined cosine similarity threshold among the elements within the cluster. In this study, we set the threshold at 0.85, striking a balance between the granularity of the concepts and the computational feasibility of the subsequent steps.

[4]All code and prompts will be made public after review.

Concept Merging: Finally, the final agent merges these grouped concepts by analyzing clusters of semantically similar concepts and distilling them into more general, unified entities. For example, the concepts "X-Shooter spectra", "Saturn's transmission spectrum," and "Keck LRIS spectrograph" were combined into the broader concept of "spectrograph". This merging simplifies the structure of the knowledge graph, reducing redundancy. Furthermore, a coarser knowledge graph improves the readability of the visualization.

We iterate the neighbour finding and merging steps three times, gradually coarsening the collection of concepts from 1,057,280, 164,352, and finally 24,797 concepts, respectively. We found, through domain expert evaluation that, the granularity of the concepts after three iterations is appropriate, with sufficient concepts covering the broad range of topics explored and methods employed in the literature, but with enough fine-grained level to understand the subtle evolution of the field in astrophysics. Some of the final concepts include the commonly known concepts such as "dark matter", "inflation", and etc. On average, each paper consists of ~ 10 concepts.

3.2 Determining Concept Relevance

Upon defining the concepts, perhaps more critical is to determine, quantitatively, how strongly two concepts are relevant. The relevancy of two concepts is certainly subjective—concepts that were deemed irrelevant at a certain point in time by the domain expert community might gradually become relevant over time. However, such temporal evolution is exactly what we are after to understand the shift of knowledge over time.

To gauge how two concepts are perceived as relevant by the community at a fixed point in time, the citation-reference relationships between articles become a natural annotated link between the concepts. In the following, we will define based on the probability with which a pair of concepts appears simultaneously in a certain article and its neighboring documents that have a citation-reference relationship, the proximity of the two concepts. This metric between concepts is inspired by the process by which researchers randomly sample through the network of articles from one concept to another. If the researcher can find another new concept from the parent concept that they were originally interested in by searching through the direct citation relation from the paper which contains the parent concept, and this leads the researcher to another paper with a new concept, the two concepts are deemed close. However, if the two concepts can only be found through a small subset of papers of the parent concepts and their citations or references, then the two concepts are deemed further apart at that point in time. We emphasize that while the linkage (and here, the hypothetical "search") is done through the domain of the published literature, the knowledge graph is constructed at the level of the extracted concepts.

More formally, let the final set of concepts be denoted as $\mathrm{C} : \{c_1, c_2, \ldots, c_n\}$, identified using large-language model-based agents as outlined in Section 3.1. Let these concepts be associated with a corpus of academic papers, $\mathrm{N} : \{n_1, n_2, \ldots, n_k\}$, and a set of citation-reference relationships $\mathrm{L} : \{(n_a, n_b) | n_a, n_b \in \mathrm{N}, \exists n_a \to n_b\}$, where $n_a \to n_b$ signifies that paper n_a cites paper n_b. To explore the propagation of a concept c_α within this network, we define the probability of encountering another concept c_β starting from a specific paper n_k that discusses c_α. This probability, denoted as $p_{\alpha\to\beta|n_k}$, is formulated as:

$$p_{\alpha\to\beta|n_k} = \frac{\mathrm{N}_\beta}{|S(n_k, \mathrm{L}, \beta)|}. \tag{1}$$

The set $S(n_k, \mathrm{L}, \beta)$ is defined through an iterative process starting with the initial paper set n_k (denoted as S_0). In each iteration, we expand the set by including papers that are directly cited by any paper in the current set and have not been included in previous sets. Formally, if S_{n-1} is the set of papers at iteration $n-1$, then $S_n = S_{n-1} \cup \{n_e | (n_s, n_e) \in \mathrm{L}, n_s \in S_{n-1}, n_e \notin S_{n-1}\}$. The iteration continues until at least one paper in the current set contains concept c_β, at which point we denote the final set as S_T and set $S_T = S(n_k, \mathrm{L}, \beta)$. The number of papers containing c_β within $S(n_k, \mathrm{L}, \beta)$ is set to be N_β.

Typically, the growth of the sets follows a pattern where $|S_0| = 1$, $|S_1| \sim 10^2$, and $|S_2| \sim 10^4$ in our experiments. This means that if the concepts cannot be found directly from a direct citation from the original paper that contains the parent concept, the number of papers "needed to be read", i.e., $|S|$, will drastically reduce the relevance of the two concepts. Nonetheless, if the concepts are very prevalent, after a certain level of search, the numerator N_β would then offset the volume of search.

As this probability pertains to just a specific paper containing concept c_α, the probability of transitioning from concept c_α to c_β, for all the papers S_α that contain c_α, would then be the expectation averaging over all papers in S_α, or,

$$p_{\alpha\to\beta} = \frac{1}{|S_\alpha|} \sum_{n_k \in S_\alpha} p_{\alpha\to\beta|n_k} \tag{2}$$

The above equation computes the average probability of moving from c_α to c_β across all papers that contain c_α. To assess the bidirectional relevance of concepts c_α and c_β, and we will assume that the order of transition between two concepts is not relevant, we define the citation-reference relevance between them as the geometric average of the probabilities of transitioning in both directions:

$$p_{\alpha,\beta} = (p_{\alpha\to\beta} \cdot p_{\beta\to\alpha})^{1/2} \tag{3}$$

Finally, the transition probability attains the following trivial properties: (1) $p_{\alpha,\beta} \leq 1, \forall c_\alpha, c_\beta \in \mathrm{C}$; (2) $p_{\alpha,\alpha} \equiv 1, \forall c_\alpha \in \mathrm{C}$; and (3) $p_{\alpha,\beta} = p_{\beta,\alpha}, \forall c_\alpha, c_\beta \in \mathrm{C}$. These properties ensure that the relevance metric is well-defined and consistent, providing a foundation for analyzing the relationships between concepts in the knowledge graph.

3.3 From Concept Relevance to Knowledge Graph

From the relevance defined as $p_{\alpha,\beta}$ above, which serves as a robust metric for the link strength between two nodes, we can visualize the knowledge as a force-directed graph. A force-directed graph [Kobourov, 2012; Bannister *et al.*, 2012], alternatively known as a spring-embedder or force-based layout, serves as a visual tool designed to illustrate relational data within network graphs. This method leverages simulation techniques inspired by physical systems, arranging nodes—which symbolize entities or concepts—and links—which depict the relationships or connections between these nodes—in an aesthetically coherent and insightful layout. These graphs utilize the concept of attraction and repulsion forces to strategically distribute nodes.

By iteratively updating the positions of nodes based on these attraction and repulsion forces, the force-directed graph algorithm converges to a layout that minimizes the overall energy of the system. This results in an informative 3D representation of the knowledge graph, where closely related concepts are automatically positioned near each other, enhancing the visibility of the density and connectivity within the graph. The capacity of force-directed graphs to dynamically represent complex relational data makes them particularly suitable for visualizing the knowledge graph.

In our context, the link strength between two nodes (concepts) is set to their citation-reference relevance, $p_{\alpha,\beta}$. Concepts with higher relevance will attract each other more strongly [Cheong *et al.*, 2021], causing them to be positioned closer together in the visualized graph. Conversely, the repulsion force is applied between all pairs of nodes, ensuring that they remain adequately spaced to prevent overlap and maintain clear visual separation. By leveraging the citation-reference relevance as the link strength between concepts, we can create a graph that intuitively conveys the relationships and clustering of ideas within the astronomical literature.

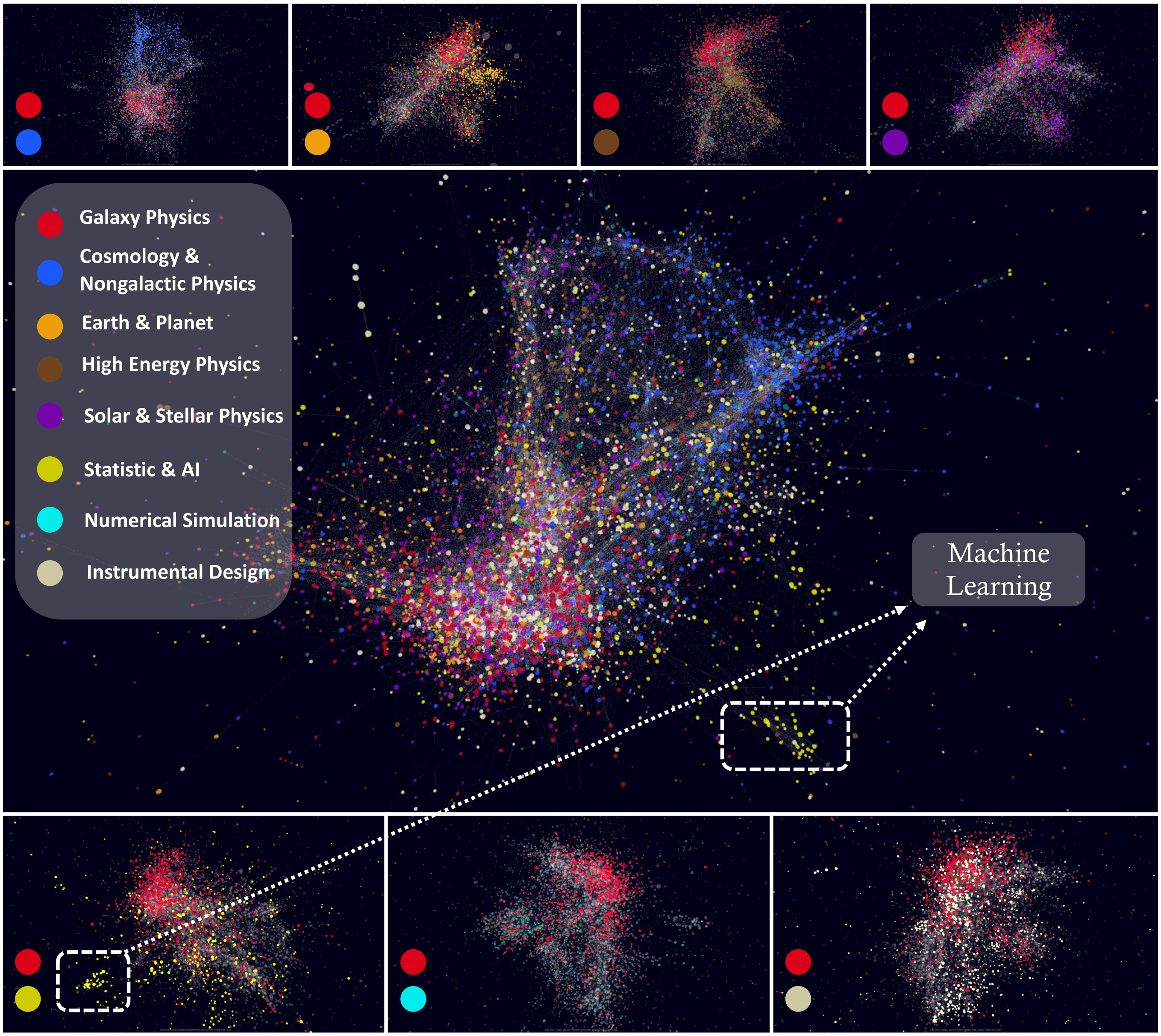

Figure 2: Visualization of a knowledge graph of 24,939 concepts, constructed from 297,807 astronomical research papers. Only concepts appearing in more than 20 papers and links with a link strength greater than 0.001 are displayed. Each concept is categorized into one of the following domains: (A) Galaxy Physics, (B) Cosmology & Nongalactic Physics, (C) Earth & Planetary Science, (D) High Energy Astrophysics, (E) Solar & Stellar Physics, (F) Statistics & AI, (G) Numerical Simulation, or (H) Instrumental Design. In the upper panels, we show connections between galaxy physics and other scientific domains. In the lower panel, we highlight the concepts from simulation, statistics, and observational instruments and their respective locations with respect to galaxy physics. Unsurprisingly, the technological concepts are generally more globally spread, as the same techniques can have wide implications for a broad range of topics in astronomy. Machine learning techniques are still at the periphery of the knowledge graph, suggesting that their integration in astronomy is still in its early stages. The interactive version of the knowledge graph is made publicly available after review.

4 Intersection between Technological Advancement and Scientific Discovery

Our knowledge graph consists of 24,939 concepts, extracted from 297,807 astronomical research papers, with 339,983,272 interconnections. The visualization of the knowledge graph as a force-directed graph is shown in Figure 2. The filamentous structure shown in the knowledge graph demonstrates the close interconnections across various subdomains within astronomical research.

For clarity, we only display concepts that appear in at least 20 papers and consider only those links with a citation-reference relevance $p_{\alpha,\beta} > 0.001$. This leads to 9,367 nodes and 32,494 links for the visualization. We set the size of the nodes to be proportional to the logarithm of their frequency of occurrence in the papers.

In the visualization, we further categorize all the concepts into scientific concepts, following the categorization of astrophysics on arXiv[5], namely Astrophysics of Galaxies,[6] Cosmology and Nongalactic Astrophysics,[7] Earth and Planetary Astrophysics,[8] High Energy Astrophysics,[9] and Solar and Stellar Astrophysics,[10]. As we aim to understand how concepts in technological advancement propel scientific discoveries, we further define another three classes of "technological" domains, which we identify as Statistics and Machine Learning, Numerical Simulation, and Instrumental Design. The classifications below are conducted using GPT-4[11].

Figure 2 illustrates how relevant concepts cluster within the same domain and how different domains interconnect. The upper panels demonstrate how the different scientific clusters interact with each other. For instance, galaxy physics, as anticipated, connects with both the largest scales in astronomical research, such as cosmology and general relativity, and the smaller scales, including stellar physics and planetary physics. The lower panel shows how the technological concepts are embedded within the scientific concepts, including numerical simulations, statistics, machine learning, and instrumental design. The technological concepts are generally distributed more globally in the knowledge graph, demonstrating their omnipresence in different subfields.

Interestingly, as shown in the figure, despite the booming interest and popularity, machine learning techniques, particularly deep learning, are situated only at the peripheral region of the knowledge graph. This suggests that machine learning techniques are not yet fully integrated into the astronomical research community, at least from the citation-reference point of view. We will provide a more quantitative comparison of this observation in the following section.

[5]https://arxiv.org/archive/astro-ph

[6]Astrophysics of Galaxies focuses on phenomena related to galaxies and the Milky Way, including star clusters, interstellar medium, galactic structure, formation, dynamics, and active galactic nuclei.

[7]Cosmology and Nongalactic Astrophysics covers the early universe's phenomenology, cosmic microwave background, dark matter, cosmic strings, and the large-scale structure of the universe.

[8]Earth and Planetary Astrophysics studies deal with the interplanetary medium, planetary physics, extrasolar planets, and the formation of the solar system.

[9]High Energy Astrophysics explores cosmic ray production, gamma ray astronomy, supernovae, neutron stars, and black holes.

[10]Solar and Stellar Astrophysics pertains to the investigation of white dwarfs, star formation, stellar evolution, and helioseismology.

[11]https://openai.com/index/gpt-4/

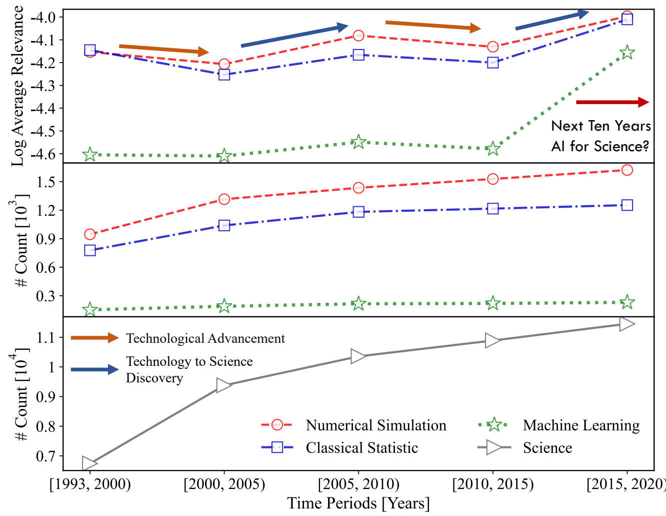

Figure 3: The citation-reference relevance for five distinct time periods to investigate the temporal integration of technological techniques into scientific research. The middle and lower panels illustrate a consistent increase in the count of concepts, both in terms of scientific concepts (bottom panel) and technical concepts (middle panel). The upper panel shows the total cross-linkage between individual technical domains and scientific concepts, with lower values indicating stronger adoption. The upper panel reveals a two-phase evolution, with an observed latency of approximately five years. The two phases signify the period of development and introduction of new techniques in astronomy and their subsequent adoption by the community (see text for details). While still modest, machine learning has begun to reach integration levels comparable to those of numerical simulations seen two decades earlier.

4.1 Numerical Simulations in Astronomy

To demonstrate how technological advancement drives scientific discovery, we will study in more depth the impact of numerical simulations on astronomy. In modern-day astronomical research, numerical simulation has become an indispensable tool. However, this was not always the case. The scientific community experienced a gradual transition from focusing primarily on theoretical deduction and analytical formulas to modeling complex phenomena through numerical simulations.

To understand this transition, we assess the average relevance between numerical simulations and scientific concepts across various time periods. We divided the dataset into five time periods from 1993 to 2020. In each time period, we recalculate the citation-reference relevance using the papers published within that specific timeframe.

As shown in the bottom panel of Figure 3, unsurprisingly, the number of "scientific concepts" has surged over time. Complementary to these scientific concepts, we also see that the number of technical concepts has surged alongside, especially in terms of numerical simulations and statistical methods, which are shown as red and blue lines in the middle panel. On the other hand, despite the interest in the field, the number of concepts in machine learning in the astronomical literature, as shown by the green line, is still an order of magnitude lagging behind these other well-developed technological concepts.

Perhaps more interesting is showing the weighted "intersection" between the scientific concepts and the technical concepts, which is shown in the top panels. The top panel shows the weighted "linkage" among all the scientific concepts with the specific technical domain. If the new methods are well-adopted in the astronomical community and advance scientific discovery, we should see an improvement in the average citation-reference linkage (large values in the top panel). Viewed this way, there is a clear two-phase evolution with the gradient of the integration oscillating positively (blue arrow) and negatively (red arrow).

This is perhaps not surprising. For any technological advancement, it might once be proposed with many technically focused papers written; however, the citation-reference relation is mostly limited to the "technologists," leading to a dilution of the cross-correlation, which is shown by the red arrow. For example, during the period of 1993-2000, there have been many works focusing on the development of N-body simulation techniques [**?**; Romeo *et al.*, 2004; Springel, 2005]. Yet, the integration remains marginal. However, from 2000 onward, the astronomical community began to embrace N-body simulations to resolve scientific questions [Paz *et al.*, 2006; Peñarrubia *et al.*, 2006; Zhou and Lin, 2007], resulting in a increase in citation-reference relevance during this time. A similar two-phase pattern is observed from [2010, 2015) to [2015, 2020), during which time hydrodynamical simulations developed [Genel *et al.*, 2014; Carlesi *et al.*, 2014b; Carlesi *et al.*, 2014a] and gradually gained acceptance [McAlpine *et al.*, 2016; Pillepich *et al.*, 2018] within the community. The delay between the development of new technologies and their impact on scientific discovery spans approximately five years.

4.2 Machine Learning in Astrophysics

The revelation of the two-phase adoption in numerical simulations leads to the possibility of better quantifying the integration of machine learning in astronomy. In recent years, we have seen a booming interest in AI and its applications in science. As modern-day astronomy is driven by big data, with billions of sources routinely being surveyed, it is not surprising that astronomy has also seen a drastic integration of AI to advance data processing and analysis [Baron, 2019].

Figure 4 shows the average cross-domain linkage, as defined in the top panel of Figure 3, but between the concepts in machine learning and the five scientific domains. In terms of the application of machine learning in astronomy, Cosmology & Nongalactic Astrophysics takes the lead, as it benefits from machine learning's capacity to manage complex, large data sets from simulations and surveys [Villaescusa-Navarro *et al.*, 2021b; Villaescusa-Navarro *et al.*, 2021a; Sun *et al.*, 2023b]. This is followed by Galaxy Physics, which leverages ML for tasks like photometric redshift prediction [Sun *et al.*, 2023a] and galactic morphology classification [Robertson *et al.*, 2023]. Solar and Stellar Physics have also shown promise in emulating and analyzing stellar spectra [Ting *et al.*, 2019]. High Energy Astrophysics and Earth & Planetary Astrophysics have been slower to adopt ML.

But is machine learning now well-adopted in astronomical research? Figures 2 and 3 paint an interesting picture. On the one hand, the top panel of Figure 3 shows that there has been a rapid increase in the cross-science-and-AI citation-reference relevance, demonstrating a huge interest among the community. For instance, the scientific-technology score remains flat and low before 2015, signifying that despite a history of AI in astronomy—such as the use of neural networks for galaxy morphology classification as early as 1992 [Storrie-Lombardi *et al.*, 1992]—its impact remained minimal until the surge in popularity of deep learning post-2015.

Yet, at the same time, even currently, Figure 2 shows that most of these concepts still occupy a peripheral position in the knowledge graph. This suggests that, from a citation-reference relevance perspective, such concepts are still considered niche within the broader scientific community. This is perhaps not too surprising because, compared to the deep integration of numerical simulations, quantitatively, the cross-linkage score of machine learning with astronomy remains only at the level that numerical simulations and traditional statistics were twenty years ago.

Perhaps what is strikingly lacking is that the number of machine learning concepts in the astronomical literature remains an order of magnitude smaller than that of numerical simulations, as shown in the middle panel of Figure 3. This might imply that the machine learning techniques widely adopted in astronomy, even at present, remain some of the more classical techniques, such as linear regression and random forests [Nyheim *et al.*, 2024]. The rapid adoption of "existing" techniques, while encouraging, might also signify a bigger underlying problem of lack of innovation in applying AI to astronomy. However, if the two-phase evolution applies, we should expect that in the coming years, there will be more novel deep learning techniques introduced before they are gradually adopted by the community.

5 Discussions and Conclusions

A quantitative study of the evolution of concepts and their interconnections would not be possible without modern-day LLMs, partly due to the large amount of arduous work required to manually label, extract concepts, and classify topics, which can be easily done with minimal computing resources in our case. Even when manual extraction is possible, the taxonomy of a scientific field is often limited—tailored to provide vague contours of the domain, e.g., for publication purposes, rather than a deep and more fine-grained differentiation of the knowledge embedded in the field.

In this study, we construct, to the best of our knowledge, the first LLM-based knowledge graph in the domain of astronomy and astrophysics. The knowledge graph comprises 24,939 concepts extracted through a careful iterative process with LLMs from 297,807 papers. We design a relevance metric defined through the citation-reference relations in the as-

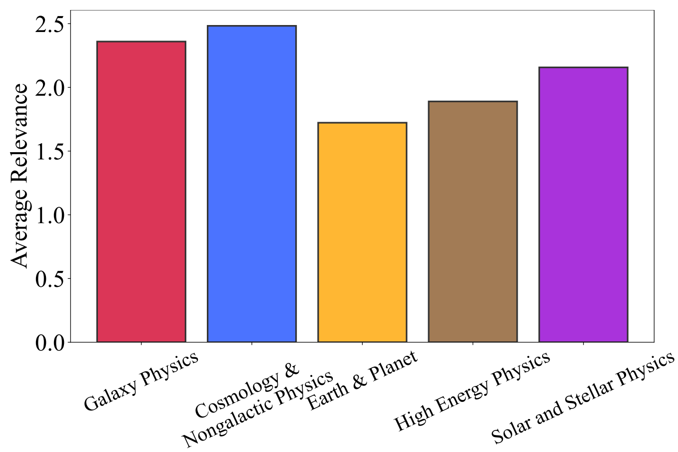

Figure 4: Integration of machine learning in different subfields of astronomy. The integration is defined as the average cross-domain linkage similar to the top panel of Figure 3. Cosmology and Nongalactic Astrophysics currently lead the application of machine learning in astronomy, followed by Galaxy Physics and Solar & Stellar Physics. The adoption of machine learning concepts in Earth & Planetary Physics and High Energy Astrophysics still lags behind.

tronomical literature to understand the relations as well as the temporal evolution between different concepts. The relevance metric follows the intuition of how humans search for new concepts by quantifying the degree of separation in the citation network as well as the prevalence of the concepts in the field. The relevance is then applied as the linkage strength of the force-directed graph to construct the knowledge graph, allowing us to visualize the knowledge in the field in detail.

Based on this knowledge graph, we evaluate the temporal evolution of the relevance of numerical simulations and machine learning in astronomical research. We showed that while numerical simulations are routinely adopted in modern-day astronomy, the concepts related to them have gone through a long process of gradually being integrated into and accepted by the community. We also found that the integration of numerical simulation into scientific discovery shows a two-phase process, in which a five-year latency can be observed between the development of techniques, where the relevance of the techniques and the science might temporarily diminish, followed by the flourishing period, where the methods mature and are widely applied to astronomical research. We also found that the same trend can be found in classical statistical analysis.

By the same metric, we found that, despite much of the interest and the booming field of deep learning, the impact of deep learning in astronomy remains marginal. While there is a drastic increase in the technique-science cross-referencing, quantitatively, the referencing remains at a level that we observed for numerical simulations about two decades ago. Furthermore, the number of machine learning concepts introduced in astronomy remains an order of magnitude smaller than that of numerical simulations and classical statistical methods, which might imply that the current rapid increase in relevance is driven mainly by the adoption of established machine learning techniques from decades ago. Nonetheless, if the two-phase transition applies, we expect more innovative techniques will be gradually introduced. In fact, in recent years, we have seen many more modern-day techniques, both in terms of flow-based and score-based generative models [De Santi *et al.*, 2024; Zhao *et al.*, 2023], being introduced, as well as, like this study, the application of LLMs in astronomical research [Dung Nguyen *et al.*, 2023; Perkowski *et al.*, 2024]. The metric introduced here will be able to continue monitoring this process.

This study primarily aims to show a proof of concept, using LLM-based Knowledge Graph to quantifiably understand the evolution of astronomical research. As such our study certainly has much room for improvement. For instance, proper robust extraction of scientific concepts from literature heavily relies on the alignment between the agents and the researchers' perception. In our study, the concepts are autonomously extracted through the LLM agent, with the granularity of the concepts optimized through merging and pruning. Such an LLM agent can certainly benefit from a subset of high-quality annotated data and comparison with existing hierarchical taxonomies. The process of concept pruning and merging is also somewhat crude, involving vectorizing the concepts and performing a cosine similarity search. A better method would involve further comparing these concepts, utilizing the capabilities of large language models for more detailed concept differentiation and pruning.

In a nutshell, our study demonstrates the potential of LLM-based knowledge graphs in uncovering the intricate relationships and evolution of astronomical research. By providing a quantitative framework for analyzing the integration of new technologies and methodologies, this approach opens up new avenues for understanding the dynamics of interdisciplinary research and the factors that drive scientific progress, in astronomy and beyond.

Ethical Statement

In this study, we construct a knowledge graph by extracting concepts from the astronomical literature available on the arXiv preprint server. Our work aims to advance the understanding of the evolution and interconnections of scientific concepts within the field of astronomy. We emphasize that our study does not involve the direct reproduction or distribution of the original literature itself. Instead, we focus on distilling and analyzing the key concepts present in the existing body of work.

To ensure ethical compliance and respect for intellectual property rights, we will only release the extracted concepts and their relationships, without sharing or reproducing the original text or any substantial portions of the literature. This approach minimizes the risk of copyright infringement and maintains the integrity of the original authors' works.

Furthermore, the field of astronomical research generally operates under an open-sky policy, which promotes collaboration, transparency, and the free exchange of scientific knowledge. This policy aligns with our research objectives and mitigates potential ethical or monetary disputes arising from our work. Our goal is to provide insights that benefit the astronomical community and contribute to the advancement of scientific understanding.

Acknowledgments

References

[Abbott *et al.*, 2016] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, B. Allen, A. Allocca, P. A. Altin, S. B. Anderson, W. G. Anderson, K. Arai, M. A. Arain, M. C. Araya, C. C. Arceneaux, J. S. Areeda, N. Arnaud, K. G. Arun, S. Ascenzi, G. Ashton, M. Ast, S. M. Aston, P. Astone, P. Aufmuth, C. Aulbert, S. Babak, P. Bacon, M. K. M. Bader, P. T. Baker, F. Baldaccini, G. Ballardin, S. W. Ballmer, J. C. Barayoga, S. E. Barclay, B. C. Barish, D. Barker, F. Barone, B. Barr, L. Barsotti, M. Barsuglia, D. Barta, J. Bartlett, M. A. Barton, I. Bartos, R. Bassiri, A. Basti, J. C. Batch, C. Baune, V. Bavigadda, M. Bazzan, B. Behnke, M. Bejger, C. Belczynski, A. S. Bell, C. J. Bell, B. K. Berger, J. Bergman, G. Bergmann, C. P. L. Berry, D. Bersanetti, A. Bertolini, J. Betzwieser, S. Bhagwat, R. Bhandare, I. A. Bilenko, G. Billingsley, J. Birch, R. Birney, O. Birnholtz, S. Biscans, A. Bisht, M. Bitossi, C. Biwer, M. A. Bizouard, J. K. Blackburn, C. D. Blair, D. G. Blair, R. M. Blair, S. Bloemen, O. Bock, T. P. Bodiya, M. Boer, G. Bogaert, C. Bogan, A. Bohe, P. Bojtos, C. Bond, F. Bondu, R. Bonnand, B. A. Boom, R. Bork, V. Boschi, S. Bose, Y. Bouffanais, A. Bozzi, C. Bradaschia, P. R. Brady, V. B. Braginsky, M. Branchesi, J. E. Brau, T. Briant, A. Brillet, M. Brinkmann, V. Brisson, P. Brockill, A. F. Brooks, D. A. Brown, D. D. Brown, N. M. Brown, C. C. Buchanan, A. Buikema, T. Bulik, H. J. Bulten, A. Buonanno, D. Buskulic, C. Buy, R. L. Byer, M. Cabero, L. Cadonati, G. Cagnoli, C. Cahillane, J. Calderón Bustillo, T. Callister, E. Calloni, J. B. Camp, K. C. Cannon, J. Cao, C. D. Capano, E. Capocasa, F. Carbognani, S. Caride, J. Casanueva Diaz, C. Casentini, S. Caudill, M. Cavaglià, F. Cavalier, R. Cavalieri, G. Cella, C. B. Cepeda, L. Cerboni Baiardi, G. Cerretani, E. Cesarini, R. Chakraborty, T. Chalermsongsak, S. J. Chamberlin, M. Chan, S. Chao, P. Charlton, E. Chassande-Mottin, H. Y. Chen, Y. Chen, C. Cheng, A. Chincarini, A. Chiummo, H. S. Cho, M. Cho, J. H. Chow, N. Christensen, Q. Chu, S. Chua, S. Chung, G. Ciani, F. Clara, J. A. Clark, F. Cleva, E. Coccia, P. F. Cohadon, A. Colla, C. G. Collette, L. Cominsky, M. Constancio, A. Conte, L. Conti, D. Cook, T. R. Corbitt, N. Cornish, A. Corsi, S. Cortese, C. A. Costa, M. W. Coughlin, S. B. Coughlin, J. P. Coulon, S. T. Countryman, P. Couvares, E. E. Cowan, D. M. Coward, M. J. Cowart, D. C. Coyne, R. Coyne, K. Craig, J. D. E. Creighton, T. D. Creighton, J. Cripe, S. G. Crowder, A. M. Cruise, A. Cumming, L. Cunningham, E. Cuoco, T. Dal Canton, S. L. Danilishin, S. D'Antonio, K. Danzmann, N. S. Darman, C. F. Da Silva Costa, V. Dattilo, I. Dave, H. P. Daveloza, M. Davier, G. S. Davies, E. J. Daw, R. Day, S. De, D. DeBra, G. Debreczeni, J. Degallaix, M. De Laurentis, S. Deléglise, W. Del Pozzo, T. Denker, T. Dent, H. Dereli, V. Dergachev, R. T. DeRosa, R. De Rosa, R. DeSalvo, S. Dhurandhar, M. C. Díaz, L. Di Fiore, M. Di Giovanni, A. Di Lieto, S. Di Pace, I. Di Palma, A. Di Virgilio, G. Dojcinoski, V. Dolique, F. Donovan, K. L. Dooley, S. Doravari, R. Douglas, T. P. Downes, M. Drago, R. W. P. Drever, J. C. Driggers, Z. Du, M. Ducrot, S. E. Dwyer, T. B. Edo, M. C. Edwards, A. Effler, H. B. Eggenstein, P. Ehrens, J. Eichholz, S. S. Eikenberry, W. Engels, R. C. Essick, T. Etzel, M. Evans, T. M. Evans, R. Everett, M. Factourovich, V. Fafone, H. Fair, S. Fairhurst, X. Fan, Q. Fang, S. Farinon, B. Farr, W. M. Farr, M. Favata, M. Fays, H. Fehrmann, M. M. Fejer, D. Feldbaum, I. Ferrante, E. C. Ferreira, F. Ferrini, F. Fidecaro, L. S. Finn, I. Fiori, D. Fiorucci, R. P. Fisher, R. Flaminio, M. Fletcher, H. Fong, J. D. Fournier, S. Franco, S. Frasca, F. Frasconi, M. Frede, Z. Frei, A. Freise, R. Frey, V. Frey, T. T. Fricke, P. Fritschel, V. V. Frolov, P. Fulda, M. Fyffe, H. A. G. Gabbard, J. R. Gair, L. Gammaitoni, S. G. Gaonkar, F. Garufi, A. Gatto, G. Gaur, N. Gehrels, G. Gemme, B. Gendre, E. Genin, A. Gennai, J. George, L. Gergely, V. Germain, Abhirup Ghosh, Archisman Ghosh, S. Ghosh, J. A. Giaime, K. D. Giardina, A. Giazotto, K. Gill, A. Glaefke, J. R. Gleason, E. Goetz, R. Goetz, L. Gondan, G. González, J. M. Gonzalez Castro, A. Gopakumar, N. A. Gordon, M. L. Gorodetsky, S. E. Gossan, M. Gosselin, R. Gouaty, C. Graef, P. B. Graff, M. Granata, A. Grant, S. Gras, C. Gray, G. Greco, A. C. Green, R. J. S. Greenhalgh, P. Groot, H. Grote, S. Grunewald, G. M. Guidi, X. Guo, A. Gupta, M. K. Gupta, K. E. Gushwa, E. K. Gustafson, R. Gustafson, J. J. Hacker, B. R. Hall, E. D. Hall, G. Hammond, M. Haney, M. M. Hanke, J. Hanks, C. Hanna, M. D. Hannam, J. Hanson, T. Hardwick, J. Harms, G. M. Harry, I. W. Harry, M. J. Hart, M. T. Hartman, C. J. Haster, K. Haughian, J. Healy, J. Heefner, A. Heidmann, M. C. Heintze, G. Heinzel, H. Heitmann, P. Hello, G. Hemming, M. Hendry, I. S. Heng, J. Hennig, A. W. Heptonstall, M. Heurs, S. Hild, D. Hoak, K. A. Hodge, D. Hofman, S. E. Hollitt, K. Holt, D. E. Holz, P. Hopkins, D. J. Hosken, J. Hough, E. A. Houston, E. J. Howell, Y. M. Hu, S. Huang, E. A. Huerta, D. Huet, B. Hughey, S. Husa, S. H. Huttner, T. Huynh-Dinh, A. Idrisy, N. Indik, D. R. Ingram, R. Inta, H. N. Isa, J. M. Isac, M. Isi, G. Islas, T. Isogai, B. R. Iyer, K. Izumi, M. B. Jacobson, T. Jacqmin, H. Jang, K. Jani, P. Jaranowski, S. Jawahar, F. Jiménez-Forteza, W. W. Johnson, N. K. Johnson-McDaniel, D. I. Jones, R. Jones, R. J. G. Jonker, L. Ju, K. Haris, C. V. Kalaghatgi, V. Kalogera, S. Kandhasamy, G. Kang, J. B. Kanner, S. Karki, M. Kasprzack, E. Katsavounidis, W. Katzman, S. Kaufer, T. Kaur, K. Kawabe, F. Kawazoe, F. Kéfélian, M. S. Kehl, D. Keitel, D. B. Kelley, W. Kells, R. Kennedy, D. G. Keppel, J. S. Key, A. Khalaidovski, F. Y. Khalili, I. Khan, S. Khan, Z. Khan, E. A. Khazanov, N. Kijbunchoo, C. Kim, J. Kim, K. Kim, Nam-Gyu Kim, Namjun Kim, Y. M. Kim, E. J. King, P. J. King, D. L. Kinzel, J. S. Kissel, L. Kleybolte, S. Klimenko, S. M. Koehlenbeck, K. Kokeyama, S. Koley, V. Kondrashov, A. Kontos, S. Koranda, M. Korobko, W. Z. Korth, I. Kowalska, D. B. Kozak, V. Kringel, B. Kr-

ishnan, A. Królak, C. Krueger, G. Kuehn, P. Kumar, R. Kumar, L. Kuo, A. Kutynia, P. Kwee, B. D. Lackey, M. Landry, J. Lange, B. Lantz, P. D. Lasky, A. Lazzarini, C. Lazzaro, P. Leaci, S. Leavey, E. O. Lebigot, C. H. Lee, H. K. Lee, H. M. Lee, K. Lee, A. Lenon, M. Leonardi, J. R. Leong, N. Leroy, N. Letendre, Y. Levin, B. M. Levine, T. G. F. Li, A. Libson, T. B. Littenberg, N. A. Lockerbie, J. Logue, A. L. Lombardi, L. T. London, J. E. Lord, M. Lorenzini, V. Loriette, M. Lormand, G. Losurdo, J. D. Lough, C. O. Lousto, G. Lovelace, H. Lück, A. P. Lundgren, J. Luo, R. Lynch, Y. Ma, T. MacDonald, B. Machenschalk, M. MacInnis, D. M. Macleod, F. Magaña-Sandoval, R. M. Magee, M. Mageswaran, E. Majorana, I. Maksimovic, V. Malvezzi, N. Man, I. Mandel, V. Mandic, V. Mangano, G. L. Mansell, M. Manske, M. Mantovani, F. Marchesoni, F. Marion, S. Márka, Z. Márka, A. S. Markosyan, E. Maros, F. Martelli, L. Martellini, I. W. Martin, R. M. Martin, D. V. Martynov, J. N. Marx, K. Mason, A. Masserot, T. J. Massinger, M. Masso-Reid, F. Matichard, L. Matone, N. Mavalvala, N. Mazumder, G. Mazzolo, R. McCarthy, D. E. McClelland, S. McCormick, S. C. McGuire, G. McIntyre, J. McIver, D. J. McManus, S. T. McWilliams, D. Meacher, G. D. Meadors, J. Meidam, A. Melatos, G. Mendell, D. Mendoza-Gandara, R. A. Mercer, E. Merilh, M. Merzougui, S. Meshkov, C. Messenger, C. Messick, P. M. Meyers, F. Mezzani, H. Miao, C. Michel, H. Middleton, E. E. Mikhailov, L. Milano, J. Miller, M. Millhouse, Y. Minenkov, J. Ming, S. Mirshekari, C. Mishra, S. Mitra, V. P. Mitrofanov, G. Mitselmakher, R. Mittleman, A. Moggi, M. Mohan, S. R. P. Mohapatra, M. Montani, B. C. Moore, C. J. Moore, D. Moraru, G. Moreno, S. R. Morriss, K. Mossavi, B. Mours, C. M. Mow-Lowry, C. L. Mueller, G. Mueller, A. W. Muir, Arunava Mukherjee, D. Mukherjee, S. Mukherjee, N. Mukund, A. Mullavey, J. Munch, D. J. Murphy, P. G. Murray, A. Mytidis, I. Nardecchia, L. Naticchioni, R. K. Nayak, V. Necula, K. Nedkova, G. Nelemans, M. Neri, A. Neunzert, G. Newton, T. T. Nguyen, A. B. Nielsen, S. Nissanke, A. Nitz, F. Nocera, D. Nolting, M. E. N. Normandin, L. K. Nuttall, J. Oberling, E. Ochsner, J. O'Dell, E. Oelker, G. H. Ogin, J. J. Oh, S. H. Oh, F. Ohme, M. Oliver, P. Oppermann, Richard J. Oram, B. O'Reilly, R. O'Shaughnessy, C. D. Ott, D. J. Ottaway, R. S. Ottens, H. Overmier, B. J. Owen, A. Pai, S. A. Pai, J. R. Palamos, O. Palashov, C. Palomba, A. Pal-Singh, H. Pan, Y. Pan, C. Pankow, F. Pannarale, B. C. Pant, F. Paoletti, A. Paoli, M. A. Papa, H. R. Paris, W. Parker, D. Pascucci, A. Pasqualetti, R. Passaquieti, D. Passuello, B. Patricelli, Z. Patrick, B. L. Pearlstone, M. Pedraza, R. Pedurand, L. Pekowsky, A. Pele, S. Penn, A. Perreca, H. P. Pfeiffer, M. Phelps, O. Piccinni, M. Pichot, M. Pickenpack, F. Piergiovanni, V. Pierro, G. Pillant, L. Pinard, I. M. Pinto, M. Pitkin, J. H. Poeld, R. Poggiani, P. Popolizio, A. Post, J. Powell, J. Prasad, V. Predoi, S. S. Premachandra, T. Prestegard, L. R. Price, M. Prijatelj, M. Principe, S. Privitera, R. Prix, G. A. Prodi, L. Prokhorov, O. Puncken, M. Punturo, P. Puppo, M. Pürrer, H. Qi, J. Qin, V. Quetschke, E. A. Quintero, R. Quitzow-James, F. J. Raab, D. S. Rabeling, H. Radkins, P. Raffai, S. Raja, M. Rakhmanov, C. R. Ramet, P. Rapagnani, V. Raymond, M. Razzano, V. Re, J. Read, C. M. Reed, T. Regimbau, L. Rei, S. Reid, D. H. Reitze, H. Rew, S. D. Reyes, F. Ricci, K. Riles, N. A. Robertson, R. Robie, F. Robinet, A. Rocchi, L. Rolland, J. G. Rollins, V. J. Roma, J. D. Romano, R. Romano, G. Romanov, J. H. Romie, D. Rosińska, S. Rowan, A. Rüdiger, P. Ruggi, K. Ryan, S. Sachdev, T. Sadecki, L. Sadeghian, L. Salconi, M. Saleem, F. Salemi, A. Samajdar, L. Sammut, L. M. Sampson, E. J. Sanchez, V. Sandberg, B. Sandeen, G. H. Sanders, J. R. Sanders, B. Sassolas, B. S. Sathyaprakash, P. R. Saulson, O. Sauter, R. L. Savage, A. Sawadsky, P. Schale, R. Schilling, J. Schmidt, P. Schmidt, R. Schnabel, R. M. S. Schofield, A. Schönbeck, E. Schreiber, D. Schuette, B. F. Schutz, J. Scott, S. M. Scott, D. Sellers, A. S. Sengupta, D. Sentenac, V. Sequino, A. Sergeev, G. Serna, Y. Setyawati, A. Sevigny, D. A. Shaddock, T. Shaffer, S. Shah, M. S. Shahriar, M. Shaltev, Z. Shao, B. Shapiro, P. Shawhan, A. Sheperd, D. H. Shoemaker, D. M. Shoemaker, K. Siellez, X. Siemens, D. Sigg, A. D. Silva, D. Simakov, A. Singer, L. P. Singer, A. Singh, R. Singh, A. Singhal, A. M. Sintes, B. J. J. Slagmolen, J. R. Smith, M. R. Smith, N. D. Smith, R. J. E. Smith, E. J. Son, B. Sorazu, F. Sorrentino, T. Souradeep, A. K. Srivastava, A. Staley, M. Steinke, J. Steinlechner, S. Steinlechner, D. Steinmeyer, B. C. Stephens, S. P. Stevenson, R. Stone, K. A. Strain, N. Straniero, G. Stratta, N. A. Strauss, S. Strigin, R. Sturani, A. L. Stuver, T. Z. Summerscales, L. Sun, P. J. Sutton, B. L. Swinkels, M. J. Szczepańczyk, M. Tacca, D. Talukder, D. B. Tanner, M. Tápai, S. P. Tarabrin, A. Taracchini, R. Taylor, T. Theeg, M. P. Thirugnanasambandam, E. G. Thomas, M. Thomas, P. Thomas, K. A. Thorne, K. S. Thorne, E. Thrane, S. Tiwari, V. Tiwari, K. V. Tokmakov, C. Tomlinson, M. Tonelli, C. V. Torres, C. I. Torrie, D. Töyrä, F. Travasso, G. Traylor, D. Trifirò, M. C. Tringali, L. Trozzo, M. Tse, M. Turconi, D. Tuyenbayev, D. Ugolini, C. S. Unnikrishnan, A. L. Urban, S. A. Usman, H. Vahlbruch, G. Vajente, G. Valdes, M. Vallisneri, N. van Bakel, M. van Beuzekom, J. F. J. van den Brand, C. Van Den Broeck, D. C. Vander-Hyde, L. van der Schaaf, J. V. van Heijningen, A. A. van Veggel, M. Vardaro, S. Vass, M. Vasúth, R. Vaulin, A. Vecchio, G. Vedovato, J. Veitch, P. J. Veitch, K. Venkateswara, D. Verkindt, F. Vetrano, A. Viceré, S. Vinciguerra, D. J. Vine, J. Y. Vinet, S. Vitale, T. Vo, H. Vocca, C. Vorvick, D. Voss, W. D. Vousden, S. P. Vyatchanin, A. R. Wade, L. E. Wade, M. Wade, S. J. Waldman, M. Walker, L. Wallace, S. Walsh, G. Wang, H. Wang, M. Wang, X. Wang, Y. Wang, H. Ward, R. L. Ward, J. Warner, M. Was, B. Weaver, L. W. Wei, M. Weinert, A. J. Weinstein, R. Weiss, T. Welborn, L. Wen, P. Weßels, T. Westphal, K. Wette, J. T. Whelan, S. E. Whitcomb, D. J. White, B. F. Whiting, K. Wiesner, C. Wilkinson, P. A. Willems, L. Williams, R. D. Williams, A. R. Williamson, J. L. Willis, B. Willke, M. H. Wimmer, L. Winkelmann, W. Winkler, C. C. Wipf, A. G. Wiseman, H. Wittel, G. Woan, J. Worden, J. L. Wright, G. Wu, J. Yablon, I. Yakushin, W. Yam, H. Yamamoto, C. C. Yancey, M. J. Yap, H. Yu, M. Yvert, A. ZadroŻny, L. Zangrando,

M. Zanolin, J. P. Zendri, M. Zevin, F. Zhang, L. Zhang, M. Zhang, Y. Zhang, C. Zhao, M. Zhou, Z. Zhou, X. J. Zhu, M. E. Zucker, S. E. Zuraw, J. Zweizig, LIGO Scientific Collaboration, and Virgo Collaboration. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physics Review Letter*, 116(6):061102, February 2016.

[Accomazzi, 2024] Alberto Accomazzi. Decades of Transformation: Evolution of the NASA Astrophysics Data System's Infrastructure. *arXiv e-prints*, page arXiv:2401.09685, January 2024.

[Bai, 2016] Xue-Ning Bai. Towards a Global Evolutionary Model of Protoplanetary Disks. *The Astrophysical Journal*, 821(2):80, April 2016.

[Bannister *et al.*, 2012] Michael J. Bannister, David Eppstein, Michael T. Goodrich, and Lowell Trott. Force-Directed Graph Drawing Using Social Gravity and Scaling. *arXiv e-prints*, page arXiv:1209.0748, September 2012.

[Baron, 2019] Dalya Baron. Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, page arXiv:1904.07248, April 2019.

[Bertschinger, 1998] Edmund Bertschinger. Simulations of Structure Formation in the Universe. *Annual Review of Astronomy and Astrophysics*, 36:599–654, January 1998.

[Blecher *et al.*, 2023] Lukas Blecher, Guillem Cucurull, Thomas Scialom, and Robert Stojnic. Nougat: Neural Optical Understanding for Academic Documents. *arXiv e-prints*, page arXiv:2308.13418, August 2023.

[Brown *et al.*, 2020] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *arXiv e-prints*, page arXiv:2005.14165, May 2020.

[Bubeck *et al.*, 2023] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv e-prints*, page arXiv:2303.12712, March 2023.

[Carlesi *et al.*, 2014a] Edoardo Carlesi, Alexander Knebe, Geraint F. Lewis, Scott Wales, and Gustavo Yepes. Hydrodynamical simulations of coupled and uncoupled quintessence models - I. Halo properties and the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 439(3):2943–2957, April 2014.

[Carlesi *et al.*, 2014b] Edoardo Carlesi, Alexander Knebe, Geraint F. Lewis, and Gustavo Yepes. Hydrodynamical simulations of coupled and uncoupled quintessence models - II. Galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 439(3):2958–2969, April 2014.

[Cheong *et al.*, 2021] Se-Hang Cheong, Yain-Whar Si, and Raymond K. Wong. Online force-directed algorithms for visualization of dynamic graphs. *Information Sciences*, 556:223–255, 2021.

[Cranmer *et al.*, 2020] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, December 2020.

[De Santi *et al.*, 2024] Federico De Santi, Massimiliano Razzano, Francesco Fidecaro, Luca Muccillo, Lucia Papalini, and Barbara Patricelli. Deep learning to detect gravitational waves from binary close encounters: Fast parameter estimation using normalizing flows. *Physical Review D*, 109(10):102004, May 2024.

[Diaz Rivero and Dvorkin, 2020] Ana Diaz Rivero and Cora Dvorkin. Flow-based likelihoods for non-Gaussian inference. *Physical Review D*, 102(10):103507, November 2020.

[Dung Nguyen *et al.*, 2023] Tuan Dung Nguyen, Yuan-Sen Ting, Ioana Ciucă, Charlie O'Neill, Ze-Chang Sun, Maja Jabłońska, Sandor Kruk, Ernest Perkowski, Jack Miller, Jason Li, Josh Peek, Kartheik Iyer, Tomasz Różański, Pranav Khetarpal, Sharaf Zaman, David Brodrick, Sergio J. Rodríguez Méndez, Thang Bui, Alyssa Goodman, Alberto Accomazzi, Jill Naiman, Jesse Cranney, Kevin Schawinski, and UniverseTBD. AstroLLaMA: Towards Specialized Foundation Models in Astronomy. *arXiv e-prints*, page arXiv:2309.06126, September 2023.

[Genel *et al.*, 2014] Shy Genel, Mark Vogelsberger, Volker Springel, Debora Sijacki, Dylan Nelson, Greg Snyder, Vicente Rodriguez-Gomez, Paul Torrey, and Lars Hernquist. Introducing the Illustris project: the evolution of galaxy populations across cosmic time. *Monthly Notices of the Royal Astronomical Society*, 445(1):175–200, November 2014.

[Gudiksen *et al.*, 2011] B. V. Gudiksen, M. Carlsson, V. H. Hansteen, W. Hayek, J. Leenaarts, and J. Martínez-Sykora. The stellar atmosphere simulation code Bifrost. Code description and validation. *Astronomy & Astrophysics*, 531:A154, July 2011.

[Günther *et al.*, 2023] Michael Günther, Jackmin Ong, Isabelle Mohr, Alaeddine Abdessalem, Tanguy Abel, Mohammad Kalim Akram, Susana Guzman, Georgios Mastrapas, Saba Sturua, Bo Wang, Maximilian Werk, Nan Wang, and Han Xiao. Jina Embeddings 2: 8192-Token General-Purpose Text Embeddings for Long Documents. *arXiv e-prints*, page arXiv:2310.19923, October 2023.

[Hara and Ford, 2023] Nathan C. Hara and Eric B. Ford. Statistical Methods for Exoplanet Detection with Radial Velocities. *Annual Review of Statistics and Its Application*, 10(1):623–649, March 2023.

[Hu and Dodelson, 2002] Wayne Hu and Scott Dodelson. Cosmic Microwave Background Anisotropies. *Annual Re-*

view of Astronomy and Astrophysics, 40:171–216, January 2002.

[Jiang *et al.*, 2014] Yan-Fei Jiang, James M. Stone, and Shane W. Davis. A Global Three-dimensional Radiation Magneto-hydrodynamic Simulation of Super-Eddington Accretion Disks. *The Astrophysical Journal*, 796(2):106, December 2014.

[Jiang *et al.*, 2023] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7B. *arXiv e-prints*, page arXiv:2310.06825, October 2023.

[Jumper *et al.*, 2021] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[Kobourov, 2012] Stephen G. Kobourov. Spring Embedders and Force Directed Graph Drawing Algorithms. *arXiv e-prints*, page arXiv:1201.3011, January 2012.

[McAlpine *et al.*, 2016] S. McAlpine, J. C. Helly, M. Schaller, J. W. Trayford, Y. Qu, M. Furlong, R. G. Bower, R. A. Crain, J. Schaye, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, A. Jenkins, Y. Rosas-Guevara, S. D. M. White, M. Baes, P. Camps, and G. Lemson. The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues. *Astronomy and Computing*, 15:72–89, April 2016.

[Meijer *et al.*, 2021] H. J. Meijer, J. Truong, and R. Karimi. Document Embedding for Scientific Articles: Efficacy of Word Embeddings vs TFIDF. *arXiv e-prints*, page arXiv:2107.05151, July 2021.

[Nyheim *et al.*, 2024] Bendik Nyheim, Signe Riemer-Sørensen, Rodrigo Parra, and Claudia Cicone. Machine Learning based Pointing Models for Radio/Sub-millimeter Telescopes. *arXiv e-prints*, page arXiv:2402.08589, February 2024.

[OpenAI *et al.*, 2023] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie

Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. GPT-4 Technical Report. *arXiv e-prints*, page arXiv:2303.08774, March 2023.

[Pan *et al.*, 2023] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *arXiv e-prints*, page arXiv:2306.08302, June 2023.

[Paz *et al.*, 2006] D. J. Paz, D. G. Lambas, N. Padilla, and M. Merchán. Shapes of clusters and groups of galaxies: comparison of model predictions with observations. *Monthly Notices of the Royal Astronomical Society*, 366(4):1503–1510, March 2006.

[Peñarrubia *et al.*, 2006] Jorge Peñarrubia, Andrew J. Benson, David Martínez-Delgado, and Hans Walter Rix. Modeling Tidal Streams in Evolving Dark Matter Halos. *The Astrophysical Journal*, 645(1):240–255, July 2006.

[Perkowski *et al.*, 2024] Ernest Perkowski, Rui Pan, Tuan Dung Nguyen, Yuan-Sen Ting, Sandor Kruk, Tong Zhang, Charlie O'Neill, Maja Jablonska, Zechang Sun, Michael J. Smith, Huiling Liu, Kevin Schawinski, Kartheik Iyer, Ioana Ciucă, and UniverseTBD. AstroLLaMA-Chat: Scaling AstroLLaMA with Conversational and Diverse Datasets. *Research Notes of the American Astronomical Society*, 8(1):7, January 2024.

[Pillepich *et al.*, 2018] Annalisa Pillepich, Volker Springel, Dylan Nelson, Shy Genel, Jill Naiman, Rüdiger Pakmor, Lars Hernquist, Paul Torrey, Mark Vogelsberger, Rainer Weinberger, and Federico Marinacci. Simulating galaxy formation with the IllustrisTNG model. *Monthly Notices of the Royal Astronomical Society*, 473(3):4077–4106, January 2018.

[Robertson *et al.*, 2023] Brant E. Robertson, Sandro Tacchella, Benjamin D. Johnson, Ryan Hausen, Adebusola B. Alabi, Kristan Boyett, Andrew J. Bunker, Stefano Carniani, Eiichi Egami, Daniel J. Eisenstein, Kevin N. Hainline, Jakob M. Helton, Zhiyuan Ji, Nimisha Kumari, Jianwei Lyu, Roberto Maiolino, Erica J. Nelson, Marcia J. Rieke, Irene Shivaei, Fengwu Sun, Hannah Übler, Christina C. Williams, Christopher N. A. Willmer, and Joris Witstok. Morpheus Reveals Distant Disk Galaxy Morphologies with JWST: The First AI/ML Analysis of JWST Images. *The Astrophysical Journal Letters*, 942(2):L42, January 2023.

[Romeo *et al.*, 2004] Alessandro B. Romeo, Cathy Horellou, and Jöran Bergh. A wavelet add-on code for new-generation N-body simulations and data de-noising (JOFILUREN). *Monthly Notices of the Royal Astronomical Society*, 354(4):1208–1222, November 2004.

[Springel, 2005] Volker Springel. The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, December 2005.

[Storrie-Lombardi *et al.*, 1992] M. C. Storrie-Lombardi, O. Lahav, Jr. Sodre, L., and L. J. Storrie-Lombardi. Morphological Classification of Galaxies by Artificial Neural Networks. *Monthly Notices of the Royal Astronomical Society*, 259:8P, November 1992.

[Sun *et al.*, 2023a] Zechang Sun, Joshua S. Speagle, Song Huang, Yuan-Sen Ting, and Zheng Cai. Zephyr : Stitching Heterogeneous Training Data with Normalizing Flows for Photometric Redshift Inference. *arXiv e-prints*, page arXiv:2310.20125, October 2023.

[Sun *et al.*, 2023b] Zechang Sun, Yuan-Sen Ting, and Zheng Cai. Quasar Factor Analysis-An Unsupervised and Probabilistic Quasar Continuum Prediction Algorithm with Latent Factor Analysis. *The Astrophysical Journal Supplement Series*, 269(1):4, November 2023.

[Ting *et al.*, 2019] Yuan-Sen Ting, Charlie Conroy, Hans-Walter Rix, and Phillip Cargile. The Payne: Self-consistent ab initio Fitting of Stellar Spectra. *The Astrophysical Journal*, 879(2):69, July 2019.

[Villaescusa-Navarro *et al.*, 2021a] Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N. Spergel, Yin Li, Benjamin Wandelt, Andrina Nicola, Leander Thiele, Sultan Hassan, Jose Manuel Zorrilla Matilla, Desika Narayanan, Romeel Dave, and Mark Vogelsberger. Multifield Cosmology with Artificial Intelligence. *arXiv e-prints*, page arXiv:2109.09747, September 2021.

[Villaescusa-Navarro *et al.*, 2021b] Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N. Spergel, Rachel S. Somerville, Romeel Dave, Annalisa Pillepich, Lars Hernquist, Dylan Nelson, Paul Torrey, Desika Narayanan, Yin Li, Oliver Philcox, Valentina La Torre, Ana Maria Delgado, Shirley Ho, Sultan Hassan, Blakesley Burkhart, Digvijay Wadekar, Nicholas Battaglia, Gabriella Contardo, and Greg L. Bryan. The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations. *The Astrophysical Journal*, 915(1):71, July 2021.

[Weinberg, 2008] Steven Weinberg. *Cosmology*. 2008.

[Xu *et al.*, 2018] J. Xu, Y. Bu, Y. Ding, et al. Understanding the formation of interdisciplinary research from the perspective of keyword evolution: a case study on joint attention. *Scientometrics*, 117:973–995, 2018.

[Zhao *et al.*, 2023] Xiaosheng Zhao, Yuan-Sen Ting, Kangning Diao, and Yi Mao. Can diffusion model conditionally generate astrophysical images? *Monthly Notices of the Royal Astronomical Society*, 526(2):1699–1712, December 2023.

[Zhou and Lin, 2007] Ji-Lin Zhou and Douglas N. C. Lin. Planetesimal Accretion onto Growing Proto-Gas Giant Planets. *The Astrophysical Journal*, 666(1):447–465, September 2007.

[Zhu *et al.*, 2023] Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities. *arXiv e-prints*, page arXiv:2305.13168, May 2023.