Exploring Advanced Hypothesis Generation in Astronomy Through the Implementation of a Mathematical Model of Linguistic Neural Networks

Ievgen Zaitsev [1], Oleksandr Golubenko [2], Olha Tkachenko [3], Oleksandr Pidmohylnyi [4] and Artem Antonenko [5]

[1] *Institute of Electrodynamics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine*
[2] *Academician Yury Bugai International Scientific and Technical University, Kyiv, Ukraine*
[3] *Taras Shevchenko Kyiv National University, Kyiv, Ukraine*
[4] *State University of Information and Communication Technologies, Kyiv, Ukraine*
[5] *National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine*[5]

Abstract

The current research conducts an exhaustive analysis into the integration of Large Language Models (LLMs), focusing on GPT-4, within the domain of Astronomy. We deploy advanced in-context and adversarial prompting methods, harnessing a database of over 1000 articles from NASA's Astrophysics Data System, aiming to augment the model's performance in domain-specific literature comprehension. This comprehensive inquiry demonstrates a marked enhancement in the model's capability for generating robust and insightful hypotheses in astronomy. Furthermore, the study highlights the role of adversarial prompting in enabling GPT-4 to sift through extensive data to yield meaningful and informed scientific hypotheses, showcasing a pioneering advancement for the implementation of LLMs in Astronomy research.

Keywords[1]

Natural Language Processing (NLP), transformer architecture, large Language Models (LLMs), GPT-4, astronomy, underlying physical processes, hypothesis generation, NASA's Astrophysics Data System, in-context and adversarial prompting, comprehensive corpus, comprehensive corpus training, astronomical knowledge, exploration in astronomy, astro-GPT

1. Introduction

The field of Natural Language Processing (NLP) has seen monumental progress propelled by advanced attention mechanisms and innovative transformer architecture. This technological evolution has culminated in the inception of Large Language Models (LLMs) like GPT-4, which stand as paragons of computational language understanding. These models unfurl a vast expanse of capabilities in not only understanding and generating human language but also in performing intricate interactions and operations with it. They transcend the boundaries of simple language tasks, venturing into the realm of making profound deductions and establishing intricate relationships across a panorama of diverse contexts.

The expansive potentials of these models have been underscored by two particular aspects that have fanned the flames of excitement in the research community. The first aspect delves into the LLMs' ability to sample the posterior means of languages. Even though this ability occasionally stumbles into the issues

Dynamical System Modeling and Stability Investigation (DSMSI-2023), December 19-21, 2023, Kyiv, Ukraine
EMAIL: zaitsev@i.ua (A. 1); golubenkooleksandr82@gmail.com (A. 2); okar@ukr.net (A. 3); alexandr.cctv@gmail.com (A. 4); artem.v.antonenko@gmail.com (A. 5)
ORCID: 0000-0003-3303-471X (A. 1); 0000-0002-1776-5160 (A. 2); 0000-0001-7983-9033 (A. 3); 0000-0001-8689-2086 (A. 4); 0000-0001-9397-1209 (A.5)

of non-trivial hallucination, a significant leap in their performance has been observed through the employment of in-context prompting [1-3]. This strategic enhancement bolsters the models' capacity to grapple with complex, domain-specific tasks, displaying an exemplary augmentation in their functionality [4].

The second facet shines light on the harmonious integration of these models with cutting-edge technologies like Langchain. This integration saturates the LLMs with extensive context, amplifying their operational bandwidth across a myriad of fields. Despite the availability of other methods like the use of adapters [5], which can significantly enhance performance for domain-specific tasks by fine-tuning the LLMs, the challenges they pose, especially for institutions constrained by limited resources, cannot be overlooked.

Venturing into the astronomical domain, this research explores the application of cost-effective in-context prompting [6]. The field of Astronomy emerges as an exceptional candidate for this study for three pivotal reasons. Firstly, despite its rich literary background, the incorporation of astronomical texts in the comprehensive corpus utilized for training GPT models is presumably limited. This limitation precipitates apparent hallucination problems when deploying naive versions of LLMs [7]. Secondly, the advancement in astronomy is often a product of synthesizing information across various subfields, a feature accentuated by the universality of underlying physical processes at diverse scales. This characteristic bolsters the assumption that extensive in-context prompting can substantially amplify hypothesis generation, especially when LLMs are preliminarily exposed to a diverse literature spectrum.

Lastly, the enduring “open sky” policy of astronomy further solidifies its position as a promising field for in-context prompting research. The policy's emphasis on public availability of most datasets, either immediately or after a brief proprietary period [8], in conjunction with a comprehensive, meticulously curated literature database, facilitates seamless engagement with the vast reservoir of astronomical knowledge. With the NASA’s Astrophysics Data System hosting a staggering excess of 15 million resources, the entirety of the astronomical literature spectrum accessible to researchers is effectively covered [9]. This extensive accessibility plays a crucial role in enhancing our interaction with the astronomy database, offering a fertile ground for further research and exploration.

In conclusion, the application of LLMs, especially with the integration of in-context and adversarial prompting, holds the promise of a revolution in various domains, including astronomy. The outlined advancements and integrations not only mitigate the challenges faced by these models but also amplify their capabilities, offering a beacon of hope for future research and applications in diverse fields.

The procedure begins with the pre-processing and embedding of Galactic Astronomy papers. A similarity search is conducted on the embedded query, and relevant document chunks are retrieved. A further contextual compression is performed to remove irrelevant information from the chunks. These compressed texts serve as input to a GPT-4 instance, which generates an idea. This idea is then critiqued by a second GPT-4 model, and the feedback is moderated by a third GPT-4 model (Fig.1).

2. Retrieval and Pre-processing of Literature

In this research, our scrutiny was centered on Galactic Astronomy. Our domain expertise was leveraged to evaluate the outcomes, selecting Galactic Astronomy for its holistic character that amalgamates knowledge from various sub-disciplines. Exploring galaxy evolution encompasses not only a foundational comprehension of stars and stellar groups, but also considers the impact of extensive cosmological environmental elements [10,11]. This multifaceted approach to understanding galaxy evolution presents both thrilling challenges and plentiful opportunities for harnessing the tacit knowledge ensconced within the expansive literary network.

For our research, a collection of 2,000 pertinent Galactic Astronomy papers were assembled from the NASA Astrophysics Data System (ADS) is a digital library portal for researchers in astronomy and physics. It is operated by the Smithsonian Astrophysical Observatory (SAO) under a NASA grant. The system is a valuable resource for accessing a wide range of publications related to astrophysics, including research articles, conference proceedings, and other relevant literature.

The ADS maintains three bibliographic databases containing more than 14 million records:

- Astronomy and Astrophysics
- Physics and Geophysics
- arXiv e-prints.

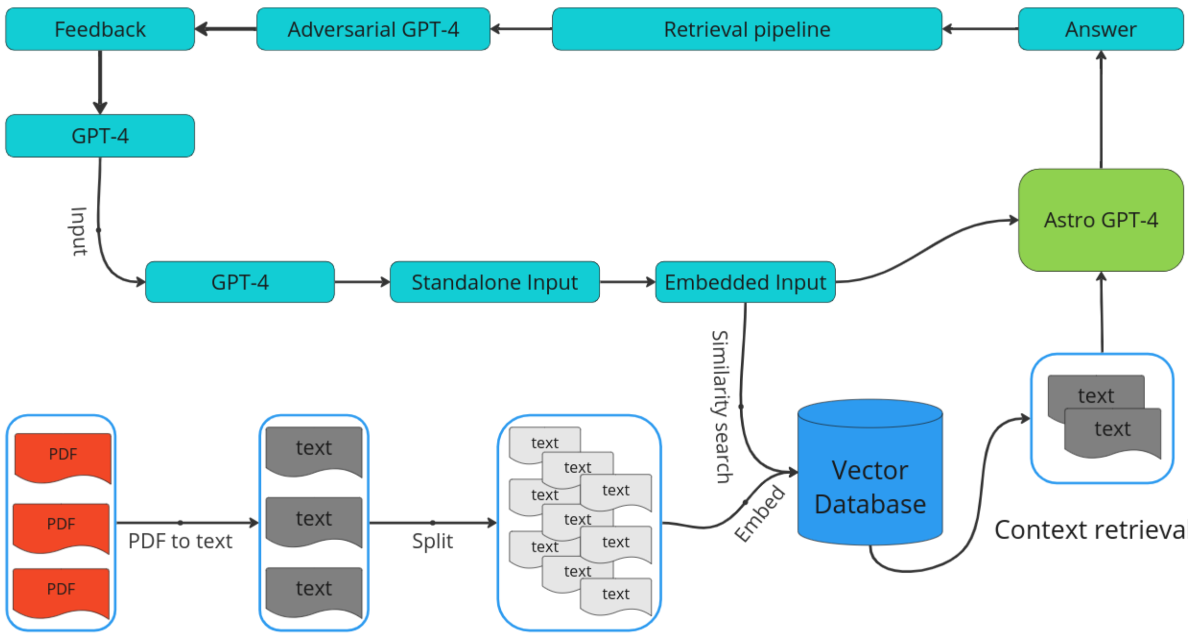

Figure 1. This figure illustrates the adversarial in-context prompting workflow using OpenAI's GPT-4 model.

The selection process involved a database query founded on specific criteria like the mention of 'Gaia' in the abstract, publications from the last decade (aligning with the Gaia launch date), peer-reviewed journal articles, and the incorporation of related keywords such as 'galaxy kinematics and dynamics', 'galaxy structure', 'galaxy disk', 'galaxy halo', 'galaxy abundances', and 'galaxy evolution'. The initial inquiry generated over 2,000 papers, prompting us to focus on the most recent contributions.

Our meticulously compiled collection encompasses various details like the ArxivID, Publication Date, Authors, Title, Abstract, Citation, and Key, forming a thorough dataset for our examination. For ensuring reproducibility, both the complete dataset and the utilized codebase in our analysis are made available here.

Astro-GPT Flow.

Our investigation harnesses the capabilities of the GPT-4 model by OpenAI. The initial step in in-context prompting involves the pre-processing of 2,000 documents from the Galactic Astronomy corpus utilizing the Langchain library. Each paper, converted from PDF to text format, is then divided into segments or 'chunks' each containing 2,000 tokens. These segments are further embedded using the text-ada-002 embedding model by OpenAI.

During the retrieval phase, chat history and input query are transformed into a singular input and then embedded. A search for similarity is executed between the embedded query and the vector database. Utilizing Langchain's contextual compression, irrelevant information is filtered out from the individual segments. The final texts, combined with the single input, lay the groundwork for the GPT-4 model, operating with a context window of around 8,000 tokens, to develop ideas.

To evaluate the model's competence, an adversarial experiment is constructed. This includes a secondary GPT-4 model tasked with analyzing the idea, pinpointing its weaknesses, and proposing possible improvements. This feedback is restructured within a feedback-question format by a third GPT-4 instance and relayed to the initial model.

In executing our experimental arrangement, we utilize Nk papers, where k is a member of the set {1, 10, 100, 1000}. Each sample is subjected to hypothesis generation by the 'Generation GPT-4' instance (our in-context prompted model on k papers). This is followed by an adversarial response from 'Adversarial GPT-4,' which is then reshaped by a moderator GPT-4 instance and returned to the generator model. This loop, producing three hypotheses and two critiques per trial, is repeated twice for each Nk and replicated five times overall. The same methodology is employed for 1,000 papers, without resampling, amassing a total of 60 hypotheses and 40 critiques.

3. Results

Human Evaluation.

Given the inherently qualitative aspect of hypothesis generation, it was crucial to adopt an assessment process that, while subjective, would align with the anticipations of human experts. Recognizing this need, we engaged two domain experts specializing in Galactic Astronomy to scrutinize the quality of the hypotheses generated. Each hypothesis was critically evaluated and rated based on the volume of incorporated papers within the domain-specific context. This method allowed us to compute an average score from these dual human evaluations for each formulated hypothesis.

The grading rubric was divided into three distinct categories – scientific accuracy, creativity, and feasibility. The hypotheses were meticulously assessed in these domains, and the average score across these three areas was considered the final score. Additionally, we conducted a comprehensive evaluation of the critiques provided by the AI judge, ensuring it had access to identical contextual information for a fair and balanced assessment. The graphical representation in the left panel of underscores the pivotal role of adversarial prompting in enhancing hypothesis generation substantially. In the absence of adversarial prompting, the quality of hypothesis generation exhibited negligible dependence on the number of consulted papers. This observation underscored that in-context prompting alone, albeit beneficial for reducing hallucination, fell short of providing an exhaustive comprehension of the corpus.

Nevertheless, the incorporation of adversarial prompting dramatically transformed this scenario. We observed a substantial escalation in the quality of hypothesis generation for both the AI generator and the AI judge. This improvement prevailed even without the need for explicitly aligning the models with human expectations. Most notably, the introduction of adversarial prompting established a robust correlation between the quality of the hypotheses and the quantity of reviewed papers, particularly at a larger context (N = 1000). It also fostered greater consistency in terms of the quality of both the hypotheses and the critiques.

The mean quality score witnessed a considerable augmentation, escalating from 2.5 (with a context of 10 papers, where a score of 3/5 corresponds to an average hypothesis by a proficient PhD student) to an almost expert level of 4/5 with the inclusion of 1,000 papers in the context. This marked enhancement accentuates the immense potential of employing adversarial prompting to bolster the quality of scientific hypothesis generation. Detailed examples underscoring these observations are elaborated in the Appendix.

Exploration of Embeddings.

In an endeavor to grasp the profound impact of adversarial prompting, we initiated the process by passing the abstracts of our set of 2,000 astronomy papers through the text-ada-002 embedding model. This was followed by their organization into a 2D TSNE projection, adeptly capturing the contextual disparities and parallels amongst the 2,000 papers.

For every hypothesis generated, an in-depth determination was conducted to ascertain the specific papers that contributed to its inception by querying the GPT model. Visually presents this 'knowledge footprint' for each hypothesis, depicted as black polygons within a green hull that signifies all the papers GPT-4 had access to. A discernible observation upper panels reveals that with the escalation in the number of papers, GPT notably expands its scope. It begins to assimilate diverse topics to construct interconnected hypotheses. It’s pertinent to highlight that even with a modest training pool of 10 papers, our experiment is methodically designed to encompass a broad spectrum of topics. However, GPT-4's lack of adequate context leads to the formulation of more generalized hypotheses (refer to Appendix for further insight).

Focusing on the scenario with 1,000 papers, the bottom panel in provides an illuminating exploration of the evolution of the knowledge footprint with varying numbers of adversarial attacks. In the initial iteration (bottom left), the judge adeptly pinpoints areas for critique, grounded on the knowledge that the original response inadvertently overlooked. This proactively propels the generator to widen its scope further, as visibly illustrated in the bottom middle and right panels. This expansion aids in appropriately addressing the surfaced criticisms.

As showcased in some instances in the Appendix, the implementation of adversarial prompting empowers the GPT-4 model to genuinely leverage a voluminous number of contextual inputs. This strategic guidance navigates the model towards a more integrated and comprehensive comprehension of the topic, rather than resulting in a mere aggregation of disjointed facts. The Appendix further elucidates specific examples of generated hypotheses and their corresponding critiques, offering more profound insights into this exploration.

Exemplary Hypotheses and Critiques Generated.

This portion showcases three kinds of hypotheses, created via our in-context and adversarial prompting methods, coupled with the related critiques offered by our adversarial GPT-4 model. This feedback resulted in notable enhancements in hypothesis formation. The initial example illustrates a hypothesis formed post subjecting the model to in-context prompting through a review of 1000 papers, succeeded by two sequences of adversarial critiques [12].

I suggest undertaking a study to explore the influence of dark matter mini-halos on the formation and progression of globular clusters (GCs) within the Milky Way. This research would merge the detailed astrometric information from Gaia EDR3 and DR3 with spectral data acquired from major terrestrial surveys such as APOGEE, GALAH, SDSS SEGUE, and LAMOST. This combination aims to analyze the movement of stars in the outer regions of GCs. The primary objective is to ascertain if these GCs are situated within dark matter mini-halos, potentially shedding light on their inception.

Building on the probabilistic method pioneered by Kuzma et al. (2021) - which evaluates the outer areas of GCs by applying a mixture model in spatial and proper motion space to delineate cluster, extra-tidal, and contaminating stellar groups - this research would incorporate the impact of dark matter mini-halos on star kinematics in the outer regions of GCs. By broadening this methodology, we can evaluate the proposition that dark matter is instrumental in shaping and developing Gcs.

Additionally, this analysis would offer a deeper comprehension of the dispersion and attributes of dark matter within the Milky Way, enriching the wider domain of proximate cosmology. Fusing this data with the existing understanding of the Milky Way's hierarchical structure and the significance of swift gas accretion in the inner Galactic disc would afford a holistic perspective on the genesis narrative of our Galaxy and the inherent distribution of dark matter.

This hypothesis garnered significant endorsement from the human panel, attributed to its extensive contextual comprehension of Galactic Astronomy, a solid thesis proposal, suitable references for conceptual foundations, and its expansion upon established methods and research. On the contrary, the subsequent hypothesis was generated when in-context prompting was confined to 10 papers. Hypotheses formed with such limited literary exposure frequently seem generic, devoid of meaningful insights and offering only trivial enhancements to existing work [13].

I propose conducting research to examine the impact of intermediate-mass gas-rich mergers on the chemical and kinematic development of galaxies similar to the Milky Way, spanning various redshifts. This investigation would expand on the discoveries made by Ciuca et al. (2023), which indicate a significant influence of the Gaia-Sausage-Enceladus (GSE) merger on the creation of the Galactic disk and the shift from the thick disk to the thin disk.

The study could incorporate data from extensive spectroscopic surveys like APOGEE, GALAH, and Gaia-ESO, alongside high-resolution cosmological simulations from projects such as the Auriga project. Central goals would encompass the characterization of chemical abundance patterns and kinematic traits of stars in the disk and halo sections during and post-mergers. It would also explore the effect of these mergers on the radial distribution and movement of stars throughout the Galactic disk.

By centering the research on intermediate-mass gas-rich mergers, it will bridge the existing knowledge void regarding the significance of these events in sculpting the form and progression of Milky Way-like galaxies. The results would further offer essential insights for the enhancement of galaxy formation models, contributing to a more exhaustive comprehension of our Galaxy's intricate history.

Without sufficient context, the model tends to extrapolate based on existing literature, resulting in incremental, less innovative suggestions. Importantly, despite the exposure to a corpus of 1000 papers, without adversarial attacks, the hypotheses can often seem like a jumble of facts, as shown in the example below:

I suggest a thorough, multi-wavelength investigation into the unexplored areas of the Galactic plane, especially focusing on the South Galactic cap, encompassing regions below a declination of -65°. The endeavor to chart these untouched regions can be realized by integrating existing extensive photometric, spectroscopic, and astrometric surveys like Gaia, LAMOST, APOGEE, and GALAH, along with new high-resolution surveys targeting the mapping of these unknown zones of the Milky Way.

Emphasis should be placed on identifying and characterizing previously undiscovered ultra-faint dwarf galaxies, chemically unusual stars, and stellar streams, coupled with assessing their origins and potential impact on our comprehension of the Milky Way's formation and evolution. This research initiative takes inspiration from the unexplored territories highlighted in Antoja et al. (2015), the suggested enhancements by Gorski and Barmby (2020), and the advancements in large-scale surveys explored in studies like Yuan et al. (2014) and An and Beers (2020).

Broadening our insight into these uncharted regions of the Galactic plane will greatly augment our understanding of the Milky Way's structure, formation, and evolution. Moreover, it will empower astronomers to uncover and evaluate previously unidentified star populations and their kinematic characteristics, further honing our understanding of the diverse elements and phenomena that mold our Galaxy.

Demonstrating adversarial prompting is critical to ensure that the GPT can harness the broad context provided in a meaningful way. Finally, an example of high-quality critique provided by the adversarial GPT-4 model is showcased. These adversarial critiques, after being primed with 1000 papers, are highly accurate and can rival the quality of expert human critique [14].

The proposed interdisciplinary method aspires to offer a thorough insight into the Milky Way's gravitational potential by merging photometric data, high-definition simulations, and innovative approaches for phase-space spirals analysis. Despite its promise, several weaknesses and limitations need attention.

Firstly, the heavy reliance on photometric data, while less susceptible to sampling biases, may encounter uncertainties from dust obscuration and crowding. The precision of photometric metallicity measurements could also fall short when compared to spectroscopic data.

Secondly, the usage of high-definition simulations for superior modeling of the Milky Way's potential comes with its drawbacks. These simulations, albeit detailed, demand extensive computational resources. The existing mass and spatial resolutions might not meet the requirements for a meticulous analysis. Additionally, the Milky Way's complexity, marked by its non-steady-state nature and the presence of various structures like spiral arms and bars, introduces additional hurdles to the modeling process.

Thirdly, employing novel techniques for analyzing phase-space spirals could be influenced by various systematic biases. The proposal somewhat overlooks the method of addressing these biases and the possible constraints of these techniques amidst the context of a multifaceted and evolving galaxy.

Lastly, the proposal inadequately covers the implications of different selection biases in spectroscopic surveys and the complications in incorporating the galaxy's complexity into extensive cosmological models. In summary, while the interdisciplinary strategy holds potential, substantial additional work is crucial to overcome these limitations and weaknesses for a more precise and thorough comprehension of the Milky Way's gravitational potential.

4. Conclusion and Future Direction

Our investigation goes beyond just the theoretical framework or limited-scope testing. It is grounded in comprehensive, real-world applications, reflecting an intent to resolve actual challenges faced in the field of Galactic Astronomy. A nuanced understanding of the specifics of in-context and adversarial prompting allows us to refine and adapt our methodologies continuously, ensuring that the resulting enhancements are not just incremental but transformative.

An essential part of our future work lies in the significant expansion of the dataset employed for in-context prompting. Utilizing a more extensive and diverse collection of documents from the NASA ADS will further hone the precision and relevance of the generated hypotheses, allowing for a richer, more comprehensive exploration of potential research areas. Moreover, the utilization of advanced analytics tools and techniques will enable more in-depth analysis and evaluation, ensuring that the generated hypotheses stand up to rigorous scrutiny from both AI and human evaluators.

A parallel line of inquiry will be pursued, focusing on the integration of the GPT-4 model with other advanced AI and machine learning technologies. This collaborative approach seeks to harness the collective strengths of multiple AI models and algorithms, enhancing the robustness and reliability of hypothesis generation. This integration will also facilitate a more seamless and efficient evaluation process, with multiple AI models working in tandem to analyze, assess, and refine the generated hypotheses.

Furthermore, the insights gleaned from our extensive analysis will serve as a robust foundation for the development of more advanced, efficient, and reliable in-context and adversarial prompting mechanisms. These improved mechanisms will not only enhance the quality and reliability of hypothesis generation but will also significantly expedite the entire process. This heightened efficiency is crucial, given the rapidly growing volume and complexity of astronomical research, ensuring that researchers can keep pace with the latest developments and insights in the field.

In conclusion, this research signifies a substantial stride towards revolutionizing hypothesis generation in Galactic Astronomy through the adept utilization of advanced AI models like GPT-4. By continually refining and enhancing our methodologies and approaches, we aim to make a meaningful and lasting contribution to the field of astronomical research, driving forward the frontier of knowledge and discovery in this critical scientific domain. Our commitment to excellence, innovation, and continuous improvement will ensure that our research remains at the cutting edge, providing valuable insights, tools, and methodologies for the global astronomical research community.

Broader Impact.

This research primarily centered on in-context prompting, a conscious choice motivated by the objective to broaden the access and application of Large Language Models (LLMs) for scientific exploration. The current versions of GPT models, with their colossal parameter counts, often make the process of fine-tuning a nonviable endeavor. In this era, characterized by the proliferation of LLMs, a vital question arises: Can all academic entities, irrespective of their scale or computational capabilities, stay abreast with these swiftly progressing technologies?

This pressing question stands at the heart of our investigation, with the results shedding a beam of optimism on the scenario. Our research underlines that employing the right strategies and involving human experts in the loop can substantially diminish the obstacles in unleashing the full potential of these sophisticated LLMs. The insights gleaned from the study propel us towards a future where institutions, regardless of their size or resources, are not only contributors to but also beneficiaries of the rapid strides in AI. This harmonious and inclusive advancement bolsters the collective mission of unearthing scientific discoveries. In addition to establishing a framework that enables wider access to LLMs, our study also contributes to the understanding of their effective implementation. It highlights the significance of human experts' involvement, underscoring the invaluable contribution of human insight in enhancing AI capabilities. This confluence of human and AI expertise stands to fuel innovative solutions and unexpected discoveries in diverse scientific domains.

Our voyage into this novel realm of LLMs is in its nascent stage, and it holds the allure of a fascinating journey teeming with unanticipated knowledge and transformative milestones. As we navigate this path, we anticipate contributing to and witnessing the democratization of AI technologies, a development that holds the potential to revolutionize diverse facets of scientific research and discovery globally [15].

5. References

[1] Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., and Zhou, D. Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171, 2022.

[2] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., and Zhou, D. Chain of thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903, 2022.

[3] Zhang, Z., Zhang, A., Li, M., and Smola, A. Automatic chain of thought prompting in large lang/ October 2022, DOI:10.48550/arXiv.2210.03493.

[4] Accomazzi, A., Eichhorn, G., Kurtz, M. J., et al. 2007, ASP Conf. Proc., 377, 69

[5] Allen, A., DuPrie, K., Berriman, B., et al. 2013, ASP Conf. Proc., 475, 387

[6] Xie, S. M., Raghunathan, A., Liang, P., and Ma, T. An explanation of in-context learning as implicit bayesian inference. arXiv preprint arXiv:2111.02080, 2021.

[7] Grant, C. S., Thompson, D. M., Chyla, R., et al. 2014, These Proceedings Egret, D., Laurenceau, A., & Accomazzi, A. 2014, These Proceedings Henneken, E. A., Accomazzi, A., Kurtz, M. J., et al. 2014, arXiv:1406.4542.

[8] Almeida A, Anderson SF, Argudo-Fernández M, Badenes C, Barger K, Barrera-Ballesteros JK et al. The Eighteenth Data Release of the Sloan Digital Sky Surveys: Targeting and First Spectra from SDSS-V. Astrophysical Journal, Supplement Series. 2023 Aug 1;267(2):44. doi: 10.3847/1538-4365/acda98

[9] Accomazzi, A., Kurtz, M. J., Henneken, E. A., Chyla, R., Luker, J., Grant, C. S., Thompson, D. M., Holachek, A., Dave, R., and Murray, S. S. ADS: The Next Generation Search Platform. In Holl, A., Lesteven, S., Dietrich, D., and Gasperini, A. (eds.), Open Science at the Frontiers of Librarianship, volume 492 of Astronomical Society of the Pacific Conference Series, pp. 189, April 2015. Doi: 10.48550/arXiv.1503.04194.

[10] Singh, A., Mahajan, S., and Bagla, J. S. Study of galaxies on large-scale filaments in simulations. Monthly Notices of the Royal Astronomical Society, 497(2):2265–2275, jul 2020. doi: 10.1093/mnras/staa1913. Vig, J. Visualizing attention in transformer-based language representation models. arXiv preprint arXiv:1904.02679, 2019.

[11] Whitney, A., Ferreira, L., Conselice, C. J., and Duncan, K. Galaxy evolution in all five CANDELS fields and IllustrisTNG: Morphological, structural, and the major merger evolution to z ∼ 3. The Astrophysical Journal, 919(2):139, oct 2021. doi: 10.3847/1538-4357/ac1422.

[12] Holmquist, J. 2014, These Proceedings Kurtz, M. J., Eichhorn, G., Accomazzi, A., et al. 2000, A&AS, 143, 41.

[13] Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., & Murray, S. S. 2002, SPIE Conf.Proc., 4847, 238.

[14] Kurtz, M. J., Eichhorn, G., Accomazzi, A., et al. 2005, Journal of the American Society for Information Science and Technology, 56, 111 Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 2.

[15] O. Kalivoshko, V. Kraevsky, K. Burdeha, I. Lyutyy and N. Kiktev, "The Role of Innovation in Economic Growth: Information and Analytical Aspect," *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 2021, pp. 120-124, doi: 10.1109/PICST54195.2021.9772201