Interpreting Multi-band Galaxy Observations with Large Language Model-Based Agents

Anonymous Author(s)
Affiliation
Address
email

Abstract

Astronomical research traditionally relies on extensive domain knowledge to interpret observations and narrow down hypotheses. We demonstrate that this process can be emulated using large language model-based agents to accelerate research workflows. We propose `mephisto`, a multi-agent collaboration framework that mimics human reasoning to interpret multi-band galaxy observations. `mephisto` interacts with the `CIGALE` codebase, which includes spectral energy distribution (SED) models to explain observations. In this open-world setting, `mephisto` learns from experience, performs tree search, and accumulates knowledge in a dynamically updated base. As a proof of concept, we apply `mephisto` to the latest data from the James Webb Space Telescope. `mephisto` attains near-human proficiency in reasoning about galaxies' physical properties, even when dealing with a recently discovered population of "Little Red Dot" galaxies. This represents the first demonstration of agentic research in astronomy, advancing towards end-to-end research via LLM agents and potentially expediting astronomical discoveries.

1 Introduction

The advent of deep learning tools and the vast amount of data routinely collected in astronomical surveys—from hundreds of millions of spectra [1, 2, 3] to tens of billions of images [4, 5, 6]—has propelled astronomical research into high gear. However, much of the development in AI-assisted astronomy often focuses on optimizing individual downstream tasks, such as building classifiers and brokers to streamline surveys [7, 8, 9], emulating expensive hydrodynamical simulations [10, 11, 12], and advancing statistical inference using generative models as posterior and likelihood surrogates [13, 14, 15]. Despite these advancements, focusing solely on individual downstream tasks limits the potential impact of AI in astronomical research.

Unlike many scientific domains, astronomy lacks the ability to repeat experiments in controlled settings. Finding interesting astronomical phenomena based on limited and noisy observations often requires researchers to sift through a vast array of possible hypotheses to find viable and rational explanations. As an exhaustive search of all possible hypotheses to match the observations is computationally infeasible, the process of reasoning, filtering, and reflecting remains one of the crowning jewels of modern human innovation [16, 17, 18, 19, 20, 21]. This process essentially allows researchers to find solutions through "intuition" and experience—capabilities that go beyond optimizing individual downstream tasks.

With the growing trend of adopting large language models to design and conduct surveys in scientific research [22, 23, 24], a key question emerges: Can AI autonomously learn from its own experiences, reflect on failures and successes like humans, and ultimately build up enough knowledge to perform reasoning on physical models against observational data in astronomy without extensive exhaustive search? To explore this, we have developed a multi-agent collaboration framework, `mephisto`. We

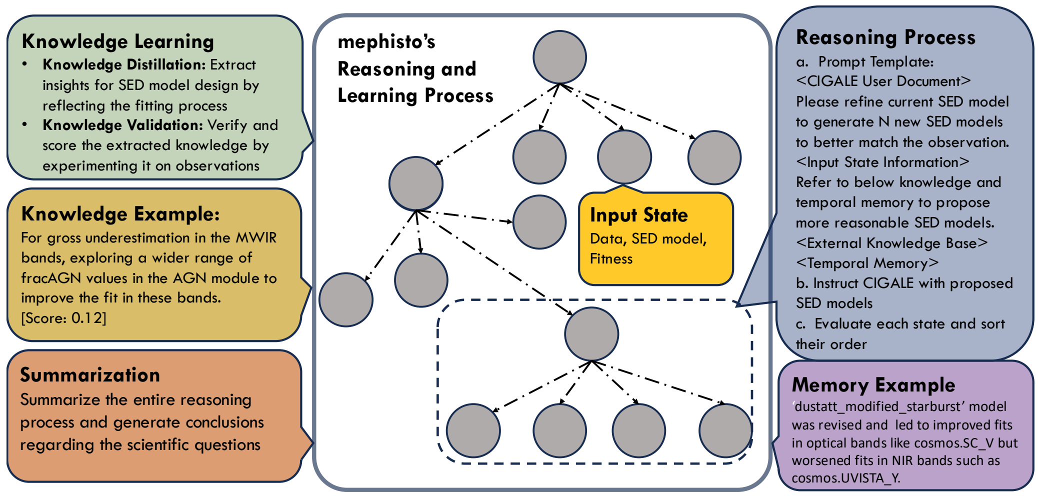


Figure 1: Schematic illustration of `mephisto`'s process for interpreting multi-band galaxy observations. Starting with a base SED model, `mephisto` assesses discrepancies between model predictions and observational data. It then generates four hypotheses (model variants) for improvement, which are input into `CIGALE` for refined fitting. `mephisto` selects the most promising unexpanded model variant based on fit quality, balancing model complexity with tree search depth. The process incorporates temporal memory to enhance efficiency and diversity of SED model proposals. Outcomes are compiled and distilled into a dynamically updated knowledge base through a learning system that autonomously extracts and validates knowledge. This iterative process mimics human reasoning, allowing `mephisto` to achieve near-human proficiency in interpreting galaxy SEDs.

focus on a specific astronomical research problem: reasoning about the task of fitting broadband photometric observations, known as Spectral Energy Distributions (SEDs), which can be seen as extremely low-resolution spectra of galaxies.

As a proof of concept, we apply `mephisto` to the latest data from the James Webb Space Telescope's JADES survey. The unique wavelength coverage and unprecedented spatial resolution of James Webb, the successor to the Hubble Space Telescope, has led to a trove of discoveries, including some of the highest redshift (earliest) galaxies ever observed [25, 26, 27], as well as a new population of galaxies known as the "Little Red Dots" [28, 29, 30, 31]. Interpreting the physics of galaxies through multi-band observations is fundamental to modern galaxy physics and cosmology. Key inquiries in galaxy evolution—including dark matter dynamics [32, 33, 34, 35] and the role of black holes [36, 37, 38, 39]—hinge on accurate interpretation of galaxy SEDs [40, 41, 42, 43]. The process of identifying model-data discrepancies, iteratively refining SED models, and developing viable explanations [44, 28, 45] requires extensive domain expertise and is time-intensive for researchers. Our work aims to emulate and accelerate this complex process using LLM-based agents.

2 Method

`mephisto` is a multi-agent collaboration framework designed to emulate human expert reasoning in fitting SED models. As illustrated in Figure 1, `mephisto` iteratively interacts with `CIGALE`[42] to construct increasingly complex SED models. These models incorporate various components including star formation histories [42], stellar populations [46, 47], dust attenuation and emissions [48, 49, 50], nebular emissions [51, 52], and active galactic nuclear emissions [53, 54]. To enhance model diversity and efficiency, we implement temporal memory and an external knowledge base, which is dynamically updated through knowledge distillation and validation, further improving the framework's performance over time, mimicking the learning process of a human expert.

Input State `mephisto` can be viewed as a reinforcement learning process without an explicit reward function, where an LLM agent implicitly evaluates the reward. The framework's goal is to assess the current SED "state" and propose actions to optimize the goodness of fit to JWST data.

Formally, the input state $s(d, m, r)$ (in json format, see Appendix A for details) comprises observation data d, described as wavelength-flux tuples due to current limitations in vision language models' ability to interpret scientific plots [55, 56]; SED model m, which includes physical model names, parameter ranges, priors, and grid sizes; and fitting results r, containing χ^2 information and auxiliary data such as the number of well-fitted photometry bands, as determined by the evaluation agent. This input structure, particularly the vague definition of well-fitted bands, mimics human practices. It allows the agent to consider overall S/N and global fit quality, potentially identifying corrupted data. This approach highlights the importance of agentic research in SED studies, where a physically implausible model with optimal χ^2 is often disfavored. The evaluation relies on human-like fuzzy logic and decision, which we aim to emulate, rather than solely on rigid statistical metrics.

Reasoning Process Unlike other machine learning applications in SED prediction [57, 58, 9, 13], `mephisto` goes beyond simple parameter estimation based on fixed models. It guides `CIGALE` through an exploration of various SED models, iteratively seeking plausible interpretations of observations. At each step, `mephisto` analyzes the state $s(d, m, r)$, reasons about discrepancies between model predictions and data, and generates $N_b = 4$ hypotheses to refine the current SED model. This iterative approach uses a tree structure to facilitate deliberate reasoning [59, 60, 61, 62] and mirrors human reasoning processes by progressing from simple to complex SED model construction [63, 64, 44]. We implement both depth-first search (DFS) and breadth-first search (BFS) strategies. After each node expansion, `mephisto` evaluates the children based on cumulative residuals, selecting the most promising unexpanded child for subsequent iterations. The SED model refinement and resulting fit quality changes are recorded in `mephisto`'s temporal memory to avoid redundant efforts.

Learning Process While large language models possess general knowledge of spectral energy distribution, they lack detailed insights into how specific `CIGALE` modules or parameters affect fit quality. To address this, we've integrated an external knowledge base into `mephisto`, following approaches in [65, 66, 67]. This knowledge base is refined through an iterative learning process and can incorporate "intuition" from experienced astronomers. `mephisto`'s learning process involves a knowledge distillation agent extracting insights from complete fitting histories, a knowledge validation agent assessing each insight's efficacy through experimentation, and a large language model integrating verified knowledge into the base. This validated knowledge is then used as context for future fits, ensuring continuous refinement and more direct action during tree searches. This process effectively allows `mephisto` to learn from previous SED fitting experiences, improving its performance over time.

3 Result

We demonstrate the performance of `mephisto` using the JWST JADES DR2 photometry catalog[26][1] in two distinct cases. For this study, we primarily use the GPT-4o model as the LLM backbone. However, our code, which will be released alongside this paper, is designed to easily accommodate any instruct LLM. We have also tested Claude-3.5-Sonnet and the more economical GLM-4 models, finding that they can perform comparably to what we demonstrate here.

To show that the model's iterative learning is performing as expected, we first project the JADES data using Self-Organizing Maps (SOM), a dimension reduction technique common in SED analysis [68, 69, 70]. We select 256 samples uniformly from the SOM latent space to ensure data diversity. Figure 2(a) illustrates how the average χ^2 of the fit improves both in terms of inference depth (at a given search) and run depth. Run depth is defined as the number of previous fits performed by `mephisto`. The results show that the dynamically updated knowledge base continues to enhance search efficiency. This allows `mephisto` to achieve similar fitting quality with much lower inference depth in subsequent runs when encountering similar spectra.

Beyond fitting the typical JWST data well, we also demonstrate `mephisto`'s ability to reason about "unknown unknown sources". We tested `mephisto` on 31 "Little Red Dots" documented in the literature [28]. These recently discovered objects were initially hypothesized to be early, dusty galaxies (hence appearing red), but alternative interpretations suggest they could be early galaxies with active galactic nuclei (AGN) [71, 29, 30, 31]. The right panel of Figure 2(a) shows the proposed solutions of `mephisto` for a representative source (with other 30 sources showing similar behavior,

[1] https://jades-survey.github.io

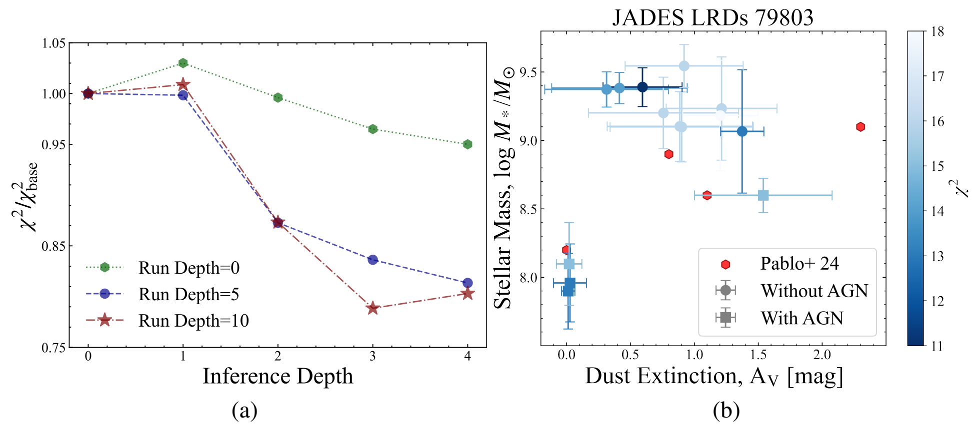


Figure 2: Performance and reasoning capabilities of `mephisto` on JWST data. Left: Average χ^2 vs. search depth for 256 SOM-selected JWST sources. Different lines represent various run depths, demonstrating how `mephisto`'s accumulated experience improves efficiency in subsequent runs when encountering similar spectra. Right: `mephisto`'s proposed solutions for a "Little Red Dot" (ID 79803) from the JADES catalog. Blue symbols represent different proposals with CIGALE statistical uncertainties, color-coded by χ^2. Two main scenarios are identified: a dusty star-forming galaxy without AGN and a dust-free galaxy with AGN, consistent with recent literature findings. The derived physical properties are consistent with those proposed in [28], shown in red.

see also Appendix C). Error bars indicate the statistical uncertainty from `CIGALE` for each particular model.

`mephisto` proposed two main solutions: a dusty star-forming galaxy devoid of an AGN, and a dust-free galaxy hosting an AGN (with lower dust extinction A_v) – see the model fits in Appendix C. The derived physical properties align with those reported in [28]. Notably, these objects were discovered after the GPT-4o training cutoff date. `mephisto`'s ability to reason about the nature (and debated characteristics) of these sources demonstrates its capacity to perform "unknown unknown" searches autonomously.

4 Broader Impact

This study presents, to our knowledge, the first end-to-end agentic research in astronomy. The SED fitting task serves as an ideal testing ground for LLM agents due to its complexity and text-parsable nature. Multiple hypotheses often satisfy the limited wavelength information from SED observations, leading to a vast array of potential physical models that preclude exhaustive search. `mephisto` demonstrates the ability to inherit understanding, distill knowledge, and develop intuition about solver routines through interactions with real data, discovering efficient paths toward multi-step reasoning akin to human researchers.

While SED fitting is a relatively straightforward task, the majority of observed sources from myriad surveys are rarely subjected to detailed human reasoning. Most are triaged based on simple χ^2 criteria through baseline models, potentially causing sources unexplainable by current physics to evade detection. The billions of SED sources collected by ongoing and upcoming surveys demand a more streamlined approach. `mephisto` represents a critical step towards the possibility of exhaustively analyzing all observed sources, potentially uncovering unknown phenomena or competing explanations for observations like the Little Red Dots.

Current limitations to scaling up to the entire sky stem from computational costs. `mephisto`'s reasoning consumes about 100K tokens, or approximately 1 US dollar per source, making it infeasible for billions of sources. However, as LLMs improve their reasoning ability and costs rapidly decrease, this study marks a crucial step toward fully utilizing all collected astronomical data.

A Input State Example

This appendix provides a detailed example of the input state used by `mephisto` in its SED fitting process. The input state, formatted in JSON, includes three key components: observational data, the `CIGALE` model configuration, and the fit quality metrics.

The dataset d encompasses redshift data from JADES and photometric details such as effective wavelengths, bandwidths, fluxes, and signal-to-noise ratios. Despite current AI limitations in interpreting scientific visuals quantitatively [55, 56], we input data numerically as tuples—a deviation from human researchers' reliance on visual analysis, yet a practical solution for AI agents.

The SED model m, constructed in `CIGALE`, integrates multiple physical components like star formation histories and dust attenuation. The fitting outcome r provides parameter estimates, residuals, computational cost, and fit quality per photometric band, assessed by the agent as good, overestimate, or underestimate. This qualitative method reduces AI numerical hallucinations [72, 73], guiding `mephisto` with textual cues for each filter evaluation. It also enables `mephisto` to apply domain knowledge to derive more physical changes, such as modifying dust attenuation physical models for suboptimal optical and UV fits.

The model input is then parsed and run by another agent designed to execute `CIGALE` codes. The results are evaluated by the reasoning prompt, where they are appended as the new state in the subsequent iteration of the tree search.

Below is an example of the input state structure:

```
{"data": {
            "redshift": 0.73,
            "photometry": [
                {
                    "name": "hst.wfc.F435W",
                    "wave_eff": 4337,
                    "bandwidth": 940,
                    "band_location_type": "OPTICAL",
                    "band_width_type": "BROAD",
                    "fit_quality": "good",
                    "signal_to_noise": "reliable"
                },
                ...]},
        "cigale_model": {
            "sfh": {
                "name": "sfhdelayed",
                "params": {
                "tau_main": [100, 500, 1000, 3000, 5000],
                "age_main": [100, 500, 1000, 3000, 5000],
                "tau_burst": [50],
                "age_burst": [20],
                "f_burst": [0.0]}},...},
        "cigale_best_parameters": {
            "sfh": {
                "name": "sfhdelayed",
                "params": {
                "tau_main": 100,
                "age_main": 500,
                "tau_burst": 50,
                "age_burst": 20,
                "f_burst": 0.0}},...},
        "cigale_fit_quality": {
        "num_of_good_fit": 5,
        "sum_chi2": 218,
        "grid_size": 450,
```

```
        "cigale_message": "CIGALE run smoothly"}}
```

B `mephisto` Reasoning Prompt

This section presents the reasoning prompt used by `mephisto` to guide its decision-making process in SED fitting. The prompt outlines the rules and knowledge base that `mephisto` adheres to when modifying CIGALE models to improve fit quality across different photometric bands. The prompt includes guidelines for module selection, parameter grid specification, and mandatory module requirements. Crucially, it incorporates two key components that enable `mephisto` to emulate human expert reasoning and learning:

1. **CIGALE SED Knowledge Base:** This component represents accumulated expertise from previous interactions with data. It enhances `mephisto`'s ability to improve fit quality for different photometric bands in the CIGALE model. The prompt instructs `mephisto` to consider this information when designing the CIGALE model. The agent dynamically filled with relevant expert knowledge about SED fitting, allowing `mephisto` to make informed decisions based on established astrophysical principles and learned insights from past analyses.

2. **Temporary Memory:** This feature captures information from the current tree search iteration. It allows `mephisto` to adapt its strategies within the ongoing analysis. The prompt instructs `mephisto` to consider these temporary memories to ensure robust diversity and superior fit quality in the updated CIGALE model. The model is updated with information from the current search process, enabling `mephisto` to refine its approach based on previous suggestions within the same analysis, avoiding redundant proposals.

The integration of the CIGALE SED Knowledge Base and Temporary Memory allows `mephisto` to generate diverse and reasonable model modifications, combining established principles with both long-term learned insights and short-term adaptive strategies. Importantly, the results of each analysis are not discarded but are instead processed to update the knowledge base. This cycle of learning and integration allows `mephisto` to continuously refine its knowledge base, improving its performance over time and across different analyses.

The prompt's output format guides `mephisto` to produce structured, diverse, and reasonable CIGALE model modifications. Each proposed modification includes a "thinking" step, module name, specific choice, and parameter grid values.

<CIGALE Documentation>
<Introduction to User Input>
Your task is to analyze user input and adhere to below rules and knowledges to modify the provided CIGALE models to improve the fit quality of different bands. You should identify the CIGALE module which should affect the fitting quality the most and carefully modify this module, e.g., use another choice, adjust the parameter grid, and etc.

- Choice Selection: For each module in the model, select either one choice or none. This decision should be based on the need to optimize the model's fit for the observational data.
- Parameter Grid Specification: For the selected choice in each module, define a parameter grid that the model will use to fit the data.
  - For discrete parameters, select grid values from a pre-defined list.
  - For continuous parameters, derive grid values within a specified range to ensure a comprehensive exploration of parameter space.
- Mandatory Modules: The model configuration must include specific settings for the following modules:
  - sfh (Star Formation History): This module is essential for modeling the rate at which a galaxy forms stars over time.

    - ssp (Simple Stellar Population): This module is crucial for understanding the collective properties of stars in a galaxy that formed at the same time and with the same metallicity.

- For 'imf' parameter in 'bc03' and 'm2005', it should be 0 or 1, not a list, i.e., "imf": 1 are accepteable, "imf": [1] are foridden
- For 'disk_type' parameter in 'fritz2006' and 'skirtor2016' it should be 0 or 1, not a list, i.e., "disk_type": 0 are acceptable, "disk_type": [1] are forbidden

##CIGALE SED Knowledge##
To enhance the fit quality for different photometric bands in the CIGALE model, consider the following addtional information when designing your CIGALE model.
%%KNOWLEDGE%%

##Temporary Memory##
Please take into account the following temporary memories to ensure a robust diversity and superior fit quality in the updated CIGALE model:
%%MEMORY%%

##Expected Output Format##
Your output format should be structured as below list constructed by four diverse and reasonable CIGALE model modifications:

```
[
    {
        "thinking": "thinking for CIGALE model modification 0 here"
        "module": "module name for CIGALE model modification 0 here",
        "name": "module choice name for CIGALE model modification 0 here",
        "parameters": [
            "parameter 1": ["parameter 1 grid here"],
            "parameter 2": ["parameter 2 grid here"],
            ...
            "parameter n": ["parameter n grid here"]
        ]
    },
    ...
]
```

C SED Fitting Result on Little Red Dots

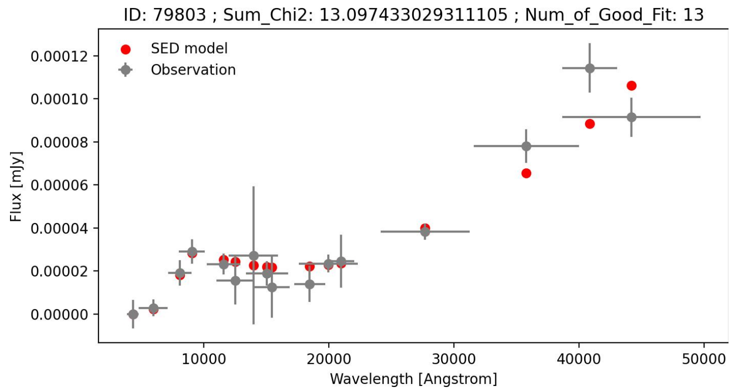
This appendix presents the SED fitting results obtained by `mephisto` for five representative JADES Little Red Dot (LRD) galaxies, demonstrating its ability to reason about complex, multimodal solutions in a manner akin to human researchers.

Little Red Dots, recently discovered by the James Webb Space Telescope, have sparked debate in the astronomical community. Initially thought to be early, dusty galaxies, they may alternatively be early galaxies with active galactic nuclei (AGN) [28, 29, 30, 31]. This uncertainty underscores the challenges in interpreting high-redshift galaxy observations and the need for sophisticated, human-like analysis techniques.

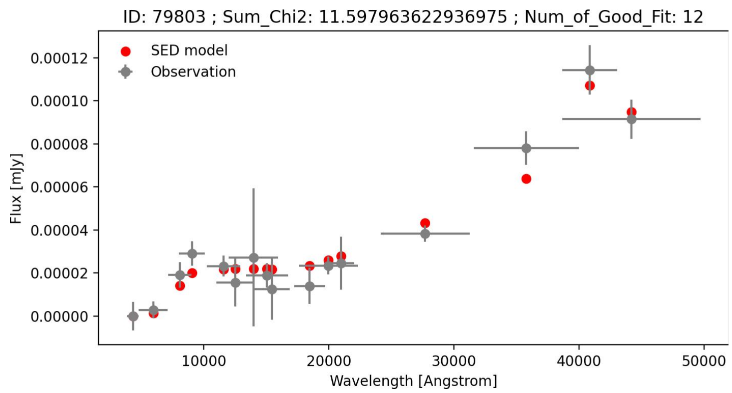
For each LRD, `mephisto` has identified two distinct, physically plausible solutions:

1. A dusty star-forming galaxy without an AGN
2. A relatively dust-free galaxy hosting an AGN

Crucially, `mephisto` arrived at these solutions through a process of reflection and model changes, not simply via hyperparameter search. This approach demonstrates its capacity for "unknown unknown" searches, proposing and evaluating multiple hypotheses to explain challenging observations.



(a) $A_v = 0.03 \pm 0.13\,\mathrm{mag}$; $\log M_*/M_\odot = 7.96 \pm 0.28$; $\mathrm{frac}_{\mathrm{AGN}} = 0.99$



(b) $A_v = 0.92 \pm 0.46\,\mathrm{mag}$; $\log M_*/M_\odot = 9.54 \pm 0.15$; $\mathrm{frac}_{\mathrm{AGN}} = 0$

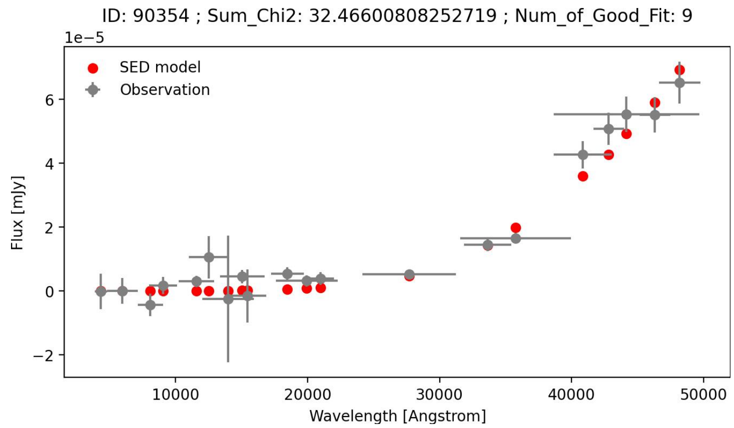
Figure 3: SED Fitting Results for JADES LRD 79803. Panel (a) shows a solution favoring a dust-free galaxy with a dominant AGN ($\mathrm{frac}_{\mathrm{AGN}} = 0.99$), while panel (b) presents an alternative scenario of a dustier ($A_v = 0.92$) star-forming galaxy without AGN contribution. These contrasting solutions, both providing good fits to the observational data, exemplify the current debate in the literature regarding the nature of Little Red Dots.

The figures illustrate these distinct solutions, showcasing `mephisto`'s ability to navigate the complex hypothesis space of SED modeling. Each subfigure provides key derived parameters: dust attenuation (A_v), stellar mass (M_*), and AGN fraction, highlighting the fundamental differences between the proposed scenarios.

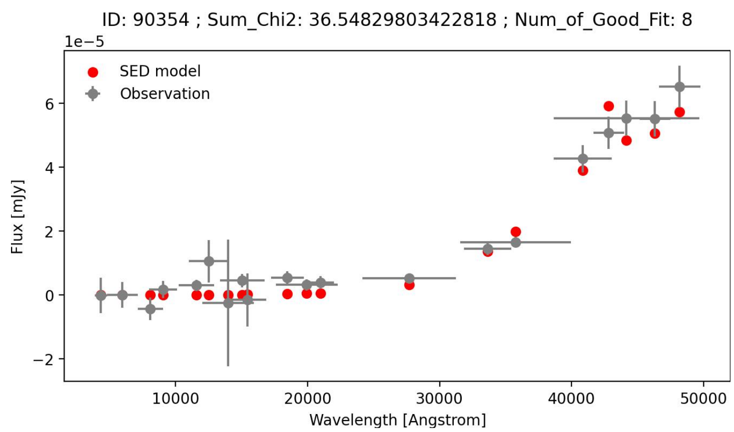
By identifying multiple valid solutions, `mephisto` contributes meaningfully to ongoing astronomical debates, providing a comprehensive view of possible physical scenarios for these objects. The contrasting solutions highlight the degeneracies inherent in SED fitting for high-redshift objects with limited photometric data, mimicking the careful consideration human experts apply to such challenging cases.

The range of physical parameters explored - from nearly dust-free to heavily obscured galaxies, and from pure star-forming systems to AGN-dominated ones - demonstrates `mephisto`'s flexibility in hypothesis generation. This capability is key to performing "unknown unknown" searches, potentially uncovering unexpected physical scenarios that human researchers might overlook.

These results showcase `mephisto`'s potential as a powerful tool for astronomical research, capable of not only fitting SEDs but also engaging in exploratory data analysis and hypothesis generation traditionally performed by human experts.



(a) $A_v = 10.1 \pm 2.2\,\mathrm{mag}$; $\log M_*/M_\odot = 12.7 \pm 0.12$; $\mathrm{frac_{AGN}} = 0.99$



(b) $A_v = 3.01 \pm 0.53\,\mathrm{mag}$; $\log M_*/M_\odot = 10.1 \pm 0.12$; $\mathrm{frac_{AGN}} = 0$

Figure 4: SED Fitting Results for JADES LRD 90354. Panel (a) depicts a heavily dust-obscured ($A_v = 10.1$) galaxy with a strong AGN component, while panel (b) shows a less dusty ($A_v = 3.01$) star-forming galaxy without AGN.

References

[1] D. G. York, J. Adelman, J. Anderson, John E., S. F. Anderson, J. Annis, and et al., "The Sloan Digital Sky Survey: Technical Summary," *The Astronomical Journal*, vol. 120, pp. 1579–1587, Sept. 2000.

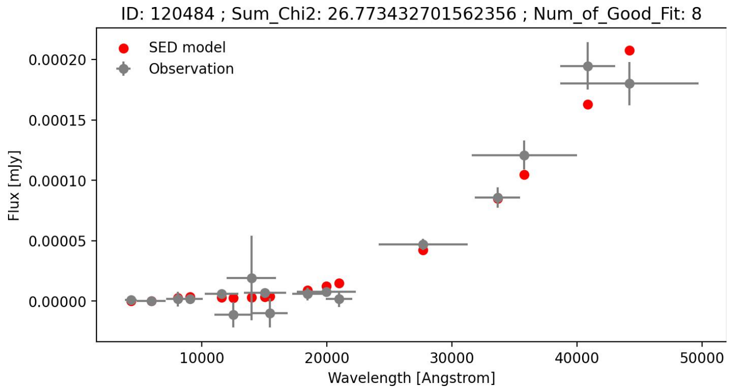
[2] R. S. de Jong, O. Agertz, A. A. Berbel, J. Aird, D. A. Alexander, and et al., "4MOST: Project overview and information for the First Call for Proposals," *The Messenger*, vol. 175, pp. 3–11, Mar. 2019.

[3] DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, and et al., "The Early Data Release of the Dark Energy Spectroscopic Instrument," *The Astronomical Journal*, vol. 168, p. 58, Aug. 2024.

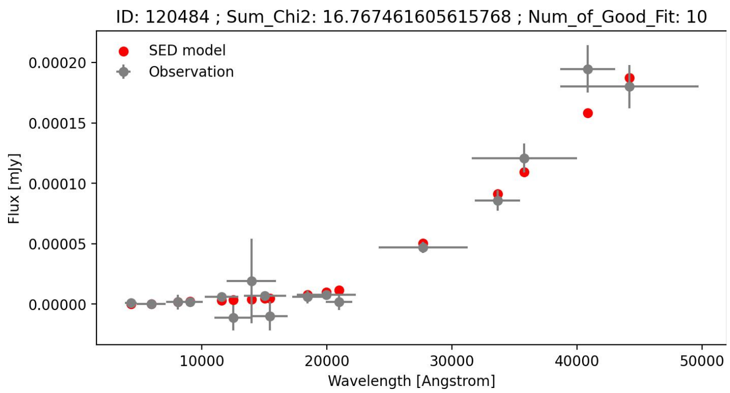
[4] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, and et al., "LSST: From Science Drivers to Reference Design and Anticipated Data Products," *The Astrophysical Journal*, vol. 873, p. 111, Mar. 2019.

[5] Euclid Collaboration, R. Scaramella, J. Amiaux, Y. Mellier, C. Burigana, and et al., "Euclid preparation. I. The Euclid Wide Survey," *Astronomy and Astrophysics*, vol. 662, p. A112, June 2022.

[6] H. Aihara, Y. AlSayyad, M. Ando, R. Armstrong, J. Bosch, and et al., "Third data release of the Hyper Suprime-Cam Subaru Strategic Program," *Publications of the Astronomical Society of Japan*, vol. 74, pp. 247–272, Apr. 2022.



(a) $A_v = 0.77 \pm 0.65\,\mathrm{mag}$; $\log M_*/M_\odot = 6.68 \pm 1.12$; $\mathrm{frac}_{\mathrm{AGN}} = 0.99$



(b) $A_v = 0.92 \pm 0.46\,\mathrm{mag}$; $\log M_*/M_\odot = 9.55 \pm 0.16$; $\mathrm{frac}_{\mathrm{AGN}} = 0$

Figure 5: SED Fitting Results for JADES LRD 120484. Panel (a) presents a low-mass galaxy dominated by AGN emission, while panel (b) shows a more massive, purely star-forming galaxy.

[7] B. Wang, J. Zou, Z. Cai, J. X. Prochaska, Z. Sun, and et al., "Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems," *The Astrophysical Journal Supplement Series*, vol. 259, p. 28, Mar. 2022.

[8] G. Stein, J. Blaum, P. Harrington, T. Medan, and Z. Lukić, "Mining for Strong Gravitational Lenses with Self-supervised Learning," *The Astrophysical Journal*, vol. 932, p. 107, June 2022.

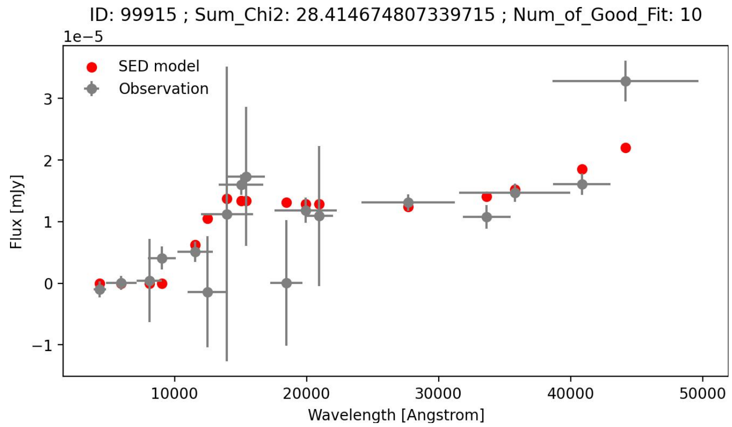
[9] B. E. Robertson, S. Tacchella, B. D. Johnson, R. Hausen, A. B. Alabi, K. Boyett, A. J. Bunker, S. Carniani, E. Egami, D. J. Eisenstein, K. N. Hainline, J. M. Helton, Z. Ji, N. Kumari, J. Lyu, R. Maiolino, E. J. Nelson, M. J. Rieke, I. Shivaei, F. Sun, H. Übler, C. C. Williams, C. N. A. Willmer, and J. Witstok, "Morpheus Reveals Distant Disk Galaxy Morphologies with JWST: The First AI/ML Analysis of JWST Images," *The Astrophysical Journal Letters*, vol. 942, p. L42, Jan. 2023.

[10] J. Chardin, G. Uhlrich, D. Aubert, N. Deparis, N. Gillet, and et al., "A deep learning model to emulate simulations of cosmic reionization," *Monthly Notices of the Royal Astronomical Society*, vol. 490, pp. 1055–1065, Nov. 2019.

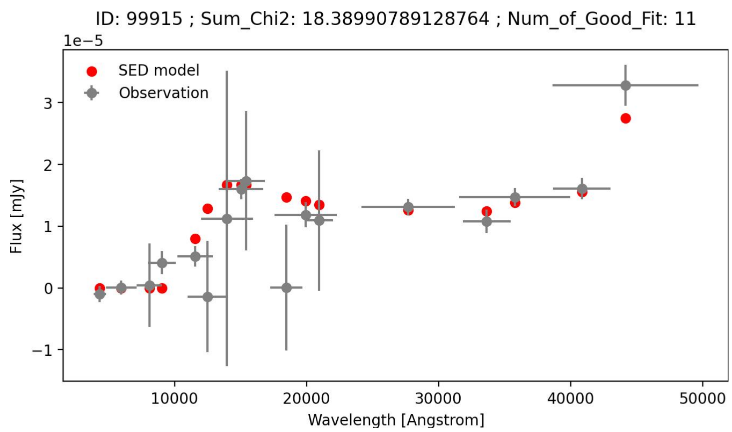
[11] S. He, Y. Li, Y. Feng, S. Ho, S. Ravanbakhsh, and et al., "Learning to predict the cosmological structure formation," *Proceedings of the National Academy of Science*, vol. 116, pp. 13825–13832, July 2019.

[12] Y. Li, Y. Ni, R. A. C. Croft, T. Di Matteo, S. Bird, and et al., "AI-assisted superresolution cosmological simulations," *Proceedings of the National Academy of Science*, vol. 118, p. e2022038118, May 2021.

[13] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proceedings of the National Academy of Science*, vol. 117, pp. 30055–30062, Dec. 2020.



(a) $A_v = 0.07 \pm 0.65\,\mathrm{mag}$; $\log M_*/M_\odot = 8.70 \pm 0.15$; $\mathrm{frac_{AGN}} = 0.99$



(b) $A_v = 0.46 \pm 0.56\,\mathrm{mag}$; $\log M_*/M_\odot = 8.2 \pm 0.27$; $\mathrm{frac_{AGN}} = 0$

Figure 6: SED Fitting Results for JADES LRD 99915. Both panels show relatively low-dust solutions, with panel (a) featuring a strong AGN component and panel (b) representing a purely star-forming galaxy. This case demonstrates that even with similar dust attenuation, the presence or absence of an AGN can significantly alter the interpretation of the galaxy's nature.

[14] X. Zhao, Y. Mao, C. Cheng, and B. D. Wandelt, "Simulation-based Inference of Reionization Parameters from 3D Tomographic 21 cm Light-cone Images," *The Astrophysical Journal*, vol. 926, p. 151, Feb. 2022.

[15] Z. Sun, Y.-S. Ting, and Z. Cai, "Quasar Factor Analysis-An Unsupervised and Probabilistic Quasar Continuum Prediction Algorithm with Latent Factor Analysis," *The Astrophysical Journal Supplement Series*, vol. 269, p. 4, Nov. 2023.

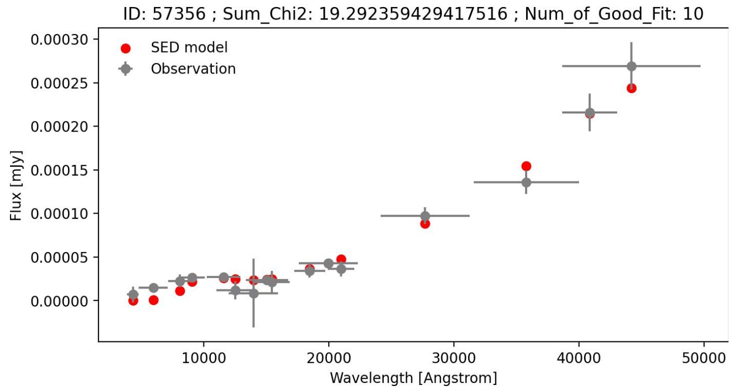
[16] A. A. Abdo, M. Ackermann, M. Arimoto, K. Asano, W. B. Atwood, and et al., "Fermi Observations of High-Energy Gamma-Ray Emission from GRB 080916C," *Science*, vol. 323, p. 1688, Mar. 2009.

[17] M. Gillon, E. Jehin, S. M. Lederer, L. Delrez, J. de Wit, and et al., "Temperate Earth-sized planets transiting a nearby ultracool dwarf star," *Nature*, vol. 533, pp. 221–224, May 2016.

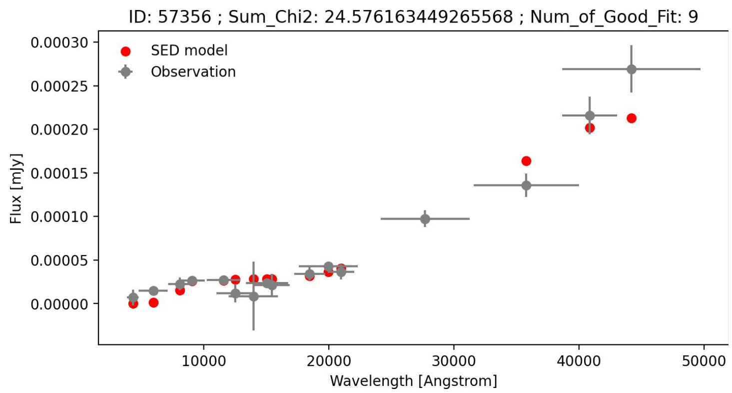
[18] L. G. Spitler, P. Scholz, J. W. T. Hessels, S. Bogdanov, A. Brazier, and et al., "A repeating fast radio burst," *Nature*, vol. 531, pp. 202–205, Mar. 2016.

[19] F. Wang, J. Yang, X. Fan, J. F. Hennawi, A. J. Barth, and et al., "A Luminous Quasar at Redshift 7.642," *The Astrophysical Journal*, vol. 907, p. L1, Jan. 2021.

[20] B. Welch, D. Coe, J. M. Diego, A. Zitrin, E. Zackrisson, and et al., "A highly magnified star at redshift 6.2," *Nature*, vol. 603, pp. 815–818, Mar. 2022.



(a) $A_v = 1.35 \pm 1.12\,\mathrm{mag}$; $\log M_*/M_\odot = 9.22 \pm 0.61$; $\mathrm{frac_{AGN}} = 0.99$



(b) $A_v = 0.36 \pm 0.48\,\mathrm{mag}$; $\log M_*/M_\odot = 9.88 \pm 0.08$; $\mathrm{frac_{AGN}} = 0$

Figure 7: SED Fitting Results for JADES LRD 57356. Panel (a) shows a dustier solution with a dominant AGN, while panel (b) presents a less dusty, purely star-forming galaxy.

[21] R. Maiolino, J. Scholtz, J. Witstok, S. Carniani, F. D'Eugenio, and et al., "A small and vigorous black hole in the early Universe," *Nature*, vol. 627, pp. 59–63, Mar. 2024.

[22] N. Yoshikawa, M. Skreta, K. Darvish, S. Arellano-Rubach, Z. Ji, L. Bjørn Kristensen, A. Z. Li, Y. Zhao, H. Xu, A. Kuramshin, A. Aspuru-Guzik, F. Shkurti, and A. Garg, "Large language models for chemistry robotics," *Autonomous Robots*, vol. 47, no. 8, pp. 1057–1086, 2023.

[23] G. Lei, R. Docherty, and S. J. Cooper, "Materials science in the era of large language models: a perspective," *arXiv e-prints*, p. arXiv:2403.06949, Mar. 2024.

[24] K. M. Jablonka, P. Schwaller, A. Ortega-Guerrero, and B. Smit, "Leveraging large language models for predictive chemistry," *Nature Machine Intelligence*, vol. 6, no. 2, pp. 161–169, 2024.

[25] C. M. Casey, J. S. Kartaltepe, N. E. Drakos, M. Franco, S. Harish, and et al., "COSMOS-Web: An Overview of the JWST Cosmic Origins Survey," *The Astrophysical Journal*, vol. 954, p. 31, Sept. 2023.

[26] D. J. Eisenstein, C. Willott, S. Alberts, S. Arribas, N. Bonaventura, and et al., "Overview of the JWST Advanced Deep Extragalactic Survey (JADES)," *arXiv e-prints*, p. arXiv:2306.02465, June 2023.

[27] J. R. Weaver, S. E. Cutler, R. Pan, K. E. Whitaker, I. Labbé, and et al., "The UNCOVER Survey: A First-look HST + JWST Catalog of 60,000 Galaxies near A2744 and beyond," *The Astrophysical Journal Supplement Series*, vol. 270, p. 7, Jan. 2024.

[28] P. G. Pérez-González, G. Barro, G. H. Rieke, J. Lyu, M. Rieke, and et al., "What Is the Nature of Little Red Dots and what Is Not, MIRI SMILES Edition," *The Astrophysical Journal*, vol. 968, p. 4, June 2024.

[29] J. Matthee, R. P. Naidu, G. Brammer, J. Chisholm, A.-C. Eilers, and et al., "Little Red Dots: An Abundant Population of Faint Active Galactic Nuclei at z $\sim$ 5 Revealed by the EIGER and FRESCO JWST Surveys," *The Astrophysical Journal*, vol. 963, p. 129, Mar. 2024.

[30] F. Gentile, C. M. Casey, H. B. Akins, M. Franco, J. McKinney, E. Berman, O. R. Cooper, N. E. Drakos, M. Hirschmann, A. S. Long, G. Magdis, A. M. Koekemoer, V. Kokorev, M. Shuntov, M. Talia, N. Allen, S. Harish, O. Ilbert, H. J. McCracken, J. S. Kartaltepe, D. Liu, L. Paquereau, J. Rhodes, M. R. Rich, B. Robertson, S. Toft, and G. Gozaliasl, "Not-so-little Red Dots: Two massive and dusty starbursts at z$\sim$5-7 pushing the limits of star formation discovered by JWST in the COSMOS-Web survey," *arXiv e-prints*, p. arXiv:2408.10305, Aug. 2024.

[31] J. F. W. Baggen, P. van Dokkum, G. Brammer, A. de Graaff, M. Franx, and et al., "The Small Sizes and High Implied Densities of 'Little Red Dots' with Balmer Breaks Could Explain Their Broad Emission Lines Without an AGN," *arXiv e-prints*, p. arXiv:2408.07745, Aug. 2024.

[32] C. A. Collins, J. P. Stott, M. Hilton, S. T. Kay, S. A. Stanford, and et al., "Early assembly of the most massive galaxies," *Nature*, vol. 458, pp. 603–606, Apr. 2009.

[33] R. Mandelbaum, "Weak Lensing for Precision Cosmology," *Annual Review of Astronomy and Astrophysics*, vol. 56, pp. 393–433, Sept. 2018.

[34] V. Desjacques, D. Jeong, and F. Schmidt, "Large-scale galaxy bias," *Physics Reports*, vol. 733, pp. 1–193, Feb. 2018.

[35] R. H. Wechsler and J. L. Tinker, "The Connection Between Galaxies and Their Dark Matter Halos," *Annual Review of Astronomy and Astrophysics*, vol. 56, pp. 435–487, Sept. 2018.

[36] A. Cattaneo, S. M. Faber, J. Binney, A. Dekel, J. Kormendy, and et al., "The role of black holes in galaxy formation and evolution," *Nature*, vol. 460, pp. 213–219, July 2009.

[37] A. C. Fabian, "Observational Evidence of Active Galactic Nuclei Feedback," *Annual Review of Astronomy and Astrophysics*, vol. 50, pp. 455–489, Sept. 2012.

[38] T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, and et al., "Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing," *Physical Review D*, vol. 105, p. 023520, Jan. 2022.

[39] T. Wang, K. Xu, Y. Wu, Y. Shi, D. Elbaz, and et al., "Black holes regulate cool gas accretion in massive galaxies," *Nature*, vol. 632, pp. 1009–1013, Aug. 2024.

[40] G. B. Brammer, P. G. van Dokkum, and P. Coppi, "EAZY: A Fast, Public Photometric Redshift Code," *The Astrophysical Journal*, vol. 686, pp. 1503–1513, Oct. 2008.

[41] J. Chevallard and S. Charlot, "Modelling and interpreting spectral energy distributions of galaxies with BEAGLE," *Monthly Notices of the Royal Astronomical Society*, vol. 462, pp. 1415–1443, Oct. 2016.

[42] M. Boquien, D. Burgarella, Y. Roehlly, V. Buat, L. Ciesla, D. Corre, A. K. Inoue, and H. Salas, "CIGALE: a python Code Investigating GALaxy Emission," *Astronomy & Astrophysics*, vol. 622, p. A103, Feb. 2019.

[43] B. D. Johnson, J. Leja, C. Conroy, and J. S. Speagle, "Stellar Population Inference with Prospector," *The Astrophysical Journal Supplement Series*, vol. 254, p. 22, June 2021.

[44] C. Pacifici, K. G. Iyer, B. Mobasher, E. da Cunha, V. Acquaviva, and et al., "The Art of Measuring Physical Parameters in Galaxies: A Critical Assessment of Spectral Energy Distribution Fitting Techniques," *The Astrophysical Journal*, vol. 944, p. 141, Feb. 2023.

[45] B. Wang, J. Leja, H. Atek, I. Labbé, Y. Li, and et al., "Quantifying the Effects of Known Unknowns on Inferred High-redshift Galaxy Properties: Burstiness, IMF, and Nebular Physics," *The Astrophysical Journal*, vol. 963, p. 74, Mar. 2024.

[46] G. Bruzual and S. Charlot, "Stellar population synthesis at the resolution of 2003," *Monthly Notices of the Royal Astronomical Society*, vol. 344, pp. 1000–1028, Oct. 2003.

[47] C. Maraston, "Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 362, pp. 799–825, Sept. 2005.

[48] C. Leitherer, I. H. Li, D. Calzetti, and T. M. Heckman, "Global Far-Ultraviolet (912-1800 Å) Properties of Star-forming Galaxies," *The Astrophysical Journal Supplement Series*, vol. 140, pp. 303–329, June 2002.

[49] D. Calzetti, L. Armus, R. C. Bohlin, A. L. Kinney, J. Koornneef, and et al., "The Dust Content and Opacity of Actively Star-forming Galaxies," *The Astrophysical Journal*, vol. 533, pp. 682–695, Apr. 2000.

[50] D. A. Dale, G. Helou, G. E. Magdis, L. Armus, T. Díaz-Santos, and et al., "A Two-parameter Model for the Infrared/Submillimeter/Radio Spectral Energy Distributions of Galaxies and Active Galactic Nuclei," *The Astrophysical Journal*, vol. 784, p. 83, Mar. 2014.

[51] G. Ferland, K. Korista, D. Verner, J. Ferguson, J. Kingdon, and E. Verner, "Cloudy 90: numerical simulation of plasmas and their spectra," *Publications of the Astronomical Society of the Pacific*, vol. 110, no. 749, p. 761, 1998.

[52] G. J. Ferland, R. L. Porter, P. A. M. van Hoof, R. J. R. Williams, N. P. Abel, and et al., "The 2013 Release of Cloudy," *Revista Mexicana de Astronomia y Astrofisica*, vol. 49, pp. 137–163, Apr. 2013.

[53] J. Fritz, A. Franceschini, and E. Hatziminaoglou, "Revisiting the infrared spectra of active galactic nuclei with a new torus emission model," *Monthly Notices of the Royal Astronomical Society*, vol. 366, pp. 767–786, Mar. 2006.

[54] P. Camps and M. Baes, "SKIRT: An advanced dust radiative transfer code with a user-friendly architecture," *Astronomy and Computing*, vol. 9, pp. 20–33, Mar. 2015.

[55] L. Li, Y. Wang, R. Xu, P. Wang, X. Feng, and et al., "Multimodal ArXiv: A Dataset for Improving Scientific Comprehension of Large Vision-Language Models," *arXiv e-prints*, p. arXiv:2403.00231, Feb. 2024.

[56] X. Yue, Y. Ni, K. Zhang, T. Zheng, R. Liu, and et al., "MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI," *arXiv e-prints*, p. arXiv:2311.16502, Nov. 2023.

[57] A. Diaz Rivero and C. Dvorkin, "Flow-based likelihoods for non-Gaussian inference," *Physical Review D*, vol. 102, p. 103507, Nov. 2020.

[58] Z. Sun, J. S. Speagle, S. Huang, Y.-S. Ting, and Z. Cai, "Zephyr : Stitching Heterogeneous Training Data with Normalizing Flows for Photometric Redshift Inference," *arXiv e-prints*, p. arXiv:2310.20125, Oct. 2023.

[59] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, and K. Narasimhan, "Tree of Thoughts: Deliberate Problem Solving with Large Language Models," *arXiv e-prints*, p. arXiv:2305.10601, May 2023.

[60] M. Besta, N. Blach, A. Kubicek, R. Gerstenberger, M. Podstawski, and et al., "Graph of Thoughts: Solving Elaborate Problems with Large Language Models," *arXiv e-prints*, p. arXiv:2308.09687, Aug. 2023.

[61] B. Sel, A. Al-Tawaha, V. Khattar, R. Jia, and M. Jin, "Algorithm of Thoughts: Enhancing Exploration of Ideas in Large Language Models," *arXiv e-prints*, p. arXiv:2308.10379, Aug. 2023.

[62] W. Chen, X. Ma, X. Wang, and W. W. Cohen, "Program of Thoughts Prompting: Disentangling Computation from Reasoning for Numerical Reasoning Tasks," *arXiv e-prints*, p. arXiv:2211.12588, Nov. 2022.

[63] J. Walcher, B. Groves, T. Budavári, and D. Dale, "Fitting the integrated spectral energy distributions of galaxies," *Astrophysics and Space Science*, vol. 331, pp. 1–52, Jan. 2011.

[64] M. Sawicki, "SEDfit: Software for Spectral Energy Distribution Fitting of Photometric Data," *Publications of the Astronomical Society of the Pacific*, vol. 124, p. 1208, Nov. 2012.

[65] G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, and et al., "Voyager: An Open-Ended Embodied Agent with Large Language Models," *arXiv e-prints*, p. arXiv:2305.16291, May 2023.

[66] A. Zhao, D. Huang, Q. Xu, M. Lin, Y.-J. Liu, and et al., "ExpeL: LLM Agents Are Experiential Learners," *arXiv e-prints*, p. arXiv:2308.10144, Aug. 2023.

[67] C. Qian, S. Liang, Y. Qin, Y. Ye, X. Cong, and et al., "Investigate-Consolidate-Exploit: A General Strategy for Inter-Task Agent Self-Evolution," *arXiv e-prints*, p. arXiv:2401.13996, Jan. 2024.

[68] M. Carrasco Kind and R. J. Brunner, "SOMz: photometric redshift PDFs with self-organizing maps and random atlas," *Monthly Notices of the Royal Astronomical Society*, vol. 438, pp. 3409–3421, Mar. 2014.

[69] J. S. Speagle, A. Leauthaud, S. Huang, C. P. Bradshaw, F. Ardila, and et al., "Galaxy-Galaxy lensing in HSC: Validation tests and the impact of heterogeneous spectroscopic training sets," *Monthly Notices of the Royal Astronomical Society*, vol. 490, pp. 5658–5677, Dec. 2019.

[70] A. H. Wright, H. Hildebrandt, J. L. van den Busch, and C. Heymans, "Photometric redshift calibration with self-organising maps," *Astronomy and Astrophysics*, vol. 637, p. A100, May 2020.

[71] I. Labbé, P. van Dokkum, E. Nelson, R. Bezanson, K. A. Suess, and et al., "A population of red candidate massive galaxies 600 Myr after the Big Bang," *Nature*, vol. 616, pp. 266–269, Apr. 2023.

[72] Z. Yuan, H. Yuan, C. Tan, W. Wang, and S. Huang, "How well do Large Language Models perform in Arithmetic tasks?," *arXiv e-prints*, p. arXiv:2304.02015, Mar. 2023.

[73] J. Ahn, R. Verma, R. Lou, D. Liu, R. Zhang, and et al., "Large Language Models for Mathematical Reasoning: Progresses and Challenges," *arXiv e-prints*, p. arXiv:2402.00157, Jan. 2024.