

Exploring AI-powered Multimodal Analogies for Science Education

Shradha Sehgal¹, Bhavya², Krishna Phani Datta³, Aditi Mallavarapu⁴ and ChengXiang Zhai⁵

¹University of Illinois Urbana-Champaign

Abstract

Analogies are an effective teaching tool for helping students understand new concepts by connecting them to familiar contexts. However, generating analogies that aid students' learning is non-trivial and requires a nuanced understanding that draws meaningful parallels between familiar concepts. Researchers have addressed this challenge by using computational models to generate textual or word-level analogies. We believe that adding visual elements to textual analogical explanations can offer greater comprehension to students than relying solely on textual analogies. Accordingly, we introduce the idea of multimodal analogies – a fusion of textual analogies and their visual counterparts to enhance understanding of scientific concepts. Further, we introduce and explore generating three types of multimodal analogies for science education, namely, general analogies; adaptive analogies tailored to the background, needs and preferences of learners; and iteratively refined analogies via human-AI and multi-agent collaboration. We leverage models like GPT-4 for text generation followed by DALL-E-3 for images and qualitatively analyze the created analogies of each of the three types. Our analysis helps identify some limitations of existing models and pinpoint future research directions in this area. Moreover, we showcase a demo system where students can engage with multimodal analogies and provide feedback. We aim to use this system to garner feedback on the AI-generated analogies and ultimately create a large-scale, high quality dataset of multi-modal analogies for scientific education.

Keywords

Analogies, Education, Large Language Models, Multimodal, CEUR-WS

1. Introduction

Analogies are comparisons that highlight similarities between two different things to clarify or explain concepts [1]. They function by transferring knowledge from a well-understood subject (the source or analogue) to one that is less familiar (the target) [2]. Analogies are useful as an educational tool as they have been proven to boost understanding and critical thinking among students [3]. By connecting new and complex concepts to familiar ones, analogies help students bridge gaps in knowledge and making the learning process more effective [4]. Generating analogies requires extensive topic knowledge and the ability to think abstractly and creatively [5]. Thus, analogies are typically created by specialists within a field who have comprehensive knowledge of the concepts [6]. To automate this process and reduce the time for generating analogies, researchers have studied the automated generation of word-level analogies like “king:man :: queen:woman” using computational methods [7, 8, 9]. Few works have investigated creating science analogies with explanations but these have been limited to using only the text modality [10, 11].

Given the effectiveness of visual elements in aiding student learning [12], we propose to augment explanation-type textual analogies with image representations to create multimodal analogies. We believe adding visual components to textual analogies can enhance the overall understandability of the content and increase student engagement. Specifically for science concepts that often include structural diagrams and complex relations, visual analogies can help students understand concepts alongside text.

To this end, we explore how to leverage LLMs and diffusion models to generate three types of multi-modal analogies, namely general analogies, adaptive analogies tailored

to students' background, needs, and preferences, and iteratively refined analogies via human-AI collaboration and multi-agent collaboration.

We leverage large language models like GPT-4 to generate a text analogy and the text-to-image model DALL-E-3 for generating the image analogy. In documenting the opportunities and limitations of multimodal analogies, we highlight the nuances of the valid and invalid instances through examples. Finally, we showcase a demo system that we plan to release to the community that can be used to leverage human-in-the-loop feedback to improve analogy generation. Our work shows potential for creating unique content for science education and lays the foundation for future research in this important area.

2. Related Work

In this section, we describe related work on computational models of analogies, application of analogies to education, and leveraging LLMs for education.

2.1. Computational Models of Analogies

Computational modeling of analogies refers to the algorithms and models for generating analogies. This section reviews the computational models used to generate analogies of different modalities - text and visual.

2.1.1. Text-based analogies

Analogies have predominantly been studied at the word-level, in the form of “A:B::C:D”, such as, “king:man::queen:woman” [7, 8, 9]. These type of proportional analogies are commonly used in entrance exams like the SAT or NCEE to test student understanding. There exist multiple ways to create word-level analogies, one prominent approach is the Structural Mapping Engine (SME) [13], which is a rule-based approach to finding analogies based on structural representations and attributes of target and source concepts. More recently, deep-learning based methods have also been developed to study such analogies [14, 15, 16].

Educational Data Mining '24 Human-Centric eXplainable AI in Education and Leveraging Large Language Models for Next-Generation Educational Technologies Workshop Joint Proceedings, July 13, 2024, Atlanta, GA

*Corresponding author.

✉ ssehgal4@illinois.edu (S. Sehgal); bhavya2@illinois.edu (Bhavya); kc62@illinois.edu (K. P. Datta); amalla5@illinois.edu (A. Mallavarapu); czhai@illinois.edu (C. Zhai)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

However, most of these works focused on word-level and proportional analogies, only recently have researchers studied generating explanations using deep learning and LLM approaches [10, 11]. Work by Bhavya et al [11] is closest to ours as they use pre-trained language models (PLMs) for complex analogy generation. However, they broadly study the applicability of LLMs for text analogy generation. Instead, our focus is on understanding the potential of the generated multimodal analogies for science education, a pivotal application domain.

2.1.2. Visual Analogies

Prior works treat visual analogies similar to word-level analogies (where A,B,C,D are images). These often include identifying the missing image portion represented by “?” in A:B::C:? [17, 18, 19] or stylistic and geometric transformations between images [19, 18]. Adding to the line of research in both textual and visual analogies, our work focuses on generating multimodal analogies, consisting of a textual explanation and a visual representation of the same. We do not relate images, but relate the target and source concepts, that are depicted in the image. Chakrabarty et al. (2023) created similar visual metaphors by elaborating on existing metaphors using LLMs. However, metaphors are much shorter in length and are more abstract than the long-form explanatory analogies we work with. Additionally, they employ existing metaphors in their study, whereas we generate both the textual analogy and the corresponding image from scratch given a target concept.

2.2. Applications of Analogies To Education

The application of analogies in education has been extensively explored across various disciplines (e.g., science, math, computer science), highlighting their significant role in enhancing learning and understanding concepts and language. Some studies [21, 22] have demonstrated how analogies can simplify complex concepts and foster problem-solving skills, particularly in science and mathematics. By linking new information to pre-existing knowledge, analogies facilitate deeper comprehension and retention [23, 24]. Vieira et al. (2022) showcase the innovative use of musical analogies to teach abstract scientific theories, thereby making challenging concepts more accessible to students. Collectively, these studies underscore the effectiveness of analogies as a powerful educational tool, capable of enhancing student engagement, understanding, and cognitive development [3, 4]. Considering the numerous advantages of employing analogies in education and the need for proactively generating analogies that aid learning in digital interactive environments, we explore the creation of scientific analogies using the combination of language and diffusion generative models.

2.3. LLMs for Education

Generative AI models (e.g., GPT [26], Claude¹, DALL-E [27], LLaMa [28]) with billions of parameters that have been trained on tremendous amounts of data have recently shown great promise on several tasks including text and image generation in multiple domains [29, 30]. The success of Large Language Models (LLMs) across diverse tasks has led researchers to explore their potential for education as well.

¹<https://claude.ai>

This includes improving teaching and learning capabilities using LLMs across several domains [31, 32], such as personalized learning [33], intelligent tutoring [34], adaptive assessment [35] and course content generation [36]. Moreover, LLMs can also be utilized to provide automated and personalized feedback to students [37]. Human-LM interaction is also being researched in the context of education, since students and teachers interact with these tutors and chatbots [38, 39]. Our study builds upon this rich body of work.

Educational resource and content creation has emerged as a key application area where LLMs have been harnessed [40, 41]. However, most work focuses on generating other kinds of content, such as educational questions [42], explanation, or assessment[37]. To the best of our knowledge, ours is the first work to explore multimodal analogy generation for educational purposes, using LLMs.

3. LLM-based multi-modal analogy generation

3.1. Analogy Generation Pipeline

In this section, we examine the potential of using LLMs and diffusion models for generating three types of multi-modal analogies (general, adaptive and iteratively refined).

For our exploration, we used the text labels in biology diagrams found on grade 6-12 educational websites², as target concepts. We seeded our search with biology diagrams since they contain visual representations of concepts and can be shown pictorially in image analogies. We collected 30 biology concepts this way.

To generate multi-modal (i.e., both text and image) analogies, we first use GPT-4³ to create a text-based analogy and feed the analogy into DALL-E-3⁴. Some of our exploration was done via the chat interface on their websites and the rest programmatically via API calls.

3.2. General Analogies

For use-cases where the intended learners are unknown or too broad, an educator might wish to generate general analogies that are broadly relevant to learners across grade-levels and backgrounds. To this end, we explored prompts like the following: “Generate a structural analogy for the biology concept <target> (part of <main_topic>).” with the GPT-4 model. We found analogies comparing the structure or the function (e.g., procedure of a science phenomenon) of the target and source concepts. For the DALL-E-3 model, we explored prompts like “Generate an image representing the scientific analogy given below.” along with the GPT-4 generated text analogy in the input.

Figure 2 and 3 present examples of the visual analogies and their corresponding texts. The authors of this paper (three graduate students with a background in Computer Science) qualitatively analyzed 30 multimodal analogies, comparing the text analogies alongside the images. We found the image analogies to be useful in providing an overview of the analogy as the text itself can sometimes be too verbose or complicated to understand. Moreover, the image analogies help visualize the analogical ideas detailed in the text.

²<https://byjus.com/biology/important-diagrams/>

³<https://openai.com/index/gpt-4-research/>

⁴<https://openai.com/index/dall-e-3/>

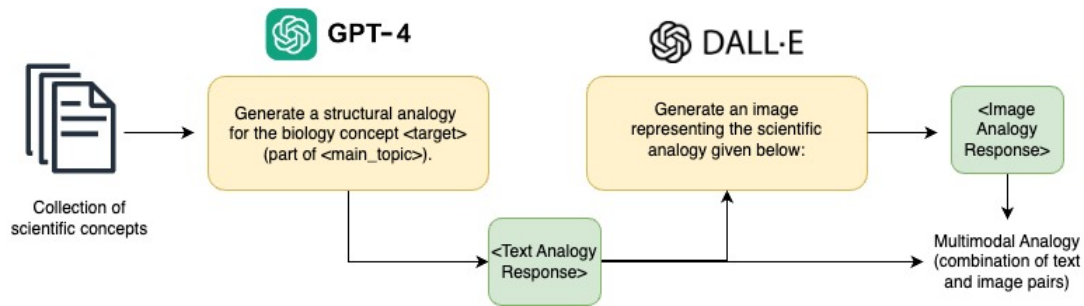


Figure 1: Pipeline for multi-modal analogy generation - first we prompt GPT-4 to obtain a text analogy that is fed into the DALL-E model to generate a corresponding image depiction.

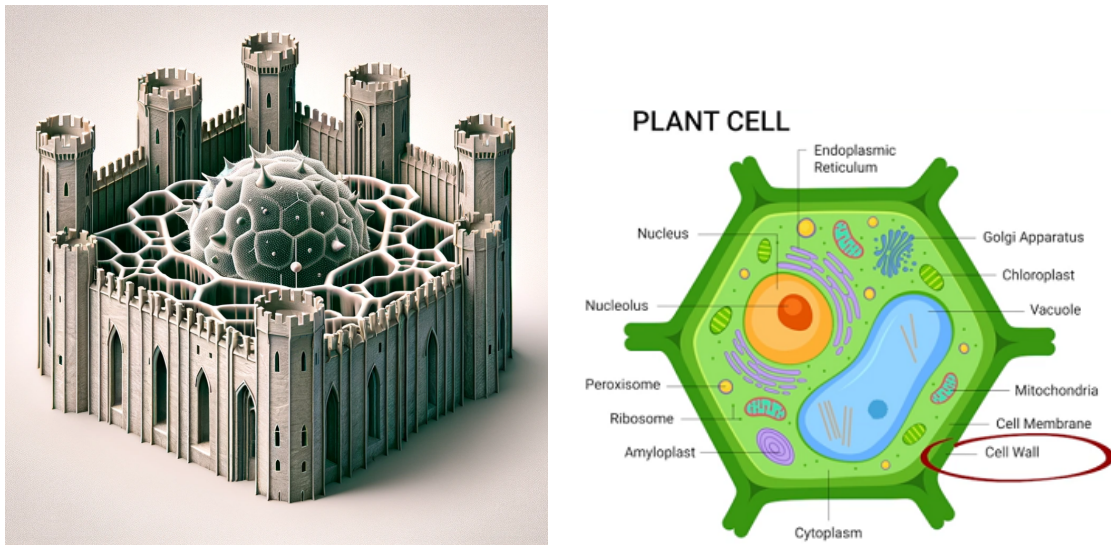


Figure 2: Visual Analogy and the actual figure for the scientific concept 'Cell Wall' - image generated by AI model DALL-E-3. Textual analogy associated generated by GPT-4: *"Imagine a plant cell as a medieval castle. The cell wall in this analogy can be compared to the high, thick stone walls that surround the castle. Just as the stone walls provide a rigid sturdy barrier protecting the castle from invaders and environmental elements, the cell wall surrounds the plant cell, giving it structure and protection. These walls are not just simple barriers but are complex, with layers and sections designed to withstand various pressures and attacks, much like the cell wall's intricate structure designed to support and protect the cell."*

This reinforces the utility of a visual representation alongside a text-based analogy to enhance its quality. We plan to release this manually validated dataset of 30 multimodal analogies of biology concepts for educators and researchers to use.

We also found some limitations with this approach. For example, the labels in images were often incoherent with the text not being in English. This is a known limitation of text-to-image models that are not good at rendering text in images [43]. Image-based analogies could benefit from better text labels as the images could then explicitly mention the similar representations between the two topics. Future work can look at how we can add text labels after image generation, as they could be useful for students to learn to draw structural diagrams. Another limitation that emerged was that the image analogies often represented multiple surrounding concepts as opposed to just the target concept (for example, the analogy for stomach also portrays other parts like esophagus and intestine). Thus, someone unfamiliar with the concept may not be able to discern which part the analogy is about. Finally, some images appeared ominous due to the nature of the target concept. For example, analo-

gies about the concept 'eye' depicted its various subparts and could appear eerie depending on the audience. Thus, we recommend that the images be sensitized by educators before presenting them to the students.

3.3. Adaptive Analogies

For more tailored use cases, one might wish to have analogies that are customized to learners' backgrounds, needs, and preferences (e.g., grade level, interest, cultural background.).

To this end, we explore prompting GPT-4 to create grade-level analogies. Figure 4 shows examples of how source concepts and analogies generated for the same scientific concept can be different based on the grade-level context. We found a difference in the relatability and complexity of the analogies based on the knowledge levels, thereby suggesting that it is possible to create more personalized and contextually appropriate analogies. We envision encoding the knowledge level for different grades in a language model, to generate these customized analogies. This can be done through providing knowledge of different subject and

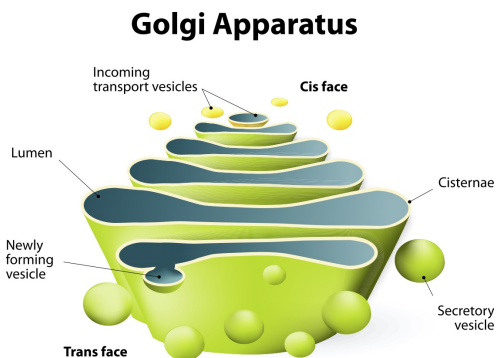
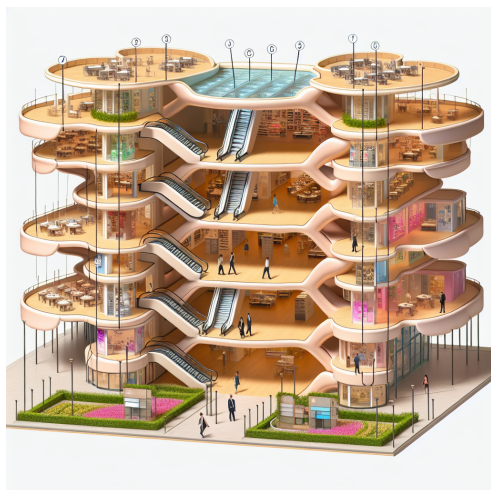
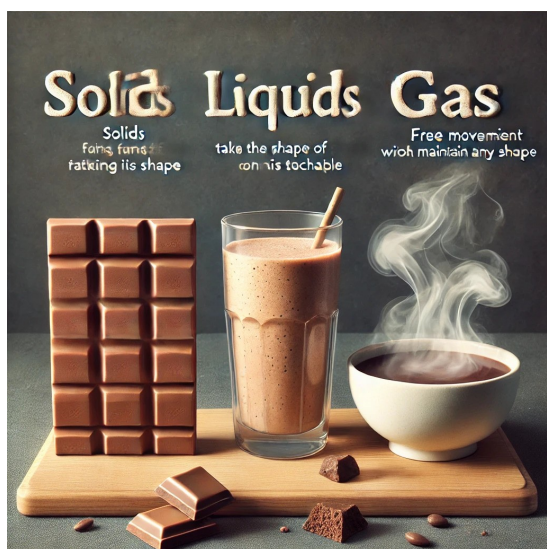
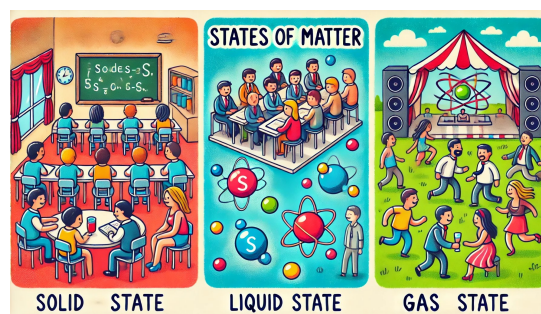


Figure 3: Visual Analogy and the actual figure for the scientific concept ‘Golgi Apparatus’. Visual analogy highlights the similar structure of a shopping mall - image generated by AI model DALL-E-3. Textual analogy associated generated by GPT-4: “Imagine a modern, multi-level shopping mall. This mall is not just a single large open space but is divided into several floors, each with its own set of shops and departments. The floors are connected by escalators and elevators, allowing for easy movement between them. The structure of the Golgi apparatus can be likened to this mall’s layout. The Golgi consists of a series of flattened, stacked pouches or cisternae, similar to how the mall has multiple levels. Just as each level of the mall has different stores, each cisterna within the Golgi is specialized for different functions or processing different molecules.”



(a) Analogy for the concept ‘States of Matter’ for Grade 3 students - image generated by AI model DALL-E-3: *Imagine matter as different types of snacks. Solids are like a bar of chocolate — firm and holding its shape. Liquids are like a smoothie — you can pour it and it takes the shape of its container, but it’s still touchable. Gases are like the steam from a hot bowl of soup — you can see it moving freely into the air, and it doesn’t keep its shape at all.*



(b) Analogy for the concept ‘States of Matter’ for Grade 12 students - image generated by AI model DALL-E-3: *Consider matter as if it were a crowd at different types of events. At a lecture, attendees sit close together, mostly stationary, like particles in a solid. At a networking event, people move around the room, mingling and shifting positions, similar to the movement of particles in a liquid. At a festival, attendees are spread out, moving freely around a large space, akin to particles in a gas that move independently and occupy any available space.*

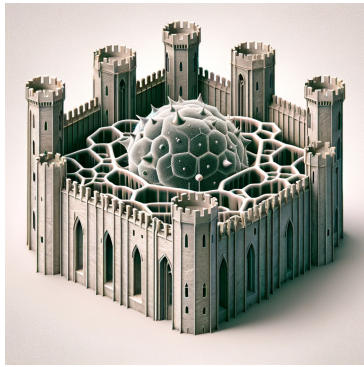
Figure 4: Grade-level multimodal analogies generated for the concept ‘States of Matter’. We see a difference in the type and complexity of source concepts and analogies generated - for example, analogy for grade 3 students employs the source concept of snacks, whereas the analogy for grade 12 students involves a comparison between classrooms, networking events, and concerts, which older students may be more aware of or have witnessed.

textbook chapters and online resources as context information in the prompt or through fine-tuning on grade-level information.

In addition to introducing adaptivity at the text level, future work could also investigate adaptivity at the image level. For example, image style (e.g., cartoon, abstract), color

palette, etc. could all be adapted to a particular learner. This could have important implications for accessible education. For example, images could be adjusted for low-vision or color-blind learners.

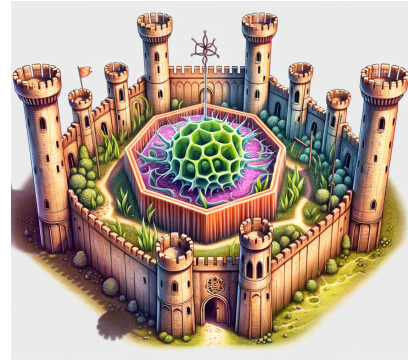
One important point to be mindful of is that adapting to certain learner traits (e.g., culture) could potentially lead to



(a) Initial image generated using the text analogy comparing Cell Wall and Castle Wall.



(b) Image generated based on image 5a and the prompt: 'Make image more colourful'.



(c) Image generated based on image 5b and the prompt 'Make castle walls more prominent'.

Figure 5: Example of iterative refinement of visual analogy through human feedback. Exact prompts for improvement are given in the sub-captions of the figure. We see that the model includes the suggested changes to improve image quality based on human feedback. All images generated by AI model DALL-E-3.

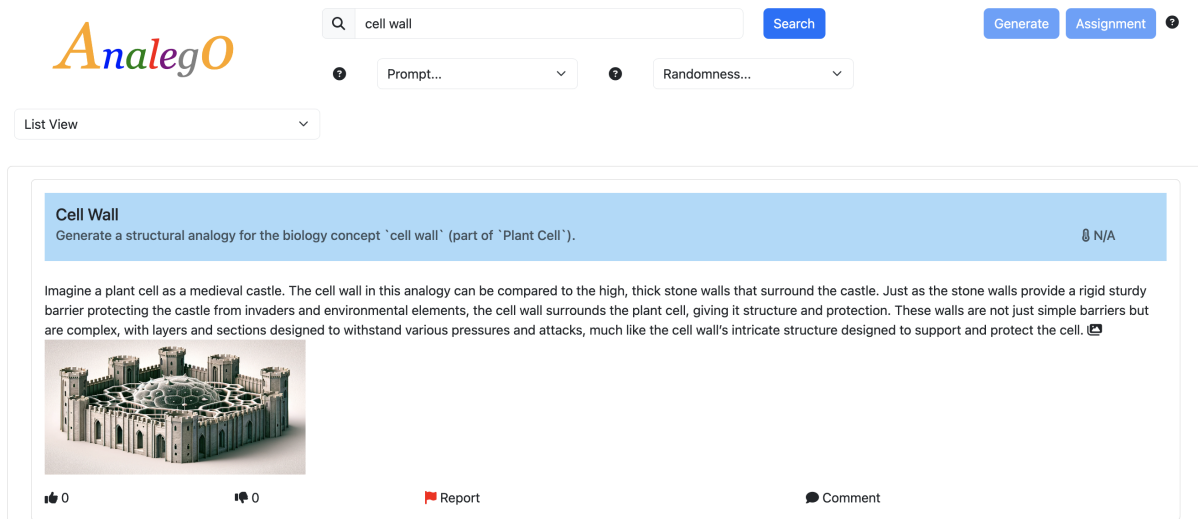


Figure 6: Demo system where students and educators can interact with multimodal analogies and provide feedback.

the generation of offensive or stereotypical analogies. Thus, practitioners must exercise caution while generating adaptive analogies and researchers should investigate methods to prevent offensive generation (e.g., safeguard models to detect such text and images).

3.4. Iteratively refined analogies

In the above two types of analogies, we've described a single round of generation. However, that might not always be sufficient to get the best or desired analogies. Naturally, we can think of an iterative approach to continually refine the generated analogies. To this end, we explored two ways of refinement: (1) human-AI collaboration where humans iteratively prompt the model to tweak the analogies, (2) multi-agent collaboration where multiple large image and language models (agents) iteratively generate and critique analogies for improvement.

3.4.1. Human-AI collaboration

We propose a human-feedback approach to improve the image analogy quality iteratively. Figure 5 showcases the example of the 'Cell Wall', where we prompt the text-to-image model DALL-E to iterate on the images based on our feedback. We find that the model adjusts the visual analogies based on human feedback, such as making the image more colourful and emphasizing different structural aspects. This shows promise for mining a large collection of multimodal analogies through human-AI collaboration. Specifically, we believe that through working closely with educators and students, we can iterate and improve the multimodal analogies, and generate a high-quality dataset tailored for educational purposes.

To realize this goal, we are currently developing a platform to enable efficient and large-scale human-AI collaboration. Figure 6 showcases a current version of the demo system where users (e.g., educators and students) can search for and provide feedback on previously generated analogies through the like, dislike, and comment features. Moreover, the system has a feature to report inappropriate or offensive

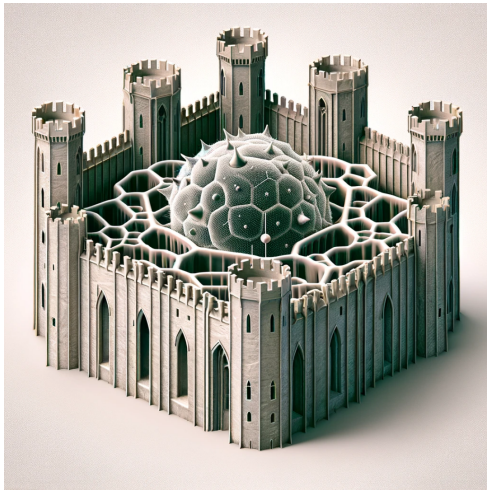


Figure 7: Example of iteration on visual through Multi-agent collaboration: The initial image analogy is provided to the Claude model to critique, the feedback of which is given back to DALL-E-3 to refine the image. Feedback given by Claude is as follows

“As a teacher, who is enhancing this analogy further, I would make a few adjustments:

- (1) Depict multiple layers: Plant cell walls consist of a primary and secondary layer. Adding a distinct inner and outer “wall” layer around the castle could represent this more accurately.*
- (2) Incorporate openings: Small gates or porous sections in the castle walls could symbolize how cell walls selectively allow the passage of certain molecules while blocking others.*
- (3) Highlight reinforcing elements: Various components like cellulose and pectin contribute to the strength of the cell wall. Incorporating reinforced sections or structural supports in the castle walls could represent this.*
- (4) Scale and perspective: Adjusting the scale and perspective to make the inner structure appear more microscopic in relation to the surrounding “walls” could better align with the actual size differences.*

By making these adjustments, the analogy becomes more representative of the layered, selectively permeable, and structurally reinforced nature of plant cell walls, while maintaining an engaging and memorable visual metaphor for students.”

DALL-E-3 then incorporates the feedback to include more multiple layers, openings, and adjusts the scale of the analogy. Thus, multi-agent collaboration can help improve the quality of the image analogy through iteration.

analogies that should not be shown in future. User feedback could then be integrated into the system via multiple ways, such as, refining the prompts and tuning the generation models via Reinforcement Learning Human Feedback (RLHF) so that they are better aligned with user needs[44]. The system could also be expanded to obtain finer-grained feedback (e.g., based on factuality, grade-level appropriateness, etc.) to enhance the quality of multimodal analogies with humans in the loop.

3.4.2. Multi-agent collaboration

In general, there could be several ways in which multiple agents collaborate together to generate high quality analogies. We explored one such approach, where we leverage the Claude3 Sonnet model⁵ to simulate a teacher and critique the GPT-4+DALL-E-3 generated analogy. The generated critique is then passed on to GPT-4 and the model is asked to refine the original analogy based on the critique. Figure 7 shows the example of multi-agent collaboration for the cell wall concept, where the Claude model provides feedback to improve the structural representation, such as adding more layers and openings in the wall, to make the analogy more scientifically accurate and understandable to students. The critique is fed to DALL-E-3 and it incorporates the suggested changes to update the image analogy.

In future, it would be interesting to explore how to optimize such a multi-agent collaboration to improve the quality of generated analogies with no or minimal human interac-

tion. For example, model-generated critique could be shown to teachers as starting points for improving the analogy. Another possibility could be to share model-generated critique with AI researchers and system engineers to distill common model failures and develop guidelines for future users generating analogies.

4. Discussion and Conclusion

We introduce the theme of multimodal analogies for science education, consisting of text and image-based analogies. We explore how to generate three types of multimodal analogies leveraging GPT-4 for textual analogy generation and feeding that into DALL-E-3 to create visual analogies. Our qualitative validation and exploration suggests that the generated image analogies successfully contain and represent text-based analogies in most cases. Furthermore, we show how they can be adapted to learners at different grade levels and can even be refined iteratively via human or multi-agent collaboration.

As next steps, we must work closely with students and educators to assess and improve the quality of multimodal analogies for science education. We showcased our demo system for displaying multimodal analogies to student learners and educators through which we hope to gather feedback. While automatically generated analogies are helpful as a starting point, incorporating human-AI collaboration and crowdsourcing with educators and practitioners can provide valuable feedback and adjustments. Such collaboration can enhance the system’s utility and ensure its analogies

⁵<https://claude.ai/>

are valid and appropriate for students and resonate with different contexts and cultures.

We have identified several interesting research challenges that still need to be solved (e.g., how to generate legible labels in images, how to mitigate generation of potentially offensive images, how to effectively support multi-agent and human-AI collaboration). Another important future work we hope to use the demo system for is to study the impact of the generated analogies on science learning amongst students.

Overall, we believe our work highlights a unique application of AI for generating multi-modal content and resources for science education. Multi-modal content [45, 46] is known to help with engaging learners and our work suggests that LLMs and diffusion models have a great potential in generating such content. Thus, our approach and findings could be widely useful to generate other kinds of multimodal, scientific and educational content (e.g., stories and dialogues), in addition to analogies, to enable more engaging learning environments.

5. Acknowledgements

This work is supported by the National Science Foundation and the Institute of Education Sciences, U.S. Department of Education, through Award # 2229612 (National AI Institute for Inclusive Intelligent Technologies for Education).

References

- [1] M. L. Minsky, *The Society of Mind*, Simon & Schuster, New York, 1988.
- [2] d. hofstadter, E. Sander, *Surfaces and Essence: : Analogy as the Fuel and Fire of Thinking*, 2013.
- [3] D. Treagust, A. Harrison, G. Venville, Teaching science effectively with analogies: An approach for pre-service and inservice teacher education, *Journal of Science Teacher Education* 9 (1998). doi:10.1023/A:1009423030880.
- [4] L. E. Richland, N. Simms, Analogy, higher order thinking, and education, *Wiley Interdisciplinary Reviews: Cognitive Science* 6 (2015) 177–192. URL: <https://doi.org/10.1002/wcs.1336>. doi:10.1002/wcs.1336.
- [5] M. B. Goldwater, D. Gentner, N. D. LaDue, J. C. Libarkin, Analogy generation in science experts and novices, *Cognitive Science* 45 (2021) e13036. doi:10.1111/cogs.13036.
- [6] D. R. Kretz, D. C. Krawczyk, Expert analogy use in a naturalistic setting, *Frontiers in Psychology* 5 (2014) 1333. doi:10.3389/fpsyg.2014.01333.
- [7] D. Jurgens, S. Mohammad, P. Turney, K. Holyoak, SemEval-2012 task 2: Measuring degrees of relational similarity, in: E. Agirre, J. Bos, M. Diab, S. Manandhar, Y. Marton, D. Yuret (Eds.), *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), Association for Computational Linguistics, Montréal, Canada, 2012, pp. 356–364. URL: <https://aclanthology.org/S12-1047>.
- [8] V. Popov, P. Hristova, R. Anders, The relational luring effect: Retrieval of relational information during associative recognition, *Journal of Experimental Psychology: General* 146 (2017) 722–745. URL: <https://api.semanticscholar.org/CorpusID:20177507>.
- [9] M. J. Kmieciak, R. J. Brisson, R. G. Morrison, The time course of semantic and relational processing during verbal analogical reasoning, *Brain and Cognition* 129 (2019) 25–34. URL: <https://api.semanticscholar.org/CorpusID:54169303>.
- [10] J. Chen, R. Xu, Z. Fu, W. Shi, Z. Li, X. Zhang, C. Sun, L. Li, Y. Xiao, H. Zhou, E-kar: A benchmark for rationalizing natural language analogical reasoning, in: *Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, 2022*. URL: <http://dx.doi.org/10.18653/v1/2022.findings-acl.311>. doi:10.18653/v1/2022.findings-acl.311.
- [11] B. Bhavya, J. Xiong, C. Zhai, Analogy generation by prompting large language models: A case study of InstructGPT, in: S. Shaikh, T. Ferreira, A. Stent (Eds.), *Proceedings of the 15th International Conference on Natural Language Generation, Association for Computational Linguistics, Waterville, Maine, USA and virtual meeting, 2022*, pp. 298–312. URL: <https://aclanthology.org/2022.inlg-main.25>. doi:10.18653/v1/2022.inlg-main.25.
- [12] E. Bobek, B. Tversky, Creating visual explanations improves learning, *Cognitive Research: Principles and Implications* 1 (2016). doi:10.1186/s41235-016-0031-6.
- [13] K. D. Forbus, R. W. Ferguson, A. M. Lovett, D. Gentner, Extending sme to handle large-scale cognitive modeling, *Cognitive science* 41 5 (2017) 1152–1201. URL: <https://api.semanticscholar.org/CorpusID:4572276>.
- [14] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, 2013. *arXiv:1301.3781*.
- [15] G. Rossiello, A. Gliozzo, R. Farrell, N. Fauceglia, M. Glass, Learning relational representations by analogy using hierarchical Siamese networks, in: J. Burstein, C. Doran, T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 3235–3245. URL: <https://aclanthology.org/N19-1327>. doi:10.18653/v1/N19-1327.
- [16] A. Ushio, L. Espinosa Anke, S. Schockaert, J. Camacho-Collados, BERT is to NLP what AlexNet is to CV: Can pre-trained language models identify analogies?, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Online, 2021, pp. 3609–3624. URL: <https://aclanthology.org/2021.acl-long.280>. doi:10.18653/v1/2021.acl-long.280.
- [17] S. E. Reed, Y. Zhang, Y. Zhang, H. Lee, Deep visual analogy-making, in: *Neural Information Processing Systems, 2015*. URL: <https://api.semanticscholar.org/CorpusID:1836951>.
- [18] Y. Bitton, R. Yosef, E. Strugo, D. Shahaf, R. Schwartz, G. Stanovsky, VASR: Visual analogies of situation recognition, in: *AAAI Conference on Artificial Intelligence, 2022*. URL: <https://api.semanticscholar.org/CorpusID:>

- 254536039.
- [19] Y. Tewel, Y. Shalev, I. Schwartz, L. Wolf, Zerocap: Zero-shot image-to-text generation for visual-semantic arithmetic, 2022. arXiv:2111.14447.
- [20] T. Chakrabarty, A. Saakyan, O. Winn, A. Panagopoulou, Y. Yang, M. Apidianaki, S. Muresan, I spy a metaphor: Large language models and diffusion models co-create visual metaphors, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 7370–7388. URL: <https://aclanthology.org/2023.findings-acl.465>. doi:10.18653/v1/2023.findings-acl.465.
- [21] P. Thagard, Analogy, explanation, and education, *Journal of Research in Science Teaching* 29 (1992) 537–544. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.3660290603>. doi:<https://doi.org/10.1002/tea.3660290603>. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.3660290603>.
- [22] L. Novick, K. Holyoak, Mathematical problem solving by analogy, *Journal of experimental psychology. Learning, memory, and cognition* 17 (1991) 398–415. doi:10.1037/0278-7393.17.3.398.
- [23] S. M. Glynn, B. K. Britton, M. Semrud-Clikeman, K. D. Muth, Analogical Reasoning and Problem Solving in Science Textbooks, Springer US, Boston, MA, 1989, pp. 383–398. URL: https://doi.org/10.1007/978-1-4757-5356-1_21. doi:10.1007/978-1-4757-5356-1_21.
- [24] R. Duit, The role of analogies and metaphors in learning science, *Science Education* 75 (1991) 649 – 672. doi:10.1002/sce.3730750606.
- [25] H. Vieira, C. Morais, Musical analogies to teach middle school students topics of the quantum model of the atom, *Journal of Chemical Education* 99 (2022). doi:10.1021/acs.jchemed.2c00289.
- [26] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, *Advances in neural information processing systems* 33 (2020) 1877–1901.
- [27] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, Zero-shot text-to-image generation, in: International conference on machine learning, Pmlr, 2021, pp. 8821–8831.
- [28] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).
- [29] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, B. Yin, X. Hu, Harnessing the power of llms in practice: A survey on chatgpt and beyond, *ACM Transactions on Knowledge Discovery from Data* 18 (2024) 1–32.
- [30] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, et al., A survey on large language models: Applications, challenges, limitations, and practical usage, *Authorea Preprints* (2023).
- [31] H. Lin, S. Wan, W. Gan, J. Chen, H.-C. Chao, Metaverse in education: Vision, opportunities, and challenges, 2022. arXiv:2211.14951.
- [32] P. Denny, S. Gulwani, N. T. Heffernan, T. Käser, S. Moore, A. N. Rafferty, A. Singla, Generative ai for education (gaid): Advances, opportunities, and challenges, 2024. arXiv:2402.01580.
- [33] T. Alqahtani, H. Badreldin, M. Alrashed, A. Alshaya, S. Alghamdi, K. Saleh, S. Alowais, O. Alshaya, I. Rahman, M. Al Yami, A. Albekairy, The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research, *Research in Social and Administrative Pharmacy* 19 (2023). doi:10.1016/j.sapharm.2023.05.016.
- [34] S. P. Chowdhury, V. Zouhar, M. Sachan, Auto-tutor meets large language models: A language model tutor with rich pedagogy and guardrails, 2024. arXiv:2402.09216.
- [35] M. Javaid, A. Haleem, R. Singh, S. Khan, I. Haleem Khan, Unlocking the opportunities through chatgpt tool towards ameliorating the education system, *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3 (2023) 100115. doi:10.1016/j.tbench.2023.100115.
- [36] D. Leiker, S. Finnigan, A. R. Gyllen, M. Cukurova, Prototyping the use of large language models (llms) for adult learning content creation at scale, 2023. arXiv:2306.01815.
- [37] J. Meyer, T. Jansen, R. Schiller, L. W. Liebenow, M. Steinbach, A. Horbach, J. Fleckenstein, Using llms to bring evidence-based feedback into the classroom: Ai-generated feedback increases secondary students’ text revision, motivation, and positive emotions, *Computers and Education: Artificial Intelligence* 6 (2024) 100199. URL: <https://www.sciencedirect.com/science/article/pii/S2666920X23000784>. doi:<https://doi.org/10.1016/j.caeai.2023.100199>.
- [38] M. Lee, M. Srivastava, A. Hardy, J. Thickstun, E. Durmus, A. Paranjape, I. Gerard-Ursin, X. L. Li, F. Ladhak, F. Rong, R. E. Wang, M. Kwon, J. S. Park, H. Cao, T. Lee, R. Bommasani, M. Bernstein, P. Liang, Evaluating human-language model interaction, 2024. arXiv:2212.09746.
- [39] J. Jeon, S. Lee, Large language models in education: A focus on the complementary relationship between human teachers and chatgpt, *Education and Information Technologies* 28 (2023) 15873–15892. URL: <https://doi.org/10.1007/s10639-023-11834-1>. doi:10.1007/s10639-023-11834-1.
- [40] W. Gan, Z. Qi, J. Wu, J. C.-W. Lin, Large language models in education: Vision and opportunities, 2023. arXiv:2311.13160.
- [41] S. Moore, R. Tong, A. Singh, Z. Liu, X. Hu, Y. Lu, J. Liang, C. Cao, H. Khosravi, P. Denny, C. Brooks, J. Stamper, Empowering education with llms: The next-gen interface and content generation, in: *Communications in Computer and Information Science*, volume 1831, Springer, 2023, pp. 32–37. doi:10.1007/978-3-031-36336-8_4.
- [42] Z. Wang, J. Valdez, D. B. Mallick, R. Baraniuk, Towards human-like educational question generation with large language models, in: *International Conference on Artificial Intelligence in Education*, 2022. URL: <https://api.semanticscholar.org/CorpusID:251137828>.
- [43] J. Betker, G. Goh, L. Jing, TimBrooks, J. Wang, L. Li, LongOuyang, JuntangZhuang, JoyceLee, YufeiGuo, WesamManassra, PrafullaDhariwal, CaseyChu, YunxinJiao, A. Ramesh, Improving image generation with

better captions, ??? URL: <https://api.semanticscholar.org/CorpusID:264403242>.

- [44] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, *Advances in neural information processing systems* 35 (2022) 27730–27744.
- [45] M. Sankey, D. Birch, M. W. Gardiner, Engaging students through multimodal learning environments: The journey continues, *Proceedings of the 27th Australasian Society for Computers in Learning in Tertiary Education* (2010) 852–863.
- [46] B. Bouchey, J. Castek, J. Thygeson, Multimodal learning, *Innovative Learning Environments in STEM Higher Education: Opportunities, Challenges, and Looking Forward* (2021) 35–54.