Using Generative AI to Enhance the Visuals of Papers

Niklas Pfützenreuter
University of Duisburg-Essen
Essen, Germany
niklas.pfuetzenreuter@uni-due.de

Abstract

Generative Artificial Intelligence (GenAI) provides new capabilities to generate figures and other visual elements for papers. This can help researchers to enhance the visual quality of their papers. However, there are also challenges when generating images or visual elements, such as the complexity of describing visual content with a text prompt. This paper describes how GenAI can be effectively used in researchers' workflows. It further discusses solutions for existing challenges in generating images from textual descriptions, like using tools like ControlNet and IP Adapter or specifically trained image generation model checkpoints. Lastly, the paper shows how GenAI can be used to generate LateX code for visual elements like tables or diagrams.

CCS Concepts

• **Human-centered computing** → *HCI theory, concepts and models.*

Keywords

Human-AI Interaction, Generative AI, Image Generation

1 Introduction

For many researchers, the integration of Generative Artificial Intelligence (GenAI) has become increasingly important in enhancing the presentation of their findings. GenAI can generate images from textual descriptions, thereby offering a new way to create figures for papers [2, 6]. By transforming textual descriptions into visual representations, GenAI aids researchers in the ideation process, allowing for more dynamic and illustrative content creation. This capability enables researchers without image manipulation experience to create their figures for their papers.

One advantage of using Artificial Intelligence (AI)-generated images is the elimination of copyright concerns. As of now, the users of GenAI models own the generated content and may use the content for non-commercial purposes [5, 8]. This ensures that researchers can enhance their work with illustrations without the risk of legal infringements. However, regional laws may apply which restrict the use of AI-generated content. This applies in particular to generated content which is very similar to existing copyrighted content [1].

Mensch und Computer 2024 – Workshopband, Gesellschaft für Informatik e.V., 01.-04. September 2024, Karlsruhe, Germany

https://doi.org/10.18420/muc2024-mci-ws09-169

Despite these advantages, achieving the desired outcome with GenAI image generation can be challenging. Often, multiple iterations are required, and fine-tuning to produce visuals that precisely match the intended concept is challenging. This iterative process can be time-consuming and may require knowledge about correctly prompt image generation networks [9].

Beyond image generation, GenAI offers additional utilities to enhance the visuals of papers, for example, by generating LateX code for diagrams of tables. These visuals are essential for clearly presenting results and complex data. Furthermore, it can be employed to design structured elements like custom boxes, e.g., for hypotheses, enhancing the organization and readability of research papers [7].

In this paper, we describe how GenAI can be used to generate figures and LateX code for visuals and what challenges exist. Further, we name lessons learned using GenAI in this context to show how to overcome existing challenges.

2 Generative AI for Figure Generation

When generating figures for papers with GenAI, we discovered two different strategies.

Generating Parts of Figures. Researchers sometimes find themselves without access to copyright-free stock material that accurately represents their concepts. Particularly for very specific scenarios, standard stock images fall short. With GenAI, researchers only need a textual description of the desired image. With that, there is no need for stock image material or capturing the images with expensive camera equipment. Further, only parts of complete figures can be generated, which can then be integrated into a larger figure.

One way to control the generation of the images instead of just using a text prompt is to use a reference image. By providing these references, the AI model can, e.g., extract the edges from the reference image, producing visuals that closely match the intended concept. This approach also makes it easier for researchers to describe their ideas, ensuring higher relevance of the generated images [4].

Once the individual parts are generated, they can be combined and edited using image manipulation software. This method enables the researchers to create more complex figures and stay in control of the generation.

Generating Complete Figures. For simpler figures, GenAI can also create the entire figure without needing to edit it in an image manipulation program. This generation process is more complex and requires a more detailed description to achieve satisfactory results.

To generate complete figures with GenAI, it is beneficial to create an initial sketch or outline of how the image should look. This preliminary sketch can be used in the same way as reference images, ensuring that the final output aligns closely with the researchers' vision.

One of the advantages of using GenAI for complete figures is that no image manipulation skills are needed. With that, researchers can focus on the content and clarity of their visuals, and the AI can generate the desired figure for them.

3 Generative AI for Paper Formatting

Apart from generating figures, GenAI has the potential to support the creation of LaTeX visuals, such as tables and other visual elements, in papers [3].

Researchers can input a description of a table, including column headers and specific formatting requirements, or also just provide an example table. The GenAI model can then produce the LaTeX code necessary to render this table in the paper. This is particularly useful for creating complex tables with multiple rows and columns, which would otherwise require complex manual coding.

Beyond tables, LLMs can assist in generating LaTeX code for a variety of visual elements, such as diagrams or charts. This can be done instead of using image generation to create diagrams or charts, as these models have problems generating scientific diagrams or charts.

Further, GenAI can generate custom visual elements like boxes for hypotheses for LateX. This often requires an iterative refinement to match the researchers' imagination.

However, some templates are limited in the availability of LateX packages. To prevent compatibility issues and ensure that the generated LaTeX code functions correctly, it is crucial to specify the allowed LaTeX packages.

4 Lessons Learned

Throughout the process of integrating GenAI to enhance visuals in papers, several key lessons have emerged that can inform and improve future efforts.

One of the primary challenges encountered was controlling the output when generating images solely from textual descriptions. In general, it is difficult to describe their visual imagination with a prompt. Thus, the generated images often do not match the researchers' imaginations. This difficulty underscores the importance of using additional techniques to refine and direct the image generation process.

To address the challenges of text-based image generation, tools like ControlNet or IP Adapters can be a good solution. These tools allow for greater control over the AI's output by incorporating initial sketches or guiding images. Depending on which features of the reference image are important, these tools can extract the desired features. ControlNet, for example, can extract edges, lines, human poses, or segments of images. The IP Adapters can extract the styles and colors of reference images and transfer them to the generated image. By providing a visual input, these tools help control the output of image generation models. Further, they allow the researchers to use a different modality to describe their imagination instead of just using words.

Another point that can enhance the results of image generation is the use of specifically trained checkpoints of Stable Diffusion. Depending on the specific requirements of a figure, these specialized models can produce more relevant and higher-quality outputs.

Lastly, using a combination of different GenAI models can be beneficial. For example, a CLIP model can be used to generate a description for an existing image that can be used to extract keywords for a new image generation prompt. This can help users to create a prompt to describe their imagination.

5 Conclusion

To summarize, GenAI offers several capabilities for enhancing the visual elements of papers. One capability is creating figures for papers by generating complete figures or parts of figures that are put together in an image manipulation program. By utilizing tools like ControlNet and IP Adapter, researchers can overcome the challenge of creating visuals that match their imagination. Additionally, leveraging specifically trained checkpoints of image generation models can help generate images for specific use cases. For LaTeX visuals, GenAI facilitates the generation of complex tables, diagrams, or custom visuals. However, it is necessary to specify the LateX packages that are allowed. As these technologies continue to evolve, their integration into academic workflows can further enrich the quality and impact of research publications.

References

[1] BYU. 2024. GenAI and Copyright. https://genai.byu.edu/copyright-page
[2] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. 2018. Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine* 35, 1 (Jan. 2018), 53–65. https://doi.org/10.1109/MSP.2017.2765202 Conference Name: IEEE Signal Processing Magazine.
[3] Anton Korinek. 2023. Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature* 61, 4 (2023), 1281–1317.
[4] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. 2024. The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3613904.3642218
[5] OpenAI. 2024. EU terms of use. https://openai.com/policies/terms-of-use/
[6] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New Orleans, LA, USA, 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042
[7] Galit Shmueli, Bianca Maria Colosimo, David Martens, Rema Padman, Maytal Saar-Tsechansky, Olivia R. Liu Sheng, W Nick Street, and Kwok-Leung Tsui. 2023. How can IJDS authors, reviewers, and editors use (and misuse) generative AI? *INFORMS Journal on Data Science* 2, 1 (2023), 1–9.
[8] StabilityAI. 2024. Professional Membership Agreement. https://stability.ai/community-license-agreement
[9] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543 [cs.CV]