



HAL
open science

Draw Me Like Your Triples: Leveraging Generative AI for the Completion of Wikidata - (Working paper)

Raia Abu, Martin Critelli, Şefika Efeoğlu, Eleonora Mancini, Célian Ringwald,
Xinyue Zhang

► To cite this version:

Raia Abu, Martin Critelli, Şefika Efeoğlu, Eleonora Mancini, Célian Ringwald, et al.. Draw Me Like Your Triples: Leveraging Generative AI for the Completion of Wikidata - (Working paper). 2023. hal-04526070

HAL Id: hal-04526070

<https://hal.science/hal-04526070v1>

Preprint submitted on 29 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Draw Me Like Your Triples: Leveraging Generative AI for the Completion of Wikidata^{*}

Raia Abu Ahmad¹[0009-0004-8720-0116], Martin Critelli²[0000-0002-8177-730X], Şefika Efeoğlu^{3,7}[0000-0002-9232-4840], Eleonora Mancini⁴[0000-0001-9205-3289], Célian Ringwald⁵[0000-0002-9232-4840], and Xinyue Zhang⁶[0000-0002-9232-4840]

¹ Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH), Germany

² University CaFoscari of Venice, Italy

³ Freie Universität Berlin, Germany

⁴ University of Bologna, Italy

⁵ Université Côte d'Azur, Inria, CNRS, I3S, France

⁶ University of Oxford, Oxford, United Kingdom

⁷ Technische Universität Berlin, Germany

Abstract. Humans are critical for the creation and maintenance of high-quality Knowledge Graphs (KGs). However, creating and maintaining large KGs only with humans does not scale, which is especially true for contributions based on multimedia (e.g. images) that are hard to find on the Web and expensive to generate by humans. Therefore, we leverage creative AI for the task of generating images for Wikidata items that do not have them. Our approach uses triples contained in Wikidata and generates natural text descriptions of items about fictional characters with missing images in multiple ways, including verbalizing them through large language models (LLMs). Then, we use those natural text descriptions as prompts for text-to-image models to generate plausible candidate images for Wikidata KG image completion. We design and implement several evaluations to verify the plausibility of our methods.

Keywords: Knowledge Graph Completion · Creative AI · Image Generation.

1 Research Questions

Our report addresses the following research questions (RQs):

- RQ1: To what extent can the output of generative AI be used for knowledge graph completion?
- RQ2: To what extent do prompts increasingly similar to natural language can be used in text-to image models to produce higher quality images?
- RQ3: To what extent do prompts increasingly similar to natural language can be used in text-to image models to preserve the emotion and sentiment between prompts and generated images?

^{*} Technical report of the task force Gryffindor from ISWS 2023 led by Albert Meroño

2 Semantic Web for Creative AI

Our definition of *Semantic Web for creative AI* is circular. On the one hand, we think that Semantic Web technologies can be used for knowledge-aware prompt engineering to be used as input for generative AI. And on the other hand, generative AI can be a powerful tool for multimodal KG construction and completion. Our approach demonstrates this full circle by using triples of Wikidata items as prompts for a text-to-image model, and then using the creative AI output to complete missing images for the same items.

3 Introduction

Recent advances in artificial intelligence (AI) have given rise to automatic systems that can generate images from textual descriptions [30]. Thus, the field of *creative AI* has witnessed an increasing number of use cases and applications. Consequently, researchers have been looking into methodologies for enhancing the textual prompts given to creative AI models, in order to exploit its inner features and receive the best results [16].

Semantic Web technologies have been incorporating AI into many of their applications and tasks. Some examples include ontology creation and enhancement, knowledge graph (KG) embeddings, and KG completion (KGC) [8]. In this report, we will focus on the task of KGC, i.e. completing missing information in KGs, by leveraging creative AI applications and using different prompt engineering approaches.

To exemplify our idea, we focus on one aspect of KGC in Wikidata [25]: image completion of fictional characters. Querying Wikidata shows that only 7% out of the 83.7K instances of the fictional character class (including its sub-classes) have an image. To resolve this, we propose the following method: 1) Extracting knowledge about fictional character entities in Wikidata in the form of triples. 2) Creating different types of prompts to portray those triples. And 3) Using a text-to-image model to create images portraying the fictional characters.

We then present two strategies for evaluating our approach. The first consists of constructing a ground-truth dataset of fictional character entities that do have an image on Wikidata, feeding those to our pipeline, and comparing the resulting images to the ground-truth ones. The second approach compares the sentiment and emotion of each prompt to that of its corresponding image. This is based on our intuition that having similar affective features of the prompt describing the character and the generated image depicting it indicates that generative AI can indeed be used for the purpose of KGC, and more specifically for image completion of characters.

This project contributes to the ever-growing area of research that leverages AI technologies for KGs. To the best of our knowledge, no previous study has explored the realm of using generative text-to-image AI for KGC.

4 Related Work

Generative AI: Current groundbreaking advances in AI enable machines to generate novel and original content based on textual prompts. Such generative applications include text-to-text [9], text-to-image, and even text-to-music. Generally, these models can capture complex patterns from the input text and produce coherent outputs. A recent survey [30] shows that text-to-image applications specifically have been emerging since 2015, when AlignDRAW [13] pioneered the field by leveraging recurrent neural networks (RNNs) to encode textual captions and produce corresponding images. Since then, end-to-end models started leveraging architectures such as deep convolutional generative adversarial networks (GANs) [18, 31, 29], autoregressive methods [17, 6, 27], and the current state-of-the-art diffusion-based methods [14, 20, 19].

Prompt Engineering: Because of the afore-mentioned advances, a novel area of *prompt-based engineering* has emerged, in which humans interact with AI in an iterative process to produce the best prompt (i.e. textual input) for a specific desired output [12]. Recent work has shed light on prompt engineering for AI art specifically [16], concluding that simple and intuitive prompts written by humans are not enough to get desired results. Rather, writing good prompts is a learned skill that is enhanced by the usage of specific prompt templates and modifiers [15]. Naturally, prompts themselves can also be automatically generated and do not necessarily require human manual writing. A novel research approach introduced mechanisms to enrich and refine prompts in order to create images with pre-defined emotion expressions [26]. This was done by extracting text features and applying strategies such as adding emotion words, context, paraphrasing, and increasing concreteness. However, to the best of our knowledge, no work has been done yet on leveraging Knowledge Graphs (KGs) for prompt engineering.

Knowledge Graph Completion: KGs are essential knowledge structures that have been increasingly used to solve many knowledge-aware tasks [8]. A large number of KGs have been constructed and are being used and edited by users such as DBpedia [11], Freebase [3], YAGO [23], and Wikidata [25]. However, a major problem of KGs remains their incompleteness, as many relations and entities are still missing [8]. As such, the task of KG completion (KGC) introduces many algorithms that can be leveraged for predicting missing parts of triples based on existing ones [22]. Researchers have leveraged some generative AI models to solve this task, however, existing studies use text-to-text generation, e.g., converting KGC to a sequence-to-sequence generation task using a pre-trained language model [28]. Another similar work tackled this problem with a similar solution by verbalizing graph structures and using flat text and using it as input for a language model [5]. In contrast to previous studies, our work proposes contributing to KGC by leveraging text-to-image generative models for missing images in Wikidata, using human fictional characters as a use-case.

5 Resources

All the code required to reproduce our results, including the evaluation framework, is openly available and can be accessed at <https://github.com/helemanc/gryffindor>.

Dataset Creation: We firstly began by designing a Wikidata SPARQL queries, that allows us to obtain all the triples related to the fictional characters⁸. In regards of the limited time we have for conducting our study we sampled a dataset of 100 randomly selected entities that have images. This first sample allowed us to have a first intuition and a first basis for conducting our analysis. As second step we computed some statistics about these triples : 100 different predicates are invested into our subsets and each entity are described by 9 different predicates. The top 10 predicates used are the following one : ‘instance of’ (99), ‘sex or gender’ (75), ‘present in work’ (71), ‘occupation’ (42), ‘given name’ (39), ‘creator’ (36), ‘country of citizenship’ (32), ‘performer’ (32), ‘from narrative universe’ (23), ‘languages spoken, written or signed’ (21) and ‘spouse’ (19). Please note that the predicate *altLabel* is discarded since only English is considered in our initial evaluation.

Prompt Generation: We designed 4 types of prompts, based on the data extracted from the knowledge graph, please refer to section 6.2 for more details. Among them, the verbalised triples are generated by the WDV verbalisation model⁹ proposed by Amaral *et al.* [1].

Image Generation: For the image generation task, we proceeded to apply DALL·E 2¹⁰ and the free version of Craiyon v.3¹¹. We decided to use two different AI models to compare them considering three main aspects: open-source availability, computational power, and output modalities. DALL·E 2 has been released in January 2022 and constitutes an improvement of DALL·E in text comprehension and image resolution. In addition to these elements, the second version allows users to load and modify their own images. In this work, it was applied via the official API that makes it possible to quickly generate one image per prompt. Craiyon is formerly a web-based mini DALL·E 2 created by Craiyon LLC. able to generate a group of nine images from the same prompt according to different stylistic effects. In addition to this, this model offers users the possibility to exclude certain concepts from the image generation process using the “negative words” slot. For the image generation task, these two models are used and the results obtained analyzed and compared.

Evaluation Framework: To assess how the affective content of the prompt influences the representation of affective content in the generated image and determine whether the character from the original image remains recognizable, we employed two distinct models: Emotion English DistilRoBERTa-base¹² and

⁸ The package *quikidata* (<https://pypi.org/project/quikidata/>) is used for triple retrieval.

⁹ <https://github.com/gabrielmaia7/wdv>

¹⁰ <https://openai.com/dall-e-2>

¹¹ <https://www.craiyon.com/>

¹² <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base>

DeepFace[21]. For inferring emotions from the four types of prompts, we utilized Emotion English DistilRoBERTa-base. This model, a distilled version of RoBERTa, was trained on a diverse range of datasets containing emotion-labeled texts from platforms like Twitter, Reddit, student self-reports, and TV dialogues. It focuses on predicting the six emotions defined in Ekman’s emotion model[7] (*anger, disgust, fear, joy, neutral, sadness*) along with *surprise*. We selected this model due to its alignment with the emotions considered in DeepFace and its demonstrated high performance. DeepFace is an agile Python framework designed for face recognition and comprehensive analysis of facial attributes. It seamlessly integrates several cutting-edge models, including VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID, ArcFace, Dlib, and SFace. In our study, we specifically harnessed the power of VGG-Face within the DeepFace framework. This allowed us to extract a wealth of valuable information from both the original and generated images, encompassing attributes such as age, gender, emotion, and race. Among this diverse array of attributes, we focused on leveraging the emotion feature for our research purposes.

6 Proposed approach

This section outlines a method for image generation corresponding to Wikidata entities currently lacking visual representation. Our primary focus for this report is the class designated as *fictional character*. The initial stage of our approach involves extraction of relevant triples pertaining to a specific entity. Subsequently, these triples are used to form various types of prompts, functioning as inputs to a sophisticated text-to-image artificial intelligence model. The ultimate goal is to generate suitable images that can serve as accurate visual identifiers for their corresponding Wikidata entities. This pipeline is shown in Fig. 1.

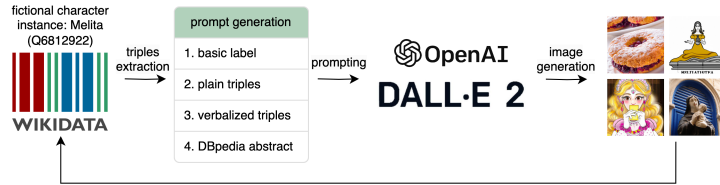


Fig. 1: The pipeline of our proposed KG completion process.

6.1 Triple Extraction

Generating an image for a specific entity requires its related triples to define the characteristics. In Wikidata, these can be obtained through SPARQL queries, yielding all triples with the entity as the subject. Moreover, since properties and entities are represented by IDs, we also extract and translate each triple from IDs to their corresponding labels.

6.2 Prompt Generation

For each unique entity, we generate several distinct prompts. The first three prompts are created utilizing the set of triples that have been extracted for the respective entity. These prompts are described as follows:

1. *Basic Label*: This prompt merely employs the “label” that Wikidata assigns to its entities. For instance, given the entity Q692395, the corresponding basic prompt is denoted as *Habinnas*.
2. *Plain Triples*: This prompt is derived by concatenating the subject, predicate, and object of a triple to form a single coherent sentence, utilizing all available triples linked to a specific entity. For instance, for the triple [Habinnas, occupation, craftsman], the sentence created from this triple is given as *Habinnas is an occupation craftsman*.
3. *Verbalised Triples*: Triple verbalisation is defined as the transformation of structured data (i.e., triples) into human-readable formats (i.e., text). The verbalised triples can serve as a paragraph of summarization of input triples.
4. *DBpedia Abstracts*: We also use DBpedia abstracts as prompts, obtained by querying the English chapter of DBpedia *et al.* [4]. Different from the above prompts, the abstract is edited by human editors.

Please note that for (2) and (3) above, the duplicate predicate of input triples except *instance of* is removed to avoid information duplication.

6.3 Image Generation

To run the text-image process, prompts were applied according to the distinction of the categories described above. First of all, Craiyon was applied without resorting “*negative words*” and selecting “*none*” for the style effect. Of the nine images generated, we opted for the one with the best resolution. DALL·E 2 has been used to apply the official API to automatically generate one image per prompt. In both generation processes, the output images were grouped by reference resource and sorted according to the resource ID provided by Wikidata. As Craiyon requires the manual insertion of prompts, we proceeded to select and organize a limited corpus of sample images to quantitatively and qualitatively compare the performances of two A.I. models.

6.4 Evaluation Framework

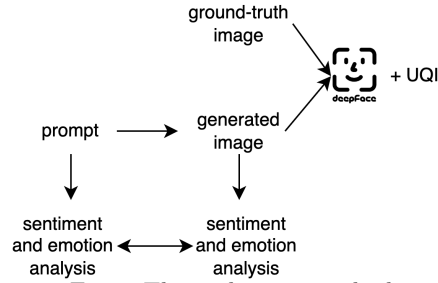


Fig. 3: Generated images for *Bevo*. (a) Ground truth from Wikidata. (b) Basic label. (c) Plain triples. (d) Verbalised triples. (e) DBpedia abstracts.

Two different evaluations are used to evaluate the characteristics of generated images, and how are they related to the types of prompts, as shown in Fig. 2.

1. The generated images are compared to the ground-truth image in Wikidata. For image similarity metrics, the Universal Quality Image Index (UQI) was utilized. UQI computes a pixel-based similarity score by comparing the generated images with their corresponding original images. Notably, since the majority of the original images are in grayscale, the similarity computation also takes into account their grayscale versions. UQI evaluates “image quality based on factors such as loss of correlation, luminance distortion, and contrast distortion” [24]. In a previous study [24], various pixel similarity metrics were compared, and the authors proposed UQI as an ideal metric. We have decided to leverage UQI metric for our evaluation distortion.
2. To assess the impact of prompt nature on generated image characteristics, we conducted a comparative analysis of prompts and resulting images. Our study specifically focused on evaluating the sentiment and emotion expressed in both elements, enabling a direct comparison. Emotion English DistilRoBERTa-base and DeepFace models were employed to predict emotions and extract information, respectively. The highest-scoring predicted emotion was considered for analysis, and sentiment analysis was performed for both prompts and image content. By following the widely accepted assumption[2] that certain emotions are perceived as positive or negative, we gained insights into the overall emotional tone conveyed by the prompts and images.

7 Results

Emotional Analysis: Table 1 shows the coherence of emotions and sentiments between prompts and their corresponding images. It investigates how well the

emotions and sentiments expressed in the prompts align with those observed in the generated images, providing valuable insights into their consistency and correlation. The coherence score is based on matching emotions/sentiments between prompts and images. The results demonstrates that the highest coherence is achieved when the image generator is prompted with verbalised triples as input.

Prompt Type	emotion sentiment	
Basic Label	10 %	53 %
Plain Triples	34 %	62 %
Verbalised Triples	35 %	86 %
DBpedia Abstract	42 %	80 %

Table 1: Coherence of Emotion and Sentiment between Prompts and Images

Image Evaluation: The DALLE-2 image generator has been used to generate images for 100 fictional characters across four prompt types. Table 2 shows the mean, minimum, and maximum UQI scores computed for the four different prompts. The basic label prompt exhibits the lowest mean, maximum, and minimum similarity scores compared to the other metrics. Conversely, the plain triples prompt yields the highest mean similarity scores, followed by the verbalised triples prompt. Consequently, relying solely on the basic label prompt is insufficient for generating images of fictional characters.

Prompt Type	min	max	avg
Basic Label	0.170945	0.792142	0.484021
Plain Triples	0.241029	0.848880	0.582811
Verbalised Triples	0.251560	0.674282	0.523799
DBpedia Abstract	0.171420	0.766946	0.508574

Table 2: Comparison between generated and ground-truth images based on UQI

8 Discussion and Conclusions

Our findings demonstrate that enhancing prompts, through verbalization or inclusion of DBpedia abstracts, significantly improves the coherence of emotions and sentiments between prompts and images. This highlights the potential of using natural language-like prompts in text-to-image models to effectively preserve emotional and sentiment aspects (addressing RQ2). Moreover, we cannot

say if enriching prompts improves image quality or not, as the evaluation metric fails to capture semantic content adequately. Thus, we cannot answer RQ3.

To address the limitation of image-similarity evaluation, future experiments will incorporate alternative measures such as DICE metric[10]. Additionally, human evaluation is crucial to better understand how prompt content and composition influence image quality (RQ2 and RQ3). Furthermore, we plan to evaluate on a larger and more diverse dataset, including testing on a dataset without ground truth, relying solely on human evaluation. Finally, we aim to expand our study to other domains, such as fictional cities and abstract concepts, for a broader understanding of prompt influence.

Acknowledgement

This project made use of the central High-Performance Computer system at Freie Universität Berlin, and we would like to express our gratitude for the resources provided. We would also like to extend our thanks to our tutor, Albert Meroño Peñuela, for generously providing us with his OpenAI API Key, but more importantly, for his invaluable input and help.

References

- [1] Gabriel Amaral, Odinaldo Rodrigues, and Elena Simperl. “WDV: A Broad Data Verbalisation Dataset Built from Wikidata”. In: *The Semantic Web–ISWC 2022: 21st International Semantic Web Conference, Virtual Event, October 23–27, 2022, Proceedings*. Springer. 2022, pages 556–574.
- [2] Sieun An, Li-Jun Ji, Michael Marks, and Zhiyong Zhang. “Two Sides of Emotion: Exploring Positivity and Negativity in Six Basic Emotions across Cultures”. In: *Frontiers in Psychology* 8 (2017). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.00610. URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00610>.
- [3] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. “Freebase: a collaboratively created graph database for structuring human knowledge”. In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. 2008, pages 1247–1250.
- [4] Martin Brümmer, Milan Dojchinovski, and Sebastian Hellmann. “DBpedia Abstracts: A Large-Scale, Open, Multilingual NLP Training Corpus”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*. Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pages 3339–3343. URL: <https://aclanthology.org/L16-1532>.
- [5] Chen Chen, Yufei Wang, Bing Li, and Kwok-Yan Lam. “Knowledge Is Flat: A Seq2Seq Generative Framework for Various Knowledge Graph Completion”. In: *arXiv preprint arXiv:2209.07299* (2022).
- [6] Ming Ding, Zhuoyi Yang, Wenyi Hong, Wendi Zheng, Chang Zhou, Da Yin, Junyang Lin, Xu Zou, Zhou Shao, Hongxia Yang, et al. “Cogview: Mastering text-to-image generation via transformers”. In: *Advances in Neural Information Processing Systems* 34 (2021), pages 19822–19835.

- [7] P. Ekman. “Basic emotions”. In: *Handbook of cognition and emotion*. Edited by T. Dalgleish and M. J. Power. 1999, pages 45–60.
- [8] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d’Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. “Knowledge graphs”. In: *ACM Computing Surveys (CSUR)* 54.4 (2021), pages 1–37.
- [9] *Introducing ChatGPT*. <https://openai.com/blog/chatgpt>. Accessed: 2023-06-15.
- [10] Shruti Jadon. “A survey of loss functions for semantic segmentation”. In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. 2020, pages 1–7. DOI: 10.1109/CIBCB48159.2020.9277638.
- [11] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. “Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia”. In: *Semantic web 6.2* (2015), pages 167–195.
- [12] Vivian Liu and Lydia B Chilton. “Design guidelines for prompt engineering text-to-image generative models”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pages 1–23.
- [13] Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. “Generating images from captions with attention”. In: *arXiv preprint arXiv:1511.02793* (2015).
- [14] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. “Glide: Towards photorealistic image generation and editing with text-guided diffusion models”. In: *arXiv preprint arXiv:2112.10741* (2021).
- [15] Jonas Oppenlaender. “A Taxonomy of Prompt Modifiers for Text-to-Image Generation”. In: *arXiv preprint arXiv:2204.13988* (2022).
- [16] Jonas Oppenlaender, Rhema Linder, and Johanna Silvennoinen. “Prompting ai art: An investigation into the creative skill of prompt engineering”. In: *arXiv preprint arXiv:2303.13534* (2023).
- [17] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. “Zero-shot text-to-image generation”. In: *International Conference on Machine Learning*. PMLR. 2021, pages 8821–8831.
- [18] Scott Reed, Zeynep Akata, Xinchun Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. “Generative adversarial text to image synthesis”. In: *International conference on machine learning*. PMLR. 2016, pages 1060–1069.
- [19] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. “High-resolution image synthesis with latent diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pages 10684–10695.

- [20] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. “Photorealistic text-to-image diffusion models with deep language understanding”. In: *Advances in Neural Information Processing Systems* 35 (2022), pages 36479–36494.
- [21] Sefik Ilkin Serengil and Alper Ozpinar. “LightFace: A Hybrid Deep Face Recognition Framework”. In: *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2020, pages 23–27. DOI: 10.1109/ASYU50717.2020.9259802. URL: <https://doi.org/10.1109/ASYU50717.2020.9259802>.
- [22] Tong Shen, Fu Zhang, and Jingwei Cheng. “A comprehensive overview of knowledge graph completion”. In: *Knowledge-Based Systems* (2022), page 109597.
- [23] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. “Yago: a core of semantic knowledge”. In: *Proceedings of the 16th international conference on World Wide Web*. 2007, pages 697–706.
- [24] Domonkos Varga. “Full-Reference Image Quality Assessment Based on an Optimal Linear Combination of Quality Measures Selected by Simulated Annealing”. In: *Journal of Imaging* 8.8 (2022). ISSN: 2313-433X. DOI: 10.3390/jimaging8080224. URL: <https://www.mdpi.com/2313-433X/8/8/224>.
- [25] Denny Vrandečić and Markus Krötzsch. “Wikidata: a free collaborative knowledgebase”. In: *Communications of the ACM* 57.10 (2014), pages 78–85.
- [26] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. “RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions”. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pages 1–29.
- [27] Chenfei Wu, Jian Liang, Lei Ji, Fan Yang, Yuejian Fang, Daxin Jiang, and Nan Duan. “Nüwa: Visual synthesis pre-training for neural visual world creation”. In: *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVI*. Springer, 2022, pages 720–736.
- [28] Xin Xie, Ningyu Zhang, Zhoubo Li, Shumin Deng, Hui Chen, Feiyu Xiong, Mosha Chen, and HuaJun Chen. “From discrimination to generation: knowledge graph completion with generative transformer”. In: *Companion Proceedings of the Web Conference 2022*. 2022, pages 162–165.
- [29] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. “Attngan: Fine-grained text to image generation with attentional generative adversarial networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pages 1316–1324.
- [30] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, and In So Kweon. “Text-to-image diffusion model in generative ai: A survey”. In: *arXiv preprint arXiv:2303.07909* (2023).

- [31] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaoolei Huang, and Dimitris N Metaxas. “Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pages 5907–5915.