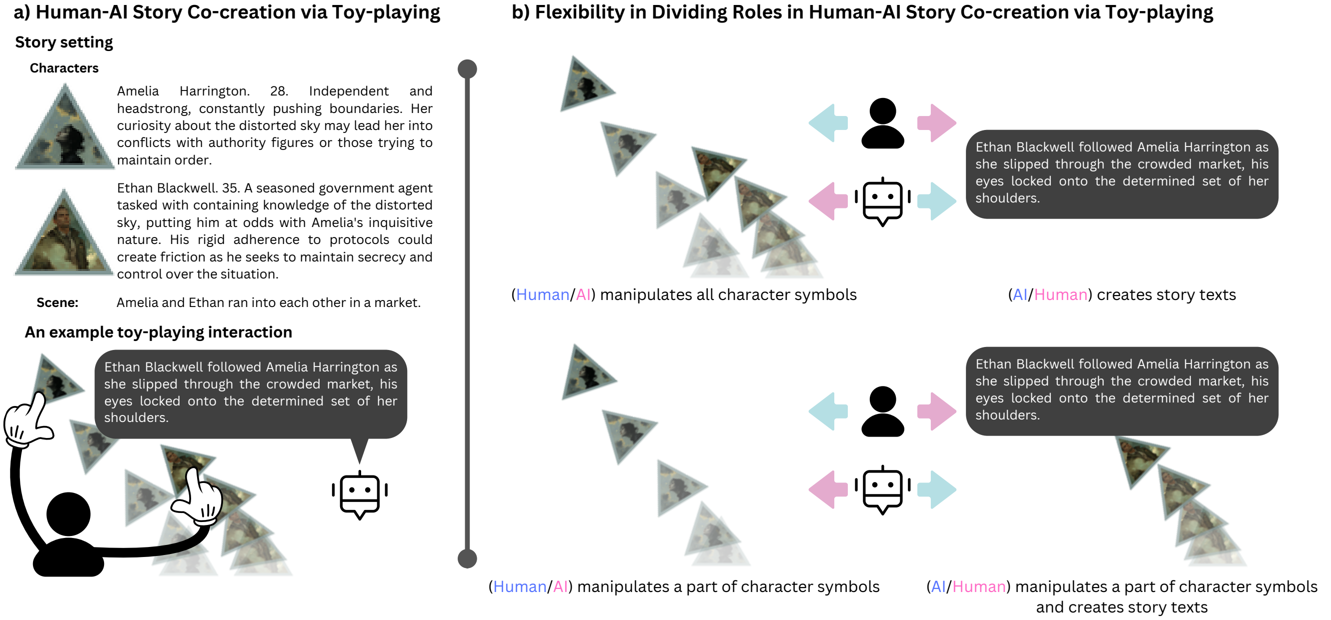
Toyteller: Toy-Playing with Character Symbols for AI-Powered Visual Storytelling

John Joon Young Chung
Midjourney
San Francisco, CA, USA
jchung@midjourney.com

Melissa Roemmele
Midjourney
San Francisco, CA, USA
mroemmele@midjourney.com

Max Kreminski
Midjourney
San Francisco, CA, USA
mkreminski@midjourney.com


a) Human-AI Story Co-creation via Toy-playing
Story setting
Characters
Scene:
An example toy-playing interaction
b) Flexibility in Dividing Roles in Human-AI Story Co-creation via Toy-playing
(Human/AI) manipulates all character symbols
(AI/Human) creates story texts
(Human/AI) manipulates a part of character symbols
(AI/Human) manipulates a part of character symbols and creates story texts

Amelia Harrington. 28. Independent and headstrong, constantly pushing boundaries. Her curiosity about the distorted sky may lead her into conflicts with authority figures or those trying to maintain order.

Ethan Blackwell. 35. A seasoned government agent tasked with containing knowledge of the distorted sky, putting him at odds with Amelia's inquisitive nature. His rigid adherence to protocols could create friction as he seeks to maintain secrecy and control over the situation.

Amelia and Ethan ran into each other in a market.

Ethan Blackwell followed Amelia Harrington as she slipped through the crowded market, his eyes locked onto the determined set of her shoulders.

Ethan Blackwell followed Amelia Harrington as she slipped through the crowded market, his eyes locked onto the determined set of her shoulders.

Ethan Blackwell followed Amelia Harrington as she slipped through the crowded market, his eyes locked onto the determined set of her shoulders.

Figure 1: Interactions in Toyteller. a) Toyteller enables visual storytelling with AI through *toy-playing* interaction with character symbols that accompany story texts. These character symbols can become both 1) a means to express the user's intention about the story unfolding and 2) a generation target that will be a part of the final visual story. For example, given a story setting defined as open-ended text, the user can physically manipulate symbols representing two of the story's characters, and Toyteller can generate a story sentence that aligns with the user-provided character motions. b) Toyteller allows users flexibility in deciding which part of the output artifact (symbol motions, story texts) is created by the user or by the AI.

ABSTRACT

We introduce Toyteller, an AI-powered storytelling system that allows users to generate a mix of story texts and visuals by directly manipulating character symbols like they are playing with toys. Anthropomorphized motions of character symbols can convey rich and nuanced social interactions between characters; Toyteller leverages these motions as (1) a means for users to steer story text generation and (2) an output format for generated visual accompaniment to user-provided story texts and user-controlled character motions. We enabled motion-steered story text generation and text-steered motion generation by mapping symbol motions and story texts onto a shared semantic vector space so that motion generation models and large language models can use it as a translational layer. We hope this demonstration sheds light on extending the range of modalities supported by generative human-AI co-creation systems.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**; **Interaction techniques**; • **Computing methodologies → Natural language generation**.

KEYWORDS

visual storytelling, toy-playing, generative AI

ACM Reference Format:

John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2024. Toyteller: Toy-Playing with Character Symbols for AI-Powered Visual Storytelling. In *The 37th Annual ACM Symposium on User Interface Software*

UIST Adjunct '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0718-6/24/10
https://doi.org/10.1145/3672539.3686781

and Technology (UIST Adjunct '24), October 13–16, 2024, Pittsburgh, PA, USA. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3672539.3686781

1 INTRODUCTION

Generative AI technologies, such as large language models (LLMs) [3, 11], have been shown to enable many new forms of AI-supported storytelling [4, 9, 14]. Many AI-powered storytelling applications, however, still rely heavily on natural language as the primary means of steering story generation—even though stories are often expressed through modalities beyond natural language [7]. In the context of human-human story co-creation, for example, children often casually create stories while they play with toys in their hands [1, 6, 13]. Existing AI systems do not effectively support these kinds of collaborative multimodal storytelling: even the most advanced AI models yet (1) exhibit limited understanding of sequential multimodal inputs [2, 15]; (2) suffer from latency that limits interactive use [10]; and/or (3) assume complex multimodal inputs (e.g., comic strips [8]) that are difficult for users to casually manipulate.

In this work, we introduce Toyteller, an AI-powered visual storytelling system that adopts the movement of character symbols as both an output modality and a steering input. Specifically, in Toyteller, the user can collaborate with AI to create the story of two characters along with their corresponding motions in symbols (Figure 1). The user manipulates the motion of one or two characters or writes down the story text, and AI fills in the rest of the content that is not filled in by the user (i.e., motion or texts). Toyteller leverages the symbolic movements of characters inspired by Heider and Simmel's experiments [5], as these anthropomorphized motions can express rich and nuanced semantics of social interactions with simple manipulation of symbolic objects. We achieve motion-to-text and text-to-motion generation by having a translational layer that maps motion inputs onto the frozen semantic vector space for texts, so that we can condition large language models and motion generation models with motions and texts, respectively. Through this demonstration, we hope to inspire the UIST community to consider more diverse interaction modalities for human-AI co-creation.

2 TOYTELLER: INTERACTION

In this section, we first motivate the use of toy-playing interaction in human-AI co-creation of visual stories and then explain interactions in Toyteller's interface.

2.1 AI-powered Storytelling via Toy-playing

With toy-playing interaction for human-AI story co-creation, we leverage humans' ability to perceive social interactions between characters even from simple movements of symbolic shapes. Heider and Simmel's experiment is one prominent example (Figure 2), where they found that people can anthropomorphize motions of triangles and a circle, considering these shapes as characters and their motions as social interactions happening between these characters [5]. The range of social interactions that can be expressed with these symbolic motions is wide—for instance, Roemmele et al. [12] categorized possible interactions between two characters into 31 classes. In fact, even motions of the same category are able to express nuanced differences through varying dynamics, such as the magnitude or velocity of movement. For example, for the motion of "hit," the velocity of a hitting character would indicate how hard the character hit the other one.

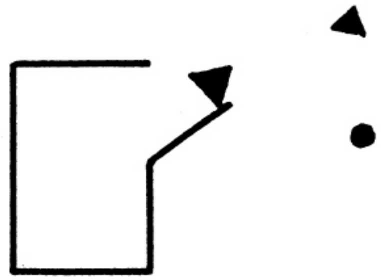


Figure 2: Heider and Simmel's experiment.

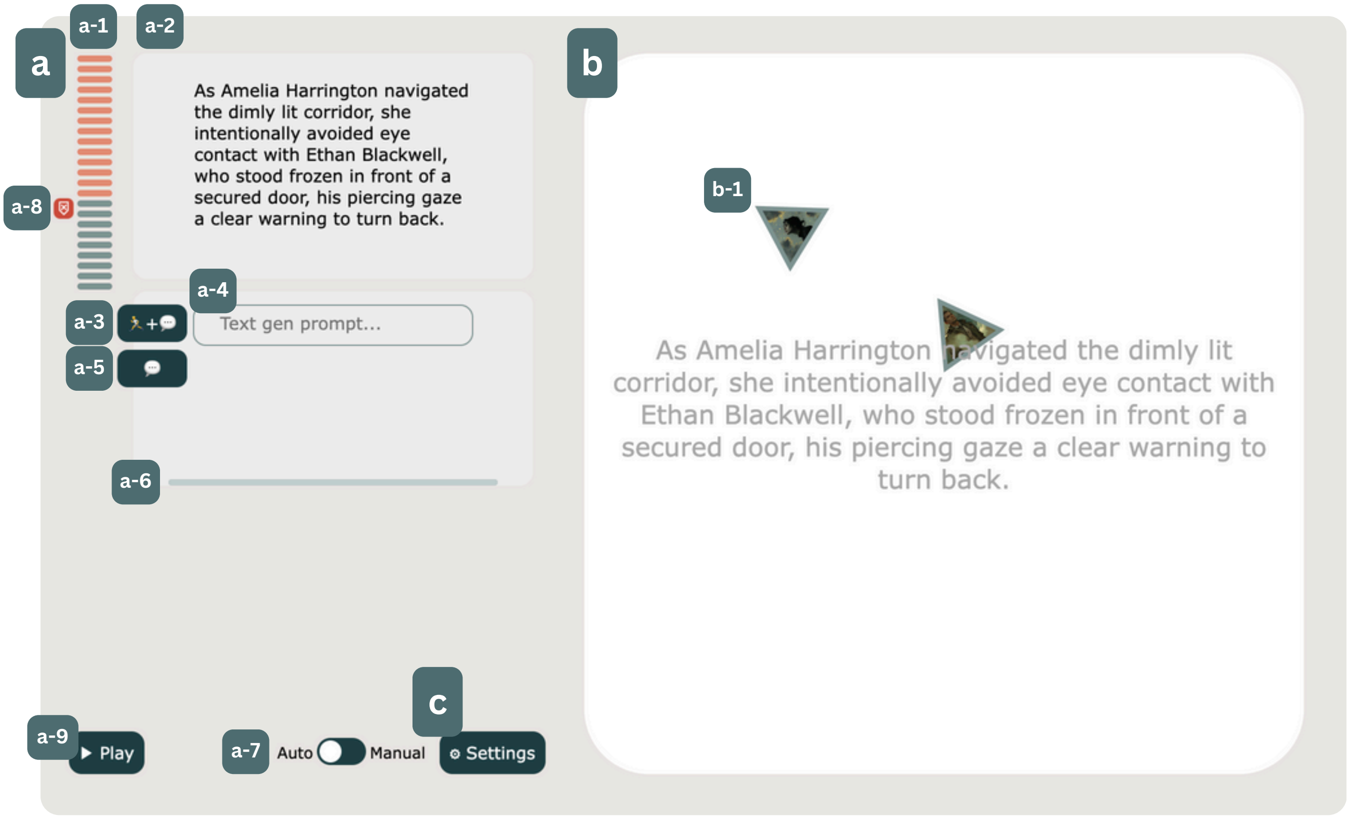
In the context of human-AI story co-creation, the movement of shapes has the benefit that it can serve both as 1) a means to steer AI generation and 2) a target of co-creation with AI. That is, as steering input, the user's manipulation of character symbols can serve as a means to condition the AI's generation of story text. This steering input offers complementary benefits to other input forms (e.g., natural language prompts), allowing users to express nuanced intentions about story events via the relatively simple interaction of gestural object manipulation. Meanwhile, movements of character symbols can also serve as part of a story artifact, as a visual complement to the story text. Hence, in human-AI co-creation contexts, these movements can be a target of AI generation—for instance, the AI can generate character motions from user-provided story text or as a reaction to the user's manipulation of other character symbols. As these character motions can serve as both input and output, the user can flexibly decide which parts they would like to contribute to visual stories, and the AI can fill in the remaining parts based on whatever input the user supplies (Figure 1b).

This *toy-playing* interaction—humans and AI manipulating visual character symbols while telling accompanying stories—can have many application scenarios. In this paper, we introduce one specific application, Toyteller, which is designed to demonstrate toy-playing interaction in a 2D screen interface.

2.2 Interface

We designed Toyteller to demonstrate toy-playing-based human-AI story co-creation in a specific setting of two characters interacting with each other. While there can be more complex story settings with more characters and props, we consider dyadic settings as users can still express a wide variety of stories with them. Moreover, Toyteller focuses on the generation of story prose text. Toyteller is designed for touchscreen and mouse-pointer user interfaces, which permit the user to manipulate at most two character symbols simultaneously.

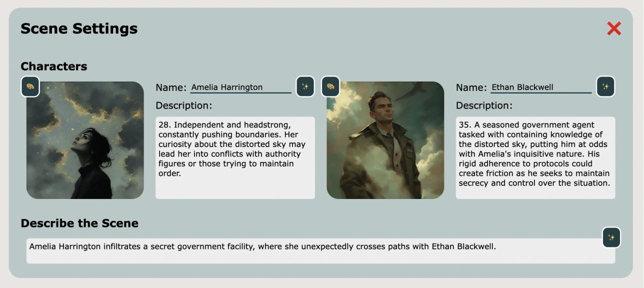
At a high level, Toyteller consists of three parts: 1) a setting page, allowing users to configure characters and story background; 2) a story timeline, showing a sequence of passages of story text; and 3) a playground, where a single passage of story text is complemented by user-manipulatable character symbols.



As Amelia Harrington navigated the dimly lit corridor, she intentionally avoided eye contact with Ethan Blackwell, who stood frozen in front of a secured door, his piercing gaze a clear warning to turn back.

As Amelia Harrington navigated the dimly lit corridor, she intentionally avoided eye contact with Ethan Blackwell, who stood frozen in front of a secured door, his piercing gaze a clear warning to turn back.

Figure 3: The Toyteller interface consists of a timeline module (a), playground (b), and a button to enter the setting page (c). The user can use the progress bar (a-1) to navigate recorded motion frames, which align with story textboxes (a-2). The user can manipulate character symbols (b-1) to record character motions. The user can ask Toyteller to generate both character motions and story texts (a-3) while giving a high-level direction as a natural language prompt (a-4). The user can also initiate the generation of story text only (a-5) and adjust the size of the last textbox (a-6). By default, Toyteller automatically reacts to the user's input, but the user can turn this off (a-7). The user can delete frames and textboxes after a selected frame with a button (a-8). After recording frames and texts, the user can play them back (a-9).



Scene Settings

Characters

Name: Amelia Harrington

Description:

28. Independent and headstrong, constantly pushing boundaries. Her curiosity about the distorted sky may lead her into conflicts with authority figures or those trying to maintain order.

Name: Ethan Blackwell

Description:

35. A seasoned government agent tasked with containing knowledge of the distorted sky, putting him at odds with Amelia's inquisitive nature. His rigid adherence to protocols could create friction as he seeks to maintain secrecy and control over the situation.

Describe the Scene

Amelia Harrington infiltrates a secret government facility, where she unexpectedly crosses paths with Ethan Blackwell.

Figure 4: Interface for setting story scene. The paint and sparkle buttons allow users to use AI to generate the setting's content images and texts, respectively.

2.2.1 Setting. Before unfolding the story with toy-playing interactions, the user needs to set up the story setting by opening a setting page with the button in Figure 3c. Specifically, the user can set 1) two characters' names, 2) specific descriptions about them, 3) their profile images, and 4) a high-level description of the scene (Figure 4). Information except profile images will be used to generate story texts, while images will be overlaid on the character symbol to indicate which symbol corresponds to which character. While users can manually type in or upload these entries, Toyteller also provides the option of generating texts and images with large language models and text-to-image models, with the sparkle and paint buttons, respectively.

2.2.2 Story Timeline and Playground. Once the user has configured the story setting, they can start unfolding the story using the story timeline (Figure 3a) and the playground (Figure 3b). The story timeline includes the progress bar that shows the recorded frames (Figure 3a-1) along with the story textboxes that correspond to certain parts of the progress bar (Figure 3a-2). The playground has two triangular character symbols, which are manipulatable by both the user and AI models (Figure 3b-1).

Motion → Text. As mentioned in Section 2.1, Toyteller allows flexible human-AI co-creation of visual story. The first approach

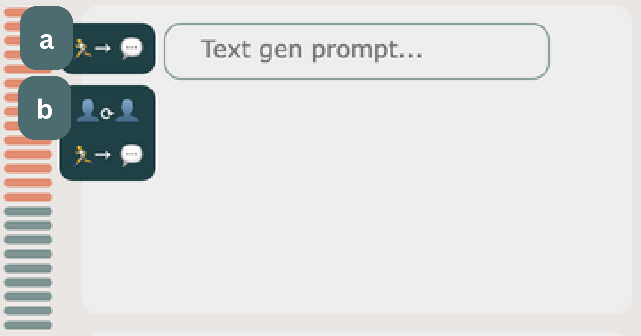
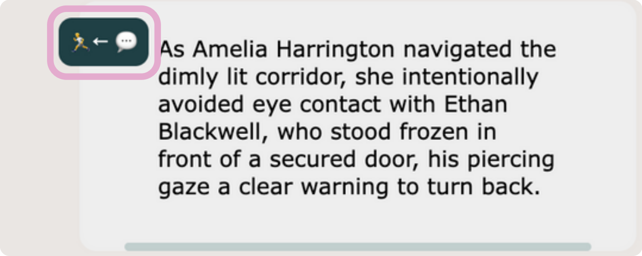


Figure 5: The user can manually initiate motion-to-text generation with a) and b). b) allows users to generate a sentence by swapping the "active" character of the event recognized by the system.

to interacting with Toyteller involves the user and AI defining character motion first, then adding story sentences that align with this character motion. When contributing character motion, the user can decide to move one or two character symbols, and the AI will concurrently generate motion for uncontrolled character symbols. Alternatively, the user can invoke AI to generate motion for both characters (Figure 3a-3). As the user or AI manipulates symbols, Toyteller records motion frames and accumulates them on the timeline. When the timeline reaches the end of the last textbox and the user is controlling one of the symbols, the size of the last textbox grows to match its end to the end of the timeline. As the user stops manipulating the symbols, Toyteller starts generating a story sentence that aligns with the provided motions. When AI generates both character motions, Toyteller starts generating the story sentence as the end of the timeline reaches the end of the last textbox. Note that for the text generation, the user can prime the high-level direction with a natural language prompt (e.g., "Write the ending of the story," Figure 3a-4).

While the above interactions assume that Toyteller reacts to the user's input automatically (e.g., as the user moves one character, AI moves the other character automatically), there might be cases in which the user wants to decide the moment for AI generation. In such cases, the user can toggle the switch in Figure 3a-7. As the user is in manual mode, if the user provides character motions first, they can manually generate story texts with buttons in Figure 5. Note that the button in Figure 5b allows the users to swap who is active or passive in the event, as there are cases where AI fails to recognize the active character in the event correctly.

Text ⟶ Motion. The second approach to interacting with Toyteller involves the user or AI contributing passages of story text first, then creating character motions. Here, the user can decide to write a story sentence by themselves or use AI to generate it (Figure 3a-5). Again, when generating text with AI, the user can steer text generation via a natural language prompt. Then, the user can move some or all of the character symbols to align them with the provided sentence, and AI will generate the motion of characters uncontrolled by the user. The user can also make the system generate all characters' motions with the button in Figure 6. When the user



As Amelia Harrington navigated the dimly lit corridor, she intentionally avoided eye contact with Ethan Blackwell, who stood frozen in front of a secured door, his piercing gaze a clear warning to turn back.

Figure 6: When text is provided before character motions, with the button highlighted in pink, the user can make AI generate those motions.

wants to adjust the size of the textbox so that fewer or more frames can correspond to it, they can use the handle in Figure 3a-6.

Revision. As the user accumulates a series of motions and story sentences, they can revise them. For motions, the user can first move to the start of the frames they want to edit and manipulate the character symbols to override previously recorded character motions. For sentences, the user can manually edit them in the text box. If they want to regenerate the story sentence, they can first delete the existing sentence and use buttons in Figure 5. The user can also delete created frames and sentences. Specifically, they can delete content after a specific frame by moving to the frame in the timeline and then clicking the delete button (Figure 3a-8). As the user is done with editing the content, they can play the recorded frames along with the aligned story sentences by clicking `Play` button (Figure 3a-9).

3 CONCLUSION

In this demonstration, we present Toyteller, which facilitates human-AI visual storytelling via toy-playing interactions. As people can easily anthropomorphize the motion of character symbols, toy-playing interactions can serve as not only a good means to express intent about character-character interactions but also a form of story content in their own right. We hope that Toyteller will inspire the UIST community to expand the range of interaction modalities supported in human-AI co-creation.

ACKNOWLEDGMENTS

Thanks to the many people at Midjourney who provided infrastructural and logistical support for this work.

REFERENCES

[1] Margaret S. Benson. 1993. The structure of four- and five-year-olds' narratives in pretend play and storytelling. *First Language* 13, 38 (1993), 203–223. https://doi.org/10.1177/014272379301303803 arXiv:https://doi.org/10.1177/014272379301303803

[2] Ankur Chemburkar, Andrew Gordon, and Andrew Feng. 2024. Evaluating Vision-Language Models on the TriangleCOPA Benchmark. *The International FLAIRS Conference Proceedings* 37, 1 (May 2024). https://journals.flvc.org/FLAIRS/article/view/135485

[3] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay

Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. PaLM: Scaling Language Modeling with Pathways. arXiv:2204.02311 [cs.CL]

[4] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. https://doi.org/10.1145/3491102.3501819

[5] Fritz Heider and Marianne Simmel. 1944. An Experimental Study of Apparent Behavior. *The American Journal of Psychology* 57, 2 (1944), 243–259. http://www.jstor.org/stable/1416950

[6] Carollee Howes. 1992. *The collaborative construction of pretend: Social pretend play functions*. State University of New York Press.

[7] Ting-Hao Kenneth Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh, Lucy Vanderwende, Michel Galley, and Margaret Mitchell. 2016. Visual Storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Kevin Knight, Ani Nenkova, and Owen Rambow (Eds.). Association for Computational Linguistics, San Diego, California, 1233–1239. https://doi.org/10.18653/v1/N16-1147

[8] Ze Jin and Zorina Song. 2023. Generating coherent comic with rich story using ChatGPT and Stable Diffusion. arXiv:2305.11067 [cs.CV] https://arxiv.org/abs/2305.11067

[9] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA.

[10] Yaniv Leviathan, Matan Kalman, and Yossi Matias. 2023. Fast Inference from Transformers via Speculative Decoding. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 19274–19286. https://proceedings.mlr.press/v202/leviathan23a.html

[11] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL] https://arxiv.org/abs/2303.08774

[12] Melissa Roemmele, Soja-Marie Morgens, Andrew S. Gordon, and Louis-Philippe Morency. 2016. Recognizing Human Actions in the Motion Trajectories of Shapes. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 271–281. https://doi.org/10.1145/2856767.2856793

[13] Nilüfer Talu. 2018. Symbolic creativity in play activity: a critique on playthings from daily life objects to toys. *International Journal of Play* 7, 1 (2018), 81–96.

[14] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3490099.3511105

[15] Jieyu Zhang, Weikai Huang, Zixian Ma, Oscar Michel, Dong He, Tanmay Gupta, Wei-Chiu Ma, Ali Farhadi, Aniruddha Kembhavi, and Ranjay Krishna. 2024. Task Me Anything. *arXiv preprint arXiv:2406.11775* (2024).