Storybranch - generating multimedia content from novels

Rushikesh Hiray
School of Computer Science
University of Birmingham
rhiray03@gmail.com

Venelin Kovatchev
School of Computer Science
University of Birmingham
v.o.kovatchev@bham.ac.uk

Abstract

We present STORYBRANCH - an automated system for generating multimedia content from long texts such as novels and fanfiction. The STORYBRANCH pipeline includes structured information extraction, text parsing and processing, content generation using Gen-AI models and syncronization of different streams (audio, video, background). Our system is highly modular and can efficiently generate three different types of multimodal content: audiobooks, simple animated videos, and visual novel text-and-image-style video games.

STORYBRANCH successfully addresses challenges such as generating unique and consistent image and voice for each character and narrator, identifying and generating background images and sounds effects, and syncronizing character expressions and lip movement with text.

As part of the STORYBRANCH , we develop and release BookNLP2 - a new open-source library for parsing and extracting information from books, based on the legacy library BookNLP.

1 Introduction

The recent advances of machine learning in the fields of speech (Casanova et al., 2021) and image (Datta et al., 2024) generation have created opportunities in multimedia content creation. Different open-source and commercial models (Sauer et al., 2024) excel at generating images, sound files, and video clips from short texts. However, generating customizable, high-quality multimedia content from long texts, such as novels and books, faces various non-trivial challenges: the system has to keep track of multiple characters, narrator(s), and constantly changing background images and noises.

In this paper, we present STORYBRANCH : a new system for multimedia content creation from long texts. We extract and process information from novels and books in a unified structured manner and leverage the capabilities of different Generative AI models to create content in different output formats: audiobooks, simple animated videos, and video games. Figure 1 shows a snapshot from a video generated by STORYBRANCH from the book Crime and Punishment by Fyodor Dostoyevsky. The scene has two characters: Raskolnikov and Alyona Ivanovna and the background is his room.

Figure 1: A scene from a STORYBRANCH video

In addition to being capable of generating complex output in different modalities, STORYBRANCH addresses multiple practical challenges: separation of human characters, animals, narrator speech, and the description of background images and sounds; consistent image and voice for each character; dynamic update of context-specific properties of characters, such as change of clothes and age; separate generation of the narrator, background effects and noises; synchronization of different media files. We encourage readers to check the demo video[1].

STORYBRANCH is highly modular and is structured as a five-step pipeline:

- **Information Extraction:** we use our new open-source library, BookNLP2, to process long texts and identify entities, properties, events, coreference, and the current speaker.

[1] Using Storybranch Demo

Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations), pages 485–493
April 30, 2025 ©2025 Association for Computational Linguistics

- **Data Processing:** we restructure the extracted data and create separate structured entries for different characters, background images and effects, and impersonal "narator" voice.
- **Prompt Engineering:** we use both structured and unstructured data to generate dynamic, chapter- and event-specific prompts for each character, the narrator and background.
- **Content Generation:** we employ open-source text-to-speech and text-to-image models to generate different multimedia elements: individual utterances, background noise, character and background images.
- **Content Synchronization:** we use the Python library 'Moviepy' to combine the different elements in a single audio, video, or game.

Multimedia solutions such as audiobooks have become a popular way to enjoy and connect with literature. Animations and image-based games can further enhance the experience and make it more interactive and engaging for children and second-language learners. STORYBRANCH can generate one or more output types at scale, using the same text-processing and content generation pipeline. As part of the project, we created and released a new text open-source library, BookNLP2[2].

2 Related Work

To ensure consistency and overall quality of the generated content, we first perform **text parsing**. The original BookNLP library is inspired by the works of Bamman et al. (2019) and Weerasundara and de Silva (2023). Our baseline and initial experiments use these approaches as well. We extend the toolbox by including the work of Yoder et al. (2021) which focuses on parsing fanfiction litrary texts. Approaches include supersense tagging (Attardi et al., 2023), entity tagging and coreference resolution (Yao et al., 2023).

Audiobook Generation is central to STORYBRANCH as both audiobooks and videos require a text-to-speech model as well as generation of special effects. ElevenLabs[3] is a highly popular commercial TTS system that focuses on the generation of human speech. Our work directly takes inspiration from popular TTS frameworks such as XTTSv2 (Casanova et al., 2024). The landscape of speech generation has seen rapid growth over the last few years.including work on expressive human-sounding narration (Barakat et al., 2024).

A current state-of-the-art model for **video Generation** is the work of (Chen et al., 2024). They present a SOTY video generation paper capable of generating high-quality videos. A key limitation to their work is the scope - they can only generate videos for short timeframes. Nonetheless, significant progress is made in improving the art of creating continuous frames (Lee et al., 2023).

The **evaluation** of multimedia generation from long texts is non-trivial, as there are no large-scale datasets for the task. Existing benchmarks for LLMs and NLP models (Liang et al., 2023; Srivastava et al., 2023; Kovatchev and Lease, 2024) are not applicable to the task. We rely on task specific benchmarks for the components of BookNLP2 (Bamman et al., 2019) and perform separate human evaluation on the final multimedia output.

3 Processing long texts - BookNLP2

Generating multimedia content from long texts in an end-to-end manner is beyond the capabilities of current open-source models. The complex structure and long-distance dependencies of a book result in high computational cost and low quality output. Instead of going end-to-end we take a hybrid divide-and-conquer approach. We first process the book and perform automatic annotation and analysis, converting the raw text into a predefined structured format. Then we combine the raw text and the structured data to generate accurate sound files and images for different characters, descriptions, and narrator. Finally, we recombine the different streams to obtain a single multimodal output.

The problem of parsing long texts has been addressed before in the Booknlp (David Bamman and Shen, 2020) library. It combines various NLP models in a configurable pipeline. Unfortunately, the library is outdated and is no longer supported.

We developed an updated version of the library called BookNLP2. We replicated the original work, updated dependencies, and deployed it on pypl for easy installation across different OS. We keep the format of input, output, and usage, enabling backwards compatibility. We have also started replacing the individual NLP models in the pipeline with state-of-the-art models using LitBank (Bamman et al., 2019) for training and evaluation. We design

[2] BookNLP2 Github
[3] ElevenLabs: Prime Voice AI

BookNLP2 as a configurable modular system that also allows users to finetune the models on specific custom domains and books. In STORYBRANCH we use the library with its default pretrained version, without in-domain finetuning, as the first step in the content generation pipeline. We release BookNLP2 as open source.

4 Extracting and Structuring data

When generating multimedia content from books, a key challenge is to correctly identify the characters, the narrator, and the context of the scene. Some aspects of the characters, such as their voice, height, and eye color, remain consistent throughout the book. Other aspects, such as their age or clothes, may change from one scene from another. While Booknlp2 does a great job extracting information about individual scenes, we need to process and enrich the data further, so that we can account for the dynamic nature of the generation of sound and animations. At the current version of STORYBRANCH we have two types of dynamically generated content: audio effects and character animations. Our approach identifies the parts of the text that correspond to either of them, extracts and stores the information.

We extract the information about the audio effects for each scene in three steps. First, we separate the scenes and perform basic linguistic processing. Then, we use zero-shot learning to identify parts of the text that describe effects and background noises. We also ask the model to provide a time duration for the effect.

For obtaining an updated description of characters, we also perform a two step approach. First, we use BookNLP2 to identify all characters in the book and to store all related information (description, coreference entities, scenes) in a csv format. We use the base description as the foundation of generating animations for the character, that is, the static character information. To account for potential change in the description, for every scene that the character participates in, we use zero-shot learning to dynamically extract information from the text chunk, using Structured Outputs. Here is an example prompt:

"" From the text below, I am trying to make an animated video. This is the person for who I need a stable diffusion prompt for person. If the text has a proper description of this person share the prompt, if you feel like you understand the essence character from the text feel free to reply using that. The prompt should talk about the face, clothing, age, facial features, body type and gender. the prompt should have 1 - 3 word descriptions followed by commas . Reply only with the prompt ""

Attribute	Description
Face	angular face, light skin,
Clothing	brown coat, jeans,
Age	mid-20s,
Facial Features	rough eyes, medium nose,
Body Type	slim, athletic,
Gender	male

Table 1: Structured LLM Output for generating dynamic character animations

Table 1 shows an example of LLM response. If the LLM returns a NOT NULL value, we overwrite parts of the static description with the dynamic description. We further experiment with keeping the dynamic descriptions across scenes and chapters.

5 Prompt generation

After we extract all static and dynamic information for each scene, as described in Section 4, we need to convert it into prompts that can be given to generative AI models. We generate the following prompts for each individual scene:

- prompts for character voices (+ narator)
- prompts for sound effects
- prompts for character description
- prompts for background animation

We generate the prompts for audio and video generation by passing the static and the dynamic information to LLMs. We have tested various LLM including OpenAI's ChatGPT 4o, Mistral, and LLama-3. We performed human experiments in Section 9 to determine user preferences.

6 Content generation

We use the prompts from Section 5 to generate multimedia content, such as audio and image. We follow the format of the prompts and generate individual characters and effects separately.

For the **audio**, we experiment with a variety of TTS and TTA models. We specifically select models that are open-sourced. Where possible, we

Figure 2: Characters from different sides

have finetuned them on voices of individuals that have allowed the usage of their voices. The fine-tuned models for character speech include XTTSv2, StyleTTS2, your-tts, and Parler-TTS. For sound effects, we used Stable-open-audio and AudioLDM.

For the **image** generation, we also favor open source models. We create "ComfyUI" workflows and chain multiple publicly available Stable Diffusion and FLUX models to generate content. The "ComfyUI" is key to consistently generating the same character in multiple angles allowing for more dynamic animations for the final video. Figure 2 show the same character seen from from multiple angles and with different clothes.

7 Synchronization

Once all the components are in place and the dialogue and images are generated for every character, we need to combine all components in a single multimedia output. First, we use the existing meta-data to assign each image and audio to the corresponding scene. Then, we compose the audio stream by merging the audio effects with the dialogues. We use the audio as the base to determine the correct timing of the scenes. The images are merged with the audio to generate videos and interactive games. First, we pair the audio with the corresponding character visuals. Then we add the background images corresponding to the scene. Finally we have the non-speaker character's opacity dropped to hint at who the speaker is. The current video generation is simplified and cartoon-like[4]. We are experimenting with more capable generative models.

All of the basic synchronization is currently done using the 'MoviePy' python library which is under the MIT license. We generate the video in chapters, but we experiment with keeping information between chapters, such as "last known dynamic character description". Once the video is generated, we can inspect the output and make iterative improvements to the extraction process, the choice of models, or directly on the CSV file, if we want to isolate sources of errors. We conducted experiments with users to assess the best pipeline and LLMs for the prompt and content creation. We also use that feedback to improve the integration of BookNLP and get additional data.

7.1 Lip Synchronization

To further improve the quality of the videos, we added a lip sync mechanism. To get accurate and robust lip synchronization, we added the Montreal Forced Aligner (MFA) (McAuliffe et al., 2017) into our pipeline. The MFA is a popular forced alignment of phonetic segments with corresponding audio frames. It receives the generated audio and the corresponding text and provides a phoneme-level alignment for every character's dialogue. We use timestamps to detect which phonemes are being spoken at any one moment in the audio. We then drive the character's lip and facial movements using this phoneme-to-time mapping so that, when the audio is played, the mouth and face expressions match the strong phonemes.

This method greatly enhances the realism of the animations since lip forms and mouth locations closely follow the development of speech sounds. Moreover, the MFA-based synchronization is robust to variations in speech rate and prosody, providing a stable and flexible foundation for enhancing the overall visual quality and believability of the final animated videos. You can find see the impact of lip syncronization here.

[4]Demo Video of Dorian Gray

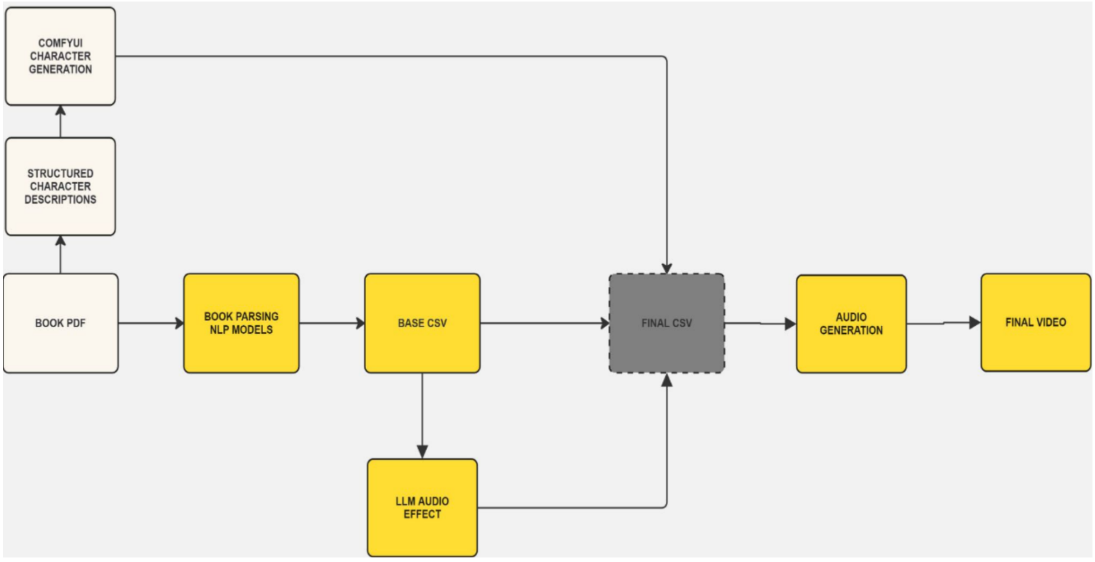

Figure 3: Overall Flowchart of STORYBRANCH

8 Visual Novel Game Generation

We use STORYBRANCH to create a new type of content in addition to audio and animation - interactive text-based games. We use the same data that we extracted for the first two steps in the pipeline, and add an additional logic component that allows for user interaction to influence the flow of the story.

Example timeline

```
[background path="res://assets/backgroudns/dialogic_factory.png"]

join Jowan left
jowan (exited): Hello and welcome to[portrait=confused]...[pause=0.5] Wait?

join Emilio (happy) right
Emilio: Well, this is is the example timeline.

Jowan: I thought this was a cool new feature?
```

Figure 4: Dialogic Script Example

We use the open-source game library Godot (Linietsky et al., 2024). The library has a visual-novel making component called dialogic that uses character images and dialogues to help make the game. We automatically convert the CSV data from STORYBRANCH into the dialogic script necessary for the visual novel games. Figure 4 shows an example of a dialog script. We are currently developing a method for adding user choices to the game that can lead to new outcomes and interactions.

9 STORYBRANCH

By combining all individual elements in a single system, we have created STORYBRANCH - a multimedia generator pipeline able to convert long-text books and novels into multiple output formats. Figure 3 shows the overall flowchart of STORYBRANCH . We start with two forms of textual analysis - static data extraction using BookNLP2 and dynamic data extraction using zero-shot learning. We use the data to first generate an audio book. The audio is then used as a base for creating animations, and the animations are converted in interactive visual-novels using Godot. The system is highly flexible and allows for customizing the models, the prompts, or directly modifying the raw data of the csv. We have experimented with several classical novels and achieved good content quality.

We performed two experiments using human participants to determine the impact of different models and settings on the final product. We recruited 12 volunteers for the task and asked them to compare audiobooks. In the first experiment, we compared audios generated by STORYBRANCH using different generative models. In the second experiment we compared STORYBRANCH to existing solutions for audiobook generation.

9.1 Assessing the impact of different components in STORYBRANCH

The modular nature of STORYBRANCH means that at every stage there are multiple alternatives for information extraction and content generation. In this experiment, we assess the impact of different components on the final product.

All participants were shown two outputs generated by different combination of open-source tools and asked to select a preference. Each participant made five comparisons. We test different models

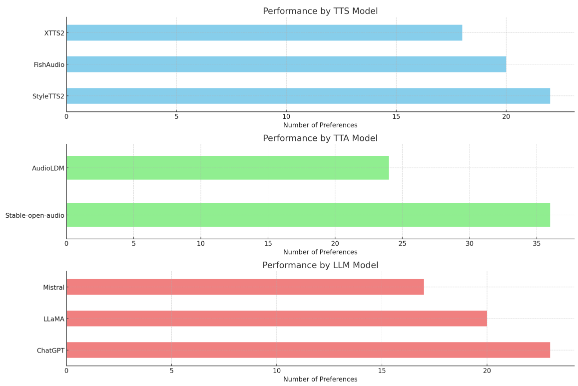

Figure 5: Comparing different generative strategies

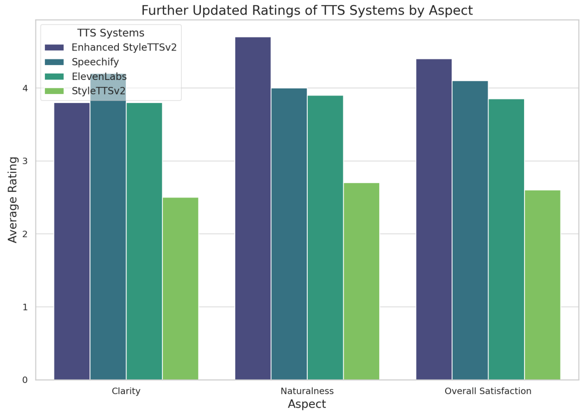

Figure 6: Comparing STORYBRANCH to other products

for: 1) **prompt generation**: ChatGPT, Llama, Mistral; 2) **text-to-speech**: XTTS2, FishAudio, and StyleTTS2; and 3) **text-to-audio**: AudioLDM and Stable-open-audio. Participants were asked to select the model that generated better audio representation. Model order was randomized.

Figure 5 shows the results of the experiments. Our user study indicates that the best combination uses ChatGPT for prompt generation, StyleTTS2 for text to speech and Stable-open-audio for sound effects. In the second set of experiments, we used this configuration for STORYBRANCH .

9.2 Comparing STORYBRANCH to commercial audiobook generation systems

In this study we compared STORYBRANCH to two popular commercial models: Speechify and ElevenLabs. We used the best STORYBRANCH configuration from the previous experiment (ChatGPT and StyleTTS2). We experimented with and without additional effects (Enchanced StyleTTS2 is StyleTTS2 + Stable-open-audio effects) to determine the impact of adding effects to the audio.

The participants were asked to make a comparison across three categories: **Speech clarity**, **Naturalness**, and **Overall Satisfaction** on a scale from 1 to 5. Figure 6 shows the results of the experiment. Without the special effects STORYBRANCH is comparable but worse than the commercial alternatives. However adding dynamic effects significantly improves users' perception of the audio and with them STORYBRANCH is clearly the best system.

We tried comparing STORYBRANCH to existing video generation systems, namely Haiper.ai and Veed.io . However, the competing models are not able to generate video longer than few seconds. We attempted generating several short videos and concatenate them, but the resulting video quality was poor and STORYBRANCH substantially outperforms those baselines.

10 Conclusions and future work

We have created a new system, STORYBRANCH , which can convert long texts such as novels and fanfiction text to multiple types of multimedia content with a high degree of success. It provides an immersive approach to experience books by combining narrative, sound effects, and animation design, transforming stories into more dynamic and entertaining format. The audiobooks, animated videos, and visual novel games improve the user experience and accessibility.

During the creation of STORYBRANCH we addressed several challenges: identifying different scenes, characters, background sounds and images. We had to balance between static properties and in-context dynamic changes in the characters and the environment. The addition of sound effects was well received by users, who rated STORYBRANCH above commercial alternatives. To the best of our knowledge, this is the first system that can consistently generate long videos from books and convert them to visual novel games. The animations and novels facilitate book access to children and second-language learners in a engaging and interactive way. We encourage the readers to check the available demonstration videos: Usage; Video output; Lip Sync; Game.

The pipeline approach of STORYBRANCH can be used as-is to automatically create multimedia content, but can also be modified and extended to new output formats specific to user needs. Our new library BookNLP2 facilitates the creation of multimedia formats and can help both researchers and practitioners. The library is open-source and has been released to the community.

As a future work, we are still working on improving the system across several different directions:

- **animations**: the current animation style is simple and directed towards children and second-language learners. We are currently testing advanced video generation
- **comic books**: we want to add a new output format - comic books. We can reuse the existing STORYBRANCH data
- **extending STORYBRANCH to non-fiction**: Currently STORYBRANCH only works on fiction books with clear descriptions. We want to expand the system to handle non-fiction books using external knowledge such as descriptions of historical figures and events

Acknowledgments

We want to thank Laraib Hasan and Azizur Rahman for their help with implementing Lip Synchronization. We also want to thank our annotators and the anonymous reviewers.

Some of the computations described in this research were performed using the Baskerville Tier 2 HPC service (https://www.baskerville.ac.uk/). Baskerville was funded by the EPSRC and UKRI through the World Class Labs scheme (EP/T022221/1) and the Digital Research Infrastructure programme (EP/W032244/1) and is operated by Advanced Research Computing at the University of Birmingham.

Ethics Statement

STORYBRANCH uses various Generative AI models to produce the final multimedia content. We recognize that generative AI models can contain biases present in their training data. We further acknowledge the risk of generating content that may be offensive or inappropriate. Our hybrid approach to structured extraction and generation aims to reduce the subjectivity in the generation and ground the model on structured representation of characters and events. By explicitly extracting and storing important text properties, we aim to improve the alignment between objectives and output.

References

Giuseppe Attardi, Luca Baronti, Stefano Dei Rossi, and Maria Simi. 2023. Supersense tagging with a maximum entropy markov model. In *SpringerLink Conference*.

David Bamman, Sejal Popat, and Sheng Shen. 2019. An annotated dataset of literary entities. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2138–2144.

Huda Barakat, Oytun Turk, and Cenk Demiroglu. 2024. Deep learning-based expressive speech synthesis: a systematic review of approaches, challenges, and resources. *EURASIP Journal on Audio, Speech, and Music Processing*, 2024(1):11.

Edresson Casanova, Kelly Davis, Eren Gölge, Görkem Göknar, Iulian Gulea, Logan Hart, Aya Aljafari, Joshua Meyer, Reuben Morais, Samuel Olayemi, and Julian Weber. 2024. Xtts: a massively multilingual zero-shot text-to-speech model. In *Interspeech 2024*, pages 4978–4982.

Edresson Casanova, Christian Shulby, Eren Gölge, N. M. Müller, Frederico Santos de Oliveira, Arnaldo Cândido Júnior, André S. Soares, Sandra Aluísio, and Moacir Ponti. 2021. Sc-glowtts: An efficient zero-shot multi-speaker text-to-speech model. In *Interspeech 2021*, pages 3645–3649.

Haoxin Chen, Yong Zhang, Xiaodong Cun, Menghan Xia, Xintao Wang, Chao Weng, and Ying Shan. 2024. Videocrafter2: Overcoming data limitations for high-quality video diffusion models. *Preprint*, arXiv:2401.09047.

Siddhartha Datta, Alexander Ku, Deepak Ramachandran, and Peter Anderson. 2024. Prompt expansion for adaptive text-to-image generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3449–3476, Bangkok, Thailand. Association for Computational Linguistics.

Sejal Popat David Bamman and Sheng Shen. 2020. Booknlp: A natural language processing pipeline for books. GitHub repository.

Venelin Kovatchev and Matthew Lease. 2024. Benchmark transparency: Measuring the impact of data on evaluation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1536–1551, Mexico City, Mexico. Association for Computational Linguistics.

T. Lee et al. 2023. Bivdiff: A training-free framework for general-purpose video synthesis. *arXiv preprint arXiv:2302.02918*.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D. Manning, Christopher Ré, Diana Acosta-Navas, Drew A. Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang, and Yuta Koreeda. 2023. Holistic evaluation of language models. *Preprint*, arXiv:2211.09110.

Juan Linietsky, Ariel Manzur, and the Godot community. 2024. Godot engine: Multi-platform 2d and 3d game engine. Open-source, MIT License.

Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal forced aligner: Trainable text-speech alignment using kaldi. In *Interspeech 2017*, pages 498–502.

Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. 2024. Fast high-resolution image synthesis with latent adversarial diffusion distillation. *Preprint*, arXiv:2403.12015.

Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W. Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda Askell, Amanda Dsouza, Ambrose Slone, Ameet Rahane, Anantharaman S. Iyer, Anders Andreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew Dai, Andrew La, Andrew Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum, Arul Menezes, Arun Kirubarajan, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut Erdem, Ayla Karakaş, B. Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej Bojanowski, Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno Stein, Berk Ekmekci, Bill Yuchen Lin, Blake Howald, Bryan Orinion, Cameron Diao, Cameron Dour, Catherine Stinson, Cedrick Argueta, César Ferri Ramírez, Chandan Singh, Charles Rathkopf, Chenlin Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Chris Waites, Christian Voigt, Christopher D. Manning, Christopher Potts, Cindy Ramirez, Clara E. Rivera, Clemencia Siro, Colin Raffel, Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman, Dan Roth, Daniel Freeman, Daniel Khashabi, Daniel Levy, Daniel Moseguí González, Danielle Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Deniz Yuret, Derek Chen, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho Mollo, Diyi Yang, Dong-Ho Lee, Dylan Schrader, Ekaterina Shutova, Ekin Dogus Cubuk, Elad Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele Rodola, Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan A. Chi, Ethan Dyer, Ethan Jerzak, Ethan Kim, Eunice Engefu Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar, Fernando Martínez-Plumed, Francesca Happé, Francois Chollet, Frieda Rong, Gaurav Mishra, Genta Indra Winata, Gerard de Melo, Germán Kruszewski, Giambattista Parascandolo, Giorgio Mariani, Gloria Wang, Gonzalo Jaimovitch-López, Gregor Betz, Guy Gur-Ari, Hana Galijasevic, Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar, Henry Shevlin, Hinrich Schütze, Hiromu Yakura, Hongming Zhang, Hugh Mee Wong, Ian Ng, Isaac Noble, Jaap Jumelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon Lee, Jaime Fernández Fisac, James B. Simon, James Koppel, James Zheng, James Zou, Jan Kocoń, Jana Thompson, Janelle Wingfield, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason Phang, Jason Wei, Jason Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy Kim, Jeroen Taal, Jesse Engel, Jesujoba Alabi, Jiacheng Xu, Jiaming Song, Jillian Tang, Joan Waweru, John Burden, John Miller, John U. Balis, Jonathan Batchelder, Jonathan Berant, Jörg Frohberg, Jos Rozen, Jose Hernandez-Orallo, Joseph Boudeman, Joseph Guerr, Joseph Jones, Joshua B. Tenenbaum, Joshua S. Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth, Karthik Gopalakrishnan, Katerina Ignatyeva, Katja Markert, Kaustubh D. Dhole, Kevin Gimpel, Kevin Omondi, Kory Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar, Kyle McDonell, Kyle Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin, Lidia Contreras-Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble, Ludwig Schmidt, Luheng He, Luis Oliveros Colón, Luke Metz, Lütfi Kerem Şenel, Maarten Bosma, Maarten Sap, Maartje ter Hoeve, Maheen Farooqi, Manaal Faruqui, Mantas Mazeika, Marco Baturan, Marco Marelli, Marco Maru, Maria Jose Ramírez Quintana, Marie Tolkiehn, Mario Giulianelli, Martha Lewis, Martin Potthast, Matthew L. Leavitt, Matthias Hagen, Mátyás Schubert, Medina Orduna Baitemirova, Melody Arnaud, Melvin McElrath, Michael A. Yee, Michael Cohen, Michael Gu, Michael Ivanitskiy, Michael Starritt, Michael Strube, Michał Swędrowski, Michele Bevilacqua, Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimee Xu, Mirac Suzgun, Mitch Walker, Mo Tiwari, Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhdeh Gheini, Mukund Varma T, Nanyun Peng, Nathan A. Chi, Nayeon Lee, Neta Gur-Ari Krakover, Nicholas Cameron, Nicholas Roberts, Nick Doiron, Nicole Martinez, Nikita Nangia, Niklas Deckers, Niklas Muennighoff, Nitish Shirish Keskar, Niveditha S. Iyer, Noah Constant, Noah Fiedel, Nuan Wen, Oliver Zhang, Omar Agha, Omar Elbaghdadi, Omer Levy, Owain Evans, Pablo Antonio Moreno Casares, Parth Doshi, Pascale Fung, Paul Pu Liang, Paul Vicol, Pegah Alipoormolabashi, Peiyuan Liao, Percy Liang, Peter Chang, Peter Eckersley, Phu Mon Htut, Pinyu Hwang, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour, Priti Oli, Qiaozhu Mei, Qing Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer Gabriel, Rahel Habacker, Ramon Risco, Raphaël Millière, Rhythm Garg, Richard Barnes, Rif A. Saurous, Riku Arakawa, Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman Sitelew, Ronan LeBras, Rosanne Liu, Rowan Jacobs, Rui Zhang, Ruslan Salakhutdinov, Ryan Chi, Ryan Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif M. Mohammad, Sajant Anand, Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel R. Bowman, Samuel S. Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah A. Rous, Sarik Ghazarian, Sayan Ghosh, Sean Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi, Shadi Hamdan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi, Shixiang Shane Gu, Shubh Pachchigar, Shubham Toshniwal, Shyam Upadhyay, Shyamolima, Debnath, Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini, Soo-Hwan Lee, Spencer Torene, Sriharsha Hatwar, Stanislas Dehaene, Stefan Divic, Stefano Ermon, Stella Biderman, Stephanie Lin, Stephen Prasad, Steven T. Piantadosi, Stuart M. Shieber, Summer Misherghi, Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li, Tao Yu, Tariq Ali, Tatsu Hashimoto, Te-Lin Wu, Théo Desbordes, Theodore Rothschild, Thomas Phan, Tianle Wang, Tiberius Nkinyili, Timo Schick, Timofei Kornev, Titus Tunduny, Tobias Gerstenberg, Trenton Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Ramasesh, Vinay Uday Prabhu, Vishakh Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Yiding Hao, Yifu Chen, Yonatan Belinkov, Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J. Wang, Zirui Wang, and Ziyi Wu. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Preprint*, arXiv:2206.04615.

Gayashan Weerasundara and Nisansa de Silva. 2023. Comparative analysis of named entity recognition in the dungeons and dragons domain. *ArXiv*.

Yao Yao, Zuchao Li, and Hai Zhao. 2023. Learning event-aware measures for event coreference resolution. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13542–13556.

Michael Yoder, Sopan Khosla, Qinlan Shen, Aakanksha Naik, Huiming Jin, Hariharan Muralidharan, and Carolyn Rosé. 2021. FanfictionNLP: A text processing pipeline for fanfiction. In *Proceedings of the Third Workshop on Narrative Understanding*, pages 13–23, Virtual. Association for Computational Linguistics.