Proceedings of CONF-CDS 2026 Symposium: Data-Centric AI Security: Securing Models, Learning Agents, and Autonomous Systems
DOI: 10.54254/2755-2721/2026.CH35660

Synthetic Mouse Behavior Video Dataset Generated Using Text-to-Video Models for Animal Behavioral Observation

Yuanyuan Yao

Department of Psychology, University of Southern California, Los Angeles, USA
lisayao@usc.edu

Abstract. Understanding animal behavior is fundamental to psychology, neuroscience, and ethology because behavior provides measurable insights into cognition, emotion, and social interaction. However, automated behavioral annotation systems remain limited by the scarcity of large-scale labeled video datasets and the high cost of manual annotation. In this study, we developed a synthetic mouse behavior video dataset using a text-to-video generative model. The dataset includes single-mouse behaviors (self-grooming, eating, drinking, and rearing) and two-mouse conditions involving social and non-social interactions. Structured prompts were used to control behavioral content while varying camera perspectives and lighting conditions to increase visual diversity. Generated videos exhibited recognizable behavioral patterns across categories and were organized into separate subsets for future supervised learning applications. These findings suggest that generative video models may provide a scalable approach for behavioral dataset development and automated behavioral annotation research.

Keywords: Mouse behavior, Text-to-video generation, Behavioral annotation, Generative AI

1. Introduction

Rodents, especially mice, have become one of the most widely used experimental animals in the field of behavioral science and translational medicine research because of their genetic ease of operation and high correlation with neuropsychiatric disease models. Researchers often use standardized behavioral detection paradigms to explore animal anxiety levels, social interaction patterns, aggressive behavior, stereotyped repetitive behavior and cognitive function performance [6, 7]. Among them, various social interactive behavior paradigms - covering the detection of proximity behavior, chasing behavior, sniffing behavior and aggressive behavior - are particularly important for building animal models of mental illness and neurodevelopmental disorders [6-8].

Although behavioral analysis is crucial, traditional behavioral labeling methods still rely heavily on manual observation and marking. Manual labeling is not only labor-intensive, time-consuming and labor-intensive, but also susceptible to subjective differences of observers, especially for complex behaviors with subtle or rapid changes [4]. As experimental research increasingly produces a large amount of high-resolution video data, the limitations of manual annotation are also becoming more and more prominent [4, 9].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.CH35660

Recent progress in the field of computational behavior has introduced automated solutions for the quantification and classification of behavior [5, 9]. Nowadays, machine learning systems can realize animal posture estimation, motion trajectory tracking and behavior pattern recognition with high accuracy. Take DeepLabCut as an example, the tool uses deep neural networks to achieve "unmarked" posture estimation, which is now widely used in the field of behavioral neuroscience [10]. In addition, other systems such as DeepPoseKit and SLEAP have also further enhanced the ability to robust posture estimation and behavior tracking of multiple animals [11, 12].

Although these tools have significantly improved the level of behavioral analysis, supervised learning methods are still inherently highly dependent on a large amount of annotated training data. Existing data sets are often limited in terms of scale, behavioral diversity or environmental consistency [13]. In addition, it is often particularly difficult to obtain annotated videos containing social interactions and rare behavioral events.

At the same time, the rapid development in the field of generative modeling makes it possible to synthesize realistic images and videos based on text descriptions. The recently emerging video diffusion model and the "text to video" model have shown the ability to generate coherent motion dynamics and structured visual scenes [14, 15]. The Sora system launched by OpenAI further confirms the significant improvement in realism and controllability of generative video technology [16]. These advances indicate a possibility that synthetic behavioral video is expected to serve the relevant applications of behavioral machine learning as an expandable resource.

However, the field of using generative video models to construct animal behavior data sets is still at a relatively unexplored stage. Specifically, it is not clear whether the generated video can reliably capture the unique behavior of rodents and their socially relevant interactions, thus providing effective support for automated behavioral annotation.

In this study, we use the structured "text-to-video" prompt engineering technology to construct a synthetic mouse behavior video data set. The data set covers a variety of behavioral situations involving a single or two animals and is classified into several independent behavioral subsets. By combining controlled prompt design with generative video synthesis technology, this study aims to provide an extensible framework for generating behavior training data sets for research in the fields of psychology, neuroscience and computational animal behavior.

2. Method

2.1. Dataset design

The synthetic data set aims to characterize typical mouse behaviors that are often studied in the fields of psychology, neuroscience and behavioral research [6, 7]. Behavioral categories include self-dressing, eating, drinking water, social interaction and non-social interaction. Each type of behavior is classified into a separate folder to facilitate the management of data sets and future supervised learning applications.

2.2. Prompt development

We have developed prompts for specific behaviors to clarify the target behavior, environmental background, mouse appearance, camera perspective and lighting conditions. All the prompts describe black experimental mice raised in a standard cage environment. In order to increase visual diversity while maintaining behavioral consistency, we have systematically changed the camera perspective and lighting conditions in different prompts. Specific examples include a variety of

DOI: 10.54254/2755-2721/2026.CH35660

observation angles such as front, side, oblique and look-down, as well as a variety of lighting conditions such as diffuse light, balanced light and controlled laboratory lighting.

2.3. Video generation

These videos are generated by using text-to-video generation models. For each behavior category, we have created a variety of hint variants to increase visual diversity and reduce dependence on specific visual configurations. Before being included in the data set, all generated videos are manually reviewed and classified according to their behavioral categories. Studies have shown that similar methods can effectively improve the robustness and diversity of synthetic visual data sets [14-16].

3. Results

3.1. Dataset overview

The symbiosis has formed 426 composite behavior videos covering five types of behaviors. The data set includes self-care (n = 100), eating (n = 100), drinking water (n = 26), double mouse social interaction (n = 100) and double mouse non-social interaction (n = 100) videos. All kinds of behavioral videos are stored in separate folders to support future classification and behavior annotation tasks.

3.2. Behavioral characteristics

Figures 1 to 5 show representative examples of various behaviors. The self-dressing video shows the repeated cleaning of the forelimbs and face (Figure 1); the video of eating and drinking shows the unique feeding behavior near food and water sources (Figure 2 and Figure 3); the social interaction video shows the close distance between animals and showing coordinated movements (Figure 4); not social interaction The video depicts the scene of two animals coexisting in the same environment without direct interaction (Figure 5).

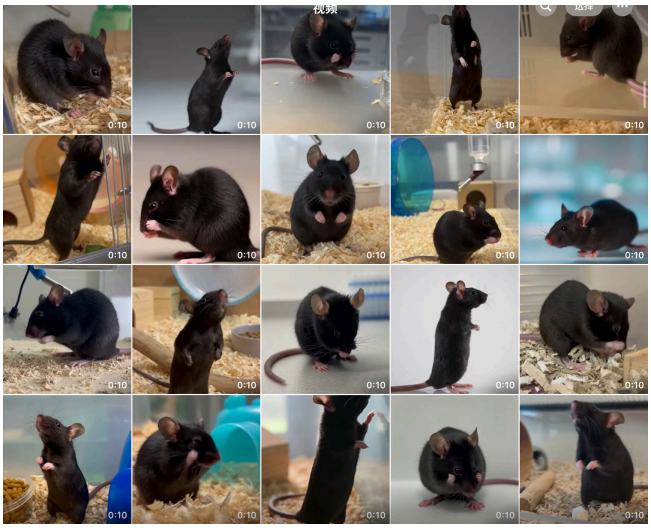

Figure 1. Representative frame of a synthetic mouse displaying self-grooming behavior

DOI: 10.54254/2755-2721/2026.CH35660

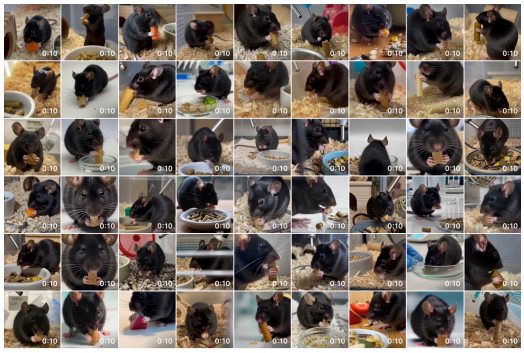

Figure 2. Representative frame of a synthetic mouse displaying eating behavior



Figure 3. Representative frame of a synthetic mouse displaying drinking behavior



Figure 4. Representative frame of two synthetic mice displaying social interaction behavior

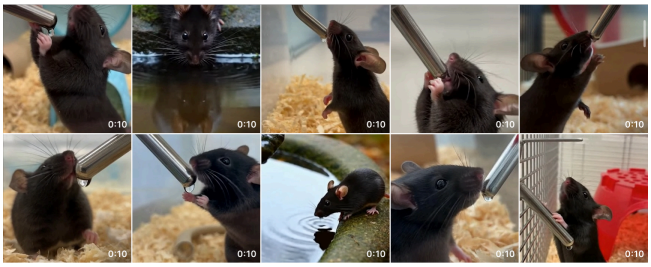

Figure 5. Representative frame of two synthetic mice displaying non-social interaction behavior

In general, the generated videos show recognizable differences between various behaviors. The visual distinction between a single animal and two animal scenarios shows that the "text-to-video" framework can generate behavioral scenes in a controlled laboratory environment. These results

DOI: 10.54254/2755-2721/2026.CH35660

provide preliminary evidence that synthetic behavioral videos are expected to become useful resources in the field of behavioral labeling and machine learning applications.

4. Discussion

This study demonstrates the feasibility of using a generative video model to construct a synthetic behavior data set of laboratory mice. By systematically generating videos covering a variety of behavioral situations, this work provides a preliminary framework for scalable behavioral data set generation and automated annotation research.

An important contribution of this research is the adoption of structured prompt engineering technology, which accurately controls the content of behavior while retaining the diversity of camera perspective and environmental presentation. Previous studies have emphasized that it is crucial to build large-scale and diversified behavioral data sets to achieve reliable automated behavioral classification [9, 5]. However, collecting and manually labeling the behavioral data of rodents is still very expensive and time-consuming, especially in research involving social interaction paradigms and fine-grained behavioral states [4, 5].

This data set also provides a controlled experimental structure, which helps to reduce mixed environmental variability during model training. Because the prompt restricts the background environment, mouse appearance and light characteristics, the behavioral differences between different categories can be maintained relatively independent. This degree of control is often difficult to achieve in laboratory recording in the real world, because in actual experiments, the lighting conditions, cage configuration and camera position tend to change significantly with different experiments [6, 7].

This data set covers both solo and social situations, thus expanding its potential application value. For automated analysis, social behavior is a very challenging field, which is mainly due to the mutual occlusion of individual bodies, dynamic interaction processes, and movement patterns that depend on specific situations [7, 8]. Therefore, the social interactive videos generated by this data set are expected to provide strong support for the relevant research of multi-animal behavior classification and social behavior analysis in the future [12, 10].

Nevertheless, it is still necessary to admit that there are some limitations. First of all, synthetic videos may not be able to fully capture the complexity and variability of real animal behavior. Details such as subtle movement dynamics, spontaneous behavioral changes, and variations in the natural environment may not be fully presented. Secondly, the number of behavioral categories covered by the current data set is limited, and the sample size is unevenly distributed under various conditions, especially in drinking water behavior.

Another limitation is the "domain gap" between synthetic video and real video. Machine learning systems that train only by generating videos may acquire some visual features or motion artifacts with obvious artificial traces, which often cannot be generalized into real laboratory data. Therefore, future research should explore the method of hybrid training to combine synthetic data sets with experimental recording data sets.

Future research can also introduce quantitative verification processes, including attitude estimation consistency analysis, classifier performance evaluation, and expert behavior scoring. With the continuous improvement of the video generation system, the value of synthetic data sets with larger scale and richer behavior diversity will become more and more prominent for research in the fields of psychology, neuroscience and computational animal behavior.

DOI: 10.54254/2755-2721/2026.CH35660

References

[1] Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

[2] Lorenz, K. (1981). *The foundations of ethology*. Springer.

[3] Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron, 93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

[4] Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. *Neuron, 84*(1), 18–31. https://doi.org/10.1016/j.neuron.2014.09.005

[5] Datta, S. R., Anderson, D. J., Branson, K., Perona, P., & Leifer, A. (2019). Computational neuroethology: A call to action. *Neuron, 104*(1), 11–24. https://doi.org/10.1016/j.neuron.2019.09.038

[6] Crawley, J. N. (2007). *What's wrong with my mouse?: Behavioral phenotyping of transgenic and knockout mice*. John Wiley & Sons.

[7] Silverman, J. L., Yang, M., Lord, C., & Crawley, J. N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nature Reviews Neuroscience, 11*(7), 490–502. https://doi.org/10.1038/nrn2851

[8] Koolhaas, J. M., Coppens, C. M., De Boer, S. F., Buwalda, B., Meerlo, P., & Timmermans, P. J. A. (2013). The resident-intruder paradigm: A standardized test for aggression, violence and social stress. *Journal of Visualized Experiments, 77*, 4367. https://doi.org/10.3791/4367

[9] Egnor, S. E. R., & Branson, K. (2016). Computational analysis of behavior. *Annual Review of Neuroscience, 39*, 217–236. https://doi.org/10.1146/annurev-neuro-070815-013845

[10] Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience, 21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

[11] Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., & Couzin, I. D. (2019). DeepPoseKit: A software toolkit for fast and robust animal pose estimation using deep learning. *eLife, 8*, e47994. https://doi.org/10.7554/eLife.47994

[12] Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., Papadoyannis, E. S., Normand, E., Deutsch, D. S., Wang, Z. Y., McKenzie-Smith, G. C., Mitelut, C. C., Castro, M. D., D'Uva, J., Kislin, M., Sanes, D. H., Kocher, S. D., Wang, S. S. H., Falkner, A. L., Shaevitz, J. W., & Murthy, M. (2022). SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods, 19*(4), 486–495. https://doi.org/10.1038/s41592-022-01426-1

[13] Ng, X. L., Ong, K. E., Zheng, Q., Ni, Y., Yeo, S. Y., & Liu, J. (2022). Animal kingdom: A large and diverse dataset for animal behavior understanding. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19023–19034. https://doi.org/10.1109/CVPR52688.2022.01844

[14] Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. (2022). Video diffusion models. *Advances in Neural Information Processing Systems, 35*, 8633–8646.

[15] Villegas, R., Babaeizadeh, M., Kindermans, P.-J., Moraldo, H., Zhang, H., Saffar, M. T., Castro, S., Kunze, J., & Erhan, D. (2022). Phenaki: Variable length video generation from open domain textual description. *arXiv preprint*, arXiv:2210.02399. https://doi.org/10.48550/arXiv.2210.02399

[16] OpenAI. (2024, February 15). Video generation models as world simulators. *OpenAI.* https://openai.com/research/video-generation-models-as-world-simulators