

Javier Rando, Daniel Paleka, David Lindner, Lennart Heim, and Florian Tramèr. Red-teaming the stable diffusion safety filter, 2022.

Charvi Rastogi, Marco Túlio Ribeiro, Nicholas King, and Saleema Amershi. Supporting Human-AI collaboration in auditing LLMs with LLMs. [arXiv preprint arXiv:2304.09991](#), 2023.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding, 2022.

Snorkel. Data-centric AI: A complete primer. URL <https://snorkel.ai/data-centric-ai-primer/>.

Tristan Thrush, Kushal Tirumala, Anmol Gupta, Max Bartolo, Pedro Rodriguez, Tariq Kane, William Gaviria Rojas, Peter Mattson, Adina Williams, and Douwe Kiela. Dynatask: A framework for creating dynamic AI benchmark tasks. In [Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations](#), pages 174–181, 2022.

Joaquin Vanschoren and Serena Yeung. Announcing the NeurIPS 2021 Datasets and Benchmarks Track | by Neural Information Processing Systems conference | medium. <https://neuripsconf.medium.com/announcing-the-neurips-2021-datasets-and-benchmarks-track-644e27c1e66c>, 2021. (Accessed on 04/19/2023).

Bertie Vidgen, Tristan Thrush, Zeerak Waseem, and Douwe Kiela. Learning from the worst: Dynamically generated datasets to improve online hate detection. In [Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing \(Volume 1: Long Papers\)](#), pages 1667–1682, 2021.

Chris Welty, Praveen Paritosh, and Lora Aroyo. Metrology for AI: From benchmarks to instruments. [arXiv preprint arXiv:1911.01875](#), 2019.

Guillaume Wenzek, Vishrav Chaudhary, Angela Fan, Sahir Gomez, Naman Goyal, Somya Jain, Douwe Kiela, Tristan Thrush, and Francisco Guzmán. Findings of the WMT 2021 shared task on large-scale multilingual machine translation. In [Proceedings of the Sixth Conference on Machine Translation](#), pages 89–99, 2021.