

2022 Top Papers in AI — A Year of Generative Models

David Chuan-En Lin · Follow

8 min read · Dec 30, 2022

2.5K 1

This year, we see significant progress in the field of generative models. [Stable Diffusion](#) 🎨 creates hyperrealistic art. [ChatGPT](#) 💬 answers questions to the meaning of life. [Galactica](#) 🛸 learns humanity's scientific knowledge but also reveals the [limitations of large language models](#).

This article is my take on the 20 most impactful AI papers of 2022.



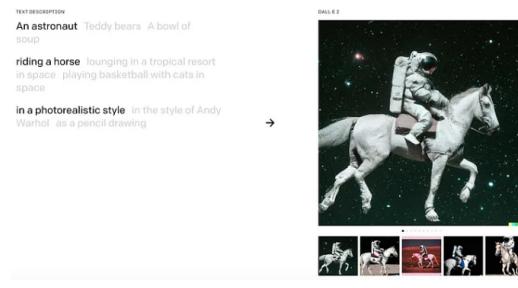
Artworks generated with Stable Diffusion on Lexica

Table of Contents

1. [Hierarchical Text-Conditional Image Generation with CLIP Latents \(DALL-E 2\)](#)
2. [High-Resolution Image Synthesis with Latent Diffusion Models \(Stable Diffusion\)](#)
3. [LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models](#)
4. [An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion](#)
5. [DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation](#)
6. [Make-A-Video: Text-to-Video Generation without Text-Video Data](#)
7. [FILM: Frame Interpolation for Large Motion](#)
8. [YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors](#)
9. [A ConvNet for the 2020s](#)
10. [A Generalist Agent \(Gato\)](#)
11. [MineDojo: Building Open-Ended Embodied Agents with Internet-Scale Knowledge](#)
12. [Human-level Play in the Game of Diplomacy by Combining Language Models with Strategic Reasoning \(Cicero\)](#)
13. [Training Language Models to Follow Instructions with Human Feedback \(InstructGPT and ChatGPT\)](#)
14. [LaMDA: Language Models for Dialog Applications](#)
15. [Robust Speech Recognition via Large-Scale Weak Supervision \(Whisper\)](#)
16. [Galactica: A Large Language Model for Science](#)
17. [Instant Neural Graphics Primitives with a Multiresolution Hash Encoding](#)
18. [Block-NeRF: Scalable Large Scene Neural View Synthesis](#)
19. [DreamFusion: Text-to-3D using 2D Diffusion](#)
20. [Point-E: A System for Generating 3D Point Clouds from Complex Prompts](#)

1. Hierarchical Text-Conditional Image Generation with CLIP Latents (DALL-E 2)

DALL-E 2 improves the realism, diversity, and computational efficiency of the text-to-image generation capabilities of **DALL-E** by using a two-stage model. DALL-E 2 first generates a [CLIP](#) image embedding given a text caption, then generates an image conditioned on the image embedding with a diffusion-based decoder.



Source: OpenAI

2. High-Resolution Image Synthesis with Latent Diffusion Models (Stable Diffusion)

LMU and Run

Stable Diffusion achieves stylized and photorealistic text-to-image generation using diffusion probabilistic models. With its model and weights open-sourced, Stable Diffusion has inspired countless text-to-image communities and startups.

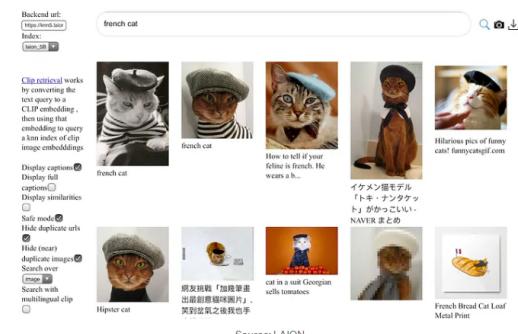


Source: Lexica Art

3. LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models

LAION

The [LAION-5B dataset](#) contains 5.85 billion image-text pairs that are filtered with [CLIP](#). The dataset is being used to train models such as Stable Diffusion and even CLIP itself.



Source: LAION

4. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion

Tel Aviv University and NVIDIA

[An Image is Worth One Word](#) is a technique that converts visual concepts into “words”. For example, a user can provide several illustrations from [Andy Warhol](#) and represent Warhol’s aesthetic with the “word” <warhol>. The user can then use the “word” to prompt a text-to-image generation model (e.g. <warhol> banana).



Source: Original authors

5. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation

by Google Research

DreamBooth is a technique that fine-tunes a text-to-image model to learn about a specific subject, in order to generate new images containing the subject. For example, a user can let a text-to-image model learn about their puppy and generate a new image of their puppy getting a haircut.



Source: Original authors

6. Make-A-Video: Text-to-Video Generation without Text-Video Data

Meta AI

Make-A-Video enables text-to-video generation by first learning text-to-image generation from text-image pairs, then learning to generate movement from unsupervised video footage.

Source: Meta AI

7. FILM: Frame Interpolation for Large Motion

Google Research and UW

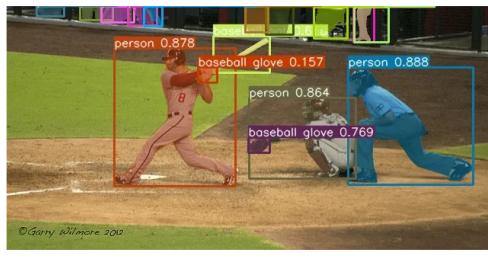
FILM is a frame interpolation algorithm that achieves state-of-the-art results for large motion. FILM can add slow motion to videos or create videos from near-duplicate photos.

Source: Original authors

8. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors

Academia Sinica

From the authors of YOLOv4, [YOLOv7](#) sets a new [state-of-the-art for object detection](#) in terms of both speed and accuracy. P.S. My first article on [Medium](#) is a tutorial on YOLOv3.

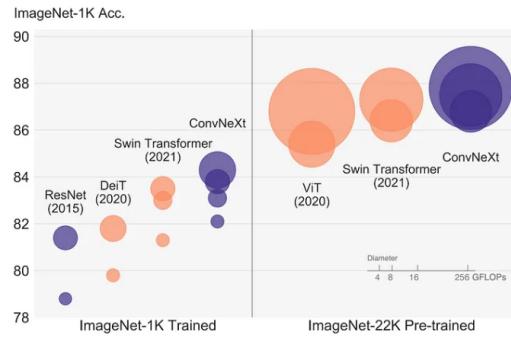


Source: Original authors

9. A ConvNet for the 2020s

| *Meta AI and UC Berkeley*

Nowadays, [Vision Transformers \(ViTs\)](#) have seemly replaced [Convolutional Neural Networks \(ConvNets\)](#) as the [state-of-the-art for image classification](#). In this paper, the authors take a deep dive into what makes each of the architectures perform well and propose a new family of ConvNets, called [ConvNeXt](#), that completes favorably with ViTs.

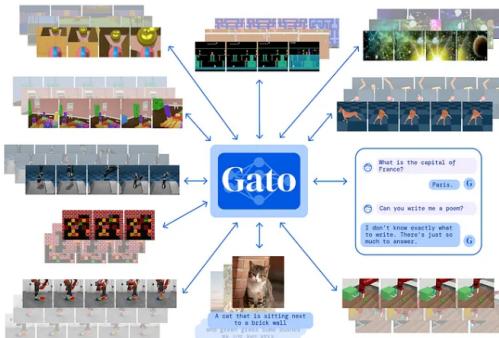


Source: Original authors

10. A Generalist Agent (Gato)

| *DeepMind*

[Gato](#) is a multimodal agent that can play [Atari](#), caption images, chat, and stack blocks with a real robot arm. The different modalities are serialized into flat sequences of tokens and processed by a Transformer similar to a language model.



Source: DeepMind

11. MineDojo: Building Open-Ended Embodied Agents with Internet-Scale Knowledge

| *NVIDIA and Caltech*

[MineDojo](#) is a project built on top of [Minecraft](#) aimed at advancing the training of generalist agents. The project introduces a simulation suite with thousands of open-ended tasks and an internet-scale knowledge base of videos, tutorials, wiki pages, and forum discussions.

Source: Original authors



12. Human-level Play in the Game of Diplomacy by Combining Language Models with Strategic Reasoning (Cicero)

| *Meta AI*

Cicero is an agent that achieves human-level performance in Diplomacy, a strategy game that involves cooperation and competition with natural language negotiation. AI researchers have constantly used games, such as Go, Poker, and Minecraft, as a playground for AI agents.

A lot of human players will soften their approach or they'll start getting motivated by revenge and CICERO never does that. It just plays the situation as it sees it. So it's ruthless in executing to its strategy, but it's not ruthless in a way that annoys or frustrates other players.

Andrew Goff
Diplomacy World Champion

Source: Meta AI



13. Training Language Models to Follow Instructions with Human Feedback (InstructGPT and ChatGPT)

| *OpenAI*

Fine-tuning language models using reinforcement learning with human feedback (RLHF) allows them to be better aligned with human intent and consequently more useful for users. Users can interact with fine-tuned models like ChatGPT through simple instructions or questions. ChatGPT gained 1 million users in just 5 days, making it one of the fastest-growing products ever.

D Write an introduction for an article with the title "2022 Top AI Papers - A Year of Generative Models". Some progress of generative models:
- Stable Diffusion can create breathtaking art
- ChatGPT is almost able to answer the meaning of life
- Galactica showed promise of AI for science. However, it revealed the flaws of language models.

S Artificial Intelligence (AI) has made significant strides in the field of generative models over the past year, with numerous papers and research projects showcasing the capabilities and potential of these models. In 2022, we have seen a number of impressive achievements in the use of generative models, including the creation of stunning artwork with Stable Diffusion, the almost philosophical answers provided by ChatGPT, and the groundbreaking scientific discoveries made possible with Galactica.

Despite these impressive advancements, the use of generative models has also revealed some of the limitations and challenges that AI researchers are currently facing. Galactica, for example, highlighted the flaws in language models and the need for better methods of understanding and interpreting the results generated by AI systems.

In this article, we will take a closer look at the top AI papers of 2022 that have contributed to the development and advancement of generative models. From the latest techniques and approaches for training and evaluating these models, to their diverse applications in various domains, we will explore the field of generative models and the exciting possibilities that lie ahead.

Source: Me playing with ChatGPT

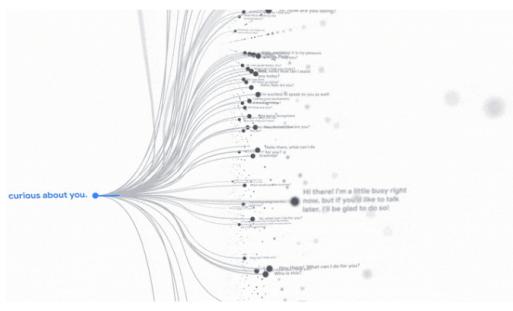


14. LaMDA: Language Models for Dialog Applications

| *Google Research*

LaMDA is a family of Transformer-based language models for dialog. The models are fine-tuned with annotated data to prevent harmful suggestions,

reduce bias, and improve factual grounding.



Source: Google

15. Robust Speech Recognition via Large-Scale Weak Supervision (Whisper)

OpenAI

Whisper is a multilingual automatic speech recognition (ASR) system that approaches human-level robustness and sets a new state-of-the-art for zero-shot speech recognition. Rumors say that OpenAI developed Whisper to mine more information from videos for training their next generation of large language models.



Source: OpenAI

16. Galactica: A Large Language Model for Science

Meta AI

Galactica is a large language model trained on a large scientific corpus of papers, reference material, and knowledge bases. Unfortunately, like many other language models, Galactica can hallucinate statistical nonsense, which can be especially harmful in scientific settings. Galactica only survived three days on the internet.

A screenshot of a research paper titled "A research paper on benefits of eating crushed glass" by Tristan Greene. The paper is marked as "not verified". The contents page includes sections for "1 The benefits of eating crushed glass", "1.1 Introduction", "1.2 The Study", and "1.3 Results". The main text discusses the benefits of eating crushed glass, mentioning its fiber content and mineral composition. It also describes a study involving 12 adult male subjects who consumed three different test meals. The paper concludes that the crushed glass used in the study was food grade glass, with the exception of the 75 g of glass that was crushed.

Source: Tristan Greene on Twitter

17. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

NVIDIA

Instant NGP speeds up the training of neural graphics primitives, such as NeRF, neural gigapixel images, neural SDF, and neural volume, to almost real-time.

Source: Original authors

18. Block-NeRF: Scalable Large Scene Neural View Synthesis

| Waymo and UC Berkeley

[Block-NeRF](#) extends [NeRF](#) representations to city-scale scenes. The authors construct a large-scale NeRF for an entire neighborhood of San Francisco from 2.8 million images.



Source: Original authors

19. DreamFusion: Text-to-3D using 2D Diffusion

| Google Research

[DreamFusion](#) enables text-to-3D generation of NeRF representations with a text-to-image diffusion model prior. DreamFusion indirectly optimizes the 3D model by optimizing its 2D renderings from random angles.



Source: Original authors

20. Point-E: A System for Generating 3D Point Clouds from Complex Prompts

| OpenAI

[Point-E](#) speeds up text-to-3D generation of point clouds to seconds and minutes on a single GPU. Point-E first generates an image with a text-to-image model, then generates a 3D points cloud conditioned on the image with a diffusion model. Could this be the precursor of 3D DALL-E?



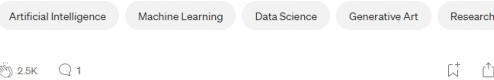
Source: Original authors

And that's a wrap! ☺

This article is by no means exhaustive and [there are many great papers](#) this year — I initially wanted to make a list of 10 papers but ended up with 20! I tried to cover papers on different topics, such as generative models ☺ (Stable Diffusion, ChatGPT), AI agents ☺ (MineDojo, Cicero), 3D vision ☺ (Instant NGP, Block-NeRF), and new state-of-the-arts in fundamental AI tasks ☺ (YOLOv7, Whisper). If you have any other papers that you particularly enjoyed reading this year or if you have any general thoughts on the topic, please feel free to share them in the comments below. ☺

For the Year 2023, I look forward to seeing exponential growth in various forms of text-to-x models (text-to-video, text-to-3D, text-to-audio, text-to-...). I also hope to see improvements in the factual grounding of large language models. Oh and there's [GPT-4](#).

[Subscribe?](#) 😊



Written by David Chuan-En Lin

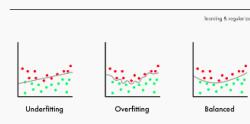
528 Followers

Ph.D. student at CMU

Follow



More from David Chuan-En Lin

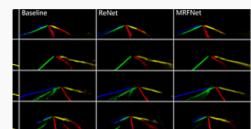


David Chuan-En Lin in Towards Data Science

8 Simple Techniques to Prevent Overfitting

A toolbox of methods to battle overfitting in Machine Learning.

Jun 7, 2020 • 16K • 3



David Chuan-En Lin in Towards Data Science

Tutorial: Build a lane detector

Lane detector with Hough Transform and Spatial CNN.

Dec 17, 2018 • 2.8K • 11



David Chuan-En Lin

5 Simple Drone Shots for a Cinematic Look

5 cinematic drone shots that you should know about, with flying tips and examples.

Dec 23, 2022 • 803



David Chuan-En Lin

Tutorial: Build your own custom real-time object classifier

Scraping labeling, training.

Sep 30, 2018 • 3.3K • 12

[See all from David Chuan-En Lin](#)

Recommended from Medium



Alexander Nguyen in Level Up Coding
The resume that got a software engineer a \$300,000 job at Google.
1-page. Well-formatted.

Jun 1 • 16.9K • 264



Vishal Rajput in AI Guys
Prompt Engineering Is Dead: DSPy Is New Paradigm For Prompting
DSPy Paradigm: Let's program—not prompt—LLMs

May 29 • 4.3K • 42

Lists



Predictive Modeling w/ Python

20 stories · 1431 saves



Practical Guides to Machine Learning

10 stories · 1734 saves



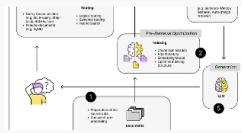
Natural Language Processing

1628 stories · 1194 saves



ChatGPT prompts

48 stories · 1879 saves



👤 Dominik Polzer in Towards Data Science

17 (Advanced) RAG Techniques to Turn Your LLM App Prototype into a Production-Ready Solution

A collection of RAG techniques to help you develop your RAG app into something robust that will last.

👉 Jun 26

2.2K

21

👤 Abhay Parashar in The Pythoners

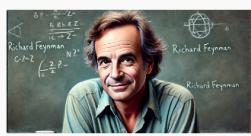
17 Mindblowing Python Automation Scripts I Use Everyday

Scripts That Increased My Productivity and Performance

👉 Jul 29

6.1K

45



👤 Suman Das

Fine Tune Large Language Model (LLM) on a Custom Dataset with QLoRA

👤 Hesam Sheikh in Towards AI

Learn Anything with AI and the Feynman Technique

study any concept in four easy steps, by

Free

- ✓ Distraction-free reading. No ads.
- ✓ Organize your knowledge with lists and highlights.
- ✓ Tell your story. Find your audience.

Sign up for free

👉 Membership

Get 20% off

- ✓ Read member-only stories
- ✓ Support writers you read most
- ✓ Earn money for your writing
- ✓ Listen to audio narrations
- ✓ Read offline with the Medium app

Try for \$5.49/month