Checkpoint recovery targets

35 historical targets across image, audio, video and language. Original research models and proprietary releases.

This is a recovery and preservation shortlist, not a list of models ready to host. It covers original checkpoints that we have not located and proprietary releases without an identified public weight source. It excludes models already retained or with clearly located complete checkpoints merely awaiting routine download.

Reading the dates. Entries run oldest to newest within each modality. Dates refer to the cited paper, announcement or version release; they are not checkpoint acquisition dates.

Reading the gaps. "No public weights identified" does not mean the creator has lost the model. "Original checkpoint unlocated" means our search has not established an authentic source. These are bounded research findings, not proof that no copy exists anywhere.

Separate verification list. 3 additional entries have publisher links or known components but unresolved access, exact-version or dependency questions. They appear after the main list and are not counted as missing weights.

Sources are clickable beside each entry. No model binaries were downloaded, inference run or contacts approached for this document.

Image	9 targets
Audio	7 targets
Video	9 targets
Language models	10 targets

Scope: selected historical releases, mainly through 2023, with the original 2024 Sora research model included. Later products with the same name are not substitutes for the specified version.

01 Image

Oldest to newest. Original checkpoints and proprietary releases.

Jul 2006

Generative deep belief network - original digit model

Geoffrey Hinton, Simon Osindero and Yee-Whye Teh / University of Toronto | ORIGINAL CHECKPOINT UNLOCATED

July 2006 · journal publication

Combined an associative top layer with directed lower layers to learn and sample a multilayer model of handwritten digits.

Gap: The author's digit demonstration and paper are available, but no original trained generative checkpoint was identified in the inspected sources or local archive. Hinton's separate Science autoencoder MATLAB project is not this model.

Sources: Author's generative digit demonstration / A Fast Learning Algorithm for Deep Belief Nets / Publication record

10 Jun 2014

Original GAN - 2014 paper checkpoints

Ian Goodfellow and coauthors / Université de Montréal | ORIGINAL CHECKPOINT UNLOCATED

10 June 2014 · first paper submission

Introduced adversarial training of a generator against a discriminator.

Gap: The author repository contains source and training configurations, but an original trained paper checkpoint was not identified. Its sampling scripts expect a serialized model; the retained 2015 DCGAN bedroom weights are a different model.

Sources: Generative Adversarial Networks - submission history / Author's original source and hyperparameters

5 Jan 2021

DALL·E 1 - full text-to-image transformer

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

5 January 2021 · announcement

Applied a large autoregressive transformer to a joint stream of text and discrete image tokens.

Gap: The official release provides the discrete VAE, and explicitly excludes the text-to-image transformer. No public original transformer checkpoint was identified; the released image tokenizer alone cannot reproduce DALL·E 1.

Sources: DALL·E announcement / Official release scope: discrete VAE only

20 Dec 2021

GLIDE - full research model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

20 December 2021 · first paper submission

Demonstrated text-conditioned diffusion generation and editing with classifier-free guidance.

Gap: The public author repository releases the small model trained on filtered data, not the full research model evaluated in the paper. The full model's original checkpoint was not identified publicly. GLIDE filtered remains a separate, locatable candidate and is excluded from this missing-weight target.

Sources: GLIDE paper and submission history / Official small filtered-model release

Feb 2022

Midjourney V1 and preceding internal prototypes

Midjourney | NO PUBLIC WEIGHTS IDENTIFIED

February 2022 · V1 default-model period begins

An early widely used service with a recognizable painterly generation style and a shared prompt-and-variation workflow.

Gap: Publisher documentation dates V1 to February-April 2022 and describes legacy service controls, but no public downloadable V1 checkpoint was identified. Pre-V1 prototype identities and dates remain unverified; service access would not provide an independent checkpoint archive.

Sources: Midjourney's legacy model history / Publisher's model-version documentation

6 Apr 2022

DALL·E 2 - original preview model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

6 April 2022 · documented preview system

Used a prior over CLIP image representations followed by diffusion decoding, supporting generation and image variations.

Gap: The publisher documents controlled access to the preview system, not a downloadable original checkpoint. No public release of the complete original prior/decoder pipeline was identified. Public CLIP weights or community reproductions do not establish recovery of DALL·E 2.

Sources: Original preview system card / Hierarchical Text-Conditional Image Generation with CLIP Latents

23 May 2022

Imagen - original 2022 research model

Google Research | NO PUBLIC WEIGHTS IDENTIFIED

23 May 2022 · first paper submission

Combined a frozen language encoder with a cascade of diffusion models for text-to-image generation.

Gap: The original project explicitly withheld the model from public release. No original checkpoint was identified. Later Imagen products and independently trained implementations are not evidence that the 2022 research cascade has been preserved.

Sources: Original Imagen project and release statement / Imagen paper and submission history

22 Jun 2022

Parti - original 20B research model

Google Research | NO PUBLIC WEIGHTS IDENTIFIED

22 June 2022 · first paper submission

Scaled autoregressive image-token generation to 20 billion parameters and evaluated complex compositional prompts.

Gap: The author project explicitly declined public release of its models. PartiPrompts is a published evaluation prompt set, not generator weights; no original 20B checkpoint was identified.

Sources: Parti project and release statement / Parti paper and submission history

3 Oct 2023

DALL·E 3 - original deployment-era model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

3 October 2023 · system-card publication

Improved image-caption fidelity and integrated text-to-image generation with conversational prompt preparation.

Gap: The publisher documents a deployed text-to-image system, but no public original checkpoint was identified. Hosted access and examples do not preserve the generator independently, and prompt rewriting can obscure the raw model input.

Sources: DALL·E 3 system card / DALL·E 3 product description

02 Audio

Oldest to newest. Original checkpoints and proprietary releases.

8 Sep 2016

WaveNet - original speech and music models

DeepMind / Google DeepMind | ORIGINAL CHECKPOINT UNLOCATED

8 Sep 2016 announcement; paper v1: 12 Sep 2016

Established sample-level neural waveform generation as a practical research direction for speech and music; a central predecessor to neural vocoders and subsequent audio generators.

Gap: The original DeepMind checkpoints behind the 2016 demonstrations were not located. Later community WaveNet implementations and Magenta's NSynth WaveNet checkpoint are different trained artifacts.

Sources: DeepMind: WaveNet announcement, 8 September 2016 / Original WaveNet paper and version history

29 Mar 2017

Tacotron - original Google checkpoint

Google | ORIGINAL CHECKPOINT UNLOCATED

29 Mar 2017 · paper v1

An influential text-to-speech system that learned the mapping from characters to spectrograms, reducing dependence on the separately engineered stages of earlier speech systems.

Gap: No original Google checkpoint was identified. Keith Ito's available tacotron-20180906.tar.gz belongs to an independent implementation and later training run; it cannot establish access to the 2017 model.

Sources: Tacotron original paper and version history / Keith Ito's independent implementation and checkpoint provenance

16 Dec 2017

Tacotron 2 - original Google model and WaveNet vocoder

Google | ORIGINAL CHECKPOINT UNLOCATED

16 Dec 2017 · paper v1; ICASSP 2018

Combined learned mel-spectrogram prediction with neural waveform synthesis and became a widely used reference design for high-quality text-to-speech.

Gap: No original Google checkpoint pair was identified. NVIDIA publishes Tacotron 2 and WaveGlow weights, but that implementation uses LJ Speech and a different vocoder from the original paper's modified WaveNet.

Sources: Tacotron 2 original paper and version history / NVIDIA Tacotron 2 implementation and pretrained links

7 Sep 2022

AudioLM - original speech and piano continuation models

Google Research / Google DeepMind | ORIGINAL CHECKPOINT UNLOCATED

7 Sep 2022 · paper v1

Combined semantic and acoustic audio tokens to continue speech and piano recordings without transcripts or symbolic music input; an important bridge from language modeling to general audio generation.

Gap: The publisher's paper and demonstration page were located, but no original runnable checkpoint set. Independent AudioLM code is not evidence that Google's trained models are available.

Sources: AudioLM paper and original submission date / Google Research AudioLM examples / Google Research SEANet publisher repository

5 Jan 2023

VALL-E - original January 2023 model

Microsoft Research | ORIGINAL CHECKPOINT UNLOCATED

5 Jan 2023 · paper v1

A prominent early demonstration of codec language modeling for text-to-speech with a short voice prompt, showing how large-scale audio-token prediction could reproduce unfamiliar speakers.

Gap: No original Microsoft checkpoint was identified. The official UNILM directory still points to the preprint and demonstration page; similarly named community implementations and later VALL-E variants are distinct artifacts.

Sources: VALL-E original paper and submission history / Microsoft UNILM VALL-E directory / Microsoft Research VALL-E demonstration

26 Jan 2023

MusicLM - original research model

Google Research / Google DeepMind | NO PUBLIC WEIGHTS IDENTIFIED

26 Jan 2023 · paper v1

Demonstrated text- and melody-conditioned music generation with sustained musical structure, extending AudioLM from continuation to descriptive control over varied music.

Gap: The January 2023 paper explicitly says the authors had no plans to release the models at that time. This audit found no later original checkpoint release. MusicCaps is a released dataset, not MusicLM weights; a hosted demo does not provide an archival checkpoint.

Sources: MusicLM v1, section 8: explicit non-release statement / MusicLM original paper date / Google Research MusicLM examples and MusicCaps distinction

16 May 2023

SoundStorm - original Google model

Google Research / Google DeepMind | ORIGINAL CHECKPOINT UNLOCATED

16 May 2023 · paper v1

Replaced AudioLM's slow acoustic-token generation with parallel decoding while preserving voice consistency, and demonstrated extended dialogue synthesis when paired with semantic conditioning.

Gap: Publisher examples and the paper are available, but no authentic original checkpoint was identified. Community SoundStorm implementations cannot substitute for the original training result.

Sources: SoundStorm original paper and submission date / Google Research SoundStorm examples

03 Video

Oldest to newest. Original checkpoints and proprietary releases.

21 Nov 2016

TGAN - original trained generators

Preferred Networks / Masaki Saito, Eiichi Matsumoto and Shunta Saito | ORIGINAL CHECKPOINT UNLOCATED

21 November 2016 preprint; ICCV 2017

An early video GAN that separated a temporal latent generator from the generator that renders individual frames.

Gap: Original training and inference code survives, but an author-issued pretrained generator was not located in the reviewed repository and project page. The linked mean2.npz is a preprocessing artifact, not the missing generator. This is an unlocated checkpoint, not evidence that the authors never released one.

Sources: TGAN paper and original submission date / Authors' TGAN repository / Authors' TGAN project page

17 Jul 2017

MoCoGAN - original trained generators

NVIDIA / Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang and Jan Kautz | ORIGINAL CHECKPOINT UNLOCATED

17 July 2017 preprint; CVPR 2018

Made content and motion separately controllable latent variables: the same subject could be generated with different motion, or the same motion with different subjects.

Gap: The official repository and training wiki provide code and instructions that expect a trained model, but an original pretrained generator download was not located. Surviving source code is not proof that the original trained weights are available.

Sources: MoCoGAN paper and submission date / Authors' MoCoGAN repository / Official training and generation instructions

15 Jul 2019

DVD-GAN

DeepMind / Aidan Clark, Jeff Donahue and Karen Simonyan | ORIGINAL CHECKPOINT UNLOCATED

15 July 2019 preprint

Scaled video GANs to more complex Kinetics-600 scenes using separate spatial and temporal discriminators, bridging small early video experiments and later general-purpose generation.

Gap: The paper survives, but no official original generator checkpoint or complete inference package was located in the reviewed primary sources. There is insufficient evidence here to claim that no original checkpoint was ever distributed privately or publicly.

Sources: Adversarial Video Generation on Complex Datasets

29 Sep 2022

Make-A-Video

Meta AI | NO PUBLIC WEIGHTS IDENTIFIED

29 September 2022 announcement

A prominent early text-to-video system that combined learning from text-image pairs with motion learned from unlabelled video, without requiring paired text-video training for the core approach.

Gap: The original announcement released research details and proposed a future demo, rather than a checkpoint package. No public original weights were found in the reviewed official announcement and project site; an API or a third-party reimplementation would not preserve this model.

Sources: Meta's dated Make-A-Video announcement / Original Make-A-Video project

5 Oct 2022

Imagen Video

Google Research, Brain Team | NO PUBLIC WEIGHTS IDENTIFIED

5 October 2022 preprint

Demonstrated a text-conditioned cascade of video diffusion and temporal/spatial super-resolution models producing 128 frames at 1280 × 768 and 24 fps.

Gap: The official project explicitly says the model and source code are not being released until its stated safety concerns are mitigated. No subsequent original checkpoint release was established in this audit.

Sources: Imagen Video official project and non-release statement / Imagen Video paper and submission date

5 Oct 2022

Phenaki - original Google model

Google Research / Ruben Villegas and collaborators | NO PUBLIC WEIGHTS IDENTIFIED

5 October 2022 preprint; ICLR 2023

Combined a causal video tokenizer with a masked transformer to generate variable-length video from a changing sequence of text prompts.

Gap: The original paper explicitly withholds the models, code, data and interactive demo at publication. No later official release was established. Community Phenaki implementations or C-ViViT checkpoints should not be represented as Google's original complete model.

Sources: Phenaki paper and submission date / ICLR paper, including model-release decision

Feb 2023

Runway Gen-1

Runway | NO PUBLIC WEIGHTS IDENTIFIED

February 2023 research introduction

An early widely noticed diffusion video-editing system that preserved the structure of an input video while changing its appearance using text or image guidance.

Gap: The original model was offered as a hosted Runway system. The official research page links demonstrations and a paper, but no original weight package was found. Hosted product access does not provide an independently runnable archive.

Sources: Runway's Gen-1 research introduction

Mar 2023

Runway Gen-2

Runway | NO PUBLIC WEIGHTS IDENTIFIED

March 2023 research introduction; June 2023 web/mobile availability

One of the early public text-to-video products; its retirement is a concrete example of a historically important generator disappearing while its examples and name remain online.

Gap: No public original checkpoint release was found. Runway confirms that the hosted Gen-2 model became unavailable to all users on 11 May 2025. The current Gen-2 landing page displays an inconsistent February date; March is supported by Runway's 22 March 2023 announcement, and June availability by its retrospective.

Sources: Runway's 22 March 2023 research-release reference / Runway retrospective: June web/mobile release / Official Gen-2 retirement notice

15 Feb 2024

Sora - February 2024 research model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

15 February 2024 research preview

A major demonstration of diffusion transformers operating on spacetime latent patches, generating video across durations, resolutions and aspect ratios, including minute-long examples.

Gap: No public original checkpoint package was found. The research report explicitly omits model and implementation details. This target is the February 2024 research model, not the later Sora Turbo product or subsequent Sora versions; access to a later service would not establish preservation of the original.

Sources: Original Sora research report, 15 February 2024 / Sora Turbo product release, distinct from the research preview

04 Language models

Oldest to newest. Original checkpoints and proprietary releases.

21 May 2015

char-rnn - original demonstration models

Andrej Karpathy | ORIGINAL CHECKPOINT UNLOCATED

21 May 2015 · author article

A widely circulated example of character-level recurrent text generation before Transformer language models.

Gap: Author code and training examples survive, but an authentic author-trained demonstration checkpoint with its character vocabulary was not identified. A training-output filename in the README is not a downloadable trained model.

Sources: Original demonstration article / Author repository

28 May 2020

GPT-3 - original 175B base model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

28 May 2020 · paper announcement

Established large-scale few-shot prompting as an alternative to task-specific fine-tuning.

Gap: The original trained checkpoint was not identified in public publisher sources. API access and later models using the GPT-3 name do not preserve the exact 2020 base model.

Sources: Publisher paper announcement / Official research repository

18 May 2021

LaMDA - original dialogue research model

Google | NO PUBLIC WEIGHTS IDENTIFIED

18 May 2021 · announcement

A dialogue-specialized Transformer presented through open-ended conversation demonstrations.

Gap: The original research model is documented, but no public original checkpoint was identified. LaMDA 2, early Bard and later Gemini systems are distinct versions or products.

Sources: Google's original announcement

8 Dec 2021

Gopher - 280B research model

DeepMind | NO PUBLIC WEIGHTS IDENTIFIED

8 December 2021 · research announcement

An early large-model scaling study spanning models from millions to hundreds of billions of parameters.

Gap: The publisher provides research and results, but an original public 280B checkpoint was not identified. The paper alone is insufficient for independent reproduction of the trained model.

Sources: DeepMind research announcement

29 Mar 2022

Chinchilla - original 70B model

DeepMind | NO PUBLIC WEIGHTS IDENTIFIED

29 March 2022 · first paper submission

Demonstrated the importance of balancing parameter count and training tokens under a fixed compute budget.

Gap: No original public Chinchilla checkpoint was identified. Later models following its compute-optimal training recommendations are not copies of the research model.

Sources: Compute-optimal training paper / Publisher analysis

4 Apr 2022

PaLM - original 540B model

Google Research | NO PUBLIC WEIGHTS IDENTIFIED

4 April 2022 · research announcement

A large dense language model used to study scaling and few-shot reasoning with the Pathways training system.

Gap: No public original 540B checkpoint was identified. Later PaLM APIs, Flan-PaLM instruction variants and PaLM 2 do not establish recovery of the original base model.

Sources: Google Research announcement

30 Nov 2022

ChatGPT - November 2022 launch model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

30 November 2022 · public research preview

Brought instruction-following dialogue and human-feedback training to a widely used public interface.

Gap: No public checkpoint of the exact launch model was identified. The continuously updated ChatGPT service, later GPT-3.5 API snapshots and saved conversations do not recover that original system.

Sources: Original ChatGPT introduction

14 Mar 2023

Claude and Claude Instant - first release

Anthropic | NO PUBLIC WEIGHTS IDENTIFIED

14 March 2023 · announcement

An early alternative conversational assistant release informed by research on model alignment.

Gap: The publisher offered chat/API access to two models, but no public original weights were identified. Later Claude generations cannot substitute for the initial Claude and Claude Instant checkpoints.

Sources: Original Claude introduction

14 Mar 2023

GPT-4 - original launch-era model

OpenAI | NO PUBLIC WEIGHTS IDENTIFIED

14 March 2023 · announcement

A text-generating model with image-input research capabilities and broader instruction-following performance.

Gap: The launch release provided service/API access, not public original weights. The exact launch checkpoint and image-input research configuration remain separate preservation targets from later GPT-4 versions.

Sources: Original GPT-4 release

6 Dec 2023

Gemini 1.0 - initial Pro and Ultra family

Google DeepMind | NO PUBLIC WEIGHTS IDENTIFIED

6 December 2023 · family announcement

Google's initial Gemini family brought multimodal training into its main language-model product line.

Gap: No public original Pro or Ultra checkpoint was identified. The December announcement is not a common availability date for every variant; service access and later Gemini versions do not preserve the initial weights.

Sources: Original Gemini announcement

Source and access verification

These entries have known publisher links or surviving components. They are not evidence of lost or unreleased weights. Grouped by modality, then date.

Audio

12 Feb 2018

WaveGAN v1 - original paper-era trained variants

Chris Donahue, Julian McAuley and Miller Puckette / UC San Diego | SOURCE / ACCESS TO VERIFY

12 Feb 2018 · paper v1; ICLR 2019

Gap: The author repository preserves a v1 tag and links a generation notebook and browser demo. Exact original checkpoint files and their correspondence to the paper variants remain unresolved in this audit. This is a source-verification task, not evidence of non-release or lost weights; current v2 must not silently replace v1.

Next: Trace the author notebook and browser asset manifest to original tensor/checkpoint files, record hashes and variant dates, and ask the authors only where provenance or a source remains unresolved.

Sources: Adversarial Audio Synthesis paper and version history / Official WaveGAN v1 source tree / Author-linked WaveGAN generation notebook

Language models

24 Feb 2023

LLaMA 1 - original authorized distribution

Meta AI | SOURCE / ACCESS TO VERIFY

24 February 2023 · announcement

Gap: Weights were distributed through an approval process. We have not established a currently usable authorized original download route; this is an access/provenance gap, not a claim that LLaMA weights are absent from the internet.

Next: Resolve original authorized access and checksums; exclude unverified leaked copies from the provenance chain.

Sources: Original code and distribution instructions

13 Mar 2023

Stanford Alpaca - original 7B reconstruction

Stanford CRFM / Tatsu Lab | SOURCE / ACCESS TO VERIFY

13 March 2023 · announcement

Gap: Author weight differences are located. Reconstruction depends on the exact legitimate LLaMA 1 base and the publisher's recovery procedure; the delta is not a complete standalone model.

Next: Resolve the base-model path, preserve the published delta and verify reconstruction against the author's expected result.

Sources: Original announcement / Published author weight differences

Preservation request contents

For a creator or rights holder: request the exact version/checkpoint, configuration, tokenizer or decoder dependencies, sampling code, representative input/output records, and permission terms for preservation, hosted inference and redistribution separately. For unavailable downloads: resolve the original object, provenance and complete file inventory before retrieval. No requests have been sent.