Public Summary of Training Content for Cosmos3

Version of the Summary: v1.1
Last update: 17/08/2026

1. General Information

1.1. Provider Identification

Provider name and contact details	NVIDIA Corporation 2788 San Tomas Expressway Santa Clara, CA 95051, USA
Authorised representative name and contact details	NVIDIA GmbH Adenauerstraße 20 A4 52146 Würselen, Germany AI_Governance@nvidia.com

1.2. Model Identification

Versioned model name(s)	• Cosmos3-Super: https://huggingface.co/nvidia/Cosmos3-Super • Cosmos3-Nano: https://huggingface.co/nvidia/Cosmos3-Nano • Cosmos3-Edge: https://huggingface.co/nvidia/Cosmos3-Edge • Cosmos3-Nano-Policy-DROID: https://huggingface.co/nvidia/Cosmos3-Nano-Policy-DROID • Cosmos3-Edge-Policy-DROID: https://huggingface.co/nvidia/Cosmos3-Edge-Policy-DROID • Cosmos3-Super-Image2Video-4Step: https://huggingface.co/nvidia/Cosmos3-Super-Image2Video-4Step • Cosmos3-Super-Text2Image-4Step: https://huggingface.co/nvidia/Cosmos3-Super-Text2Image-4Step • Cosmos3-Super-Image2Video: https://huggingface.co/nvidia/Cosmos3-Super-Image2Video

	• Cosmos3-Super-Text2Image: https://huggingface.co/nvidia/Cosmos3-Super-Text2Image
Model dependencies	Cosmos3 models are not fine-tuned or modifications of a previously placed general-purpose AI model.
Date of placement of the model on the Union market	**Released on July 20, 2026:** • Cosmos3-Edge • Cosmos3-Edge-Policy-DROID • Cosmos3-Super-Image2Video-4Step • Cosmos3-Super-Text2Image-4Step **Released on May 31, 2026:** • Cosmos3-Nano • Cosmos3-Super • Cosmos3-Nano-Policy-DROID • Cosmos3-Super-Image2Video • Cosmos3-Super-Text2Image

1.3. Modalities, Overall Training Data Size and Other Characteristics

Modality	Training data size	Types of content
☑ Text	☐ Less than 1 billion tokens ☑ 1billion to 10 trillions tokens ☐ More than 10 trillions tokens	General language text, physical AI task descriptions, reasoning traces, robot planning instructions, scene descriptions, captions.

☑ Image	☐ Less than 1 million images ☑ 1 Million to 1 billion images ☐ More than 1 billion images	Photographs and synthetic images of indoor and outdoor scenes, including industrial and factory environments, and daily life activities.
☐ Audio	☐Less than 10,000 hours ☐ 10, 000 to 1 million hours ☐ More than 1 million hours	
☑ Video	☐ Less than 10,000 hours ☐ 10,000 to 1 million hours ☑ More than 1 million hours	Video clips of indoor and outdoor scenes, including driving footage, robotic manipulation, daily life activities, industrial environments, nature, objects, animals, and technology; aerial footage of landscapes, cities, transportation, and other environments.
☑ Other: Action Trajectories	8 million samples (generation)	Robot and agent action trajectories in JSON format: per-frame sequences of robot or agent state/control values (e.g. joint positions, gripper state, camera pose, hand poses).

Latest date of data acquisition/collection for model training	May 2026
Description of the linguistic characteristics of the overall training data	For text and caption data, the training data is primarily English language. The video, image, and action trajectory data are not language-specific by nature.

Other relevant characteristics of the overall training data	The corpus spans diverse multimodal sources across robotics, autonomous driving, industrial and factory environments, indoor and outdoor scenes.
Additional comments (optional)	Technical report available here: https://research.nvidia.com/labs/cosmos-lab/cosmos3/technical-report.pdf

2. List of Data Sources

2.1. Publicly Available Datasets

Have you used publicly available datasets to train the model?	☑ Yes ☐ No
If yes, specify the modality(ies) of the content covered by the datasets concerned:	☑ Text ☑ Image ☑ Video ☑ Audio ☐ Other
List of large publicly available datasets:	OpenImage (https://storage.googleapis.com/openimages/web/index.html), Coyo700M (https://github.com/kakaobrain/coyo-dataset), UMI (Universal Manipulation Interface) (https://umi-data.github.io/).

General description of other publicly available datasets not listed above	Other publicly available datasets include broad, multi-domain image and video datasets.
Additional comments (optional)	

2.2. Private Non-Publicly Available Datasets Obtained from Third Parties

2.2.1. Datasets Commercially Licensed by Rightsholders or Their Representatives

Have you concluded transactional commercial licensing agreement(s) with rightsholder(s) or with their representatives?	☑ Yes ☐ No
If yes, specify the modality(ies) of the content covered by the datasets concerned:	☐ Text ☐ Image ☑ Video ☐ Audio ☐ Other

2.2.2. Private Datasets Obtained from Other Third Parties

Have you obtained private datasets from third parties that are not licensed as described in Section 2.2.1?	☑ Yes ☐ No
If yes, specify the modality(ies) of the content covered by the datasets concerned:	☐ Text ☑ Image ☑ Video ☐ Audio ☐ Other

If publicly known, list private datasets obtained from other third parties:	Nexar, AgiBot, HOI (Hand-Object Interaction), Egocentric
General description of non-publicly known private datasets obtained from third parties	Image and video content (including autonomous driving videos, robotics manipulation videos, industrial scene videos), archives, and metadata.
Additional comments (optional):	

2.3. Data Crawled and Scraped from Online Sources

Were crawlers used by the provider or on behalf of?	☑ Yes ☐ No
Crawler Names/Identifier:	NVIDIA Crawler Identifier: nvidiabot Additional first- and third-party crawler identifiers unknown.
Purposes of the crawler(s):	Crawlers were used to collect diverse video content from publicly available internet sources for training the model's video generation and understanding components.
General description of crawler behaviour:	Automated crawlers were used to collect publicly available media from selected websites. The crawlers navigated publicly accessible pages, extracted relevant links and associated metadata, and collected eligible content for model training and evaluation.
Period of data collection:	2024–2026
Comprehensive description of the type of content and online sources crawled:	Photographs and videos covering daily life activities, nature, travel, urban and indoor environments, objects, animals, technology, and industrial scenes; aerial footage of landscapes, cities, transportation, and other environments.

Type of modality covered:	☑ Video ☑ Audio ☑ Text ☑ Image
Summary of the most relevant domain names crawled:	youtube.com
Additional comments (optional)	

2.4. User Data

Was data from user interactions with the AI model (e.g. user input and prompts) used to train the model?	☐ Yes ☑ No
Was data collected from user interactions with the provider's other services or products used to train the model?	☐Yes ☑ No
Additional comments (optional)	

2.5. Synthetic Data

Was synthetic AI-generated data created by the provider or on their behalf to train the model?	☑ Yes ☐ No
If yes, modality of the synthetic data:	☐ Text ☑ Image ☐ Video ☐ Audio ☐ Other
If yes, specify the general-purpose AI	Qwen-Image-2512

model(s) used to generate the synthetic data if available on the market:	
Information about other AI models, including provider's own AI models not available on the market, used to generate synthetic data:	Qwen3-VL, HiDream-I1
Additional comments (optional)	

2.6. Other Sources of Data

Have data sources other than those described in Sections 2.1 to 2.5 been used to train the model?	☑ Yes ☐ No
If yes, provide a narrative description of these data sources and the data:	NVIDIA-created or collected data, including images and videos from robotics, autonomous vehicles, and industrial testing environments.
Additional comments (optional)	

3. Data Processing Aspects

3.1. Respect of Reservation of Rights from Text and Data Mining Exception or Limitation

Are you a Signatory to the Code of Practice for general-purpose AI models that includes commitments to respect reservations of rights from the TDM exception or limitation?	☐ Yes ☑No
Describe the measures implemented before model training to respect reservations of rights from the TDM exception or limitation before and during data collection, including the opt-out protocols and solutions honoured by the provider or, as applicable, by third parties from which datasets have been obtained:	NVIDIA implements measures to respect EU rights reservations relevant to text and data mining by: (1) respecting robots.txt directives at the domains accessed where those directives constituted a reservation of rights under Article 4(3) of Directive (EU) 2019/790; and (2) filtering datasets on any actionable metadata identifiers provided by rightsholders.
Additional comments (optional)	

3.2. Removal of Illegal Content

General description of measures taken:	Training datasets passed through multiple layers of automated and manual safeguards designed to reduce the presence of harmful or policy-violating content across categories including weapons and weapons-related instructional content, criminal planning, child sexual abuse material (CSAM), non-consensual intimate imagery (NCII), sexual content involving minors, harassment, hate speech, profanity, threats and incitement to violence, self-harm or suicide-related content, and graphic violence. Data sources are reviewed for licensing compatibility, provenance, and alignment with internal data governance and safety policies before admission into training corpora. Automated filtering pipelines combine multiple detection strategies: hash-matching against known CSAM and NCII reference databases; classifier-based moderation models trained for explicit sexual content, hate speech, violence, weapons imagery, and other restricted categories; keyword and regex-based screening for criminal-planning, threats, and self-harm phrases in text data; metadata and provenance heuristics for source-level risk signals; and embedding-based anomaly detection to surface samples that fall outside expected distributions. Human review and targeted audits supplement automated filtering for selected datasets, benchmark construction, and safety-sensitive evaluation. For multimodal Physical AI data (robotics, autonomous driving, industrial scenes), additional filtering targets invalid action trajectories, physically implausible interactions, and unsafe control sequences. Synthetic and simulation-generated data are evaluated through internal validation before inclusion. Benchmark evaluations and red-team testing are applied post-training to surface remaining safety gaps across world generation, reasoning, audio, and action tasks. Ongoing monitoring and dataset review continue post-release.

3.3. Other Information (Optional)

Other relevant information about data processing (optional):	N/A