

# Public Summary of Training Content for Cosmos3-Nano-Sim-Bimanual

Version of the Summary: v0.1

Last update: 20 October 2026

## 1. General information

### 1.1. Provider identification

NVIDIA Corporation

Provider name and contact details: 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-Nano-Sim-Bimanual — V1.0  
<https://huggingface.co/nvidia/Cosmos3-Nano-Sim-Bimanual>

Model dependencies: Post-trained, causal-distilled variant of Cosmos3-Nano. This summary covers only additional training data first introduced after the Cosmos3 baseline. Inherited training content is described in the [Cosmos3 public summary](#), available at [https://docs.nvidia.com/cosmos/latest/\\_downloads/e482b7114ce8dbfbb07d2d4b42cafe4e/training-content-Cosmos-3.pdf](https://docs.nvidia.com/cosmos/latest/_downloads/e482b7114ce8dbfbb07d2d4b42cafe4e/training-content-Cosmos-3.pdf).

Date of placement of the model on the Union market: 2026-10-20

### 1.3 Modalities, overall training data size and other characteristics

| Modality                                 | Training data size                                                                                                                                                                                                                                                                     | Types of content                                                                           |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> Text | <input type="checkbox"/> Less than 1 billion tokens<br><input checked="" type="checkbox"/> 1 billion to 10 trillion tokens<br><input type="checkbox"/> More than 10 trillion tokens<br>Alternatively, specify the approximate size in a different measurement unit:<br>Not applicable. | English captions and task/action descriptions associated with robotic manipulation videos. |
| <input type="checkbox"/> Image           | <input type="checkbox"/> Less than 1 million images<br><input type="checkbox"/> 1 million to 1 billion images<br><input type="checkbox"/> More than 1 billion images                                                                                                                   |                                                                                            |

| Modality                                  | Training data size                                                                                                                                                     | Types of content                                                                                       |
|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Audio            | <input type="checkbox"/> Less than 10,000 hours<br><input type="checkbox"/> 10,000 to 1 million hours<br><input type="checkbox"/> More than 1 million hours            |                                                                                                        |
| <input checked="" type="checkbox"/> Video | <input type="checkbox"/> Less than 10,000 hours<br><input checked="" type="checkbox"/> 10,000 to 1 million hours<br><input type="checkbox"/> More than 1 million hours | Real and synthetic robotic manipulation videos paired with frame-aligned robot state and control data. |
| <input type="checkbox"/> Other            |                                                                                                                                                                        |                                                                                                        |

Latest date of data acquisition/collection for model training:

2026-07. The model is not continuously trained on new or dynamic data after this date.

Description of the linguistic characteristics of the overall training data:

The post-training video datasets have associated English captions. Robot action and camera-trajectory values are not inherently language-specific.

Other relevant characteristics of the overall training data:

The additional inventory comprises vendor-supplied real action data, publicly available real action datasets, and a publicly available dataset marked synthetic.

Additional comments (optional):

## 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 ☒ Video

List of large publicly available datasets:

MolmoAct2-BimanualYAM ([allenai/MolmoAct2-BimanualYAM-Dataset](https://huggingface.co/datasets/allenai/MolmoAct2-BimanualYAM-Dataset) · [Datasets at Hugging Face](https://huggingface.co/datasets/allenai/MolmoAct2-BimanualYAM-Dataset); available at <https://huggingface.co/datasets/allenai/MolmoAct2-BimanualYAM-Dataset>)

ABC-130k: the RealSense and ZED-X training entries ([ABC-130k](https://huggingface.co/datasets/XDOF/ABC-130k) · [Datasets at Hugging Face](https://huggingface.co/datasets/XDOF/ABC-130k); available at <https://huggingface.co/datasets/XDOF/ABC-130k>)

Behavior-1K: BEHAVIOR-1K 2025 Challenge Demos ([behavior-1k/2025-challenge-demos](https://huggingface.co/datasets/behavior-1k/2025-challenge-demos) · [Datasets at Hugging Face](https://huggingface.co/datasets/behavior-1k/2025-challenge-demos); available at <https://huggingface.co/datasets/behavior-1k/2025-challenge-demos>)

General description of other publicly available datasets not listed above:

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, such as data obtained from providers of private databases, or data intermediaries?

☐ Yes ☒ No

If yes, specify the modality(ies) of the content covered by the datasets concerned:

If publicly known, list private datasets obtained from other third parties:

General description of non-publicly known private datasets obtained from third parties

Additional comments (optional):

## 2.3 Data crawled and scraped from online sources

Were crawlers used by the provider or on behalf of?

☐ Yes ☒ No

If yes, specify crawler name(s)/identifier(s):

Not applicable.

Purposes of the crawler(s):

General description of crawler behaviour:

Period of data collection:

Comprehensive description of the type of content and online sources crawled:

Type of modality covered:

Summary of the most relevant domain names crawled:

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

If yes, provide a general description of the provider's services or products that were used to collect the user data:

Not applicable.

Type of modality covered:

Not applicable.

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 model(s) used to generate the synthetic data if available on the market:

Information about other AI models, including provider's own AI model(s) not available on the market, used to generate synthetic data to train the model to which this Summary applies:

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:

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 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):

---