Public Summary of Training Content for Voice

Version of the Summary: V1
Last update: 29 July 2026

1. General information

1.1. Provider identification

Provider name and contact details: xAI LLC, 1450 Page Mill Rd, Palo Alto, CA 94304

Authorised representative name and contact details: EDSR, Valukoja 8/2, 2nd floor, Tallinn, Estonia, EE-11415

1.2. Model identification

Versioned model name(s): Grok Voice Think Fast 2.0

Model dependencies: N/A

Date of placement of the model on the Union market: 29 July 2026

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	*Grok Voice Think Fast 2.0 was trained using a carefully designated data recipe that incorporates a diverse corpus of publicly available information, including, for example, certain scientific text, legal and official documents, and X social media posts. The training data also encompasses content produced by third parties, including data from users and contractors, as well as internally generated data. Training data was curated and filtered for quality and safety.*
☐ Image	☐ Less than 1 million images ☐ 1Million to1 billion images ☐ More than 1 billion images	N/A
☒ Audio[1]	☐ Less than 10 000 hours ☐ 10 000 to1 million hours ☒ More than 1 million hours	*Grok Voice Think Fast 2.0 was trained using a carefully designated data recipe that incorporates a diverse corpus of publicly available audio data, including, for example, public speech, media, voice application recordings, noise, audio communications, certain clips and audio from social media videos. The training data may also encompass audio content produced by third parties, including from users and contractors, as*

[1] Excluding audio that is part of video, as this should be reported under the "video" modality instead. Furthermore, the Commission understands the modality of 'audio' to include 'speech'.

		well as internally generated content. Training data was curated and filtered for quality and safety.
☒ Video	☒ Less than 10 000 hours ☐ 10 000 to1 million hours ☐ More than 1 million hours	*Grok Voice Think Fast 2.0 was trained on publicly available audiovisual content, including, for example, certain clips and X social media videos.* *The training data may also encompass audiovisual content produced by third parties, including from users and contractors, as well as internally generated content. Training data was curated and filtered for quality and safety.*
☐ Other	N/A	N/A

Latest date of data acquisition/collection for model training:	The data used to train Grok Voice Think Fast 2.0 includes different datasets from varying time periods, with data collected no later than June 2026.
Description of the linguistic characteristics of the overall training data:	Multilingual, with strong English coverage and substantial representation across EU official languages and other languages from around the world.
Other relevant characteristics of the overall training data:	Grok's training corpus is designed for a highly speech-to-speech model with strong multimodal understanding. It draws from a range of written material, audio and audiovisual content. The corpus is intentionally broad, aiming for extensive coverage across topics, geographies, languages, and formats.
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 ☐ Image ☒ Video ☒ Audio ☐ Other

- List of large publicly available datasets:
  - ***Common Voice:*** *a free, open-source crowdsourcing project started by Mozilla to create a public domain database of diverse human voices. It helps developers train speech-to-text and text-to-speech AI applications without high costs or proprietary (dates of data collection not known)*
  - ***LibriSpeech:*** *LibriSpeech is a large speech dataset containing about 1,000 hours of English audio, designed for training and testing speech recognition systems. The audio comes from read audiobooks in the LibriVox project and is sampled at 16 kHz (dates of data collection not known)*

	• ***VoxPopuli:*** *VoxPopuli is a large-scale multilingual speech corpus for representation learning, semi-supervised learning and interpretation. The raw data is collected from 2009-2020 European Parliament event recordings.* • ***Fleurs****: speech version of the FLoRes machine translation benchmark. Uses 2009 n-way parallel sentences from the FLoRes dev and devtest publicly available sets, in 102 languages (dates of data collection not known).*
General description of other publicly available datasets not listed above:	Other publicly available datasets used by Grok Voice Think Fast 2.0 consisted of text and audio, content and audiovisual content, in various different languages, from other publicly available sources (dates of data collection not known).
Additional comments (optional):	Audio data is packaged for diverse audio processing tasks (speech recognition, speaker recognition, speech generation, acoustic event analysis)

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:	☐ Text ☐ Image ☐ Video ☒ Audio ☐ Other
If publicly known, list private datasets obtained from other third parties:	
General description of non-publicly known private datasets obtained from third parties	Private datasets used to train Grok Voice Think Fast 2.0 that are not publicly known consisted of audio, including conversations on various different topics, recorded in clean and high quality environments.
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):	xAI Web Crawler

Purposes of the crawler(s):	xAI uses a web crawler to crawl certain publicly available internet pages to help improve language understanding, reasoning, and overall capabilities.
General description of crawler behaviour:	xAI uses a web crawler to discover and scan certain publicly available internet pages and rescan for updates and analysis.
Period of data collection:	From January 2024 to June 2026
Comprehensive description of the type of content and online sources crawled:	Content encompasses a wide variety of publicly available internet data including educational, scientific, technical, government, and institutional sources. These materials may consist of text, audio, and video.
Type of modality covered:	☒ Text ☐ Image ☒ Video ☒ Audio ☐ Other
Summary of the most relevant domain names crawled:	The domains accounting for the largest share of content crawled and used to train Grok Voice Think Fast 2.0, may comprise a mix of publicly available information, including larger social media and content sharing platforms, educational repositories, government and institutional resources, community and document-sharing sites. Audio and audiovisual content may be accompanied by publicly available transcripts, captions, titles, descriptions, and other associated metadata. This associated text provides contextual and linguistic information relevant to speech recognition, speech generation, and multilingual understanding.
Additional comments (optional):	xAI uses the U.S. Trade Representative (USTR) Notorious Markets for Counterfeiting and Piracy list as a signal when deciding to exclude data from certain websites that have been recognized as persistently and repeatedly infringing copyright.

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:	
Type of modality covered:	☒ Text ☐ Image ☐ Video ☒ Audio ☐ Other
Additional comments (optional):	Subject to privacy settings, controls, user requests, opt-outs, and our policies, xAI may use interactions with Grok to help train and improve our models, including Grok Voice Think Fast 2.0. For more information on how we use data to improve model performance, please visit: https://x.ai/privacy. xAI uses advanced data filtering processes to reduce personal information from training data. We also use advanced data processes to reduce the amount of personal data in our training data.

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:	xAI generates synthetic data using its own general-purpose AI models, including Grok 4.3, Grok 4.5.
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:	xAI uses other models to generate synthetic data for targeted training objectives, including augmenting data in domains, languages, tasks, or formats where specialized training data is comparatively scarce. These models may be used to generate examples for instruction following, reasoning, coding, multimodal understanding, safety, and evaluation.
Additional comments (optional):	xAI's tutors used xAI's own general-purpose AI models to synthetic data, text and audio conversations that could be used to train Grok Voice Think Fast 2.0.

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:	xAI also creates training data through employees and tutors where suitable existing data is not available. For Grok Voice Think Fast 2.0, this included human recorded conversations and simulated interactions across a range of topics and contexts.
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:	xAI implements measures to respect rights reservations and opt-out signals relevant to data. For example, xAI does not train on X data from X users who have opted out of training. For web data used for training, xAI respects instructions that identified crawled content should not be used to train xAI's generative foundation models.
Additional comments (optional):	

3.2 Removal of illegal content

General description of measures taken:	xAI applies preprocessing and screening measures designed to avoid or remove content that may be illegal from our training data. These measures include filtering, keyword-based rules and model-based classifiers.

3.3. Other information (optional)

Other relevant information about data processing (optional):	