

Kyutai Unmute Voice Donation Project Privacy Policy

Welcome to the Unmute Voice Donation Project

What is the Unmute Voice Donation Project? The Unmute Voice Donation Project is an open-sourced project developed by Kyutai (“we”, “us”, “our”), that allows AI models to listen and speak. It is made possible by our new speech-to-text and text-to-speech models optimized for real-time usage.

As part of this Unmute Voice Donation Project, we invite users (“you”, “your”), with their consent, to record their voice to help improve open research and development in this field.

We care about your privacy and want to be upfront about how we treat your personal data: what we do collect, why we collect it, where it all goes, and for how long. So we wrote this privacy policy (the “**Privacy Policy**”) to explain it to you in the clearest, simplest way we know how. In addition to this Privacy Policy, please read our [Terms](#) which apply to your participation in the Unmute Voice Donation Project.

⚠️ Your participation in the Unmute Voice Donation Project may entail risks to your privacy, so please read carefully this Privacy Policy.

When you record your voice, you will be asked if you want to consent to the transmission of your voice recording to us to (i) develop and train our text-to-speech AI models (“**TTS AI Models**”), and to provide these TTS AI Models to the public, and/or (ii) to constitute an open-source dataset that includes multiple voice recordings and embeddings, including yours, which will be released and made publicly available on the Internet. You may not participate in the Unmute Voice Donation Project if you do not consent to at least one of these purposes, but rest assured: it will not have any consequences on you.

We may publish a voice dataset in open-source because we believe that open-source publication is the best means to spread knowledge and allow everyone to carry out research and build tools that could be helpful and beneficial to everyone and society.

💡 When the voice dataset is released in open-source, it means that it is publicly available to anyone on the Internet and therefore anyone may be able to process your voice for multiple purposes such as research and development, training of AI models or even recreational use. We cannot exclude that some users of the dataset may even replicate your voice and use it to say things that you never said.

While we impose on any person accessing the open-source voice dataset an Acceptable Use Policy prohibiting them to use your voice in any unlawful manner, we will have very limited means to enforce this policy due to the open-source nature of the voice dataset.

⚠️ If you do not want people to reuse your voice and reproduce it for any purpose that you may think harmful to you (e.g. voice replication), you should not consent to the public release of your voice in open-source or the use of your voice for the training of our TTS AI Models.

We care about your privacy so if you want at any time to delete your voice from the open-source voice dataset to which you previously consented, we will delete it as soon as possible by updating all the open-source datasets that we released on the Internet.

And as always, if you have any questions or suggestions, give us a shout at unmute@kyutai.org.

Overview of our processing of your personal data

Purpose of our processing	Personal data that we process	Data subject affected by the processing	Legal basis of our processing	Retention period of your personal data	Recipient of your personal data
Training and developing our TTS AI Models and making it available to the public	Your voice recording Embedding of your voice Synthetic recordings of your voice. Time and date of your voice recording	You	Your consent	15 years from the voice recording (you may withdraw your consent at any time so that we delete it)	Any undertaking taking over our activities and business in the future, if any. Professional advisors. Authorities (e.g. courts, law enforcement etc).
Constitution and public release on the Internet of an open-source voice dataset	Your voice recording Embedding of your voice Time and date of your voice recording	You	Your consent	As long as the voice dataset is available in open-source	Any person accessing the open-source voice dataset on the Internet The platforms hosting the voice dataset on the Internet. The voice datasets will be published on the following platforms: Hugging Face, Freesound.org. Professional advisors. Authorities (e.g. courts, law enforcement etc).
Managing your rights and documenting your consent	Your voice recording Embedding of your voice Time and date of your voice recording Your email address Consent	You	Our legal obligation	As long as the voice dataset is available in open-source, or for at least 1 year from the date of the voice recording	Any undertaking taking over our activities and business in the future, if any. Professional advisors. Authorities (e.g. courts, law enforcement etc).
Ensuring that the voice recording is from you and not an existing audio recording	Your verification sentences Your voice recording	You	Our legitimate interest to ensure that you are indeed the source of recording	15 years	Any undertaking taking over our activities and business in the future, if any. Professional advisors. Authorities (e.g. courts, law enforcement etc).

Please find more details about our processing of your personal data below

1. Who are we and how can you contact us?

We are Kyutai, a not-for-profit open-science French AI lab. Our mission is to build and democratize artificial general intelligence through open science. For more information on our activities, you can visit our website available here: <https://kyutai.org/>.

We are the entity responsible for the processing of your personal data under this Privacy Policy (as a data controller, where provided under applicable data protection laws such as the EU General Data Protection Regulation (GDPR)).

We are located at: 16 Rue de la Ville-l'Évêque, 75008 Paris, France, and you can contact us for any question regarding this Privacy Policy and our processing of personal data at the following email address: unmute@kyutai.org.

2. Why do we process your personal data?

We process your personal data for the following purposes:

- Ensuring that the voice recording is from you and not a random audio sample. We do so by requesting you to pronounce random verification sentences such as "*Her designs extended from foliage and plants to narrative scenes. The dots are only on the back.*" Then, we verify that the voice of the recording is the same as the voice of the verification sentences.
- Training and developing our TTS AI Models and making it available to the public. We use your voice recording and your voice embedding to build, train and develop our TTS AI Models, with the aim to make it available for use to the public in the future. As part of this purpose, we may use your voice recording and your voice embedding to create synthetic recordings (i.e. artificial voice recording based on your voice) so that we have additional data to train our TTS AI Models, and that its training be based on more diverse data to improve our TTS AI Models.
- Constituting and publicly releasing an open-source voice dataset. We will constitute a dataset that will include notably the voice recordings and voice embeddings of any user that consented to record their voice for such purpose. This dataset will be made publicly available on the Internet in open-source to help promote science, research and development in AI models.
- Managing your rights and documenting your consent. We collect your email address and your consent to any of the aforementioned processing requiring such consent to ensure that any voice recording or embedding that we use for the training of our TTS AI Models or to constitute an open-source voice dataset has been subject to an appropriate consent. We also keep one means of identification (e.g. email) to manage any right request you may have, for example if you would like to withdraw your consent or have your personal data deleted from our database or any open-source database.

3. What personal data do we process and how do we collect them?

We process the following personal data from you:

- Your email address, before you proceed to your voice recording.
- Your consent to any of our processing that requires it.
- Your verification sentences, which are random sentences such as "*Her designs extended from foliage and plants to narrative scenes. The dots are only on the back.*"
- Your voice recording.

You may record your voice through our dedicated website available or through a third party platform identified on our dedicated website, which may also require you to create an account. We do not collect or receive your account information from these third-party platforms.

We also generate the following personal data based on your voice recording:

- An embedding of your voice. An embedding is a numerical representation generated by automatic tools that captures distinctive characteristics of your voice. These embeddings may be used to identify, mimic, or differentiate between speakers, and will be used by text-to-speech AI models to faithfully reproduce any voice recorded.
- Any synthetic recording of your voice (i.e. artificial voice recording based on your voice).
- Time and date of your voice recording.

4. What is the lawful basis of our processing of your personal data?

We process your personal data with the following lawful basis in accordance with applicable data protection laws:

- Ensuring that the voice recording is from you and not a random audio sample. We carry out this processing based on our legitimate interest. Our legitimate interest is that the voice recording actually comes from the person doing the voice recording, and not from a random audio sample. This verification allows us to only select voice recordings (i.e. your voice recording) that will be helpful for research and development in text-to-speech and speech-to-text AI models.

- Training and developing our TTS AI Models and making it available to the public. This processing is based on your express consent. As explained in Section 8, you may withdraw your consent at any time.
- Constitute and release an open-source voice dataset. This processing is based on your express consent. As explained in Section 8, you may withdraw your consent at any time.
- Managing your rights and documenting your consent. This processing is based on our legal obligation pursuant to applicable data protection laws, including GDPR, to document any consent collected for our processing of personal data, and respect your data subjects rights detailed in Section 8 below.

5. For how long do we keep your personal data?

We keep your personal data only for the duration necessary to fulfil the purpose of our processing. The retention periods for our processing of your personal data are the following:

- Ensuring that the voice recording is from you and not a random audio sample. We keep the verification sentences and the associated voice recording for 15 years after the date of your voice recording. After assessment of the voice recording and the verification sentences, the verification sentences and voice recordings are archived and removed from our day-to-day databases.
- Training and developing our TTS AI Models and making it available to the public. We keep your voice recording, your voice embedding and the time and date of recording for 15 years after the date of your voice recording.
- Constitute and release an open-source voice dataset. We keep your voice recording, your voice embedding and the time and date of recording for as long as the open-source voice dataset that we made available on the Internet remains publicly available on the platforms on which we uploaded it.
- Managing your rights and documenting your consent. We keep your voice recording, your voice embedding, the time and date of recording, your consent and your email address for as long as the open-source voice dataset remains publicly available on the web platforms on which we uploaded it, and for the retention period of your personal data in our database, for 15 years. These data are immediately archived after collection, and will only be accessed by Kyutai's personnel on a need-to-know basis (e.g. privacy or legal teams) to document your consent and respond to your right requests or any other lawful request from the authorities.

6. Who are the recipients of your personal data ?

We do not use any data processor within the meaning of data protection laws, including GDPR, which may have access to your personal data for the purposes of training and making available our TTS AI Models or ensuring that the voice recording is from you.

The following recipients may have access to your personal data, and will process your personal data as independent data controller within the meaning of data protection laws, including GDPR:

- Any person accessing the open-source voice dataset. Despite our Acceptable Use Policy in place and applicable to such persons, we are not and cannot be responsible for the processing of your personal data by these persons, who act as separate and independent data controllers.
- Any platform hosting the open-source voice dataset. The platforms hosting the voice dataset on the Internet. The voice datasets will be published on the following platforms: Hugging Face, Freesound.org.
- Our professional advisors such as lawyers, auditors, bankers, and insurers, where necessary in the course of the professional services that they render to us.
- Authorities and others, such as law enforcement, government authorities, and private parties, as we believe in good faith to be necessary or appropriate to comply with our legal obligations.
- Any undertaking taking over our activities and any affiliate thereof, through any legal means such as corporate divestiture, merger, consolidation, acquisition, reorganisation, sale, asset contribution, or other disposition of all or any portion of the business or assets of, or equity interests in Kyutai (including, in connection with a bankruptcy or similar proceedings), including any corporation that may receive the assets and resources of Kyutai.

7. Transfers outside the EU

We do not transfer your personal data outside the European Union.

However, as the open-source voice dataset will be publicly available on the Internet, your personal data included in the open-source voice dataset may be transferred outside the EU by any user accessing your personal data in such open-source voice dataset. We impose on any user of the open-source voice dataset the obligation to process your personal data in compliance with any applicable data protection laws.

8. What are your rights on your personal data ?

You can exercise the following rights on your personal data:

- **Access**. You have the right to know if we process your personal data. You also have the right to request a copy of such personal data and to obtain further information about the way we process your personal data.
- **Rectification**. You have the right to ask us to update or correct your personal data if your personal data is inaccurate.
- **Objection**. You have the right to object to the processing of your personal data. This right does not apply when we have a legal obligation to process your personal data.
- **Consent withdrawal**. You have the right to withdraw your consent to the processing of your personal data at any time. In such a case, we will delete your personal data that we processed based on your consent, from our internal database, and will delete your personal data from existing available open-source voice dataset that we publicly released on the Internet. We will do our best efforts to have your personal data deleted from the existing available open-source voice dataset, for example by deleting your personal data from said dataset, but we cannot guarantee that users of said open-source datasets, who may have accessed your personal data before their deletion, will stop using your personal data, even though they are obliged by our Acceptable Use Policy to regularly review any update to the open-source voice dataset so to identify any consent withdrawal requests.
- **Deletion**. You have the right to delete and/or ask us to delete your personal data, unless we need to keep your personal data to comply with our legal obligations or protect our rights or third party rights. In such a case, we will delete your personal data that we processed from our internal database, and will delete your personal data from existing available open-source voice dataset that we publicly released on the Internet. We will do our best efforts to have your personal data deleted from the existing available open-source voice dataset, for example by deleting your personal data from said dataset, but we cannot guarantee that users of said open-source datasets, who may have accessed your personal data before their deletion, will stop using your personal data, even though they are obliged by our Acceptable Use Policy to regularly review any update to the open-source voice datasets so to identify any deletion requests.
- **Limitation**. You have the right to ask us to freeze the processing of your personal data.
- **Automated decision**. You have the right to not be subject to an automated decision (including profiling) and to appeal such a decision. We do not engage in profiling or automated decision-making in the processing of your personal data.
- **Portability**. You have the right to obtain and transfer your personal data to another entity.
- **Post mortem**. You have the right to tell us how you would like us to process your personal data after your death.
- **Lodge a complaint**. You have the right to lodge a complaint before the competent data protection authority, including the French data protection authority (the CNIL).

We will take every step to make sure we reply to your requests. However, when your request concerns the open-source release of your personal data, it is important to note that your rights have technical limitations and we may not be able to compel and enforce your rights towards any user of the open-source voice dataset.

You can exercise these rights at all times:

- By contacting us at the following email address: unmute@kyutai.org.
- By sending us a letter to our following postal address: 16 Rue de la Ville-l'Évêque, 75008 Paris, France.

9. Changes to this Privacy Policy

We may amend this Privacy Policy from time to time as our processing of personal data continuously evolve. Please make sure to check this Privacy Policy frequently.