

Using digital tools for transcription

Overview and Whisper AI demonstration

Laura Armstrong & Dr Nidhi Gowdra
Centre for eResearch

July 2024



Valuing inclusion

Ensuring all individuals feel respected, accepted, and valued.

- Manaakitanga - show respect, care and support for others
- Whanaungatanga – foster an environment where all in our community have a place
- Kotahitanga – built unity and partnership
- Kaitiakitanga - recognise our responsibilities as kaitiaki (guardians) to protect and respect our environment, traditions, knowledge, culture, languages and other taonga.

Mute yourself, use Zoom Chat or Reactions to 'raise a hand' to ask a question.

Camera on or off – up to you.



<https://www.auckland.ac.nz/en/on-campus/life-on-campus/code-of-conduct.html>



Chat: Introduce yourself and share what you are hoping to hear about today?

Today

- Considerations for choosing transcription tools
- Whisper demonstration
- Questions

Transcription in research

- Data transcription plays a crucial role in conducting qualitative research and analysing qualitative data
 - Interviews
 - Group discussion/Focus groups
- Qualitative research is commonly sensitive
 - Personally identifiable?
 - Subject to ethics approval?
 - Privacy Act 2020?



Chat: What tools do you use or plan to use for transcribing your research data?

Why do you use these tools?

AI tools in transcription

- What is artificial intelligence – AI?
Artificial intelligence (AI) refers to computer systems capable of performing complex tasks that historically only a human could do, such as reasoning, making decisions, or solving problems.
- Benefits of using an AI tool for transcription
 - Accessible and easy to use
 - Quick
 - Cheaper
 - Accurate
 - Translation options

Risks and considerations

Where does your data go, and who has access to the data and for what purpose?

- Ethical
- Privacy
- Security
- Sovereignty

Read the terms of use



Solutions ▾

Pricing

Apps & Integrations

Resources ▾

Schedule a Demo

Log In

Start for Free

2. HOW WE USE YOUR PERSONAL INFORMATION

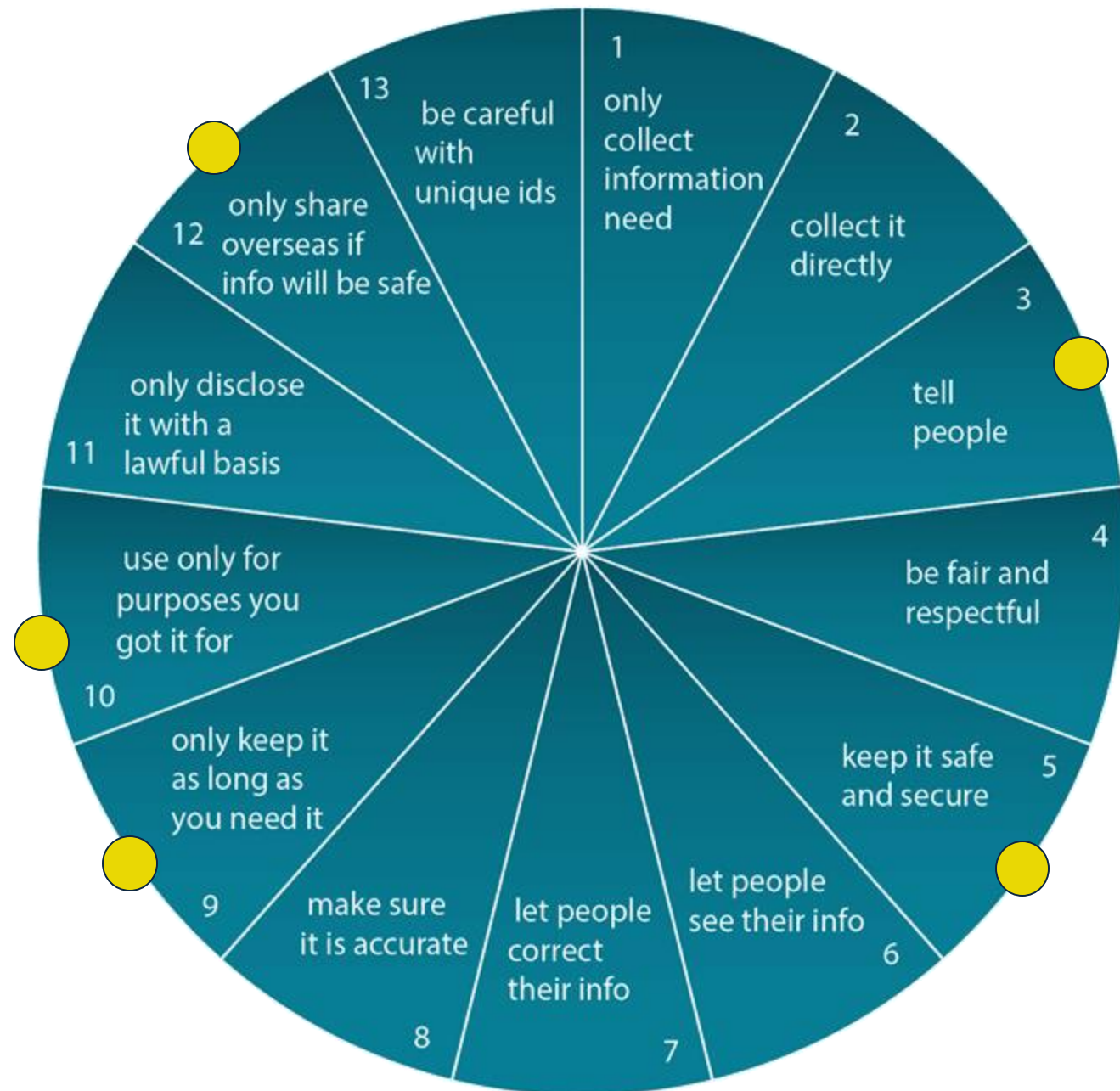
We use your Personal Information to:

- **Set up your account.** We use your registration information, device information and information received from third parties (such as your username, email address) in order to set up an account for you to use our Services. We do so in accordance with our contractual and precontractual obligations to you in order to provide you with an account to use the Services.
- **Provide you with the Services.** We use your audio recordings, usage information and platform information in order to provide you with the Services. In addition, we use your communication information to facilitate support (e.g. retrieval of a forgotten password). We do so in accordance with our contractual obligations to you in order to provide you with the Services.
- **Improve and monitor the Services.** We use information we automatically collect or generate about you when you use the Services, as well as information about your device such as device manufacturer, model and operating system, and the amount of free space on your device, to analyze the use of and improve our Services. We train our proprietary artificial intelligence technology on de-identified audio recordings. We also train our technology on transcriptions to provide more accurate services, which may contain Personal Information. We obtain explicit permission (e.g. when you rate the transcript quality and check the box to give Otter.ai and its third-party service provider(s) permission to access the conversation for training and product improvement purposes) for manual review of specific audio recordings to further refine our model training data.
- **Communicate with you.** If you contact us, we will use your contact information to communicate with you and, if applicable, your usage information to support your use of the Services.

<https://otter.ai/privacy-policy>

Privacy Act 2020

- Consent?
- Secure?
- Retained?
- For another purpose?
- Held in other jurisdictions?



Approved tools

Use institution-supported software or tools, whenever possible.

These applications are often security tested and approved ('Authority to Operate').

1. [MS Word](#)
2. [Zoom](#)
3. [MS Teams](#)



What is transcription?

Transcribing audio into text is part of the research process for many researchers. Manually transcribing text can be time consuming, so an increasing number of researchers are using software to transcribe, for example, interview or focus group audio recordings.

Considerations for selecting a transcription method

Approved Transcription Software

Members of the University community have access to the following enterprise software that has been approved by the Chief Information Security Office for use with **internal and sensitive data**, in compliance with the [Research data classification standard](#) and [Research Data Management Policy](#) and [guidance](#).

Transcribe audio files with:

- **Microsoft Word** (web and desktop application). Instructions are provided [here](#).

Record and transcribe with:

- **Zoom** (auckland.zoom.us). Go to [Zoom Support](#) for more information about automated transcription of audio from Zoom meetings or webinars.
- **MS Teams**. Go to [MS Teams Recording options](#) for more information about automated transcription and editing of audio from Teams meetings.

Other software options (e.g. Otter, Rev, Scribie) are not approved for use with internal or sensitive research data.

External transcription services

Researchers may want to use a paid transcription service where the transcriber is based in Aotearoa New Zealand. Faculties, LRSI, and Schools may maintain lists of service providers.

If using an external transcription service, be aware of the need for appropriate file transfer and agreements, see [Research data classification standard](#) and [Research data management policy guidance](#).



Chat: Would you use an AI transcription tool that didn't share the sensitive data beyond your organisation, project team or computer?



Automatic Speech Recognition (ASR) OpenAI-Whisper



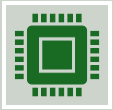
What is Whisper?

- A general-purpose speech recognition Neural Network (specifically, an encoder-decoder transformer network).
- Trained on a large dataset of diverse multi-lingual audio (98 languages).
- Able to perform multi-lingual speech recognition, speech translation, and language identification.

Types of models



Whisper is released in multiple model sizes



Tiny, Base (39M, 74 M parameters)
requires ~1-2GB of RAM.

Less accurate but faster to run



Small, Medium (244M, 769 M
parameters) requires ~2-8GB of RAM.

More accurate but slower to run relative to
above.



Large (1550 M parameters) requires
~10-16GB of RAM.

The most accurate but slowest to run.

Languages supported

Although 98 languages are supported, the following languages are with an accuracy of 50% or more.

Afrikaans, Arabic, Armenian, Azerbaijani, Belarusian, Bosnian, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, Galician, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Japanese, Kannada, Kazakh, Korean, Latvian, Lithuanian, Macedonian, Malay, Marathi, **Māori**, Nepali, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swahili, Swedish, Tagalog, Tamil, Thai, Turkish, Ukrainian, Urdu, Vietnamese, and Welsh.



Features

- All whisper models can be ***deployed locally*** with no talk back to base (i.e. all data will be stored on-device and not shared with openai.com).
- Transcription:
 - Ignores silence, easy to use and deploy AI models.
- Translation:
 - Speech-to-text, any-to-English translation (Accuracy may vary based on input language).
- Language identification:
 - Identifies the 98 supported languages.





DEMO

- *Different Model Sizes*
- *Speed of translation and accuracy*
- *Local and Virtual Machine*



Resources:

Whisper's website:

- <https://openai.com/index/whisper/>

Other ASR models:

- https://huggingface.co/spaces/hf-audio/open_asr_leaderboard

Vibe application:

- <https://thewh1teagle.github.io/vibe/>



Questions?

researchdata@auckland.ac.nz

July 2024

