G-Cloud 14

Multilingual Speech to Text

Service definition document document

Contents

1. Service Information

Transform unstructured speech data into accurate transcriptions. Gain deeper insights into speech data with our state-of-the-art AI powered speech-to-text technology.

Our Speech to Text Performance

Emotech's Speech to Text is state-of-the-art with consistently high accuracy across English, Modern Standard Arabic (MSA) and major dialects, including Saudi, Gulf and Egyptian.

Easy to Use

Our API was built by developers, for developers. Extensive and thorough documentation makes it easy for any-one to integrate our AI into their applications.

Latest AI Breakthroughs

Our team of expert Research Scientists and Engineers are continuously updating our models to ensure continued state-of-the-art performance.

1.1 Use Cases:

1.1.1 Government

Analyse at scale, whilst keeping your data safe

Unparalleled accuracy and efficiency in transcribing crucial meetings and communications, with the option of on-premise deployments to ensure your data never leaves your environment.

1.1.2 Smart Offices

Maximise your team's efficiency and collaboration

Generate efficient meeting documentation and allow for easy retrieval of key meeting actions and insights.

1.1.3 Call Centres

Enhance the efficiency of your call centres

Gain better insights about your customers, whilst ensuring precise transcriptions for improved service quality and compliance.

1.1 Supported Languages

The platform currently supports the following languages:

1. English
2. Arabic - Modern Standard Arabic (MSA)
3. Arabic - Saudi Arabia (Dialect)
4. Arabic - UAE (Dialect)
5. Arabic - Egypt (Dialect)

1.2 Audio format supported

The platform has 2 modes: File upload and Real-time transcription.

- File upload:

  The file upload feature supports 2 audio formats: MP3 and WAV

  It supports up to 2 channels in a file.

- Real-time:

  The platform supports real-time transcription. See 1.3 Access through the UI for more information

1.3 Access through the UI

1.3.1 Sign-up and Log-in

Accessing the multilingual speech platform is straightforward. Users only need to use the browser such as Google Chrome or Microsoft Edge to visit the URL of the platform, where the platform will prompt the user to Sign-up or login (Figure 1). New users are able to sign up using email addresses, and set up passwords.

For users covered by the contract, they will be directed directly to the user interface and be able to transcribe as soon as they are logged in.

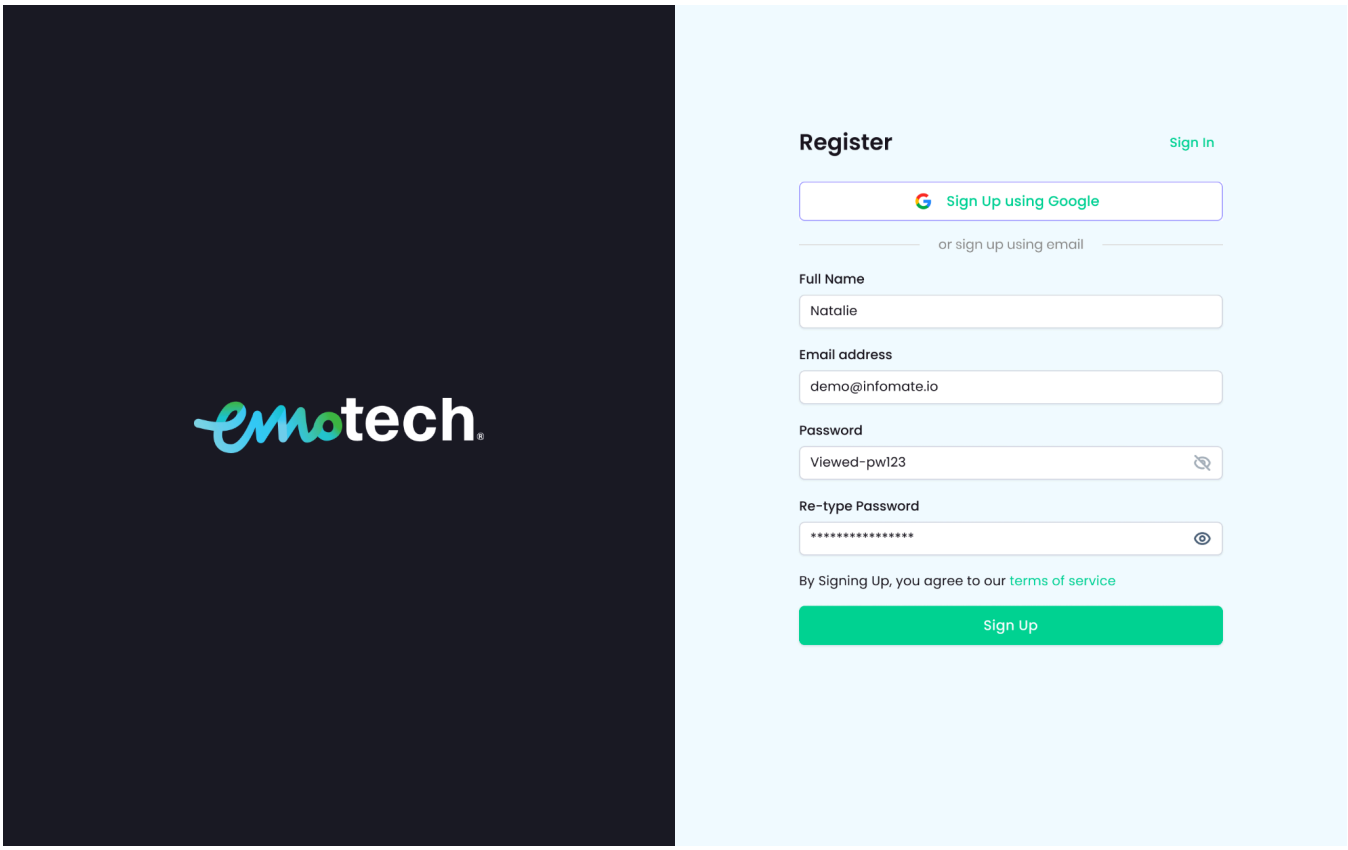

Figure 1: Sign-up and Log-in

1.3.2 Speech Recognition - Upload File & Streaming

Once entering the user interface designed for the transcribing feature, users will be able to perform speech recognition tasks. There are two modes.

1. File upload

   It allows the user to upload the recordings they would like to transcribe. MP3 and WAV are supported. If the recording uploaded includes 2 audio channels, both channels will be transcribed and listed. As shown in Figure 2

2. Streaming

   If the user would like to transcribe their speech in real-time as a voice memo, users can select the “Streaming” option, where the platform will transcribe user speech and show results on the fly.

   To perform the streaming task, the browser will prompt the user to give permission on the microphone usage.

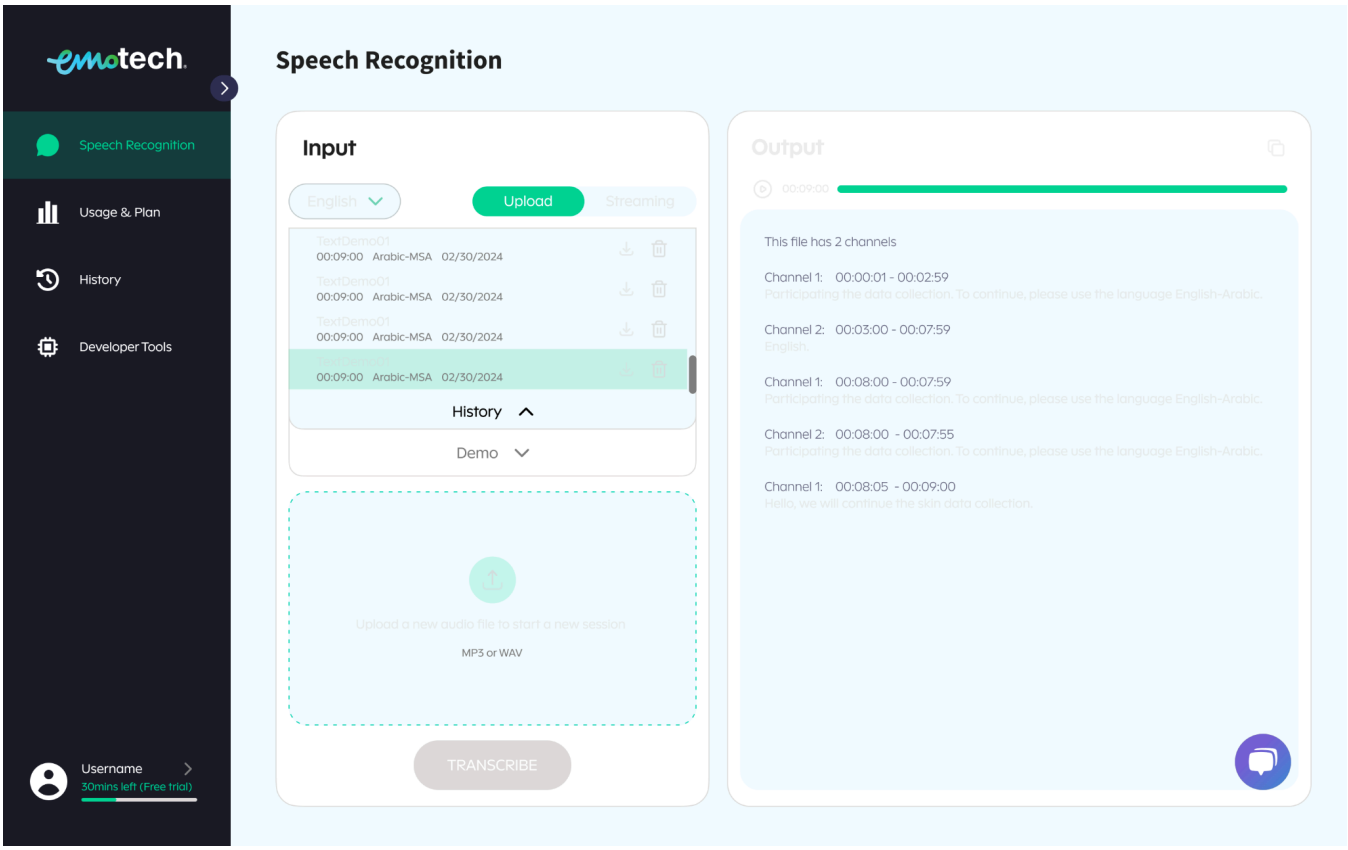

Figure 2: Speech Recognition Interface

1.3.3 History

To review previously performed transcription tasks and delete unwanted records, users can visit the History section. Here the user will see the list of transcription tasks, download the audio recordings, review transcription, and delete the record (Figure 3).

1.4 Access with API (Developer Tools)

For bulk access and integration, API access is available. The API access key and the access information are available at the “Developer Tools” section (Figure 4).

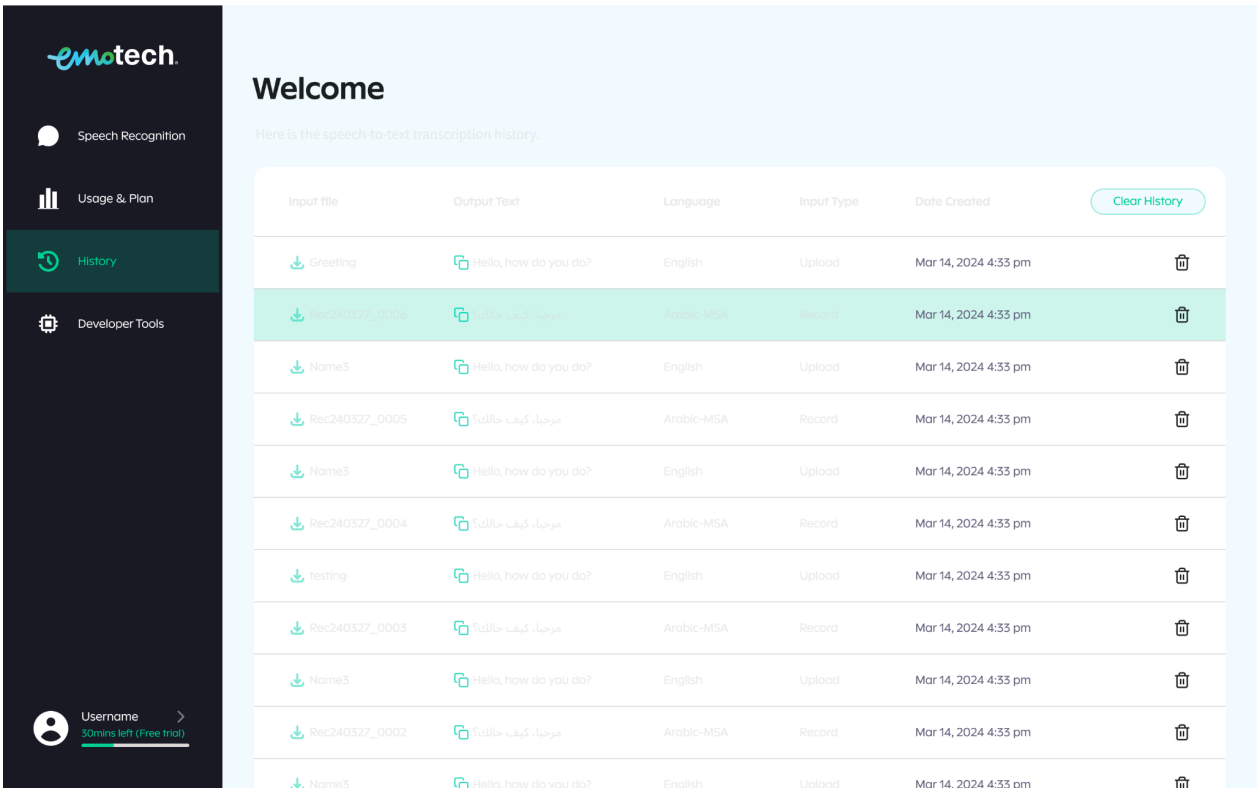

Figure 3: History of previous transcription tasks

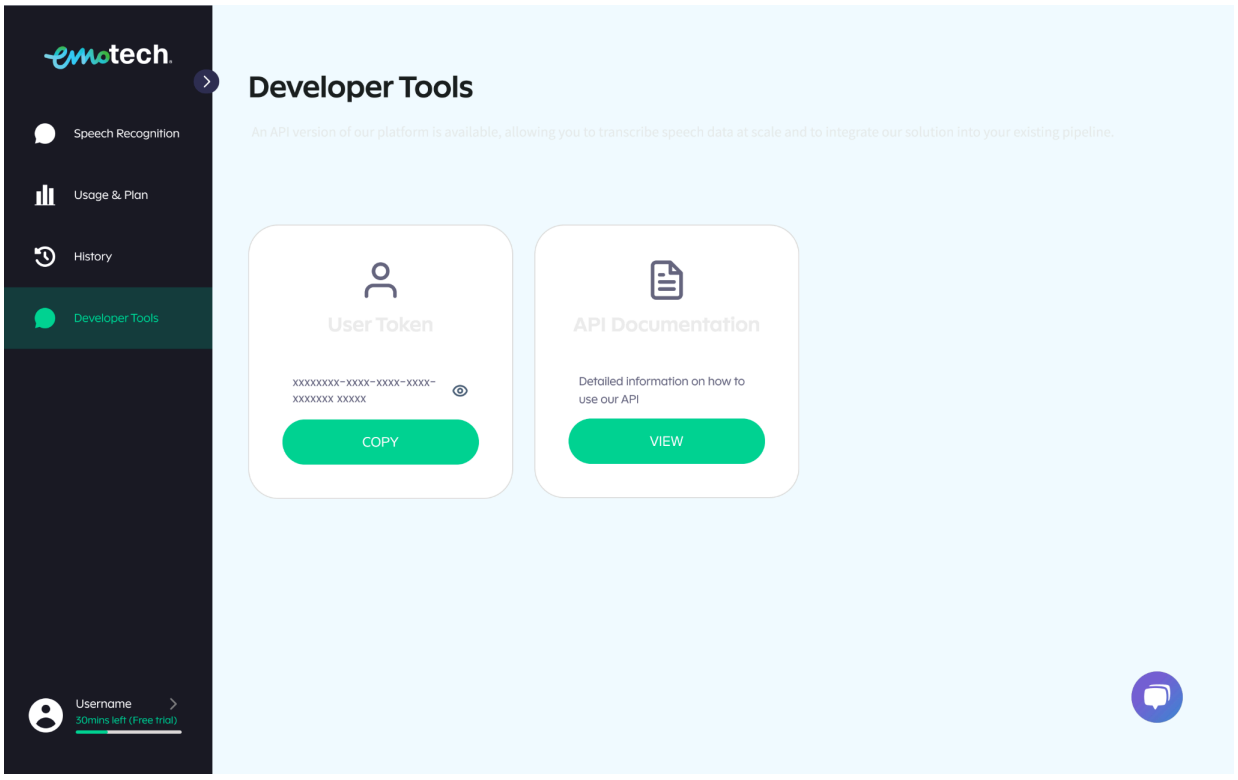

Figure 4: Developer Tools for API access

2. Solution Advantages

Emotech stands at the forefront with its innovative Interactive Digital Avatars. Utilising state-of-the-art artificial intelligence, including the latest Large Language Models (LLMs) and Retrievable Augmented Generation (RAG) technologies, these avatars enhance digital interactions across public sector interfaces, offering a dynamic and user-friendly solution that revolutionises how public services communicate with citizens.

2.1 Advanced AI and Latest Technologies

Emotech's Digital Ambassadors are powered by advanced Speech-to-Text (STT) and Text-to-Speech (TTS) capabilities and are further enhanced by the latest developments in LLMs and RAG. This integration allows the avatars to process and generate responses with a depth of understanding and contextual relevance previously unattainable, facilitating real-time interactions that closely mimic human nuances with exceptional accuracy. The use of SYNC.AI technology ensures that these interactions are not only realistic but also engaging.

The flexibility of Emotech's avatars enables them to perform a wide range of functions, from fielding enquiries to facilitating complex transactions and data visualisations. This versatility makes them indispensable across various governmental departments, enhancing accessibility and improving efficiency.

2.2 Bespoke Customisation

A key advantage of Emotech's Digital Ambassadors is their high degree of customisation. Each avatar is meticulously crafted to reflect the branding and ethos of the specific governmental department it serves, ensuring it resonates well with its intended audience. This customisation extends beyond visual aesthetics to include tailored functionalities that align with unique operational needs, seamlessly integrating into existing digital infrastructures.

2.3 Uncompromised Security

Recognising the critical importance of security, Emotech offers robust security options tailored to each deployment, with flexible solutions available for both cloud-based and on-premises implementations. These systems are designed to meet stringent security standards, safeguarding sensitive data while providing scalable and reliable performance. Emotech's infrastructure supports comprehensive data integration capabilities, enabling seamless connections with existing customer databases and ensuring data integrity is maintained at all times.

2.4 Diverse Sector Applications

Emotech's Interactive Digital Avatars are transforming public sector engagements by making services more accessible and interactive. From healthcare to education and local government, these avatars are setting new standards for service delivery. For instance, in healthcare, avatars can guide patients through treatment options and appointment scheduling, while in education, they can support learning by interacting dynamically with students.

2.5 Multilingual

Emotech's Interactive Digital Avatars are designed to operate seamlessly across more than 20 languages, making them an ideal solution for diverse cultural and linguistic environments within the UK and beyond. This multilingual capability ensures that all users, regardless of their language, can interact effectively with public services. By incorporating extensive language support, Emotech's avatars can deliver personalised and inclusive experiences, thus broadening access and enhancing communication across various sectors of government.

2.6 Low Latency Performance

A critical feature of Emotech's Interactive Digital Avatars is their low-latency response times, which are essential for maintaining smooth and efficient user interactions. This is only achievable with Emotech as Emotech has full control of the pipeline from Speech-to-Text, LLMs, Text-to-Speech, and Avatar animation. This rapid performance allows the avatars to provide real-time feedback and support to users. Such speed is crucial in dynamic environments where timely information and swift service are paramount, ensuring that user engagements are not only effective but also satisfying.

2.7 Accurate Lip Sync Technology

The realism of Emotech's avatars is significantly enhanced by their accurate lip-sync capabilities, thanks to advanced SYNC.AI technology. This feature ensures that the avatars' lip movements are perfectly synchronised with their spoken output, enhancing the natural feel of interactions and boosting the credibility of the digital representatives. Accurate lip-syncing is particularly important in delivering clear and comprehensible communications, which can improve user trust and engagement in digital government services.

2.8 Future-Proofing Public Services

As Emotech continues to innovate, the potential for these Interactive Digital Avatars within the G-Cloud framework is expansive. Future developments are expected to include even greater emotional intelligence and deeper learning capabilities, which will further personalise and enhance the user experience, making public service interactions as intuitive and natural as talking to a human.

2.9 Conclusion

Emotech's commitment to developing these sophisticated digital solutions reflects its dedication to improving public sector communication. As digital needs evolve, Emotech's avatars stand ready to meet these challenges, promising a smarter, more connected future for public services.

Governmental bodies are encouraged to explore how Emotech's Interactive Digital Avatars can revolutionise their service offerings. For further details or to arrange a demonstration, please visit Emotech's website or contact their dedicated public sector team.

3. Emotech introduction

3.1 Introduction

Emotech has more than eight years of experience in developing cutting-edge AI Solutions including Speech Technologies, Language Understanding and Digital Avatars. We have long-term collaborations with world-leading academic institutions such as University College London, Cambridge, Oxford, etc. Our solutions are serving many clients in the Middle East including UAE and KSA. For our products we received numerous awards including four innovation awards at the most prestigious Consumer Electronics Show (CES) in Las Vegas.

At the heart of our organisation is an exceptional group of more than 40 individuals, embodying the dynamic spirit of entrepreneurship. Emotech has three exceptional co-founders covering three main pillars of any successful AI business: Product, Investments and Technology. We have noticeable academic advisors such as John Shawe Taylor who is also a chairman of the UNESCO AI and Yvonne Rogers who is a Renowned researcher in the field of human-computer interactions and a winner of the British Academy of Engineering and Physical Sciences Award. Both from the world-leading university UCL in London. Our team members come from diverse backgrounds and they have experience working in companies such as Apple, Santander, Ogilvy.

For more information, please visit our website at https://www.emotech.ai/

3.2 EMOTECH'S ADVANTAGES

Below are highlighted several Emotech's advantages but not limited to. The comparison is related to "Industrial supplier" which can be any other AI company or "Academic supplier" that represents research institutes which usually have published fancy technologies. However, their technologies are not proven to be working in the commercial environments. Therefore, it is very risky to count on them to turn their internal research projects into workable solutions solving real-world use-cases.

	Emotech	Industrial supplier	Academic supplier	Note
Deployed NLP Solutions	YES	LIMITED	NO	Emotech has significant experience running NLP deployments across the Middle East including UAE and KSA.
Understanding use-case	HIGH	MED	LOW	From working for several years in the NLP industry and

				through real deployments of NLP solutions, Emotech has gained valuable insights of market needs
Level of support	HIGH	MID	MID	Emotech provides exceptional support by leading experts in the industry.
Delivery time of a working solution	8 months	12 months	18 months	Emotech technology readiness is above anyone else. Phase 1 solution is scheduled to be delivered within 2.5 months.
Production-designed systems	HIGH	HIGH	LOW	It requires a lot of expertise and industrial experience to build robust engineering solutions that are highly scalable, modular and secure.
Industrial/user experience	HIGH	HIGH	LOW	One of our academic advisors is Yvonne Rogers who is a Renowned researcher in the field of human-computer interactions and a winner of the British Academy of Engineering and Physical Sciences Award
Product Design	HIGH	MID	NO	It's not just about technology. Emotech offers experienced product design to build solutions that will benefit users. We received numerous awards for our products including four times CES innovation award.
Years of experience in AI industries	60	?	?	Emotech's researchers have accumulated more than 60 years of combined experience working in AI industries.

Arabic Speech Technologies	HIGH	MID	HIGH	The Arabic language is very complex including many dialects and different written forms. It's not common to have as many insights as Emotech.
Languages spoken by the team	>20	?	?	More than 20 different languages are spoken within the Emotech multinational team.
RTF (real-time factor)	LOW	MID	HIGH	RTF is very crucial to ensure a responsive and smooth user experience. Emotech's strong engineering team has many years of experience by optimising service execution times.