



AZURE DAY ROME 2024



Un avatar con Azure OpenAI Services il
next step di ChatGPT?



Alessio Iafrate



Microsoft



avanade



eng

software one





Thanks to

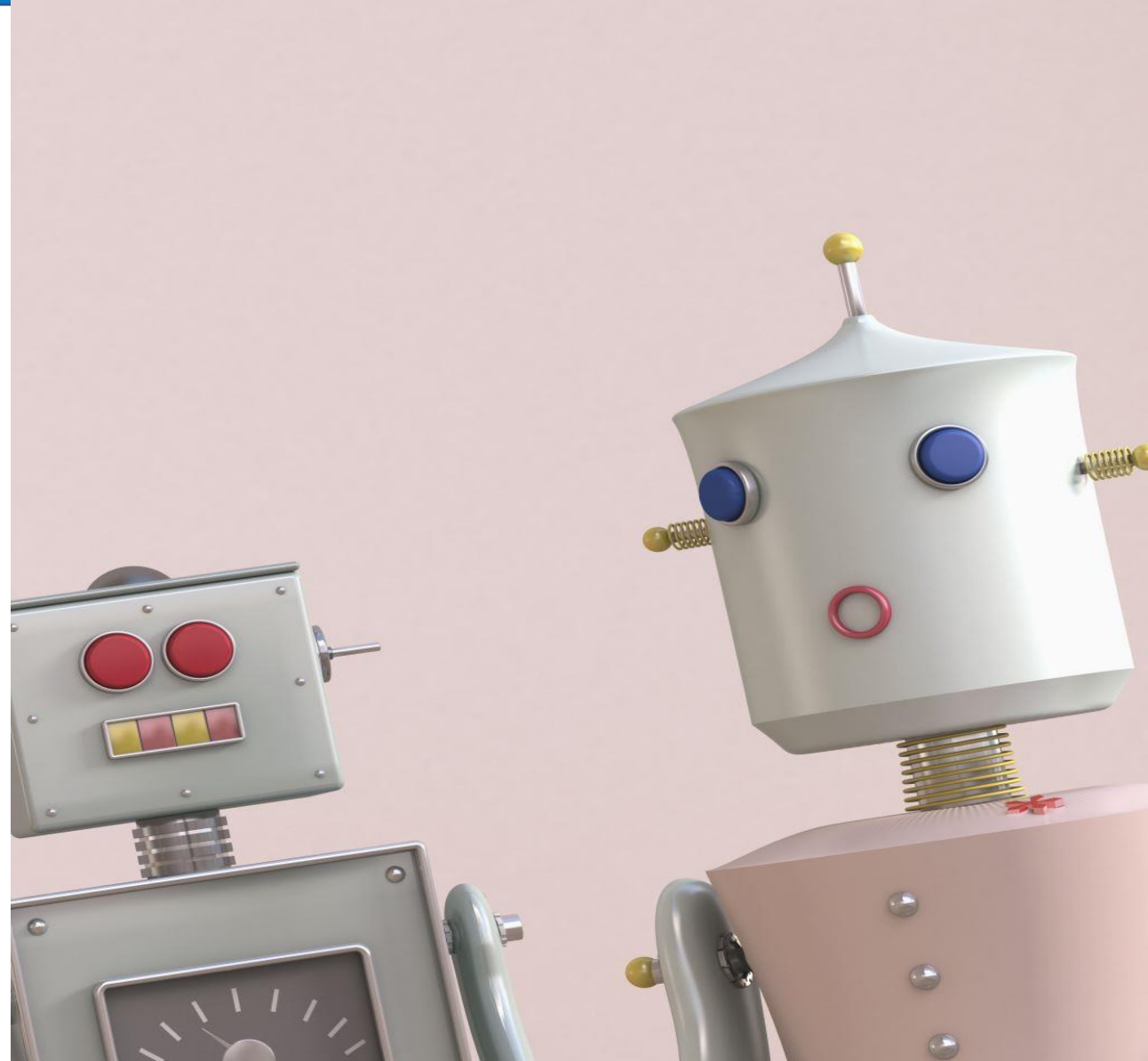




AZURE DAY

Disclaimer

In questa
sessione nessun
bot di ChatGpt o
Azure OpenAI è
stato maltrattato





AZURE DAY

Azure AI services (ex cognitive services)

Product	Features
Azure OpenAI Service	Power your apps with large-scale AI models.
Azure AI Search	Enterprise scale search for app development.
Azure AI Content Safety	Use AI to monitor text and image content for safety.
Azure AI Document Intelligence	Accelerate information extraction from documents.
Azure AI Speech	Transcribe, translate and generate spoken audio.
Azure AI Language	Identify, analyze and summarize text with natural language processing.
Azure AI Translator	Real-time machine translation for documents and text.
Azure AI Vision	Analyze, extract and categorize information from images.



AZURE DAY

Azure AI Speech



Convert speech to text

Quickly and accurately transcribe audio in more than 100 languages and variants. Gain customer insights with call center transcription, improve experiences with voice-enabled assistants, capture key discussions in meetings and more.

[Learn more about Speech to Text >](#)

[Get started transcribing speech today >](#)



Give your app a voice

Use text to speech to create apps and services that speak conversationally. Create natural-sounding [audio content](#), improve accessibility with read-aloud functionality, and create custom voice assistants.

[Learn more about Text to Speech >](#)

[Learn how to convert text to audio >](#)



Translate speech in real time

Translate audio from more than 30 languages and customize translations for your organization's specific terms—all in your preferred programming language.

[Learn more about Speech Translation >](#)

[Get started translating speech in real time >](#)



Verify and recognize speakers

Confirm a person's identity or recognize who's speaking in a meeting by adding speaker verification and identification to your app.

[Learn more about Speaker Recognition >](#)

[Learn how to recognize speakers in your app >](#)



Activate your assistant or IoT device with a custom keyword

Create a custom keyword for IoT devices and voice-enabled assistants to set your brand apart—making it more personal, personable, and secure.

[Learn how to create a custom keyword >](#)

[Get started building voice-controlled apps >](#)



Add voice commands for hands-free scenarios

Build a touchless, voice-first experience to improve safety and support back-to-work scenarios.

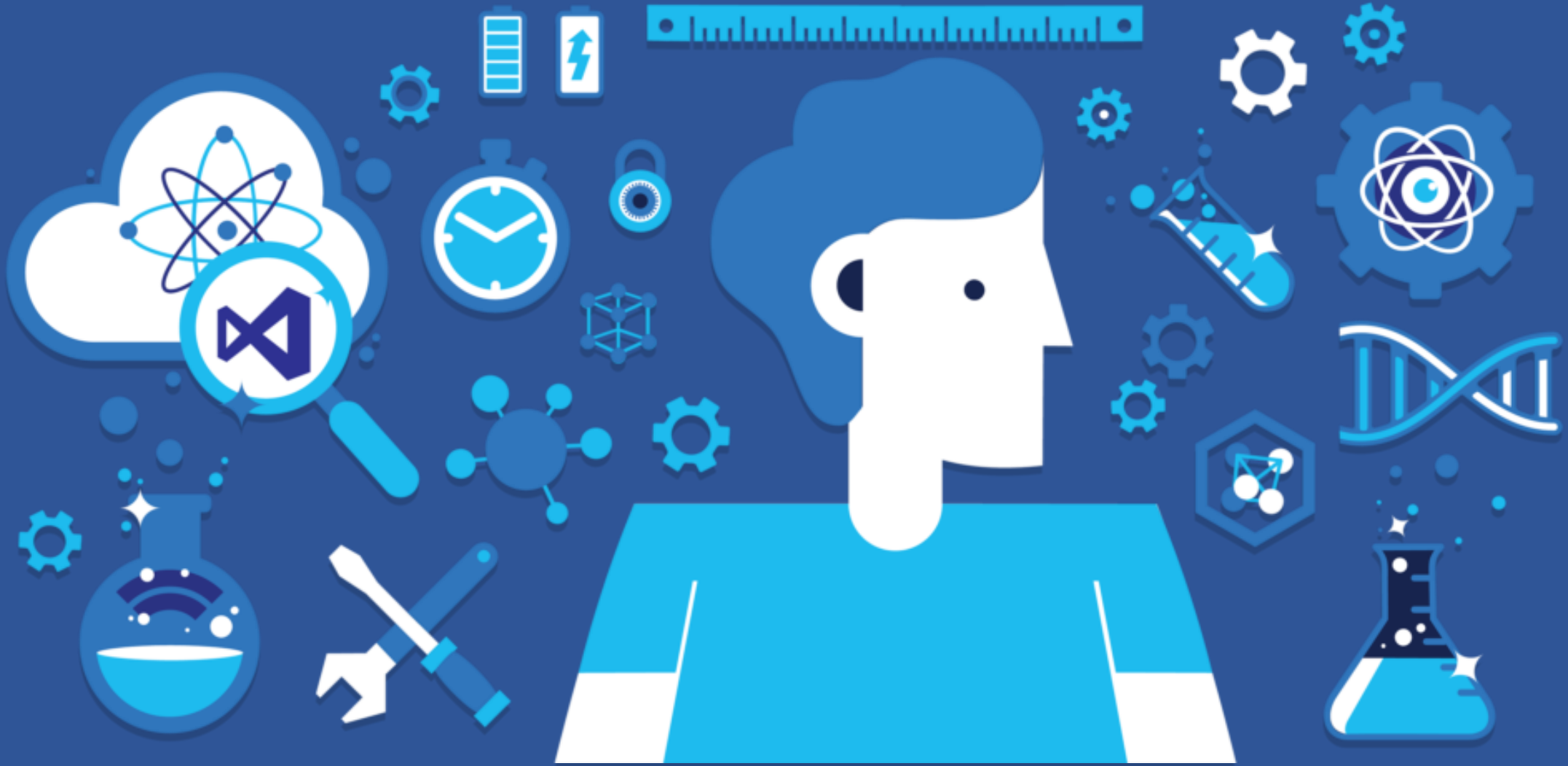
[Learn more about Custom Commands >](#)

[Get started adding Custom Commands >](#)



AZURE DAY

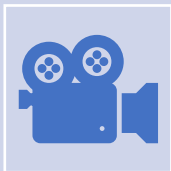
Azure AI Speech text to speech avatar (preview)





AZURE DAY

Quando utilizzare un avatar



La creazione tradizionale di contenuti video richiede molto tempo e budget, inclusa la configurazione dell'ambiente di ripresa video, la ripresa di video, l'editing, ecc. Con l'avatar di sintesi vocale, gli utenti possono creare video in modo più efficiente. Gli utenti possono utilizzare l'avatar per creare video di formazione, presentazioni di prodotti, testimonianze di clienti, ecc., semplicemente inserendo testo.



Con il rilascio del servizio Azure OpenAI e della sintesi vocale neurale, la conversazione interattiva è più naturale di prima. Con l'avatar di sintesi vocale, gli utenti possono creare interazioni digitali più coinvolgenti. Puoi utilizzare l'avatar per creare agenti conversazionali, assistenti virtuali, chatbot e altro ancora.



AZURE DAY

Workflow



analizzatore di testo

fornisce l'output sotto forma di sequenza di fonemi



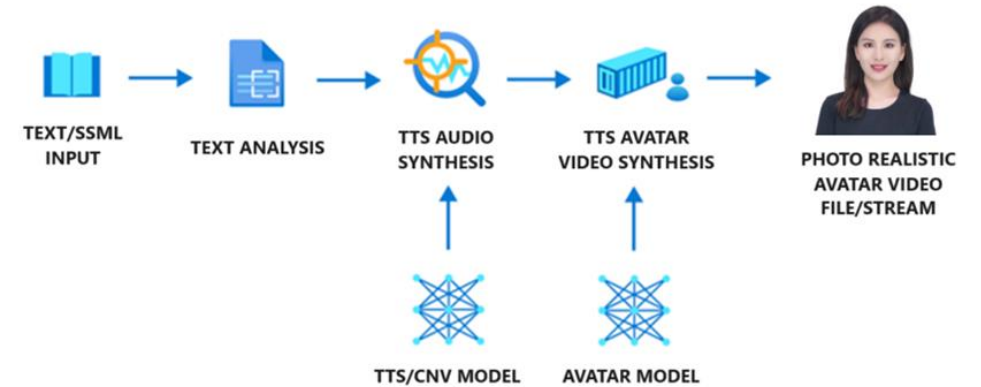
il sintetizzatore audio TTS

predice le caratteristiche acustiche del testo di input e sintetizza la voce



sintetizzatore video avatar TTS

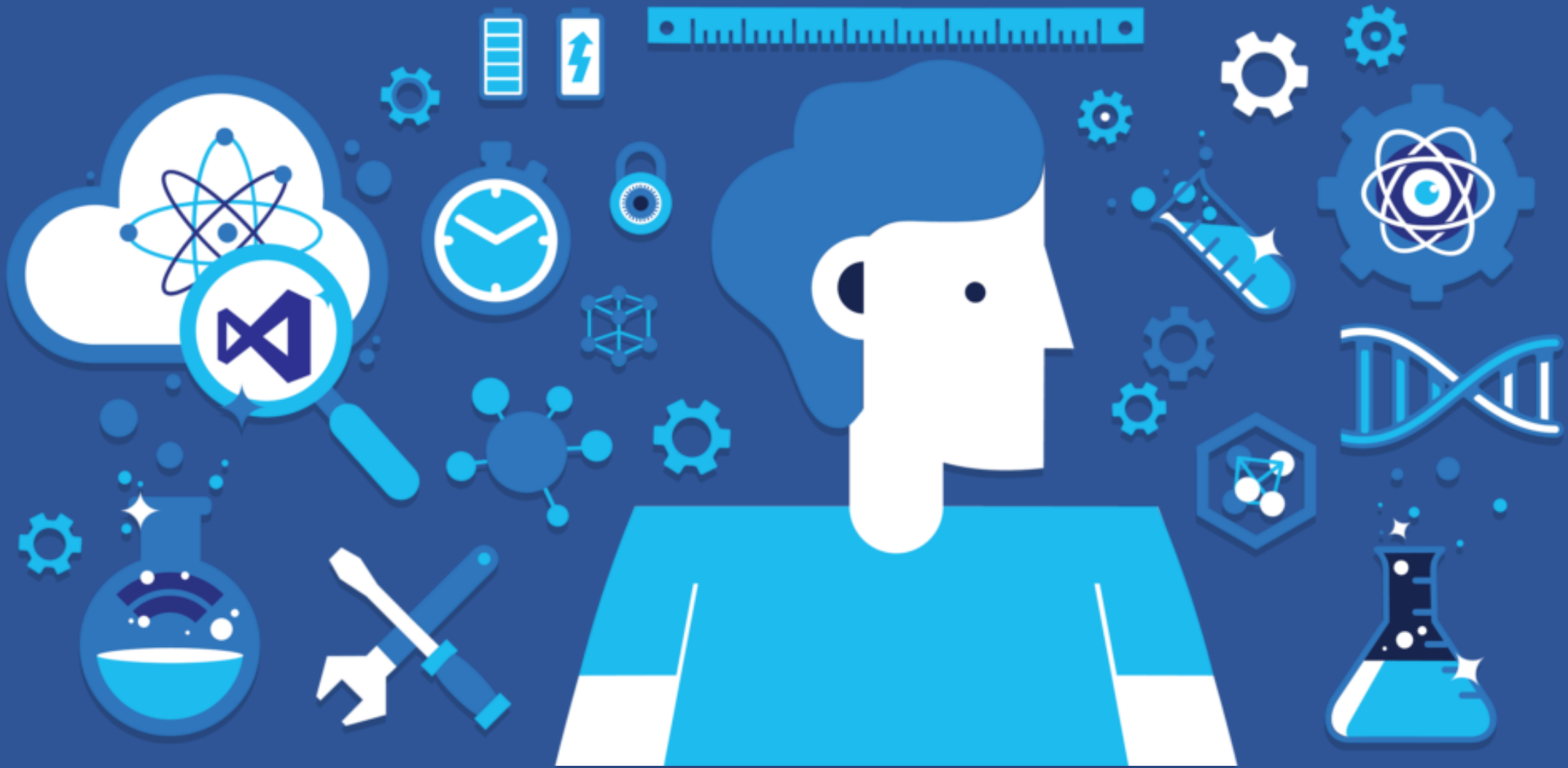
prevede l'immagine della sincronizzazione labiale con le caratteristiche acustiche, in modo da generare il video sintetico





AZURE DAY

Demo





AZURE DAY

AI text to speech avatar featur

Converte il testo in un video digitale di un essere umano fotorealistico che parla con voci dal suono naturale basate sulla sintesi vocale di Azure AI.

Fornisce una raccolta di avatar predefiniti.

La voce dell'avatar viene generata dalla sintesi vocale di Azure AI. Per ulteriori informazioni, consulta Voce e lingua dell'Avatar.

Sintetizza video avatar sintesi vocale in modo asincrono con l'API di sintesi batch o in tempo reale.

Fornisce uno strumento di creazione di contenuti in Speech Studio per creare contenuti video senza codifica.

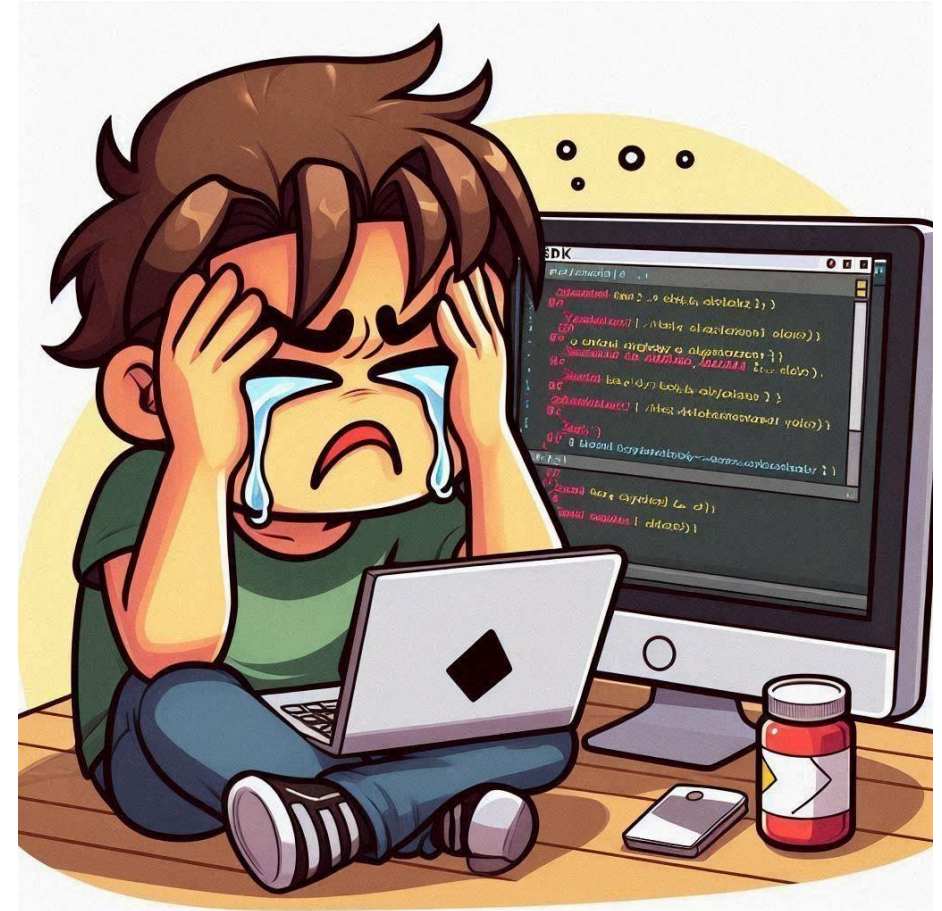
Abilita conversazioni avatar in tempo reale tramite lo strumento avatar della chat dal vivo in Speech Studio.



AZURE DAY

Sdk

Solo Javascript al momento!





AZURE DAY

Compatibilità sdk javascript

Platform	Chrome	Microsoft Edge	Safari	Firefox	Opera
Windows	Y	Y	N/A	Y ¹	Y
Android	Y	Y	N/A	Y ¹²	N
iOS	Y	Y	Y	Y	Y
macOS	Y	Y	Y	Y ¹	Y


1 It doesn't work with ICE server by Communication Service but works with Coturn.


2 Background transparency doesn't work.





AZURE DAY

Creaiamo la risorsa

 **Azure Synapse Analytics**
[Create](#) | [Docs](#) | [MS Learn](#)

 **Azure AI services**
[Create](#) | [Docs](#) | [MS Learn](#)

 **Speech**
[Create](#) | [Docs](#) | [MS Learn](#)

 **Language service**
[Create](#) | [Docs](#) | [MS Learn](#)



azureavatardemo | Keys and Endpoint ☆ ...

Azure AI services multi-service account






 Search










Regenerate Key1




Regenerate Key2

-  Overview
-  Activity log
-  Access control (IAM)
-  Tags
-  Diagnose and solve problems

Resource Management

-  **Keys and Endpoint**
-  Pricing tier
-  Networking
-  Identity
-  Cost analysis
-  Properties
-  Locks

 These keys are used to access your Azure AI services API. Do not share your keys. Store them securely– for example, using Azure Key Vault. We also recommend regenerating these keys regularly. Only one key is necessary to make an API call. When regenerating the first key, you can use the second key for continued access to the service.

Show Keys

KEY 1

.....



KEY 2

.....



Location/Region ⓘ

westeurope



Endpoint

https://azureavatardemo.cognitiveservices.azure.com/





AZURE DAY

Disponibilità

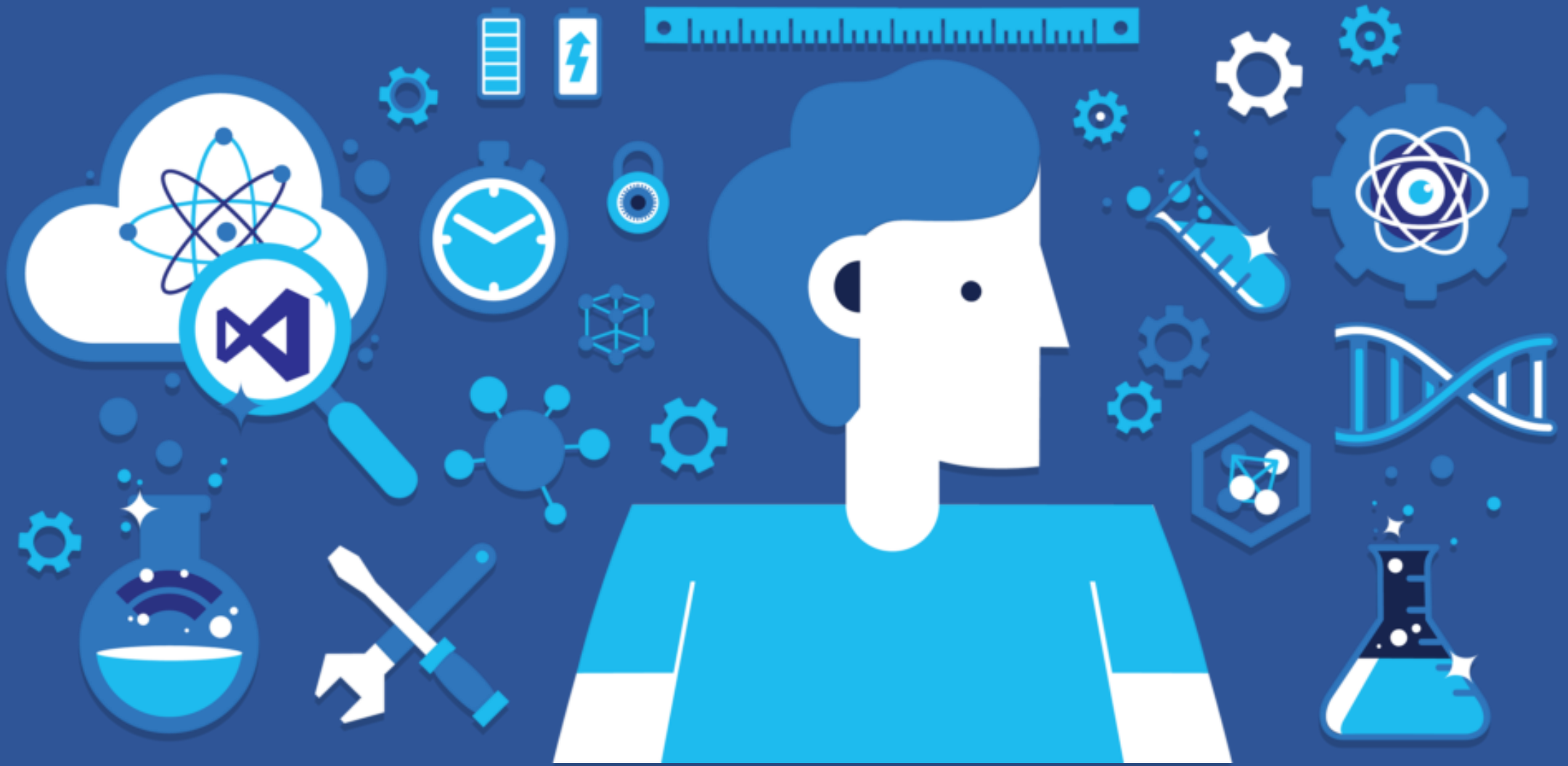
- West US 2
- West Europe
- Southeast Asia





AZURE DAY

Demo sdk





AZURE DAY

SDK - Configurazione

```
const speechConfig = SpeechSDK.SpeechConfig.fromSubscription("YourSpeechKey", "YourSpeechRegion");  
// Set either the `SpeechSynthesisVoiceName` or `SpeechSynthesisLanguage`.  
speechConfig.speechSynthesisLanguage = "en-US";  
speechConfig.speechSynthesisVoiceName = "en-US-AvaMultilingualNeural";
```

```
const avatarConfig = new SpeechSDK.AvatarConfig(  
    "lisa", // Set avatar character here.  
    "casual-sitting", // Set avatar style here.  
);
```



AZURE DAY

SDK - Connessione

```
GET /cognitiveservices/avatar/relay/token/v1 HTTP/1.1
```

```
Host: westus2.tts.speech.microsoft.com
```

```
Ocp-Apim-Subscription-Key: YOUR_RESOURCE_KEY
```

```
// Create WebRTC peer connection
peerConnection = new RTCPeerConnection({
  iceServers: [{
    urls: [ "Your ICE server URL" ],
    username: "Your ICE server username",
    credential: "Your ICE server credential"
  }]
})
```



AZURE DAY

WebRTC

WebRTC (Web Real-Time Communication) is a [free and open-source](#) project providing [web browsers](#) and [mobile applications](#) with [real-time communication](#) (RTC) via [application programming interfaces](#) (APIs). It allows audio and video communication and streaming to work inside web pages by allowing direct [peer-to-peer](#) communication, eliminating the need to install [plugins](#) or download native apps.^[3]

Supported by [Apple](#), [Google](#), [Microsoft](#), [Mozilla](#), and [Opera](#), WebRTC specifications have been published by the [World Wide Web Consortium](#) (W3C) and the [Internet Engineering Task Force](#) (IETF).^{[4][5]}

According to the [webrtc.org](#) website, the purpose of the project is to "enable rich, high-quality RTC applications to be developed for the browser, mobile platforms, and [IoT](#) devices, and allow them all to communicate via a common set of [protocols](#)".^[6]



AZURE DAY

SDK – Start session

```
// Fetch WebRTC video/audio streams and mount them to HTML video/audio player elements
peerConnection.ontrack = function (event) {
  if (event.track.kind === 'video') {
    const videoElement = document.createElement(event.track.kind)
    videoElement.id = 'videoPlayer'
    videoElement.srcObject = event.streams[0]
    videoElement.autoplay = true
  }

  if (event.track.kind === 'audio') {
    const audioElement = document.createElement(event.track.kind)
    audioElement.id = 'audioPlayer'
    audioElement.srcObject = event.streams[0]
    audioElement.autoplay = true
  }
}

// Offer to receive one video track, and one audio track
peerConnection.addTransceiver('video', { direction: 'sendrecv' })
peerConnection.addTransceiver('audio', { direction: 'sendrecv' })
```

```
// Create avatar synthesizer
var avatarSynthesizer = new SpeechSDK.AvatarSynthesizer(speechConfig, avatarConfig)

// Start avatar and establish WebRTC connection
avatarSynthesizer.startAvatarAsync(peerConnection).then(
  (r) => { console.log("Avatar started.") }
).catch(
  (error) => { console.log("Avatar failed to start. Error: " + error) }
);
```



AZURE DAY

SDK – Let's talk

```
var spokenText = "I'm excited to try text to speech avatar."
avatarSynthesizer.speakTextAsync(spokenText).then(
  (result) => {
    if (result.reason === SpeechSDK.ResultReason.SynthesizingAudioCompleted) {
      console.log("Speech and avatar synthesized to video stream.")
    } else {
      console.log("Unable to speak. Result ID: " + result.resultId)
      if (result.reason === SpeechSDK.ResultReason.Canceled) {
        let cancellationDetails = SpeechSDK.CancellationDetails.fromResult(result)
        console.log(cancellationDetails.reason)
        if (cancellationDetails.reason === SpeechSDK.CancellationReason.Error) {
          console.log(cancellationDetails.errorDetails)
        }
      }
    }
  })
  .catch((error) => {
    console.log(error)
    avatarSynthesizer.close()
  });
```



AZURE DAY

Custom TTS

Il modello di avatar di sintesi vocale personalizzato contiene tre componenti: analizzatore di testo, sintetizzatore audio di sintesi vocale e renderer video avatar di sintesi vocale.

Per generare un file video avatar o uno streaming con il modello avatar, il testo viene prima inserito nell'analizzatore di testo, che fornisce l'output sotto forma di sequenza di fonemi.

Il sintetizzatore audio sintetizza l'audio vocale per il testo di input e queste due parti sono fornite da sintesi vocale o modelli vocali neurali personalizzati.

Infine, il modello neurale dell'avatar text-to-speech prevede l'immagine della sincronizzazione labiale con l'audio del parlato, in modo da generare il video sintetico.

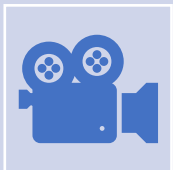


AZURE DAY

Custom TTS



L'avatar di sintesi vocale personalizzato ti consente di creare un avatar parlante sintetico personalizzato e unico nel suo genere per la tua applicazione. Con l'avatar personalizzato di sintesi vocale, puoi creare un avatar unico e dall'aspetto naturale per il tuo prodotto o marchio fornendo i dati di registrazione video degli attori selezionati. Se crei anche una voce neurale personalizzata per lo stesso attore e la usi come voce dell'avatar, l'avatar sarà ancora più realistico.



La creazione di un avatar di sintesi vocale personalizzato richiede almeno 10 minuti di registrazione video del talento avatar come dati di addestramento ed è necessario prima ottenere il consenso del talento attore.



AZURE DAY

Custom TTS steps

Ottieni video di consenso

- ottieni una registrazione video della dichiarazione di consenso. La dichiarazione di consenso è una registrazione video del talento avatar che legge una dichiarazione, dando il consenso all'utilizzo della propria immagine e dei dati vocali per addestrare un modello di avatar vocale personalizzato.

Preparare i dati di allenamento

- assicurarsi che la registrazione video sia nel formato corretto. È una buona idea girare la registrazione video in uno studio di ripresa video di qualità professionale per ottenere un'immagine di sfondo pulita. La qualità dell'avatar risultante dipende fortemente dal video registrato utilizzato per l'allenamento. Fattori come la velocità di parola, la postura del corpo, l'espressione facciale, i gesti delle mani, la coerenza nella posizione dell'attore e l'illuminazione della registrazione video sono essenziali per creare un accattivante avatar di sintesi vocale personalizzato.

Addestrare il modello avatar

- inizieremo ad addestrare il modello di sintesi vocale personalizzato dopo aver verificato la dichiarazione di consenso del talento avatar. Nella fase di anteprima di questo servizio, questo passaggio verrà eseguito manualmente da Microsoft. Riceverai una notifica una volta che il modello sarà stato addestrato con successo.

Distribuisci e utilizza il tuo modello di avatar nelle tue APP

<https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech-avatar/custom-avatar-record-video-samples>

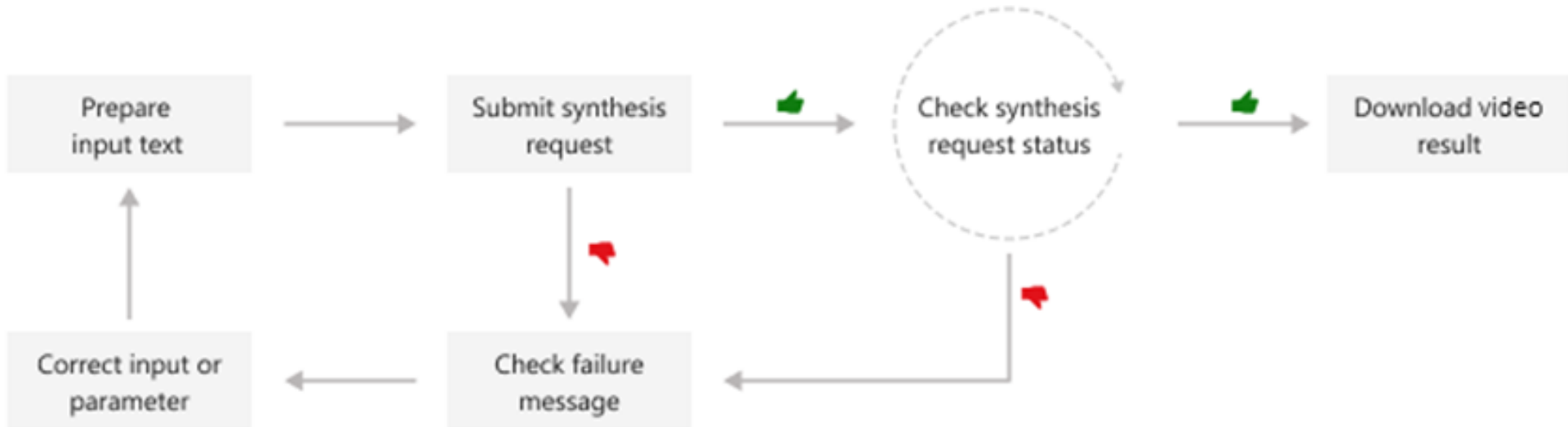


AZURE DAY

Batch synthesis

L'API di sintesi batch per avatar di sintesi vocale (anteprima) consente la sintesi asincrona del testo in un avatar parlante come file video. Gli editori e le piattaforme di contenuti video possono utilizzare questa API per creare contenuti video avatar in batch. Questo approccio può essere adatto a vari casi d'uso come materiali di formazione, presentazioni o pubblicità.

Il video avatar sintetico verrà generato in modo asincrono dopo che il sistema avrà ricevuto l'input di testo. L'output video generato può essere scaricato in modalità di sintesi batch. Invii il testo per la sintesi, esegui il polling sullo stato della sintesi e scarichi l'output video quando lo stato indica il successo. I formati di input del testo devono essere testo normale o testo SSML (Speech Synthesis Markup Language).





AZURE DAY

Batch synthesis - API

Operation	Method	REST API call
Create batch synthesis	PUT	avatar/batchsyntheses/{SynthesisId}? api-version=2024-04-15-preview
Get batch synthesis	GET	avatar/batchsyntheses/{SynthesisId}? api-version=2024-04-15-preview
List batch synthesis	GET	avatar/batchsyntheses/?api- version=2024-04-15-preview
Delete batch synthesis	DELETE	avatar/batchsyntheses/{SynthesisId}? api-version=2024-04-15-preview



AZURE DAY

Prezzi (WE)

Text to Speech⁸

Standard Voice

Neural: **€13.871** per 1M characters

Neural HD⁴: N/A per 1M characters

Custom Voice

Professional Voice:

Synthesis: **€22.194** per 1M characters

Voice model training: **€48.086** per compute hour, up to
€4,616.239 per training

Endpoint hosting: **€3.73** per model per hour

Personal Voice⁶:

Synthesis: **€22.194** per 1M characters

Voice creation: Free

Voice profile storage: **€554.837** per 1,000 voice profiles
per month

Enhanced Add-on feature: Avatar^{Preview}

Standard: **€0.925** per minute

Custom:

Real-time synthesis: **€0.925** per minute

Batch synthesis: **€1.850** per minute

Endpoint hosting: **€0.6** per model per hour



AZURE DAY

Avatar video output specs

	Batch synthesis	Real-Time synthesis
Resolution	1920 x 1080	1920 x 1080
FPS	25	25
Codec	h264/h265/vp9	h264



AZURE DAY

Risorse

Demo

- <https://github.com/a-iafrate/MauiAvatarAzDay24>

Documentazione learn

- [Text to speech avatar overview - Speech service - Azure AI services | Microsoft Learn](#)

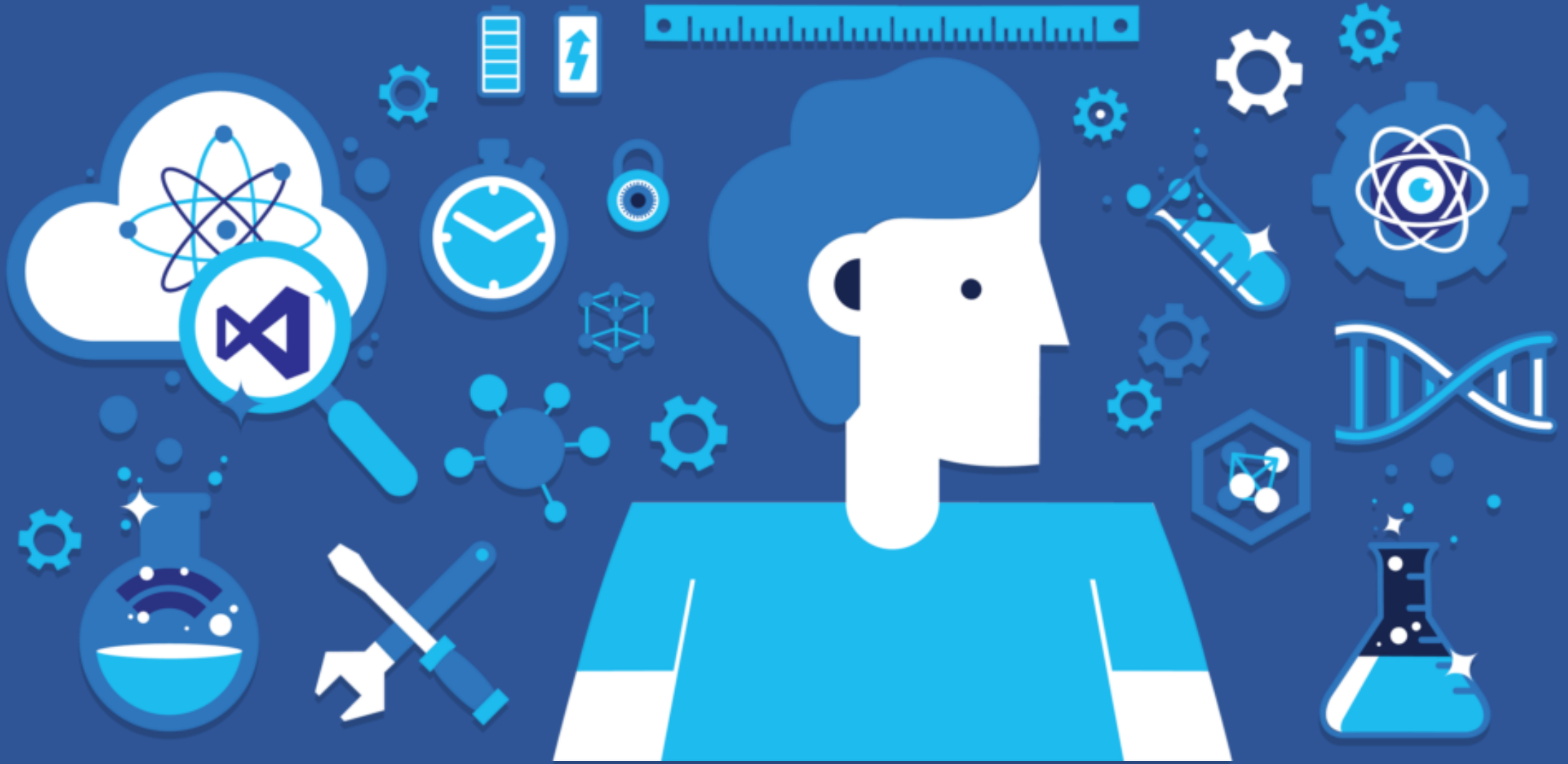
SDK js

- [cognitive-services-speech-sdk/samples/js/browser/avatar at master · Azure-Samples/cognitive-services-speech-sdk · GitHub](#)



AZURE DAY

Demo





AZURE DAY

Thank you!

Contatti

- X: @alessioiafrate
- Email: alessioiafrate@hotmail.com
- LinkedIn: alessio-iafrate
- Github: a-iafrate

Vote my session





AZURE DAY ROME 2024



VOTA LA SESSIONE

Facci sapere se questa sessione ti piace.

Inquadra il QR code e esprimi una tua opinione.

Ci aiuterai a migliorare.



Microsoft



avanade



eng

software **one**



4wardPRO
AN IMPRESOFT COMPANY

LXOBRA

UNIKEY

Packt

