



# NVIDIA Announces Availability of Jarvis Interactive Conversational AI Framework

## Pre-Trained Deep Learning Models and Software Tools Enable Developers to Adapt Jarvis for All Industries; Easily Deployed from Any Cloud to Edge

**GTC** -- NVIDIA today announced availability of the [NVIDIA Jarvis framework](#), providing developers with state-of-the-art pre-trained deep learning models and software tools to create interactive conversational AI services that are easily adaptable for every industry and domain.

With billions of hours of phone calls, web meetings and streaming broadcast video content generated daily, NVIDIA Jarvis models offer highly accurate automatic speech recognition, as well as superhuman language understanding, real-time translations for multiple languages, and new text-to-speech capabilities to create expressive conversational AI agents.

Utilizing GPU acceleration, the end-to-end speech pipeline can be run in under 100 milliseconds — listening, understanding and generating a response faster than the blink of a human eye — and can be deployed in the cloud, in the data center or at the edge, instantly scaling to millions of users.

“Conversational AI is in many ways the ultimate AI,” said Jensen Huang, founder and CEO of NVIDIA. “Deep learning breakthroughs in speech recognition, language understanding and speech synthesis have enabled engaging cloud services. NVIDIA Jarvis brings this state-of-the-art conversational AI out of the cloud for customers to host AI services anywhere.”

NVIDIA Jarvis will enable a new wave of language-based applications previously not possible, improving interactions with humans and machines. It opens the door to the creation of such services as digital nurses to help monitor patients around the clock, relieving overloaded medical staff; online assistants to understand what consumers are looking for and recommend the best products; and real-time translations to improve cross-border workplace collaboration and enable viewers to enjoy live content in their own language.

Jarvis has been built using models trained for several million GPU hours on over 1 billion pages of text, 60,000 hours of speech data, and in different languages, accents, environments and lingos to achieve world-class accuracy. For the first time, developers can use [NVIDIA TAO](#), a framework to train, adapt and optimize these models for any task, any industry and on any system with ease.

Developers can select a Jarvis pre-trained model from NVIDIA's [NGC™ catalog](#), fine-tune it using their own data with the [NVIDIA Transfer Learning Toolkit](#), optimize it for maximum throughput and minimum latency in real-time speech services, and then easily deploy the model with just a few lines of code so there is no need for deep AI expertise.

### Broad Industry Support

Since Jarvis' early access program began last May, thousands of companies have asked to join. Among early users is T-Mobile, the U.S. telecom giant, which is looking to AI to further augment its machine learning products using natural language processing to provide real-time insights and recommendations.

“With NVIDIA Jarvis services, fine-tuned using T-Mobile data, we're building products to help us resolve customer issues in real time,” said Matthew Davis, vice president of product and technology at T-Mobile. “After evaluating several automatic speech recognition solutions, T-Mobile has found Jarvis to deliver a quality model at extremely low latency, enabling experiences our customers love.”

NVIDIA is also partnering with Mozilla Common Voice, an open source collection of voice data for startups, researchers and developers to train voice-enabled apps, services and devices. The world's largest multi-language, public domain voice dataset, Common Voice contains over 9,000 total hours of contributed voice data in 60 different languages. NVIDIA is using Jarvis to develop pre-trained models with the dataset, and then offer them back to the community for free.

“We launched Common Voice to teach machines how real people speak in their unique languages, accents and speech patterns,” said Mark Surman, executive director at Mozilla. “NVIDIA and Mozilla have a common vision of democratizing voice technology — and ensuring that it reflects the rich diversity of people and voices that make up the internet.”

NVIDIA's conversational AI tools have had more than 45,000 downloads. These can be combined with technology from hundreds of partners and support leading software libraries, allowing developers worldwide to build innovative and intuitive conversational AI applications.

“Jarvis has a wide selection of pre-trained models, making it a truly end-to-end pipeline for conversational AI — from automatic speech recognition, natural language processing and text-to-speech,” said Harrison Kinsley, YouTuber and

founder of PythonProgramming.net. "All of the models are shockingly fast and well optimized and the API is easy for developers to use with examples that apply to many conversational AI tasks."

### **Availability**

Newly announced features will be released in the second quarter as part of the ongoing NVIDIA Jarvis open beta program. Developers can download it today from NGC with more information available [here](#).

[Register for free](#) to learn more about NVIDIA Jarvis during GTC21, taking place online April 12-16. Tune in to watch Huang's [GTC21 keynote](#) address streaming live on April 12 starting at 8:30 a.m. PT.

### **About NVIDIA**

NVIDIA's (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market and has redefined modern computer graphics, high performance computing and artificial intelligence. The company's pioneering work in accelerated computing and AI is reshaping trillion-dollar industries, such as transportation, healthcare and manufacturing, and fueling the growth of many others. More information at <https://nvidianews.nvidia.com/>.

Certain statements in this press release including, but not limited to, statements as to: the features, performance, benefits and availability of the NVIDIA Jarvis framework; conversational AI as the ultimate AI; NVIDIA and Mozilla having a common vision of democratizing conversational AI by giving developers powerful tools to build voice recognition applications; and the impact of NVIDIA's conversational AI tools in combination with technology from partners are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo and NGC are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Kristin Uchiyama  
Enterprise and Edge Computing  
+1-408-486-2248  
[kuchiyama@nvidia.com](mailto:kuchiyama@nvidia.com)