

A decorative graphic consisting of several overlapping, light blue, wavy lines that resemble sound waves or data patterns, set against a dark blue background.

# Proving the Art of the Possible with Natural Language Processing

Research at the Dell Technologies HPC & AI Innovation Lab is showcasing the art of the possible with deep learning for language-to-language translation and text-to-voice translation.

## **ABSTRACT**

Dell Technologies has an active research program focused on helping organizations explore, develop and adopt natural language processing applications. This research is carried out by a data sciences team in the Dell Technologies HPC & AI Innovation Lab in Austin, Texas. This white paper explores two groundbreaking projects now under way in the lab. One focuses on language-to-language translation and the other focuses on text-to-voice translation.

September 2020

## TABLE OF CONTENTS

<b>NATURAL LANGUAGE PROCESSING</b> . . . . .	<b>1</b>
<b>LANGUAGE-TO-LANGUAGE TRANSLATION</b> . . . . .	<b>1</b>
Computing resources . . . . .	2
Results . . . . .	2
<b>TEXT-TO-VOICE TRANSLATION</b> . . . . .	<b>2</b>
Computing resources . . . . .	3
Results . . . . .	3
<b>TIPS FOR YOUR PROJECT</b> . . . . .	<b>3</b>
<b>KEY TAKEAWAYS</b> . . . . .	<b>4</b>
<b>TO LEARN MORE</b> . . . . .	<b>4</b>

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying and distribution of any software described in this publication require an applicable software license.

Copyright © 2020 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, Dell Technologies, EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners.

Dell Technologies believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Published in the USA 09/20.

## NATURAL LANGUAGE PROCESSING

Natural language processing is a form of artificial intelligence that allows a computer application to understand human language, either spoken or written. The concept of NLP encompasses coding, understanding, interpreting and manipulating human language. NLP applications use computers to translate languages, convert voice to text and text to voice, and create human-like conversational agents to help customers, employees and others deal with issues, questions and concerns.

In recent years, the field of NLP has been transformed by the shift from statistical machine learning methods to the use of neural networks and deep learning. With these new approaches, it is now possible to build automated systems that can interact with people more naturally than ever before. And forward-looking businesses are seizing the day, incorporating NLP into a wide range of their processes for both customer-facing activities and internal operations.

To help organizations capitalize on this trend, Dell Technologies has an active research program focused on advancing the technologies and methodologies for the development of NLP applications. This research is carried out by a data sciences team in the Dell Technologies HPC & AI Innovation Lab in Austin, Texas. The lab currently has two key projects under way in this realm. One focuses on language-to-language translation and the other focuses on text-to-voice translation.

## LANGUAGE-TO-LANGUAGE TRANSLATION

In the lab's research project focused on language-to-language translation, our data scientists are working to solve key problems associated with translating from one human language to another using a neural network. This is a process that involves taking inputs from a source language and converting it to a target language.

In this process, the translation model first reads a sentence in a source language and then passes it to an encoder, which builds an intermediate representation. This intermediate representation is then passed to a decoder, which processes the intermediate representation to produce the translated sentence in the target language.

For our language-to-language translation project, we started with a stock topology created by Google, and then improved some of the underlying mathematics, so we could parallelize the workflows more efficiently. Our goal was to run our model on hundreds of compute nodes at the same time to get to a solution more quickly.

In this optimization process, which spanned several months, our team looked at how the system was using memory, how the system was performing computation and how we could improve certain parts of the system. We then ran tests to make sure the model we got in the end was just as accurate as what Google could achieve. This validation of the accuracy of the model gave us the assurance that in our efforts to speed up the computation, we didn't end up with lower-quality answers.



## About the lab

The Dell Technologies HPC & AI Innovation Lab encompasses a 13,000 square foot data center in Austin, Texas, devoted to high-performance computing and artificial intelligence (AI). It houses thousands of servers, a TOP500 cluster, and a wide range of storage and network systems.

In addition to providing access to world class infrastructure, the lab brings together HPC operational excellence and expertise. It is staffed by a dedicated group of computer scientists, engineers and subject matter experts who actively partner and collaborate with customers and other members of the HPC community. Among other activities, the team gets early access to new technologies, integrates and tunes clusters, benchmarks applications, develops best practices, and publishes white papers.

## COMPUTING RESOURCES

For this project, our research team used Dell EMC systems with 2nd-generation Intel® Xeon® Scalable processors in the [Zenith supercomputer](#), one of three HPC clusters in our HPC & AI Innovation Lab. This TOP500 system, which resulted from a partnership between Dell Technologies and Intel, serves as a benchmarking system for internal teams, as well as a showcase resource for evaluations.

In addition, the team leveraged the processing power of the [Stampede2 supercomputer](#) at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin. This Intel-based system, ranked in the [TOP500 list](#), serves as a strategic national resource that provides HPC capabilities to thousands of researchers across the United States.

## RESULTS

In this project, we demonstrated that the process of training models for language-to-language translation could be scaled to an extreme level — up to 512 nodes — without impacts on the quality of the results. This result suggests that these models can now be trained at a much faster pace and at a much large scale without breaking the current state of the art.

## TEXT-TO-VOICE TRANSLATION

Text-to-voice translation takes written words and converts them to audio. The objective is to generate a complete audio wave form synthetically — while not using the mechanized, clip recordings that we have been used to hearing on telephone systems for the last 20 years.

With these more advanced approaches, developers use training data that consists of a transcript and clips of a voice actor reading that transcript. These resources serve as the training foundation for the creation of a voice that a computer will mimic. The developers then train the neural network to produce a voice that sounds extremely similar to the actor's voice, although it's not that person speaking. It's a neural network creating that voice completely from scratch.

For the text-to-voice translation project initiated in August 2019, we used a two-part process, with two deep learning models:

- We began by taking text and converting it to a spectrogram image, and that takes one deep learning model. This spectrogram image is a frequency distribution of the letters and sounds that are expected to be produced in the resulting voice.
- We then created a second model that takes the spectrogram and generates a complete audio waveform that uses a completely synthetic voice that pretends to be the voice actor that was used in the training process. Again, we're not just stitching together voice clips of an actor talking. We're creating a synthetic voice that sounds a lot like the original.

In this ongoing project, we are now working to accelerate the process of producing the audio waveforms.

## COMPUTING RESOURCES

For this work, we are parallelizing the work across NVIDIA V100 GPUs in the [Rattler supercomputer](#), which is housed in our HPC & AI Innovation Lab. The Rattler cluster is the result of a partnership among Dell Technologies, Mellanox, Bright Computing and NVIDIA. It is designed to showcase extreme scalability, as well as provide application-specific benchmarking and characterizations.

## RESULTS

In our ongoing lab research we have demonstrated that we can create a voice that sounds like any voice we want to mimic, and that we can use parallelization to create the model for this task in a relatively short period of time. We have reduced the process of producing a realistic voice model from more than a month to less than three days, just by parallelizing the process on the Rattler supercomputer.

## TIPS FOR YOUR PROJECT

At the Dell Technologies HPC & AI Innovation Lab, we work actively to share our learnings, insights and best practices with organizations seeking to capitalize on the technologies for high performance computing and artificial intelligence. With that thought in mind, here are some of our thoughts on how your organization can get on the path to a successful NLP project.

### DON'T TRY TO REINVENT THE WHEEL.

Build on the work that others have done. For example, in our research we work with open source data that your organization can access should you want to try to replicate our results in proofs of concept and other projects.

### DON'T THINK YOU WILL GET THE RIGHT ANSWER THE FIRST TIME.

The training process for a deep learning application is highly iterative. You go down one path and see what sort of results you get. And then you go down another path, and another path after that. In our research in the HPC & AI Innovation Lab, we spend months training and tweaking our models.

### BE WILLING TO FAIL.

The development of an NLP application is not a one-and-done undertaking. You need to be willing to fail in the short term in order to achieve success in the long run. Be ready to abandon unproductive approaches, to rethink things that you thought you knew for sure, and to go back to the drawing board to map out a new path forward.

### START WITH THE EASY STUFF.

Don't try to achieve your big vision right out of the gate. Start small. Try to figure out some easy things that you can tease out of your dataset to prove what's possible with your machine learning and deep learning tools. Get some initial wins, and then build on them.

### KEEP YOUR EYES ON THE PRIZE.

The development on an NLP application is a complex undertaking from beginning to end. You're trying to create a mathematical model that mimics the human brain. This isn't going to happen overnight. The key is to recognize what's possible, and always work toward the big goal — an application you can put to work to drive your business forward.

### KEY TAKEAWAYS

Natural language processing is a potentially powerful tool for enterprises and other organizations that want to streamline their interactions with customers, employees, partners and others. To help organizations capitalize on this opportunity, researchers in the Dell Technologies HPC & AI Innovation Lab are working actively to advance the technologies and methodologies for the development of language-to-language translation and text-to-voice translation applications.

### TO LEARN MORE

- To learn more about the resources available through the Dell Technologies HPC & AI Innovation Lab, visit [delltechnologies.com/innovationlab](https://delltechnologies.com/innovationlab).
- For a broader look at NLP systems, see the article "[Natural Language Processing Could Be Key to Your Company's Digital Transformation](#)" by Dell Technologies data scientist Lucas Wilson, Ph.D.
- For an inside look at a recent neural machine translation project the Dell Technologies HPC & AI Innovation Lab was involved with, read the white paper "[Densifying Assumed-sparse Tensors: Improving Memory Efficiency and MPI Collective Performance during Tensor Accumulation for Parallelized Training of Neural Machine Translation Models.](#)"
- To explore new HPC solutions for powering AI-driven applications, visit [delltechnologies.com/ai](https://delltechnologies.com/ai).

To learn more, visit [hpcatdell.com](https://hpcatdell.com).