

# An AI assistant for material discovery

14 February 2022, by Aaron Dubrow



For a given materials performance metric, several JARVIS (Joint Automated Repository for Various Integrated Simulations) components can work together to design optimized or completely new materials. Credit: Kamal Choudhary, NIST

When Tony Stark needs to travel to space in the original Iron Man movie, he asks his artificial intelligent (AI) assistant J.A.R.V.I.S. to make a suit that can survive harsh conditions.

As AI specialist Kamal Choudhary explains: "The way I see it, what J.A.R.V.I.S. did is, it had a database of materials, scanned the database, found a suitable material, tested it, then synthesized an alloy that could survive space conditions.

"That's what we want our system to do, and that's why we called it JARVIS."

Choudhary, a researcher at the National Institute of Standards and Technology (NIST), is the founder and developer of JARVIS (Joint Automated Repository for Various Integrated Simulations)—an open dataset designed to automate materials discovery and optimization.

Writing in *npj Computational Materials* in December 2021, Choudhary and Brian DeCost (NIST) described the latest enhancements to

JARVIS that apply AI to speed discovery.

Combining graph [neural networks](#) with chemical and structural knowledge about materials, their Atomistic Line Graph Neural Network (ALIGNN) outperforms previously reported models on atomistic prediction tasks with very high accuracy and better or comparable model training speed.

"ALIGNN can predict characteristics in seconds instead of months," Choudhary said.

Beyond the inspiration from Iron Man, there was the Materials Genome Initiative. Originated in 2011 under President Obama, the initiative is a multi-federal agency effort to discover, manufacture, and deploy [advanced materials](#) twice as fast and at a fraction of the cost of traditional methods.

NIST's original contribution to the initiative was the creation of a database of materials and their characteristics, obtained rigorously, using standardized, cutting-edge computing methods.

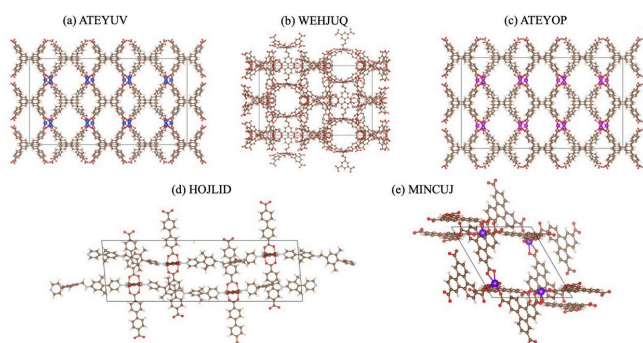
Several such databases have been established, but "what's particular about the JARVIS database is that it contains modules for various kinds of computational approaches," according to David Vanderbilt, professor of physics at Rutgers University, member of the National Academy of Sciences, and a contributor to the project. "There are many different theoretical levels on which you can approach the field. JARVIS is unusual in that it spans more levels than other databases."

The original data for JARVIS was drawn from density function theory (or DFT) calculations. "DFT is the standard way that most people compute properties of a material at an atomistic level," Vanderbilt explained. "They're first-principal calculations, where there's no experimental input and the results are derived from theory from the ground up according to the laws of quantum mechanics."

This paradigm has been incredibly effective, "however if you look at the periodic table, there are

billions of possible combinations of elements—more than we can ever generate data for," said Choudhary. "This is where machine learning comes in."

If quantum mechanical calculations can act as a screening tool for physical experiments, Choudhary reasoned, machine learning can act as a [screening tool](#) for expensive calculations.



The Atomistic Line Graph Neural Network (ALIGNN) predicted examples of the high-CO<sub>2</sub> adsorbing Metal Organic Frameworks. Credit: Kamal Choudhary, Taner Yildirim, Daniel Siderius, A. Gilad Kusne, Austin McDannald, Diana L. Ortiz-Montalvo

But first, such a system needs to be trained. Neural networks like ALIGNN, require massive amounts of training data to be effective. Standing behind Choudhary's cutting-edge AI model are DFT simulations of 70,000 materials and counting. This growing database was used to train the neural network, which in turn can rapidly characterize new materials or screen for materials with specific properties.

"It's the dream of the Materials Genome Initiative come to life," Choudhary said.

Writing in arXiv, Choudhary and his collaborators provided an example of how the system can speed discovery. They used ALIGNN to predict the CO<sub>2</sub> adsorption properties of Metal Organic Frameworks, a class of porous materials that can remove CO<sub>2</sub> from the atmosphere, and to

computationally rank leading candidates for experimental synthesis.

The JARVIS dataset was generated primarily on supercomputers at NIST, which have been working on this effort for nearly five years. More recently, Choudhary gained access to the Frontera and Stampede2 supercomputers at the Texas Advanced Computing Center (TACC), which have also contributed to the dataset.

"The machine learning field has been around since the 1980s, but the main problem was well-curated datasets," Choudhary said. "We're now approaching 100,000 materials in our database and that was only possible because of Frontera and NIST. That's what helped us bridge that gap."

With a large number of training samples available, and knowledge from chemistry and physics hard-coded into the neural network, Choudhary was able to greatly improve the accuracy of his machine learning model. "The more domain knowledge you can use the better. I think physics and AI should not be competitors to each other; they should be friends and collaborators."

The ALIGNN tool, like those for DFT calculations and other [machine learning](#) methods, are incorporated into JARVIS and made available to researchers worldwide. Choudhary estimates that 8,000 chemists and biologists make use of the repository each year. Recently, it has enabled scientists at Argonne National Laboratory to study topological magnetic materials, and helped Northwestern University researchers study transfer learning for materials.

Choudhary is also collaborating with David Vanderbilt to develop "beyond-DFT" methods, apply them to quantum materials, and integrate those methods and datasets into JARVIS.

"DFT has some significant approximations in it," Vanderbilt said. "Because electrons are treated as independent, you miss some of the very special and interesting behavior in quantum materials, which lead to effects that are beyond the normal expectation of ordinary theory."

These include, but are not limited to, unconventional superconductivity, the quantum hall effect, and topological magnetic structure. "For these classes of material, ordinary DFT doesn't work well enough," he continued. "Our database adopts three or four higher level beyond-DFT approaches to give the community a sense of how the answers may differ based on the underlying approach."

By establishing a database of possible materials and developing tools to automate screening, Choudhary hopes to speed up the pipeline of discovery, bringing Iron Man-like capabilities closer to reality.

"Imagine the day when a model that can predict a new material, a new medicine—and say, 'out of one million molecules, try this one first.'" Choudhary said. "That is the golden age of materials science."

**More information:** Kamal Choudhary et al, Atomistic Line Graph Neural Network for improved materials property predictions, *npj Computational Materials* (2021). [DOI: 10.1038/s41524-021-00650-1](https://doi.org/10.1038/s41524-021-00650-1)

Provided by Texas Advanced Computing Center  
APA citation: An AI assistant for material discovery (2022, February 14) retrieved 9 August 2022 from <https://techxplore.com/news/2022-02-ai-material-discovery.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*