



# ServiceNow, Hugging Face, and NVIDIA Release New Open-Access LLMs to Help Developers Tap Generative AI for Building Enterprise Applications

## StarCoder2 — Created With the BigCode Community and Trained on 600+ Programming Languages — Advances Code Generation, Transparency, Governance, and Innovation

ServiceNow (NYSE: NOW), Hugging Face, and NVIDIA today announced the release of StarCoder2, a family of open-access large language models for code generation that sets new standards for performance, transparency, and cost-effectiveness.

StarCoder2 was developed in partnership with the BigCode Community, managed by [ServiceNow](#), the leading digital workflow company making the world work better for everyone, and [Hugging Face](#), the most-used open-source platform, where the machine learning community collaborates on models, datasets, and applications.

Trained on 619 programming languages, StarCoder2 can be further trained and embedded in enterprise applications to perform specialized tasks such as application source code generation, workflow generation, text summarization, and more. Developers can use its code completion, advanced code summarization, code snippets retrieval, and other capabilities to accelerate innovation and improve productivity.

StarCoder2 offers three model sizes: a 3-billion-parameter model trained by ServiceNow; a 7-billion-parameter model trained by Hugging Face; and a 15-billion-parameter model built by NVIDIA with [NVIDIA NeMo](#) and trained on NVIDIA accelerated infrastructure. The smaller variants provide powerful performance while saving on compute costs, as fewer parameters require less computing during inference. In fact, the new 3-billion-parameter model matches the performance of the original StarCoder 15-billion-parameter model.

“StarCoder2 stands as a testament to the combined power of open scientific collaboration and responsible AI practices with an ethical data supply chain,” emphasized Harm de Vries, lead of ServiceNow’s StarCoder2 development team and co-lead of BigCode. “The state-of-the-art open-access model improves on prior generative AI performance to increase developer productivity and provides developers equal access to the benefits of code generation AI, which in turn enables organizations of any size to more easily meet their full business potential.”

“The joint efforts led by Hugging Face, ServiceNow, and NVIDIA enable the release of powerful base models that empower the community to build a wide range of applications more efficiently with full data and training transparency,” said Leandro von Werra, machine learning engineer at Hugging Face and co-lead of BigCode. “StarCoder2 is a testament to the potential of open source and open science as we work toward democratizing responsible AI.”

“Since every software ecosystem has a proprietary programming language, code LLMs can drive breakthroughs in efficiency and innovation in every industry,” said Jonathan Cohen, vice president of applied research at NVIDIA. “NVIDIA’s collaboration with ServiceNow and Hugging Face introduces secure, responsibly developed models and supports broader access to accountable generative AI that we believe will benefit the global community.”

### StarCoder2 Models Supercharge Custom Application Development

StarCoder2 models share a state-of-the-art architecture and carefully curated data sources from BigCode that prioritize transparency and [open governance](#) to enable responsible innovation at scale.

StarCoder2 advances the potential of future AI-driven coding applications, including text-to-code and text-to-workflow capabilities. With broader, deeper programming training, it provides repository context, enabling accurate, context-aware predictions. These advancements serve seasoned software engineers and citizen developers alike, accelerating business value and digital transformation.

The foundation of StarCoder2 is a new code dataset called [Stack v2](#), which is more than 7x larger than Stack v1. In addition to the advanced dataset, new training techniques help the model understand low-resource programming languages (such as COBOL), mathematics, and program source code discussions.

### Fine-Tuning Advances Capabilities With Business-Specific Data

Users can fine-tune the open-access StarCoder2 models with industry- or organization-specific data using open-source tools such as NVIDIA NeMo or [Hugging Face TRL](#). They can create advanced chatbots to handle more complex summarization or classification tasks, develop personalized coding assistants that can quickly and easily complete programming tasks,

retrieve relevant code snippets, and enable text-to-workflow capabilities.

Organizations have already begun to fine-tune the foundational StarCoder model to create specialized task-specific capabilities for their businesses.

ServiceNow's text-to-code Now LLM was purpose-built on a specialized version of the 15-billion-parameter StarCoder LLM, fine-tuned and trained for its workflow patterns, use cases, and processes. Hugging Face has also used the model to create its StarChat assistant.

### **BigCode Fosters Open Scientific Collaboration in AI**

BigCode represents an open scientific collaboration led by Hugging Face and ServiceNow, dedicated to the responsible development of LLMs for code.

The BigCode community actively participated in the technical aspects of the StarCoder2 project through working groups and task forces, leveraging ServiceNow's Fast LLM framework to train the 3-billion-parameter model, Hugging Face's nanotron framework for the 7-billion-parameter model and the NVIDIA NeMo cloud-native framework and [NVIDIA TensorRT-LLM](#) software to train and optimize the 15-billion-parameter model.

Fostering responsible innovation is at the core of BigCode's purpose, demonstrated through its open governance, transparent supply chain, use of open-source software, and the ability for developers to opt data out of training. StarCoder2 was built using responsibly sourced data under license from the digital commons of [Software Heritage](#), hosted by [Inria](#).

"StarCoder2 is the first code generation AI model developed using the Software Heritage source code archive and built to align with our policies for responsible development of models for code," stated Roberto Di Cosmo, director at Software Heritage. "The collaboration of ServiceNow, Hugging Face, and NVIDIA exemplifies a shared commitment to ethical AI development, advancing technology for the greater good."

StarCoder2, like its predecessor, will be made available under the BigCode Open RAIL-M license, allowing royalty-free access and use. Further fostering transparency and collaboration, the model's supporting code will continue to reside on the BigCode project's GitHub page.

All StarCoder2 models will also be available for download from Hugging Face, and the StarCoder2 15-billion-parameter model is available on [NVIDIA AI Foundation models](#) for developers to experiment with directly from their browser, or through an API endpoint.

For more information on StarCoder2, visit <https://huggingface.co/bigcode>.

### **About NVIDIA**

Since its founding in 1993, NVIDIA (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing infrastructure company with data-center-scale offerings that are reshaping industry. More information at <https://nvidianews.nvidia.com/>.

### **About ServiceNow**

ServiceNow (NYSE: NOW) makes the world work better for everyone. Our cloud based platform and solutions help digitize and unify organizations so that they can find smarter, faster, better ways to make work flow. So employees and customers can be more connected, more innovative, and more agile. And we can all create the future we imagine. The world works with ServiceNowTM. For more information, visit: [www.servicenow.com](http://www.servicenow.com).

### **About Hugging Face**

Hugging Face is the leading open source and community driven AI platform, providing tools that enable users to build, explore, deploy and train machine learning models and datasets. For more information, visit: [www.huggingface.co](http://www.huggingface.co).

### **NVIDIA Forward-Looking Statements**

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, and performance of NVIDIA's products, services, and technologies, including NVIDIA NeMo and NVIDIA TensorRT-LLM; the benefits and impact of NVIDIA's collaboration with ServiceNow and Hugging Face, and the features and availability of its services and offerings; and the ability of code LLMs to drive breakthroughs in efficiency and innovation in every industry 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.

© 2024 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, NVIDIA NeMo and NVIDIA TensorRT-LLM 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. Features, pricing, availability and specifications are subject to change without notice.

Jacqueline Velasco  
[press@servicenow.com](mailto:press@servicenow.com)  
Anna Kiachian  
Senior PR Manager  
NVIDIA Corporation  
+1-650-224-9820  
[akiachian@nvidia.com](mailto:akiachian@nvidia.com)  
Brigitte Tousignant  
[brigitte@huggingface.co](mailto:brigitte@huggingface.co)