# Right on Track: NVIDIA Open-Source Software Helps Developers Add Guardrails to AI Chatbots

NeMo Guardrails helps enterprises keep applications built on large language models aligned with their safety and security requirements.

Author: Jonathan Cohen

Newly released open-source software can help developers guide generative AI applications to create impressive text responses that stay on track.

NeMo Guardrails will help ensure smart applications powered by large language models (LLMs) are accurate, appropriate, on topic and secure. The software includes all the code, examples and documentation businesses need to add safety to AI apps that generate text.

Today's release comes as many industries are adopting LLMs, the powerful engines behind these AI apps. They're answering customers' questions, summarizing lengthy documents, even writing software and accelerating drug design.

NeMo Guardrails is designed to help users keep this new class of AI-powered applications safe.

Safety in generative AI is an industry-wide concern. NVIDIA designed NeMo Guardrails to work with all LLMs, such as OpenAI's ChatGPT.

The software lets developers align LLM-powered apps so they're safe and stay within the domains of a company's expertise.

NeMo Guardrails enables developers to set up three kinds of boundaries:

Topical guardrails prevent apps from veering off into undesired areas. For example, they keep customer service assistants from answering questions about the weather.

Safety guardrails ensure apps respond with accurate, appropriate information. They can filter out unwanted language and enforce that references are made only to credible sources.

Security guardrails restrict apps to making connections only to external third-party applications known to be safe.

Virtually every software developer can use NeMo Guardrails — no need to be a machine learning expert or data scientist. They can create new rules quickly with a few lines of code.

Since NeMo Guardrails is open source, it can work with all the tools that enterprise app developers use.

For example, it can run on top of LangChain , an open-source toolkit that developers are rapidly adopting to plug third-party applications into the power of LLMs.

"Users can easily add NeMo Guardrails to LangChain workflows to quickly put safe boundaries around their AI-powered apps," said Harrison Chase, who created the LangChain toolkit and a startup that bears its name.

In addition, NeMo Guardrails is designed to be able to work with a broad range of LLM-enabled applications, such as Zapier . Zapier is an automation platform used by over 2 million businesses, and it's seen first-hand how users are integrating AI into their work.

"Safety, security, and trust are the cornerstones of responsible AI development, and we're excited about NVIDIA's proactive approach to embed these guardrails into AI systems," said Reid Robinson, lead product manager of AI at Zapier.

“We look forward to the good that will come from making AI a dependable and trusted part of the future.”

NVIDIA is incorporating NeMo Guardrails into the NVIDIA NeMo framework, which includes everything users need to train and tune language models using a company’s proprietary data.

Much of the NeMo framework is already available as open source code on GitHub.  Enterprises also can get it as a complete and supported package, part of the NVIDIA AI Enterprise software platform.

NeMo is also available as a service . It’s part of NVIDIA AI Foundations , a family of cloud services for businesses that want to create and run custom generative AI models based on their own datasets and domain knowledge.

Using NeMo, South Korea’s leading mobile operator built an intelligent assistant that’s had 8 million conversations with its customers. A research team in Sweden employed NeMo to create LLMs that can automate text functions for the country’s hospitals, government and business offices.

Building good guardrails for generative AI is a hard problem that will require lots of ongoing research as AI evolves.

NVIDIA made NeMo Guardrails — the product of several years’ research — open source to contribute to the developer community’s tremendous energy and work on AI safety.

Together, our efforts on guardrails will help companies keep their smart services aligned with safety, privacy and security requirements so these engines of innovation stay on track.

For more details on NeMo Guardrails and to get started, see our technical blog , and watch the video below.