



TELUS

AI STUDIO

Developer-First AI Cloud at Self-Service Speed, Powering AI Innovation for Canada



WHY TELUS BUILT A DEVELOPER-FIRST AI STUDIO

Across Canada, developers and enterprises were ready to build with AI, yet most lacked the compute access, governance, and agility needed to bring ideas to market. TELUS recognized a structural gap: Canada needed a sovereign AI platform that put developers at the center.

TELUS saw this as both a national opportunity and a strategic imperative. The company's network, data center footprint, and trust position uniquely equipped it to create a platform that could meet two non-negotiable demands simultaneously: the speed developers expect and the sovereignty regulators require.

Four insights guided TELUS' decision to build AI Studio:

^ **DEVELOPERS NEEDED VELOCITY, NOT HARDWARE.**

Canada's AI ecosystem had access to GPUs, but not to the experience layer that turns compute into rapid innovation. AI builders spent weeks configuring clusters, networking, and dependencies before a single experiment could start. TELUS sought to collapse that timeline into minutes through full self-service provisioning and ready-to-use environments.

^ **THE SELF-SERVICE GAP WAS STIFLING INNOVATION.**

Hyperscalers offered scale but not autonomy; internal enterprise platforms offered control but not speed. Developers were caught in the middle. TELUS recognized the need for a platform that combined the agility of cloud with the rigor of enterprise governance, enabling every developer to build, deploy, and iterate independently.

^ **SOVEREIGNTY AND AGILITY COULD NO LONGER BE TRADE-OFFS.**

Regulated sectors from healthcare to public services required strict data residency, yet innovation demanded fast iteration cycles. TELUS' strategy was to make sovereignty invisible to the developer: all data, models, and workloads remain within Canadian borders, governed by TELUS' infrastructure, without adding friction to the build-deploy loop.

^ **COST TRANSPARENCY WAS ESSENTIAL FOR SCALE.**

As GPU resources became strategic capital assets, business leaders demanded clarity. TELUS identified the need for real-time metering and chargeback, enabling organizations to scale AI workloads confidently while maintaining accountability for consumption and cost.

TELUS set out to remove these frictions with a sovereign-by-design platform, backed by NVIDIA's accelerated computing and Rafay's orchestration layer, that transforms idle GPU infrastructure into a self-service, production-ready experience for AI builders.

INTRODUCING THE TELUS AI STUDIO

The result is Canada's first sovereign, developer-ready AI Studio. A production-grade environment designed to accelerate how enterprises, startups, and research institutions build, fine-tune, and deploy AI applications. Through the TELUS AI Studio, developers gain on-demand access to GPU-powered compute, curated model catalogs, and reusable AI blueprints, all hosted entirely within Canadian borders to ensure complete data sovereignty and compliance.

KEY HIGHLIGHTS

SOVEREIGN BY DESIGN

TELUS' AI Cloud is architected to ensure that all data, in storage or in transit, remains entirely within Canadian borders. Leveraging its owned and operated fiber-optic network and national AI data center footprint, TELUS enforces multiple layers of security and compliance so that sensitive workloads across government, healthcare, and enterprise sectors are governed exclusively under Canadian law, without exposure to foreign jurisdictions.

A FIRST-MOVER IN NORTH AMERICA

TELUS is the first North American telecommunications provider in the NVIDIA Cloud Partner Program delivering a fully validated AI stack that significantly reduces deployment time and cost for enterprises and startups building production-grade AI solutions.

DEVELOPER-READY AI STUDIO

AI Studio provides a self-service workspace where developers and data scientists can build, fine-tune, and deploy applications using NVIDIA NIM microservices, NVIDIA NeMo toolkits, and NVIDIA AI Blueprints orchestrated through the Rafay Platform. This unified environment lowers barriers to innovation and accelerates time-to-market for AI applications.



TRUE SELF-SERVICE CONSUMPTION

Through rich APIs and a modern web portal, developers can provision and manage GPU-powered compute, including Kubernetes clusters and virtual machines, on demand. Policy-based governance, quota enforcement, and chargeback ensure that agility never comes at the expense of control.

PRODUCTION-READY AI WORKFLOWS

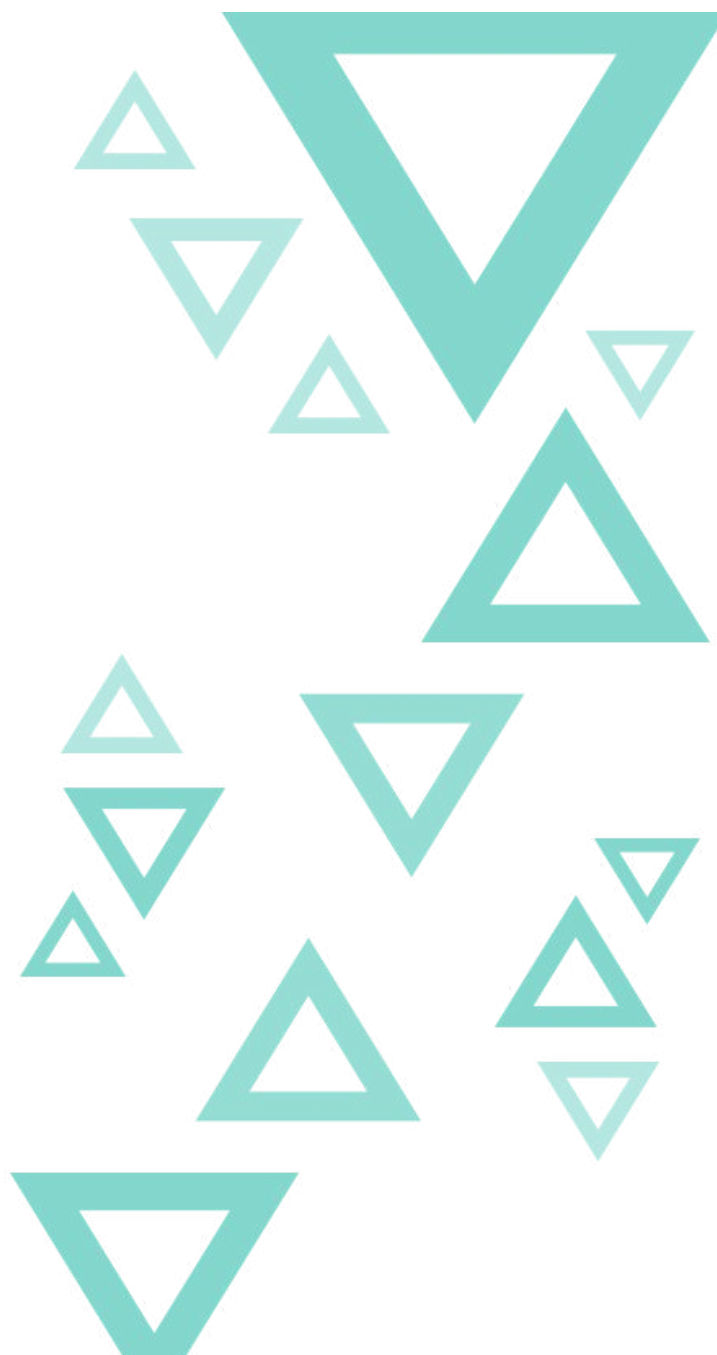
NVIDIA NIM microservices, NVIDIA AI Blueprints, and NVIDIA NeMo capabilities together create a repeatable, efficient path from prototype to production in the TELUS AI Factory. With the addition of NemotTron, NVIDIA's family of datasets and open models for advanced reasoning, math, coding, visual understanding, agentic tasks, AI safety, and information retrieval retrieval-augmented generation (RAG), developers gain immediate access to enterprise-ready, high-performance models.

MARKETPLACE FOR AI INNOVATION

TELUS curates a governed catalog of approved models, agents, and applications that can be published and consumed as SKUs. This marketplace experience provides transparent usage visibility and clear economics, transforming sovereign infrastructure into a scalable revenue engine, powered by Rafay's orchestration.

BUILT TO SCALE, READY FOR MULTI-TENANCY

Underpinned by Rafay's multi-tenant platform, TELUS operates a fully governed environment with fine-grained RBAC, usage metering, and cost transparency. Enterprises and teams can scale workloads securely and predictably, confident that sovereignty and performance are preserved at every level.



HOW IT WORKS

^ PROVISION: LAUNCH IN AI STUDIO

Developers log into TELUS' AI Studio Portal to self-provision GPU-powered compute in a variety of consumption models, from bare-metal nodes to Kubernetes or virtual clusters, all orchestrated through Rafay's multi-tenant controller. Projects inherit enterprise guardrails such as RBAC, quotas, and cost policies. Within this secure, governed environment, users gain instant access to NVIDIA NIM microservices, NVIDIA NeMo, and NVIDIA AI Blueprints for a consistent, governed start to every workflow.

^ BUILD: ACCELERATE WITH BLUEPRINTS

Developers can begin with NVIDIA AI Blueprints that codify agentic workflows (reason-plan-act) or select pre-validated base models from the NVIDIA NIM catalog. Rafay automates the environment setup, networking, and dependency management so developers can focus purely on model logic, not infrastructure.

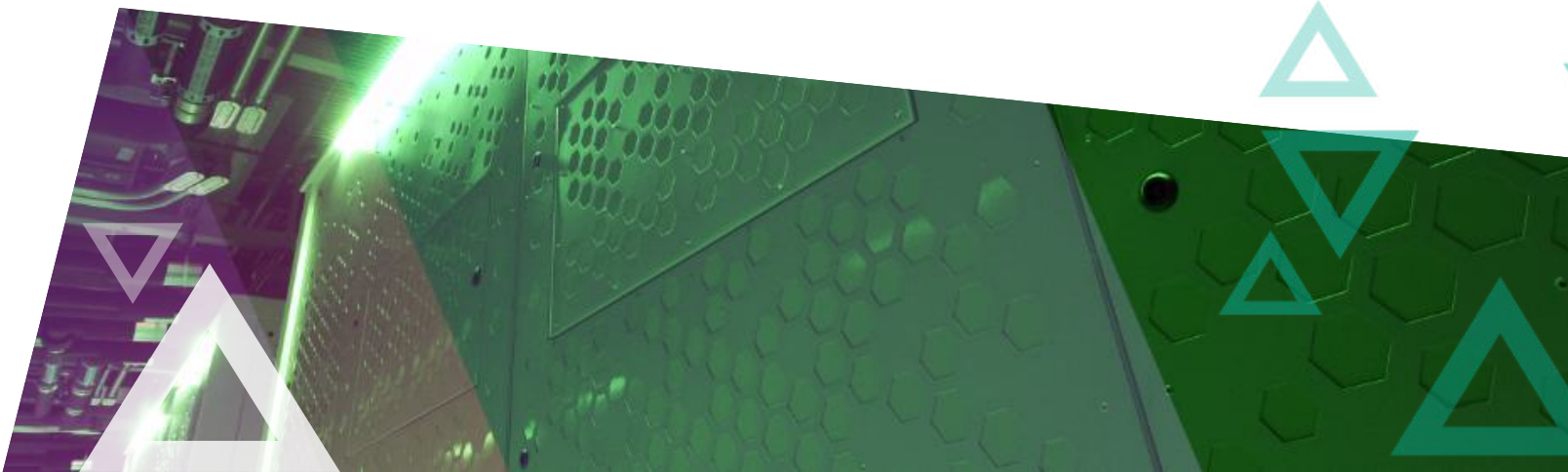
^ CUSTOMIZE & ASSURE: REFINE WITH NVIDIA NEMO

Using NVIDIA NeMo, developers can fine-tune models with domain data, apply guardrails, and run evaluations to validate accuracy and safety. Once validated, models are deployed as NIM microservices for production-grade inference behind stable APIs, ensuring repeatability and compliance across tenants.

^ PUBLISH & OPERATE: MONETIZE AND SCALE

Production-ready models, agents, and applications are published into TELUS' marketplace-style catalog as SKUs. Through chargeback and usage metering, consumption is tracked in real time for transparency and revenue attribution. Customers and partners can subscribe on demand, turning TELUS' AI Studio into a continuously expanding ecosystem of AI services.

Together, these steps form a complete value chain, from provisioning to production to monetization, transforming AI Studio into Canada's launchpad for enterprise and developer innovation.



OUTCOMES:

WHAT IT MEANS FOR DEVELOPERS AND CANADIAN AI BUILDERS

TELUS' AI Studio is redefining how Canadian developers and organizations bring AI ideas to life, delivering sovereignty, speed, and simplicity in one unified platform.

^ INSTANT ACCESS, IMMEDIATE PRODUCTIVITY

Developers can create and deploy governed AI environments within minutes. Through self-service portals and APIs, teams gain immediate access to GPU compute and pre-built model blueprints without waiting for manual approvals.

^ FASTER TIME-TO-PRODUCTION

Projects that once required extensive setup can now progress from prototype to production-ready environments in a fraction of the time. TELUS' validated blueprints and automated orchestration remove operational friction, allowing developers to focus entirely on innovation.

^ SOVEREIGNTY AND TRUST BY DEFAULT

Every workload runs entirely within Canada's borders under TELUS' data governance framework. Developers and enterprises alike can innovate confidently, knowing their data and intellectual property are protected by Canadian privacy and AI regulations.

^ TRANSPARENT ECONOMICS

Built-in cost metering and quota controls ensure predictable usage and spending for every project. Enterprises can scale GPU consumption while maintaining clear accountability and chargeback across teams.

^ COLLABORATION ACROSS CANADA'S AI ECOSYSTEM

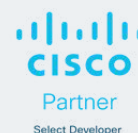
By opening AI Studio to partners, startups, and research institutions, TELUS enables secure, multi-tenant collaboration. This unified environment accelerates shared innovation while maintaining strict workload isolation and policy compliance.





Discover how your teams can provision, build, customize, and publish AI applications with the speed of self-service and the confidence of enterprise-grade infrastructure.

Built on NVIDIA-validated software, orchestrated by the Rafay Platform, and operated on TELUS' sovereign Canadian infrastructure, the AI Studio represents a new standard for how organizations innovate securely, sustainably, and at scale.



Rafay Systems, Inc. | 530 Lakeside Dr, Ste 210 Sunnyvale, CA 94085 | 669.247.2551 | rafay.co

© 2025 Rafay Systems, Inc. All rights reserved. All other trademarks and service marks are property of their respective owners.