STAC NEW YORK

LLM /GENERATIVE AI FOR SYSTEMATIC ALGO/DISCRETIONARY TRADING INCL. ALTERNATIVE DATA

BY PRABHU RAMAMOORTHY, NVIDIA FINANCIAL SERVICES & TECHNOLOGY TEAM
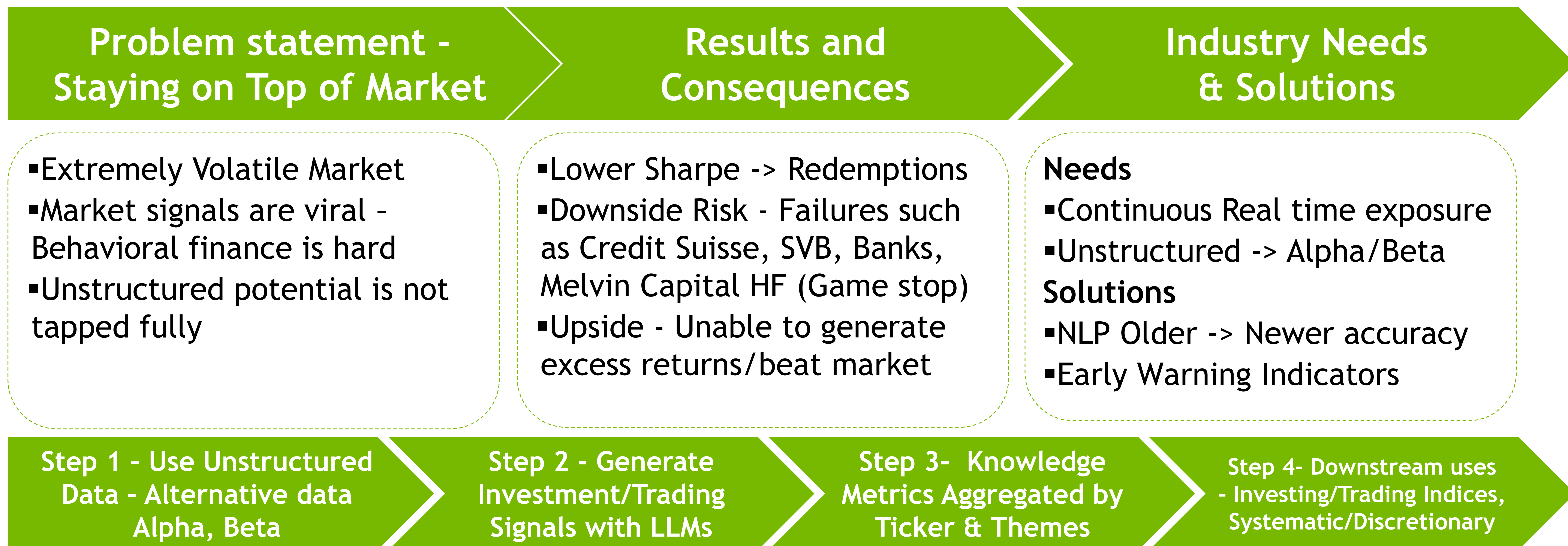CFA, FRM, CAIA

NVIDIA AI LEADER FOR LLM/GENERATIVE AI

APPLICATION FRAMEWORKS

MODULUS · CLARA MONAI · RIVA · MAXINE · NEMO · MERLIN · AVATAR · DRIVE · ISAAC · METROPOLIS · HOLOSCAN

PLATFORM

NVIDIA HPC · NVIDIA AI ENTERPRISE · NVIDIA OMNIVERSE

SYSTEM SOFTWARE

RTX · CUDA-X · PHYSX

UCF · DOCA · MAG · BASE CMD · FLEET CMD · AERIAL

ACCELERATED HARDWARE

RTX · DGX · HGX · EGX · OVX · SUPER POD · AGX

GPU · CPU · DPU · NIC · SWITCH · SOC

Full Stack - Software + Hardware

Data Center Scale

(Cloud/Onprem)

2,700 Accelerated Applications

450 SDKs, AI Models

WHY LLMs/GENERATIVE AI in Capital Markets

Used in Systematic Trading/Discretionary Trading

McKinsey reports - Advanced analytics in asset management beyond the buzz, Generative AI is here to change biz

Problem statement - Staying on Top of Market	Results and Consequences	Industry Needs & Solutions
▪Extremely Volatile Market ▪Market signals are viral – Behavioral finance is hard ▪Unstructured potential is not tapped fully	▪Lower Sharpe -> Redemptions ▪Downside Risk - Failures such as Credit Suisse, SVB, Banks, Melvin Capital HF (Game stop) ▪Upside - Unable to generate excess returns/beat market	**Needs** ▪Continuous Real time exposure ▪Unstructured -> Alpha/Beta **Solutions** ▪NLP Older -> Newer accuracy ▪Early Warning Indicators

Step 1 - Use Unstructured Data – Alternative data Alpha, Beta	Step 2 - Generate Investment/Trading Signals with LLMs	Step 3- Knowledge Metrics Aggregated by Ticker & Themes	Step 4- Downstream uses - Investing/Trading Indices, Systematic/Discretionary

NVIDIA.

Leverage Accelerated workloads in Capital Markets for you

One Stop Platform - Not only LLMs but use it with Quant Finance, ETL/ML, DL Algos

AI (Neural nets) – LLMs/Generative AI/others	Quant Finance Aka HPC	Data Processing ETL/ML (Non neural nets)
• AI Unstructured Data using NLP with LLMs, Other Systematic Trading Algos • Framework –PyTorch/TensorFlow, NVIDIA NeMo LLM, NVIDIA RIVA	• Pricing, Risk (MC Sim, Margin, FRTB, CVA, SIMM, XVA) & Back testing • Framework – CUDA C/C++, Parallel Algorithms C++, NVIDIA Accelerated Python – RAPIDS, Open ACC	• Feature Engineering, Data Prep, & Data Science (e.g., XGBOOST) • Framework – NVIDIA Accelerated Python - RAPIDS, Spark on GPU

Forrester Wave™: AI Infrastructure, Q4 2021- "NVIDIA's DNA is in every other AI infrastructure solution we evaluated. It's an understatement to say that NVIDIA GPUs are synonymous with AI infrastructure."

NVIDIA

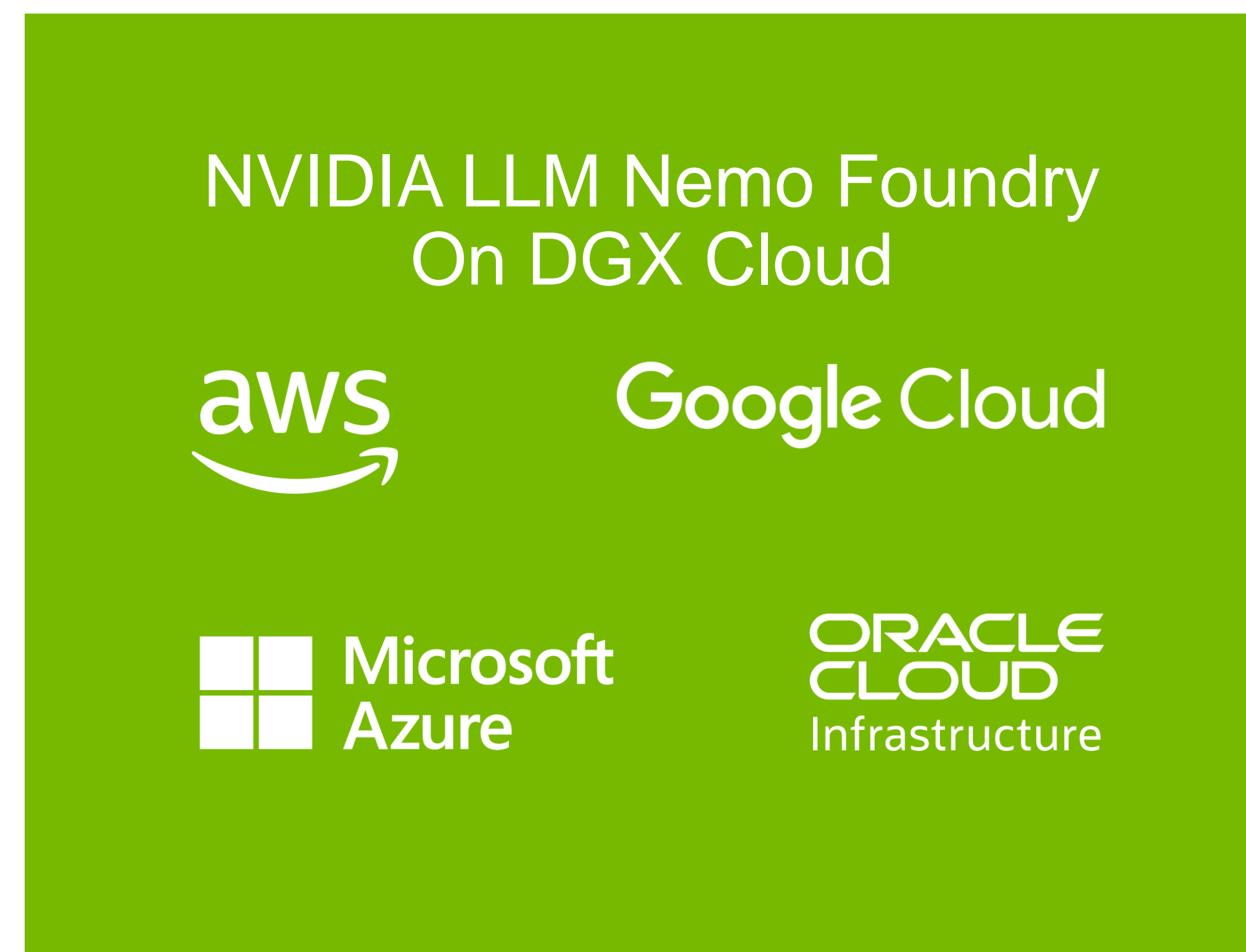# How to Get Started With LLMs/Generative AI at your firm

Generate your AI ROI

McKinsey Modelling Impact of AI - Front-runner companies could "Double their cash flow"
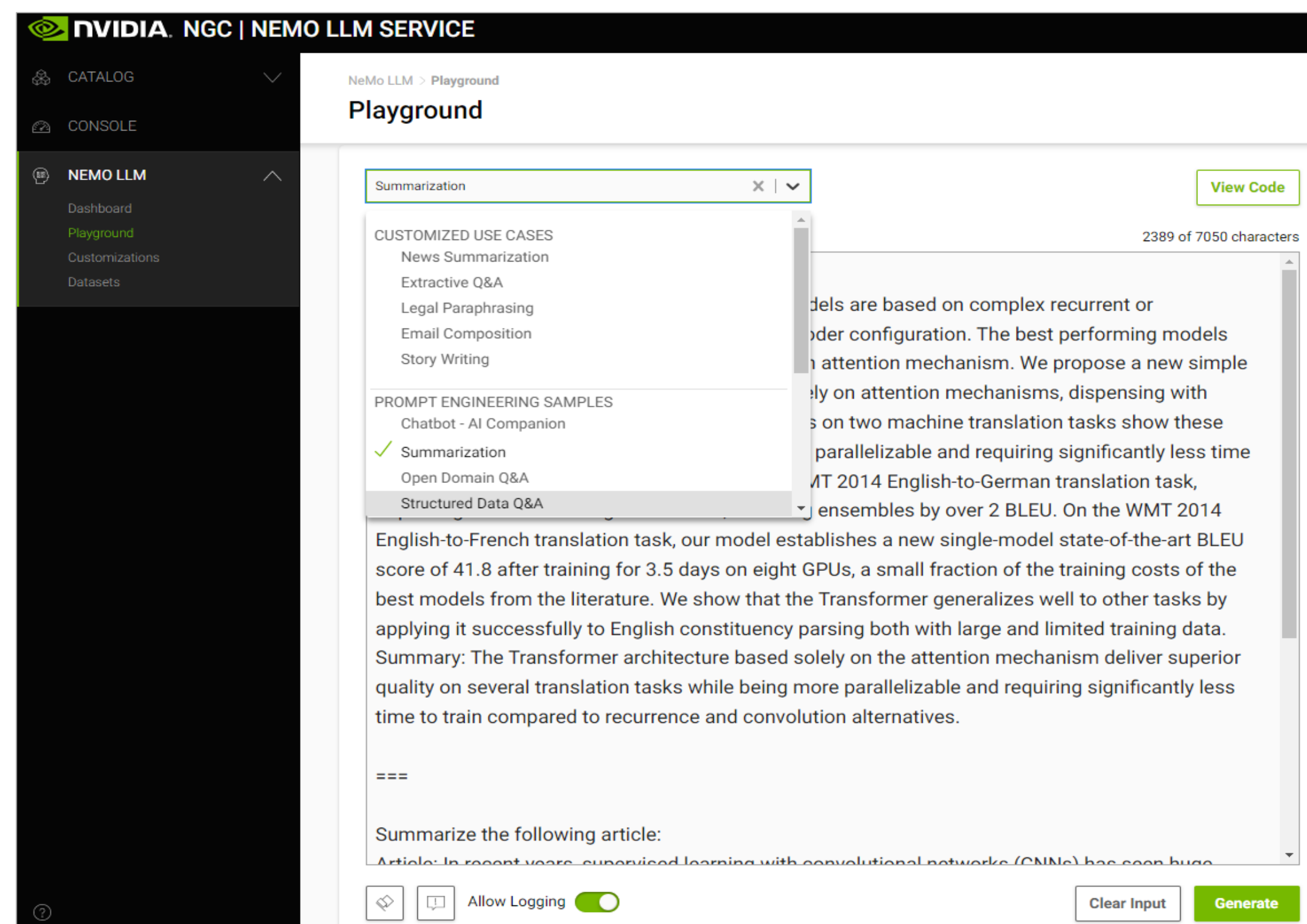
NVIDIA LLM Nemo – Prompt tuning, guard rails

or

NVIDIA LLM Nemo Foundry On DGX Cloud

aws

Google Cloud

Microsoft Azure

ORACLE CLOUD Infrastructure

Test/Deploy on Platform of choice

NLP - News, Docs   Sentiment & Q/A   Alt Data

LLM/Generative -> AI ROI

- Step 1: Workshops to prioritize, & define the problem statement, success criteria in LLM/Generative AI (2-4 hours)

- Step 2: Leverage NVIDIA LLM NeMo Foundry on cloud to prototype with your proprietary data (2-3 weeks)

- Step 3: Test & document (1 weeks)

- Step 4: Present findings & ROI to key stakeholders (1 week)

What Would an LLM/Gen AI Journey Look Like?

THANK YOU