

# AI Workshop - oneAPI AI and HPC DevSummit 2023

Gen AI: Unleash the power of GenAI and LLM using  
Intel Developer Cloud (IDC)

**Kazi Haque**



# Agenda

- oneAPI Industry Initiative
  - Why oneAPI
  - Intel® oneAPI Toolkits
- GenAI as a complete backend / production system for Prompt Generation and API Integration  
<https://llamastudio.dev/>
- AI Hands - on Workshop on
  - Intel Developer Cloud (IDC) sign up and registration
  - Text-To-Image Stable Diffusion with Stability AI and CompVis models  
Demo with notebook deployment in Intel Developer Cloud (IDC) using Intel latest GPU
  - LLM Custom Chatbot development with Huggingface opensource Mistral 7B model  
Complete application development on Intel® Max Series GPU in Intel Developer Cloud (IDC)

# oneAPI Industry Momentum

## End Users



## National Labs



## ISVs & OSVs



## OEMs & Sis



## Universities & Research Institutes



## CSPs & Frameworks

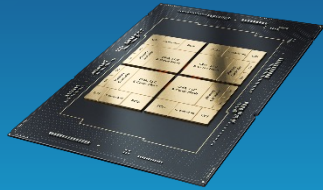


These organizations support the oneAPI initiative for a single, unified programming model for cross-architecture development. It does not indicate any agreement to purchase or use of Intel's products. \*Other names and brands may be claimed as the property of others.

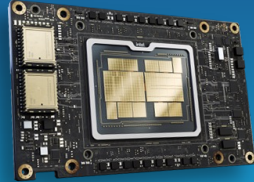
# Diverse Compute Requirements

Diverse accelerators needed to meet today's performance requirements:  
48% of developers target heterogeneous systems  
that use more than one kind of processor or core<sup>1</sup>

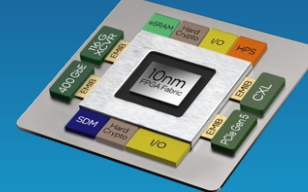
CPU



GPU



FPGA



Other Accelerators



Developer Challenges: Multiple Architectures, Vendors, and Programming Models



Open, Standards-based, Multiarchitecture Programming

1 - <https://evansdata.com/reports/viewRelease.php?reportID=40>

# oneAPI Industry Initiative

Break the Chains of Proprietary Lock-in

Freedom to Make Your Best Choice

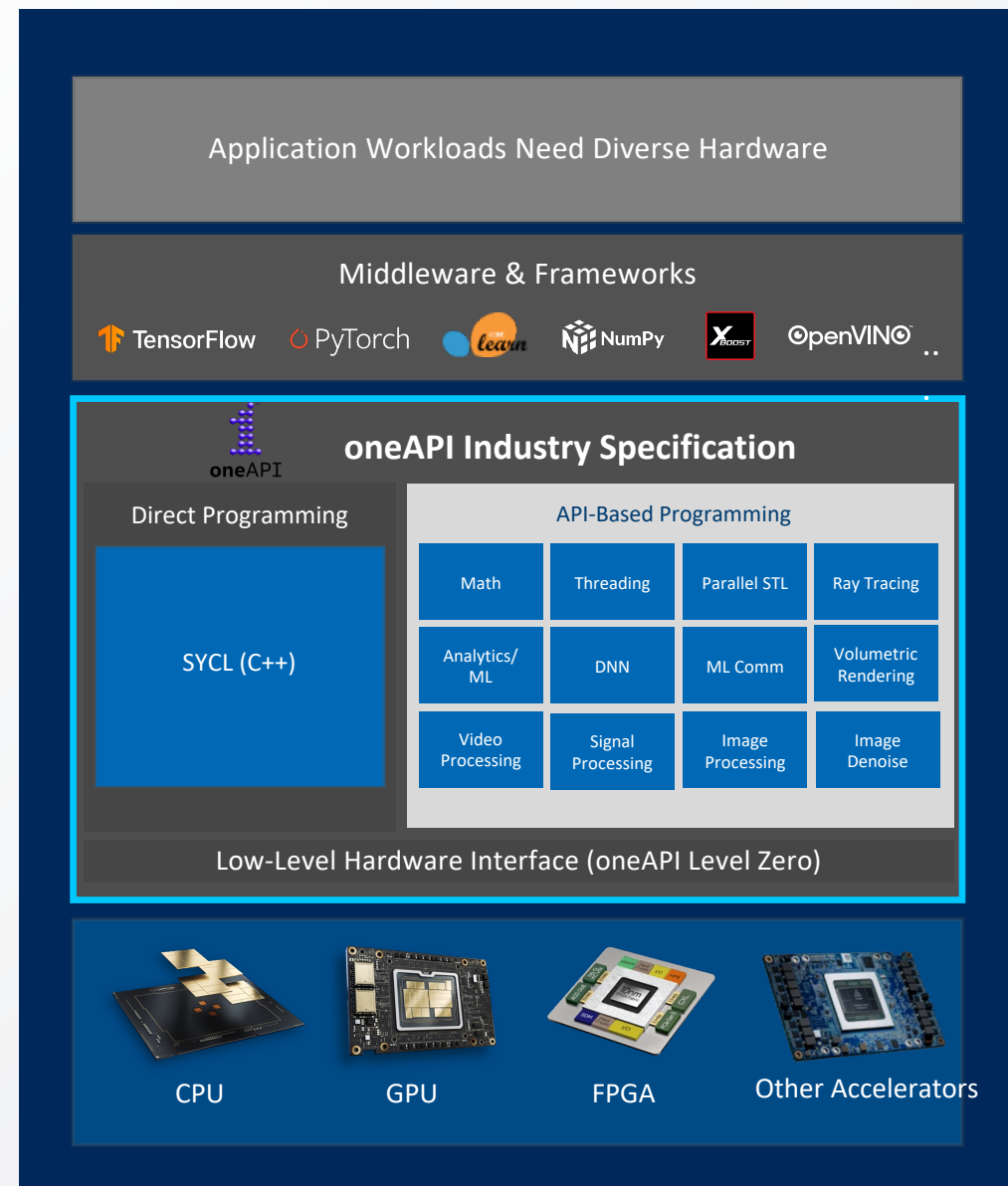
- C++ programming model for multiple architectures and vendors
- Cross-architecture code reuse for freedom from vendor lock-in

Realize all the Hardware Value

- Performance across CPU, GPUs, FPGAs, and other accelerators
- Expose and exploit cutting-edge features of the latest hardware

Develop & Deploy Software with Peace of Mind

- Open industry standards provide a safe, clear path to the future
- Interoperable with familiar languages and programming models including Fortran, Python, OpenMP, and MPI
- Powerful libraries for acceleration of domain-specific functions



The productive, smart path to freedom for accelerated computing from the economic and technical burdens of proprietary programming models

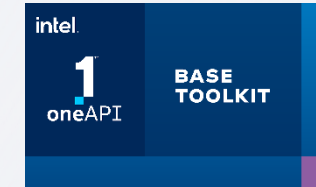


# 1 Intel® oneAPI Toolkits

A complete set of proven developer tools expanded from CPU to Accelerators

## Intel® oneAPI Base Toolkit

A core set of high-performance libraries and tools for building C++, SYCL and Python applications

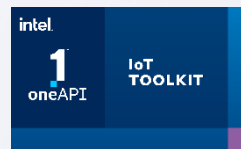


## Add-on Domain-specific Toolkits



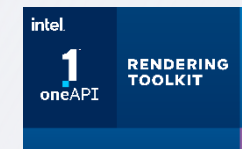
### Intel® oneAPI Tools for HPC

Deliver fast Fortran, OpenMP & MPI applications that scale



### Intel® oneAPI Tools for IoT

Build efficient, reliable solutions that run at network's edge



### Intel® oneAPI Rendering Toolkit

Create performant, high-fidelity visualization applications

## Toolkits powered by oneAPI



### Intel® AI Analytics Toolkit

Accelerate machine learning & data science pipelines end-to-end with optimized DL frameworks & high-performing Python libraries



### Intel® Distribution of OpenVINO™ Toolkit

Deploy high performance inference & applications from edge to cloud

# Intel® AI Analytics Toolkit

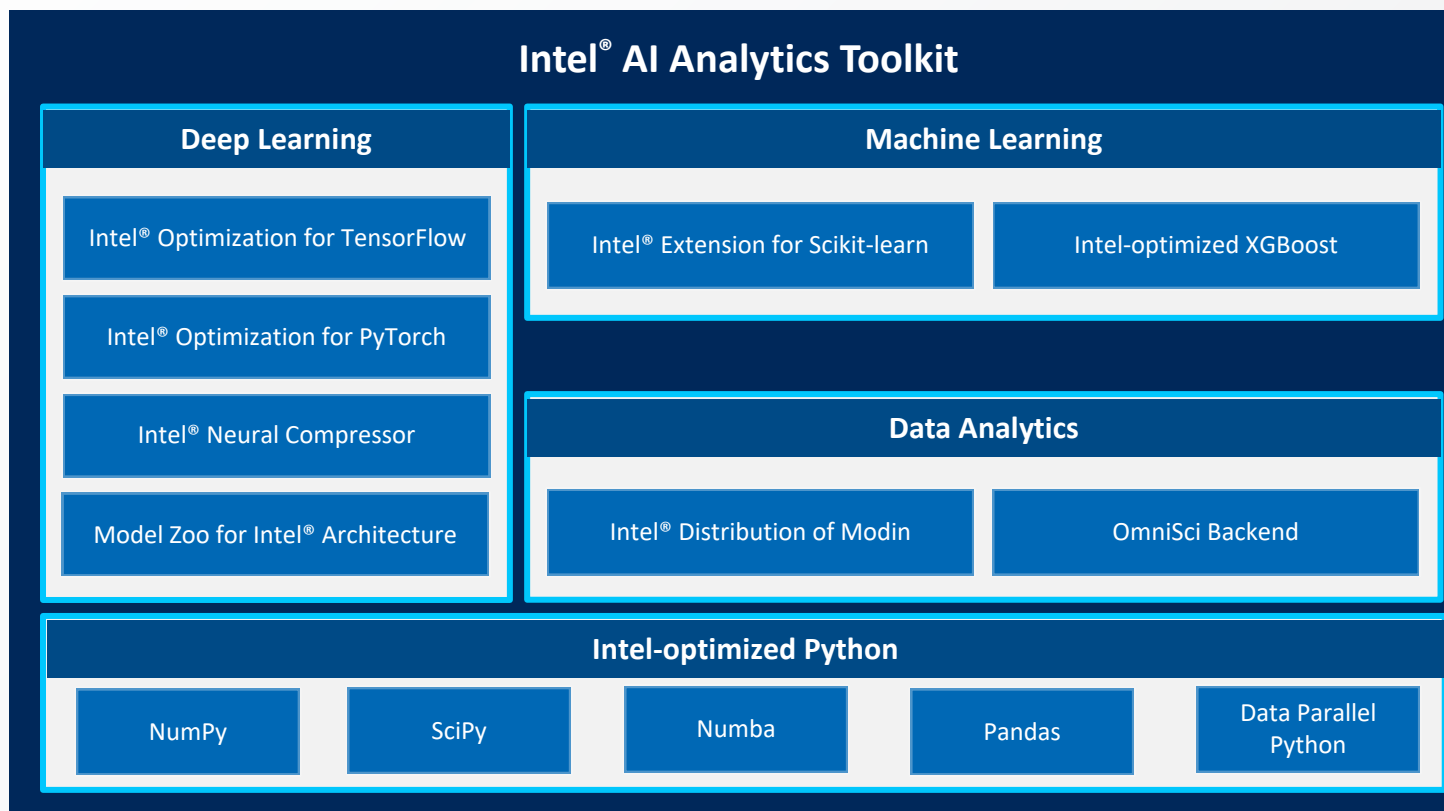
Accelerate end-to-end AI and data analytics pipelines with libraries optimized for Intel® architectures

## Who needs this product?

Data scientists, AI researchers, ML and DL developers, AI application developers

## Top Features/Benefits

- Deep learning performance for training and inference with Intel optimized DL frameworks and tools
- Drop-in acceleration for data analytics and machine learning workflows with compute-intensive Python packages



CPU



GPU

Hardware support varies by individual tool. Architecture support will be expanded over time.

Get the Toolkit [HERE](#) or via these locations

<https://software.intel.com/content/www/us/en/develop/tools/oneapi/download.html#aikit>

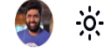
[Intel Installer](#)

[Docker](#)

[Apt, Yum](#)

[Conda](#)

[Intel® DevCloud](#)



# Gen AI as a Backend

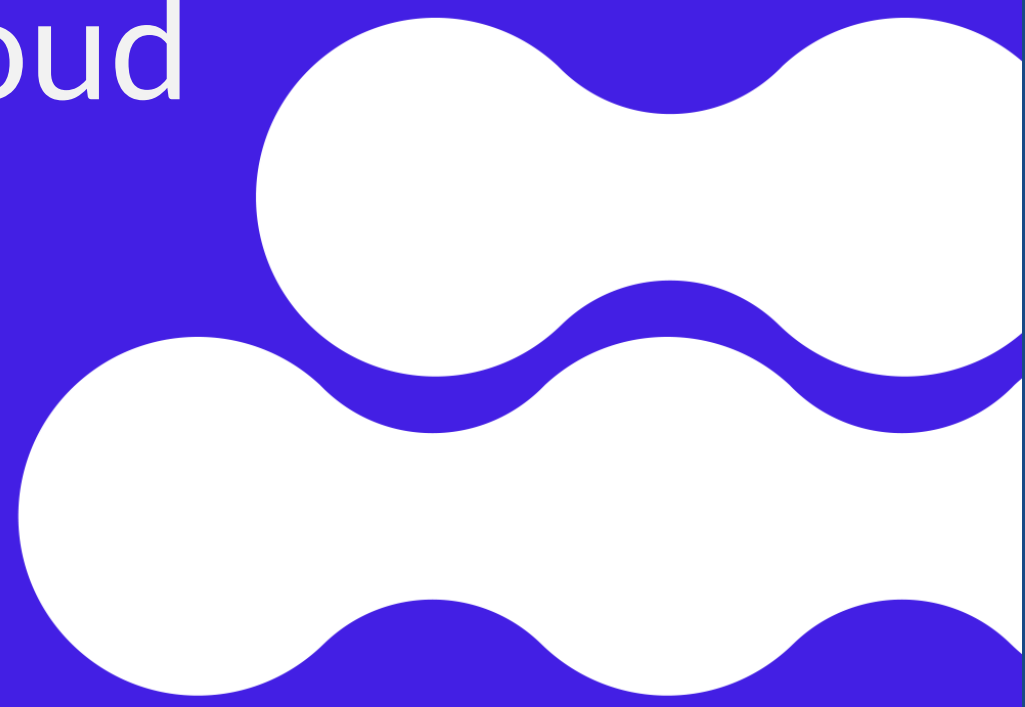
**Start →**

What will you create today?





# Intel® Developer Cloud



# Workshop Pre-requisites

- Register for accessing Intel Developer Cloud (10 mins)  
visit → [cloud.intel.com](https://cloud.intel.com)  
Sign up --> Create Account

Ensure Developer cloud Jupyter access works

Sign-in to Intel® oneAPI Developer cloud → Get Started → (scroll down)

Launch JupyterLab\*

# Intel® Developer Cloud

a service platform for developing and running workloads in Intel®-optimized deployment environments with the latest Intel® processors

Landing page : <https://developer.intel.com/GetStartedDevCloud>



# Intel® Developer Cloud

a service platform for developing and running workloads in Intel®-optimized deployment environments with the latest Intel® processors

## Available Intel Hardware

### Available Intel Hardware

Test and evaluate your application on a variety of virtual machines, bare metal systems, edge devices, and platforms designed for AI training.



# Intel® Developer Cloud



Developer Cloud

us-region-1

Help



## Console Home

### Quick Start

Hardware Catalog

Software Catalog

Training and Workshops

Cloud Credits

### Learning and Support



#### Getting started

Learn the fundamentals to get the Most out of the Intel developer cloud



#### Tutorials

Browse how to create better solutions using Intel developer cloud



#### What's new?

Learn the fundamentals to get the Most out of the Intel developer cloud

### Notifications



No notifications yet

Stay tuned for exciting updates!  
No new notifications at the moment.



# Intel® Developer Cloud



Developer Cloud

us-region-1

Help



## Training and Workshops

Launch JupyterLab



### AI

#### AI Kit XGBoost Predictive Modeling

Learn predictive modeling with decision trees using Intel® AI Analytics Toolkit

Launch

#### Heterogeneous Programming Using Data Parallel Extension for Numba® for AI and HPC

Data Parallel Extension for Numba accelerates Python® code on Intel® XPU<sub>s</sub>

Launch

#### Machine Learning Using oneAPI

Intel® AI Analytics Toolkit accelerates data science and analytics with Python®

Launch

### C++ SYCL

#### Essentials of SYCL

Learn to write performant and portable code using oneAPI and SYCL C++

Launch

#### Performance, Portability and Productivity

Learn to write performant and portable HPC code for multiple platforms with oneAPI and SYCL C++

Launch

#### Introduction to GPU Optimization

Learn GPU optimization techniques using SYCL.

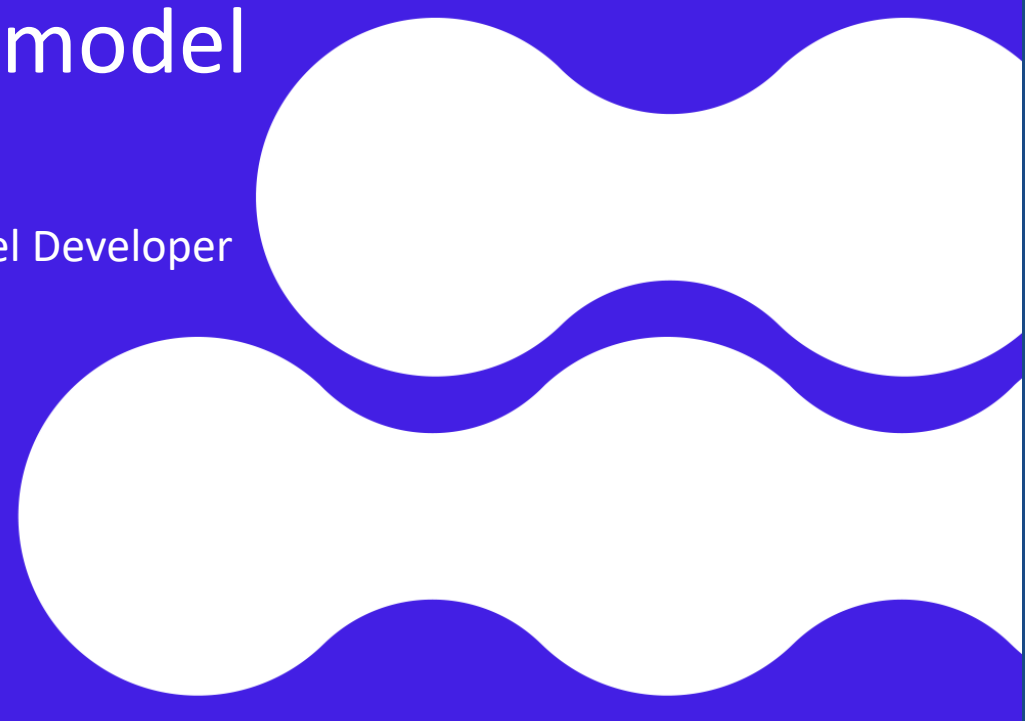
Launch



# Intel® Developer Cloud

## LLM Custom Chatbot development with Huggingface opensource Mistral 7B model

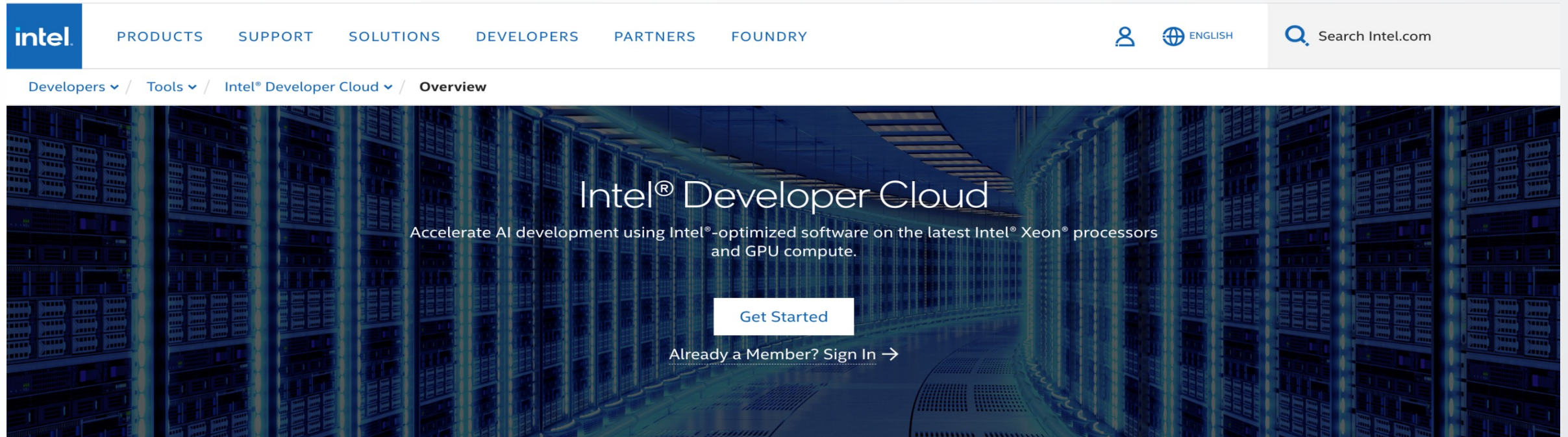
Complete application development on Intel® Max Series GPU in Intel Developer Cloud (IDC)



# LLM Custom Chatbot development with Huggingface opensource Mistral 7B model

Please follow the steps below to get your complimentary access to Intel® Developer Cloud.

1. Go to Intel Developer Cloud: <https://cloud.intel.com/>
2. Click “**Get Started**”



The screenshot shows the Intel Developer Cloud website. At the top left is the Intel logo. A navigation bar contains links for PRODUCTS, SUPPORT, SOLUTIONS, DEVELOPERS, PARTNERS, and FOUNDRY. On the right, there are icons for user profile, language (ENGLISH), and a search bar labeled "Search Intel.com". Below the navigation bar is a breadcrumb trail: "Developers ▾ / Tools ▾ / Intel® Developer Cloud ▾ / Overview". The main content area features a large background image of server racks in a data center. Overlaid on this image is the text "Intel® Developer Cloud" in a large white font, followed by the subtext "Accelerate AI development using Intel®-optimized software on the latest Intel® Xeon® processors and GPU compute." Below this text is a prominent white "Get Started" button. Underneath the button is the text "Already a Member? Sign In →".

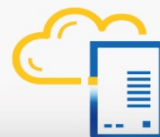
Software

Platforms

Resources

Customers

Sign Up



# LLM Custom Chatbot development with Huggingface opensource Mistral 7B model

Please follow the steps below to get your complimentary access to Intel® Developer Cloud.

3. Subscribe to “**Standard**” service tier and complete cloud registration.

Software

Platforms

Resources

Customers

Sign Up

## Service Tiers

### Standard - Free

Explore and evaluate the latest Intel® AI products plus:

- Develop AI skills.
- Access cutting edge learning resources.
- Get support from the Intel community.

Subscribe

### Premium - Pay as You Go

Get single-user access to the latest Intel products, Standard tier benefits, plus:

- Early prerelease hardware
- AI and machine learning software toolkits
- Intel Premium support

Subscribe

### Enterprise - Team Access

In addition to Premium level benefits, an Enterprise subscription includes discounts and team access to the latest Intel products.

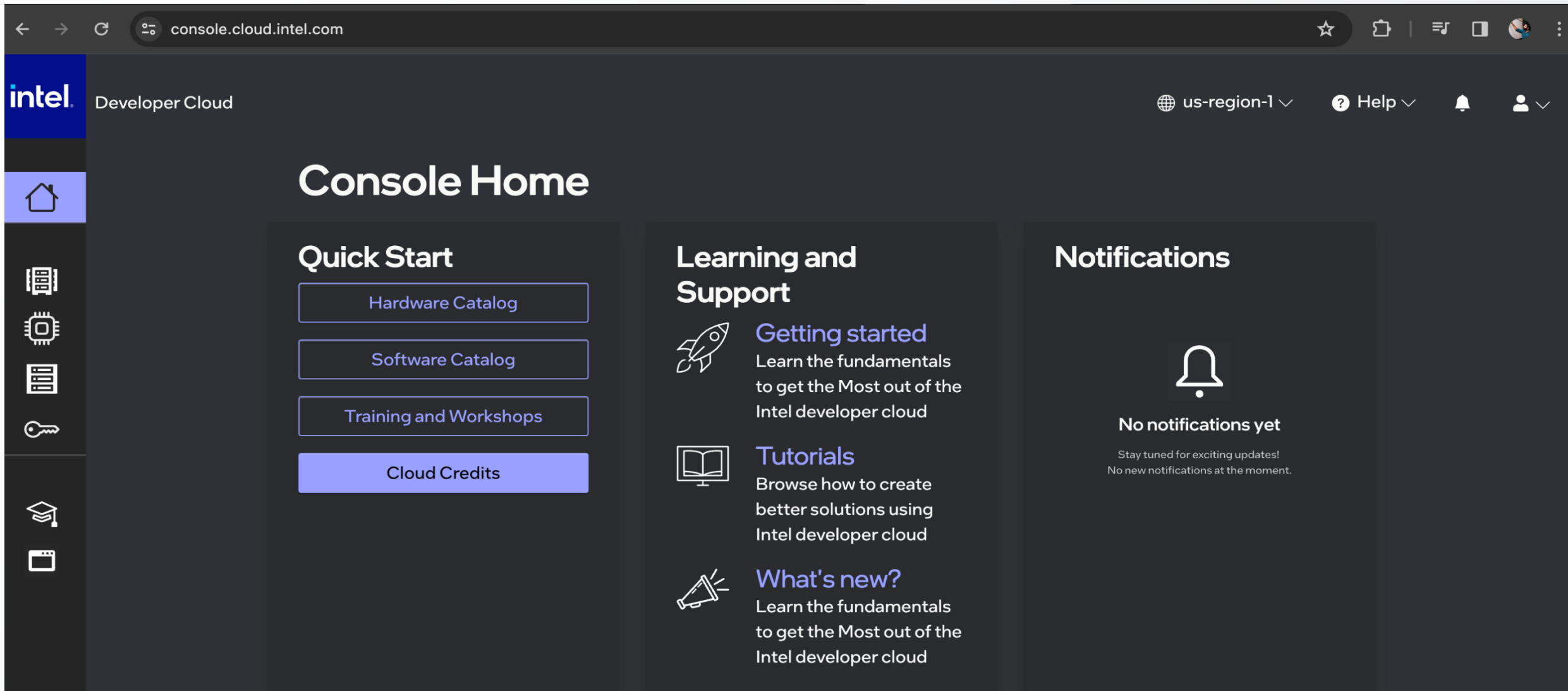
- Billed subscription for teams
- Deploy inference on your own infrastructure
- Use CPU, GPU, and AI accelerators
- 24 x 7 Intel premium plus support

Subscribe

# LLM Custom Chatbot development with Huggingface opensource Mistral 7B model

Please follow the steps below to get your complimentary access to Intel® Developer Cloud.

## 4. Select “Cloud Credits”



The screenshot shows the Intel Developer Cloud console home page. The browser address bar displays 'console.cloud.intel.com'. The page features a dark theme with a blue sidebar on the left containing the Intel logo and navigation icons for Home, Hardware, Software, Training, and Cloud Credits. The main content area is titled 'Console Home' and is divided into three columns: 'Quick Start', 'Learning and Support', and 'Notifications'. The 'Quick Start' column contains buttons for 'Hardware Catalog', 'Software Catalog', 'Training and Workshops', and 'Cloud Credits'. The 'Learning and Support' column includes sections for 'Getting started', 'Tutorials', and 'What's new?'. The 'Notifications' column shows a bell icon and the message 'No notifications yet'.

intel Developer Cloud

us-region-1 Help

## Console Home

### Quick Start

- Hardware Catalog
- Software Catalog
- Training and Workshops
- Cloud Credits**

### Learning and Support

- Getting started**  
Learn the fundamentals to get the Most out of the Intel developer cloud
- Tutorials**  
Browse how to create better solutions using Intel developer cloud
- What's new?**  
Learn the fundamentals to get the Most out of the Intel developer cloud

### Notifications

No notifications yet

Stay tuned for exciting updates!  
No new notifications at the moment.

# intel<sup>®</sup> Developer Cloud

Coupon code: BZM8-86MM-PMFS

Amount: \$50

Expires on 12/31/2023



Scan to access

Disclaimer: "Not a legal tender. Cash value of \$0.001."

# intel<sup>®</sup> Developer Cloud

Please follow the steps below to get your complimentary access to Intel<sup>®</sup> Xeon<sup>®</sup> virtual Server, Intel<sup>®</sup> Max Series GPU and Intel<sup>®</sup> Gaudi2<sup>®</sup> Deep Learning Server:

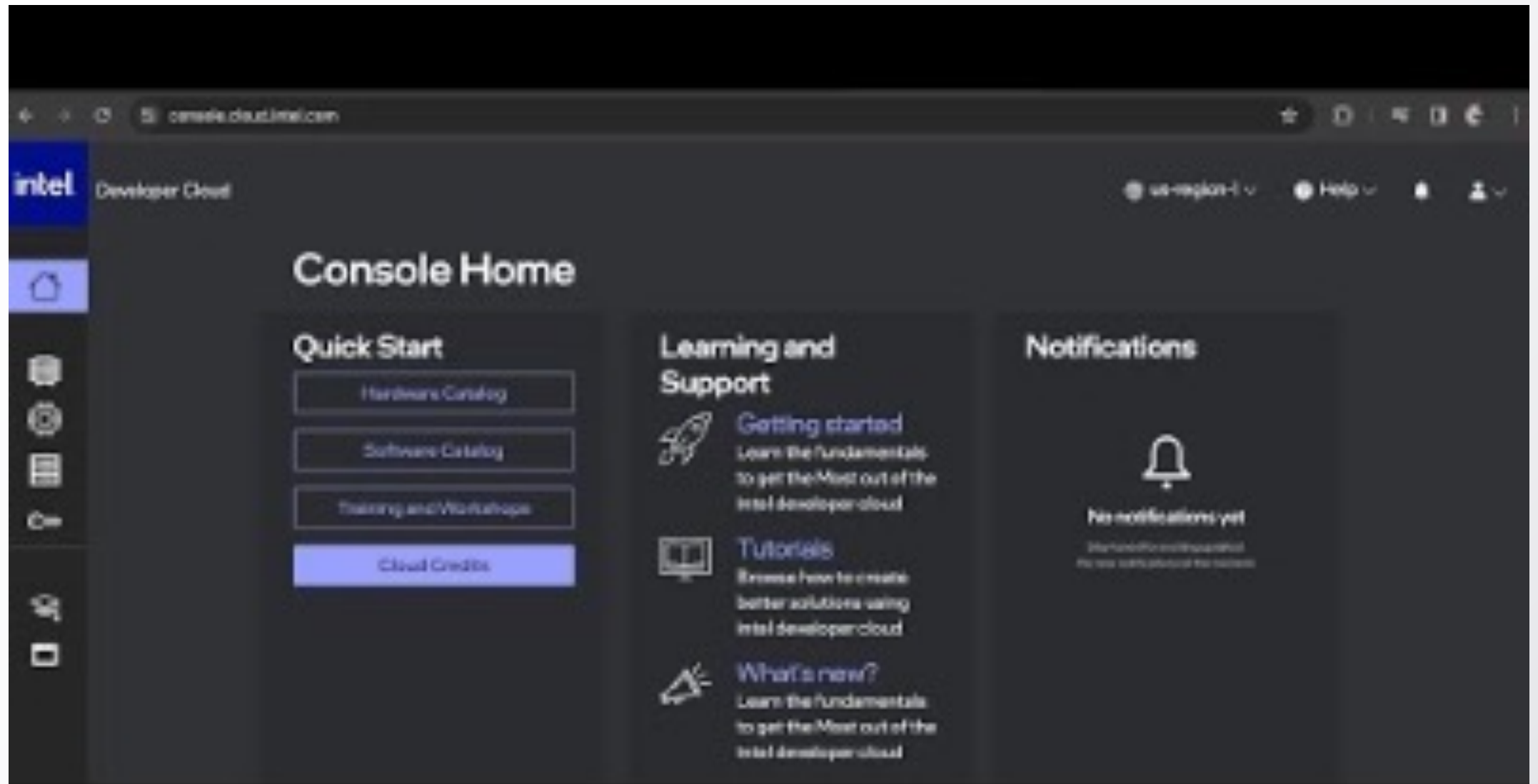
1. Go to Intel Developer Cloud: <https://cloud.intel.com/>
2. Click Get Started
3. Subscribe to “Standard” service tier and complete cloud registration
4. Select “Cloud Credits”
5. Select “Redeem coupon”
6. Enter your Intel Developer Cloud code: BZM8-86MM-PMFS



Scan to access



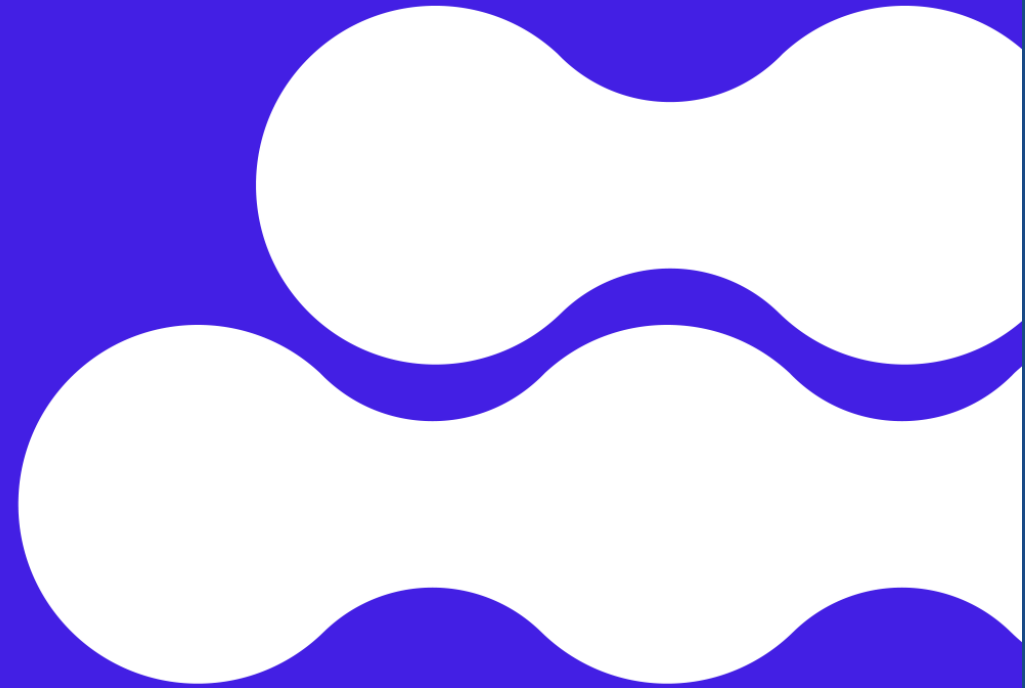
# LLM Custom Chatbot development with Huggingface opensource Mistral 7B model



# FAQ !!

As next step join us on Discord >Intel DevMesh>Global Communities > oneAPIAPJcommunity

<https://discord.gg/ycwqTP6>



Linked In :

<https://www.linkedin.com/in/haquekazi/>

Medium :

<https://medium.com/@kazithaque22>



**Kazi Haque**

Co-Founder @Dynopii Inc.

Topic

Tell Stories / Share  
MEMEs



Thankyou!!

