

# IBM z16 and AI

## IMAGE AND NATURAL LANGUAGE PROCESSING

### Highlights



- Achieve low latency inferencing with first-in-industry IBM z16™ integrated accelerator



- Achieve high throughput while analyzing real-time transactions with IBM z16's on chip AI accelerator

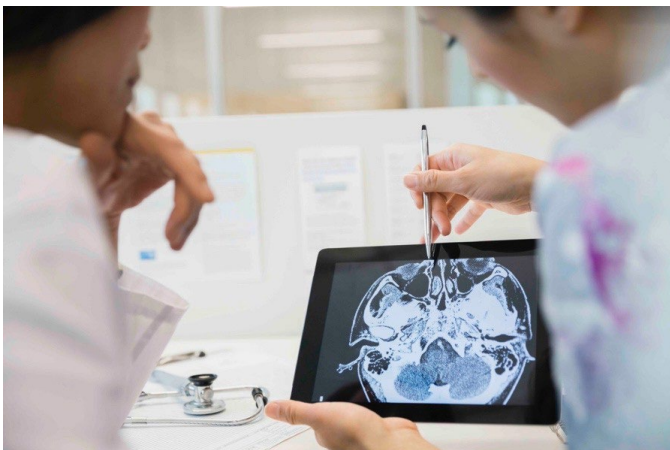


- Design and deploy AI applications on IBM z16, a platform optimized for AI inferencing that supports industry leading open-source tooling and frameworks

Image processing and Natural Language processing form a very important subfield of machine learning. This gives rise to interesting areas like speech recognition, voice generation, auto generating reports, conversational chatbots and many more. Activities like processing images to train AI models can be a very resource intensive task and very energy consuming

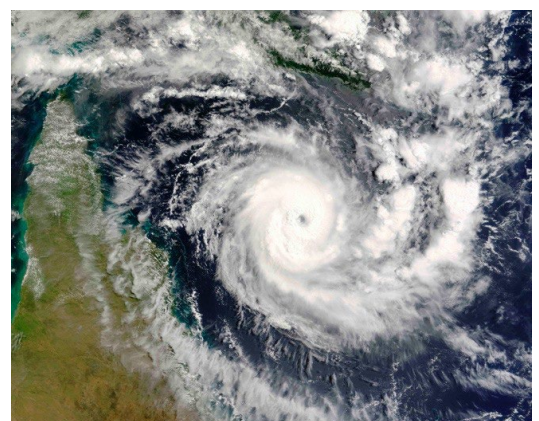
Using AI applications and frameworks running on IBM z16 and LinuxONE, not only achieve energy efficiency and greenAI but also get uncompromised model accuracy

## Go green and yet be efficient







- Enhanced Model Accuracy – Schedule periodic re-evaluations of new data to monitor model accuracy over time and receive alerts when performance deteriorates
- Enterprise-ready AI model deployment- Operationalize models within transactional applications to enable real-time insight. Choose from several scoring approaches, including RESTful APIs, Java™ and CICS® integration, optimized for the highest performance.

- Enhanced Model Accuracy – Schedule periodic re-evaluations of new data to monitor model accuracy over time and receive alerts when performance deteriorates
- Train anywhere deploy on IBM z16 – Train the model on Public Clouds, Private clouds, On-Premises but deploy them on platform



## Advantages

-  Flexible and scalable platform to deploy fraud models and data
-  Score all transactions and leave none behind yet meet all stringent SLAs
-  Detect patterns, adapt more dynamically to different types of images across industries
-  Go green with lesser energy consumption by training models on LinuxONE



 PyTorch



ONNX &



IBM  
Deep Learning  
Compiler (DLC)

### LinuxONE

Model training and deployment

### AI on IBM z16 Solution

On chip accelerator for real-time inferencing



**Image recognition and Natural Language Processing on IBM z16**

### Deep Learning Compiler and ONNX

Enables multiple Deep Learning Frameworks

### PyTorch on IBM z16

Delivers acceleration in an industry standard serving environment

Want to know more? Contact [aionz@us.ibm.com](mailto:aionz@us.ibm.com)