

Case Study 1: AI Stack Development for “AI stack” development platform (AI/ML – Dev)

Business Challenge

AI Stack development for AI Processors



Requirement

- Development of different layers of model execution path
model conversion,
optimization (quantization, pruning)
compilation
- Development of debugging tool



Technology & Tools

- C++, Python
- Graph Compilers (GLOW)
- TensorFlow, PyTorch, ONNX



Value add by Mirafra

- Team worked for various layers like,
 - Model Onboarding.
 - Model optimization (Quantization, Pruning)
- Model conversion
- Compilation and generate graph IR
- Tool development to debugging model execution
- Operator or subgraph level performance analysis

Case Study 2: Information Extraction using OpenAI (AI/ML –

Dev)

Business Challenge

Improvement of recruitment cycle time by auto-extraction from profiles

Requirement

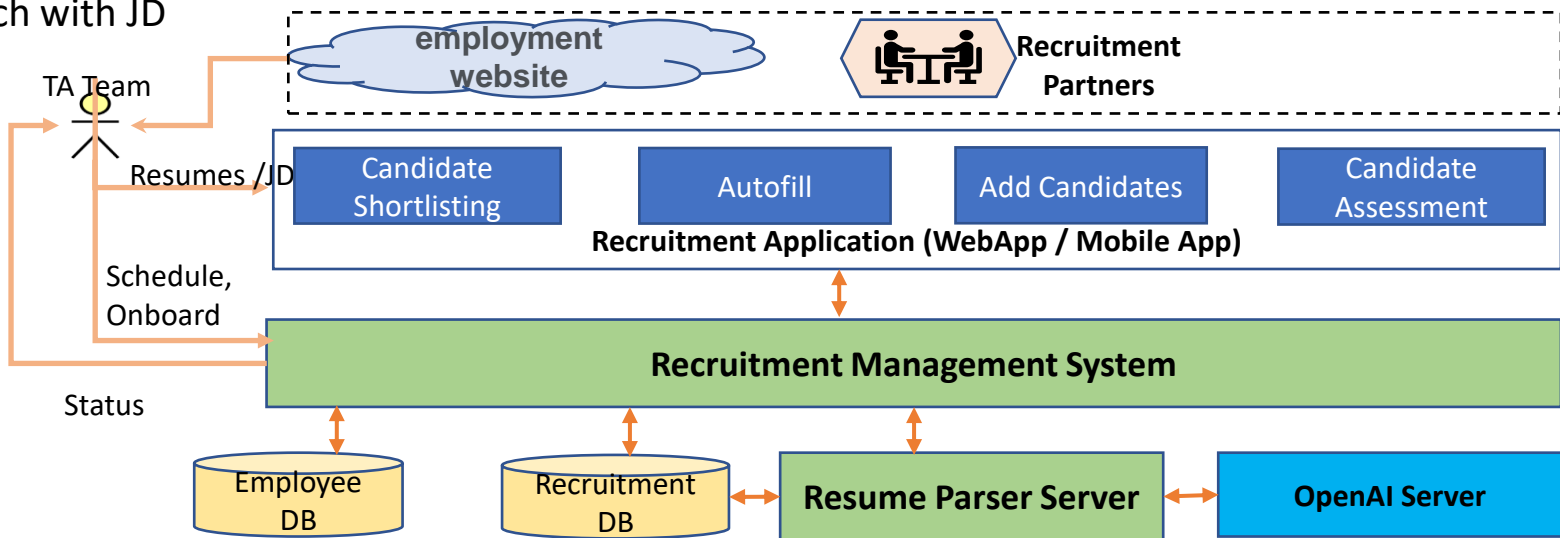
- Develop AI based resume parser
- Provide microservices to the recruitment App server
- Extract information from resume
- Match with JD

Technology & Tools

- Python, Pandas, PyMuPDF, DocX
- OpenAI API, Prompt Engineering
- LLMs, generative AI, PyTorch
- FastAPI, MySQL, MongoDB, Docker, Git

Value add by Mirafra

- Reduce resumes selection and submission process
- Integrate with OpenAI
- Develop prompts to get better response from OpenAI
- Support word and PDF files



Case Study 3: AI Inference Engine (AI/ML – Dev)

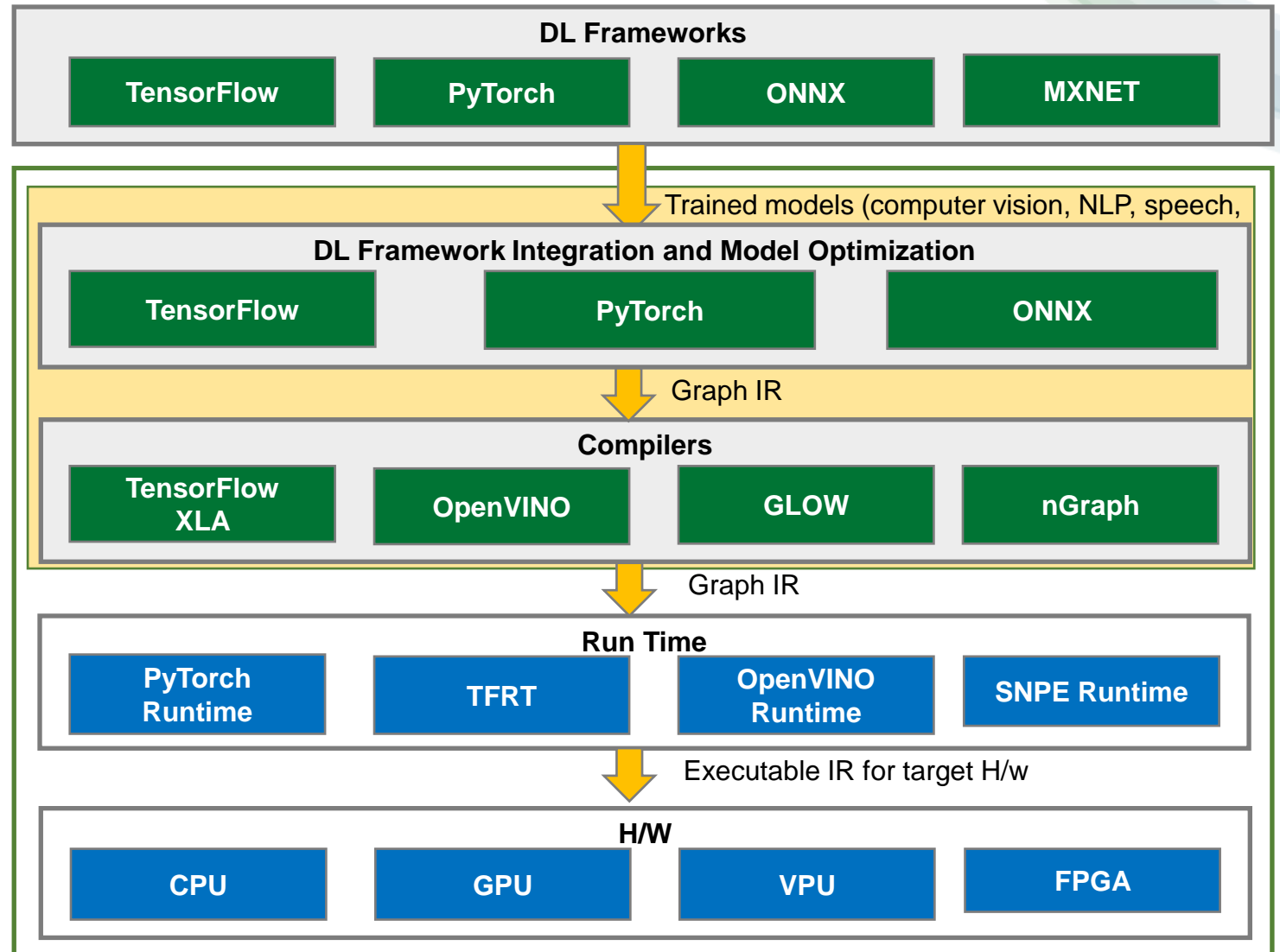
Our Goal:

- Assist various clients to Develop AI Accelerated Inference Engine
- Assist our clients to build next generation AI Chips

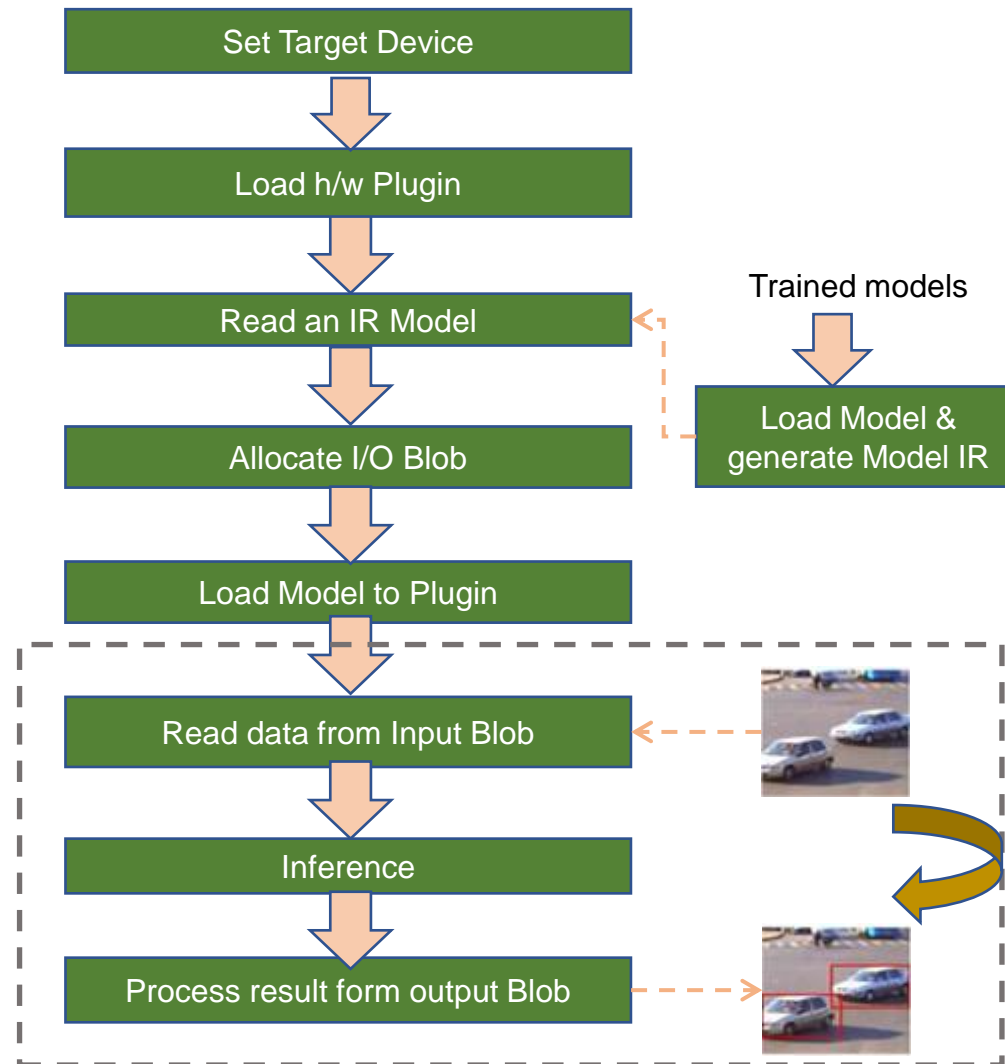
AI Inference Engine:

- AI Inference Engine to boost performance of deep learning models on various h/w platforms (CPU, GPU, VPU, FPGA ...), which in turns accelerate artificial intelligence & machine learning applications.
- AI accelerated chips to support deep learning applications on edge devices.

Deep Learning Model Training to Inference



AI Inference Engine Workflow



Inference Workflow

Case Study 4: Development of AI Enabled Processors for Cars (AI/ML – Dev)

Mirafra's Role:

- Integration of different DL frameworks (TensorFlow, PyTorch, ONNX) with AI Inference Engine
- Model optimization (quantization & pruning), model conversion, model deployment on AI chipset
- Development of Application specific components on top of AI stack
- Model Validation Benchmarking for different target AI platform
- Lane detection

Mirafra Expertise:

Development (Tools and Technologies) :

- Deep Learning frameworks like TensorFlow, Caffe and PyTorch
- Train and Test Deep Learning models for lane
- Cross platform supports using ONNX
- C++
- CUDA, OpenCL
- OpenVX, dlib
- OpenCV
- Computer Vision & Image processing

Case Study 5: OpenVINO Enhancement (AI/ML – Dev)

Mirafra's Role:

- Integration of different DL frameworks (TensorFlow, PyTorch, ONNX) with OpenVINO
- Analyze deep learning models in operator levels, develop plugins if it is not supported in OpenVINO
- Enhance Model Optimizer (Quantization, Pruning, optimization based on Static Analysis)
- AI Workflow for Cloud Integration (Habana-OpenVINO Workflow)
- Model Validation Benchmarking for different target AI platform

Tools and Technologies:

- Deep Learning frameworks like TensorFlow, PyTorch, ONNX
- Analysis different pretrained models
- C++, Python
- nGraph
- OpenVINO

Case Study 6: Failure Prediction in Semiconductor Manufacturing (AI/ML – Dev)

- ❖ Finds critical parameters that have high impact on failure
- ❖ Identify the value range for critical parameters for real-time monitoring in production
- ❖ Finds critical conditions that lead to failure

Rules	Lift	Confidence
V22 > 0.4998335, V512 > 1.168714, V527 <= 1.30043 ---> Fail	2	0.97
V22 <= 0.4998335, V23 <= -0.53650, V286 <= 0.7725662, V441 > -0.617767 ---> Fail	2	0.966
V42 <= -0.1351485, V441 > -0.5314914, V527 <= 0.1197254, V572 > 1.007129 ---> Fail	1.9	0.96
V42 <= 0.3976235, V64 > 0.624203, V521 > -0.01903021, V523 <= -0.3313859 ---> Fail	1.9	0.96
V22 > 0.4998335, V42 > -0.4502337, V121 <= -0.8006536 ---> Fail	1.9	0.952
V11 > 0.5702185, V28 <= 0.5971413, V39 <= -0.3845876, V42 <= -0.1351485, V60 > -0.0733163 ---> Fail	1.9	0.94

Classifier	Recall	Accuracy	Precision	F1-Score	AUC
Random Forest	97	95	97	95	95
SVM	98	96	98	96	96
Xgboost	93	94	93	94	94
Decision Tree	88	87	88	86	87
KNN	99	98	99	98	98
Naive Bayes	51	69	51	61	68
Logistic Regression	76	74	76	74	74

Tools and Technology

- Python, NumPy, Pandas
- R

Algorithms

- Random Forest
- Logistic Regression
- Support Vector Machine
- Association Rule Mining
- Apriori Algorithm

Case Study 7: Automation Testing of AI Chipset (AI/ML – Ver)

Business Challenge

To develop a python based automation framework and testing AI chipset

Requirement

- Develop python based test framework for automation testing of h/w boards
- Integrate with Axiom
- Perform Regression and stress test

Technology & Tools

- Python, Pandas
- Unix Shell scrips
- Jenkins
- Axiom

Value add by Mirafra

- Develop Automation framework & Testing
- Automate testing different AI chips across geography locations
- Model Validation & Benchmarking for different target AI chipset
- Regression Testing & Stress Testing

