Why ONNX Runtime Matters for Deploying AI in Institution

Xavier Tao, Data Engineer, Banque De France
Banque de France is the Central Bank of France, but not only

- Banque de France primary mission is the management of the money supply
- But, we also do regulation of banks and insurances companies
- As well as debt relief for individuals and SMEs, national economic forecast
Artificial Intelligence has disrupted Banque de France operations

- From detecting fake banknotes
- Enhancing auditing capabilities of banks and insurances
- To discovering fraudulent public spending
Institutions AI software mix several ML/DL technologies that is hard to maintain.

Many ML/DL packages

Leads to

- Debugging hell
- Package version incompatibilities
- Security issues
- Overall complexity
Institutions infrastructure generates friction for deployment of AI

- Randomness of result
- Deployment issues
  - « But it worked on my machine »

 Leads to...

CPU / GPU / Cloud
Institutions organizational mix makes it hard to build production AI

**DATA SCIENTIST STARTER PACK...**

- jupyter
- R Studio

With a background in...

- Economics
- Sociology
- Business Administration
- Statistics
- Math

**DATA ENGINEER/OPS STARTER PACK...**

- Docker
- Kubernetes
- Jenkins
- Java

With a background in...

- Support
- DevSecOps
- Software Engineering
- IT
ONNX as a solution to the software, hardware, and organization problems
ONNX is the Conway interface between Data Scientist and Data Engineer

Any organization that designs a system will produce a design whose structure is a copy of the organization.
— Melvin E. Conway
Keeping safespaces between teams, with a well defined interface

- Training of Model
- Data Exploration
- Hyper parameter tuning
- AI Production Pipeline

Data Scientist Safespace

ONNX Model

Data Engineer / Ops Safespace
ONNX simplify AI and makes it more Unix like

Write programs that do one thing and do it well.
Write programs to work together.
Write programs to handle text streams, because that is a universal interface.

- Unix Philosophy
Open format enables new open MLOps tools
Key Takeaways

■ ONNX as interface between:
  ▪ Softwares
  ▪ Hardwares
  ▪ Teams: Data Scientist and Data Engineer

■ With other nice features such as:
  ▪ Simpler to run than frameworks
  ▪ Open format -> Extensible

■ Current edge cases:
  ▪ Ecosystem is still pretty young.
  ▪ Maybe a bit over sophisticated for small POCS / MVP.
  ▪ Maybe a bit over sophisticated for ML/DL research.
Annex Case Study: A PDF Analyzer using AI

Banks

Loan Contracts

Checking Loan motives
Checking Loan Amounts Validity
Checking signatures

SaaS Application

DATA SCIENTIST SCOPE
SWE/DATA ENGINEER/OPS SCOPE