

---

# WhatNext Tech 1.0

## MetaFlow @ Netflix

### Tavolo 1

**Cristina Duma, Stefano Stalio, Federica Legger,  
Tommaso Tedeschi, Tommaso Boccali  
Stefano Bagnasco (Facilitator)**

**Giugno 2019**

# Presentazione del problema

- Le tecniche di **Machine Learning (ML)** e **Deep Learning (DL)** si stanno diffondendo sempre di più negli esperimenti
  - analisi di dati, data quality, trigger, tracking, etc.
- Vasta gamma di ML/DL frameworks disponibile sul mercato.
  - Tipicamente forniscono accesso a uno o più librerie di ML e possono essere facilmente integrate con sistemi di calcolo distribuito tipo Spark
- Quello che spesso manca a questi framework:
  - versioning
  - riproducibilità
  - strumenti collaborativi
  - strumenti per fare il deploy dei modelli in ambienti di produzione

# Descrizione dell'idea del prodotto

- **Metaflow** - framework sviluppato da Netflix che sarà rilasciato open-source nei prossimi mesi
  - architettura a grafi
  - python/R
  - interfaccia alle principali librerie di ML
  - semplici comandi (decorazioni) per eseguire codice su risorse diverse (portatile, cloud, GPU) e fare deploy
  - versioning
  - condivisione di parti di codice tramite namespaces
  - persistenza automatica del risultato di ogni grafo
  - deploy di microservizi web via curl

# Mercato

- “Acquirente”:
  - gli esperimenti (LHC e non)
- Consumatore:
  - utenti dell’esperimento (sviluppatori, fisici, comunità di detector, trigger, software, ecc)
- Ogni qualvolta ci sia necessità di:
  - sviluppo di modelli ML a più mani, riproducibili, riutilizzabili
  - deploy di modelli ML in un ambiente di produzione

# Validazione del concetto da parte delle principali parti interessate

- **Netflix** proclama che il tempo medio di messa in produzione dei modelli ML sviluppati dai loro data scientist è **sceso da 4 mesi a 7 giorni** dopo l'introduzione di Metaflow
- L'idea è stata discussa nel contesto dell'attività di Operational Intelligence
- La **riproducibilità** e **condivisione** delle nostre analisi è una necessità conclamata degli esperimenti
- Per molte comunità (detector, trigger, software) i modelli ML devono essere **deployati in un ambiente di produzione**, con possibilità di attuare velocemente **hot fixes, rollback**, etc

# Cosa serve per perfezionare il concetto?

- Mettere le mani sul **sorgente**
  - siamo in contatto con Netflix per partecipare a una fase di beta test
  - valutazione del prodotto e delle sue reali capacità di adattarsi alle nostre infrastrutture/tecnologie
  - implementazioni di show case d'uso negli esperimenti