

- Versioning : 100 Gg

Modeling

- train/ dev/ test → ~~90%~~ 85, 80%
 train
 Domain ← → 90% → Deploy 70%
- Multiple metric → (Accuracy, Latency)
 max(Accuracy) | Latency < 0.1
 Deep → 10s | 80% ✓
 Lining → 0.01s | 78%
- Model selection (Deep vs classic) text : BERT, Trueta, (Regx)
 Call reason → 0, 12 Wikiped C
- Error Analysis 80% 50
 trainset, dev
- Data → noise label
- Experiment tracking
- Deployment Ready

Deployment

(50% startup fail to deploy)

- latency / throughput (QPS) ASR : 30K x 5min = 150K min = 20k h
- Logging / monitoring (feedback) / Alert CA : ASR ↓
 → KPI data drift
- Agile / time consuming (MLOps)
- versioning 1.0.2 1.2.3
 ↓
 2
- Exception handling KYC : video format, fimpug 10k