

# Artificial Intelligence needs Backend & DevOps to reach the real-world

Nerea Luis @sailormerqury · SNGULAR Data&AI





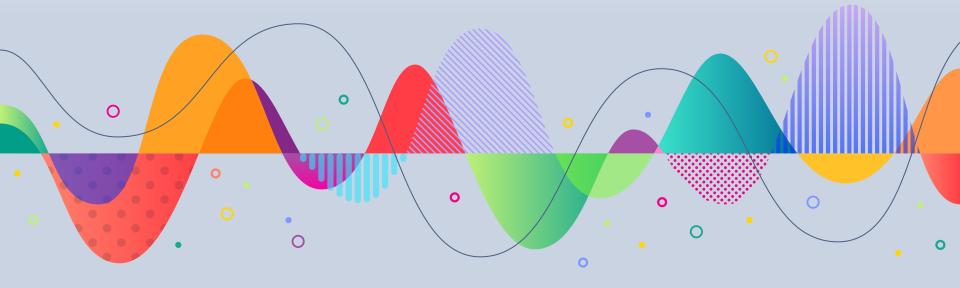
Nerea Luis (nerealuis.es)

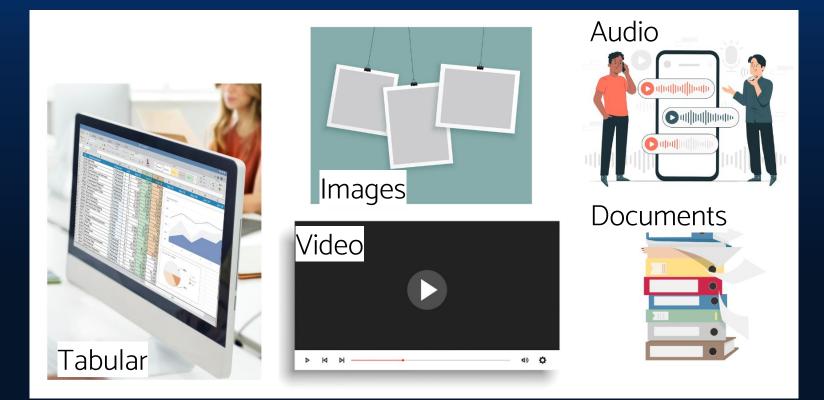
Artificial Intelligence lead @ SNGULAR PhD Artificial Intelligence

Sci-comm, women-in-tech advocate

Template: Slides Carnival

# **1. The Data-Scientist work**







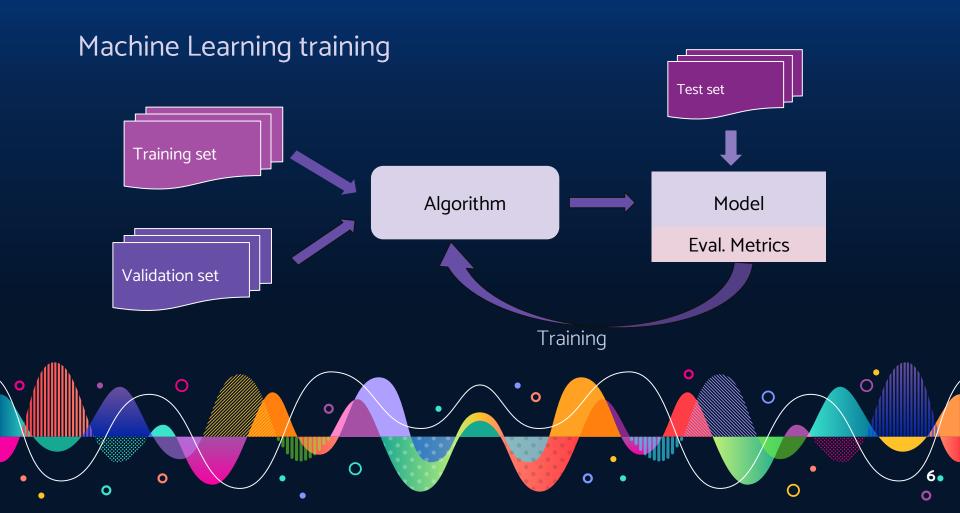


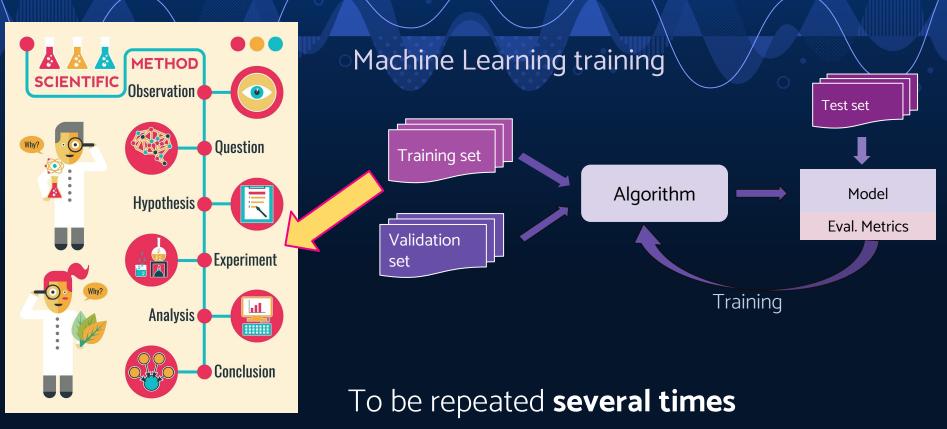


Exploring data sources

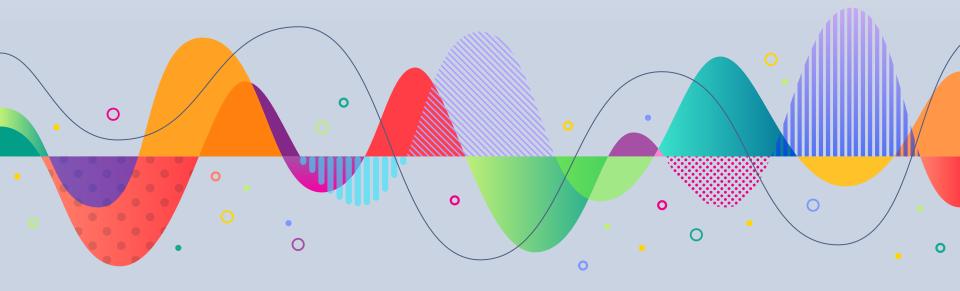
Dataset preprocessing Exploratory Data Analysis Data labeling, data aggregation







# 2. I already have an Al model. What's Next?



### 2015 830 cites

### Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips

{dsculley,gholt,dgg,edavydov,toddphillips}@google.com Google,Inc.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison {ebner, vchaudhary, mwyoung, jfcrespo, dennison}@google.com Google, Inc.

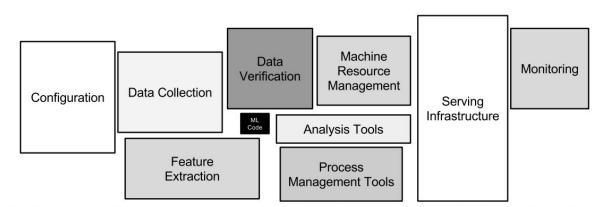
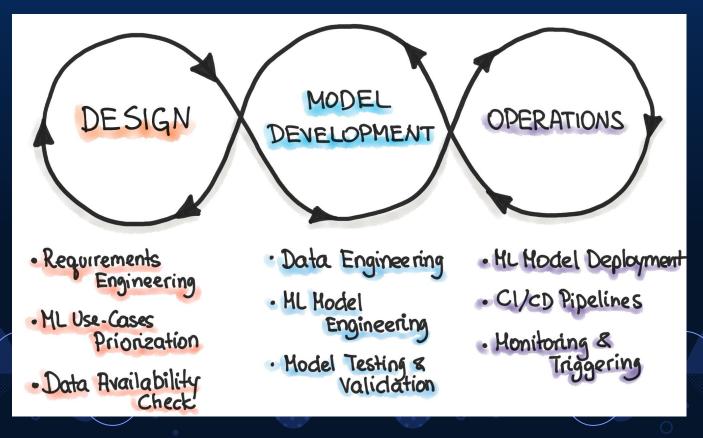


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

### Machine Learning "Operations"

10



C

### **Continuous Integration**

Cl is no longer only about testing and validating code and components, but also testing and validating data, data schemas, and models. CD

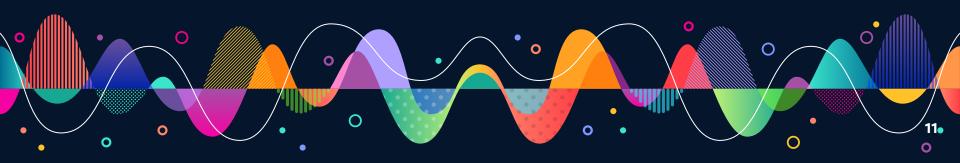
#### **Continuous Delivery**

CD is no longer about a single software package or a service, but a system (an ML training pipeline) that should automatically deploy another service (model prediction service).

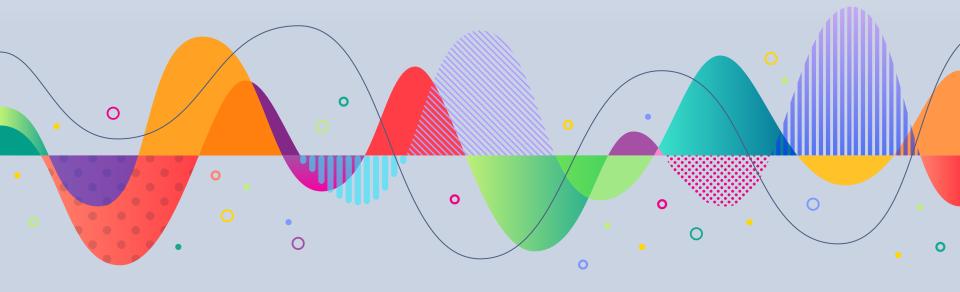
#### **Continuous Training**

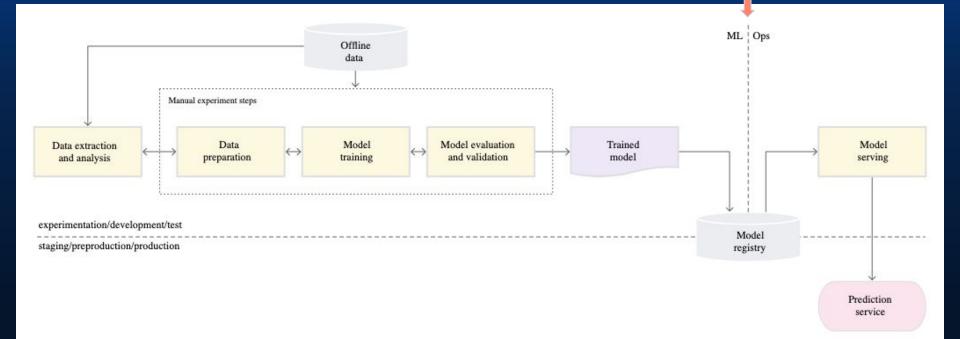
CT is a new property, unique to ML systems, that's concerned with automatically retraining and serving the models.

<u>Reference</u>



# **3. Accelerating AI in production**

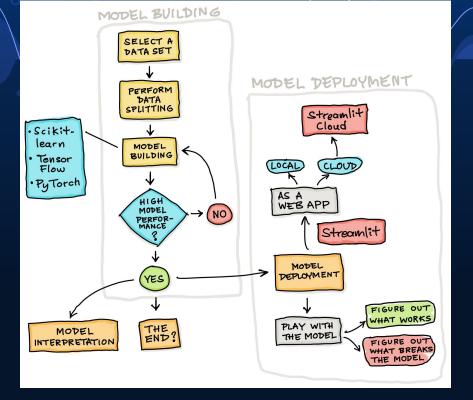








Q 0





#### **Streamlit Face-GAN Demo**

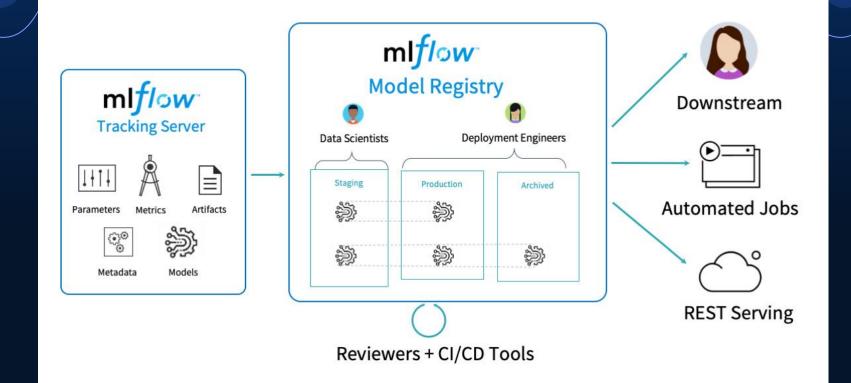
This demo demonstrates using <u>twidia's Progressive Growing of GANs</u> and Shaobo Guan's <u>Transparent</u> <u>Latent-space GAN method</u> for tuning the output face's characteristics. For more information, check out the tutorial on <u>Towards Data Science</u>.



Let's work on some GUI

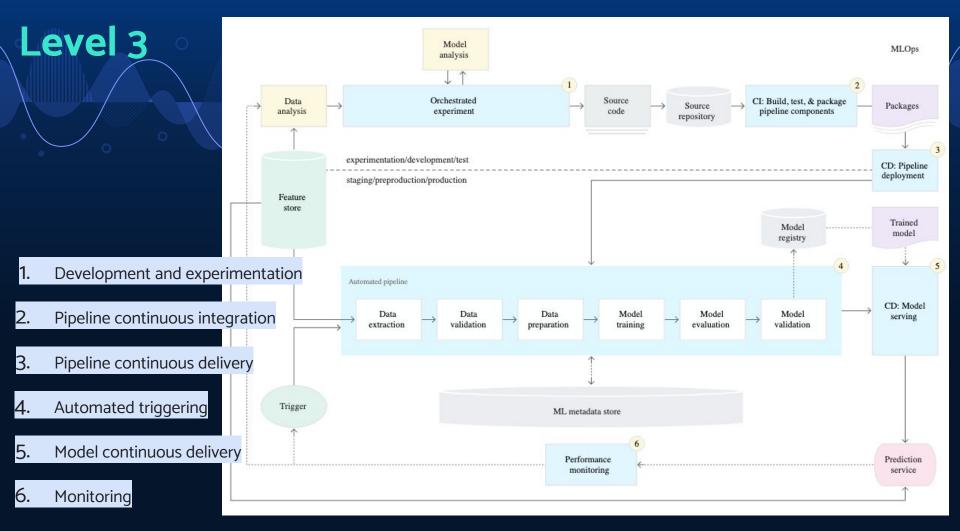


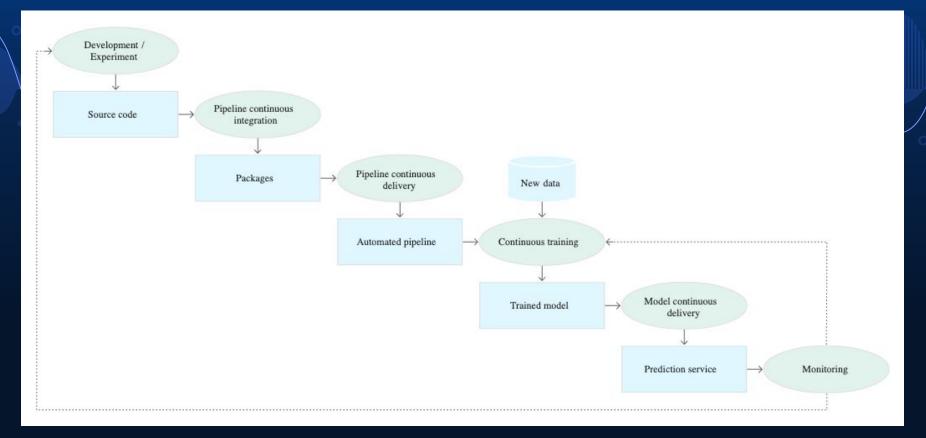
## Level 2 - MLFlow



dels													(	GitHub Do
Defa	ault 🗍													Sha
0	Track ma	achine learning t	raining runs	in an experime	int. Learn more									
	riment ID:													
	escriptio			Decenter of										
5	Refresh	Compare	Deinte	Downio	ad CSV	↓ Start Time ∨	All time	~						
<i>€</i> ) ≣	Refresh	@Columns		w differences (				ams.model = "tree"	Search	<b>-</b> Filter	Clear			
=	•								Search		clear Aetrics		Parar	neters
=	•	Columns										-	Paran	neters
=	<b>⊞</b> ing 5 mat	Columns								N		Recall	and the second second	
E Show	ing 5 mat	@Columns	Only sho	w differences (	•	<b>Q</b> metrica.mae				Metrics >	1etrics	Recall	Parameters	
Show	Ing 5 mat I Star ⊘	Columns ching runs	Only sho Duration	w differences ( Run Name	User	Q .metrica.mue	< 1 and part Version	anna model = "tres" Models		Metrics > Accuracy	Precision		Parameters max_depth	n_estimat
Show	ing 5 mat	Columns ching runs t Time 4 seconds age	Only sho Duration 2.6s	w differences ( Run Name	User smighani	Q metrica.mue	< 1 and per Version	annumodel = "tree" Models		Metrics > Accuracy 0.968	Precision 0.987	0.947	Parameters max_depth None	n_estimat 200
Show	I Star	Columns ching runs t Time 4 seconds ago 51 seconds ago	Only sho Duration 2.6s 2.3s	w differences ( Run Name -	User smighani smighani	Q metricumme Source ipykernel_ ipykernel_ ipykernel_	< 1 and part Version	annu model = "tree" Models R sklearn S sklearn		Metrics > Accuracy 0.968 0.765	Precision 0.987 0.712	0.947 0.891	Parameters max_depth None 2	n_estimate 200 150
Show	I Star	Columns ching runs t Time 4 seconds ago 51 seconds ago 1 minute ago	Duration 2.6s 2.3s 2.3s	w differences ( Run Name - -	User smighani smighani smighani	Q metricul mae	< 1 and part Version - -	annu model = "tree" Models Sidearn Sidearn Sidearn		Metrics > Accuracy 0.968 0.765 0.786	Precision 0.987 0.712 0.746	0.947 0.891 0.868	Parameters max_depth None 2 2	n_estimate 200 150 150

MLFlow Main Dashboard



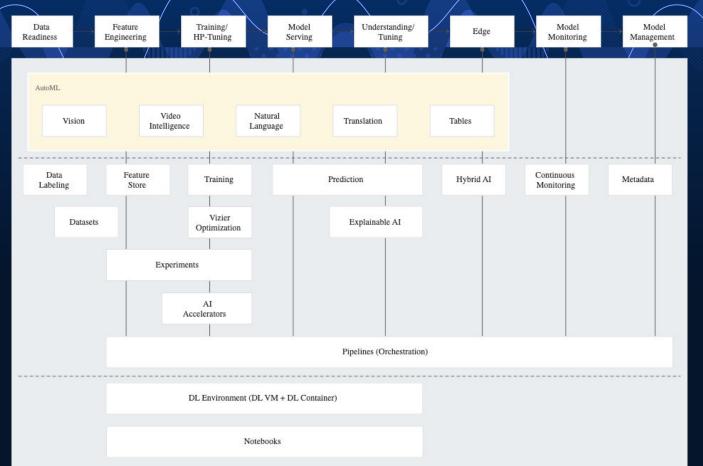


## ML-Ops State Machine

# Level 3 - Kubeflow



# Level 3 - VertexAl (GCP)



# Main Conclusions

- Data Scientists and Data Engineers cannot work alone anymore. These profiles should be complemented with Backend, DevOps, QA
- ML-Ops is still a WIP field. The need to move AI models to production environments is increasing. The field will grow, lot of work to be expected here
- We can start by providing to Data Scientist simple API+Frontend to experiment with Model Serving. Then, Model Registry and Tracking Server
- Whole experience includes Continuous Delivery, Continuous Integration and Continuous Training



### nerea.luis@sngular.com

### ai.sngular.com

Template: Slides Carnival