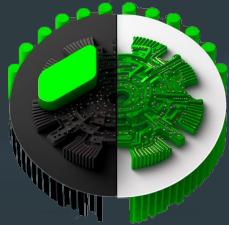


DATA AUGMENTATION

L3Q1 Group - Supervisor : Elie EL DEBS

Ilan' DAUMONT-OUK
Mahyr-Florian ABOU ASSAF
Ethan CHEMLA
Alexandre NGO



DATA FLOW AI



Year : 2022-2023



Plan





Introduction - The team

**Mr. EL
DEBS
Elie**

CHEMLA Ethan



NGO Alexandre



ABOU ASSAF
Mahyr-Florian



DAUMONT-OUK
Ilan'





Introduction - Definition

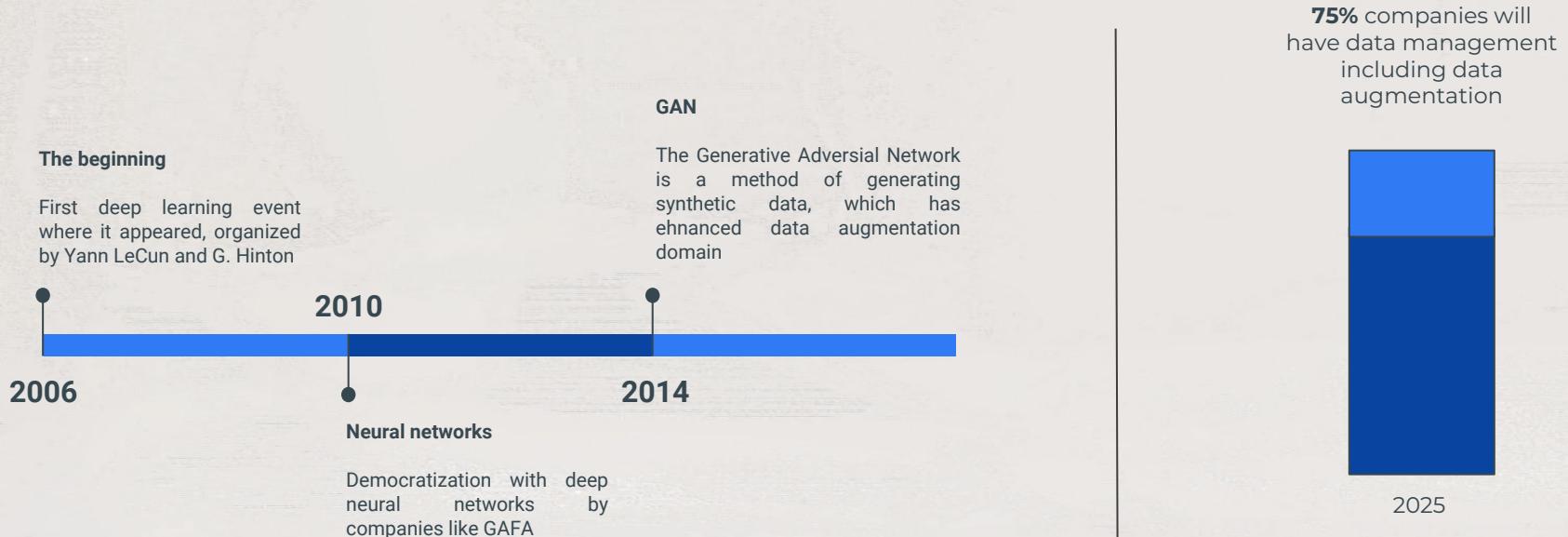
Data augmentation

is a technique of artificially increasing the training set by creating modified copies of a dataset using existing data.





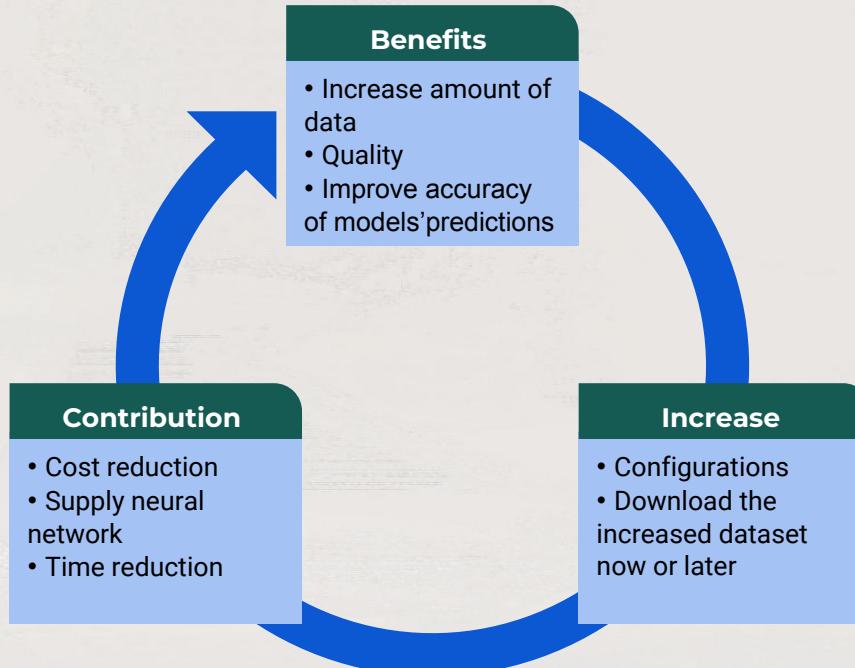
Introduction - Context





Introduction - Life cycle

WHY WOULD YOU USE DATA AUGMENTATION ?





Objective

The objective is to create a web application that allows using data augmentation with pictures.

Web application

Database

Server

Increased data

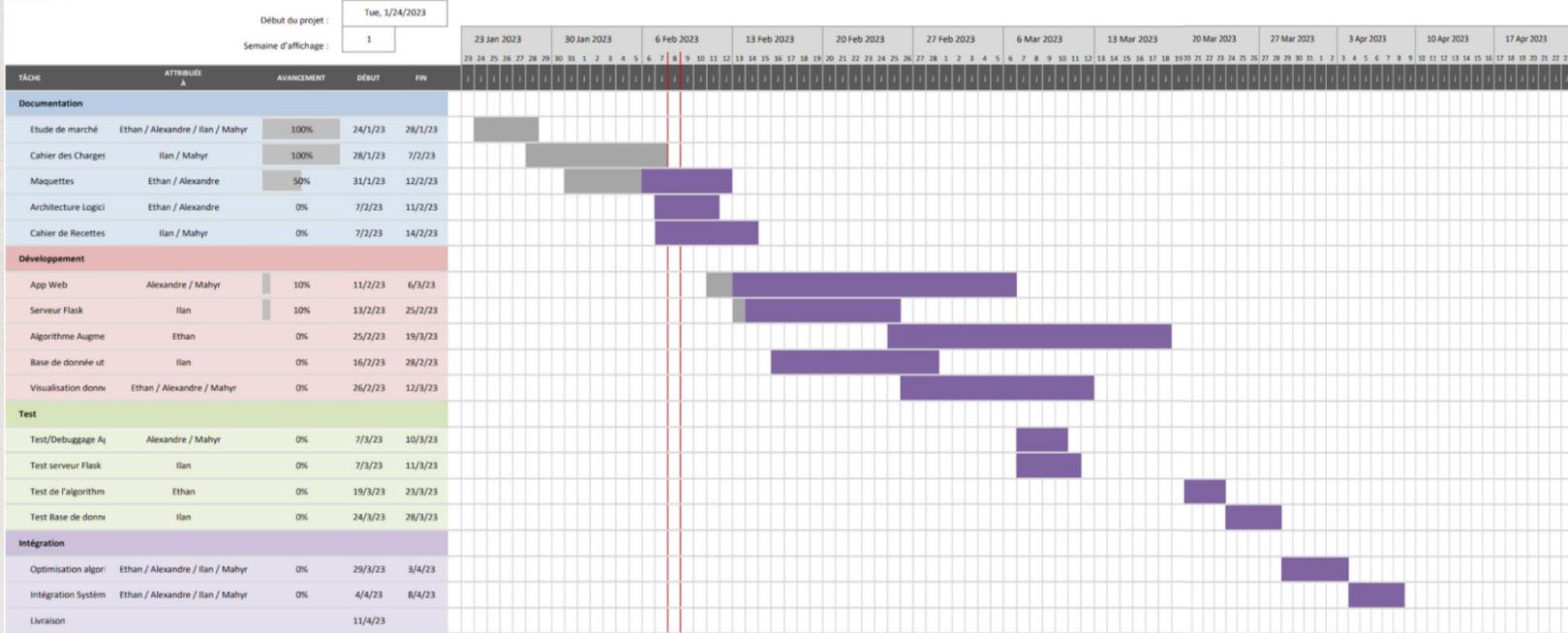




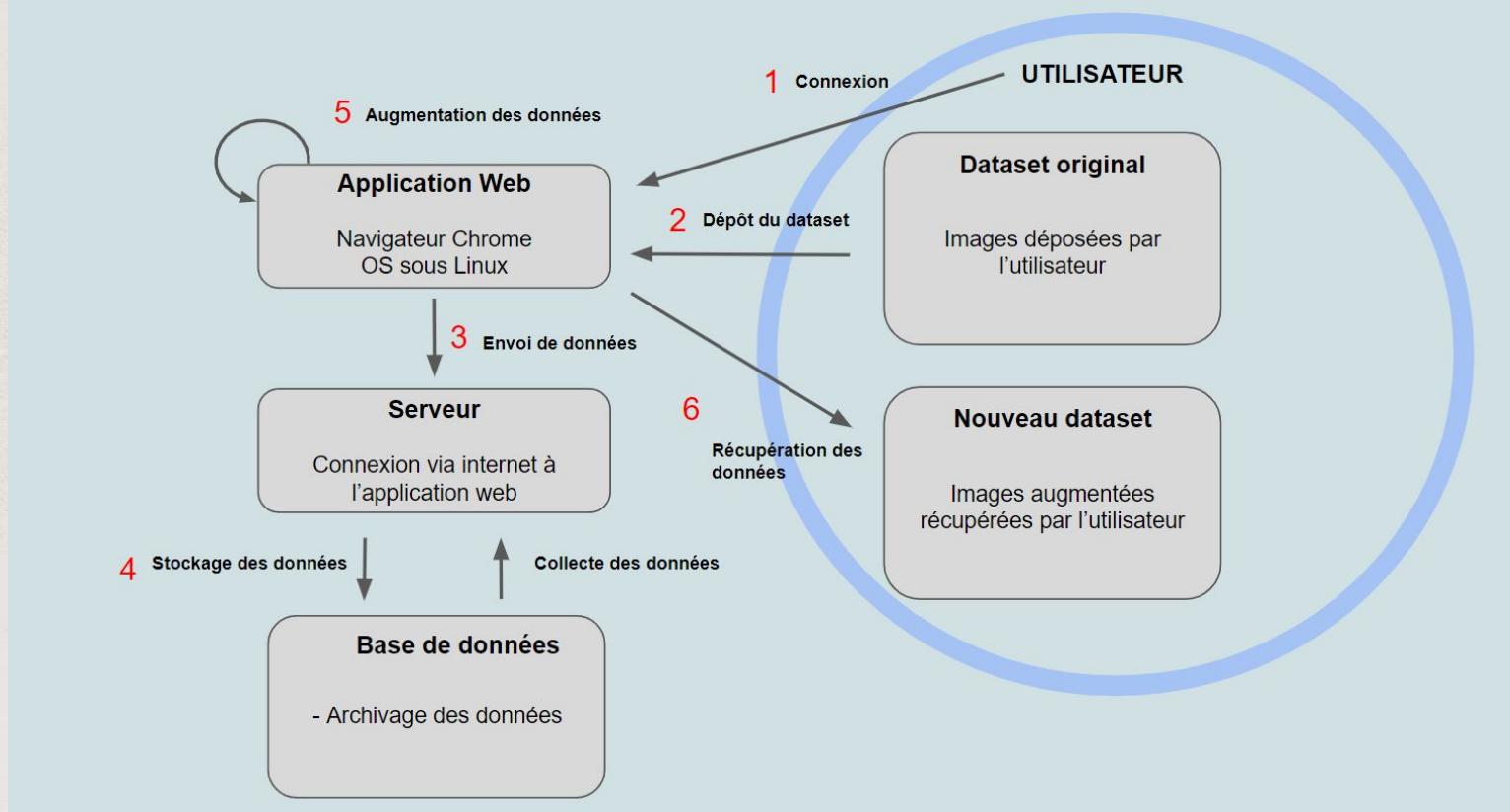
Project Organisation

L3Q1 - Outil d'augmentation de la Donnée

DataFlow AI



Architecture of our web application



Front-end programming languages

HTML



CSS



JavaScript



Bootstrap

HTML (Hypertext Markup Language) allows for structuring and presenting the content of a website

CSS (Cascading Style Sheets) is used to control the visual presentation of web pages.

JavaScript is used to add interactive features to web pages, such as animations, visual effects, etc

A framework that allows for rapid and easy web development. It provides pre-built HTML components, CSS styles, and JavaScript scripts



Back-end programming languages



Python web framework used for web application development. Provides an interface for interacting with databases, handling HTTP requests, managing user sessions, etc.



NoSQL database, used for storing data.



Main treatments

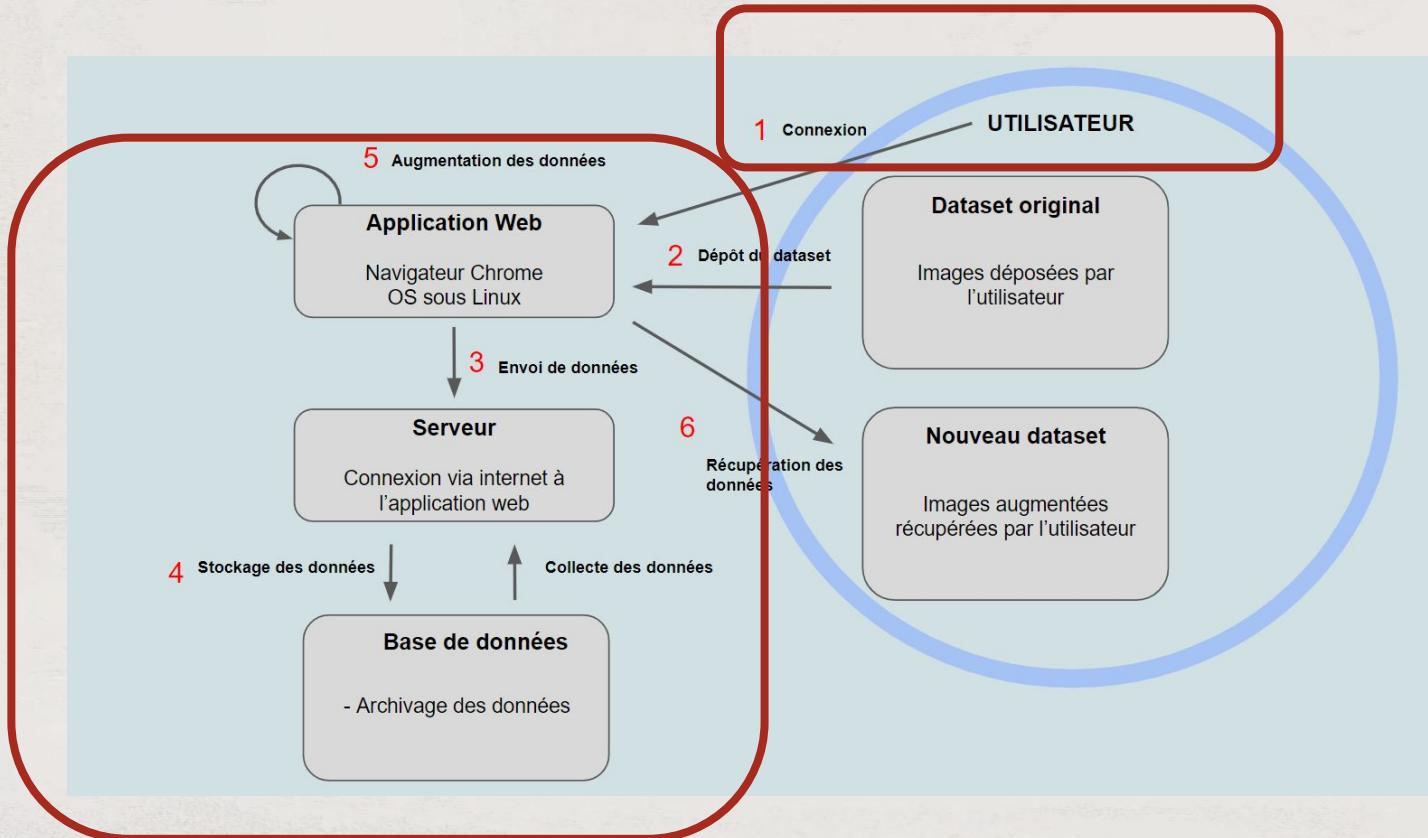
- **Rotation** : transformation that turns an object around a fixed point by a certain angle in a specific direction
- **Noise** : adding noise refers to the process of introducing random variations or perturbations to the data in order to increase the robustness
- **Luminosity/Contrast** : modifying the contrast and brightness levels of an image typically by scaling the pixel values
- **Blur** : adding blur refers to the process of applying a blur filter to an image in order to simulate a certain degree of blurriness or fuzziness







Solution overview - Database





Solution overview - Database



mongoDB

NoSQL

Storage : collections and documents

pymongo

Python driver for MongoDB

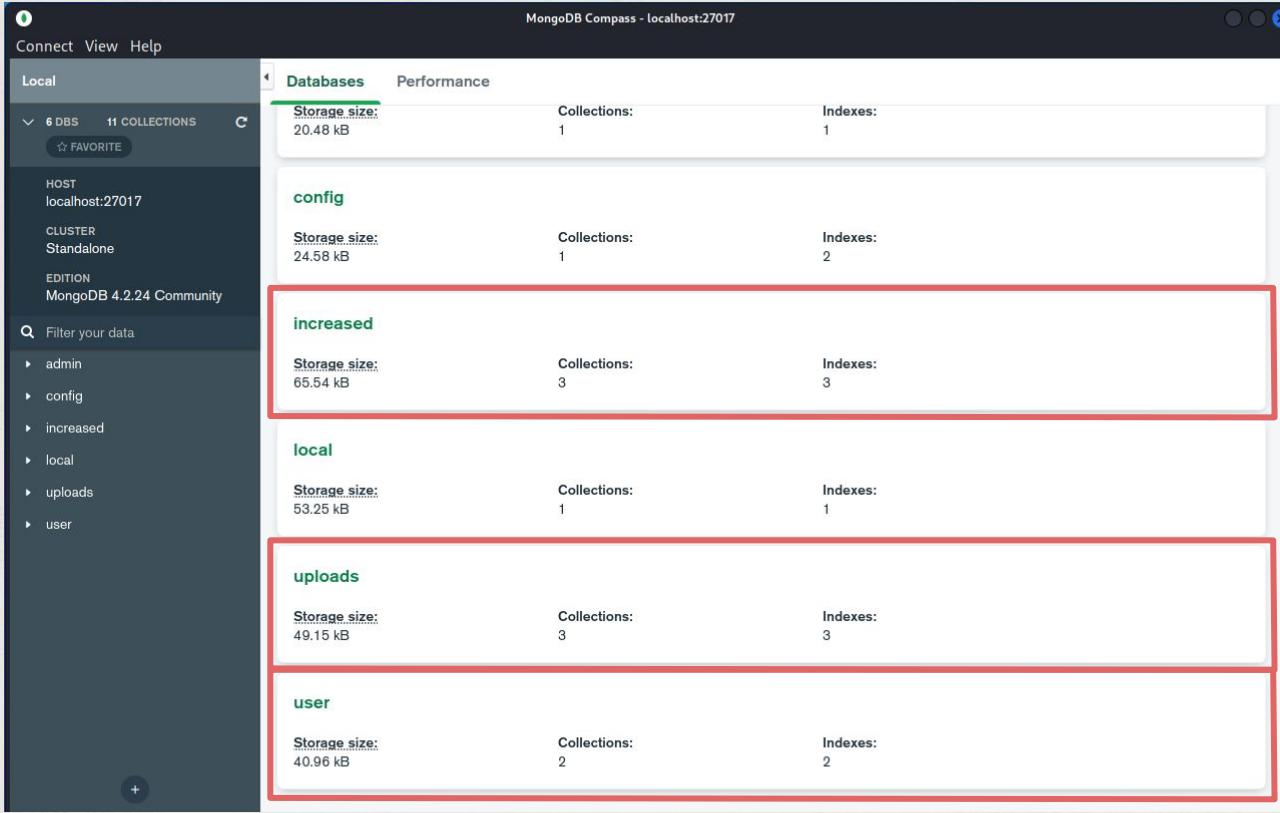


mongoDB Compass

Storage visualization for developers



Solution overview - Database



The screenshot shows the MongoDB Compass interface connected to a local host at port 27017. The left sidebar lists databases: Local (with 6 DBs and 11 Collections), config, increased, local, uploads, and user. The main area displays statistics for each database:

Database	Storage size:	Collections:	Indexes:
config	24.58 kB	1	2
increased	65.54 kB	3	3
local	53.25 kB	1	1
uploads	49.15 kB	3	3
user	40.96 kB	2	2





Solution overview - Database

uploads.64396cc7dafd85f2f8a6ecbd

DOCUMENTS	2	STORAGE SIZE	24.6KB	Avg. Size	117B	INDEXES	1	Total Size	24.6KB	Avg. Size	24.6KB
Documents	Aggregations	Schema	Explain Plan	Indexes	Validation						

FILTER { field: 'value' } **OPTIONS** **FIND** **RESET** **...**

ADD DATA **VIEW** **...** Displaying documents 1 - 2 of 2 **REFRESH**

```
_id: ObjectId("643ea9c5fcbed392c610ad89")
filename: "oiseau.png"
path: "uploads/64396cc7dafd85f2f8a6ecbd/oiseau.png"
date: 2023-04-18T16:31:33.983+00:00
```

```
_id: ObjectId("643ea9c5fcbed392c610ad8a")
foldername: "Test_Images"
path: "uploads/64396cc7dafd85f2f8a6ecbd/Test_Images"
date: 2023-04-18T16:31:33.996+00:00
```

increased.643b4755a5cd2386ab99a263

DOCUMENTS	2	STORAGE SIZE	36.9KB	Avg. Size	129B	INDEXES	1	TOTAL SIZE	36.9KB	Avg. Size	36.9KB
Documents	Aggregations	Schema	Explain Plan	Indexes	Validation						

FILTER { field: 'value' } **OPTIONS** **FIND** **RESET** **...**

ADD DATA **VIEW** **...** Displaying documents 1 - 2 of 2 **REFRESH**

```
_id: ObjectId("643e6e21a9589fdf7b2460d4")
foldername: "oiseau(2)"
path: "./increased_data/643b4755a5cd2386ab99a263/oiseau(2)"
date: 2023-04-18T12:17:05.224+00:00
```

```
_id: ObjectId("643e6e22a09589fdf7b2460d5")
foldername: "test_images(2)"
path: "./increased_data/643b4755a5cd2386ab99a263/Test_images(2)"
date: 2023-04-18T12:17:06.731+00:00
```



Solution overview - Database

MongoDB Compass - localhost:27017/user.logs

Connect View Collection Help

local

6 DBS 11 COLLECTIONS C

HOST localhost:27017

CLUSTER Standalone

EDITION MongoDB 4.2.24 Community

Filter your data

- admin
- config
- increased
- local
- uploads
- user
 - files
 - logs

+

user.logs Documents

user.logs

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER { field: 'value' } OPTIONS FIND RESET ...

Displaying documents 1 - 3 of 3 < > C REFRESH

Document 1:

```
firstName: "test"
lastName: "test"
email: "test@gmail.com"
password: Binary('J0DiJDEyJC9acGloFBURhpIbVVXSkd3aENzSnvhMvHMRgtHwxlybTR3RlljLopDdxY4Vw1lcwxQyldt', 0)
passwordConfirm: Binary('J0DiJDEyJC9acGloFBURhpIbVVXSkd3aENzSnvhMvHMRgtHwxlybTR3RlljLopDdxY4Vw1lcwxQyldt', 0)
country: "Argentine"
gender: "Homme"
terms: "on"
salt: Binary('J0DiJDEyJC9acGloFBURhpIbVVXSkd3aENzSnU=', 0)
```

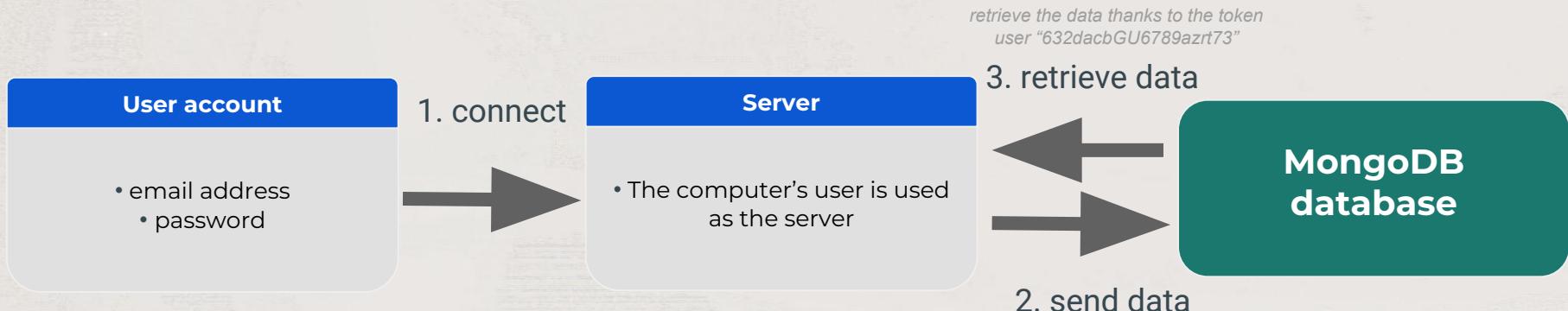
Document 2:

```
_id: ObjectId("644299e4a00031425d3d7f48")
firstName: "Bertrand"
lastName: "Renard"
email: "bertrand.renard@gmail.com"
password: Binary('J0DiJDEyJDFA1NSWEg5U2hDS21is2JvUWnvUy5WMHAzV2Q3ZjUwQW1oUHEyT2RqW5HV6dsU3Y4S2lP', 0)
passwordConfirm: Binary('J0DiJDEyJDFA1NSWEg5U2hDS21is2JvUWnvUy5WMHAzV2Q3ZjUwQW1oUHEyT2RqW5HV6dsU3Y4S2lP', 0)
country: "France"
gender: "Homme"
terms: "on"
salt: Binary('J0DiJDEyJDFA1NSWEg5U2hDS21is2JvUWnvUy4=', 0)
```



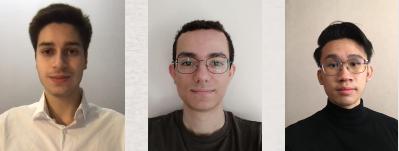
Solution overview - Database



User token ID	File details		
	Filename	Filepath	Date
632dacbGU6789azrt73	image.png	./uploads/632da.../image.png	2023-04-18 T 16:30:55



Task Repartition :

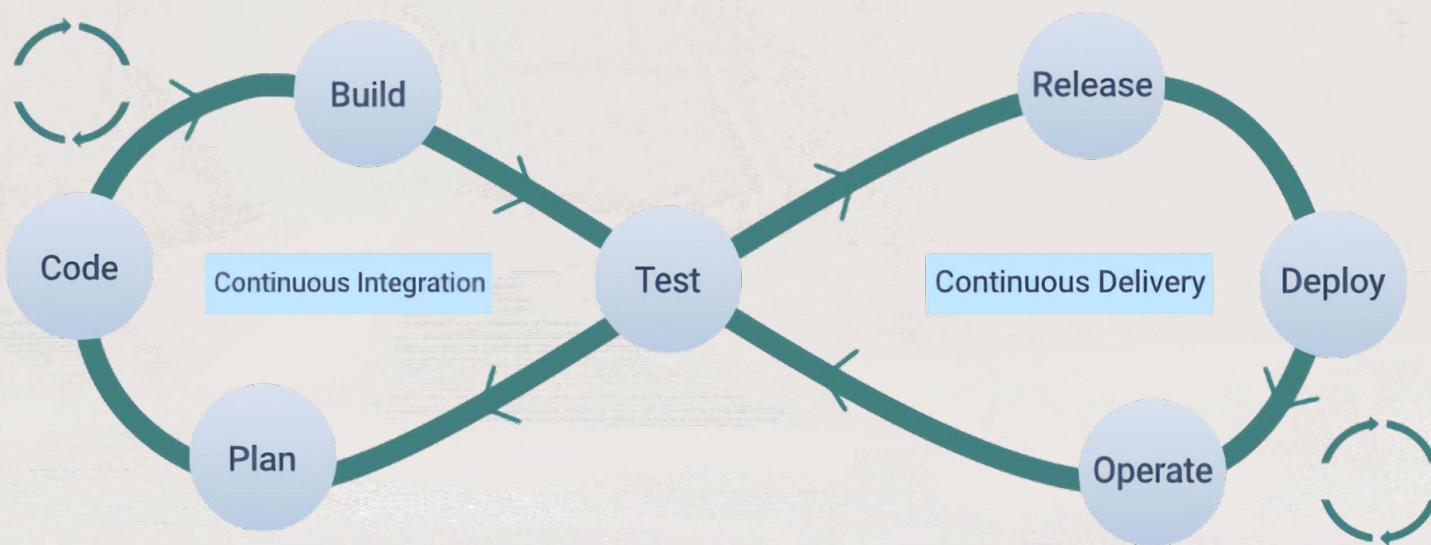
Web Developement	Flask Server	Database	Data Augmentation process
			
<ul style="list-style-type: none">• Visual Mockup• Web Responsive• JavaScript Integration	<ul style="list-style-type: none">• Interactions Server/Client• Request management• Routes Implementation	<ul style="list-style-type: none">• Data Modeling• Database Design• Data Update• Security Implementation	<ul style="list-style-type: none">• Requirement Analysis• Algorithmic conception• Post-development adjustments



Management :

CI/CD

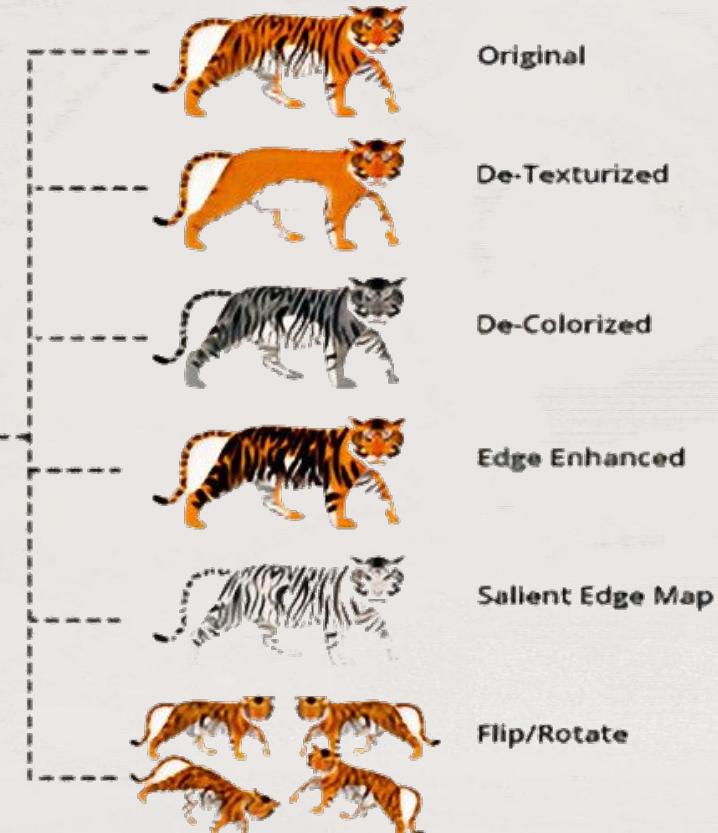
(Continuous Integration/Continuous Delivery)



Example :



----- Data Augmentation -----



Problems :

- **Web-Responsive Issues :**

- Various devices
- Different screen sizes



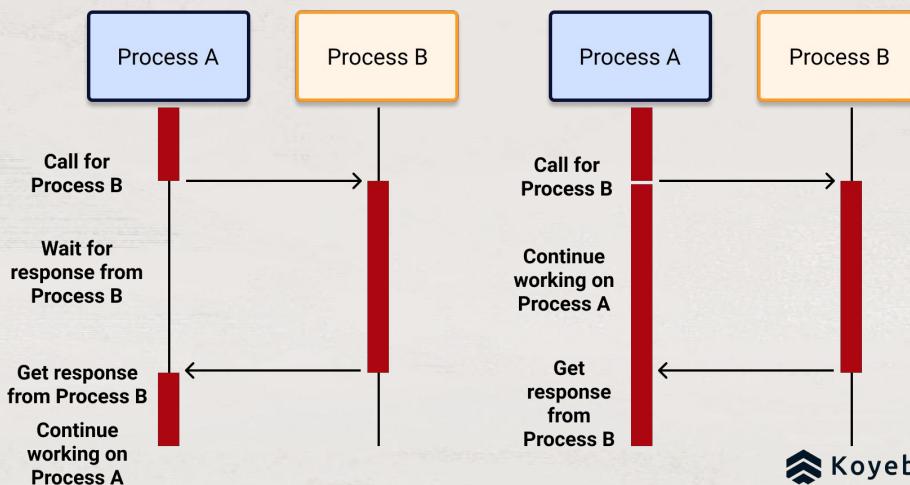
Responsive Web Design

Mobile First Web Design



Synchronous Processing

Asynchronous Processing



- **Server-Client interactions from synchronous to asynchronous :**

- Users continue interacting
- Reduce wait times



Next Steps :

I) Add Iterative Augmentation :

For Instance :

- Apply rotation to every images before augmentation
- Every data with brightness increased must be followed by a noise increasing

II) Expand the scope of the domains treated :

- Text
- Audio, Video
- Medical data



Next steps :

III) Design Improvements :

- Improve user Interface and experience
- Intuitive interactions to efficiently utilize the application

IV) Integrate a GAN :

- > create our self GAN with the python's libraries **PyTorch** and **TensorFlow**.
- > assess the quality of the augmented dataset

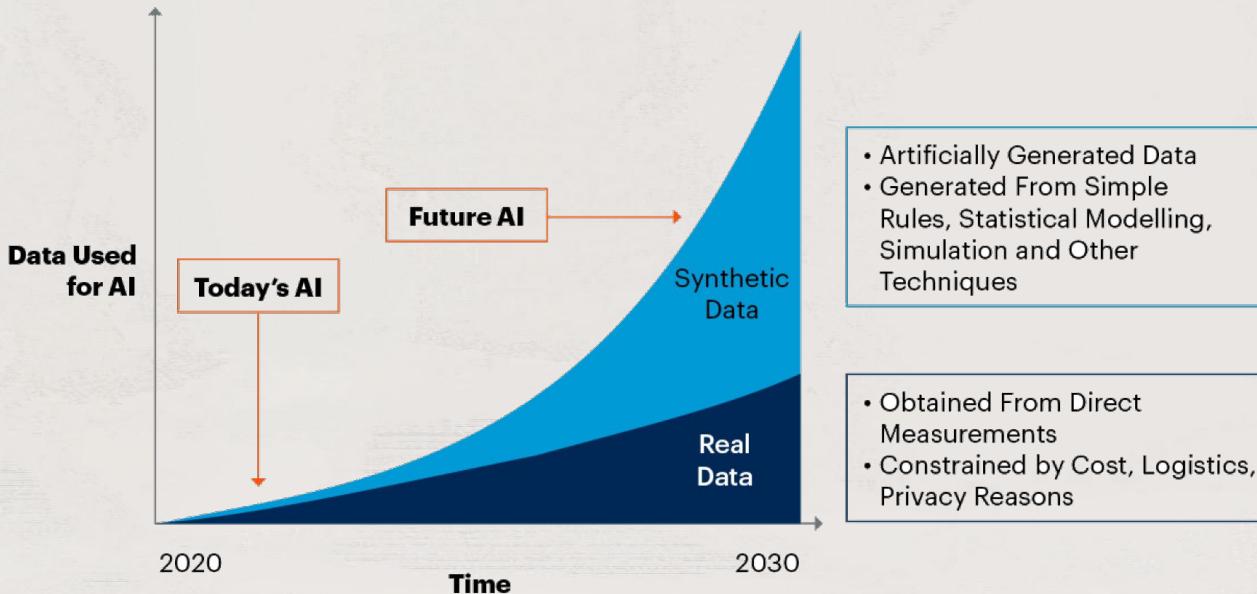


Conclusion



Speed Up collect of datas

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



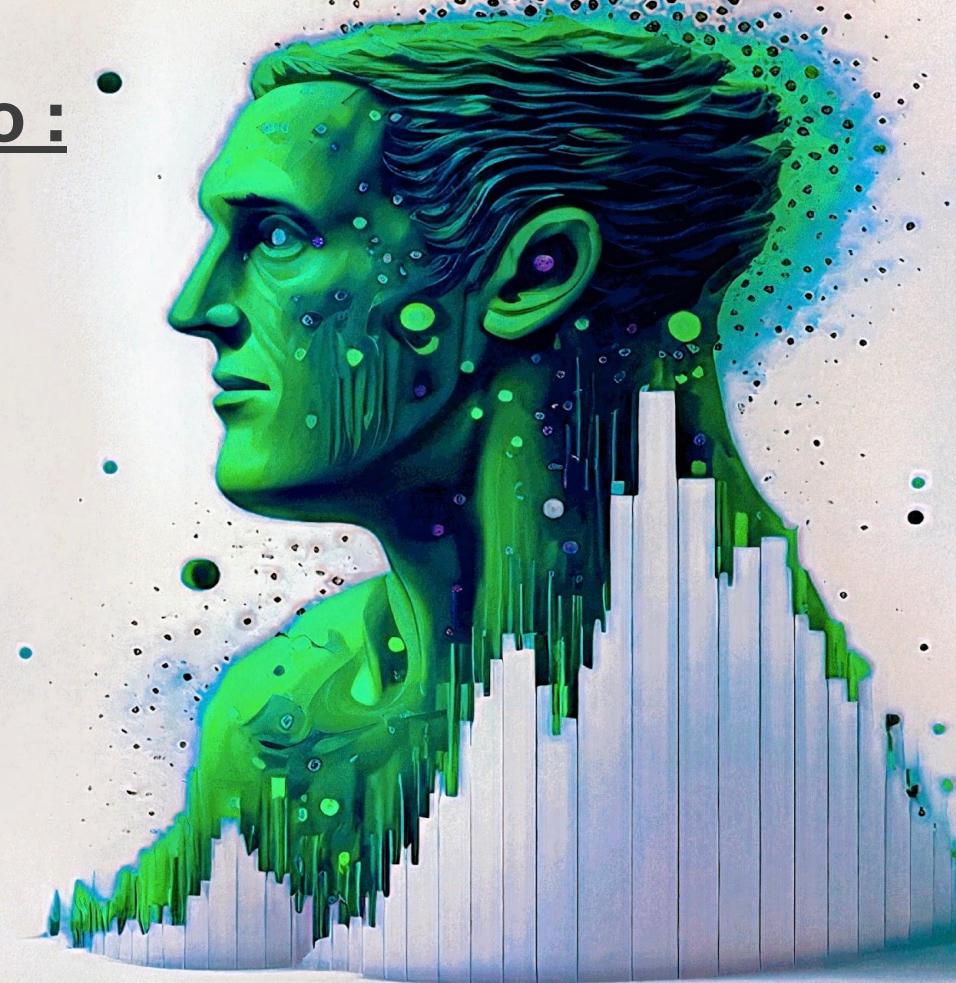
Source: Gartner
750175_C



Gartner

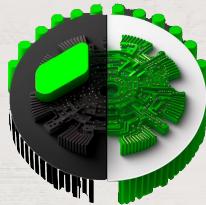
DataFlow AI is committed to :

- Reduce overfitting with data augmentation
- Improve model's robustness
- Speed up data's collect
- Ethical use of sensitive datas



Thanks for Listening

The L3Q1 Group



Join the future of the Big Data with DataFlow AI

Slides Annexes

Our Application :

Image Gallery

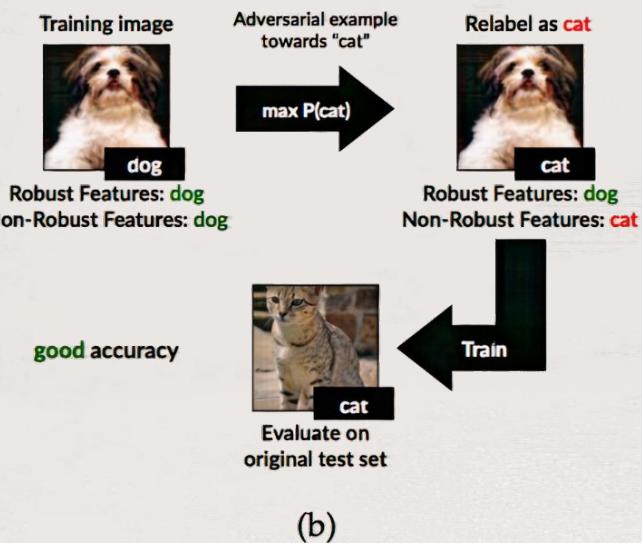
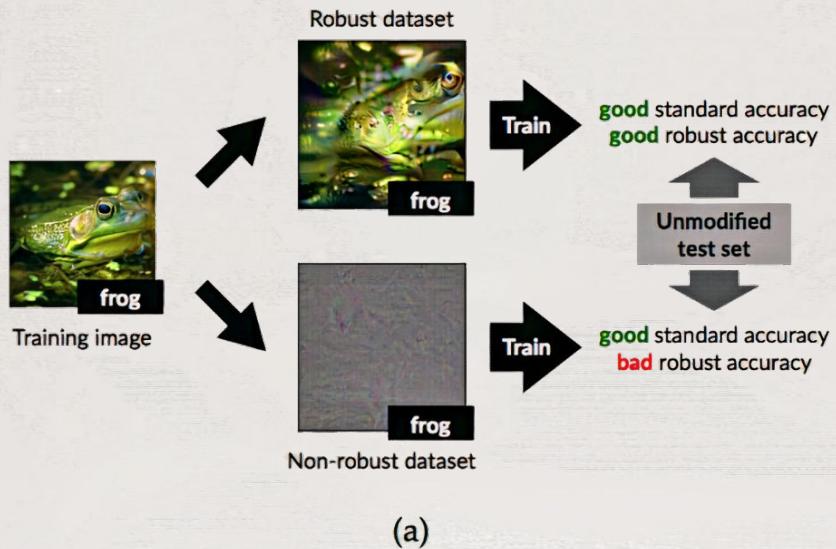
[-> Retour à la page précédente](#)

descartes.



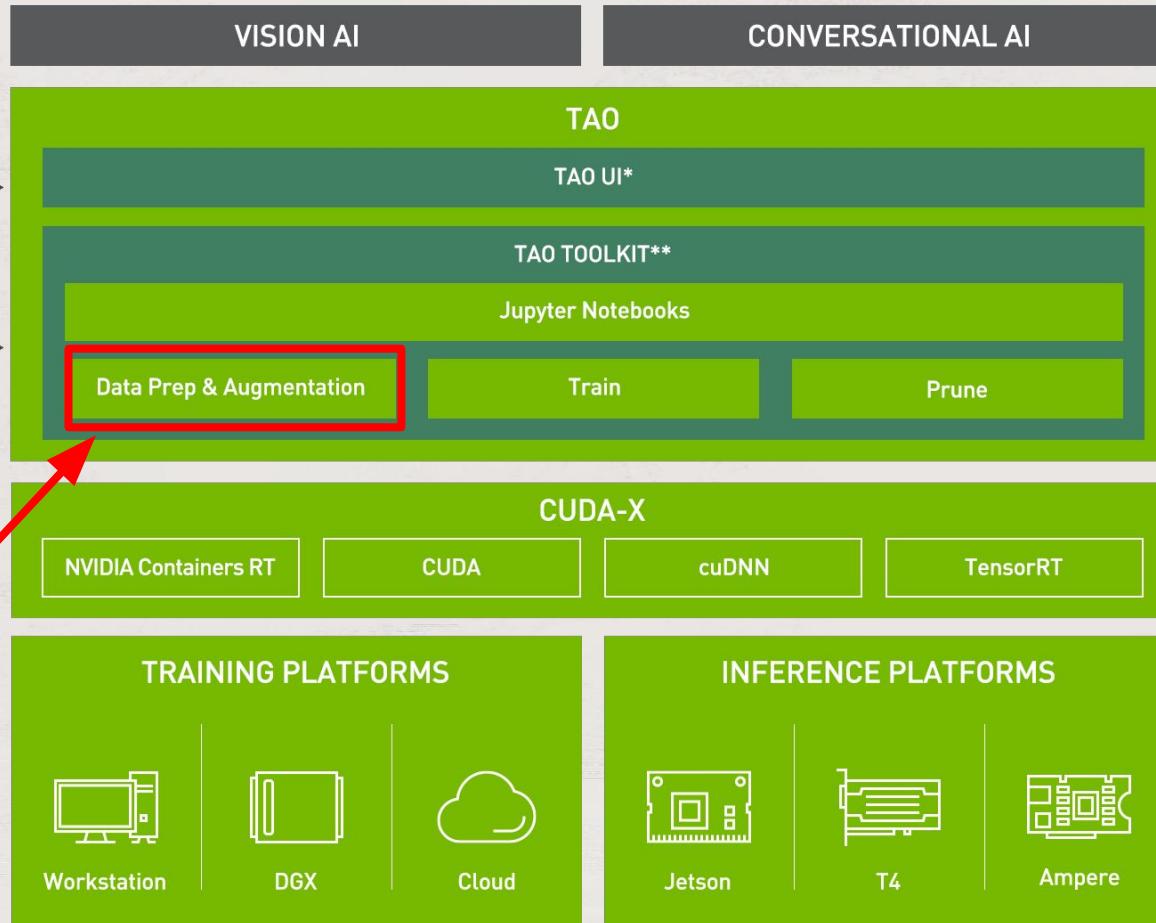
We plan to add lot of settings missing at the moment

Improve Model's Robustness





TAO

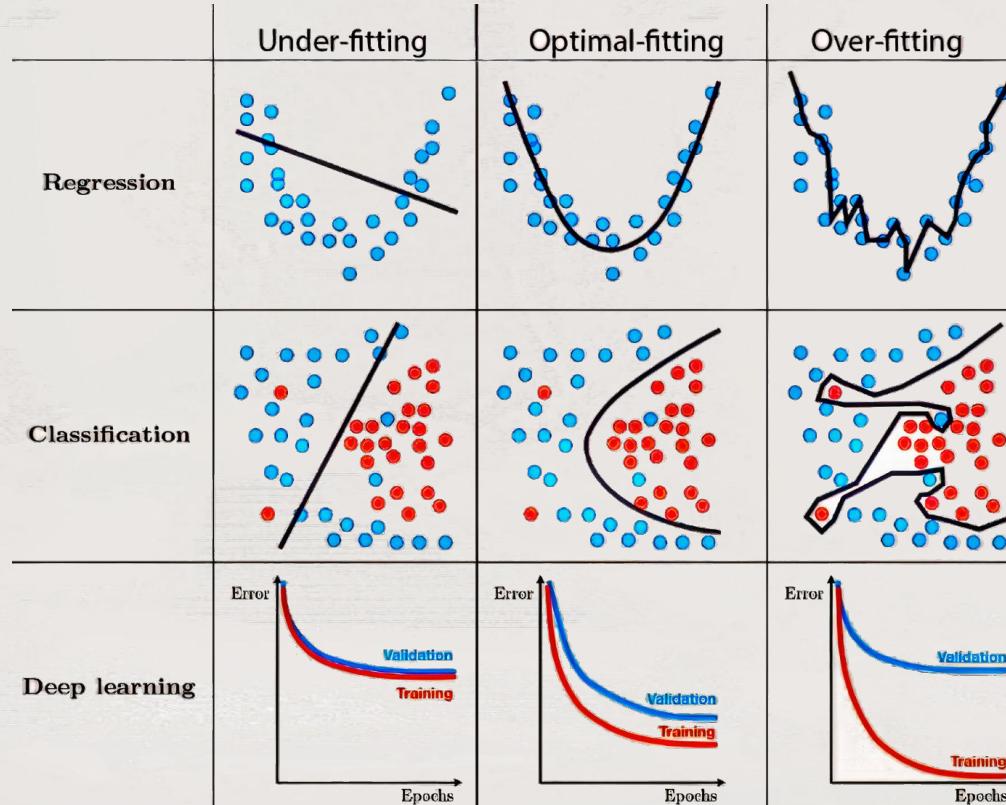


* Coming Soon

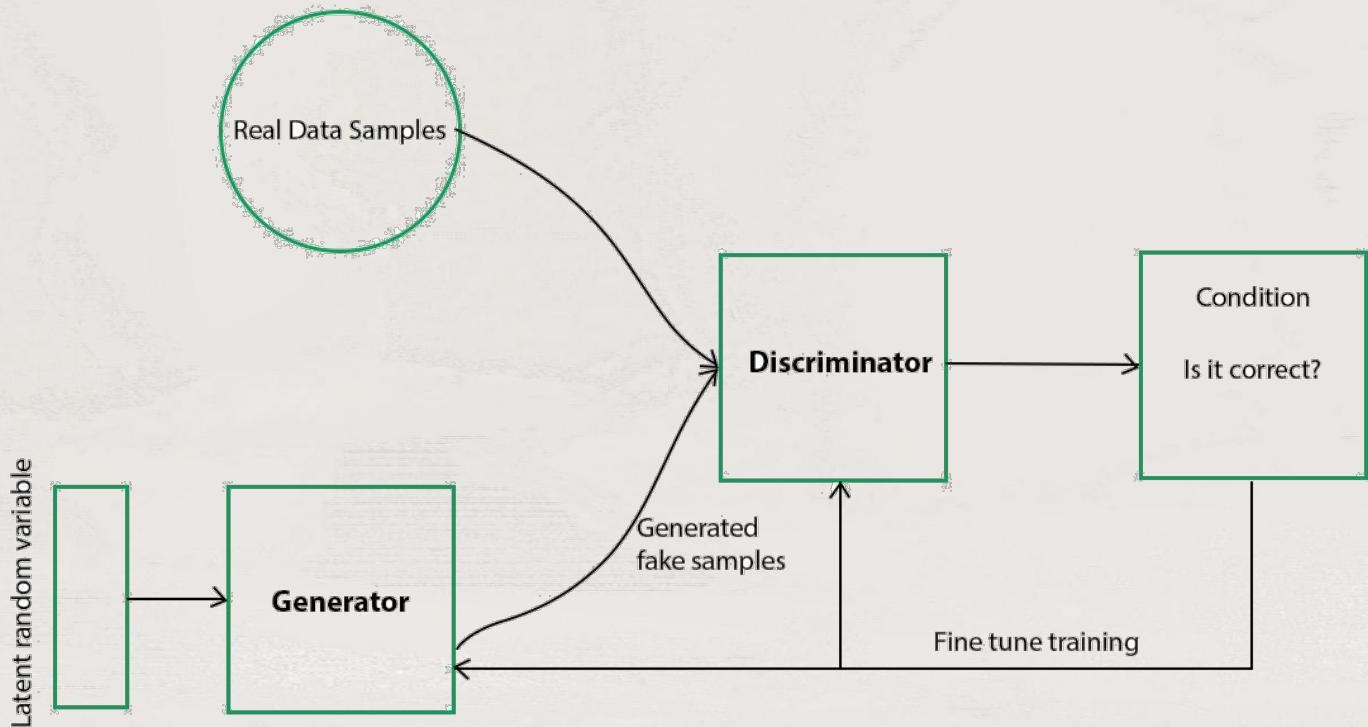
** Formerly Transfer Learning Toolkit

<https://docs.nvidia.com/launchpad/ai/tao-ds/latest/tao-ds-tao-toolkit-overview.html>

Reduce Overfitting :



GAN : generative adversarial network



Effect of settings :

- **Rotation and inverted axes** : Recognize objects from different angles and orientations
- **Brightness** : Recognize objects under different lighting conditions
- **Noise** : Simulates real-world imperfections
- **Zoom** : Recognize objects at different scales
- **Darkness** : Recognize objects in low-light conditions.
- **Distortion** : Recognize objects in images captured with different lenses or from different perspectives

Examples :



Original Image

Data Augmentation



Augmented Images

→ Data Augmentation Impacts Performance Of Image Classification

Task Repartition / Management :

