

# Machine Learning presentation

---

07-09-2017

# Table of contents

---

- ▶ Introduction to Deep Learning
- ▶ Text analysis on CNAF ticketing system
- ▶ Deep Learning School
- ▶ Sophia Conf - Artificial Intelligence

# Videos in French

---

▶ Intro

- <https://www.youtube.com/watch?v=c7mIXLJESOk&feature=youtu.be>

▶ Deep Learning for CNAF

- <https://www.youtube.com/watch?v=1DZkdrGwSSE&feature=youtu.be>

▶ Feedback Deep Learning school

- <https://youtu.be/vjtVkDFVbNE>

▶ Feedback Sophia conf – AI

- <https://www.youtube.com/watch?v=Ad77CAYFr2U&feature=youtu.be>

1

# Introduction to Deep Learning

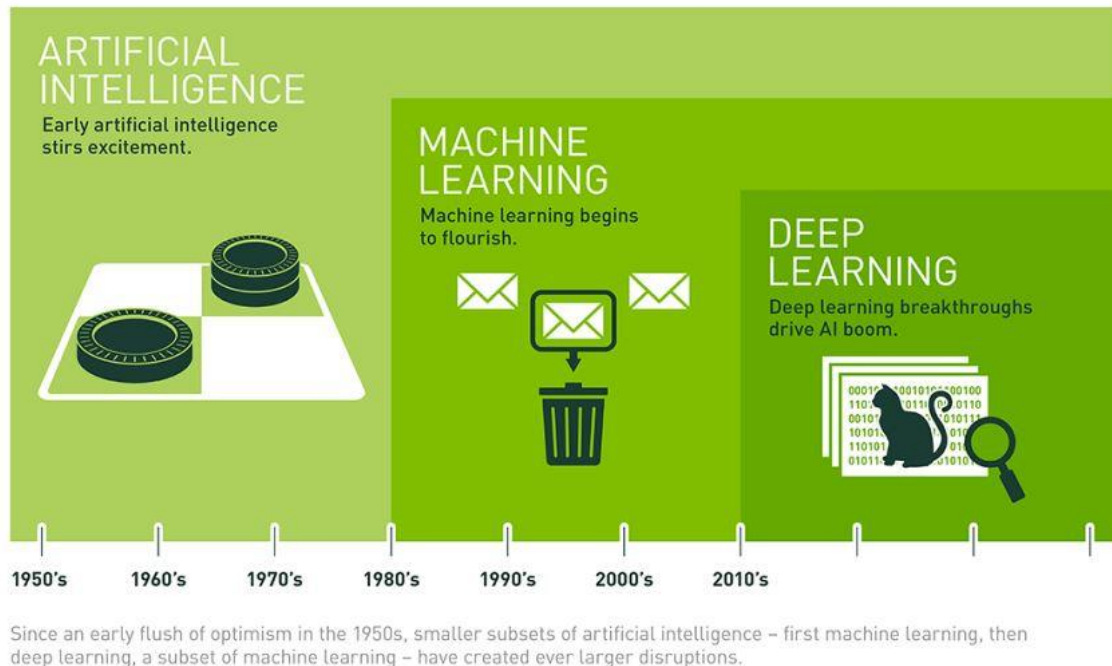
# Deep learning ?

## Part of Artificial Intelligence

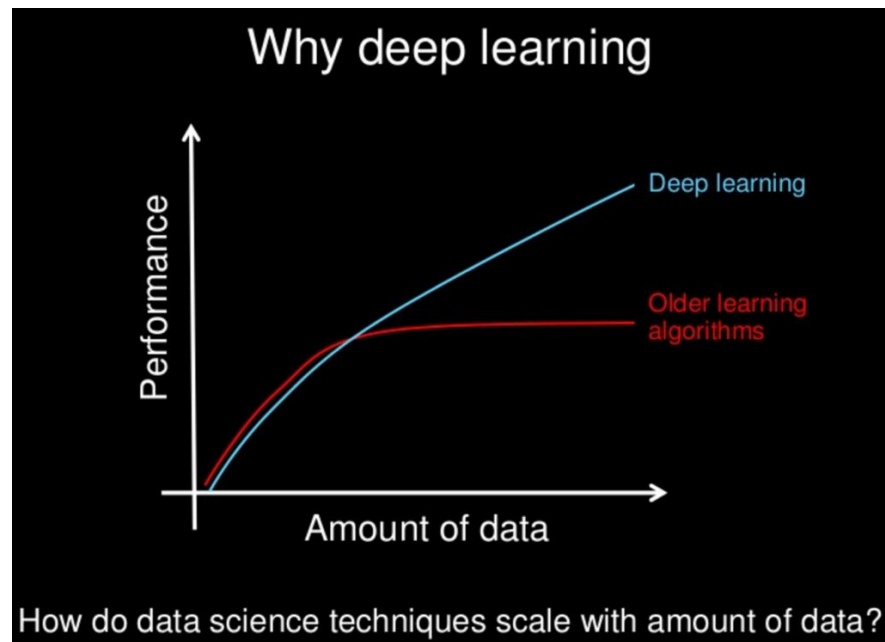
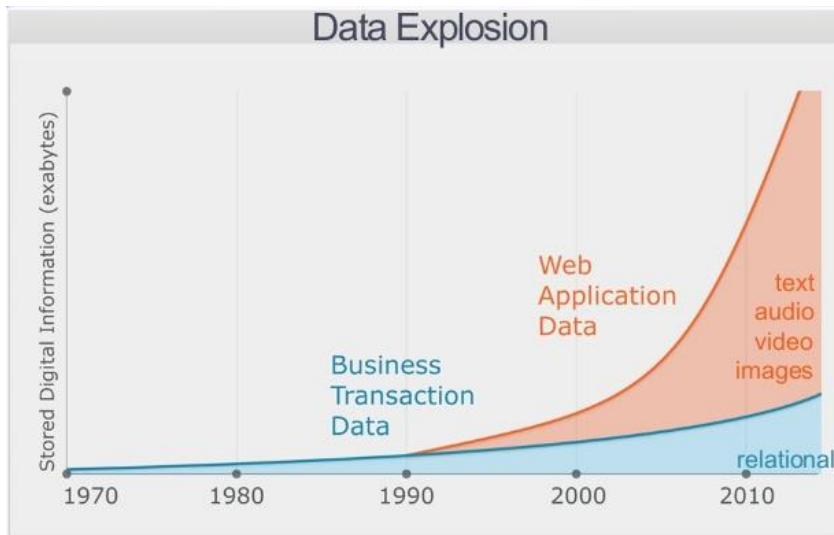
**Artificial Intelligence** – Computers with the ability to reason as humans

**Machine Learning** – Computers with the ability to learn without being explicitly programmed

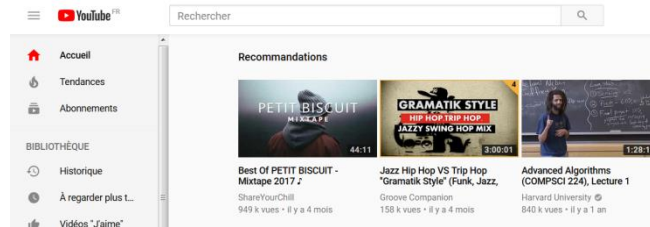
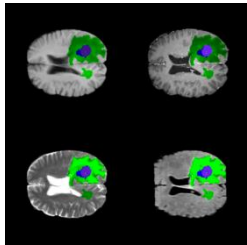
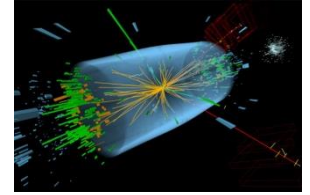
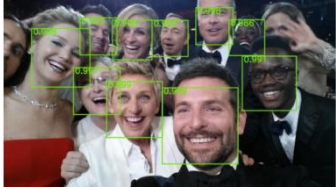
**Deep Learning** – Network capable of adapting itself to new data



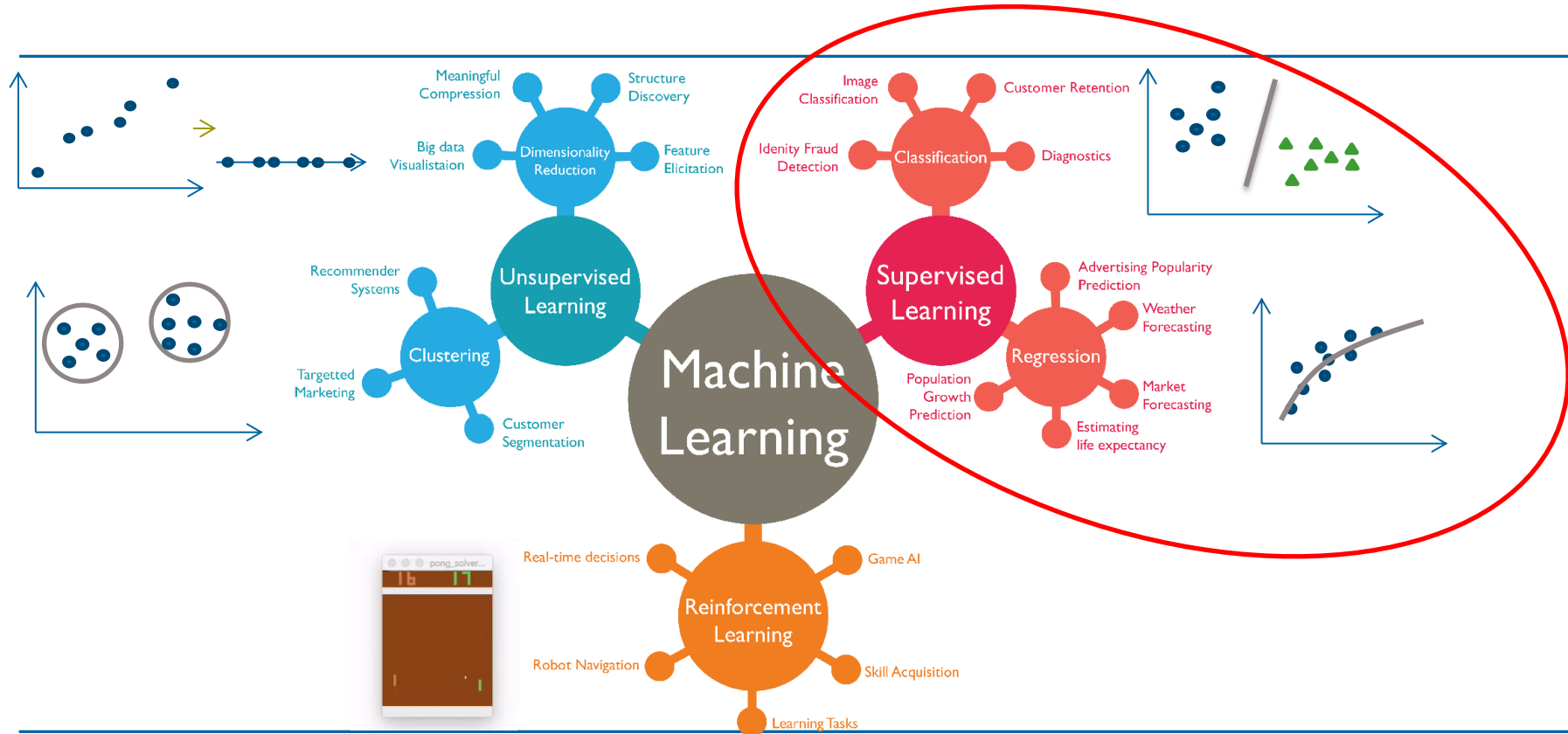
# Why deep learning ?



# Applications



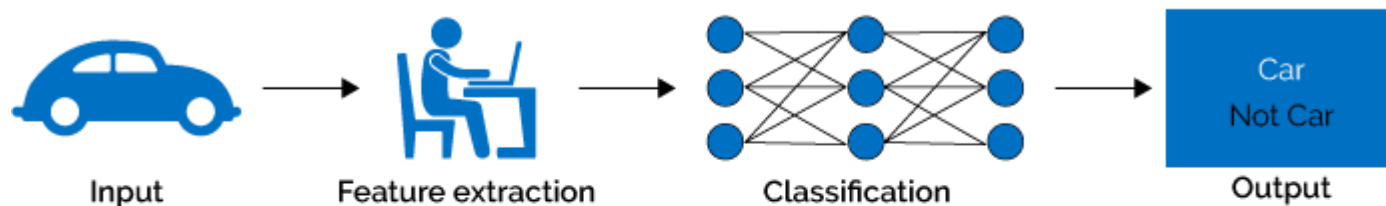
# Machine learning basics



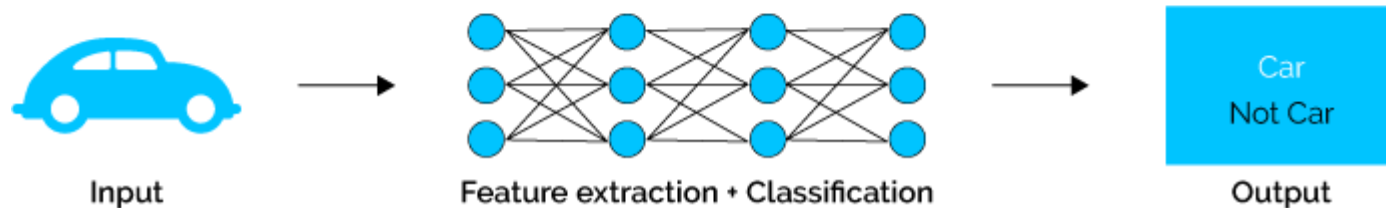


# Machine Learning vs Deep Learning

## Machine Learning



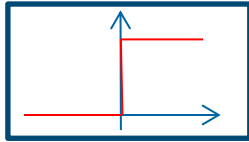
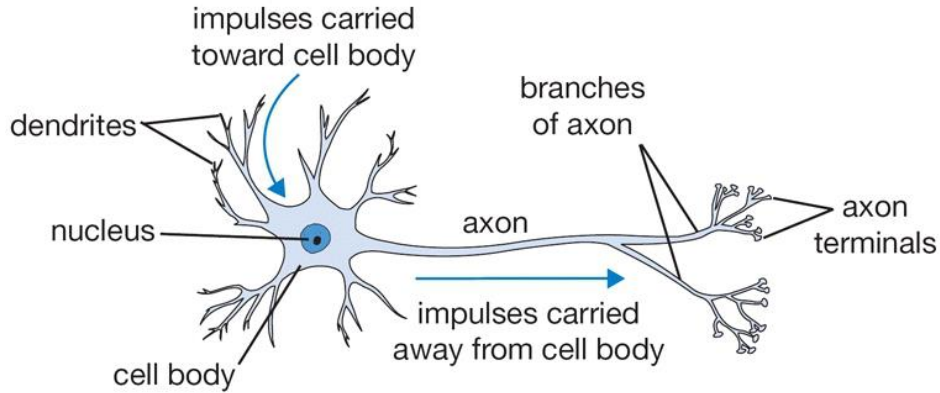
## Deep Learning



# Basics in deep learning

## Neuron, called Perceptron

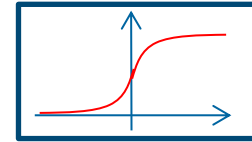
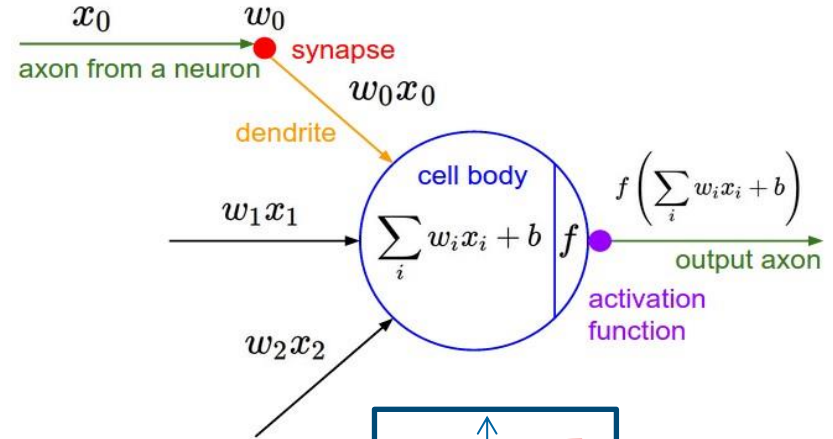
*Biological neuron*



*cell body activation*



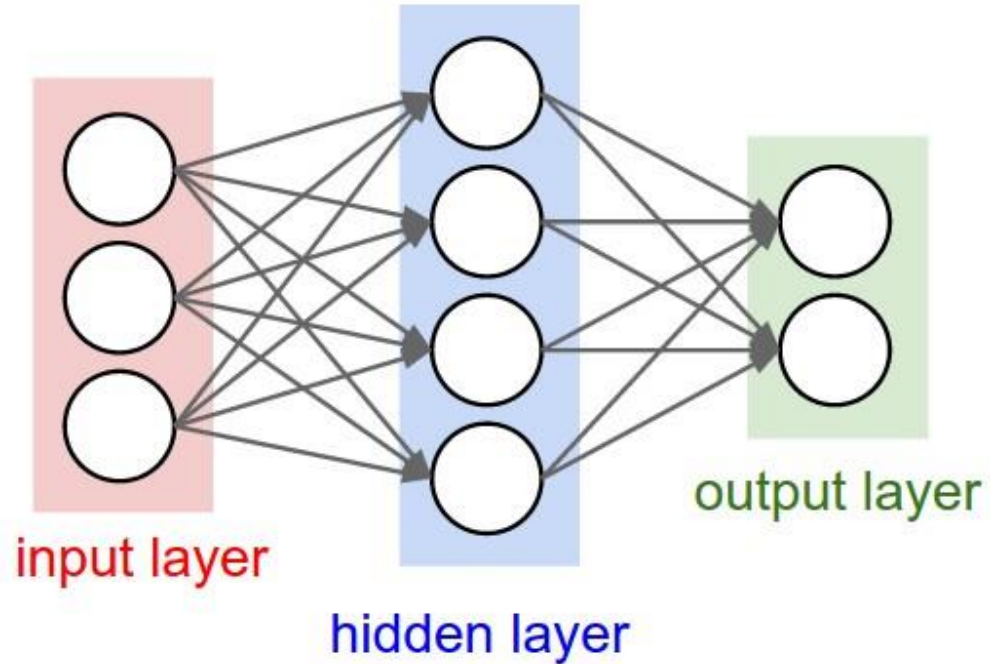
*Artificial neuron*



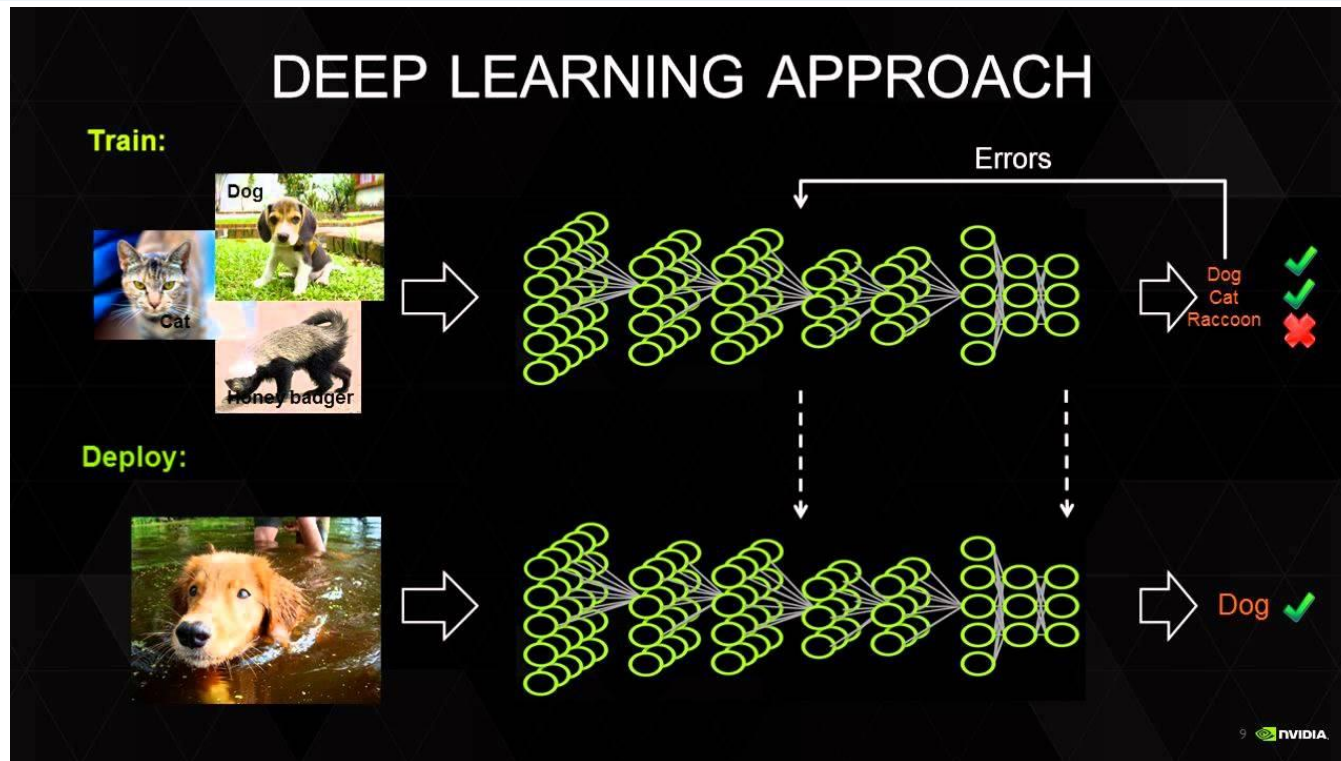
*cell body activation*

# Neural network

---



# Deep learning



2

Text analysis on CNAF  
ticketing system

# Ticketing system

ALLOCATION  
FRANCE 06

OUTILS

DONNÉES

ASSOCIÉS

Pièces jointes (0)

AGIR

Clôture technique

Service Mgt. > Référentiel Actifs Niveaux de service Sollicitation Problème FAQ / Plan d'action Changement

TRAITEMENT SOLLICITATION

Informations

N° sollicitation

Priorité PO 0 j et 00:30

Date engage. 04/09/2017 08:19:51

Statut Ouvert

Etat

Equipe Equipe CDR Contentieux

Intervenant

Equipe précédente

Seuil d'alerte

Description

Demander \* ZERROUKI, Said, 06E

Equipe utilisateur \* DNE\_PPS\_Qualification\_Authentification

Type \* Demande métier

Nature \* Fonctionnel

Service \*

Catégorie \*

Desc. courte \* Evolution technique

Macro-processus

Processus

Description \*

Bonjour,

Pouvez vous ajouter un bouton "Valider" sur la page d'authentification du portail partenaires.

Merci

Diagnostic

Impact \*

Livraison

Version

Composant

Sécurité \* Non

Cell. Crise \* Non

Aide au diag.

Résolution

Réponse interne

Disponibilité du service

Prod. prestations familiales \* Oui

Accueil Physique \* Oui

Vérification comptable \* Oui

Plate-forme téléphonique \* Oui

Action Sociale (SIAS) \* Oui

Gestion CORALI \* Oui

Gestion GRH \* Oui

Gestion MAGIC \* Oui

Date de début

Date de fin

Localisation

Origine

Site(s) impacté(s)

Siège \* Non

Toutes les antennes \* Non

Antenne(s) impactée(s)

Pièces jointes

Impact

Financier

Volume

# Problem identified

---

- ▶ 300 000 tickets on 5 years
- ▶ Multiple duplicates and not always identified or sometimes late in the process  
*Solution → SIMILARITY PREDICTION*
- ▶ Process to redirect ticket to team able to solve the ticket.
- ▶ Redirection is not always easy and some tickets are long to solve because of it.  
*Solution → TEAM PREDICTION*

The purpose of every solution suggested is to help decision

---

# Using only text of a ticket

---

- ▶ One challenge → use only text
- ▶ Makes it reusable for other applications at CNAF
- ▶ Another approach in progress which is not presented here uses also categorical variables
- ▶ Scientific approach:
  - My method using Deep Learning will be compared with a classical approach in Machine learning (*and Natural Language Processing*)



# Solutions proposed

Using only text

Team  
prediction

Text



Deep  
learning



Team

Similarity  
prediction

Text 1



Text 2

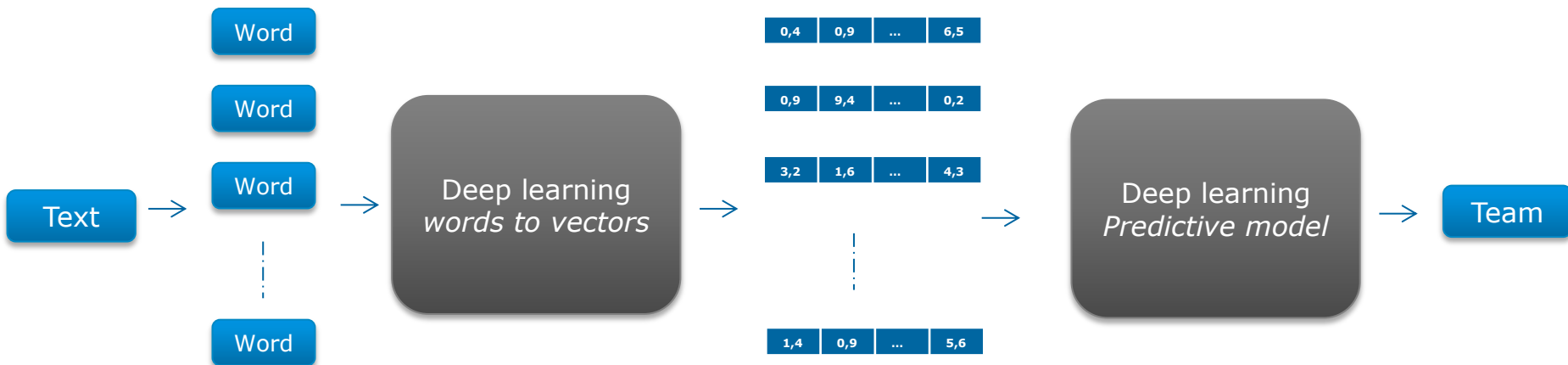


Deep  
learning



Similar ?

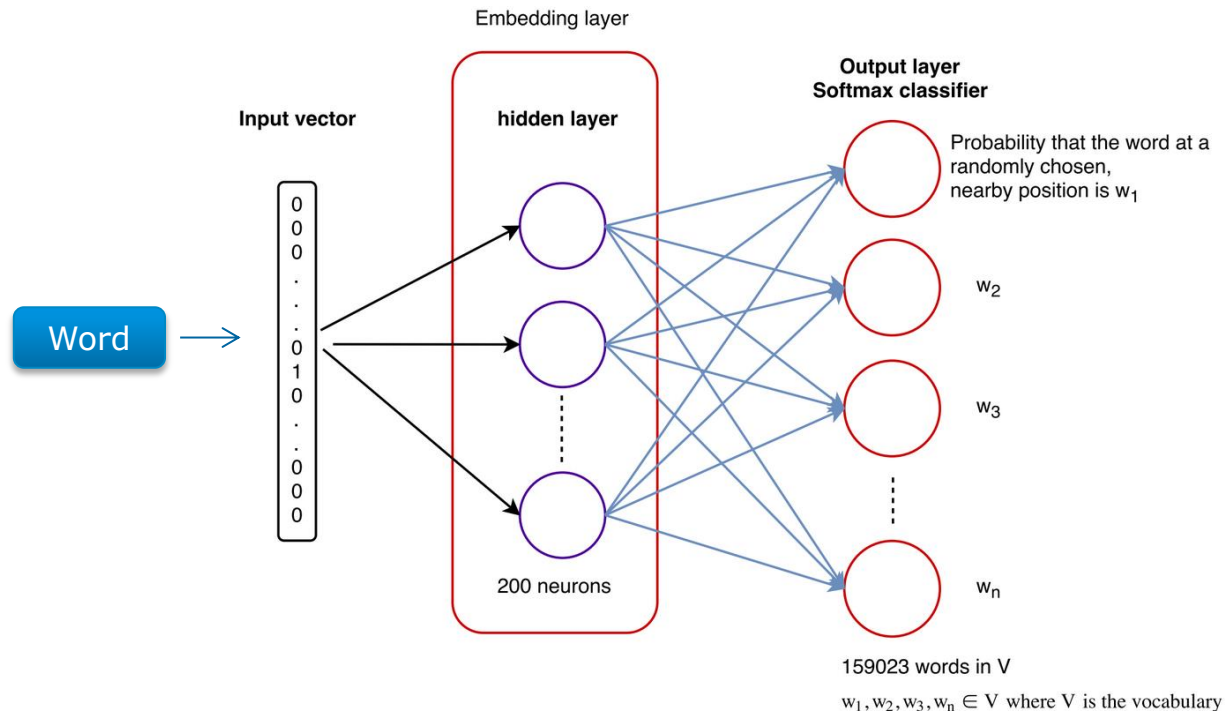
# Deep learning solution proposed



# From words to vectors

## Embedding model (Skip gram)

**NOT DETAILED IN THIS  
PRESENTATION**

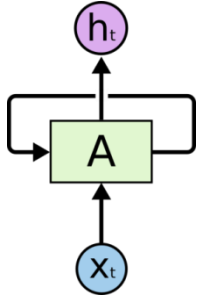


- Fake task: predict surrounding words of the input
- hidden layer is the representation by a vector of reals of the input word

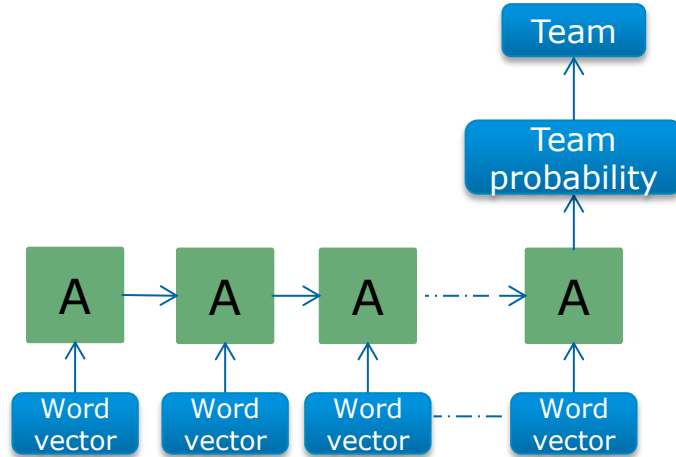
# Predictive model

## Recurrent Neural Network (RNN) and bidirectional RNN

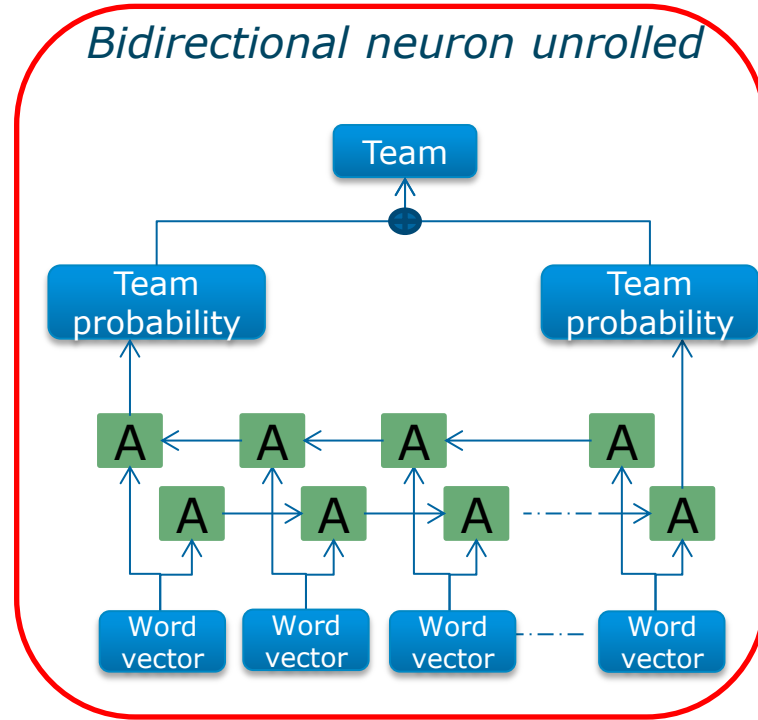
*Recurrent neuron*



*Recurrent neuron unrolled*

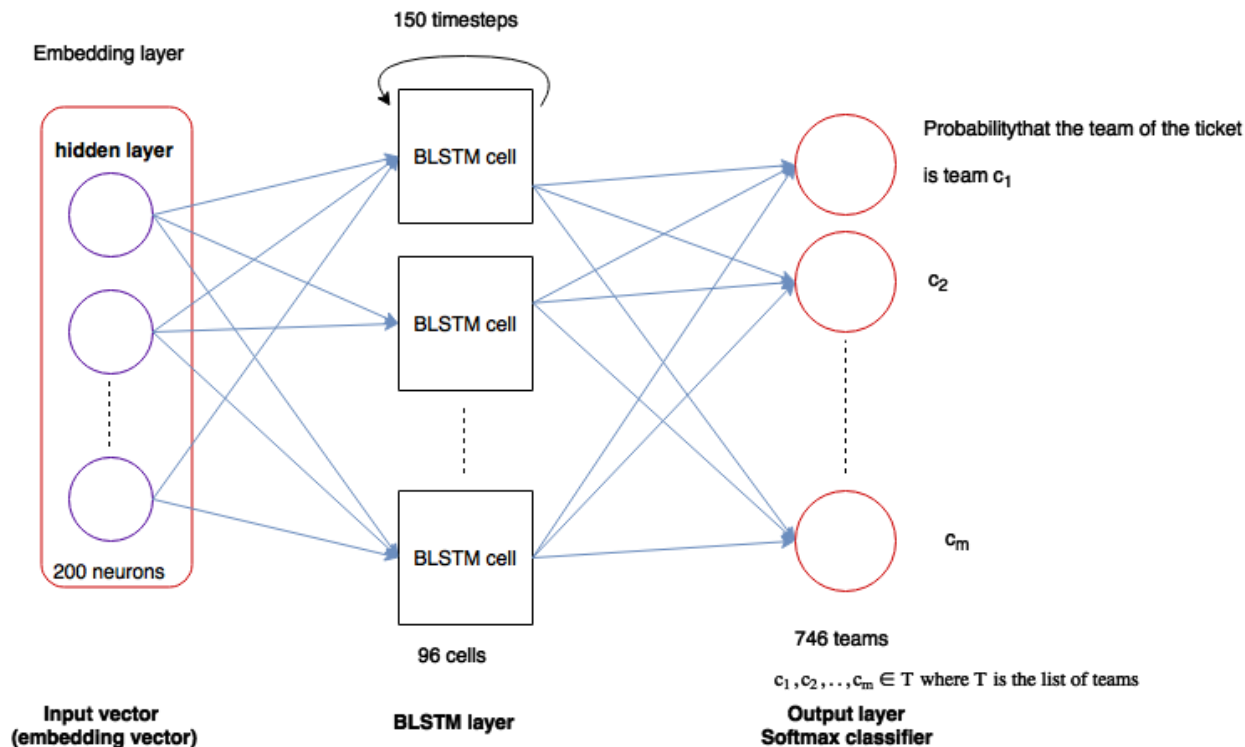


*Bidirectional neuron unrolled*



# Team prediction

**NOT DETAILED IN THIS  
PRESENTATION**



# Results

---

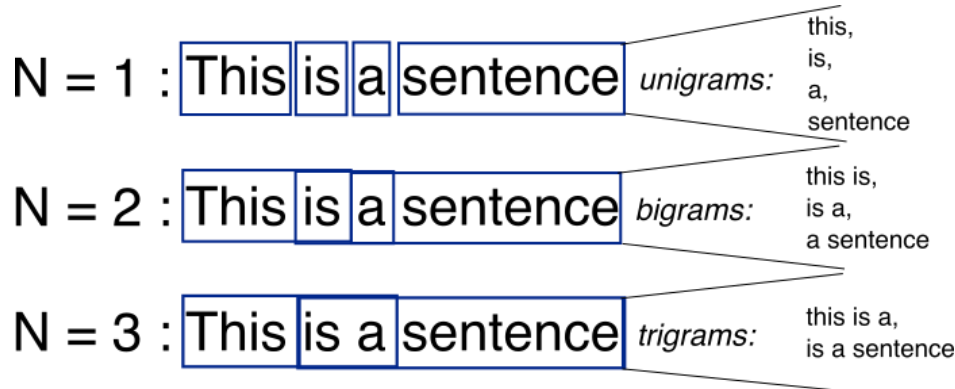
- ▶ +700 teams

Model	Test accuracy (%)
Hazard following distribution	1.21
Most represented team	4.54
Traditional approach (no deep learning)	23.26
Proposed approach	47.12

- ▶ Good results and research project to be continued...

# Traditional method ?

NOT DETAILED IN THIS PRESENTATION



$$\text{tf-idf}_{(t,d)} = \text{tf}_{(t,d)} \times \text{idf}_{(t)}$$

t = term

d = document

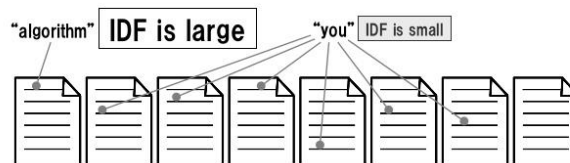
Term-Frequency Metric	Term(1)	Term(2)	Term(3)	Term(4)	Term(5)	Term(6)	Term(n)
Document(1)	225	300	0	0	0	25	0
Document(2)	78	87	0	92	0	175	0
Document(3)	58	137	0	0	237	0	21
Document(4)	0	12	101	0	0	0	0
Document(5)	3	15	0	24	0	48	87
Document(6)	0	0	71	0	0	0	0
Document(n)	109	0	901	221	331	441	551

## Inverse Document Frequency (IDF)

Give **more weight** to a term occurring in **less documents**

$$\text{IDF}(t) = \log \frac{|D|}{df(t)}$$

$t$  : Term  
 $df(t)$  : Document frequency of  $t$   
 $|D|$  : Number of documents in  $D$



# Project conclusion

---

- ▶ Github in progress (soon available): <https://github.com/turpaultn/CnafSAXO>
- ▶ Research report available (deep learning basic knowledge required)  
send a mail to: [turpaultn@gmail.com](mailto:turpaultn@gmail.com)



3

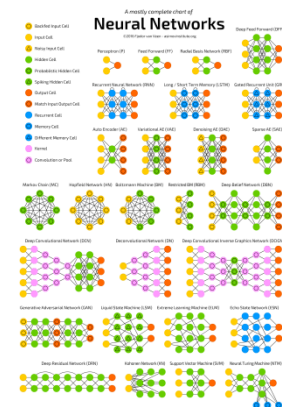
Deep Learning School  
June, 12-15

---

# Lundi 12 Juin

## Introduction and getting started workshop

- ▶ Presented by Stéphane CANU and Frédéric PRECIOSO
- ▶ Mathematical basics
- ▶ Deep learning concepts with presentation of Convolutional Neural Network (CNN)
- ▶ Workshop already using a neural network
- ▶ Good introduction and gives basics to rapidly come on deep learning



### Example: Handwritten digit recognition

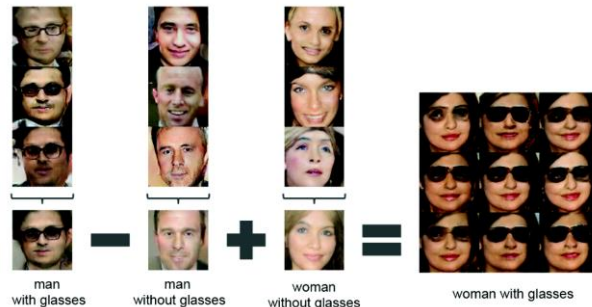
- The goal: SW to recognize the digit in each image (**Classifier**)
- Source: "MNIST database of handwritten digits", 60,000 examples
- Typical human error: 2.5%. Common confusion between {2, 7}, {4, 9}



# Mardi 13 Juin

## Deep architectures

- ▶ Soufiane BELHARBI & Mélanie DUCOFFE
- ▶ Presentation of some of their thesis work
- ▶ Presentation of GANs and Autoencoders with applications
- ▶ From basics to research applications
- ▶ Gives better understanding of some architectures with some tricky points
- ▶ Lab on medical data segmentation with GAN

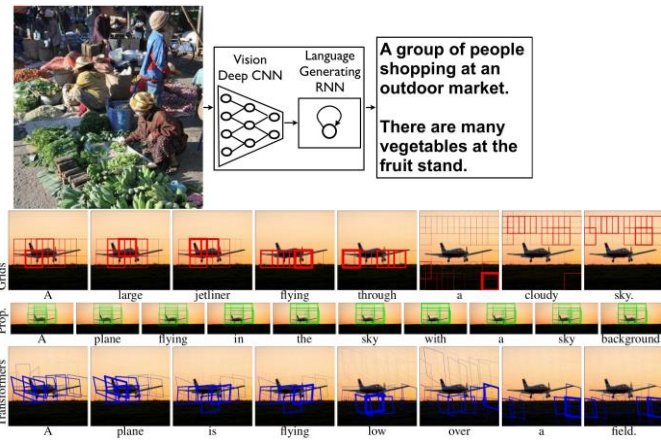
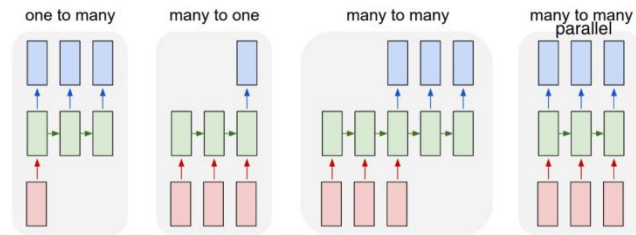


Fictional celebrity faces generated by a variational autoencoder.

# Mercredi 14 Juin

## Dealing with time

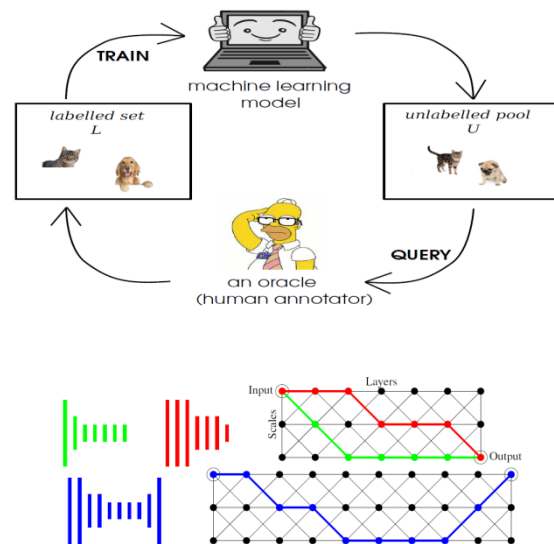
- ▶ Rémi CADENE and Jakob VERBEEK
- ▶ Introduction to Recurrent Neural Network
- ▶ Annotate content of an image
- ▶ Using CNN and RNN
- ▶ Presentations which represent research work, good vulgarization for beginners and some details for others
- ▶ Lab on modelling time series with RNN



# Jeudi 15 Juin

## Building and training a Deep Network

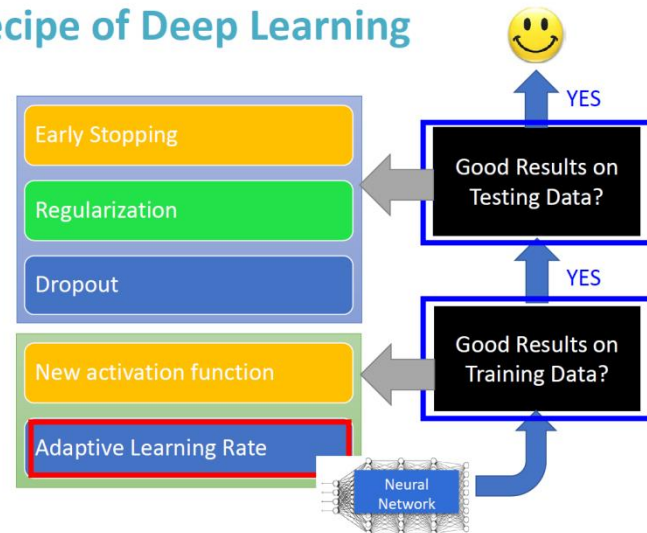
- ▶ Frédéric Précioso, Mélanie Ducoffe and Jakob VERBEEK
- ▶ Active learning presentation
- ▶ Explanation of research experiments
- ▶ Optimization process with techniques and tricks
- ▶ Good conclusion explaining some difficulties to train deep networks.
- ▶ It needs time to find good parameters on the optimization process



# Conclusion and resources

- ▶ <http://univ-cotedazur.fr/events/deep-learning-school>
- ▶ Mostly industrial public
- ▶ Good initiation for beginners
- ▶ And labs designed for them
- ▶ Good levels of details for others
- ▶ Big success for this 1<sup>st</sup> Deep learning school in France
- ▶ To be reproduced next year and extended to other cities

## Recipe of Deep Learning





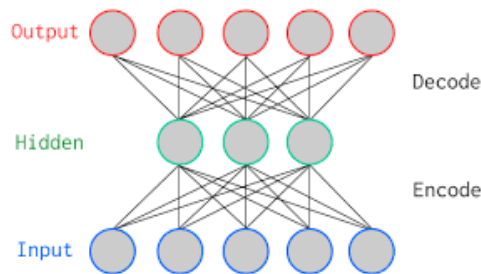
Sophia Conf – July, 5  
Artificial Intelligence

# Anomaly detection

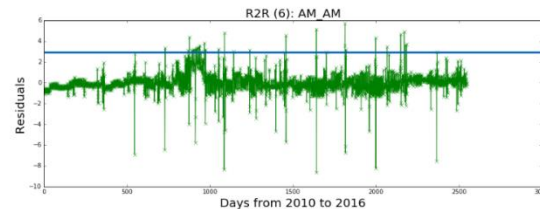
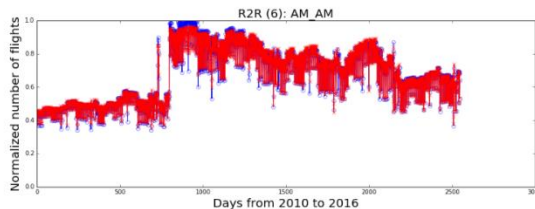
Asmaa FILLATRE - Amadeus

► <http://www.telecom-valley.fr/wp-content/uploads/2017/05/FILLATRE.pdf>

► **Autoencoder** for anomaly detection



► **Unsupervised learning**

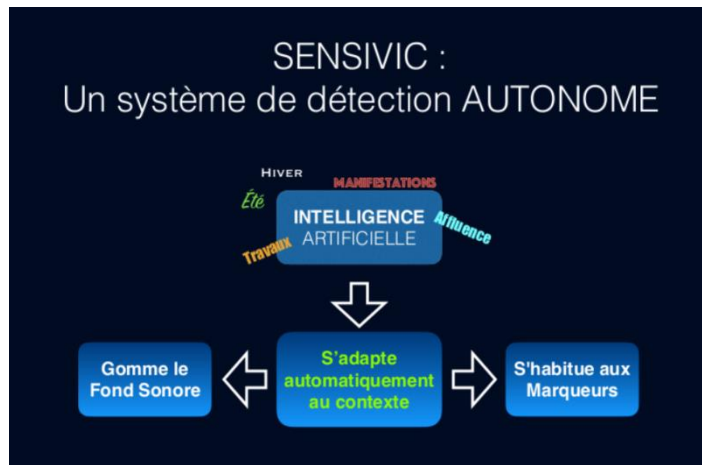




# Anomaly detection in sound for security

Jean DEMARTINI – USS SENSIVIC

- ▶ [http://www.telecom-valley.fr/wp-content/uploads/2017/05/DEMARTINI-20170705\\_Sophia\\_Conf\\_DAAS\\_IA\\_pratique.pdf](http://www.telecom-valley.fr/wp-content/uploads/2017/05/DEMARTINI-20170705_Sophia_Conf_DAAS_IA_pratique.pdf)
- ▶ Add microphone with camera in security to detect anomalies camera can not see.
- ▶ No algorithm given

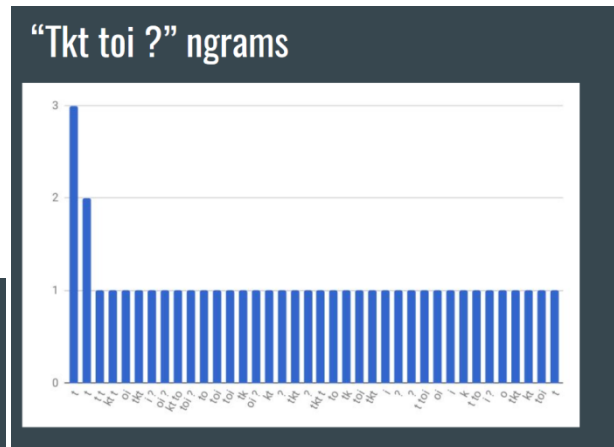


# Language identification in (very) short texts

# Mathieu LACAGE - Alcmeón

- ▶ <http://www.telecom-valley.fr/wp-content/uploads/2017/05/LACAGE.pdf>
- ▶ **Classification** problem using ngrams
- ▶ No deep learning
- ▶ Learning by **examples** available on github

English	87300
Dutch	87541
French	87553
Italian	87753
German	87926
Spanish	88444
Russian	88845
Arabic	90000



# Photo editing with GANs

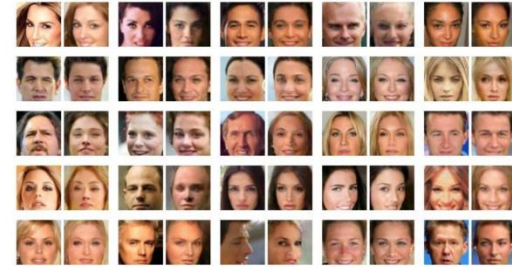
Greg HEINRICH - Nvidia

► <http://www.telecom-valley.fr/wp-content/uploads/2017/05/HEINRICH.pdf>

- GANs: Generative Adversarial Network
- Basics of the algorithm very well explained
- Famous algorithm
- Learning to generate the right data, in this case images

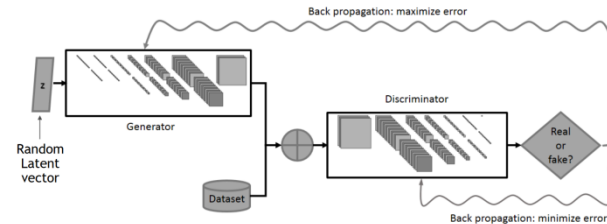
## IMAGE RECONSTRUCTIONS

Visualizing  $G(E(\text{image}))$



## THE GAN SET-UP

Connecting the Discriminator to the Generator and the Dataset

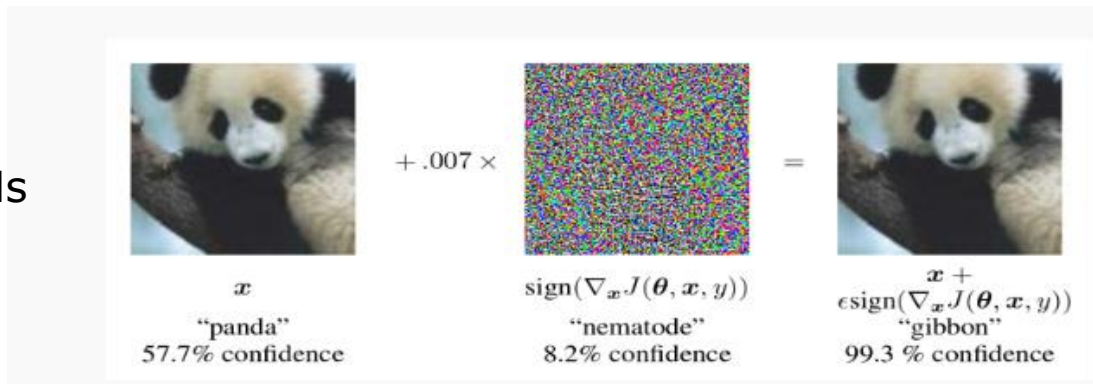


# Fooling Deep Networks

Guillaume DEBARD – Laboratoire I3S

► <http://www.telecom-valley.fr/wp-content/uploads/2017/05/DEBARD.pdf>

- More technical
- Show some limitations of GANs
- How to attack a GAN
- Very good presentation which reduces the claim of Deep learning



5

---

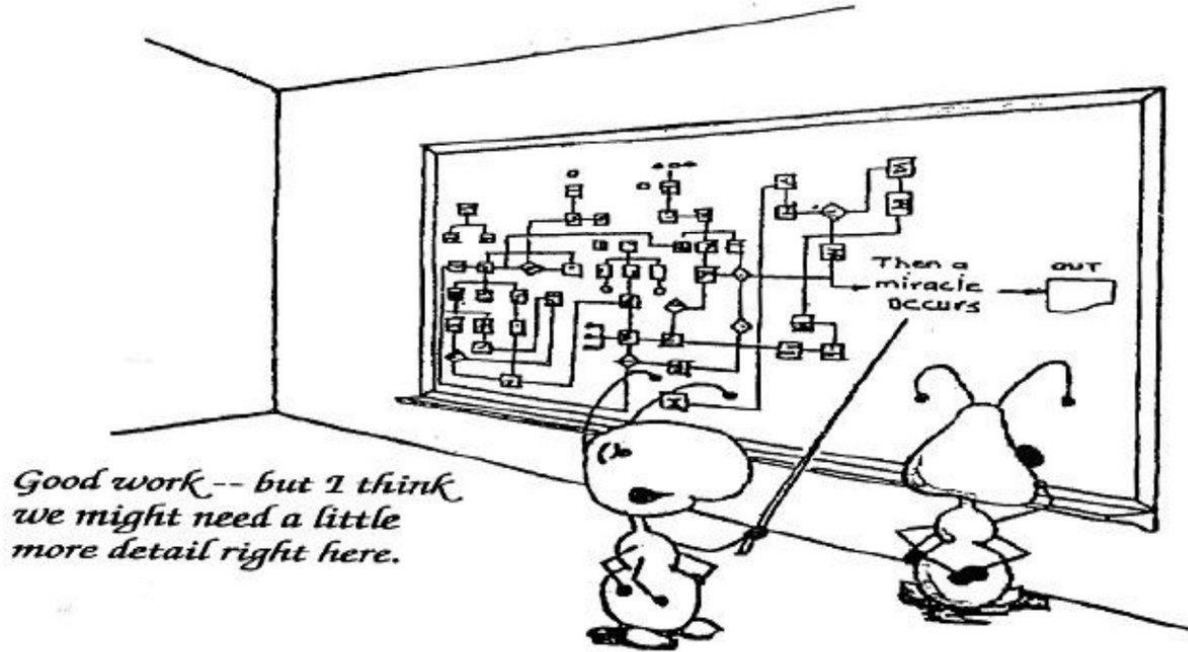
Conclusion

# Deep learning

---

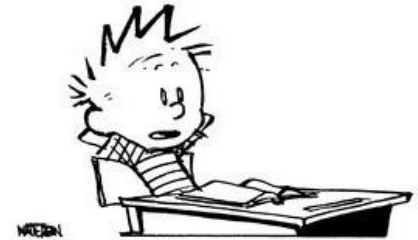
- ▶ First steps easy to implement
- ▶ Can solve new problems our outperforms previous method
- ▶ Not only for robots
- ▶ It is an **optimization process** relevant for many applications
- ▶ Why don't you try ?

# Any questions?



GIVEN THE PACE OF  
TECHNOLOGY, I PROPOSE  
WE LEAVE MATH TO THE  
MACHINES AND GO PLAY  
OUTSIDE.

/



6

References

---



# References

## Images

---

- ▶ <https://leverton.de/blog/2016/12/21/3-deep-learning-trends-2017/>
- ▶ <https://www.mongodb.com/blog/post/deep-learning-and-the-artificial-intelligence-revolution-part-2>
- ▶ <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>
- ▶ <https://www.technologyreview.com/s/535201/the-face-detection-algorithm-set-to-revolutionize-image-search/>
- ▶ <https://www.analyticsvidhya.com/blog/2017/01/introduction-to-reinforcement-learning-implementation/>
- ▶ <http://med.stanford.edu/gevaertlab.html>
- ▶ <http://cs231n.github.io/neural-networks-1/>
- ▶ <https://www.youtube.com/watch?v=4A14mOsT6vQ>
- ▶ <http://www.businesswire.com/news/home/20170118005301/en/Neurala-Announces-14-Million-Series-Bring-Deep>
- ▶ <https://www.singularityweblog.com/alphago-deepmind-intelligence/>
- ▶ <https://www.linkedin.com/pulse/text-mining-deep-learning-part-1-shamane-siriwardhana>
- ▶ <https://www.ibm.com/blogs/insights-on-business/government/watson-in-defense-not-just-for-trivia/>
- ▶ <http://www.asimovinstitute.org/neural-network-zoo/>
- ▶ <https://www.youtube.com/watch?v=XNZIN7Jh3Sg>
- ▶ <https://www.slideshare.net/RoiBlanco/big-datasantiagov2>
- ▶ <https://www.linkedin.com/pulse/term-frequency-inverse-document-sanjay-singh-1>
- ▶ <https://www.slideshare.net/MasumiShirakawa/www-48698138>
- ▶ <http://keetmalin.wixsite.com/keetmalin/single-post/2017/06/05/TF-IDF-in-the-Field-of-Information-Retrieval>

# Thanks

---

For more information please contact:

T+ 33 1 98765432

F+ 33 1 88888888

M+ 33 6 44445678

firstname.lastname@atos.net

Atos, the Atos logo, Atos Codex, Atos Consulting, Atos Worldgrid, Bull, Canopy, equensWorldline, Unify, Worldline and Zero Email are registered trademarks of the Atos group. August 2017. © 2017 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.

