

RESSOURCES DE DÉVELOPPEMENT D'UN PROJET IA

Neila Mezghani
(Hiver 2025)

Plan du cours

- Ressources informatiques
 - *Langage de programmation*
 - *L'informatique en nuage*
- Ressources bibliographiques
- Plateforme d'IA
- Étapes d'un projet IA

Plan du cours

- Ressources informatiques
 - *Langage de programmation*
 - *L'informatique en nuage*
- Ressources bibliographiques
- Plateforme d'IA
- Étapes d'un projet IA

Top 5 des langages d'IA

- Python est le leader incontesté de l'Intelligence Artificielle
- Java est un autre langage de programmation qui permet de coder l'intelligence artificielle. Cet outil est avec Python l'un des codes les plus utilisés pour l'IA. De fait, cette technologie est assez demandée dans les recrutements du secteur.
- C++ est le langage le plus rapide, sa rapidité est très appréciée pour les projets d'intelligence artificielle qui sont limités dans le temps.
- LISP est un vieux langage de programmation et des améliorations ont été apportées pour en faire un langage puissant et dynamique.
- R est l'un des langages de programmation les plus efficaces pour analyser et manipuler des données dans un but statistique. En utilisant R, on peut facilement produire de belles publications grâce à la data visualisation, avec des formules si besoin.

Python et l'IA

- Python est le langage le plus populaire dans le monde de l'intelligence artificielle.
- C'est un langage de programmation orienté objet et se veut relativement facile d'accès.
- Il est très utilisé au sein de la communauté scientifique et particulièrement dans le domaine de l'intelligence artificielle.
- Les principaux frameworks de machine learning et deep learning sont disponibles en Python.
- Python est un langage de programmation qui peut s'utiliser dans de nombreux contextes et s'adapter à tout type d'utilisation grâce à des bibliothèques spécialisées.

Python et l'IA

- Depuis 2014, l'Institute of Electrical and Electronics Engineers (IEEE), la plus grande association mondiale de professionnels techniques, s'est érigée en baromètre des langages de programmation en livrant au deuxième semestre de chaque année son classement des meilleurs langages selon divers critères :
 - popularité générale,
 - langages en forte croissance,
 - langages les plus demandés par les employeurs,
 - les meilleurs langages pour le développement de sites et applications web,

Top 10 du classement général des langages en 2021

Rank	Language	Type	Score
1	Python	🌐💻⚙️	100.0
2	Java	🌐📱💻	95.4
3	C	📱💻⚙️	94.7
4	C++	📱💻⚙️	92.4
5	JavaScript	🌐	88.1
6	C#	🌐📱💻⚙️	82.4
7	R	💻	81.7
8	Go	🌐💻	77.7
9	HTML	🌐	75.4
10	Swift	📱💻	70.4

<https://programmation.developpez.com/actu/317910/Meilleurs-langages-de-programmation-en-2021-selon-l-IEEE-Python-leader-pour-la-5e-annee-consecutive-il-s-impose-dans-tous-les-domaines-dans-lesquels-il-est-utilise-du-web-a-l-embarque/>

Top 10 des langages les plus demandés par les employeurs

Language Ranking: Jobs

Rank	Language	Type	Score
1	Python	🌐💻⚙️	100.0
2	C	📱💻⚙️	96.0
3	Java	🌐📱💻	95.9
4	JavaScript	🌐	89.6
5	C++	📱💻⚙️	88.3
6	Go	🌐💻	87.3
7	R	💻	85.7
8	HTML	🌐	81.3
9	C#	🌐📱💻⚙️	79.8
10	SQL	💻	71.9

<https://programmation.developpez.com/actu/317910/Meilleurs-langages-de-programmation-en-2021-selon-l-IEEE-Python-leader-pour-la-5e-annee-consecutive-il-s-impose-dans-tous-les-domaines-dans-lesquels-il-est-utilise-du-web-a-l-embarque/>

Librairies d'IA

- [Annoy](#), librairie extrêmement rapide implémentant la recherche des plus proches voisins
- [Caffe](#), Deep learning framework
- [Chainer](#), Framework intuitif pour les réseaux de neurones
- [neon](#), Deep Learning framework extrêmement performant
- [NuPIC](#), Plateforme d'IA implémentant les [algorithmes d'apprentissage HTM](#)
- [Shogun](#), Large Scale Machine Learning Toolbox
- [TensorFlow](#), Réseau de neurones disposant d'une API de haut niveau
- [Theano](#), Librairie d'apprentissage automatique destinée à évaluer et optimiser des expressions mathématiques
- [Torch](#), Framework d'algorithmes d'apprentissage très performant disposant de binding Python
- [Theanets](#), deep learning
- [Scikit-Learn](#), est probablement la plus populaire des librairies disponibles

Librairies d'IA

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Analyse de données et visualisation

- Pandas
- Seaborn
- Statsmodels

Machine Learning

- Sklearn
- Hdf5
- TensorFlow, Keras et PyTorch

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Numpy

- numpy est un package pour Python spécialisé dans la manipulation des tableaux (array), plus spécifiquement les vecteurs et les matrices
- Les tableaux « numpy » ne gèrent que les objets de même type
- Le package propose un grand nombre de routines pour un accès rapide aux données (ex. recherche, extraction), pour les manipulations diverses (ex. tri), pour les calculs (ex. calcul statistique)
- Les tableaux « numpy » sont plus performants (rapidité, gestion de la volumétrie) que les outils usuels de Python
- Les tableaux « numpy » sont sous-jacents à de nombreux packages dédiés au calcul scientifique sous Python.

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Matplotlib

- matplotlib est une bibliothèque Python capable de produire des graphes de qualité.
- Elle permet la gestion des visualisations graphiques très simples comme Barplot, line plot, histogrammes, scatter plot, courbe de densité.
- Elle permet la manipulation des couleurs et des types de points facilement.
- Elle permet aussi la définition des axes et des titres, l'enchaînement de graphiques dans la même fenêtre visuelle.

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Scipy (1/2)

- Scipy est une librairie de calcul scientifique pour Python.
- Elle s'appuie sur les structures de données de NumPy (vecteurs, matrices).
- Utilisée principalement pour:
 - Mise en œuvre des techniques statistiques inférentielles
 - Mise en œuvre des techniques d'analyse exploratoires

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Scipy (2/2)

- Mise en œuvre des techniques statistiques inférentielles:
 - *tests de comparaisons de populations*
 - *tests pour échantillons appariés,*
 - *mesures d'association*
- Mise en œuvre des techniques d'analyse exploratoires:
 - *essentiellement la classification,*
 - *k-means,*
 - *classification ascendante hiérarchique*
 - *affichage du dendrogramme*
- Lien utile

<http://scipy-lectures.org>

Librairies d'IA

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Analyse de données et visualisation

- Pandas
- Seaborn
- Statsmodels

Machine Learning

- Sklearn
- Hdf5
- TensorFlow, Keras et PyTorch

Analyse de données et visualisation

- Pandas
- Seaborn
- Statsmodels

Pandas

- Pandas est une librairie Python spécialisée dans l'analyse des données.
- Un objet de type «data frame » permet de réaliser de nombreuses opérations de filtrage, prétraitements, avant de ce procéder à l'apprentissage machine.
- Liens utiles:
 - *<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html>*

Analyse de données et visualisation

- Pandas
- Seaborn
- Statsmodels

Seaborn

- Seaborn est une bibliothèque permettant de réaliser des graphiques statistiques en Python.
- Elle s'appuie sur matplotlib et s'intègre étroitement aux structures de données pandas.
- Seaborn aide à explorer et à comprendre les données.
- Ses fonctions de traçage opèrent sur des dataframes et des tableaux contenant des ensembles de données entiers et effectuent en interne la mise en correspondance sémantique et l'agrégation statistique nécessaires pour produire des tracés informatifs.
- Liens utiles:
 - <https://seaborn.pydata.org>

Analyse de données et visualisation

- Pandas
- Seaborn
- Statsmodels

statsmodels

- statsmodels est un module Python qui fournit des classes et des fonctions pour l'estimation de nombreux modèles statistiques différents, ainsi que pour la réalisation de tests statistiques et l'exploration de données statistiques.
- Les résultats sont testés par rapport aux paquets statistiques existants pour s'assurer qu'ils sont corrects.
- Liens utiles:
 - <https://www.statsmodels.org/stable/index.html>

Librairies d'IA

Les modules scientifiques

- Numpy
- Matplotlib
- Scipy

Analyse de données et visualisation

- Pandas
- Seaborn
- Statsmodels

Machine Learning

- Sklearn
- Hdf5
- TensorFlow, Keras et PyTorch

Machine Learning

- Scikit-learn
- Hdf5
- TensorFlow, Keras et PyTorch

Scikit-learn

- Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique.
- C'est une bibliothèque contenant des outils simples et efficaces pour l'analyse prédictive des données
- Elles accessible facilement et réutilisable dans différents contextes
- Elle s'appuie sur NumPy, SciPy, et matplotlib
- Lien utile:

<https://scikit-learn.org/stable/>

Machine Learning

- Scikit-learn
- Hdf5
- TensorFlow, Keras et PyTorch

Hdf5

- Le Hierarchical Data Format (HDF) est un ensemble de formats de fichiers permettant de sauvegarder et de structurer des fichiers contenant de très grandes quantités de données.
- Un fichier HDF est un conteneur de fichiers.
- Il existe principalement deux formats HDF : HDF4 et HDF5.
- Le HDF5 améliore le HDF4.
- Il simplifie la structure des fichiers pour proposer seulement deux types d'objets :
 - *Les ensembles de données (datasets), qui sont des tableaux multidimensionnels contenant des données d'un même type (pré-défini ou dérivé) ;*
 - *Les groupes (groups), qui contiennent, ou regroupent, des ensembles de données et d'autres groupes.*

Machine Learning

- Scikit-learn
- Hdf5
- **TensorFlow, Keras et PyTorch**

TensorFlow, Keras et PyTorch

- TensorFlow, Keras et PyTorch sont des bibliothèques très répandues
- Tensorflow:
 - *C'est une bibliothèque open-source développée par l'équipe Google Brain qui l'utilisait initialement en interne.*
 - *Elle implémente plusieurs méthodes d'apprentissage machine basées sur le principe des réseaux de neurones profonds (deep learning).*

Machine Learning

- Scikit-learn
- Hdf5
- **TensorFlow, Keras et PyTorch**

TensorFlow, Keras et PyTorch

- TensorFlow, Keras et PyTorch sont des bibliothèques très répandues
- Keras:
 - *C'est une librairie Python qui encapsule l'accès aux fonctions proposées par plusieurs librairies de machine learning, en particulier Tensorflow.*
 - *Pourquoi ajouter une couche supplémentaire ?*
 - *C'est parce qu'elle permet de construire, d'entraîner, d'évaluer et d'exécuter facilement toutes sortes de réseaux de neurones.*

Machine Learning

- Scikit-learn
- Hdf5
- **TensorFlow, Keras et PyTorch**

TensorFlow, Keras et PyTorch

- TensorFlow, Keras et PyTorch sont des bibliothèques très répandues
- PyTorch :
 - *C'est une librairie Python open-source d'apprentissage machine qui s'appuie sur Torch (en) développée par Facebook.*
 - *PyTorch permet d'effectuer les calculs tensoriels nécessaires notamment pour l'apprentissage profond.*
 - *La popularité de PyTorch a considérablement augmenté en 2018, essentiellement grâce à sa simplicité et à son excellente documentation contrairement à la première version de Tensorflow.*

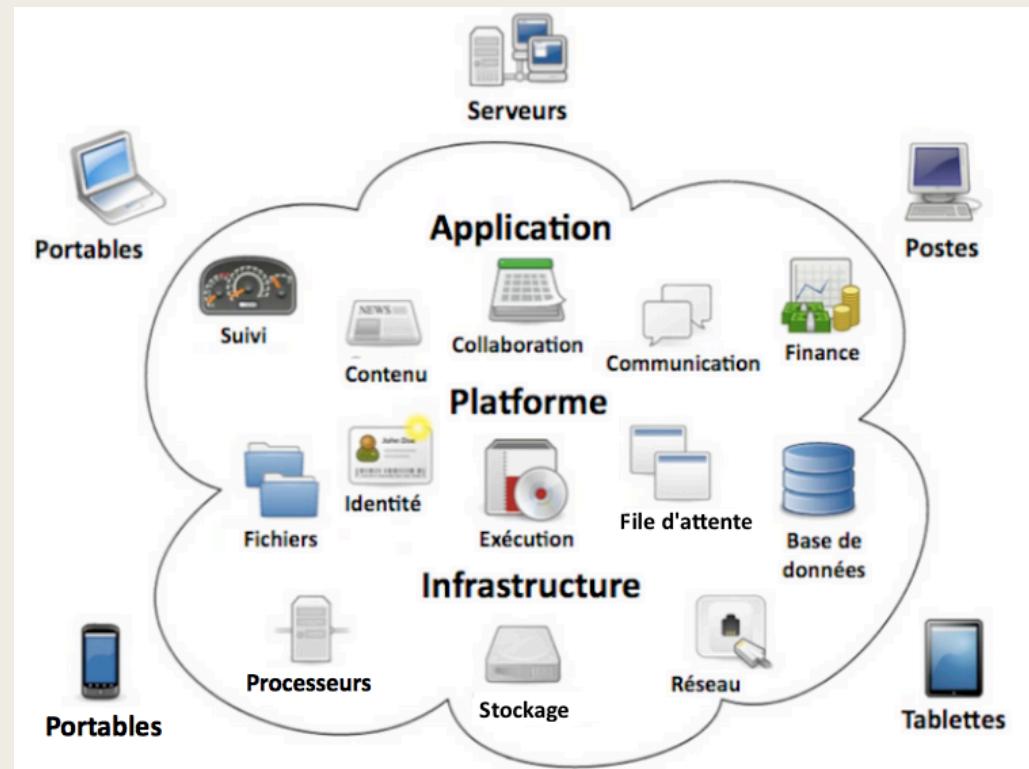
Plan du cours

- Ressources informatiques
 - *Langage de programmation*
 - *L'informatique en nuage*
- Ressources bibliographiques
- Plateforme d'IA
- Étapes d'un projet IA

L'informatique en nuage

1- Définition

- Le cloud computing correspond à l'accès à des services informatiques (serveurs, stockage, mise en réseau, logiciels) via Internet (le « cloud » ou « nuage »).
- Il existe trois types de modèles de déploiement de cloud computing.
- Il existe aussi trois principaux types de services de cloud computing.

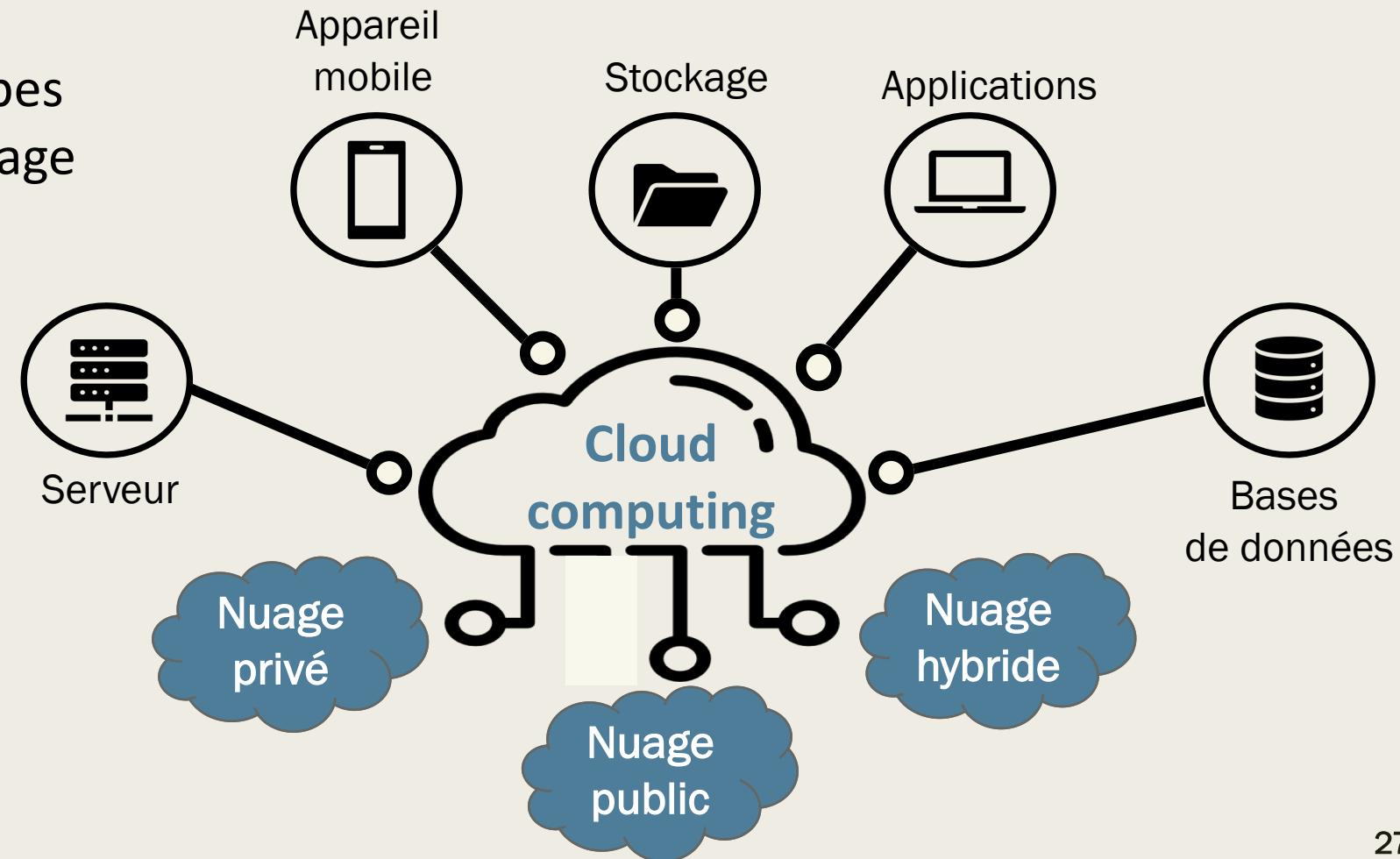


Par Sam Johnston — modification of the Wikipedia file

L'informatique en nuage

2- Types de modèles de cloud computing (1/5)

Il existe trois principaux types de cloud computing : le nuage public, le nuage privé et le nuage hybride.



L'informatique en nuage

2- Types de modèles de cloud computing (2/5)

- Le choix d'un type de Nuage est important .
- Ce choix dépend de plusieurs critères dont:
 - *L'évolutivité et la flexibilité*
 - *La sécurité et la conformité*
 - *Le cout*

L'informatique en nuage

2- Types de modèles de cloud computing (3/5)

- Nuage Public : Lorsqu'une organisation déploie un modèle de nuage public, tout le matériel, les logiciels et l'infrastructure de soutien sont détenus et gérés par un fournisseur de services, qui est fourni exclusivement sur Internet.
 - *Exemples: Amazon Elastic Compute Cloud (EC2), Microsoft Azure, IBM's Blue Cloud, Sun Cloud et Google Cloud.*

L'informatique en nuage

1- Types de modèles de cloud computing (4/5)

- Nuage privé: Le modèle de nuage privé fait référence aux ressources de l'informatique en nuage utilisées exclusivement par une seule entreprise ou organisation.
 - *Exemples : Cisco, Siemens*
- Nuage hybride: Un modèle d'informatique en nuage hybride combine les nuages publics et privés pour partager des données et des applications.
 - *Exemples: Amazon, Microsoft Azure, Cisco et Google.*

L'informatique en nuage

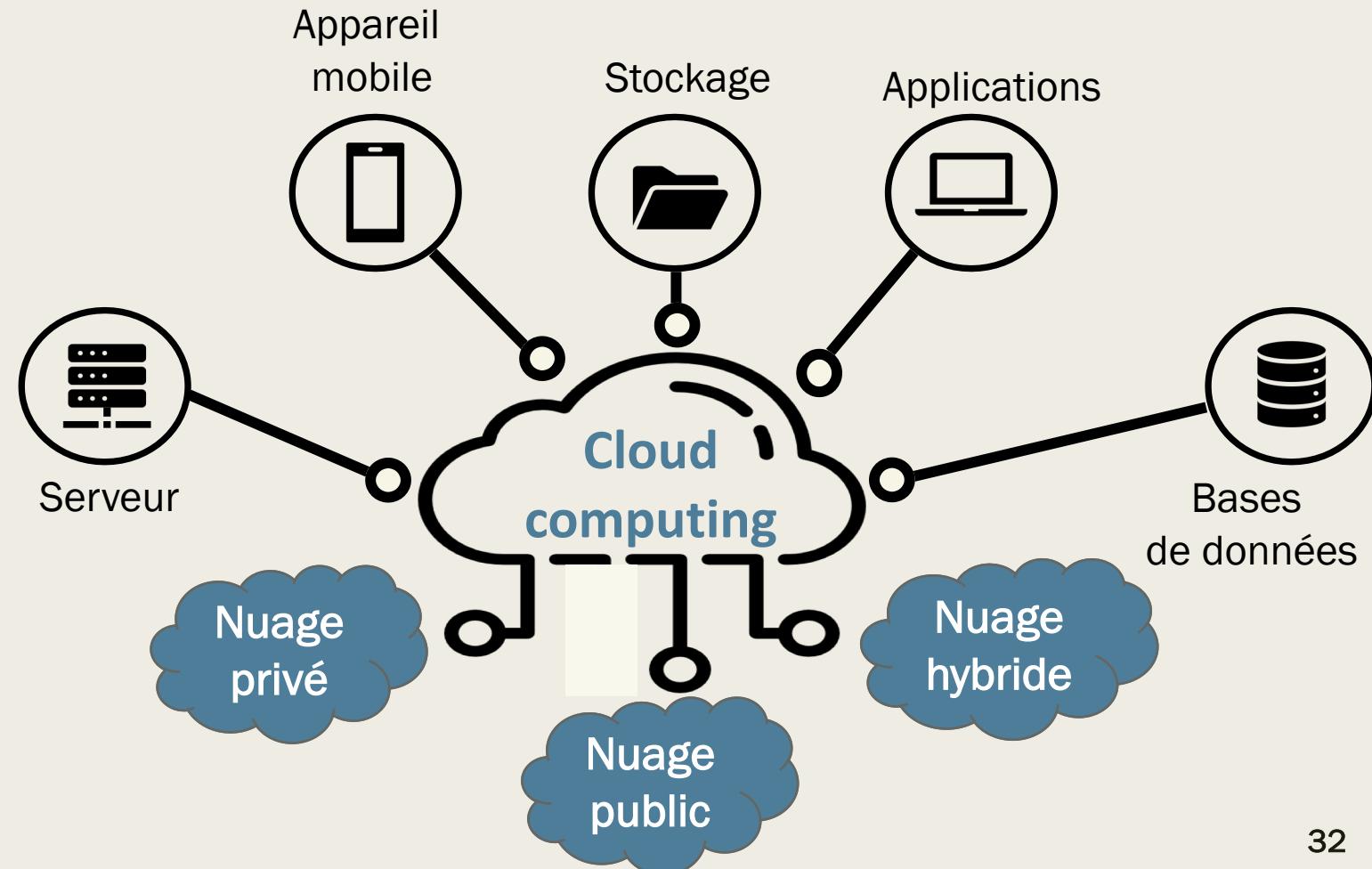
1- Types de modèles de cloud computing (5/5)

- Complément d'information - Multiclouds : Le multicloud est une approche du cloud qui s'appuie sur plusieurs services cloud et sur plusieurs fournisseurs de cloud, public ou privé.
 - *Tous les clouds hybrides sont des multiclouds, mais tous les multiclouds ne sont pas des clouds hybrides.*

L'informatique en nuage

3- Types de services de cloud computing (1/3)

- Les services cloud peuvent prendre la forme d'une infrastructure, d'une plateforme ou d'un logiciel mis à disposition des utilisateurs par Internet.
- Il existe trois principaux types de solutions « aaS » (as-a-Service) : IaaS, PaaS et SaaS.



L'informatique en nuage

3- Types de services de cloud computing (2/3)

Infrastructure en tant que service (IaaS):

- **Description :** IaaS fournit une infrastructure informatique complète sur le cloud, y compris les serveurs, le stockage, les réseaux et les systèmes d'exploitation. Les utilisateurs peuvent gérer ces ressources de manière flexible sans avoir à acheter ni entretenir de matériel physique.
- **Avantages :** Flexibilité élevée, évolutivité, contrôle complet sur les ressources.
- **Fournisseurs :** Amazon Web Services (AWS), Microsoft Azure, Google Compute Engine.

L'informatique en nuage

3- Types de services de cloud computing (2/3)

Plateforme en tant que service (PaaS):

- **Description :** PaaS fournit un environnement de développement complet avec les outils nécessaires pour créer, tester et déployer des applications. Le fournisseur de services gère l'infrastructure sous-jacente, permettant aux développeurs de se concentrer sur le code.
- **Avantages :** Moins de gestion de l'infrastructure, gain de temps pour les développeurs, facile à intégrer avec des outils de développement.
- **Fournisseurs :** Google App Engine, Heroku, Microsoft Azure App Service.

L'informatique en nuage

3- Types de services de cloud computing (2/3)

Logiciel en tant que service (SaaS) :

- **Description** : SaaS propose des applications prêtes à l'emploi accessibles via internet, souvent via un abonnement. L'infrastructure, les mises à jour et la maintenance de l'application sont entièrement gérées par le fournisseur.
- **Avantages** : Facilité d'accès, maintenance simplifiée, accessible depuis n'importe quel appareil connecté.
- **Fournisseurs** : Salesforce, Microsoft 365, Google Workspace.

Plan du cours

- Ressources informatiques
 - *Langage de programmation*
 - *L'informatique en nuage*
- Ressources bibliographiques
- Plateforme d'IA
- Étapes d'un projet IA

Ressources bibliographiques

1- Accès à la bibliothèque de BdeB

The screenshot shows the homepage of the BiblioMondo website. At the top, there's a dark header with the BiblioMondo logo and a search bar. Below it is a yellow navigation bar with links for Tout, Numérique, Livres, Revues, Bandes dessinées, and Films. A search bar with the placeholder "Recherche au catalogue" and a count of "92 657 contenus" is also present. The main content area features a large image of a student in a library, a section titled "Heures d'ouverture" (Opening hours) with details about opening times and closed days, a "Nouvelles" (New) section with a thumbnail of a person at a computer, and a "Services offerts sur place" (Services offered on site) section with a printer icon and a "Crédits d'impression : marche à suivre" (Printing credits: steps to follow) section.

Les étudiants ont accès à diverses plates-formes via le site web de la bibliothèque : <https://biblio.bdeb.qc.ca>

Il suffit de se connecter à son dossier en ligne avec ses accès réseau. La démarche à suivre est disponible [ICI](#).

This screenshot shows a specific page from the BiblioMondo website: "Accès à distance aux ressources en ligne de la bibliothèque". The page has a yellow header with the BiblioMondo logo and a search bar. Below the header, there's a navigation bar with links for Tout, Numérique, Livres, Revues, Bandes dessinées, and Films. A sidebar on the right is titled "ÉVÉNEMENT" and features an illustration of people interacting with digital devices. The main content area contains a section titled "Accès à distance aux ressources en ligne de la bibliothèque" with a detailed description of how students can access resources online using their college email and password. It includes a link to the "marche à suivre" (steps to follow) and a note about printing credits. At the bottom, there's a "Commentaires d'événement" (Event comments) section with a note that anonymous comments are not allowed.

Ressources bibliographiques

2- Accès libre à des livres de IA/sciences de données (1/3)

- 1) [Deep Learning](#) by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
- 2) [Natural Language Processing with Python](#) by Steven Bird, Ewan Klein, and Edward Loper
- 3) [Neural Networks and Deep Learning](#) by Michael Nielsen
- 4) [Automate the Boring Stuff with Python](#) by Al Sweigart
- 5) [An Introduction to Statistical Learning](#) by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani
- 6) [Learning Deep Architectures for AI](#) by Yoshua Bengio
- 7) [Python for Data Science Handbook](#) by Jake VanderPlas

Ressources bibliographiques

2- Accès libre à des livres de IA/sciences de données (2/3)

- 1) [The Hundred-Page Machine Learning Book](#) by Andriy Burkov.
- 2) [A Course in Machine Learning](#) by Hal Daumé III
- 3) [Intuitive ML and Big Data in C++, Scala, Java, and Python](#) by Kareem Alkaseer
- 4) [Python Notes for Professionals book](#)
- 5) [Learning Pandas](#)
- 6) [Machine Learning - A First Course for Engineers and Scientists](#) by Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten, and Thomas B. Schön
- 7) [Dive into Deep Learning](#) by Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola
- 8) [A Comprehensive Guide to Machine Learning](#) Soroush Nasiriany, Garrett Thomas, William Wang, Alex Yang, Jennifer Listgarten, Anant Sahai

Ressources bibliographiques

2- Accès libre à des livres de IA/sciences de données (3/3)

- 1) [Data Mining and Analysis](#) by Mohammed J. Zaki and Wagner Meira Jr.
- 2) [SQL Notes for Professionals book](#)
- 3) [The Data Science Handbook](#) by Carl Shan, Henry Wang, William Chen, and Max Song
- 4) [Understanding Machine Learning: From Theory to Algorithms](#) by Shai Shalev-Shwartz and Shai Ben-David
- 5) [Algorithms Notes for Professionals book](#)
- 6) [Bayesian Reasoning and Machine Learning](#) by David Barber
- 7) [800+ Q&As about: Stats, Python, ML, DL, NLP, CV, MLOps](#) by Steve Nouri
- 8) [An Introduction/A History of Data Science](#)
- 9) [Introduction to Machine Learning](#) by Alex Smola and S.V.N. Vishwanathan

Ressources bibliographiques

3- Accès à des plateformes de publications scientifiques

- **arXiv** est une archive ouverte de prépublications électroniques d'articles scientifiques dans plusieurs domaines comme la physique, des mathématiques, de l'informatique, la statistique, de l'ingénierie électrique et des systèmes, et de l'économie
 - Elle est accessible gratuitement via <https://arxiv.org>
- **Google Scholar** est un service de Google permettant la recherche d'articles et de publications scientifiques.

Plan du cours

- Ressources informatiques
 - *Langage de programmation*
 - *L'informatique en nuage*
- Ressources bibliographiques
- Plateforme d'IA
- Étapes d'un projet IA

Kaggle

- Kaggle est une plateforme web organisant des compétitions en science des données appartenant à Google.
- Sur cette plateforme, les entreprises proposent des problèmes en science des données et offrent un prix aux datalistes obtenant les meilleures performances.
- L'entreprise a été fondée en 2010 par Anthony Goldbloom
- Plusieurs bases de données sont disponibles

Base de données MNIST

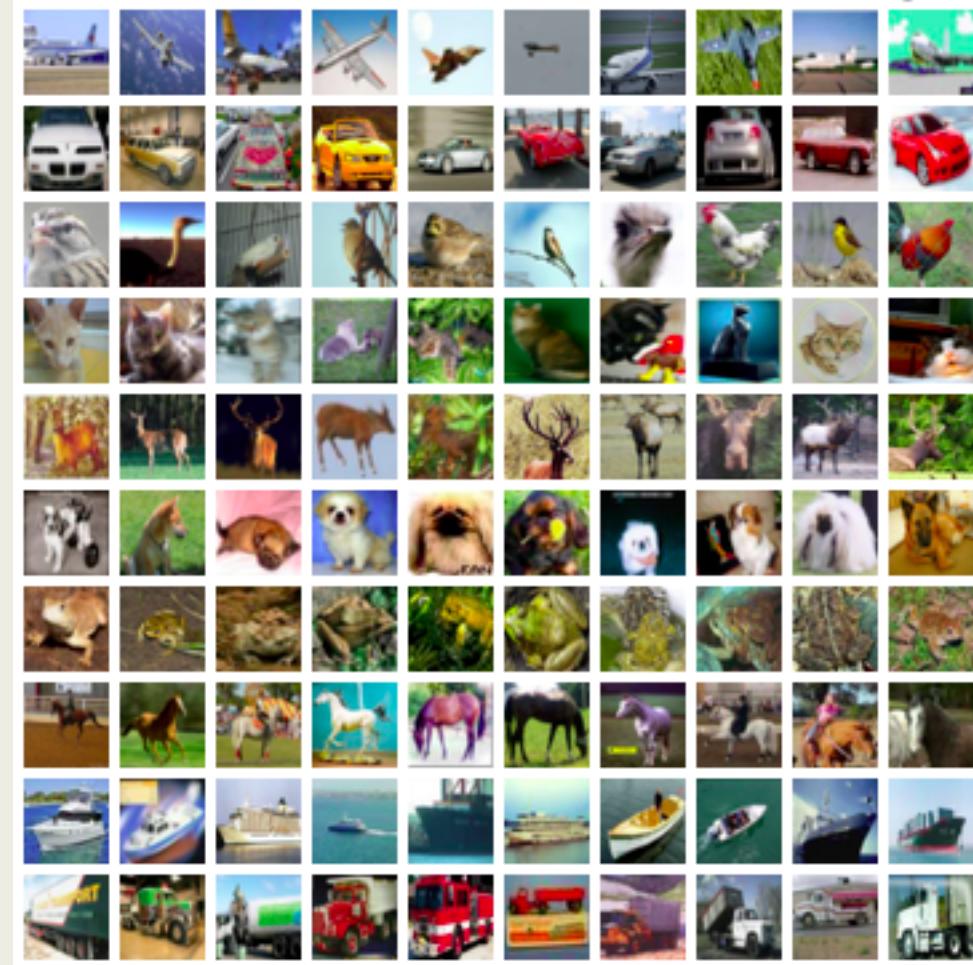
- La base de données MNIST
= Modified ou Mixed National Institute of Standards and Technology.
- C'est une base de données de chiffres écrits à la main.
- C'est un jeu de données très utilisé en IA



https://fr.wikipedia.org/wiki/Base_de_données_MNIST

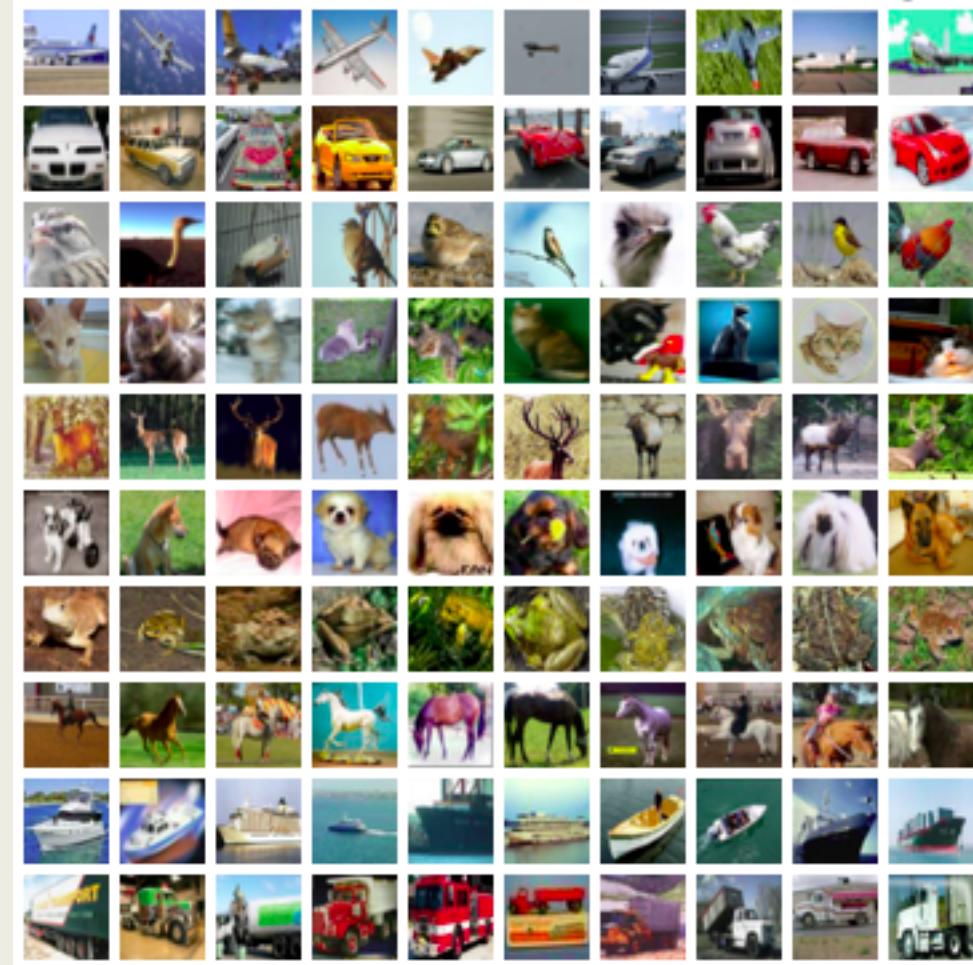
CIFAR10 (1/2)

- Le jeu de données CIFAR-10 (Canadian Institute For Advanced Research) est un ensemble d'images souvent utilisé pour entraîner les algorithmes d'apprentissage machine et de vision par ordinateur.
- Il s'agit de l'un des ensembles de données les plus utilisés pour la recherche.



CIFAR10 (2/2)

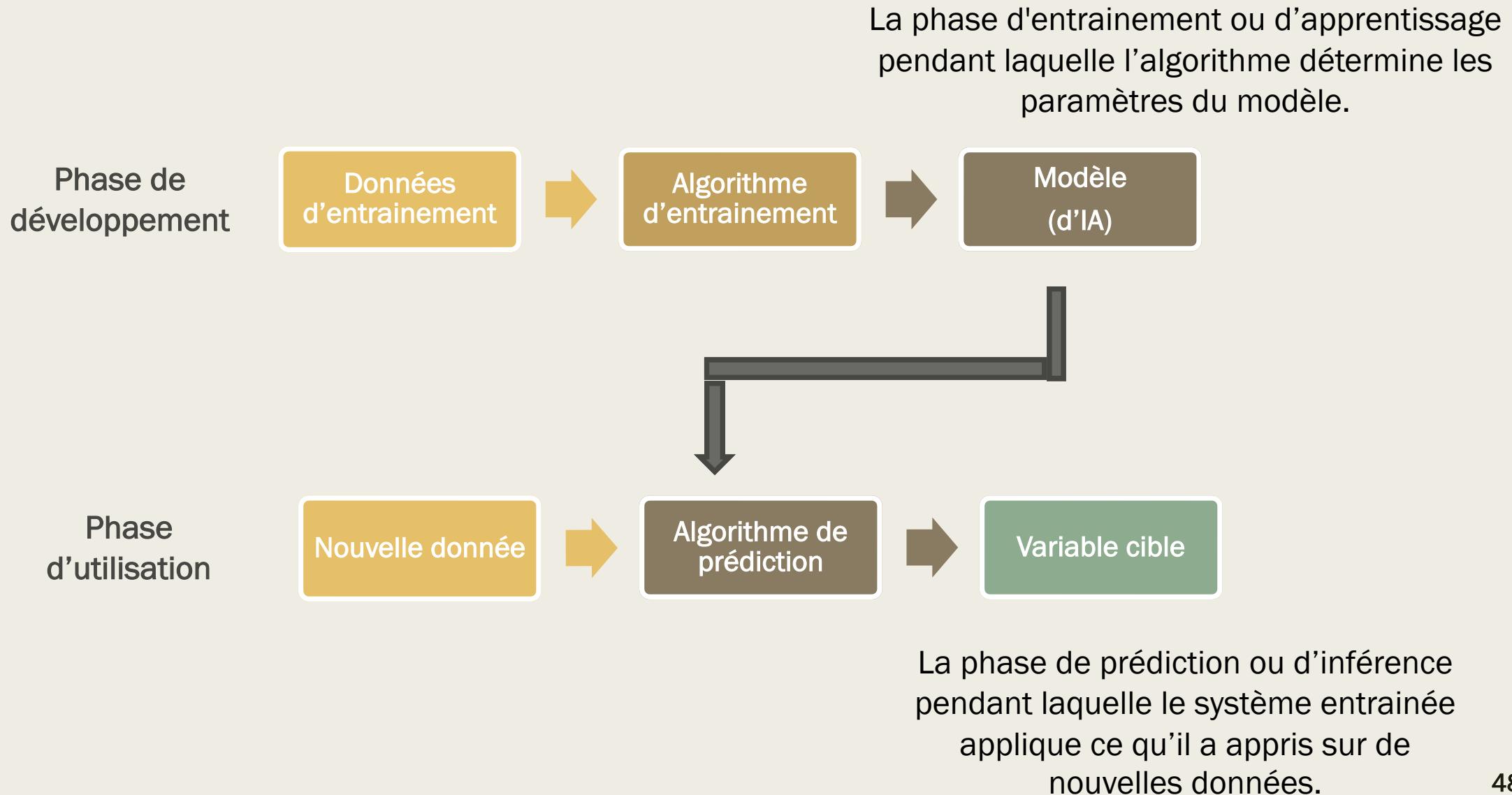
- L'ensemble de données CIFAR-10 contient 60000 images (6000/classe) couleur 32x32 dans 10 classes différentes.
- Les 10 classes différentes représentent des avions, des voitures, des oiseaux, des chats, des cerfs, des chiens, des grenouilles, des chevaux, des bateaux et des camions.



Plan du cours

- Ressources informatiques
 - *Langage de programmation*
 - *L'informatique en nuage*
- Ressources bibliographiques
- Plateforme d'IA
- Étapes d'un projet IA

Étapes d'un projet d'IA



Étapes de développement d'un projet d'IA

Différentes étapes : (1) Analyse, (2) Collecte de données, (3) Nettoyage, (4) Exploration et modélisation des données, (5) Évaluation et interprétation du modèle établi, (6) Mise en production et maintenance.



Étapes d'un projet d'IA



1- Analyse

L'analyse de besoin se fait par une équipe multidisciplinaire qui regroupe plusieurs acteurs.

- **Les experts métiers** : des spécialistes du domaine d'application dont la connaissance est nécessaire pour développer une solution pertinente

- Par exemple:

- Des cliniciens dans le domaine de la santé.
- Des partenaires industriels

- **Des architectes logiciels et des développeurs** qui vont réaliser l'implémentation des outils.

- **Les spécialistes de l'IA** : des experts en IA et des scientifiques de données ►

- **Les acteurs de gouvernance** : Ce sont les décideurs.

- Par exemple:

- Un représentant du service juridique pour la gestion légale des données et l'accès aux données privées.
- Un responsable de l'entreprises pour définir les objectifs économique, écologiques, à atteindre.



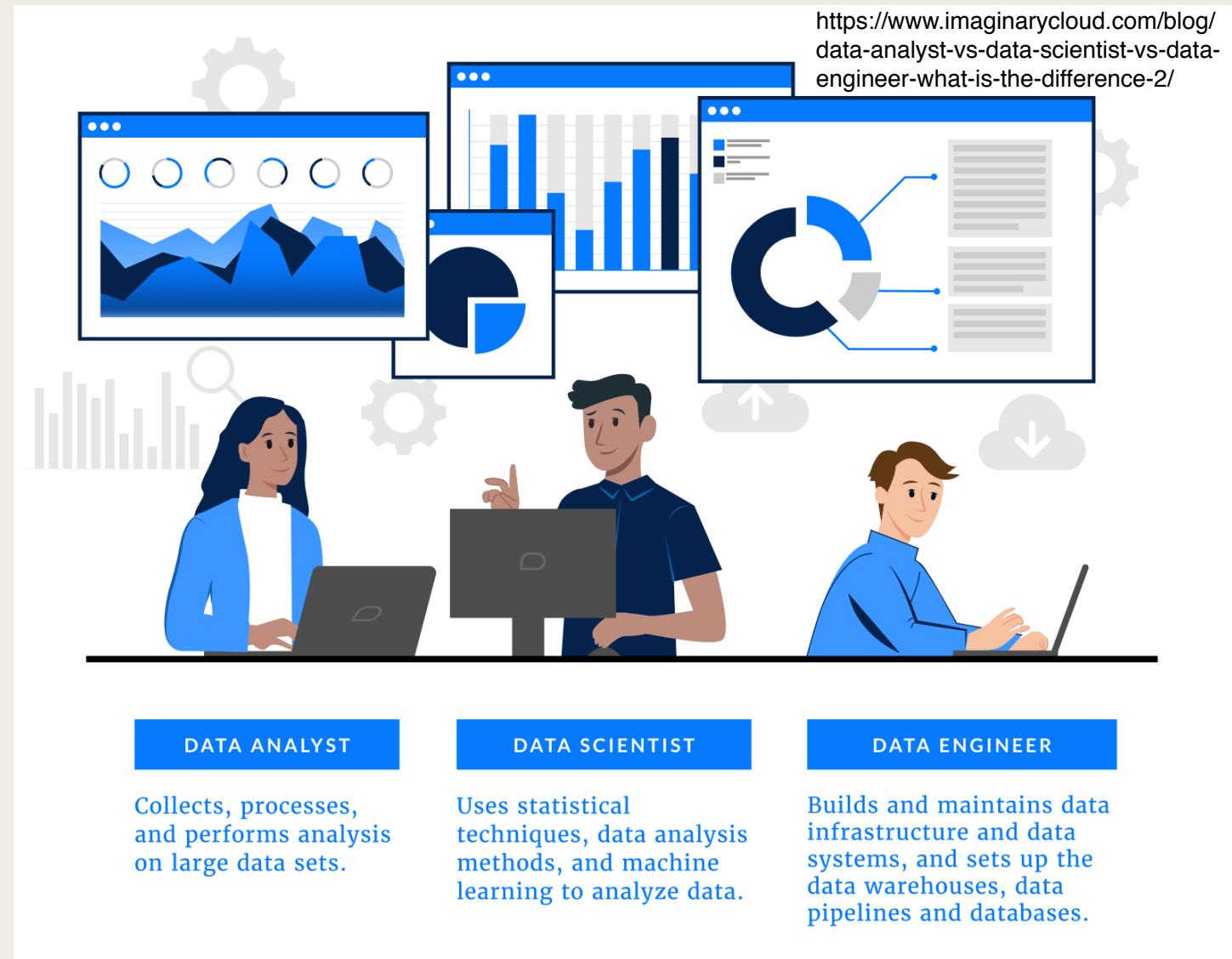
Analyste de données

- Recueillir des données à partir de diverses bases de données et entrepôts par le biais de requêtes ;
- Rédiger des requêtes SQL et des scripts complexes pour recueillir et extraire des informations ;
- Filtrer et nettoyer les données pour obtenir les informations requises ;
- L'exploration de données afin d'en extraire de nouvelles informations ;
- Identifier et analyser les tendances dans des ensembles de données complexes à l'aide d'outils statistiques ;
- Créer des rapports de synthèse à l'aide d'outils de visualisation des données pour que les équipes de direction puissent prendre des décisions opportunes.



Scientifique de données

- Recueillir de grands ensembles de données structurées et non structurées provenant de sources multiples ;
- Interpréter les données à l'aide de méthodes statistiques, en concevant et en évaluant des modèles statistiques avancés pour travailler sur les big data ;
- Construire régulièrement des modèles prédictifs et des algorithmes d'apprentissage automatique pour travailler sur des volumes de données passés ;
- Utiliser des paquets et des outils de visualisation pour créer des rapports et des tableaux de bord pour les parties prenantes concernées ;
- Travailler côte à côte avec des analystes de données et des ingénieurs de données pour formuler les résultats de l'analyse.



Ingénieur de données

- Construire et maintenir les pipelines ETL (Extract, Transform, and Load) et l'infrastructure de données ;
- Informatique en nuage ;
- Créer et intégrer des API
- Déploiement et intégration de l'apprentissage automatique ;
- Développer, construire, tester et maintenir l'architecture des systèmes de traitement à grande échelle et des bases de données pour s'assurer que les besoins de l'entreprise sont satisfaits ;
- Fournir et mettre en œuvre des moyens d'améliorer la fiabilité, l'efficacité et la qualité des données.

Étapes d'un projet d'IA



2- Collecte de données

- Les données sont l'or noir d'un projet IA → Sans données, impossible d'assurer le succès d'un projet en IA!
- L'objectif est de collecter le plus de données possibles, même si on n'est pas sûr de les utiliser par la suite.
- Exemple d'outils de collectes de données:
 - Applications mobiles
 - Capteurs
 - Entretiens approfondis
 - Observations

Étapes d'un projet d'IA



3- Nettoyage

Avant de pouvoir les exploiter, on doit s'assurer que ces données sont bien fiables.

Voici deux exemples typiques de contrôle :

- **Données manquantes.** On parle de données manquantes lorsque des valeurs de variables manquent pour une observation
 - Par exemple, à la suite d'une panne informatique, certains capteurs ont été défaillants dans l'enregistrement de données.
- **Données aberrantes.** Il s'agit d'observations qui sont « distante = éloignées = isolées » des autres observations
 - Par exemple, en observant les données de température, on observe pour une journée des températures très élevées qui ne peuvent pas être celles réellement observées.

Étapes d'un projet d'IA



4- Exploration et Modélisation

- Pour l'étape d'exploration des données, le scientifique de données travaille avec des experts métiers pour:
 - Mieux comprendre les données.
 - Transformer les données brutes en informations précieuses (connaissances)
 - Par exemple: Les organisations explorent les habitudes d'achat des consommateurs afin de générer des stratégies de marketing personnalisées et d'accroître les résultats financiers.
- Pour la modélisation, on utilise des techniques d'AM pour déterminer le modèle

Étapes d'un projet d'IA



5- Évaluation et interprétation

- On doit évaluer le modèle obtenu, c'est-à-dire confirmer qu'il est pertinent
- Pour ce faire, on doit tester la robustesse de l'algorithme d'apprentissage,
- Exemple: L'erreur de prédiction du modèle généré.

6- Mise en production et maintenance

Le système d'intelligence artificielle est prêt à être mis en place.

Travail à faire (Évaluation orale)

- Ce travail aura la forme d'une présentation orale en format power point
- Choisissez une application d'intelligence artificielle.
- Décrivez cette application (son objectif, ses utilisateurs, ...).
- Utilisez la terminologie appris dans le cours (apprentissage supervisé, non supervisé, par renforcement ...).
- Décrivez les différentes étapes de cette application: quelles sont les données utilisées, quelle est le modèle mis en œuvre.
- Décrivez les limites et donnez des pistes d'améliorations (s'il y a lieu).
- Sources : articles scientifiques, applications commercialisées.
- Préparez une présentation (**5 min max de présentation + 3 min de questions**) dans laquelle vous présentez cette application.

Il est important de respecter le temps.