

INF 1901 : Introduction et réflexion sur l'IA

Titres possibles

- Fondements et perspectives de l'intelligence artificielle
- Panorama de l'intelligence artificielle : concepts et réflexions
- Comprendre et questionner l'intelligence artificielle

Objectifs généraux du cours

- Développer une perspective historique de l'IA, des années 50 à maintenant
- Démêler les grands courants et leur articulation logique et scientifique
- Développer une maîtrise minimale de certains sujets plus techniques
- Réfléchir à des enjeux philosophiques et éthiques

Objectifs du module 1

- Explorer l'arc historique de l'IA classique et symbolique (GOFAI, systèmes experts, etc)
- Explorer le développement de l'apprentissage machine, en parallèle de l'IA classique
- Expliquer la différence entre l'intelligence artificielle classique et l'apprentissage machine
- Expliquer la différence entre l'IA et le NLP (traitement automatique de la langue)
- Expliquer la différence entre l'apprentissage machine et la programmation traditionnelle
- Distinguer et positionner (dans le temps) les jalons classiques : Eliza, DeepBlue, Watson, AlphaGo, ChatGPT

Travail noté : ligne du temps (personnages, innovations et jalons) et diagramme représentant les relations logiques entre les différents courants de l'IA

Objectifs du module 2

- Expliquer ce que sont les données, ainsi que leur rôle en apprentissage machine
- Clarifier les racines statistiques de l'apprentissage machine
- Décrire les composantes principales d'un algorithme d'apprentissage
- Expliquer la différence entre l'entraînement d'un modèle et l'inférence
- Décrire les différents paradigmes de l'apprentissage machine
- Fournir des exemples de problèmes pour chacun
- Décrire comment on évalue la performance d'un modèle d'apprentissage machine
- Décrire l'importance de la capacité de généralisation pour un algorithme d'apprentissage

Travail noté : construction guidée d'un algorithme de classification naïve bayésienne pour les courriels dans Google Sheets, suivi de questions interprétatives

Objectifs du module 3

- Décrire les différentes composantes d'un réseau de neurones
- Expliquer en quoi le réseau de neurones est une métaphore du fonctionnement d'un cerveau (connexionnisme)
- Expliquer l'idée de base de l'apprentissage profond : les couches successives d'un réseau de neurones encodent une hiérarchie d'abstractions
- Expliquer le rôle des GPUs et des frameworks de différentiation automatique (Tensorflow, PyTorch, etc) dans l'essor de l'apprentissage profond
- Répertorier les différents types de topologie (réseaux à convolution, transformers, réseaux récurrents, etc), et expliquer leurs avantages et inconvénients
- Expliquer le succès relatif des réseaux de neurones, par rapport aux autres algorithmes d'apprentissage

Travail noté : exercice de manipulation et d'interprétation de l'application interactive Tensorflow Playground (en cours de conception)

Objectifs du module 4

- Décrire le fonctionnement d'un modèle de langue classique
- Décrire la différence entre la classification naïve bayésienne (modèle probabiliste discriminatif) et un modèle de langue (modèle probabiliste génératif)
- Expliquer la notion et l'utilité des "word embeddings"
- Décrire les composantes d'un transformer (auto-attention, etc).
- Décrire comment on peut passer d'un modèle prédictif des mots à un "chatbot" (qui peut répondre aux questions, etc)
- Décrire le fonctionnement du RHLF
- Décrire la notion d'agent en IA

Travail noté : reste à déterminer

Objectifs du module 5

- Réfléchir à certaines questions relatives à la philosophie de l'esprit et des sciences cognitives : est-ce que l'IA a un "modèle du monde"? Est-ce possible d'en acquérir un seulement par la lecture de textes? Est-ce que l'IA est capable de raisonnement? Est-ce que l'IA a une conscience de soi, peut avoir des émotions?
- Réfléchir en particulier aux différentes conceptions en matière de langage : les idées de Chomsky, et l'avènement de ChatGPT (empirisme versus rationalisme)
- Réfléchir au problème de "l'alignement de l'IA"
- Décrire l'idée de la singularité
- Réfléchir aux conséquences sur la société de l'avènement d'une IA supra-humaine, en matière d'intelligence
- Décrire les différentes écoles de pensée et d'opinions en matière d'IA : les "doomers" (ceux qui croient que l'IA est néfaste), les "accélérationnistes" (qui croient le contraire), etc.

Travail noté : traiter un de ces problèmes (questions) et établir un protocole d'exploration (pour tenter d'y répondre) à l'aide d'un chatbot