

Modélisation des savoirs musicaux relevant de l'oralité

Marc Chemillier (EHESS, CAMS)
13 novembre 2024

Intelligence artificielle et savoirs musicaux

Règles et apprentissage

Qu'est-ce que l'intelligence artificielle ?

IA maigre et agnosticisme



European Research Council (ERC) REACH project

Règles et apprentissage

Vidéo du logiciel d'improvisation Djazz avec le célèbre cithariste malgache Justin Vali (qui a joué avec Peter Gabriel et Kate Bush)

<http://digitaljazz.fr/2023/06/19/videos-djazz-avec-justin-vali/#masoala>

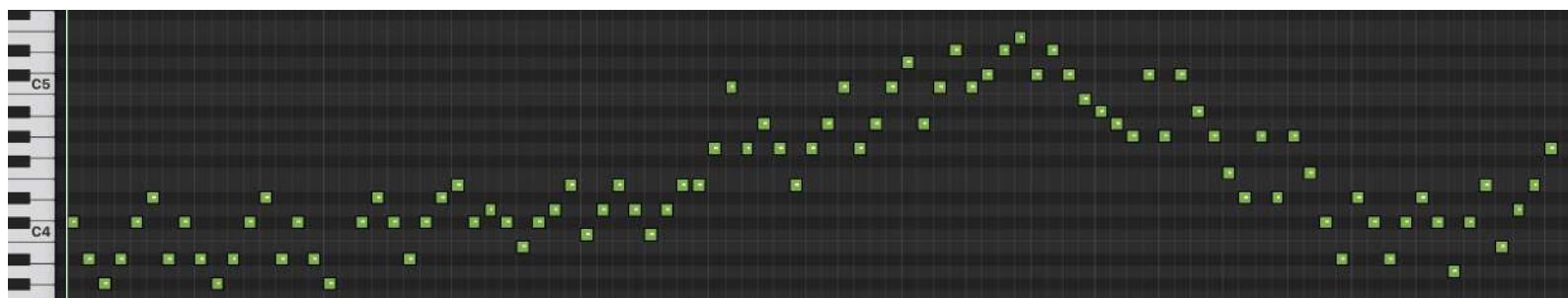


Djazz est basé sur une pulsation régulière :

- soit une battue manuelle (cf. ralenti à la fin)
- soit un clic de percussions d'Ableton Live

Le logiciel improvise en utilisant des données de solos pré-enregistrés. Il peut créer des hybrides mélangeant les données de musiciens différents. Dans la vidéo précédente, le solo à 2:50 provient d'une transcription de Joe Zawinul en concert au Munich Piano Summer en 1985 (à 13:03, Partie I) :

<https://www.youtube.com/watch?v=6WQfeplhzT4&t=783s>





KV 516f, quelques mesures écrites par Mozart en marge de son manuscrit de l'Adagio du Quintette à cordes K 516.

Démo du jeu de dés de Mozart :
<https://www.buschs.de/Mozart/index.html>

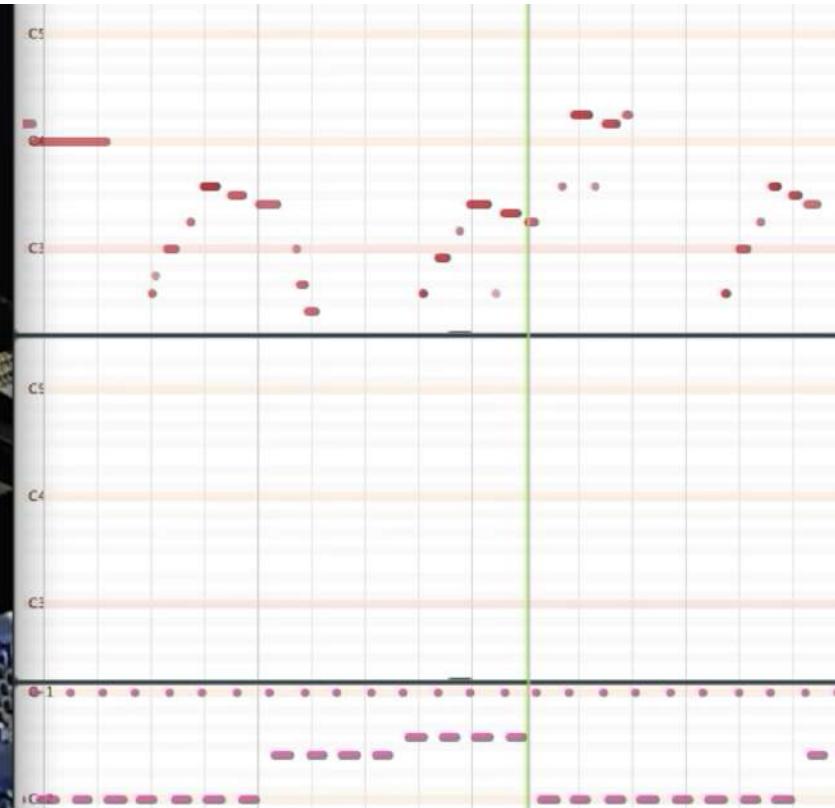
Les deux tables indiquent pour chacune des 8 mesures du menuet (première et deuxième parties) quelle carte il faut tirer en fonction du résultat du lancer des dés. Les connaissances musicales sont donc à la fois dans les règles et dans les données :

- données = cartes contenant les mesures pré-composées
- règles = tables définissant quelle carte peut succéder à quelle autre

*Avec Djazz, les données peuvent être captées en direct : c'est de l'**apprentissage automatique**.*

Elles sont recombinées par Djazz pour improviser en tenant compte du rythme et des harmonies.

<https://www.youtube.com/watch?v=J52SEoHvx1o>



Qu'est-ce que l'intelligence artificielle ?

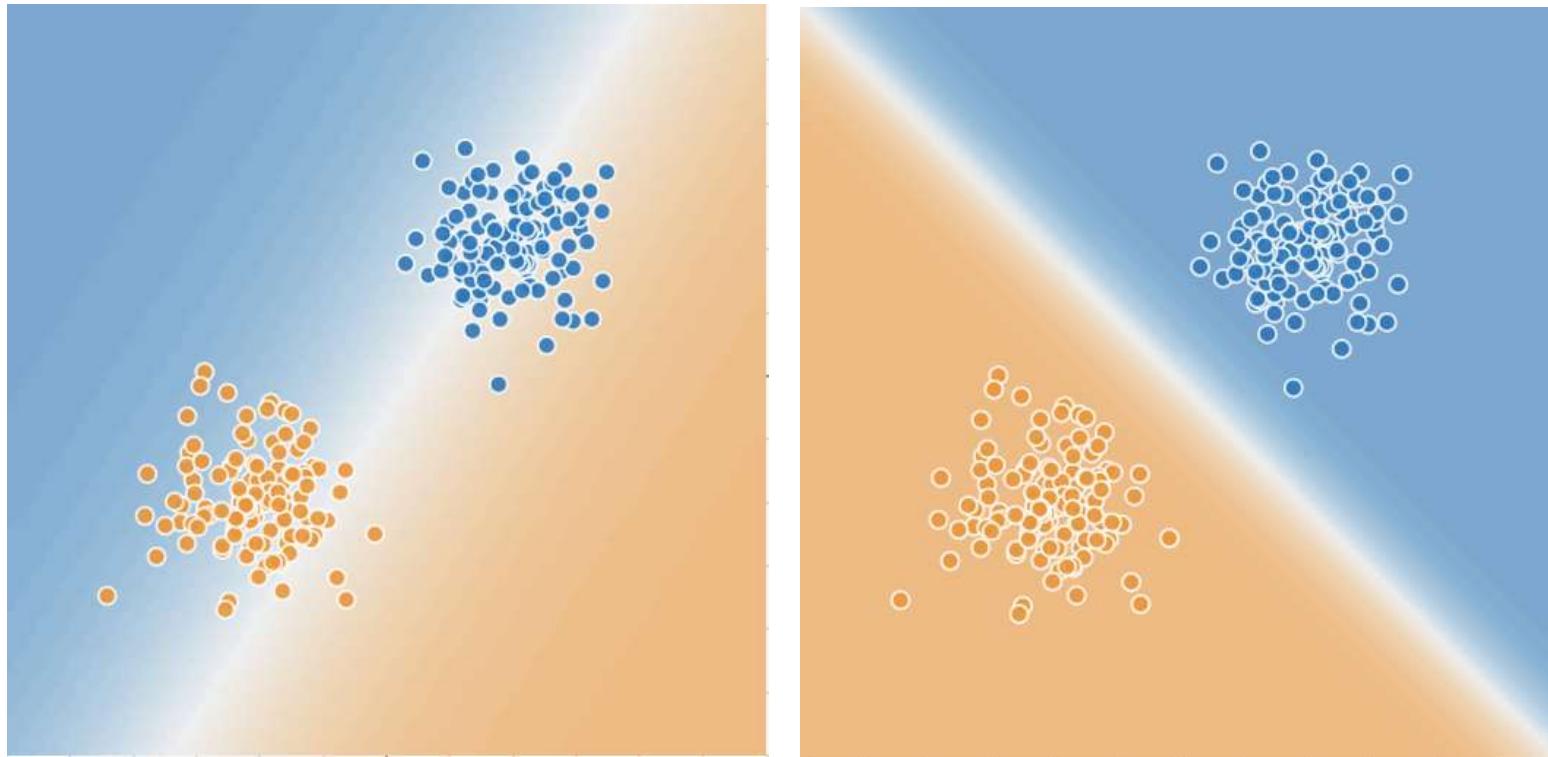
1956 : *John McCarthy invente le terme "intelligence artificielle" = réaliser des programmes informatiques capables de simuler l'intelligence humaine dans l'exécution de tâches complexes.*

- 1950-1980 : **règles** (« systèmes-experts »)
- 1990-2000 : **apprentissage automatique**
(données d'apprentissage étiquetées à la main)
- 2010-2020 : **apprentissage profond** (big data, données brutes et abondantes)

En 1990-2000, l'augmentation de la taille mémoire des ordinateurs a permis l'apprentissage automatique (machine learning).

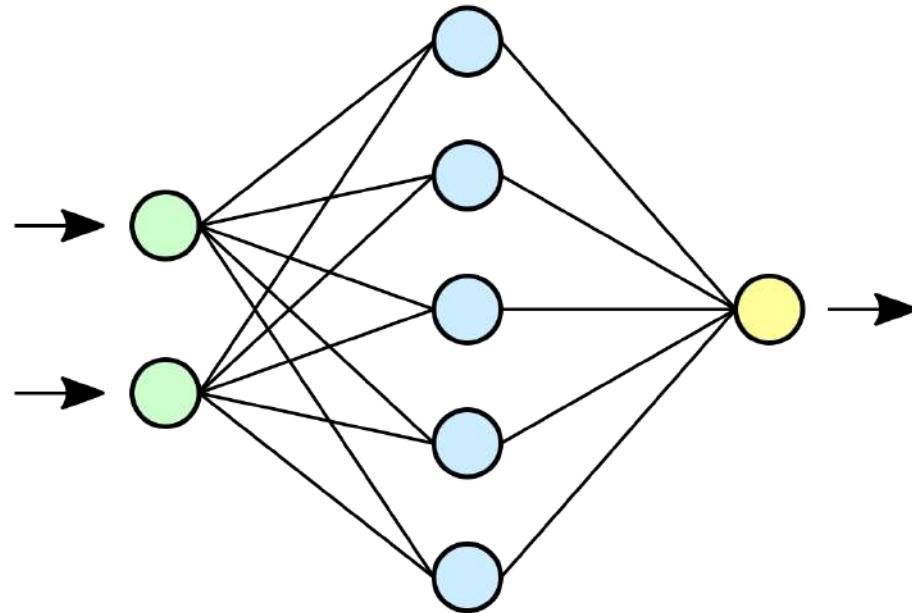
L'un des modèles d'apprentissage automatique est le réseau de neurones. Il utilise la régression logistique, qui consiste à classer des données de façon linéaire. Ex : grouper des points en deux classes séparées par une ligne droite. On ajuste progressivement les paramètres de la droite ($y = ax + b$, paramètres a et b) par entraînement, jusqu'à obtenir la bonne droite qui sépare les points orange et bleu.

Démo sur les réseaux de neurones:
<https://playground.tensorflow.org>



Si les ensembles de points ne sont pas séparables linéairement, on combine plusieurs unités de régression logistique dans un réseau de neurones.

Dans un réseau de neurones simple, on a 3 couches : entrée, couche cachée, sortie

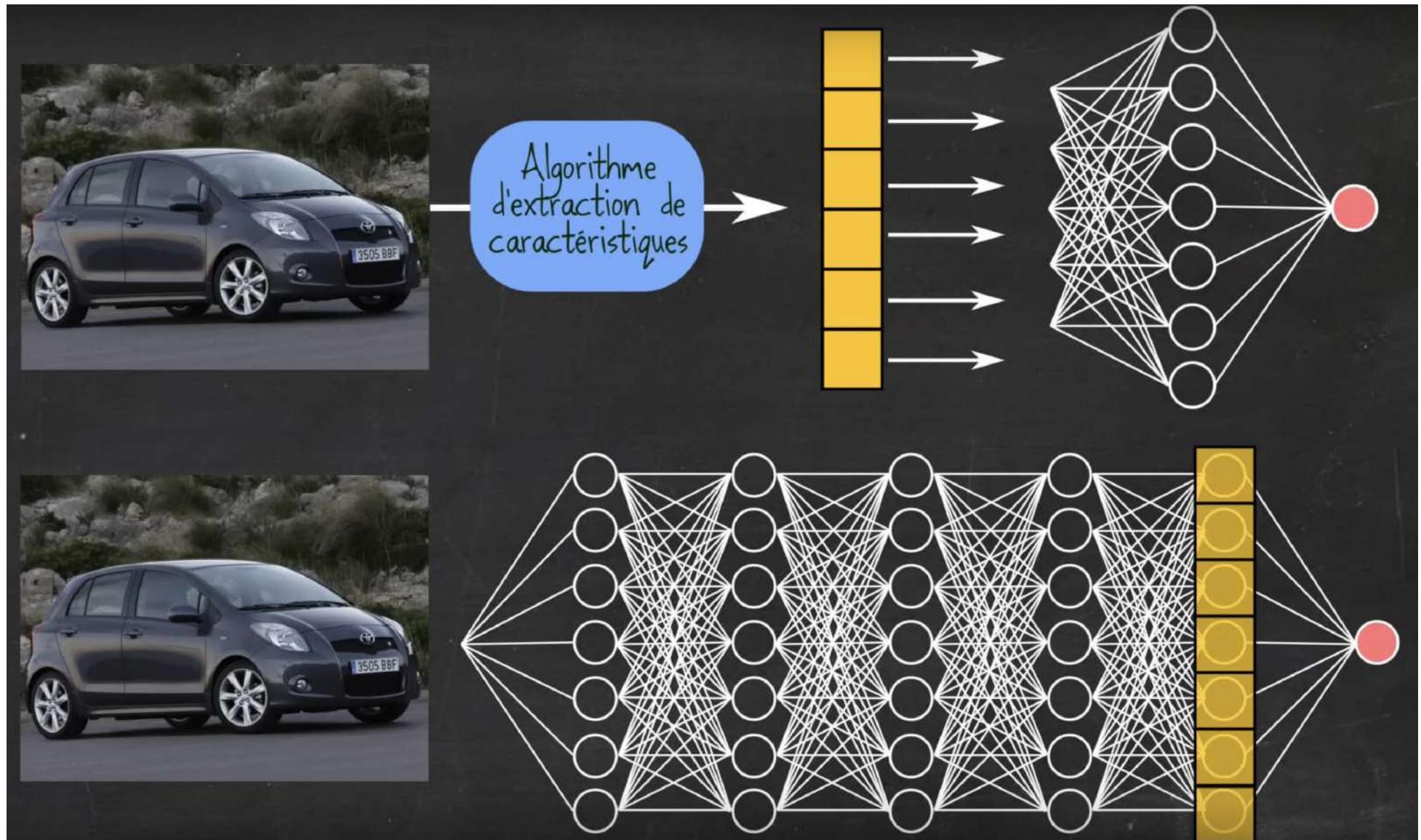


Pour des images numérisées en pixels, il faut simplifier les données par extraction préalable de caractéristiques (abstractions d'images) qu'on donne en entrée au réseau (et non les pixels).

2012 : succès spectaculaire de SuperVision (Yann Le Cun, recruté ensuite par Facebook).

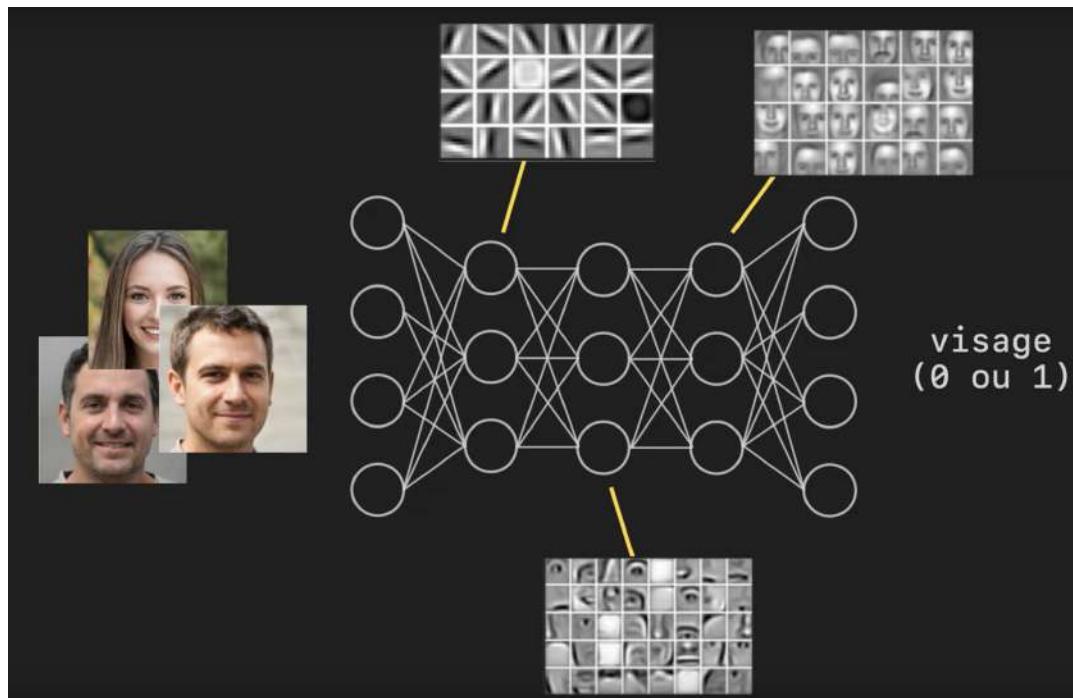
= pas d'extraction préalable de caractéristiques, il traite directement les pixels, mais avec un très grand nombre de couches. Ce nouveau modèle est appelé l'**apprentissage profond** (= grande profondeur de couches).

Du point de vue de l'anthropologie de la connaissance, on a supprimé l'étape d'extraction de caractéristiques simplement en ajoutant des couches supplémentaires. Mais ces « caractéristiques » sont recréées par le réseau lui-même dans son avant-dernière couche.



David Louapre (ScienceEtonnante): Le deep learning
<https://www.youtube.com/watch?v=trWrEWfhTVg>

*Détection de visage : 1ère couche = traits simples, avant-dernière couche = formes élaborées (yeux, nez,...). Le réseau redécouvre par lui-même ces **concepts** : si une entrée ne contient que le concept, on voit quel neurone est activé.*



Alexandre TL : Comprendre les réseaux de neurones
<https://www.youtube.com/watch?v=bkoNI7ImPBU>

IA maigre et agnosticisme

*L'apprentissage profond utilise un grand nombre de couches de neurones, donc il lui faut beaucoup de données d'entraînement (années 2010 grandes bases d'images classifiées comme ImageNet de Stanford). Le **big data** est l'utilisation (pas très contrôlée...) des données disponibles en ligne.*

*Depuis quelques années, tendance au **small data**.*

- *augmentation de la puissance des ordinateurs : pas gros serveurs pour calculs distribués*
- *parcimonie : pas big data pour toutes les tâches*

The Small Data Manifesto (14 juillet 2024) :

<https://motherduck.com/blog/small-data-manifesto/>

*Djazz utilise une **IA maigre**. Les données d'entraînement du modèle **ne proviennent que***

- *du musicien interagissant avec Djazz*
 - *d'un corpus de transcriptions sélectionnées*
- Contrairement au small data, l'IA maigre ajoute*
- *écologie : pas de gros serveur (stockage, calcul)*
 - *esthétique : intention artistique dans le choix des données*

*Le modèle d'apprentissage de Djazz **n'est pas un réseau de neurones** (oracle des facteurs). Donc **Djazz ne fait pas d'apprentissage profond.***

Concernant l'anthropologie de la connaissance, on a vu que l'apprentissage profond n'utilise **pas de caractéristiques extraites au préalable**.

agnosticisme : il n'y a **pas de connaissance a priori hors de l'expérience** (métaphysique, religion)

Dans Djazz, il y a étiquetage préalable (pulsations, harmonie). Mais **pas de connaissance musicale a priori sur ces étiquettes** (= simples marques). Donc Djazz est agnostique, d'où l'**adaptabilité à différents contextes culturels**.

Générateur de chanson avec big data : <https://www.suno.ai/>