

# SAINT Curriculum

UNITÉ 3 : Application de l'IA à la parole et à la vision

---

Livrable : WP2/2.2



## SAINT

HANDS ON INTRODUCTION TO ARTIFICIAL  
INTELLIGENCE IN PRIMARY EDUCATION  
USING MINECRAFT

AVRIL 2023

---

EKVASIS

Auteurs : EKVASIS Team

Numéro de projet : 2022-1-FR01-KA220-SCH-000087794



**Co-funded by  
the European Union**

Le soutien de la Commission européenne à la production de cette publication ne constitue pas une approbation de son contenu, qui reflète uniquement les opinions des auteurs, et la Commission ne peut être tenue responsable de l'utilisation qui pourrait être faite des informations contenues dans cette publication.

## HISTORIQUE DES RÉVISIONS

Version	Date	Auteur	Description	Action	Pages
1.0	24/11/2022	TARAN	Creation	C	TBS

(\* ) Action: C = Création, I = Insertion, U = Mise à jour, R = Remplacement, D = Suppression

## DOCUMENTS DE RÉFÉRENCE

ID	Référence		Titre
1	2022-1-FR01-KA220-SCH-000087794		SAINT candidature
2			

## DOCUMENTS APPLICABLES

ID	Référence		Titre
1			
2			

## Table des matières

1	Introduction du projet.....	4
1.1	La portée du projet .....	4
1.2	Les groupes cibles.....	4
1.3	L'objet du présent document.....	5
2	Glossaire de l'unité .....	5
3	Introduction de l'Unité.....	6
3.1	Description .....	6
3.2	Objectifs et résultats de l'apprentissage.....	7
3.3	Estimation du temps à passer.....	7
4	Contenu du cours de l'Unité.....	7
4.1	Introduction.....	7
4.2	Idée 1 : Perception .....	8
4.3	Idée 2 : Représentation et raisonnement .....	9
4.4	Idée 3 : Apprentissage.....	10
4.5	Idée 4 : Interaction naturelle .....	12
4.6	Idée 5 : Impact sociétal.....	14
4.7	Études de cas et exemples de réussite .....	15
4.7.1	Exemple 1 : L'IA DeepMind de Google pour la lecture labiale.....	15
4.7.2	Exemple 2 : Watson d'IBM pour l'oncologie .....	16
4.7.3	Exemple 3 : Alexa d'Amazon pour la reconnaissance vocale .....	16
5	Matériel et ressources supplémentaires .....	17
6	Synthèse .....	18
7	Quiz.....	19
8	Références.....	20

# 1 Introduction du projet

## 1.1 La portée du projet

Ce projet, qui constitue un environnement d'apprentissage numérique idéal pour enseigner aux enfants les applications pratiques de l'IA sur la base des lignes directrices du projet AI4K12, est motivé par les objectifs suivants :

- Présenter aux élèves, aux enseignants et aux éducateurs les concepts de l'IA, son impact sur notre société et ses applications pratiques,
- Répondre au besoin croissant de développer des solutions d'apprentissage à distance facilitant l'engagement des élèves, leur créativité, leur capacité à résoudre des problèmes et à prendre des décisions,
- Améliorer les compétences des enseignants et des éducateurs avec de nouveaux ensembles de compétences (PBL, IA, gamification, etc.) développées grâce à des méthodes d'enseignement innovantes,
- Améliorer les taux d'engagement des enfants grâce à l'utilisation d'une méthode d'enseignement innovante, en aidant les enfants à développer leur créativité,
- Réduire l'écart entre les besoins et la disponibilité des compétences liées à l'IA.

AI Adventures in Minecraft enseigne les compétences liées à l'IA aux enfants âgés de 9 à 12 ans, en utilisant un monde Minecraft. Nous créons ainsi un environnement d'apprentissage amusant, interactif et créatif par le biais d'activités et de défis spécifiques alignés sur les lignes directrices AI4K12 (ai4ka12.org) et les 5 grandes idées de l'IA : 1) Perception, 2) Représentation et raisonnement, 3) Apprentissage, 4) Interactions naturelles, 5) Impact sociétal.

À cette fin, le projet développe et promeut les résultats tangibles suivants :

- Ce curriculum : un cours d'apprentissage complet pour introduire l'IA dans l'enseignement scolaire, basé sur les 5 grandes idées du cadre AI4K12. Le cours diffuse des connaissances sur les lignes directrices de l'éducation à l'IA AI4K12 et les 5 grandes idées, explore l'impact de l'IA dans notre société et améliore la compréhension des concepts pertinents.
- Un monde Minecraft sur mesure (AI Adventures World) proposant des défis éducatifs basés sur le cours d'apprentissage. Il utilise le concept de salle d'évasion et propose des activités d'apprentissage par problèmes. Un défi pour chaque unité ou leçon.
- L'espace virtuel de la fonderie, qui soutient une communauté croissante d'utilisateurs de SAINT et guide la maintenance corrective/perfective et évolutive de l'ensemble de la formation.

## 1.2 Les groupes cibles

Le projet prévoit la participation directe d'enseignants, principalement des enseignants d'enfants âgés de 9 à 12 ans ou du personnel de l'enseignement supérieur impliqué dans l'enseignement des

éducateurs. Ces enseignants enseignent des matières STIM ou ont des connaissances et un intérêt pour l'IA et/ou Minecraft.

En ce qui concerne les groupes cibles indirects identifiés, les personnes suivantes peuvent être impliquées :

- Les centres STEM qui cherchent à développer leur catalogue de technologies d'enseignement innovantes ou leur catalogue de produits améliorant les connaissances en matière d'IA,
- Les établissements d'enseignement supérieur qui collaborent avec des entreprises ou des autorités publiques engagées dans la création de matériel pédagogique,
- Les organisations, associations ou réseaux cherchant à fournir aux parents et/ou aux éducateurs du matériel éducatif sur l'IA : tels que les clubs de codage, les centres d'apprentissage pour adultes, les services de coaching entrepreneurial, les centres de formation continue, etc.

### 1.3 L'objet du présent document

Le work package n°2 - AI4K12 Educational Programme se concentre sur la production d'un cours complet sur l'IA avec un ensemble de 5 défis dans le monde Minecraft pour illustrer la mise en œuvre pratique de la technologie.

Ce programme d'enseignement de l'IA est composé d'un total de 5 unités de matériel pédagogique basé sur les lignes directrices de l'enseignement AI4K12 et les objectifs d'apprentissage mis en lumière à la suite d'enquêtes nationales :

1. Application de l'IA à l'apprentissage automatique,
2. Application de l'IA au travail et à l'esprit d'entreprise,
3. Application de l'IA à la parole et à la vision,
4. Application de l'IA dans les jeux et les puzzles,
5. Application de l'IA dans la vie quotidienne.

En outre, un glossaire est créé dans chaque unité afin de faciliter l'adoption du logiciel SAINT par les enseignants et les écoles.

## 2 Glossaire de l'unité

Mots	Définition
<b>Apprentissage automatique</b>	Sous-ensemble de l'IA qui utilise des algorithmes et des modèles statistiques pour permettre aux systèmes informatiques d'améliorer leurs performances dans une tâche spécifique, en fonction des données fournies, sans être explicitement programmés.
<b>Reconnaissance de la parole</b>	La capacité d'une machine ou d'un programme à identifier et à comprendre la parole humaine, souvent utilisée dans des applications telles que les assistants virtuels, le service client automatisé et la transcription.

<b>Vision par ordinateur</b>		Un domaine d'étude visant à permettre aux ordinateurs d'interpréter et de comprendre les données visuelles du monde qui les entoure, souvent utilisé dans des applications telles que la reconnaissance d'objets, la reconnaissance faciale et les véhicules autonomes.
<b>Apprentissage profond</b>	<b>en</b>	Un sous-ensemble de l'apprentissage automatique qui utilise des réseaux neuronaux artificiels pour modéliser des modèles complexes dans les données, souvent utilisé dans des applications telles que la reconnaissance d'images et de la parole.
<b>Traitement du langage naturel (NLP)</b>		Un sous-ensemble de l'IA qui vise à permettre aux ordinateurs de comprendre, d'interpréter et de générer du langage humain, souvent utilisé dans des applications telles que les chatbots, la traduction linguistique et les véhicules autonomes. Les réseaux d'apprentissage automatique permettent de modéliser des modèles complexes dans les données, souvent utilisés dans des applications telles que la reconnaissance d'images et de discours.
<b>Apprentissage renforcement</b>	<b>par</b>	Un type d'apprentissage automatique dans lequel un agent apprend à prendre des décisions par essais et erreurs, en recevant des récompenses ou des punitions en fonction de ses actions. de ses actions.
<b>Réalité augmentée</b>		Une expérience interactive dans laquelle des informations numériques sont superposées au monde physique. Les données numériques sont superposées au monde physique et sont souvent utilisées dans des applications telles que les jeux, l'éducation et le marketing.
<b>Réalité virtuelle</b>		Un environnement généré par ordinateur qui simule la présence physique d'un utilisateur et permet l'interaction, souvent utilisé dans des applications telles que les jeux, la formation et la thérapie. les jeux, la formation et la thérapie.

## 3 Introduction de l'Unité

### 3.1 Description

Cette unité introduira les apprenants au domaine passionnant de l'intelligence artificielle et à ses applications dans les domaines de la parole et de la vision. Elle explore les différents types d'apprentissage automatique, y compris l'apprentissage supervisé, non supervisé et par renforcement, et la manière dont ils sont utilisés dans la reconnaissance vocale, la classification d'images et d'autres applications du monde réel. L'unité se penchera également sur l'interaction naturelle, notamment sur la manière dont l'IA interagit avec les humains et dont la réalité virtuelle et augmentée modifie la manière dont nous interagissons avec la technologie. En outre, l'unité abordera l'impact sociétal de l'IA, y compris les considérations éthiques et les avantages et inconvénients de la technologie de l'IA. À la fin de l'unité, les apprenants devraient avoir une compréhension fondamentale de l'IA dans les domaines de la parole et de la vision, ainsi que de son potentiel pour résoudre les problèmes du monde réel.

## 3.2 Objectifs et résultats de l'apprentissage

Dans cette unité, les apprenants se familiariseront avec l'application de l'IA à la parole et à la vision. Ils apprendront les principes fondamentaux de l'apprentissage automatique, y compris l'apprentissage supervisé, non supervisé et par renforcement. L'unité couvrira également l'interaction naturelle et sa pertinence dans l'IA, ainsi que l'impact sociétal et les considérations éthiques associés à l'utilisation de l'IA.

À l'issue de cette unité, les apprenants :

- Devraient être capables d'identifier les différents types d'apprentissage automatique et leurs applications dans les domaines de la parole et de la vision.
- Pourront expliquer l'importance de l'interaction naturelle et son rôle dans le développement des systèmes d'IA.
- Auront une solide compréhension de l'impact sociétal de l'IA et des considérations éthiques qui doivent être prises en compte dans le développement et le déploiement de ces systèmes.
- Seront capables d'évaluer les applications réelles de l'IA dans les domaines de la parole et de la vision et de reconnaître leur potentiel pour résoudre des problèmes complexes.

## 3.3 Estimation du temps à passer

L'achèvement du module et la mise en œuvre des connaissances acquises prendront environ 5 à 6 heures, en fonction de la quantité de contenu et de quiz fournis dans le module. Toutefois, le temps réel nécessaire aux apprenants pour terminer le module et appliquer les connaissances acquises peut varier en fonction de leur rythme d'apprentissage et de leur familiarité avec le sujet.

# 4 Contenu du cours de l'Unité

## 4.1 Introduction

Bienvenue dans l'unité sur l'application de l'IA à la parole et à la vision. Dans cette unité, nous allons explorer comment l'IA est utilisée pour transformer les domaines de la reconnaissance vocale et de la vision par ordinateur. Nous commencerons par discuter des bases de l'apprentissage automatique et des différentes techniques utilisées dans les applications de la parole et de la vision, y compris l'apprentissage supervisé, non supervisé et par renforcement. Ensuite, nous nous pencherons sur l'interaction naturelle et sur la manière dont l'IA modifie la façon dont nous interagissons avec les machines, y compris l'interaction vocale et visuelle. Nous examinerons également l'impact sociétal de l'IA dans les applications de la parole et de la vision, y compris ses avantages et inconvénients et les considérations éthiques. Enfin, nous mettrons en lumière certaines des études de cas et des réussites les plus passionnantes dans ce domaine, en montrant comment l'IA est utilisée pour résoudre des problèmes concrets dans les domaines de la santé, de l'éducation et du divertissement. À la fin de cette unité, vous aurez une solide compréhension de la façon dont l'IA transforme les domaines de la parole et de la vision et de l'impact potentiel qu'elle peut avoir sur la société dans son ensemble.

## 4.2 Idée 1 : Perception

L'intelligence artificielle (IA) est un domaine qui évolue rapidement et qui transforme nos vies à bien des égards. De la reconnaissance vocale à l'analyse d'images, l'IA fait désormais partie intégrante de notre vie quotidienne. Cependant, enseigner les concepts de l'IA aux enfants peut s'avérer difficile, notamment en raison de la nature complexe de la technologie. Pour relever ce défi, le cadre AI4K12 a été développé comme un ensemble de lignes directrices pour l'enseignement des concepts de l'IA et de ses applications pratiques aux élèves de la maternelle à la terminale. Ce cadre repose sur cinq grandes idées, dont la première est la perception. Dans cette unité, nous explorerons la première grande idée du cadre AI4K12, la perception, qui implique de comprendre comment l'IA perçoit le monde qui nous entoure. On ne saurait trop insister sur l'importance de la perception dans l'IA, car c'est elle qui permet aux machines d'interpréter le monde qui les entoure et d'y répondre d'une manière similaire à celle des humains. Pour percevoir le monde qui les entoure, les systèmes d'IA s'appuient sur une série de capteurs, chacun étant conçu pour collecter des données provenant de différentes sources. Ces capteurs jouent un rôle essentiel en permettant aux machines de comprendre leur environnement et de prendre des décisions sur la base de ces informations. Cette section présente les différents types de capteurs utilisés dans l'IA et leur fonctionnement.

### Exemples de capteurs :

**Microphones** : L'un des types de capteurs les plus courants utilisés en IA est le microphone, qui sert à la reconnaissance vocale. Les microphones captent les ondes sonores et les convertissent en signaux numériques qui peuvent être traités par des algorithmes d'IA. Les assistants vocaux comme Alexa d'Amazon et Siri d'Apple s'appuient sur des microphones pour comprendre les commandes vocales et y répondre.

**Caméras** : Un autre type de capteur essentiel utilisé dans l'IA est la caméra, qui sert à la vision par ordinateur. Les caméras capturent des images ou des trames vidéo et utilisent des algorithmes pour en extraire des informations. Ces informations peuvent inclure la détection d'objets, la reconnaissance faciale et le suivi des mouvements. Les caméras sont utilisées dans de nombreuses applications du monde réel, notamment les systèmes de surveillance, les véhicules autonomes et les caméras de sécurité.

**Capteurs infrarouges** : Les capteurs infrarouges détectent le rayonnement infrarouge, qui est émis par tous les objets dont la température est supérieure au zéro absolu. Ces capteurs sont utilisés dans des applications telles que les systèmes de sécurité domestique et les caméras de vision nocturne. Ils détectent la signature thermique des objets, ce qui permet aux systèmes d'intelligence artificielle de détecter et de suivre les mouvements dans l'obscurité.

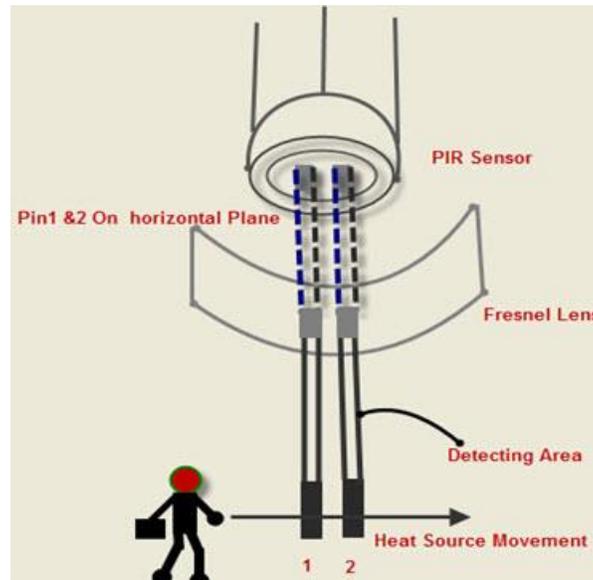


Figure 1: Capteur infrarouge

**Capteurs tactiles :** Les capteurs tactiles sont utilisés dans des applications telles que les écrans tactiles, les pinces robotiques et les prothèses. Les capteurs tactiles peuvent détecter la pression, la température et d'autres propriétés physiques, ce qui permet aux systèmes d'intelligence artificielle d'interagir avec le monde physique.

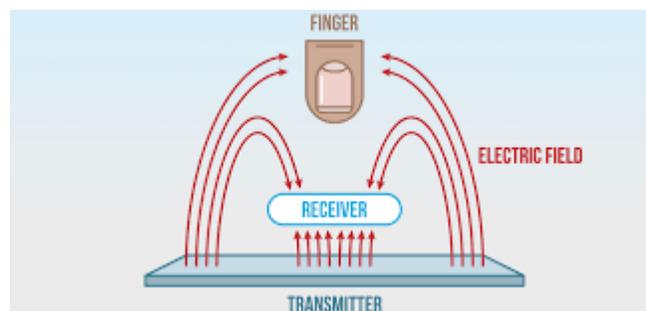


Figure 2: Capteurs tactiles

## 4.3 Idée 2 : Représentation et raisonnement

La représentation fait référence à la manière dont les données sont stockées et traitées par les systèmes d'IA. Cette représentation permet au système d'IA d'analyser et d'extraire les caractéristiques du son, telles que les phonèmes ou les mots. En IA, les données sont représentées de manière à ce que les machines puissent les comprendre et les traiter efficacement. La façon la plus courante dont les données sont représentées dans l'IA est par des symboles ou des nombres. Par exemple, l'image d'un chat peut être représentée sous la forme d'une matrice de nombres correspondant à la couleur de chaque pixel de l'image.

Le raisonnement consiste à tirer des conclusions à partir des données collectées et représentées. Les systèmes d'IA utilisent différentes techniques de raisonnement pour traiter et analyser les données, telles que le raisonnement déductif, le raisonnement inductif et le raisonnement abductif. Ces techniques aident les machines à faire des prédictions, à reconnaître des modèles et à résoudre des problèmes.

### Exemples :

**Reconnaissance de la parole** : Les systèmes d'IA représentent le langage parlé comme une séquence de symboles ou de nombres, puis utilisent des techniques de raisonnement pour interpréter et comprendre le sens du discours. Le système d'IA peut utiliser le raisonnement pour faire correspondre les données audio à une base de données de mots ou de phrases connus, ou pour identifier l'intention d'une commande vocale.

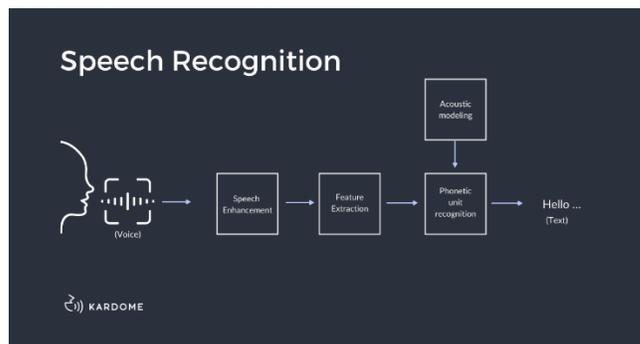


Figure 3: Reconnaissance de la parole

**Vision par ordinateur** : Les systèmes d'IA représentent les images comme des matrices de nombres, puis utilisent des techniques de raisonnement pour reconnaître les objets et les modèles dans les images. Le raisonnement peut impliquer l'utilisation d'algorithmes de reconnaissance d'images pour identifier des objets dans une scène ou pour suivre le mouvement d'objets dans le temps.

**Domaine médical** : L'IA peut analyser de grandes quantités de données sur les patients afin d'identifier des modèles et de prédire des problèmes de santé potentiels. Cela peut aider les médecins à prendre des décisions plus éclairées sur les soins à apporter aux patients.

**Finances** : L'IA peut analyser les tendances et les données du marché afin d'identifier des modèles et de prédire les mouvements futurs du marché. Cela peut aider les investisseurs à prendre des décisions plus éclairées en matière d'investissement et à gérer les risques.

## 4.4 Idée 3 : Apprentissage

L'apprentissage est un aspect essentiel de l'IA et c'est ce qui la distingue de la programmation informatique traditionnelle. Les algorithmes d'apprentissage automatique permettent aux systèmes d'IA d'apprendre et de s'améliorer à partir de leur expérience sans être explicitement programmés. Dans le

contexte des applications vocales et visuelles, l'apprentissage joue un rôle crucial en permettant aux machines de reconnaître des modèles et de faire des prédictions basées sur des données. On ne saurait trop insister sur l'importance de l'apprentissage dans les applications vocales et visuelles. En apprenant à partir de grandes quantités de données, les machines peuvent reconnaître des modèles complexes et faire des prédictions plus précises, ce qui permet d'améliorer les performances dans des applications réelles telles que la reconnaissance vocale, la détection d'objets et la conduite autonome.

### **Les trois principaux types d'apprentissage automatique sont les suivants :**

**L'apprentissage supervisé :** L'apprentissage supervisé est utilisé pour des tâches telles que la reconnaissance de la parole et de l'image. L'apprentissage supervisé est un type d'apprentissage automatique dans lequel l'algorithme est formé sur un ensemble de données étiquetées. L'ensemble de données étiquetées se compose de données d'entrée (par exemple, des images ou des fichiers audio) et d'étiquettes de sortie correspondantes (par exemple, classification de textes ou d'images). L'objectif de l'algorithme est d'apprendre une fonction de correspondance entre les données d'entrée et les étiquettes de sortie. Dans la reconnaissance vocale, l'apprentissage supervisé est utilisé pour entraîner l'algorithme à reconnaître les mots prononcés et à les convertir en texte. Les données d'entrée sont des fichiers audio et les étiquettes de sortie sont les transcriptions correspondantes. L'algorithme est entraîné sur un grand ensemble de données de fichiers audio et de transcriptions étiquetés, et il apprend à reconnaître les modèles et les relations entre les données audio et le texte correspondant. Les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux récurrents (RNN) sont des algorithmes d'apprentissage supervisé couramment utilisés dans les applications vocales et visuelles. Les CNN sont particulièrement bien adaptés aux tâches de reconnaissance d'images, car ils peuvent apprendre automatiquement des représentations hiérarchiques des caractéristiques visuelles. Les RNN sont particulièrement adaptés aux tâches de reconnaissance vocale, car ils peuvent modéliser les dépendances temporelles entre les données audio et le texte correspondant.

**Apprentissage non supervisé :** L'apprentissage non supervisé est un type d'apprentissage automatique qui implique l'apprentissage de modèles et de relations dans les données sans avoir besoin d'étiquettes explicites ou de supervision. L'apprentissage non supervisé est particulièrement utile dans les applications de vision artificielle où de grandes quantités de données visuelles non structurées doivent être analysées pour identifier des modèles et des relations significatifs. Par exemple, les algorithmes de regroupement peuvent être utilisés pour regrouper des images ou des objets similaires dans un ensemble de données sur la base de leurs caractéristiques visuelles, sans qu'il soit nécessaire de procéder à un étiquetage explicite. L'autoencodeur est un autre exemple d'algorithme d'apprentissage non supervisé. Un autoencodeur est un réseau neuronal entraîné à reconstruire les données d'entrée à partir d'une représentation de dimension inférieure de ces données. La partie codeur du réseau réduit les données d'entrée à un espace latent de dimension inférieure, et la partie décodeur du réseau tente de reconstruire l'entrée originale à partir de cette représentation latente. Les autoencodeurs peuvent être utilisés pour des tâches telles que le débruitage d'images, la réduction de la dimensionnalité et l'apprentissage de caractéristiques. Un autre exemple d'algorithme d'apprentissage non supervisé est le réseau accusatoire génératif (GAN).

**Apprentissage par renforcement** : Dans les applications vocales et visuelles, l'apprentissage par renforcement est utilisé dans des domaines tels que la robotique, la conduite autonome et les jeux. Par exemple, l'apprentissage par renforcement peut être utilisé pour apprendre à un robot à naviguer dans un environnement encombré ou pour contrôler les mouvements d'une voiture à conduite autonome. Dans le domaine des jeux, l'apprentissage par renforcement peut être utilisé pour développer des agents capables de jouer à des jeux comme les échecs ou le Go à un niveau surhumain. L'un des algorithmes d'apprentissage par renforcement les plus populaires est l'apprentissage Q, qui utilise une table Q pour stocker la récompense attendue pour chaque action dans chaque état. L'agent met à jour la table Q en fonction des récompenses qu'il reçoit et utilise la table mise à jour pour choisir les actions à venir.

Actions :    ↑    →    ↓    ←

Start	0	0	0	0
Nothing / Blank	0	0	0	0
Power	0	0	0	0
Mines	0	0	0	0
END	0	0	0	0

Figure 4: table Q

## 4.5 Idée 4 : Interaction naturelle

L'interaction naturelle dans l'IA et les robots est la capacité de comprendre et de répondre au langage naturel humain et aux actions, telles que la voix et les gestes. Ceci est particulièrement pertinent dans les applications de parole et de vision, où l'objectif est de créer des interfaces capables de comprendre le langage humain et les repères visuels et de répondre d'une manière compréhensible, intuitive et utile. Les technologies d'interaction vocale et visuelle, telles que la reconnaissance vocale, la reconnaissance gestuelle et la reconnaissance faciale, transforment la façon dont nous interagissons avec les machines et ouvrent de nouvelles possibilités d'applications dans les jeux, la robotique, la sécurité, etc.

**Interaction vocale** : L'interaction vocale est l'utilisation du langage naturel pour communiquer entre les humains et les systèmes d'IA, et elle comprend la reconnaissance vocale et la synthèse vocale. La reconnaissance vocale convertit la langue parlée en texte à l'aide d'algorithmes, tandis que la synthèse vocale génère une parole artificielle à partir de texte à l'aide de systèmes TTS. L'une des principales

applications de l'interaction vocale dans l'IA est les assistants virtuels, tels que Siri d'Apple, Alexa d'Amazon et Google Assistant. Ces assistants virtuels utilisent la reconnaissance vocale et la PNL pour comprendre les demandes des utilisateurs et répondre avec des informations ou des actions pertinentes. L'interaction vocale est également utilisée dans les centres d'appels, où les systèmes vocaux automatisés peuvent acheminer les appels et répondre aux questions de base, libérant ainsi des agents humains pour des tâches plus complexes.

**Interaction visuelle** : l'interaction visuelle fait référence à la manière dont les humains peuvent interagir avec les systèmes d'IA à l'aide d'indices visuels, de gestes et d'autres méthodes de communication non verbales. Cela inclut des technologies telles que la reconnaissance faciale, le suivi du regard et la capture de mouvement. Ces technologies ont de nombreuses applications pratiques, telles que le contrôle des appareils avec des gestes de la main, le suivi des mouvements oculaires et la reconnaissance des expressions faciales et des émotions.

**Réalité virtuelle et augmentée** : La réalité virtuelle et augmentée (VR et AR) sont des technologies utilisées dans l'IA pour créer des interactions immersives entre les humains et les machines. La réalité virtuelle crée un environnement entièrement virtuel à l'aide de casques et de contrôleurs, tandis que la réalité augmentée superpose des objets virtuels à la vue de l'utilisateur sur le monde réel. Ces technologies ont des limites, telles que la nécessité de matériel spécialisé et le risque de mal des transports, mais elles évoluent et ont de nombreuses applications dans des domaines tels que l'éducation, la santé et le divertissement.



Figure 5: Réalité virtuelle



Figure 6: Réalité augmentée

## 4.6 Idée 5 : Impact sociétal

L'impact sociétal de l'IA dans les applications de parole et de vision est devenu un sujet de discussion crucial ces dernières années. Bien que ces technologies aient le potentiel d'apporter des avancées significatives dans divers domaines, elles ont également des implications éthiques, sociales et économiques qui doivent être prises en compte.

Avantages de l'IA dans les applications de la parole et de la vision : l'IA dans la parole et la vision présente des avantages et des inconvénients pour la société. Du côté positif, il a amélioré les soins de santé, l'éducation et la sécurité publique. Par exemple, la reconnaissance vocale aide les médecins à transcrire avec précision les dossiers médicaux et les technologies d'assistance aident les personnes handicapées. Les applications de vision contribuent à la reconnaissance faciale et aux véhicules autonomes. Cependant, des considérations éthiques doivent être prises en compte pour le développement et la mise en œuvre responsables de la technologie de l'IA.

Inconvénients de l'IA dans les applications vocales et visuelles : l'IA dans la parole et la vision présente des inconvénients potentiels. L'une des principales préoccupations est le potentiel de biais et de discrimination, car les algorithmes d'IA peuvent reproduire les préjugés de la société. Il existe également un risque de suppression d'emplois et de disparités économiques en raison du coût élevé des systèmes d'IA. La confidentialité et la surveillance sont des préoccupations éthiques, l'IA collectant des données personnelles. De plus, il existe une préoccupation plus large concernant l'impact à long terme de l'IA, à mesure qu'elle devient plus avancée et autonome. Il est essentiel de répondre à ces préoccupations et de créer des systèmes plus transparents, équitables et responsables qui profitent à la société.

- **Considérations éthiques** : bien que les technologies d'IA aient le potentiel d'apporter des avantages significatifs, plusieurs considérations éthiques doivent également être prises en compte. Certaines des principales questions éthiques associées à l'IA dans les applications Speech and Vision incluent la confidentialité, la transparence, les préjugés et la responsabilité.
- **Confidentialité** : l'utilisation de l'IA dans les applications Speech and Vision soulève des inquiétudes quant à la confidentialité. Avec la quantité croissante de données collectées, il existe un risque que les informations personnelles soient utilisées à mauvais escient ou mal traitées. En tant que tel, il est important que les développeurs et les organisations s'assurent que des protections de confidentialité appropriées sont en place pour protéger les données personnelles.
- **Transparence** : une autre préoccupation éthique liée à l'IA dans les applications de parole et de vision est la question de la transparence. Il peut être difficile de comprendre comment un système d'IA arrive à ses décisions, ce qui peut compliquer l'évaluation de l'équité et de l'exactitude de ses résultats. Pour répondre à cette préoccupation, il est important que les

développeurs accordent la priorité à la transparence de leurs systèmes d'IA et s'assurent qu'ils peuvent être audités et expliqués.

- **Biais** : le biais est un autre problème éthique qui peut survenir dans les applications d'IA dans la parole et la vision. Si les données utilisées pour former un système d'IA sont biaisées, les sorties du système peuvent également être biaisées. Cela peut avoir des implications importantes dans des domaines tels que l'embauche, les prêts et la justice pénale. Pour résoudre ce problème, les développeurs doivent prendre des mesures pour s'assurer que leurs ensembles de données sont représentatifs et impartiaux.
- **Responsabilité** : Enfin, l'utilisation de l'IA dans les applications Speech and Vision soulève des questions de responsabilité. Qui est responsable si un système d'IA commet une erreur ou cause un préjudice ? À mesure que l'IA s'intègre de plus en plus dans nos vies, il est important de s'assurer qu'il existe des lignes claires de responsabilité et de responsabilité en place.

## 4.7 Études de cas et exemples de réussite

Des études de cas présentent des applications pratiques de l'IA dans la parole et la vision. L'IA DeepMind de Google pour la lecture labiale améliore la reconnaissance vocale pour les personnes malentendantes, tandis que Watson pour l'oncologie d'IBM aide les oncologues à créer des plans de traitement du cancer personnalisés. Alexa pour la reconnaissance vocale d'Amazon a révolutionné la façon dont les gens interagissent avec la technologie. Parmi les autres applications d'intelligence artificielle notables dans Speech and Vision, citons l'identifiant de visage d'Apple et le pilote automatique de Tesla. Ces exemples démontrent le potentiel transformateur de l'IA pour améliorer nos vies.

### 4.7.1 Exemple 1 : L'IA DeepMind de Google pour la lecture labiale

DeepMind de Google a créé un système d'IA formé sur 5 000 heures de séquences télévisées qui a atteint un taux de précision de 46,8 % en lecture labiale, dépassant les 12,4 % des experts humains. Il a des applications potentielles dans la reconnaissance vocale, la sécurité et l'aide aux personnes malentendantes. La lecture labiale est difficile, et même les meilleurs lecteurs labiaux humains ne peuvent comprendre avec précision qu'environ 30 % des mots prononcés. L'IA de DeepMind a le potentiel de révolutionner la communication pour les personnes malentendantes. Il existe des problèmes de confidentialité liés à la collecte et à l'analyse de grandes quantités de séquences vidéo, et le potentiel d'utilisation abusive dans des contextes de surveillance et d'application de la loi. Dans l'ensemble, cette technologie démontre comment l'IA peut améliorer la communication pour les personnes malentendantes et souligne la nécessité de considérations éthiques dans le développement et l'utilisation de l'IA.

## 4.7.2 Exemple 2 : Watson d'IBM pour l'oncologie

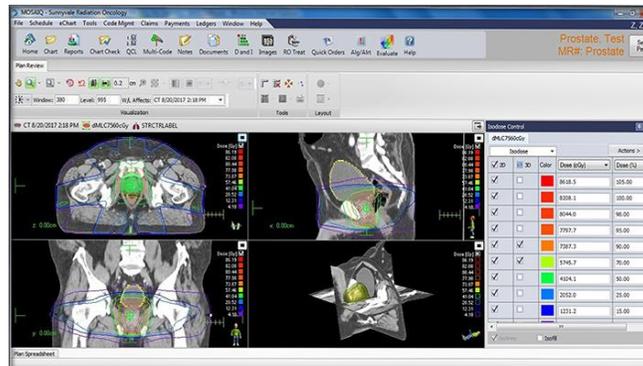


Figure 7: IBM Watson

Watson for Oncology d'IBM est un système d'IA qui aide les oncologues à prendre des décisions de traitement personnalisées pour les patients atteints de cancer. Le système utilise le traitement du langage naturel et l'apprentissage automatique pour analyser de grandes quantités de données médicales et fournir des recommandations de traitement basées sur les circonstances uniques d'un patient. Bien qu'il y ait eu quelques défis et critiques du système, il a été utilisé avec succès dans les établissements de santé et représente un exemple prometteur de la façon dont l'IA peut améliorer les résultats pour les patients.

## 4.7.3 Exemple 3 : Alexa d'Amazon pour la reconnaissance vocale

Alexa d'Amazon est un assistant virtuel populaire qui utilise la technologie de reconnaissance vocale pour permettre le fonctionnement mains libres de divers appareils. Il reconnaît et répond aux commandes en langage naturel et utilise l'apprentissage automatique pour fournir des recommandations personnalisées. Alexa a également aidé les personnes handicapées en leur permettant d'effectuer plus facilement les tâches quotidiennes. Cependant, les problèmes de confidentialité et de sécurité doivent être pris en compte. Le succès d'Alexa démontre le potentiel de la technologie de reconnaissance vocale pour améliorer notre vie quotidienne, mais les considérations éthiques et les risques associés à l'IA doivent être pris en compte.

## 5 Matériel et ressources supplémentaires

Type de ressource	Titre	Sujet	Lien
Article	« Reconnaissance vocale : l'avenir, c'est maintenant »	Reconnaissance de la parole	<a href="https://www.wired.com/story/speech-recognition-future-is-now/">https://www.wired.com/story/speech-recognition-future-is-now/</a>
Article	"Computer Vision : l'avenir de l'intelligence artificielle"	Vision par ordinateur	<a href="https://www.forbes.com/sites/forbestechcouncil/2021/01/05/computer-vision-the-future-of-artificial-intelligence/?sh=39c773b742f4">https://www.forbes.com/sites/forbestechcouncil/2021/01/05/computer-vision-the-future-of-artificial-intelligence/?sh=39c773b742f4</a>
Article	"Comment le Deep Learning change l'avenir du contenu visuel"	L'apprentissage en profondeur	<a href="https://blog.hubspot.com/marketing/deep-learning-visual-content">https://blog.hubspot.com/marketing/deep-learning-visual-content</a>
Article	"Les avantages et les inconvénients de l'IA dans les soins de santé"	L'IA dans la santé	<a href="https://www.healthtechmagazine.net/article/2020/03/pros-and-cons-ai-healthcare">https://www.healthtechmagazine.net/article/2020/03/pros-and-cons-ai-healthcare</a>
Vidéo	"TEDTalk : comment l'IA facilite le diagnostic des maladies"	Avantages de la vision et de la parole dans les soins de santé	<a href="https://www.ted.com/talks/pratik_shah_how_ai_is_making_it_easier_to_diagnose_disease">https://www.ted.com/talks/pratik_shah_how_ai_is_making_it_easier_to_diagnose_disease</a>
Vidéo	"Comment l'IA change notre façon de voir le monde"	Les effets bouleversants du développement et de l'utilisation de l'IA	<a href="https://www.youtube.com/watch?v=BDBTJOGvCv4">https://www.youtube.com/watch?v=BDBTJOGvCv4</a>
Lien	TensorFlow	TensorFlow : une plateforme d'apprentissage automatique open source pour la création de modèles	<a href="https://www.tensorflow.org/">https://www.tensorflow.org/</a>

		d'apprentissage en profondeur	
Lien	OpenCV	OpenCV : une bibliothèque de vision par ordinateur open source pour la création d'applications de vision par ordinateur en temps réel	<a href="https://opencv.org/">https://opencv.org/</a>
Lien	TorchePy	PyTorch : une bibliothèque d'apprentissage automatique open source pour la création de modèles d'apprentissage en profondeur	<a href="https://pytorch.org/">https://pytorch.org/</a>
Lien	Nvidia	Institut d'apprentissage en profondeur NVIDIA	<a href="https://www.nvidia.com/en-us/deep-learning-ai/education/">https://www.nvidia.com/en-us/deep-learning-ai/education/</a>

## 6 Synthèse

En conclusion, l'unité sur l'application de l'IA à la parole et à la vision a couvert divers sujets, notamment le traitement du langage naturel, la vision par ordinateur, l'apprentissage automatique et l'impact sociétal. Nous avons exploré des études de cas telles que DeepMind AI de Google pour la lecture labiale, Watson d'IBM pour l'oncologie et Alexa d'Amazon pour la reconnaissance vocale, qui illustrent le potentiel de l'IA pour résoudre des problèmes du monde réel. Cependant, les considérations éthiques et le potentiel de biais et de limites de l'IA doivent également être soigneusement examinés. Avec ces connaissances, les apprenants peuvent continuer à explorer les vastes opportunités et défis dans le domaine de l'IA.

## 7 Quiz

Question 1 : Vrai/faux

- a) L'apprentissage en profondeur est un sous-ensemble de l'apprentissage automatique. (**Vrai/faux**)  
b) Les biais et le surajustement sont des défis courants dans l'apprentissage automatique. (**Vrai faux**)

Question 2 : Associez les types d'apprentissage automatique suivants à leurs descriptions :

a. Enseignement supervisé	i. Apprendre avec des données étiquetées
b. Apprentissage non supervisé	ii. Apprendre sans données étiquetées
c. Apprentissage par renforcement	iii. Apprendre avec un système de récompense

Réponses : a. i, b. ii, c. iii

Question 3 : Associez les algorithmes d'apprentissage automatique suivants à leurs applications :

a. Réseaux de neurones convolutifs (CNN)	i. Reconnaissance de la parole
b. Auto-encodeurs	ii. Classement des images
c. Q-apprentissage	iii. Classement des images

Réponses : a. ii, b. I, c. iii

Question 4 : Quelle est la principale différence entre l'apprentissage supervisé et non supervisé?

- a) L'apprentissage supervisé nécessite des données étiquetées, contrairement à l'apprentissage non supervisé.**  
b) L'apprentissage non supervisé nécessite des données étiquetées, contrairement à l'apprentissage supervisé.  
c) Les deux types d'apprentissage nécessitent des données étiquetées.  
d) Il n'y a pas de différence entre l'apprentissage supervisé et non supervisé.

Question 5 : Quel type d'apprentissage est utilisé pour la robotique et le jeu ?

- a) Apprentissage supervisé  
b) Apprentissage non supervisé  
**c) Apprentissage par renforcement**  
d) Aucune des réponses ci-dessus

Question 6 : Quelle est la principale limite de l'apprentissage automatique dans les applications de parole et de vision ?

- a) Manque de puissance de calcul
- b) Manque de données étiquetées
- c) Biais et surajustement**
- d) Manque d'interprétabilité

Question 7 : Quel est le principal avantage d'utiliser le Deep Learning ?

- a) Meilleure précision sur les tâches complexes**
- b) Besoins de calcul inférieurs
- c) Moins sujet au surajustement
- d) Plus facile à interpréter

Question 8 : Quelle est la principale considération éthique dans le développement de l'IA pour un impact sociétal ?

- a) Assurer la sûreté et la sécurité des utilisateurs
- b) Garantir l'équité et éviter les préjugés**
- c) Assurer la rentabilité de l'entreprise
- d) Assurer la compatibilité avec les systèmes existants

Question 9 : Quel est le principal avantage de l'utilisation de la réalité augmentée en combinaison avec l'IA ?

- a) Amélioration de l'expérience utilisateur**
- b) Réduction des exigences de calcul
- c) Améliorer la précision des tâches complexes
- d) Rendre l'IA plus interprétable

## 8 Références

- Kelleher, J. D., & Tierney, B. (2018). Data science: An introduction. CRC Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
- Jurafsky, D., & Martin, J. H. (2020). Speech and language processing (3rd ed.). Pearson.
- Zhang, L., Tan, T., & Li, Y. (2016). Deep learning-based face recognition: A survey. International Journal of Automation and Computing, 13(4), 261-279.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484- 489.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.
- Weng, L., Zhang, Y., Xue, X., & Chen, K. (2021). The applications of artificial intelligence in computer vision. In 2021 International Conference on Computer Science and Artificial Intelligence (CSAI) (pp. 139-143). IEEE.

- 
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
  - Dhillon, G., & Moores, D. (2018). Natural language processing in AI: past, present and future. In Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval (pp. 1-10).
  - Canny, J. (2021). Ethics and artificial intelligence. Communications of the ACM, 64(1), 18-20.
  - Li, X., Liang, J., & Li, Y. (2021). Research on the impact of augmented reality technology on the experience of cultural tourism. Journal of Physics: Conference Series, 1838(1), 012064.
  - Yaqoob, I., Ahmed, M. M., Gani, A., Imran, M., Guizani, M., & Hitha, H. (2019). Virtual reality for education:A survey. Journal of Network and Com