IA DANS LES SOINS DE SANTÉ

dwengo

IA L'ECOLE

FICHES D'INFORMATION

Ce projet a été réalisé avec le soutien de la Fondation Roi Baudouin, de la Loterie Nationale et de plusieurs partenaires.

UNIVERSITEIT GENT

provincie Oost-Vlaanderen

Carte	Fiches d'information pertinentes
1	Anonymisation, Authentification, Banques de données, Cryptage, Hachage
2	Anonymisation, Banques de données, Cryptage, Hachage, Traitement du langage
3	Authentification, Capteurs, Communication de données, Cryptage, IoT
4	Authentification, Capteurs, Communication de données, Cryptage, IoT
5	Capteurs, Images digitales, Reconnaissance d'images, Robots sociaux
6	Biomimétisme, Robots sociaux
7	Biomimétisme, Capteurs, Parole, Robots sociaux, Traitement du langage
8	Décisions, Parole, Robots sociaux, Traitement du langage
9	Capteurs, Images digitales, Reconnaissance d'images, Robots sociaux, Parole
10	Capteurs, Images digitales, Reconnaissance d'images, Robots sociaux, Parole
11	Décisions, Traitement du langage
12	Authentification, Capteurs, Communication de données, Cryptage, IoT
13	Capteurs, Communication de données
14	Capteurs, Images digitales, Reconnaissance d'images, Robots sociaux, Parole
15	Décisions, Images digitales, Parole, Reconnaissance d'images, Robots sociaux
16	Cryptage, Décisions, Hachage, Traitement du langage
17	Cryptage, Hachage, Parole, Traitement du langage
18	Banques de données, Communication de données, IoT, Robots sociaux
19	Capteurs, Images digitales, Parole, Reconnaissance d'images, Robots sociaux
20	Biomimétisme, Capteurs, Parole, Robots sociaux
21	Banques de données, Images digitales, Reconnaissance d'images
22	Parole, Robots sociaux, Traitement du langage
23	Images digitales, Parole, Robots sociaux, Traitement du langage
24	Images digitales, Parole, Reconnaissance d'images, Robots sociaux
25	Banques de données, Décisions, Tracking
26	Décisions, Traitement du langage
27	Anonymisation, Parole, Traitement du langage
28	Authentification, Banques de données, Cryptage, Décisions, Hachage
29	Décisions, Parole, Reconnaissance d'images, Robots sociaux
30	Biomimétisme, Capteurs, Images digitales, Reconnaissance d'images
31	Biomimétisme, Capteurs, Images digitales, Reconnaissance d'images, Robots sociaux
32	Capteurs, Décisions, Parole, Reconnaissance d'images, Traitement du langage
33	Biomimétisme, Capteurs, Robots sociaux

Anonymisation des données

Tu connais peut-être quelqu'un qui a déjà participé à une étude clinique. Ces études sont faites pour tester l'efficacité d'un médicament ou d'un vaccin. Pendant l'étude, on te demande généralement tes ***données personnelles***. Ce sont des ***données qui permettent de t'identifier***. Tu donnes par exemple ton ***nom et ton adresse e-mail*** afin que les chercheurs puissent te contacter pour te communiquer des informations relatives à l'étude. Outre ces coordonnées, les chercheurs ***collectent aussi souvent des données nécessaires à l'étude***. Il s'agit généralement d'informations très personnelles, comme ***la taille, le poids, le groupe sanguin, les allergies***... Pour garantir la protection de ta vie privée, les chercheurs veillent à ce que ces données ne puissent pas être reliées à toi. Tes coordonnées comme ton adresse e-mail, ton nom, ton numéro de téléphone, ton adresse... ne peuvent donc pas être enregistrées avec les données de l'étude. Autrement dit, ***les données de l'étude sont anonymisées***.

Pourquoi faut-il apprendre cela ? Les entreprises et les écoles ont besoin de certaines de tes données, par exemple pour livrer des colis ou envoyer des factures. En comprenant comment fonctionne l'anonymisation, tu comprendras comment ils veillent à ce que tes informations restent en sécurité et tu apprendras comment protéger ta propre vie privée.

Défi : il n'est pas toujours simple d'anonymiser correctement des données. Tu vois ci-dessous deux tableaux. Ils contiennent les données des participants à une étude sur l'efficacité d'un vaccin. Un tableau contient les coordonnées des participants, l'autre les données anonymes nécessaires à l'étude. Peux-tu relier les personnes avec leurs données pour l'étude ?

Les coordonnées

Prénom	Nom	Numéro de téléphone	Adresse	E-mail	Mineur	E-mail du tuteur
Emma	De Vries	555123456	Mechelsesteenweg 10, Anvers	emma.devries@mail.com	Non	
Finn	Janssen	555123457	Brusselsestraat 15, Louvain	finn.janssen@mail.com	Oui	jan.janssen@mail.com
Sophie	Vermeulen	555123458	Kerkstraat 25, Gand	sophie.vermeulen@mail.com	Oui	karel.vermeulen@mail.com
Gustaaf	Peters	555123459	Maison de repos Regina Celi, Rijlaan 21, Hasselt, Chambre 67	Gustaaf.peters@mail.com	Non	
Finn	Bosch	555123461	Lange Munt 30, Gand	finn.bosch@mail.com	Non	
Ginette	Smits	555123462	Zeedijk 12, bus 20, Ostend	ginette.smits@mail.com	Non	

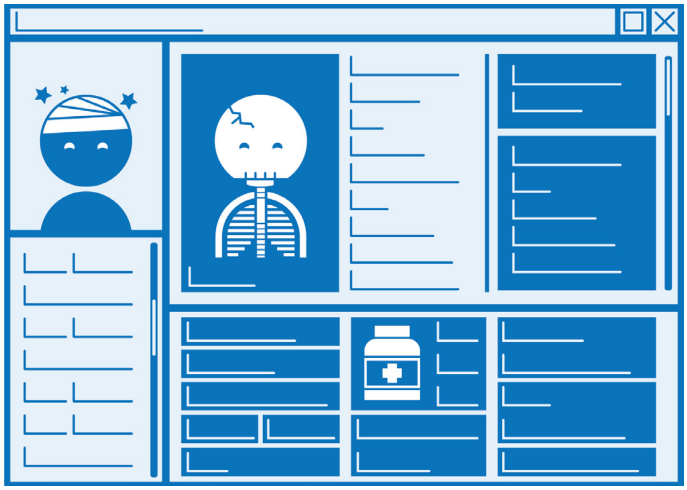
Données pour l'étude

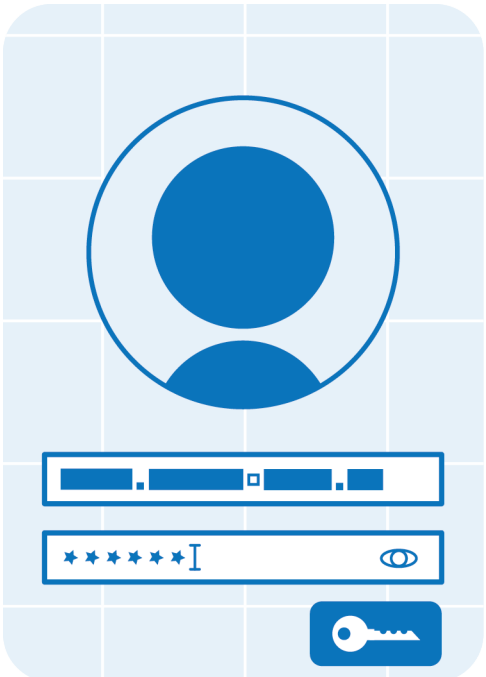
Âge	Genre	Groupe sanguin	Allergies	Maladies
12	Femme	O+	Aucune	Pipi au lit
13	Homme	B+	Aucune	Aucune
19	Homme	O-	Aucune	Diabète type 2
33	Femme	A+	Pollen	Maladie de Crohn
79	Femme	B-	Aucune	Aucune
82	Homme	AB+	Arachides	Incontinence

Authentification et autorisation

Tu te connectes chaque jour à différentes plateformes comme Smartschool, Instagram, TikTok, OneDrive... Mais t'es-tu déjà demandé pourquoi il était utile de se connecter à ces plateformes ?
La connexion peut être nécessaire pour deux raisons importantes : l'authentification et l'autorisation. ***L'authentification est nécessaire pour que la plateforme à laquelle tu te connectes sache qui tu es***.
Lorsque la plateforme sait qui tu es, elle te donne accès à certaines données (ex. les informations de ton profil Facebook). ***Le choix des données auxquelles tu peux accéder s'appelle l'autorisation***. L'autorisation est nécessaire parce que tu ne peux pas accéder aux données que tu n'as pas le droit de voir, comme les données du profil Facebook d'une personne avec laquelle tu n'es pas ami.e.

L'authentification peut se faire de différentes manières, par exemple en combinant un nom d'utilisateur et un mot de passe unique, en utilisant un document que tu es seul.e à posséder (ex. ta carte d'identité électronique) ou via une application d'authentification (ex. itsme). ***L'autorisation*** peut se faire, par exemple, en enregistrant dans un tableau qui a accès à quel fichier.





L'authentification et l'autorisation revêtent une importance capitale dans le ***secteur de la santé***. De plus en plus d'***informations sur les patients*** sont stockées sous forme digitale, par exemple dans un dossier médical partagé. À ta demande, ta ou ton médecin généraliste enregistre toutes tes données médicales dans un dossier médical global (DMG). Ce DMG contient un aperçu de toutes tes données médicales. Certaines de ces données peuvent être échangées via un dossier médical partagé. Différents soignants peuvent consulter ou modifier ce dossier médical partagé. Tu ne

veux toutefois pas que tous les soignants puissent lire toutes les informations contenues dans ton dossier médical partagé. Il est donc primordial que seules les personnes qui doivent avoir accès à ces informations personnelles y aient effectivement accès. Pour cela, il est nécessaire que l'application du dossier médical partagé puisse savoir qui demande certaines informations (***authentification***) et donner à ces personnes uniquement accès aux données dont elles ont besoin pour effectuer leur tâche médicale (***autorisation***). Ainsi, un.e kiné pourra ajouter des informations sur un traitement, mais pas accéder aux résultats d'une prise de sang.

Défi : crée un fichier sur Google Drive, Microsoft OneDrive ou une autre plateforme similaire. Partage ce fichier avec deux personnes. Une personne doit pouvoir lire le fichier mais pas le modifier, tandis que l'autre doit pouvoir le modifier.

*L'authentification et l'autorisation sont souvent associées au **cryptage** et au **hachage**. Le cryptage est utilisé pour protéger les données d'un.e utilisateur.rice. Le hachage peut être utilisé pour vérifier l'identité d'un.e utilisateur.rice. Consulte les fiches sur le cryptage et le hachage pour en savoir plus sur ces techniques.*

Banques de données

En 2024, plus de ***50 zettaoctets*** (Zo) de données ont été stockées dans le monde. Un zettaoctet équivaut à 10^{21} octets. Cela signifie que ***6,25 téraoctets (To) de données sont stockés en moyenne par personne sur terre***. Cette quantité de données augmente très rapidement d'année en année. On estime que la quantité de données stockées doublera environ tous les deux ans.

Pour gérer ces énormes quantités de données, nous avons besoin de techniques pour stocker les données de manière ***compacte*** tout en ***permettant d'effectuer efficacement des recherches***. Les informaticiens ont mis au point plusieurs techniques et les ont rassemblées sous la forme d'un ***système de banque de données***. Un ***système de banque de données***, ou, plus simplement, une ***banque de données*** est donc un ensemble de techniques qui nous permettent de stocker des données de manière compacte tout en permettant d'effectuer efficacement des recherches.

Voici quelques exemples de ces techniques : suppression des données superflues, tri des données et recherche binaire de données.

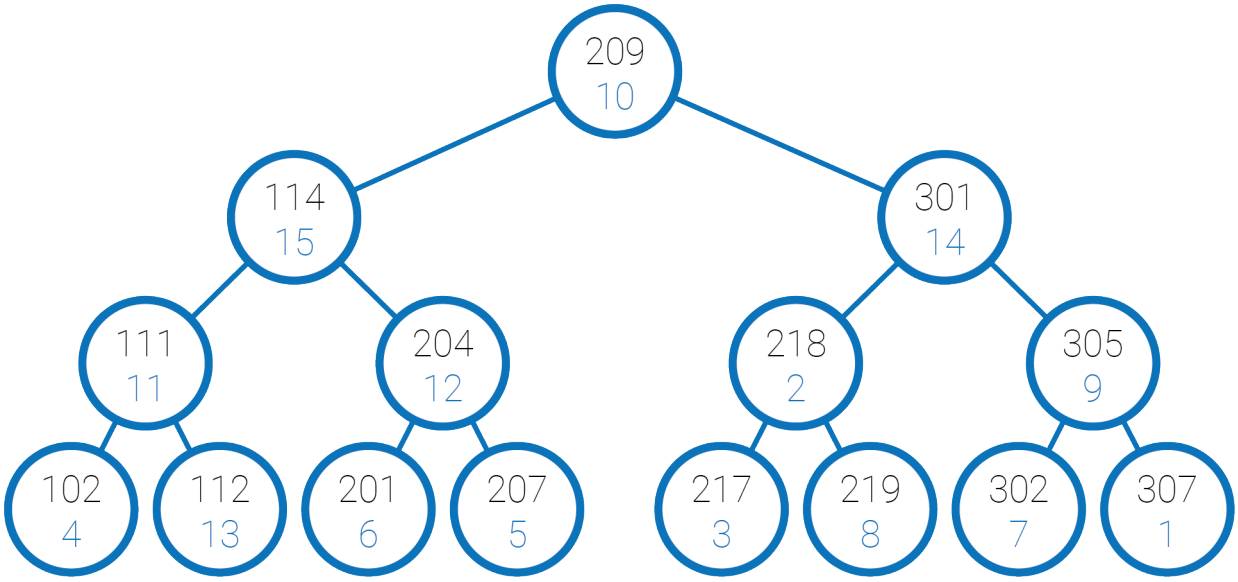
L'***indexation*** des données est une technique importante des systèmes de banque de données. Elle permet de trouver les informations plus rapidement. On peut la comparer à un index situé au début ou à la fin d'un livre, qui permet de savoir facilement sur quelles pages apparaît un terme.

Ce tableau contient les numéros de chambre et les patients d'un hôpital. Regarde le tableau. Quel patient a 66 ans ?

N°	Nom	Prénom	Âge	Chambre
1	Berendzen	Douwe	40	307
2	Burckhard	Karl	41	218
3	Claussen	Lies	62	217
4	De Ronde	Huib	31	102
5	Eyman	Eelco	64	207
6	Halle	Teun	32	201
7	Henneke	Arnoldus	44	302
8	Hinken	Marie	51	219
9	Jagers	Wim	70	305
10	Koen	Felix	58	209
11	Roode	Klaus	37	111
12	Rooth	Wijnand	66	204
13	Rosh	Amira	41	112
14	Robrecht	Rein	51	301
15	Veerkamp	Lukas	41	114

Tu trouveras ci-dessous un index, représenté sous la forme d'un arbre binaire. Grâce à cet arbre, tu peux effectuer une recherche rapide dans la liste. Cet arbre permet de retrouver facilement quel patient se trouve dans quelle chambre, sans avoir à parcourir toute la liste des patients.

L'arbre se compose de cercles (nœuds) et de connexions entre ces cercles (arcs). Un ***nœud*** contient le ***numéro de chambre*** avec, en dessous, le ***numéro du patient***. Tu peux retrouver quel patient se trouve dans une chambre en partant du nœud supérieur de l'arbre (la racine). Dans ce nœud, trois actions sont possibles.
1) Le numéro de chambre que tu cherches ***correspond*** au numéro de chambre de ce nœud. Dans ce cas, le numéro du patient que tu cherches se trouve également dans ce nœud. Il est donc inutile de chercher plus loin.
2) Le numéro de chambre que tu cherches est ***plus petit*** que celui de ce nœud. Passe au nœud suivant par l'arc de gauche. Repasse en revue ces trois actions pour ce nœud.
3) Le numéro de chambre que tu cherches est ***plus grand*** que celui de ce nœud. Passe au nœud suivant par l'arc de droite. Repasse en revue ces trois actions pour ce nœud.



Défi : utilise l'arbre binaire pour trouver quel patient se trouve dans la chambre 204. Combien de numéros de chambre as-tu dû regarder avant de trouver le nom du patient ? Combien aurais-tu dû en regarder si tu avais parcouru le tableau ligne par ligne ? Imagine que la liste contienne 1024 noms, combien de numéros de chambre au maximum aurais-tu dû parcourir dans l'arbre binaire ?

Le savais-tu ?
Pour créer cet arbre, on a appliqué le concept de recherche binaire. Toutes les chambres sont ordonnées de gauche à droite : 102, 111, 112, 114, 201, 204, ..., 219, 301, ..., 307. La chambre du milieu est choisie comme racine de l'arbre. Ensuite, pour les nœuds suivants, on choisit la chambre du milieu de la partie gauche et de la partie droite, etc.

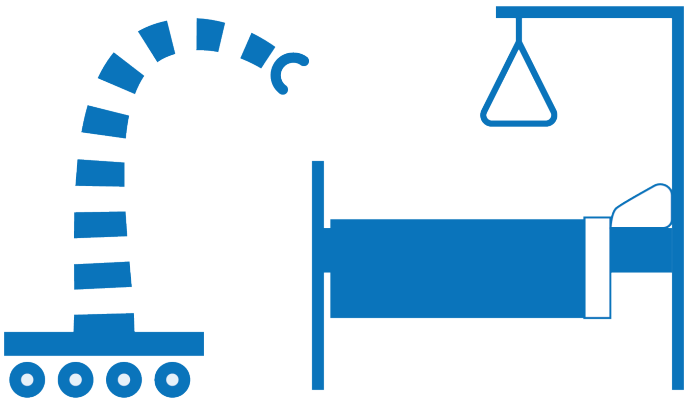
Biomimétisme

Le personnel infirmier et soignant doit souvent soulever des patients tous les jours. Sortir un.e patient.e de son lit, le mettre dans une chaise roulante, l'aider à aller aux toilettes... toutes ces tâches requièrent un important effort physique de la part des soignants. La technologie essaie de les soutenir autant que possible. Il existe par exemple des lève-personnes, des harnais de levage et des couvertures de transfert qui tentent d'alléger le travail du personnel soignant. Malgré ces équipements, le personnel soignant doit toujours fournir d'importants efforts physiques pour s'occuper correctement des patients.

Les ingénieurs élaborent de nouveaux systèmes pour répondre à ces besoins. Ils s'inspirent souvent de la nature, imitant en quelque sorte les solutions qu'elle propose pour résoudre certains problèmes. Les ingénieurs cherchent par exemple à savoir s'il est possible de construire des robots capables de soulever des patients sans leur faire mal. Ces robots doivent être suffisamment forts pour porter un.e patient.e, mais suffisamment souples et flexibles pour ne pas lui faire mal.

Dans le laboratoire AIRO « intelligence artificielle et robotique » du professeur Francis Wyffels, on développe des robots inspirés de la queue d'un hippocampe. Ce robot pourrait par exemple soulever des personnes, mais aussi les soutenir ou leur apporter un verre d'eau.

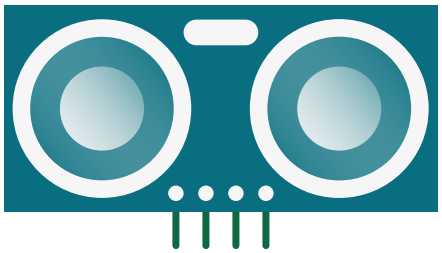
Les hippocampes ont une queue préhensile, non seulement solide mais aussi agile. Cette queue intéresse les ingénieurs qui veulent développer des systèmes préhensiles innovants en robotique, aux applications multiples. Pour rendre les robots plus agiles, ils s'inspirent des parties biologiques de l'anatomie de la queue, de la manière dont elles permettent aux hippocampes de saisir quelque chose avec leur queue et de la façon dont ils contrôlent ce mouvement (Source : EOS Special, Iedereen Uitvinder, 2024).



Défi : connais-tu d'autres applications de biomimétisme ? Et retrouves-tu ces applications dans ta vie quotidienne ?

Capteurs

De nos jours, on trouve des capteurs partout. Il en existe de différentes formes et tailles. Pense par exemple au ***détecteur de mouvement*** qui commande l'éclairage, au ***capteur de vitesse*** dans un radar ou au ***capteur en lumière*** dans ton smartphone, qui ajuste automatiquement la luminosité de l'écran. Il existe donc de nombreux types de capteurs, mais quelle propriété fondamentale ces capteurs ont-ils en commun ? ***Chaque capteur convertit un phénomène physique (ex. la température, la lumière, la glycémie) en un signal électrique***. Un ordinateur peut ensuite mesurer ce signal électrique, puis traiter, enregistrer et analyser les valeurs.



Les capteurs sont souvent bon marché, c'est pourquoi ils sont intégrés dans de nombreux produits. Toutefois, si tu souhaites une grande précision ou une large plage (ex. de quelques mg à quelques kg), le capteur peut coûter cher. Les ***voitures électriques*** modernes, par exemple, renferment souvent ***plus de 100 capteurs***. Ils mesurent des propriétés telles que la pression des pneus ou la température de la batterie.

Le secteur de la ***santé*** a lui aussi largement recours aux capteurs, par exemple pour surveiller ***les constantes vitales d'un patient***. Mais cela ne se limite pas à enregistrer des propriétés telles que le rythme cardiaque, la pression artérielle et la température corporelle à l'aide de capteurs ; les chercheurs travaillent également d'arrache-pied sur les ***biocapteurs*** capables de détecter la présence de certaines molécules. L'un des biocapteurs les plus connus mesure la glycémie et peut aider les patients diabétiques dans leur traitement. D'autres types de biocapteurs peuvent être utilisés pour détecter la présence de certains virus dans l'air.

En plus de mesurer un phénomène physique, de nombreux capteurs traitent déjà partiellement ces valeurs mesurées. Le capteur filtre par exemple le bruit ou convertit la valeur mesurée en unités SI. Cela permet d'interpréter plus facilement la valeur du capteur. Le traitement de la valeur mesurée peut être très complexe. La start-up gantoise IntelliProve crée par exemple une app qui te permet de mesurer un rythme cardiaque à l'aide de l'appareil photo de ton smartphone. Cet appareil photo se compose de millions de capteurs de lumière qui enregistrent la lumière qui arrive sur ton appareil photo. Ces capteurs de lumière sont de trois types, chacun étant sensible à une couleur de lumière différente (lumière rouge, verte et bleue). Ton smartphone convertit ces valeurs mesurées en pixels avec une valeur RGB[1]. L'app de la start-up gantoise analyse ensuite les valeurs RGB des images successives. IntelliProve a réussi à détecter avec suffisamment de précision une variation de la couleur de la peau sur ces images pour pouvoir mesurer, à l'aide de l'app, le rythme cardiaque d'une personne rien qu'en la filmant.

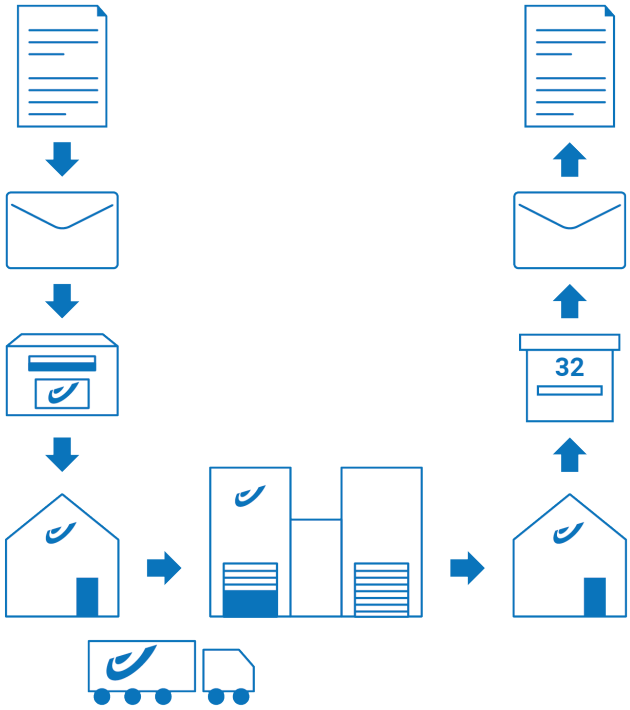
Défi : tu connais peut-être quelqu'un qui a une montre connectée. Quels capteurs intègre cette montre ?

1: Pour en savoir plus sur les valeurs RGB, consultez la fiche sur les images digitales.

Communication de données

T'es-tu déjà demandé comment le texte de tes e-mails arrive sur l'ordinateur du destinataire? Ou comment une photo prise par l'un.e de tes ami.e.s en voyage arrive sur ton smartphone? Plusieurs techniques issues de l'informatique se cachent derrière ces deux applications quotidiennes. Pour garder une vue d'ensemble de ce qui se passe lors de l'envoi de données, ***les informaticiens répartissent ce processus en couches ; chaque couche étant responsable d'une partie spécifique de la communication***. Le message sera par exemple converti en bits puis divisé en paquets.

On peut comparer cela aux différentes personnes nécessaires à l'envoi d'une lettre. Tu dois écrire la lettre. Quand elle est écrite, tu la glisses dans une enveloppe que tu déposes dans une boîte aux lettres. Un.e préposé.e sort le courrier de la boîte aux lettres et l'emmène au bureau de poste. Là, la lettre est prise en charge par un camion et emmenée au centre de tri. De là, la lettre est emmenée dans un autre bureau de poste où un facteur ou une factrice trie le courrier pour la tournée. Cette personne dépose ensuite la lettre dans la boîte aux lettres du destinataire. Ce dernier peut alors ouvrir et lire la lettre.



Que se passe-t-il, par exemple, lorsque tu envoies un e-mail à partir de ton ordinateur ? Comme toujours sur un ordinateur, le programme de messagerie transmet le texte que tu tapes au système d'exploitation ; cela se fait sous la forme de nombres binaires.

Le système d'exploitation divise le message en paquets distincts. Le nombre de paquets dépend de la longueur de ton e-mail. Le système d'exploitation ajoute des informations supplémentaires aux paquets. Ces paquets reçoivent par exemple chacun un numéro qui détermine l'ordre des paquets afin de pouvoir être remis dans le bon ordre de l'autre côté. Le système d'exploitation ajoute aussi à qui le message doit parvenir. Les paquets sont ensuite transmis un par un à la carte réseau de ton ordinateur. Celle-ci convertira les nombres binaires du paquet en signaux électriques envoyés via un câble réseau ou un routeur sans fil. Du côté du destinataire, la carte réseau reconvertit le signal électrique en nombres binaires et les transmet au système d'exploitation. Ce système d'exploitation attend d'avoir reçu tous les paquets et les rassemble en un grand nombre binaire. Ce

nombre est transmis à un programme de messagerie, qui le convertit en texte de l'e-mail.

Tu as sans doute déjà entendu parler du morse. Ce code permet d'envoyer des messages, composés de lettres, de signes de ponctuation et de chiffres, en envoyant deux types de signaux de durée différente (courte ou longue). Il s'agit donc d'un code binaire[1].
Ce tableau te montre une partie de l'alphabet morse.
Le morse comporte deux symboles : les points et les tirets. La longueur du signal avec lequel un point est envoyé détermine la vitesse à laquelle le message est transmis et sert d'« unité ». Par convention, un tiret dure trois fois plus longtemps qu'un point. Les espaces entre les points et les tirets d'un code de lettres ont la longueur d'un point. Les espaces entre les lettres d'un mot ont la longueur de 3 points, et les espaces entre les mots ont la longueur de 7 points.

Lettre	Morse	Lettre	Morse
A	• –	N	– •
B	– • • •	O	– – –
C	– • – •	P	• – – •
D	– • •	Q	– – • –
E	•	R	• – •
F	• • – •	S	• • •
G	– – •	T	–
H	• • • •	U	• • –
I	• •	V	• • • –
J	• – – –	W	• – –
K	– • –	X	– • • –
L	• – • •	Y	– • – –
M	– –	Z	– – • •

Chiffre	Morse	Chiffre	Morse
0	– – – – –	5	• • • • •
1	• – – – –	6	– • • • •
2	• • – – –	7	– – • • •
3	• • • – –	8	– – – • •
4	• • • • –	9	– – – – •

Défi : tapote le message suivant sur ton banc à l'aide de ton crayon : SOS SOS SOS.

Tu connais à présent le principe des couches de la communication réseau. Dans la pratique, c'est bien plus complexe que dans l'exemple donné à propos de l'e-mail. En réalité, on utilise le ***modèle OSI*** *(Open Systems Interconnection model). Tu trouveras sur Wikipédia les couches qui composent le modèle OSI.*

1: https://fr.wikipedia.org/wiki/Code_Morse_international

Cryptage

Tu enregistres chaque jour des données personnelles sur ton ordinateur et dans le cloud.

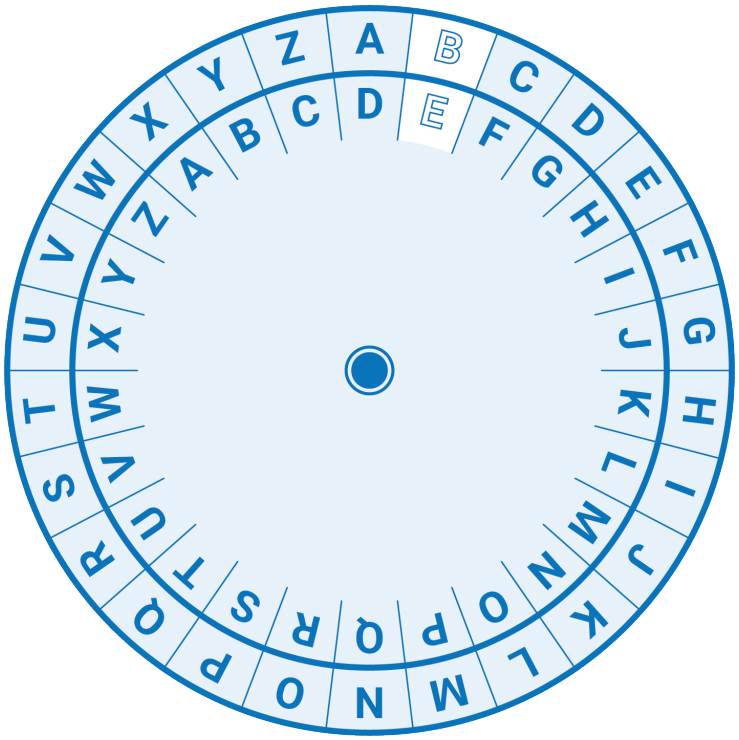
Comment veiller à ce que personne d'autre ne puisse lire tes données personnelles ? Comment te protéger contre l'utilisation abusive de tes données ?

Les informaticiens ont imaginé des techniques pour s'assurer que toi seul.e, ou une personne en qui tu as confiance, a la possibilité de lire tes données. On appelle ces ***techniques de*** stockage sécurisé des données ***« techniques de cryptage »***.

Lorsque tu ***cryptes*** des données, tu les transformes en une représentation difficile à déchiffrer, par le biais d'une formule. Pour ***décrypter*** le message et le rendre à nouveau lisible, tu dois connaître la formule dans son intégralité.

Un exemple simple de cryptage est le ***chiffre de César***, également connu sous le nom de ***code César***. Cette technique de cryptage était déjà utilisée par Jules César pour communiquer par écrit en langage secret avec ses généraux. Le chiffre de César indique de combien de places dans l'alphabet tu déplaces chaque lettre du message.

Tu vois ci-dessous, par exemple, un cryptage avec le chiffre de César 4.



Défi : sais-tu déchiffrer le message suivant ?

Jgty tuzxk sutjk jomozgr, rk inollxksktz kyz kyyktzokr !

Astuce : ce message a été crypté avec le chiffre de César 6.

Avec le chiffre de César, tu peux découvrir assez facilement comment fonctionne le cryptage. Les techniques modernes de cryptage sont conçues pour qu'il soit pratiquement impossible de découvrir à quel texte correspond un message crypté. Ces techniques utilisent une ***clé de cryptage***. Tu as besoin de cette clé pour crypter le message mais aussi pour le rendre à nouveau lisible. La clé de cryptage fait partie de la formule mathématique qui rend les données illisibles. Tant que tu es seul.e à connaître cette clé, personne d'autre ne peut lire le message crypté.

*L'une des techniques de cryptage les plus connues et les plus utilisées est **l'algorithme AES**, également connu sous le nom d'**algorithme Rijndael**. **AES** signifie **Advanced Encryption Standard** (norme de cryptage avancée). Cette norme internationale est utilisée, entre autres, par la National Security Agency (NSA) des États-Unis pour crypter les informations les plus sensibles du gouvernement américain.*

L'algorithme AES a été inventé par deux informaticiens belges, Joan Daemen et Vincent Rijmen. Ils ont développé l'algorithme entre 1997 et 2000 et l'ont présenté lors d'un concours organisé par l'Institut national américain de normalisation et de technologie. Les deux Belges ayant gagné le concours, leur algorithme est devenu une norme internationale et il permet de sécuriser les informations personnelles de milliards de personnes.

Hachage

Le ***hachage*** – ou ***hashing*** en anglais – est utile lorsque, par exemple, tu veux enregistrer un mot de passe et empêcher que d'autres personnes puissent le lire. Le hachage convertit le texte, c'est-à-dire le mot de passe, en un ***grand nombre***. Ce nombre est enregistré à la place du mot de passe lui-même. De plus, le nombre haché ne peut être reconverti en texte original.

Tu sais probablement que noter un mot de passe est une mauvaise idée. À partir du moment où tu écris un mot de passe physiquement ou digitalement, il existe un risque que d'autres personnes puissent lire ton mot de passe. Pour éviter que cela se produise, mieux vaut retenir tes mots de passe. Cela soulève naturellement la question suivante : comment les sites web et les applications peuvent-ils enregistrer ton mot de passe sans qu'il soit noté (digitalement) quelque part ? Grâce au ***hachage***.

Le ***hachage*** est une technique qui permet de convertir un ***texte***, par exemple un mot de passe, en un ***grand nombre*** grâce à un ***algorithme***. Cependant, contrairement au cryptage, le hachage se fait à sens unique. Il est donc impossible de reconvertir le nombre en texte original. En enregistrant non pas le mot de passe, mais le nombre haché, il n'y a aucun risque que des personnes puissent lire ton mot de passe. Au pire, elles peuvent lire le nombre, mais sans le convertir en ton mot de passe. Cela signifie également qu'au moment où tu saisis ton mot de passe, le site web ou l'application va d'abord hacher le mot de passe saisi, puis comparer le nombre obtenu avec celui qui est enregistré.

Une méthode simple pour hacher un texte consiste à ***convertir les lettres en nombres***, à ***multiplier tous ces nombres entre eux***, puis à ***calculer le reste*** en divisant le nombre obtenu par la valeur maximale de hachage.

Voici un exemple de la façon de procéder avec les cinq premières lettres de l'alphabet et une valeur de hachage maximale de 100 :

A	***B***	***C***	***D***	***E***
2	3	5	7	11

Ce tableau associe un nombre à chaque lettre. Tu calcules le hachage du texte ***ACCE*** en faisant le calcul suivant :

(2 x 5 x 5 x 11) mod 100
= 550 mod 100
= 50.

Le tableau de droite montre la valeur de hachage de plusieurs mots de passe. Note que plusieurs mots de passe peuvent avoir la même valeur de hachage, ce qui garantit que tu ne seras jamais sûr.e à 100 % de retrouver le mot de passe original à partir d'une valeur de hachage. Dans la pratique, la probabilité d'avoir des valeurs de hachage identiques est toutefois très faible.

Note que cette méthode utilise des ***nombres premiers***. La multiplication des nombres premiers entre eux est très facile pour un ordinateur. L'opération inverse, à savoir trouver les facteurs premiers d'un très grand nombre (en pratique plus grand que 10^{600}) est très difficile.

Mot de passe	***Hachage***
DAACD	80
CBDAB	30
DBCEB	65
ACBEA	60
ACCDD	50
EEDBC	5
BCCDC	25
CEBAE	30
BABCA	80
CDECE	75

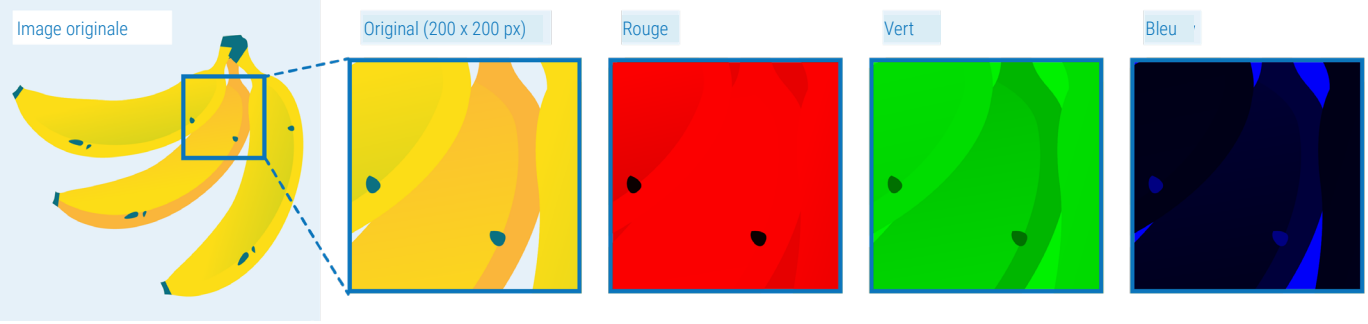
Défi : sais-tu calculer le hachage du mot de passe BBCDAA ?

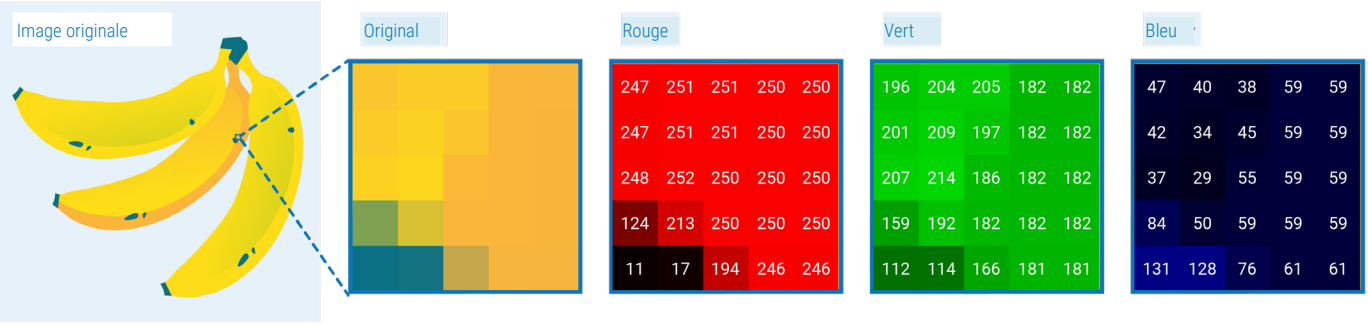
Les algorithmes modernes de hachage fonctionnent sur un principe similaire mais sont beaucoup plus sûrs et bien plus complexes.
L'un des algorithmes de hachage les plus utilisés est le ***Secure Hash Algorithm (SHA256)****. Cet algorithme convertit un texte en un nombre binaire de 256 bits. Il y a donc $2^{256} \approx 10^{77}$ valeurs possibles pour notre hachage. Note qu'il y a environ 10^{80} protons et neutrons dans l'univers observable.*

Images digitales

Un ordinateur représente les images digitales sous la forme d'une ***grille de pixels***. Le mot ***pixel*** vient de la description ***picture element***, que l'on peut traduire par ***élément d'une image***. Chaque pixel a une ***couleur*** particulière, ***représentée par une combinaison de nuances de rouge, de vert et de bleu (RGB pour Red, Green, Blue)***. L'ordinateur représente généralement un pixel par trois nombres allant de 0 à 255, un nombre pour chaque couleur.

Les images ci-dessous montrent à chaque fois sur la gauche une image de bananes sur laquelle une zone est indiquée par un cadre rouge. Les valeurs RGB de cette zone sont indiquées sur la droite.





Un ordinateur peut détecter bien mieux qu'un humain de très petits changements dans une photo. Un ordinateur est en effet capable de « voir » chaque pixel séparément. C'est pourquoi les radiologues peuvent compter sur les systèmes d'IA pour les aider à examiner et à interpréter plus en détail les ***images médicales***.

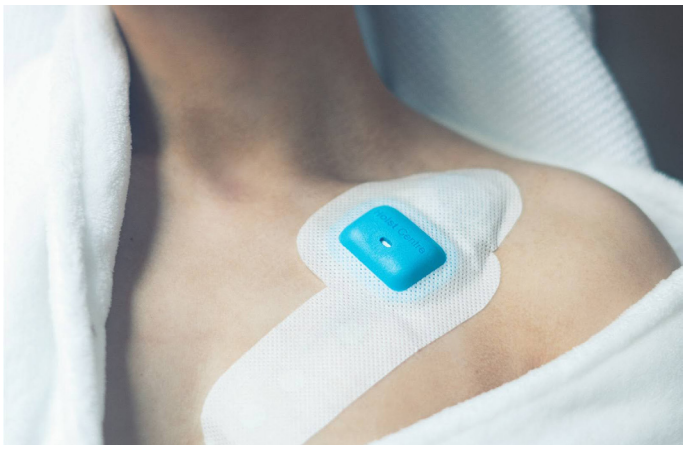
Défi : détecte les pixels ! Chaque écran utilise des pixels pour représenter des images. Utilise l'appareil photo d'un smartphone pour photographier l'écran de ton ordinateur en te rapprochant le plus possible de l'écran, tout en veillant à ce que le contenu de l'écran soit bien net. Ensuite, zoome sur la photo avec le smartphone. Vois-tu les pixels ?

Internet des choses

Presque tous les appareils électriques contiennent aujourd'hui de l'électronique. C'est notamment le cas d'un lave-vaisselle, d'une porte automatique ou du thermostat de ton chauffage.
L'électronique contenue dans ces appareils a de plus en plus souvent la possibilité de se connecter à internet. Les appareils sont ainsi en mesure de communiquer avec toi et avec d'autres appareils. Tu peux par exemple recevoir une notification sur ton téléphone lorsque ton lave-vaisselle a fini son programme ou, grâce à l'électronique, tu peux faire en sorte que le lave-vaisselle ne démarre que lorsque tes panneaux solaires génèrent suffisamment d'énergie. ***Tous ces appareils connectés à internet forment ensemble « l'internet des choses» ou Internet of Things (IoT)***.

Cette électronique est chargée de piloter l'appareil. Dans un lave-vaisselle, par exemple, l'électronique veille à ce que l'eau chauffe à la bonne température et que la pastille de lavage tombe dans la machine au bon moment. Dans une porte automatique, l'électronique interprète ce que mesure un détecteur de mouvement et ouvre la porte sur la base de ces informations.

Dans le ***monde médical*** aussi, de plus en plus d'appareils sont connectés à internet. Un.e radiologue peut par exemple utiliser un échographe qui envoie automatiquement les clichés médicaux à une banque de données. Ou un.e infirmier.e peut recevoir une notification en cas de problème avec les constantes vitales mesurées d'un.e patient.e.
Pour que les patients puissent sortir plus tôt de l'hôpital, ils sont de plus en plus souvent suivis par des dispositifs portables (wearables). Les données collectées par un dispositif portable doivent parvenir au médecin. Plusieurs concepts entrent en jeu, comme l'authentification, les banques de données, le cryptage, la communication des données et les capteurs.

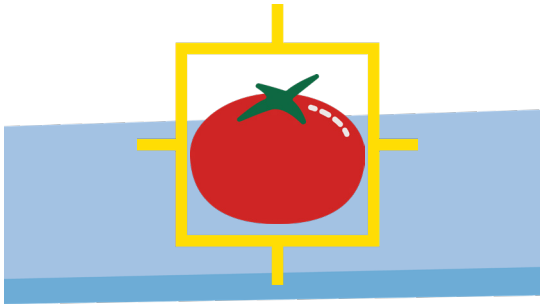


© imec

Défi : cherche, dans ton environnement quotidien, cinq appareils différents qui sont connectés à internet. Les ordinateurs, les tablettes et les smartphones ne comptent pas.

Reconnaissance d'images

Impossible d'échapper aux systèmes à reconnaissance d'images (***computer vision***). Ils sont partout : Facebook donne des suggestions pour taguer des personnes sur tes photos, une voiture avertit le conducteur lorsqu'il fatigue, une sonnette intelligente te dit qui est à la porte…
La reconnaissance d'images est très puissante, mais aussi complexe. Pour qu'un ***ordinateur détecte un objet dans une image, cet ordinateur doit pouvoir suivre des règles claires***. Et la création de ces règles n'est pas une mince affaire, notamment parce qu'une image n'est qu'une grille de pixels pour un ordinateur. Grâce aux techniques d'apprentissage automatique, il est possible d'établir ces règles à l'aide de l'ordinateur. On développe ainsi des ***algorithmes*** qui permettent à l'ordinateur de « voir ».



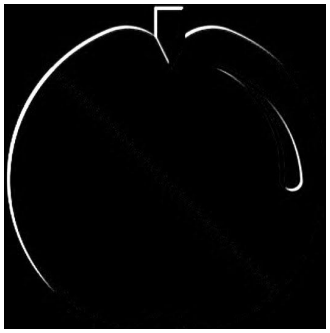
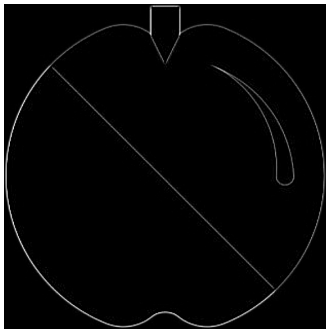
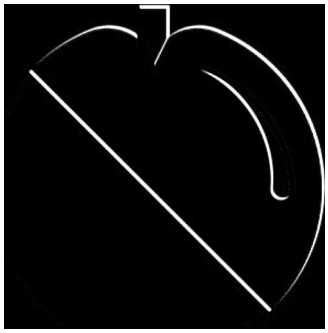
Pour reconnaître une tomate sur une image, tu peux regarder la couleur des pixels. Si le rouge prédomine, tu as déjà une indication qu'il s'agit d'une tomate. Tu peux aussi regarder la forme : si elle est plus ou moins ronde, il y a encore plus de chances que l'image représente une tomate. Tu peux également regarder s'il y a une couronne verte dans le haut. En imaginant de plus en plus de règles, tu reconnais de mieux en mieux une tomate. Cependant, si l'on te montre l'image d'une tomate jaune, tu ne parviendras pas à la reconnaître avec ces règles. Dans ce cas, les règles de détection ne fonctionneront pas.

Étant donné qu'il est pratiquement impossible de créer soi-même des règles de détection pour chaque image de chaque type de tomate, les informaticiens ont conçu des algorithmes permettant à un ordinateur d'apprendre automatiquement ces règles en s'appuyant sur des exemples. C'est ce qu'on appelle ***l'apprentissage automatique***.
La reconnaissance d'images est importante ***dans le domaine de la santé***. Dans la recherche sur les troubles du spectre autistique, on utilise parfois des robots sociaux pour communiquer avec les enfants autistes. Une étude a en effet montré que les robots sociaux peuvent avoir un effet positif sur le fonctionnement social des enfants autistes. Il est en effet très important, à cet égard, de pouvoir reconnaître les émotions des enfants. En sachant reconnaître correctement les émotions, le système peut interagir correctement avec l'enfant.

1: Kouroupa A, Laws KR, Irvine K, Mengoni SE, Baird A, Sharma S. The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis. PLoS One. 2022 Jun 22;17(6):e0269800. doi: 10.1371/journal.pone.0269800. PMID: 35731805; PMCID: PMC9216612.

Défi : tu aimerais savoir comment apprendre à un ordinateur à détecter les émotions ?
Scanne le code QR ou surfe sur la page web en suivant le lien ci-dessous.
Pour cette mission, il est préférable d'utiliser un ordinateur ou une tablette.
https://dwengo.org/basics_ai/emoties_fr

La reconnaissance d'images se fait via des réseaux de neurones profonds convolutifs. À l'aide de filtres sur les images, l'ordinateur détectera les caractéristiques des objets dans les images. À partir de ces caractéristiques, le système décidera quels objets il détecte et avec quelle certitude. Lis-en plus à ce sujet sur le site web de Dwengo.

Reconnaissance vocale et synthèse vocale

Peut-être qu'un robot circule dans ton supermarché, à qui tu peux demander où se trouve la limonade. La technologie de reconnaissance vocale (speech to text) et les logiciels de synthèse vocale (text to speech) permettent de communiquer par le langage parlé avec un robot physique ou un assistant digital comme Siri, Alexa et l'assistant vocal de la voiture.
La technologie de reconnaissance vocale convertit un enregistrement sonore en un fichier texte numérique. Ce que tu dis est automatiquement converti en texte écrit. Un logiciel de « traitement automatique du langage naturel » interprète ce texte et formule une réponse écrite. Cette réponse est ensuite convertie en son à l'aide d'un logiciel de synthèse vocale, afin que le robot prononce la réponse.

Le robot « comprend » donc ce que tu dis grâce au logiciel de reconnaissance vocale combiné au « traitement automatique du langage naturel ». Et le logiciel de synthèse vocale est là pour faire parler le robot. Ce n'est pas simple. Le système doit par exemple être capable de distinguer la parole humaine d'un éventuel bruit de fond. Il est également difficile pour un robot de « comprendre » un texte, car chacun s'exprime à sa manière, même dans une même langue. Pourtant, de grands modèles de langage parviennent parfois déjà à détecter l'ironie ou à distinguer les dialectes d'une même langue. La synthèse vocale est également confrontée à des défis, comme la vocalisation réaliste : une voix ayant peu d'intonations n'est pas agréable à écouter.

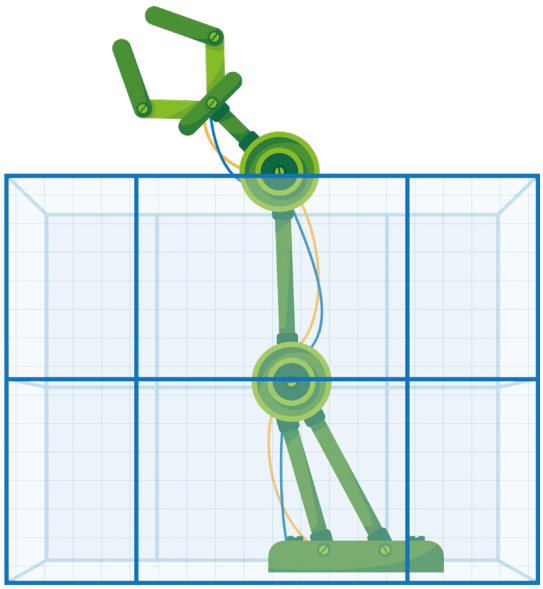
Les médecins utilisent par exemple un logiciel de reconnaissance vocale pour enregistrer les résultats d'un examen dans le ***dossier électronique du patient*** au lieu de les taper au clavier.
La synthèse vocale a également été utilisée pour aider ***un acteur ayant perdu sa voix à cause d'un cancer de la gorge à parler à nouveau***. L'acteur Val Kilmer a été touché par un cancer de la gorge en 2014. En raison du traitement, Val Kilmer est difficile à comprendre et ne peut pas communiquer et être créatif comme il le souhaiterait. À l'aide d'enregistrements sonores existants de la voix de Val Kilmer, Sonantic a construit un modèle d'IA pour se rapprocher le plus possible de la voix de l'acteur[1].

Défi : en utilisant NotebookLM de Google, tu peux générer un podcast à l'aide d'un texte, par exemple le règlement scolaire. Surfe sur dwengo.org/zorg/kaartjes/notebooklm et essaie !

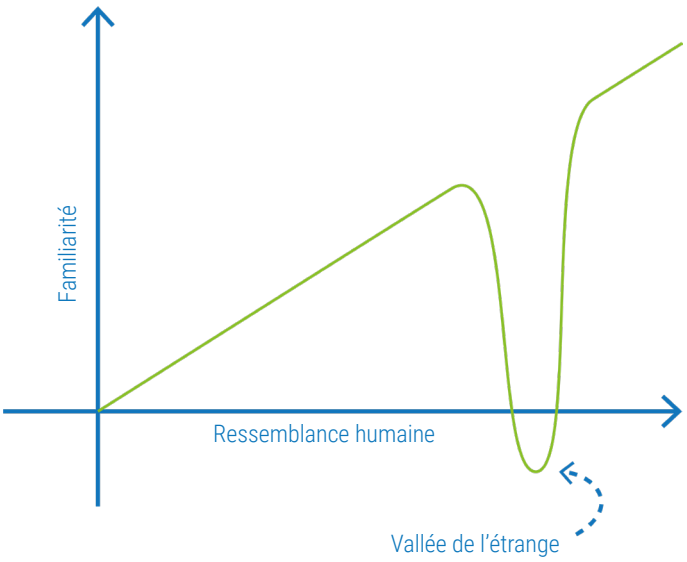
1: Flynn, J. (2021). Helping actor Val Kilmer reclaim his voice [Blogpost]. Geraadpleegd op 19 juni 2022 via https://www.sonantic.io/blog/helping-actor-val-kilmer-reclaim-his-voice.

Robots sociaux

Tu as probablement déjà vu des images de grands bras robotisés qui assemblent des voitures dans une usine. Ces ***robots industriels*** peuvent ***effectuer une tâche spécifique avec rapidité et précision***. Préprogrammés pour effectuer toujours les mêmes mouvements, ils ne savent pas s'adapter à leur environnement. C'est pourquoi il est très dangereux de s'approcher de ce type de robot. ***En effet, un bras robotisé ne s'arrêtera pas si tu te trouves sur sa trajectoire préprogrammée et peut donc te blesser grièvement***. Ces robots ne peuvent pas être utilisés pour interagir avec les humains.



Les robots sociaux sont spécifiquement conçus pour interagir avec les humains. Étant donné qu'un environnement quotidien est bien plus chaotique qu'une chaîne de production industrielle contrôlée, il est beaucoup plus complexe de faire fonctionner correctement ces robots sociaux. Ils ont par exemple besoin de nombreux capteurs pour éviter de heurter des objets ou des personnes. De plus, on attend d'eux qu'ils « comprennent » les gens qui les entourent, tant par la communication verbale que non verbale, et qu'ils interagissent amicalement avec eux. Et surtout, ils doivent être capables d'effectuer différentes tâches (ex. mettre un patient au lit ou lui apporter un verre d'eau).



Les constructeurs de robots sociaux essaient souvent de faire ressembler le plus possible le robot à un humain. Cette apparence crée des attentes : on attend d'un robot qui a des yeux qu'il voie, d'un robot qui a des roues qu'il conduise, d'un robot androïde (un robot qui a l'air très humain) qu'il fasse des actions humaines. Lorsqu'un robot ne répond pas à ces attentes, cela peut être décevant et le robot risque d'être perçu comme négatif. Il est déroutant de voir un animal robot parler ou un robot partiellement recouvert de « peau », mais dont les parties mécaniques sont également visibles. Les gens aiment généralement qu'un robot ait l'air humain, sauf si le robot ressemble trop à un humain, auquel cas ils le trouvent effrayant. Ce phénomène s'appelle la ***Vallée de l'étrange*** : plus un robot ressemble à un humain, plus les gens s'attendent à des comportements humains et plus le robot risque de les décevoir.

Défi : cherche des photos de robots dans la Vallée de l'étrange.

Tracking

De nombreuses entreprises internet comme Google, YouTube et Facebook tirent la majeure partie de leurs ***revenus de la publicité***. Leurs clients ne sont donc pas les utilisateurs de leurs sites web, mais plutôt les entreprises qui placent des publicités chez eux. ***Les géants d'internet vendent donc à ces entreprises une partie de ton attention***. Tu es en quelque sorte leur produit.
Pour faire payer plus cher les publicités, elles doivent pouvoir les cibler. Il est vrai qu'une publicité pour des chaussures de foot a plus d'effet sur les personnes qui jouent au football que sur celles qui n'y jouent pas... Pour gagner plus, les géants d'internet veulent donc en ***savoir le plus possible sur toi***. Pour recueillir ces informations, ***ces entreprises essaient de te suivre (presque) partout***. C'est ce qu'on appelle le ***tracking***.

Le tracking peut se faire de différentes manières. Les sites web utilisent généralement des cookies. Les cookies sont de petits fichiers qui peuvent être enregistrés dans ton navigateur internet quand tu visites un site web. ***Les cookies sont souvent utiles*** : grâce à eux, tu ne dois pas indiquer ton nom d'utilisateur et ton mot de passe à chaque fois que tu visites un site web. ***Il existe toutefois un type spécifique de cookie qui peut être utilisé pour te suivre sur internet***. On les appelle les ***third party cookies*** (ou « cookies tiers »). Ces cookies ne sont pas enregistrés dans ton navigateur par un site web à proprement parler mais, par exemple, par une publicité publiée sur ce site web. Cette publicité vient d'un autre endroit sur internet (ex. Google). Ce système permet à Google de savoir quels sites web tu visites. Si tu visites souvent le site web du club de foot local, tu auras plus de chances de voir des publicités pour des chaussures de foot.

Défi : quels sont les cookies enregistrés dans ton navigateur internet ? Va dans les paramètres de ton navigateur et cherche la liste de cookies enregistrés. Sais-tu pourquoi tous ces sites web enregistrent des cookies ?

Le savais-tu ? Dans la plupart des navigateurs, tu peux paramétrer le blocage automatique des cookies tiers. Ce paramètre est toutefois désactivé par défaut dans de nombreux navigateurs. Tu dois l'activer toi-même si tu veux bloquer ces cookies.

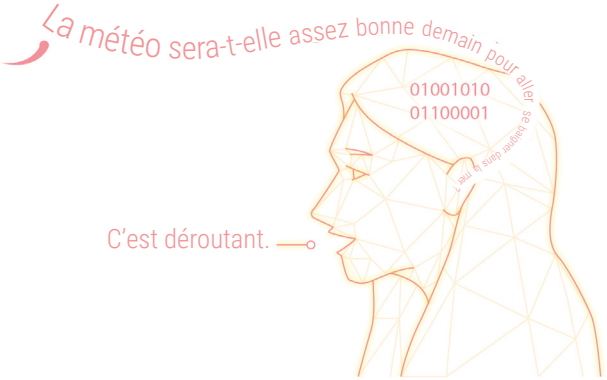
Il existe d'autres moyens de te suivre que les cookies. Les plateformes de réseaux sociaux, par exemple, enregistrent ton comportement pour savoir quels sont tes centres d'intérêt. Il existe également des apps qui peuvent stocker ta localisation. En découvrant où tu te trouves souvent (ex. un restaurant spécifique), une entreprise en apprend également plus sur toi. Lis-en davantage sur le profilage sur le site web de Dwengo.

Traitement automatique du langage naturel, grands modèles de langage et chatbots

Tu as sans doute déjà utilisé la ***technologie linguistique***, par exemple pour traduire automatiquement la carte d'un restaurant lors d'un voyage à l'étranger, pour demander à Microsoft Copilot de rédiger un e-mail ou pour trouver ton chemin avec ton GPS.
Pour « comprendre » automatiquement un texte, on utilise un logiciel de « traitement automatique du langage naturel ». Grâce à des systèmes à base de connaissances ou à des techniques d'apprentissage automatique, comme l'entraînement de modèles à la reconnaissance de textes ou de fréquences sonores, les ordinateurs sont capables de travailler intelligemment avec la langue.

Les ***chatbots*** sont des systèmes logiciels capables d'interagir avec des personnes par le biais de messages textuels tapés au clavier. Un chatbot n'est donc rien d'autre qu'un ordinateur avec lequel tu peux chatter.

Certains médecins spécialistes demandent à un chatbot de mener l'***entretien d'anamnèse***. Le chatbot résume cette conversation, puis le ou la médecin parcourt ce résumé avant la consultation, afin de gagner du temps par patient. Le chatbot utilise le «***traitement automatique du langage naturel***» (***Natural Language Processing*** ou NLP) pour extraire les informations utiles de la conversation. Il est par exemple également possible d'appliquer une analyse automatique des sentiments sur un texte tapé au clavier. Un chatbot peut ainsi détecter que tu commences à t'irriter quand une conversation est laborieuse et te transférer vers un.e collègue humain.e. Comme tu peux le lire sur la fiche consacrée à la reconnaissance vocale et à la synthèse vocale, les robots physiques et les assistants vocaux font également cela.



Les chatbots les plus simples, dont les premiers chatbots comme ELIZA, sont des ***systèmes à base de connaissances***. Dans ces chatbots, différents scénarios sont explicitement programmés, ce qui engendre automatiquement des limites : si tu poses par exemple une question au chatbot en des termes différents de ceux qui sont programmés, le chatbot ne reconnaîtra pas la question. Même s'ils sont souvent impopulaires, ces chatbots sont encore largement utilisés.
Les chatbots modernes ont recours à l'apprentissage automatique, en particulier ***à l'apprentissage profond (deep learning)***.

ChatGPT et Google Gemini reposent sur un ***grand modèle de langage*** (***Large Language Model*** ou LLM). Capables de prédire le mot suivant dans un texte en se basant sur les mots précédents, ces systèmes permettent à l'ordinateur de mieux « comprendre » le langage. Grâce à des techniques

d'apprentissage automatique, les ***grands modèles de langage*** utilisent le word embedding : les mots d'un texte sont convertis en nombres, et un maximum d'informations sur les mots sont stockées via ces nombres : la signification du mot, les autres mots souvent utilisés dans des phrases avec le mot, la place du mot dans une phrase, le contexte d'une phrase contenant le mot, etc.

Les chatbots ne sont pas la seule application du NLP. Le traitement automatique du langage naturel est également utilisé dans le ***secteur de la santé***, où il connaît un succès croissant depuis l'apparition de ce qu'on appelle les grands modèles de langage. Le NLP est par exemple utilisé pour extraire des informations d'un ***dossier électronique du patient***. La ***traduction automatique*** s'améliore aussi de plus en plus. Heureusement, car aux USA, les services d'urgence y ont souvent recours. À leur sortie des urgences, les patients reçoivent parfois des instructions à emporter chez eux concernant les médicaments à prendre ou leur prochaine consultation. Tous les patients ne maîtrisant pas suffisamment l'anglais, les soignants donnent ces instructions dans la langue maternelle du patient[1].

Défi : dans les années 60 du siècle dernier, le logiciel ELIZA a été conçu au Massachusetts Institute of Technology (MIT) dans le cadre d'une expérience visant à déterminer si ELIZA pouvait convaincre en tant que psychothérapeute.
Chatte avec ELIZA via dwengo.org/zorg/kaartjes/eliza

Les développements en technologie linguistique fournissent de nouveaux outils pour détecter les fake news, lutter contre la criminalité et accroître l'efficacité des soins de santé. Il existe également de nombreuses applications dans les entreprises.
Apprends-en davantage à ce sujet sur le site web de Dwengo : à propos de l'analyse des sentiments, à propos de ChatGPT & co et à propos des grands modèles de langage.

1: Taira, B., Kreger, V., Orue, A., & Diamond, L. (2021). A Pragmatic Assessment of Google Translate for Emergency Department Instructions. Journal of General Internal Medicine, 36(11), 3361–3365.