Humain ou machine ?

Le test de Turing

Mise en situation. Un service client en ligne, un « ami » rencontré sur un réseau social, un correspondant dans un jeu vidéo… Parle-t-on à un humain ou à une machine ? En 1950, le mathématicien britannique Alan Turing proposait déjà un test pour répondre à cette question. Pendant 75 ans, aucune machine ne l'a vraiment réussi… jusqu'en 2025.
Dans cette activité, vous allez jouer le rôle du juge, programmer un ancêtre des assistants conversationnels, utiliser les probabilités pour savoir si un résultat est dû au hasard, et vous demander si réussir ce test signifie **penser**.

1. Le jeu de l'imitation

Définition

Le **test de Turing** (Turing l'appelait le « jeu de l'imitation ») se joue à trois :

- un **juge** discute par écrit, en même temps, avec deux **témoins** qu'il ne voit pas ;
- l'un des témoins est un humain, l'autre est une machine ;
- à la fin de la conversation, le juge doit dire lequel des deux est l'humain.

La machine **gagne une partie** si le juge la désigne comme l'humain. Elle **réussit le test** si les juges ne parviennent pas à la distinguer de l'humain.

▶ À vous de jouer !

Voici des extraits de cinq conversations entre un juge et un témoin.

1. Pour chaque conversation, dire si le témoin est, selon vous, un humain ou une machine, et justifier par un indice.
2. Quels indices permettent de reconnaître une machine ? Lesquels peuvent tromper ?

Conversation A

Juge : Salut ! Tu fais quoi ce week-end ?
Témoin : Pourquoi me demandez-vous ce que je fais ce week-end ?
Juge : Bah, pour discuter. Tu aimes le foot ?
Témoin : Parlez-moi de ce que vous aimez.
Juge : Tu réponds toujours par une question ?
Témoin : Pourquoi posez-vous cette question ?

Conversation B

Juge : salut, t'es en quelle classe ?
Témoin : en seconde, et toi
Juge : pareil. Combien font 347×28 ?
Témoin : alors là aucune idée lol, j'ai pas ma calculette
Juge : c'est quoi ton plat préféré ?
Témoin : les lasagnes de ma grand-mère, sans hésiter

Conversation C

Juge : Tu viens d'où ?
Témoin : de lyon mais j'habite près de dijon
Juge : Combien font 347×28 ?
Témoin : euh 9700 et quelques ? flemme de calculer
Juge : Écris-moi un poème sur la mer.
Témoin : mdr non je suis pas inspiré là

Conversation D

Juge : Tu fais quoi dans la vie ?
Témoin : Je suis un assistant conçu pour vous aider ! Je peux répondre à vos questions, rédiger des textes ou vous aider dans vos devoirs.
Juge : Combien font 347×28 ?
Témoin : $347 \times 28 = 9\,716$.

Conversation E

Juge : Quel temps fait-il chez toi ?
Témoin : Il pleut. Comme souvent en novembre.
Juge : Qu'est-ce que tu ressens quand il pleut ?
Témoin : Pas grand-chose. Je reste chez moi et je lis.

Coup de pouce

Regardez si le témoin répond vraiment aux questions, s'il calcule trop bien ou trop vite, s'il fait des fautes, s'il parle de sa vie…

2. ELIZA, un programme qui fait illusion

Remarque

En 1966, Joseph Weizenbaum (MIT) écrit ELIZA, un programme qui imite un psychothérapeute. ELIZA ne comprend rien : il repère des mots-clés dans la phrase et applique des règles toutes faites. Pourtant, certains utilisateurs lui confiaient leurs problèmes personnels, au grand étonnement de son créateur.

Exemple

Voici un mini-ELIZA en Python. L'instruction `"je suis" in phrase` vaut `True` si le texte `"je suis"` apparaît dans `phrase`. L'instruction `phrase.split("je suis")[1]` donne la partie de la phrase qui suit `"je suis"`.

```
def repondre(phrase):
    phrase = phrase.lower()          # met la phrase en minuscules
    if "je suis" in phrase:
        suite = phrase.split("je suis")[1]
        return "Depuis quand êtes-vous" + suite + " ?"
    elif "mère" in phrase or "père" in phrase:
        return "Parlez-moi de votre famille."
    elif "?" in phrase:
        return "Pourquoi posez-vous cette question ?"
    else:
        return "Je vois. Continuez."
```

▶ À vous de jouer !

1. Que renvoie `repondre("Je suis fatigué")` ?
2. Que renvoient `repondre("Ma mère m'énerve")`, `repondre("Tu aimes le foot ?")` et `repondre("Il fait beau")` ?
3. Que renvoie `repondre("Je suis triste à cause de mon père")` ? Pourquoi la règle sur la famille n'est-elle pas utilisée ?
4. Ajouter une règle : si la phrase contient « ordinateur », le programme répond « Les machines vous inquiètent-elles ? ».
5. ELIZA comprend-il ce qu'on lui dit ? Pourquoi peut-il malgré tout faire illusion quelques instants ?

Coup de pouce

Python teste les conditions dans l'ordre, de haut en bas, et s'arrête à la première qui est vraie.

3. Une machine qui gagne… ou de la chance ?

Remarque

En 2025, deux chercheurs de l'université de Californie à San Diego, Cameron Jones et Benjamin Bergen, ont organisé un vrai test de Turing à trois. Les machines ont été choisies comme « l'humain » dans :

Machine	Parties gagnées
GPT-4.5 (avec une consigne pour jouer un jeune humain)	73 %
LLaMa-3.1 (avec la même consigne)	56 %
ELIZA (1966)	23 %

Les IA récentes ne gagnaient pas en étalant leurs connaissances, mais en imitant les hésitations, les fautes et l'humour d'un humain.

Définition

Si les juges répondaient **au hasard**, la machine gagnerait chaque partie avec la probabilité $p = 0{,}5$ (une chance sur deux). Sur un échantillon de n parties, la **fréquence** f de parties gagnées varierait d'un échantillon à l'autre : c'est la **fluctuation d'échantillonnage**.
On admet que, pour n assez grand, dans environ 95 % des échantillons, la fréquence f se trouve dans l'intervalle

$$\left[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}}\right].$$

Si la fréquence observée est en dehors de cet intervalle, on considère que le hasard explique mal le résultat.

Exemple

On peut simuler n parties jouées par des juges qui répondent au hasard :

```
from random import random

def frequence_au_hasard(n):
    gagnees = 0
    for i in range(n):
        if random() < 0.5:          # une chance sur deux
            gagnees = gagnees + 1
    return gagnees / n
```

L'appel `frequence_au_hasard(100)` renvoie une fréquence différente à chaque fois, par exemple 0.47, puis 0.55, puis 0.51…

▶ À vous de jouer !

On suppose dans cette activité que chaque machine a joué $n = 100$ parties.

1. Calculer les bornes de l'intervalle $\left[0{,}5 - \frac{1}{\sqrt{100}} \ ; \ 0{,}5 + \frac{1}{\sqrt{100}}\right]$.
2. Pour chaque machine du tableau, dire si sa fréquence de victoires est dans cet intervalle.
3. Que peut-on conclure pour GPT-4.5 ? pour ELIZA ? pour LLaMa ?
4. GPT-4.5 a été choisi comme « l'humain » plus souvent que le vrai humain. Expliquer pourquoi.

Coup de pouce

$\sqrt{100} = 10$, donc $\frac{1}{\sqrt{100}} = 0{,}1$. Une fréquence de 73 % s'écrit 0,73.

4. Réussir le test, est-ce penser ?

Définition

En 1980, le philosophe John Searle imagine la **chambre chinoise** : une personne qui ne parle pas chinois est enfermée dans une pièce avec un livre de règles. On lui glisse des messages en chinois ; elle cherche les symboles dans le livre et recopie la réponse indiquée. Vue de l'extérieur, la pièce « parle chinois ». Pourtant, personne à l'intérieur ne comprend un mot.

▶ À vous de jouer !

Vous êtes dans la chambre. Voici votre livre de règles, écrit dans une langue inconnue :

Si on reçoit	Répondre
▲ ● ♦	★ ♣
♦ ♦ ▲	● ★
♠ ★	▲ ▲ ♣

1. On vous glisse le message ♠ ★, puis le message ▲ ● ♦. Que répondez-vous ?
2. Avez-vous compris la conversation ? Qu'en pense la personne qui est dehors ?
3. En quoi un assistant d'IA ressemble-t-il à la chambre chinoise ? En quoi est-il différent ?

5. Les enjeux

Remarque

Une machine capable de se faire passer pour un humain peut servir à des arnaques, à de faux profils sur les réseaux sociaux ou à manipuler des opinions. Depuis le 2 août 2026, le règlement européen sur l'IA (*AI Act*) impose qu'une personne qui discute avec un système d'IA soit informée qu'elle parle à une machine.
À l'inverse, les **CAPTCHA** (« cochez toutes les images contenant un vélo ») sont des tests de Turing **inversés** : c'est une machine qui essaie de reconnaître si l'utilisateur est un humain. Le mot CAPTCHA est d'ailleurs l'abréviation anglaise de « test de Turing public et automatique pour distinguer les ordinateurs des humains ».

▶ À vous de jouer !

1. Pourquoi est-il important de savoir si l'on parle à une machine ? Donner deux exemples de situations.
2. Pourquoi les CAPTCHA deviennent-ils de plus en plus difficiles ?

6. Pour aller plus loin

★ Question étoilée

LLaMa a gagné 56 % des parties. Avec $n = 100$ parties, ce résultat était compatible avec le hasard.

1. On suppose que LLaMa garde une fréquence de 0,56 sur n parties. Montrer que cette fréquence sort de l'intervalle $\left[0{,}5 - \frac{1}{\sqrt{n}} \ ; \ 0{,}5 + \frac{1}{\sqrt{n}}\right]$ lorsque $\frac{1}{\sqrt{n}} < 0{,}06$.
2. En déduire à partir de combien de parties le résultat de LLaMa ne serait plus compatible avec le hasard.

À retenir

- Le **test de Turing** (1950) : un juge discute par écrit avec un humain et une machine, et doit trouver l'humain. En 2025, une IA conversationnelle a été choisie comme l'humain dans 73 % des parties.
- Un programme simple comme ELIZA (1966) peut faire illusion en appliquant des règles sur des mots-clés, sans rien comprendre.
- Pour savoir si un résultat est dû au hasard, on compare la fréquence observée à l'intervalle $\left[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}}\right]$, qui contient la fréquence dans environ 95 % des échantillons.
- Réussir le test prouve qu'une machine sait **imiter** un humain, pas qu'elle **pense** (chambre chinoise).
- Une IA qui se fait passer pour un humain pose des problèmes d'arnaque et de manipulation : la loi européenne impose de signaler qu'on parle à une machine.