

Conscience artificielle « Nosce te Ipsum »

Machine de Turing, Problème de décision, Langage, Hypothèse I.A. Test de Turing, Chambre de Searle, Processus privés, Problème de Hilbert, Indécidabilité de Gödel, Quelques problèmes indécidables.



Sensibilité :

Référence :

Version de travail : rentrée 2022

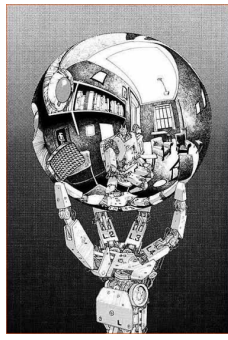
<https://www.youtube.com/watch?v=5LWvS12Yk14>

Olivier Raynaud

1



MC Escher, Main avec sphère réfléchissante (1935)



Katsuhiko Otomo (大友 克洋) né en 1954

2

Espace des questions

Que peut-on dire de ces trois ensembles de questions ?

- A) Ensemble de questions que se pose l'homme.
- B) Ensemble de questions dont la réponse est accessible à l'homme.
- C) Ensemble de questions dont la réponse est accessible à une machine.

- ✓ Quels sont les domaines d'étude de ces ensembles;
- ✓ Existe-t-il des relations d'inclusion claires entre ces ensembles?
- ✓ Savons nous exactement ce qui sera accessible à une machine?

3

Machine de Turing

☐ 1936 Alan Turing imagine de formaliser la notion de machine en décomposant son fonctionnement en termes élémentaires.

« Turing considérait le cerveau humain comme un exemple de machine. Quelles que pussent être les activités mises en œuvre par des mathématiciens humains, celles-ci devaient pouvoir tomber sous la rubrique procédure mécanique »

Penrose 92

ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGSPROBLEM
Turing 1936

4

Machine de Turing

« Nous pouvons comparer un homme en train de calculer un nombre réel à une machine qui ne peut prendre qu'un nombre fini d'états q_1, q_2, q_n, \dots ; qui seront appelés "m-configurations". La machine comporte un "ruban" (l'analogue de papier) ... »

Traduction de D. Leavitt DUNOD

5

Décidabilité

☐ La notion de décidabilité est intrinsèquement liée à celle de calcul

Définition : Un problème est **décidable** s'il existe un processus, descriptible en un nombre fini d'étapes, tel que, une fois ce processus achevé, on connaisse la réponse au problème.

Définition : Un problème **indécidable** est un problème pour lequel on démontre qu'il n'existe aucune méthode de résolution.

☐ Quelques énoncés

- ✓ L'entier n est-il un nombre premier?
- ✓ Quelle est la n ième décimale de π ?
- ✓ Etant données une phrase en anglais, peut-elle être traduite en français ?

6

Thèse de Church-Turing

☐ Church et Turing formalise au cours des mêmes années ce qu'est une procédure effective de calcul.

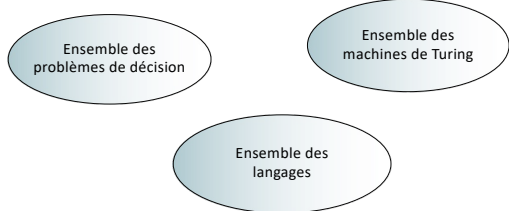
Principe d'équivalence: Soit P est un problème
 Il existe une fonction qui en calcule toutes les réponses;
 Il existe un programme sur une machine de Turing universelle qui y répond;
 Il existe un algorithme qui décrit le processus de calcul de toutes les réponses;
 Il existe une machine qui sait y répondre.

7

Combinatoire et Infinis

☐ Ensembles dénombrables et indénombrables
 ✓ Dénombrable si je peux être mis en bijection avec N ou une partie de N.
 ✓ Indénombrable sinon.

Théorème de Cantor : L'ensemble E est strictement inclus dans P(E).

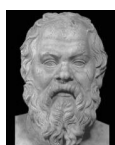


8

Revenons à nos espaces des questions

☐ Que peut-on dire de ces trois ensembles de questions ?

- A) Ensemble de question que se pose l'homme.
- B) Ensemble de question dont la réponse est accessible à l'homme.
- C) Ensemble de question dont la réponse est accessible à une machine.



9

Hypothèse de l'I.A.

On peut exprimer les positions sous 4 points de vue :

- 1) **Thèse de l'I.A. forte** : Toute pensée se réduit à un calcul; en particulier le sentiment de connaissance immédiate consciente naît simplement de l'exécution de calculs appropriés.
- 2) **Thèse de l'I.A. faible**: La connaissance immédiate est un produit de l'activité physique du cerveau; mais bien que toute action physique puisse être simulée par un calcul, une telle simulation ne peut, par elle-même, susciter la connaissance immédiate.
- 3) La connaissance immédiate est suscitée par une action physique du cerveau, mais aucun calcul ne peut simuler cette action physique;
- 4) On ne peut pas expliquer la connaissance immédiate à l'aide du langage, de l'informatique, ni de quelque autre discipline scientifique que ce soit.

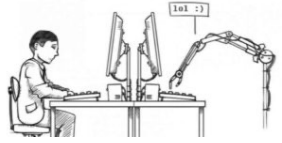
10

Imitation Game

Test de Turing

✓ On ne s'intéresse plus à la nature de la conscience mais à la capacité à la reconnaître grâce au jugement subjectif d'un être considéré conscient.

Le **Test de Turing** consiste à mettre un humain en confrontation verbale à l'aveugle avec un ordinateur et un autre humain. Si la personne qui engage les conversations n'est pas capable de dire lequel de ses interlocuteurs est un ordinateur, on peut considérer que le logiciel de l'ordinateur a passé avec succès le test.



[image extraite de <https://www.chibco.com/fr/actualite-231102/imitation-game-2016-tour-de-informatique.html>]

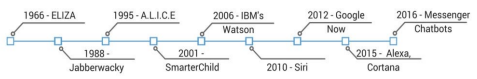
11

Test de Turing

Des rencontres annuelles

- ✓ Le prix Huh Loebner depuis 1990 en partenariat avec l'Université de Cambridge récompense les meilleurs efforts logiciels dans le domaine.
- ✓ Logiciel **Alice** (Artificial Linguistic Internet Computer Entity) remporte de nombreuses fois le prix.
- ✓ Logiciel **Cleverbot** en 2011 arrive à convaincre une majorité de votant.

Brief History of Chatbots



[image extraite de <https://www.altosoft.com/>]

12

La spécificité humaine

La spécificité humaine: Le cerveau humain possède des propriétés, caractéristiques, qu'aucune machine ne pourra jamais avoir. La conscience est une entité distincte du monde physique et ne saurait s'y dissoudre.

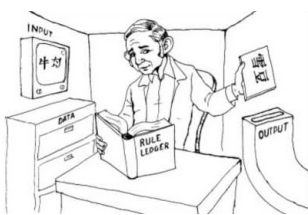
- ✓ Seul le cerveau peut avoir accès au sens. Et cet accès n'est pas réductible à un traitement ou processus, même très élaboré.
- ✓ La nature des perceptions pose le problème des impressions subjectives (Qualia);



13

Chambre de Searle

- ☐ *La chambre chinoise* : analogie réfutant l'hypothèse de l'I.A. forte
- ✓ Les programmes informatiques sont syntaxiques (ce ne sont que des outils à manipuler des symboles);
- ✓ Les pensées humaines ont un contenu sémantique (un sens est attaché au mot manipulés, ...)
- ✓ **Donc** jamais les programmes ne penseront ni ne seront intelligents.



14

Thèse de Church-Turing à propos de l'I.A.

Version privée :

Supposons qu'il existe une méthode utilisée par un être sensible pour trier les nombres en 2 classes, que cette méthode produise toujours une réponse en temps fini et qu'elle donne toujours la même réponse pour un même nombre.

Alors il existe quelque programme à aboutissement certain qui retourne précisément les mêmes réponses que la méthode utilisée par cet être sensible.

[Version dite Privée de la thèse de l'I.A.(Hofstadter)]

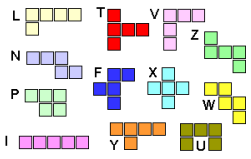
15

Les pentominos

Problème de pavages

Un **pentomino** est une figure géométrique constituée de 5 carrés accolés par un de leurs côtés.

✓ Il faut déterminer comment paver une surface donnée avec des objets de cette forme.



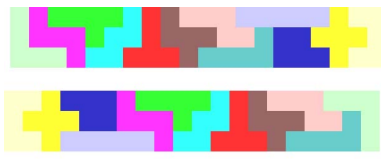
[Extrait de <https://fr.wikipedia.org/wiki/Pentomino>]

16

Pavage rectangle par des pentominos

Quelques résultats de dénombrement pour les surfaces rectangles

- ✓ Rectangle 6*10 : 2339 solutions
- ✓ Rectangle 5*12 : 1010 solutions
- ✓ Rectangle 4*15 : 368 solutions
- ✓ Rectangle 3*20 : 2 solutions




17

Sensibilité à la beauté

Argument de la primauté des compétence inconscientes (H. Dreyfus)

- ✓ *L'intelligence et l'expertise repose essentiellement sur des instincts inconscients.*
- ✓ *Les compétences inconscientes ne sauraient être capturées dans des règles formelles.*



[Image extraite de [https://fr.wikipedia.org/wiki/La_Naissance_de_V%C3%A9nus_\(Botticelli\)](https://fr.wikipedia.org/wiki/La_Naissance_de_V%C3%A9nus_(Botticelli))]

18

Thèse de Church-Turing à propos de l'I.A.

Version publique :

Supposons qu'il existe une méthode utilisée par un être sensible pour trier les nombres en 2 classes, que cette méthode produise toujours une réponse en temps fini et qu'elle donne toujours la même réponse pour un même nombre.

Supposons que cette méthode puisse être communiquée de façon fiable à une autre personne à l'aide du langage.

Alors il existe quelque programme à aboutissement certain qui retourne précisément les mêmes réponses que la méthode utilisée par cet être sensible.

[Version dite Publique de la thèse de l'I.A.(Hofstadter)]

19

Problème de décision

□ Recherche d'un consensus de valeur universelle

« Lorsqu'il s'agit de poser les principes d'une science, l'on doit établir un système d'axiomes renfermant une description complète et exacte des relations entre les concepts élémentaires de cette science (...). Aucune affirmation relative à la science dont nous examinons les principes fondamentaux ne sera admise comme exacte, à moins qu'on ne puisse la déduire des axiomes au moyen d'un nombre fini de déductions. »

Hilbert 1900, Congrès de mathématique

Problème de décision (*Entscheidungsproblem*) : Existe-t-il une procédure générale de résolution des questions mathématiques ?

20

Théorème d'incomplétude de Gödel

□ Premier théorème d'incomplétude

Publication de K. Gödel 1931 : « Sur les propositions formellement indécidables des Principia Mathematica et des systèmes apparentés »

Premier théorème : Dans n'importe quelle théorie récursivement axiomatisable, cohérente et capable de « formaliser l'arithmétique », on peut construire un énoncé arithmétique qui ne peut être ni démontré ni réfuté dans cette théorie.?

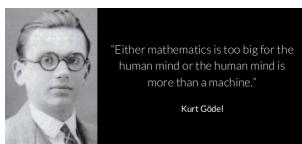


[image extraite de <http://dilemme.sterof.fr/62/veevmm/0exposdicoimpl.htm>]

21

Argument Anti-mécanisme

- *Interprétation du théorème d'incomplétude Gödel*
- ✓ En 1951 Gödel lui-même interprète ses théorème d'incomplétude en termes philosophiques (conférence Gibbs).



- ✓ Le philosophe J. Lucas (1961) et le logicien R. Penrose (1989) défendent cet argument anti mécaniste.

Argument Gödelien : Un système de mathématiciens est à la fois suffisamment cohérent et puissant pour reconnaître sa propre cohérence. Puisque cela est impossible pour une machine de Turing, le Gödelien conclut que le raisonnement humain doit être non mécanique.

22

Quelles conclusions

- *Il semble par nature très hasardeux de réduire l'esprit humain à une procédure universelle de calculs. Pourtant*
- ✓ Les arguments de Searle sont peu convaincants et pourraient s'appliquer dans des contextes plus larges.
- ✓ L'argument de Dreyfus avait été prédit par Turing lui-même et décrit comme étant justifié par l'informalité d'un comportement donné. Mais que les modèles de description des comportements avaient le temps de progresser.
- ✓ L'interprétation du théorème de Gödel – argument anti-mécanisme - est très critiquée. L'esprit humain semble en particulier intuitivement très peu cohérent.

Bilan : Aucune des thèses correspondants au différents points de vue n'est ni démontrée ni réfutée à ce jour.

23

Approche méthodologique

- *Comment construire des machines dites conscientes ?*
 - ✓ Identifier les propriétés saillantes de la conscience.
 - ✓ Réaliser des phases d'ingénierie sur ces propriétés.
- Un avion ne ressemble en rien à un pigeon et les pigeons ne confondent pas les avions avec leur congénères.

24

Complexité et Décidabilité

☐ La machine de Turing permet d'induire les classes de problème

Indécidabilité

25

Problèmes indécidables

☐ En mathématique et logiques de nombreux problèmes ont été montrés indécidables

✓ Le jeu de la vie : Etant donnée une configuration initiale, cette configuration perdure-t-elle indéfiniment ou finie-t-elle par disparaître?

✓ Problème de correspondance de Post

bba	ab	bba	a
bb	aa	bb	baa

26

Indécidabilité physique

☐ L'indécidabilité prend aussi d'autres formes comme en mécanique physique.

✓ Mécanique classique / Problème des 3 corps : il s'agit de trouver toutes les solutions mathématiques possibles des équations différentielles décrivant les mouvements de trois astres s'attirant les uns les autres sous l'effet de la gravitation.

✓ Mécanique quantique / Principe d'incertitude de Heisenberg: on ne peut calculer, à la fois, la position et la vitesse d'une particule.

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

27

Pour résumer

☐ *Quelques points essentiels de l'exposé*

- ✓ **Thèse de l'I.A.** : 4 thèses sont généralement retenues pour décrire les objectifs de l' I.A (de la thèse de l'I.A . forte au nihilisme).
- ✓ **Modèle de Turing** : Le modèle de Turing décrit les limites du décidable et semble capturer certains aspects du raisonnement humain;
- ✓ **Théorème d'incomplétude** : Il existe une infinité de questions dites indécidables (non calculables par une machine). De plus tous système axiomatique est incomplet.

Question :

Existe-il des questions mathématiques et logiques accessibles aux hommes et non aux machines?
