

# Natural Language Processing (NLP)

## 1 — Introduction

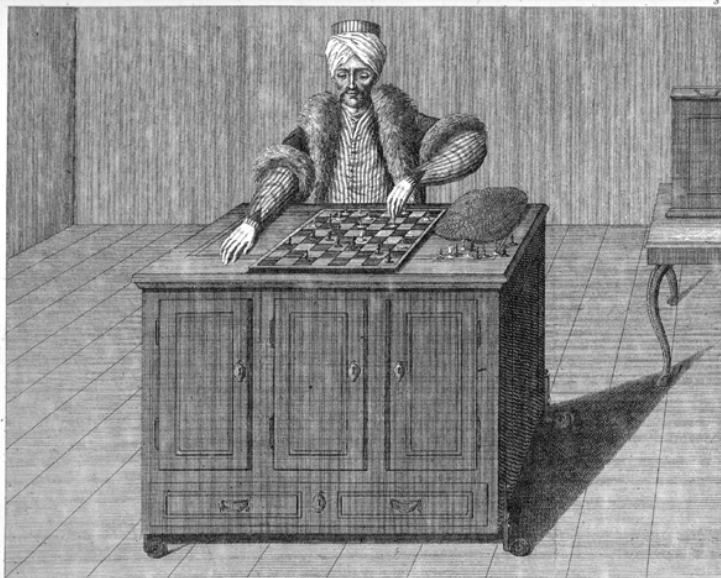
---

Georges-André Silber

<https://mines.paris/nlp>

Session 2025/2026 — ES3A\_MES-07

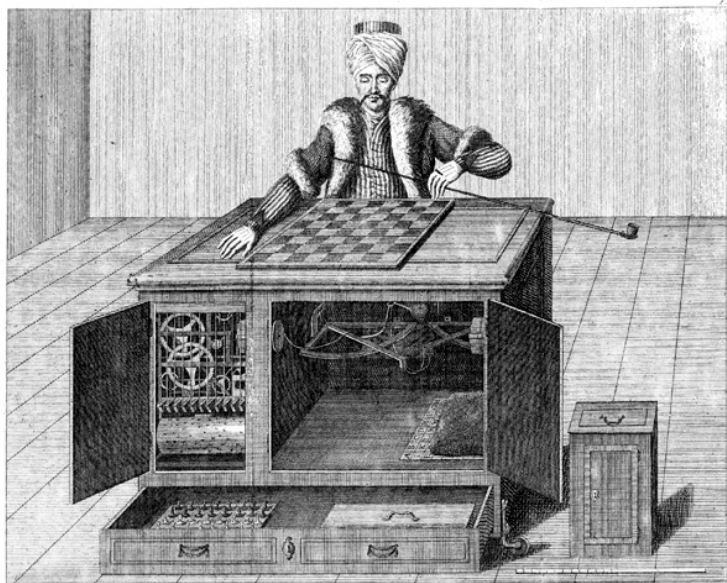
École des mines de Paris



W. de Kempelen, del. | Chr. à Meckel, excudit, Basileae. | P. G. Pinet, sculp.  
Der Schachspieler im Spiel begriffen. | Le Joueur d'Échecs tel qu'on le voit pendant le jeu.

## Turc mécanique (1770)

Source : Karl Gottlieb von Windisch,  
domaine public.



W. de Kempelen del. Che à Mechel, exécut. Basileus. P.G. Pintz, fecit.  
*Der Schach-Spieler, wie er vor dem Spiel gezeigt wird von vorne. Le Joueur d'Échecs, tel qu'on le montre avant le jeu par devant.*

## Turc mécanique (1770)

Source : Wolfgang von Kempelen, [domaine public](#).

M I N D  
A QUARTERLY REVIEW  
OF  
PSYCHOLOGY AND PHILOSOPHY



I.—COMPUTING MACHINERY AND  
INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

## Test de Turing (1950)

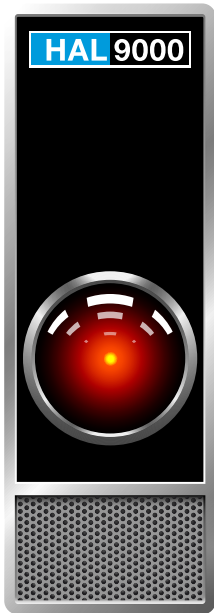
Turing, A.M. 1950. *Computing machinery and intelligence*. Mind; a quarterly review of psychology and philosophy LIX, 236, 433–460



## **Blade Runner** (1982)

Ridley Scott. Los Angeles en 2019. Test Voight-Kampff pour détecter les *réplicants*.

*"Do Androids Dream of Electric Sheeps?"*, Philip K. Dick (1966).



## **HAL 9000** (1968)

Stanley Kubrick, *"2001, l'odyssée de l'espace"*.



## **Terminator** (1984)

James Cameron

Skynet : **Risque**  
existentiel



# WARGAMES

**Wargames** (1984)

John Badham

WOPR (War Operation  
Plan Response)

[extrait 1](#), [extrait 2](#)

The image shows a movie poster for the film 'Her' (2013). It features a solid red background with the word 'her' written in a large, white, lowercase, sans-serif font, centered horizontally and vertically.

**Her** (2013)

Spike Jonze

- Apprendre les langues aux machines
- **Applications** (transparents MVA 2024, de 8 à 27) (**transparentes originaux**)
- **Challenges** (transparents MVA 2024, de 63 à 87)
- "SOTA de la plupart de ces tâches : ML (LLM)"
- Traduction : "les techniques à l'état de l'art (SOTA) pour résoudre ces tâches utilisent du *machine learning* (ML), notamment des grands modèles de langues (LLM)"

- 1950, traduction automatique (contexte de guerre froide)
- 1950, *Computing Machinery and Intelligence* (A. Turing)
- 1954, expérience [Georgetown-IBM](#), traduction du russe vers l'anglais
- 1955, introduction du terme "[Artificial Intelligence](#)" par John Mac Carthy
- 1966, [ELIZA](#) (Joseph Weizenbaum)
- 1968, [SHRDLU](#) (PhD de [Terry Winograd](#) au MIT)
- 1970–2000, « ontologies conceptuelles »
- 2018, [BERT](#) (Google)
- 2020, [GPT-3](#) (OpenAI)
- 2022, [ChatGPT](#) (OpenAI)
- 2023, Poids et code ouvert : [llama](#) (Meta)
- 2025, Multimodalité : [Gemini](#) (Google), [Mistral AI](#), [Claude](#) (Anthropic), [GPT 5](#) (OpenAI)

- 1943, Notion de neurone artificiel ([McCulloch & Pitts](#))
- 1957/1958, Apprentissage supervisé, Perceptron (Rosenblatt, [1957](#), [1958](#))
- 1962, Plusieurs couches en propagation avant ([Rosenblatt](#))
- 1986, Rétropropagation du gradient ([Rumelhart, Hinton, Williams](#))
- 1989, Réseaux convolutifs ([Le Cun et al.](#))
- 1990, Réseaux récurrents ([Elman](#))
- 1997, LSTM ([Hochreiter](#))
- 2006, *Deep Learning*,  $c \geq 3$  ([Hinton, Bengio](#))
- 2017, Architecture *Transformer* ([Vaswani et al.](#))

Extrait de la [leçon inaugurale](#) de Benoît Sagot au Collège de France (11/2023) :

1. Écriture : stockage des informations de manière externe et pérenne. Outil d'accès à l'information
2. Imprimerie : externalisation et diffusion facilités
3. Web : numérisation massive, moteurs de recherche. Automatisation de l'identification des sources.
4. IA : restitution des informations et capacité externe de raisonnement