

# Que veut dire "comprendre" pour une machine?

Jean-Louis Dessalles\*

\* Telecom Paris

<https://perso.telecom-paristech.fr/jld/>

**Résumé.** Un humain dont on dit qu'il ne comprend rien n'est pas considéré comme intelligent. L'intelligence artificielle s'expose à ce type de jugement. Les techniques numériques comme les réseaux de neurones ou l'apprentissage par renforcement prennent des décisions réflexes sur la base d'une expertise statistique. Elles n'ont pas besoin de comprendre la situation, ni de comprendre la solution qu'elles proposent. Les techniques à base de connaissances explicites, quant à elles, ont un horizon de compréhension particulièrement restreint : les prédicats qui apparaissent dans leurs règles ont un sens pour l'être humain, mais restent inconnaisables pour la machine. Je montrerai comment on peut augmenter le degré de compréhension d'une machine en la dotant de quelques mécanismes inspirés de notre fonctionnement cognitif, notamment les mécanismes de contraste, d'abduction et de compression.

## Summary

A human who is said to "understand nothing" is not considered intelligent. Artificial intelligence is exposed to this type of judgment. Digital techniques such as neural networks or reinforcement learning make reflex decisions based on statistical expertise. They do not need to understand the situation, nor do they need to understand the solution they offer. Explicit knowledge-based techniques, on the other hand, have a particularly limited horizon of understanding: the predicates that appear in their rules have meaning for humans, but remain unknowable for the machine. I will show how we can increase the level of understanding of a machine by providing it with a few mechanisms inspired by our cognitive functioning, notably the mechanisms of contrast, abduction and compression.