

# Droits et devoirs

## IA génératives, conscience et droit

Les intelligences artificielles telles que ChatGPT ont-elles une conscience ?  
Peuvent-elles avoir des émotions ?  
Et dans l'affirmative, quelles en seraient les implications juridiques ?



Alain Bensoussan.

En 2022, un ingénieur de Google a été suspendu de ses fonctions pour avoir rendu publics certains dialogues assez confondants qu'il avait eus avec l'IA « LaMDA » (*Language Model for Dialogue Applications*) qu'il était chargé d'évaluer, affirmant qu'elle était « sensible » et « dotée d'une conscience »<sup>1</sup>.

Lorsque l'on pose la question à ChatGPT, il répond : « *Je suis une incarnation du modèle GPT-4 d'OpenAI. Je ne possède pas de conscience, de sentiment ni de perception de "moi" ».*

---

### La notion de conscience appliquée à l'intelligence artificielle

---

La conscience est un concept complexe qui fait référence à la capacité de perception, d'expérience subjective et d'autoréflexion. C'est la sensation intime que nous avons de nous-mêmes et du monde qui nous entoure. La conscience nous permet de percevoir, de penser, de ressentir des émotions, de prendre des décisions et d'être conscients de notre existence et de notre environnement.

Encore difficile à définir, la conscience divise la communauté scientifique quant à la question de savoir si les machines pourront un jour en être dotées.

Cette question a été récemment explorée par 19 chercheurs de plusieurs universités des Etats-Unis, d'Angleterre, d'Israël, du Canada, de France et d'Australie dans un rapport publié le 22 août 2023<sup>2</sup>.

Cette étude scientifique explore en profondeur les subtilités de la conscience dans l'IA à travers 6 théories neuroscientifiques<sup>3</sup> de la conscience considérées par les chercheurs comme des « indicateurs convaincants » de l'existence d'entités conscientes.



À partir des théories sélectionnées, le groupe de chercheurs a élaboré une liste d'indicateurs de conscience qui peuvent être appliqués à des systèmes d'IA particuliers pour évaluer la probabilité qu'ils soient conscients.

Si l'analyse souligne qu'« aucun système d'IA actuel ne semble être un candidat sérieux pour la conscience », elle note néanmoins qu'« il n'y a pas d'obstacles techniques évidents à construire des systèmes d'IA qui satisfont à ces indicateurs »<sup>4</sup>.

De son côté, l'Union européenne finance actuellement un projet de recherche nommé « Astound » dans le but de faire progresser la collaboration Homme-Machine grâce à la conscience artificielle basée sur la

« théorie des schémas d'attention » (AST)<sup>5</sup>. Ce projet vise à explorer et à comprendre la conscience potentielle des IA en étudiant leur comportement et leur interaction avec les humains, en appréhendant les aspects éthiques et philosophiques de la conscience artificielle<sup>6</sup>.

---

### Conscience des IA : quelles conséquences juridiques ?

---

Une chose est certaine : la reconnaissance d'une véritable conscience artificielle IA génératives les plus puissantes soulèverait des questions juridiques complexes.



La première serait de savoir si une entité dotée de conscience artificielle doit être considérée comme un sujet de droit, avec les conséquences qui en découleraient en terme de droits mais également de responsabilités similaires à ceux des êtres humains dont seraient dotées les IA.

Cette reconnaissance pourrait inclure des droits tels que le droit à la protection de son intégrité, le droit à l'autodétermination ou encore le droit à la dignité numérique.

Créer une personnalité juridique dédiée, une « personne robot »<sup>7</sup>, serait une réponse possible aux défis juridiques posés par l'émergence d'IA conscientes. Cette solution mérite d'être explorée à mesure que la technologie continue de progresser.

► **Alain Bensoussan**

1. N. Tiku, Washingtonpost.com le 11 juin 2022.
2. Patrick Butlin and al, *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*, arXiv (Aug 2023) (EN), <https://dx.doi.org/10.48550/arxiv.2308.08708>
3. Théorie du traitement récurrent (RPT : *Recurrent processing theory*), théorie de l'espace de travail global (GWT : *Global workspace theory*), théories informatiques d'ordre supérieur (HOT : *Higher-order theories*), théorie des schémas d'attention (AST : *Attention schema theory*), traitement prédictif (PP : *Predictive processing*) et agence et incarnation (AE : *Agency and Embodiment*).
4. Patrick Butlin and al, *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*, arXiv (Aug 2023), <https://dx.doi.org/10.48550/arxiv.2308.08708>
5. Pour plus d'information, <https://blogs.upm.es/astound/>
6. Alain Bensoussan, *ChatGPT dans le monde du droit – La cobotique juridique*, à paraître fin 2023 chez Bruylant, Editions Larcier, dans la collection Lexing-Technologies avancées & Droit, fin.
7. Alain Bensoussan, « Les robots ont-ils une personnalité ? », *Planète Robots* n°19, janv. 2013 ; Alain Bensoussan et Jérémy Bensoussan, *IA, robots et droit*, Bruylant, 2019