



Grokking (apprentissage automatique)

2 langues

[Article](#) [Discussion](#)

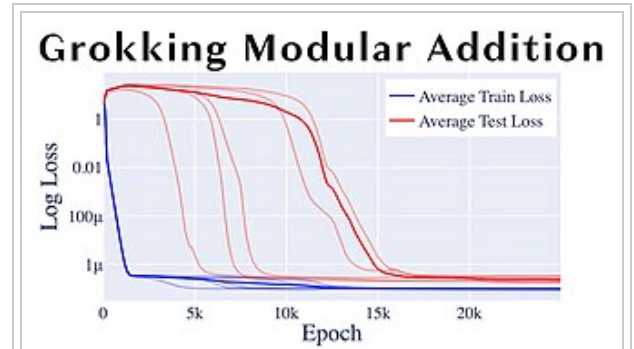
[Lire](#) [Modifier](#) [Modifier le code](#) [Voir l'historique](#) [Outils](#)

En [apprentissage automatique](#), le **grokking** (ou **généralisation différée**), est un phénomène où, après avoir commencé en mémorisant simplement les réponses attendues, le modèle apprend soudainement une solution qui se généralise à des tâches similaires^{2,3,4}.

Le terme a été introduit en janvier 2022 par des chercheurs d'[OpenAI](#) qui étudiaient comment les réseaux de neurones effectuent des additions. Le terme dérive du mot *grok* inventé par Robert Heinlein dans son roman *Stranger in a Strange Land*¹.

Le *grokking* peut être compris comme une [transition de phase](#) au cours du processus d'entraînement⁵. Bien que le grokking ait été considéré comme un phénomène largement lié à des modèles relativement superficiels, le grokking a été observé dans des réseaux neuronaux profonds et des modèles non neuronaux, et fait l'objet de recherches^{6,7,8,9}.

Une explications potentielle serait que la normalisation (un composant de la fonction de perte qui pénalise les valeurs élevées de paramètres) favorise la solution générale, qui implique des valeurs moins élevées de paramètres, mais qui est aussi plus difficile à trouver. Selon Neel Nanda, l'apprentissage de la solution générale pourrait en fait être graduelle, même si la transition vers la solution générale arrive ensuite de manière relativement soudaine¹.



Les courbes de perte bleues représentent une mémorisation précoce des données d'entraînement (similaire au [surapprentissage](#)), et les courbes rouges montrent la généralisation tardive, avec l'apprentissage d'un algorithme d'addition modulaire qui fonctionne pour d'autres nombres¹.

Références

[[modifier](#) | [modifier le code](#)]

(**en**) Cet article est partiellement ou en totalité issu de l’article de Wikipédia en anglais intitulé « [Grokking \(machine learning\)](#) » (voir la liste des auteurs).

- ↑ ^a ^b et ^c (**en**) Anil Ananthaswamy, « [How Do Machines ‘Grok’ Data?](#) [archive] », sur *Quanta Magazine*, 12 avril 2024 (consulté le 21 janvier 2025)
- ↑ (**en**) Adam Pearce et Asma Ghandeharioun, « [Do Machine Learning Models Memorize or Generalize?](#) [archive] », sur *pair.withgoogle.com* (consulté le 19 janvier 2025)
- ↑ (**en**) « [Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets](#) [archive] », 2022 (arXiv 2201.02177)
- ↑ (**en**) « [Bridging Lottery ticket and Grokking: Is Weight Norm Sufficient to Explain Delayed Generalization?](#) [archive] », 2023 (arXiv 2310.19470)

5.

↑

(en)

« Towards Understanding Grokking: An Effective Theory of Representation Learning », *NeurIPS*, 2022
([arXiv 2205.10343](#), [lire en ligne](#) [[archive](#)])

6.

↑

(en)

« Deep Grokking: Would Deep Neural Networks Generalize Better? », *ICML*, 2024 ([arXiv 2405.19454](#))

7.

↑

(en)

« Grokking Beyond Neural Networks: An Empirical Exploration with Model Complexity », *TMLR*, 2024
([arXiv 2310.17247](#))

8.

↑

(en)

« Omnigrok: Grokking Beyond Algorithmic Data », *ICLR*, 2023 ([arXiv 2210.01117](#))

9.

↑

(en)

« Grokking-like effects in counterfactual inference », *IJCNN*, 2022 ([lire en ligne](#) [[archive](#)])



Portail de l'intelligence artificielle

Catégories : [Phénomène](#) | [Apprentissage automatique](#) [+]

La dernière modification de cette page a été faite le 27 janvier 2025 à 01:37.

Droit d'auteur : les textes sont disponibles sous [licence Creative Commons attribution, partage dans les mêmes conditions](#) ; d'autres conditions peuvent s'appliquer. Voyez les [conditions d'utilisation](#) pour plus de détails, ainsi que les [crédits graphiques](#). En cas de réutilisation des textes de cette page, voyez [comment citer les auteurs et mentionner la licence](#).

Wikipedia® est une marque déposée de la [Wikimedia Foundation, Inc.](#), organisation de bienfaisance régie par le paragraphe [501\(c\)\(3\)](#) du code fiscal des États-Unis.

[Politique de confidentialité](#) [À propos de Wikipédia](#) [Avertissements](#) [Contact](#) [Code de conduite](#) [Développeurs](#) [Statistiques](#)

[Déclaration sur les témoins \(cookies\)](#) [Version mobile](#)

