

AIEd : Intelligence Artificielle pour l'Éducation

Guillaume Cleuziou

Journée NSI-SNT

Nancy, 3 avril 2024

- 🏠 Introduction à Python
 - latest
 - Search docs
- TP1 - Prise en main de Python
- TP1 - Exercices complémentaires
- ☐ TD2 - Boucles et fonctions
 - Exercice 1.1 : Somme
 - Exercice 1.2 : Maximum
 - Exercice 1.3 : Nombre de voyelles
 - Exercice 2.1 : Minimum**
 - Exercice 2.2 : Moyenne
 - Exercice 2.3 : Ecart
 - Exercice 3.1 : Nombre d'occurrences d'une lettre
 - Exercice 3.2 : Première occurrence d'une lettre
 - Exercice 4.1 : Prononçable
 - Exercice 4.2 : Nombre de syllabes
- TP2 - Conditionnelles et boucles
- TD3 - Chaînes et Listes
- TP3 - Chaînes/Listes et boucle While
- TD4 - Les boucles while
- Le "Défi - Code"
- TP4 - API d'une liste de joueurs

Exercice 2.1 : Minimum

Ecrire une fonction `minimum` qui retourne le minimum d'une liste de nombres passée en paramètre.

- Sur l'entrée `[12, 1, 25, 7]` votre solution doit renvoyer `1`.
- Sur l'entrée `[-3, 7, 18, -12]` votre solution doit renvoyer `-12`.
- Sur l'entrée `[-12, -1, -25, -7]` votre solution doit renvoyer `-25`.
- Sur l'entrée `[]` votre solution doit renvoyer `None`.

```
1 def minimum(liste):
2     res = 0
3     for i in range(1, len(liste)):
4         if liste[i] < res :
5             res = liste[i]
6     return res
```

Envoyer

Votre fonction ne fait pas ce qui est attendu

- Sur l'entrée `([12, 1, 25, 7],)` votre programme a renvoyé `0` alors qu'on attendait `1`
- Sur l'entrée `([],)` votre programme a renvoyé `0` alors qu'on attendait `None`
- Votre fonction ne renvoie pas ce qui est attendu pour au moins une entrée invisible

Qu'est-ce que l'AIEd?

“L'AIEd fait référence à l'utilisation des technologies de l'Intelligence Artificielle dans le domaine de l'éducation pour faciliter l'enseignement, l'apprentissage ou la prise de décision.”

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001.

Qu'est-ce que l'AIEd?

... et au développement

“L'AIEd fait référence à l'utilisation des technologies de l'Intelligence Artificielle dans le domaine de l'éducation pour faciliter l'enseignement, l'apprentissage ou la prise de décision.”

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001.

Qu'est-ce que l'AIEd?

... et au développement

“L'AIEd fait référence à l'utilisation des technologies de l'Intelligence Artificielle dans le domaine de l'éducation pour faciliter l'enseignement, l'apprentissage ou la prise de décision.”

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001.

Plan de la présentation

1. **Le passé** : évolutions des recherches en AIEd et principaux champs d'études
2. **Le présent** : l'AIEd à l'ère du Deep Learning
→ focus sur les **embeddings** de programmes
3. **Le futur** : AIEd et IA génératives

Plan de la présentation

1. **Le passé** : évolutions des recherches en AIEd et principaux champs d'études
2. **Le présent** : l'AIEd à l'ère du Deep Learning
→ focus sur les **embeddings** de programmes
3. **Le futur** : AIEd et IA génératives

Application : l'enseignement de la programmation

Le "Passé"

Évolutions des recherches en AIEd
et principaux champs d'études

Historique

Depuis une 30aine d'années (seulement!) l'AIEd est un champ de recherche à part entière qui évolue à mesure des progrès de l'Intelligence Artificielle.

L'AIEd fait se rencontrer principalement deux communautés de chercheurs :

- en informatique & statistiques
- en sciences de l'éducation

Historique

Depuis une 30aine d'années (seulement!) l'AIEd est un champ de recherche à part entière qui évolue à mesure des progrès de l'Intelligence Artificielle.

L'AIEd fait se rencontrer principalement deux communautés de chercheurs :

- en informatique & statistiques
- en sciences de l'éducation

Autour de quelques défis majeurs auxquels l'IA contribue :

- Automatisation de certaines tâches (ex. correction)
- Personnalisation des apprentissages
- Amélioration de l'engagement des apprenants
- Meilleure compréhension des mécanismes d'apprentissage

AIEd & Ethique

Des questionnements éthiques se posent, principalement lorsque l'exploitation des données s'éloigne des enjeux pédagogiques initiaux pour répondre à des enjeux financiers :

*“Des centaines d'établissements d'enseignement supérieur sont abonnés à des plateformes privées qui [utilisent les données des classes antérieures] pour **évaluer les candidats à l'admission** en fonction de facteurs tels que la probabilité qu'ils s'inscrivent, le montant de l'aide financière dont ils auront besoin, la probabilité qu'ils obtiennent leur diplôme, etc.”*

Kopenski M. (2021). From admissions to teaching to grading, AI is infiltrating higher education. *Online publication www.gsra.org.uk*

AIEd & Ethique

Des questionnements éthiques se posent, principalement lorsque l'exploitation des données s'éloigne des enjeux pédagogiques initiaux pour répondre à des enjeux financiers :

*“Des centaines d'établissements d'enseignement supérieur sont abonnés à des plateformes privées qui [utilisent les données des classes antérieures] pour **évaluer les candidats à l'admission** en fonction de facteurs tels que la probabilité qu'ils s'inscrivent, le montant de l'aide financière dont ils auront besoin, la probabilité qu'ils obtiennent leur diplôme, etc.”*

*“L'université du Texas à Austin a abandonné le système GRADE en 2020 estimant qu'il risquait de **reproduire des biais superficiels** dans la notation”*

Kopenski M. (2021). From admissions to teaching to grading, AI is infiltrating higher education. *Online publication www.gsra.org.uk*

AIEd & Ethique

Des questionnements éthiques se posent, principalement lorsque l'exploitation des données s'éloigne des enjeux pédagogiques initiaux pour répondre à des enjeux financiers :

*“De nombreux établissements d'enseignement supérieur ont de plus en plus recours à l'IA pour noter les étudiants, car **les classes en ligne deviennent trop nombreuses** pour que les enseignants puissent les gérer correctement”*

Kopenski M. (2021). From admissions to teaching to grading, AI is infiltrating higher education. *Online publication www.gsra.org.uk*

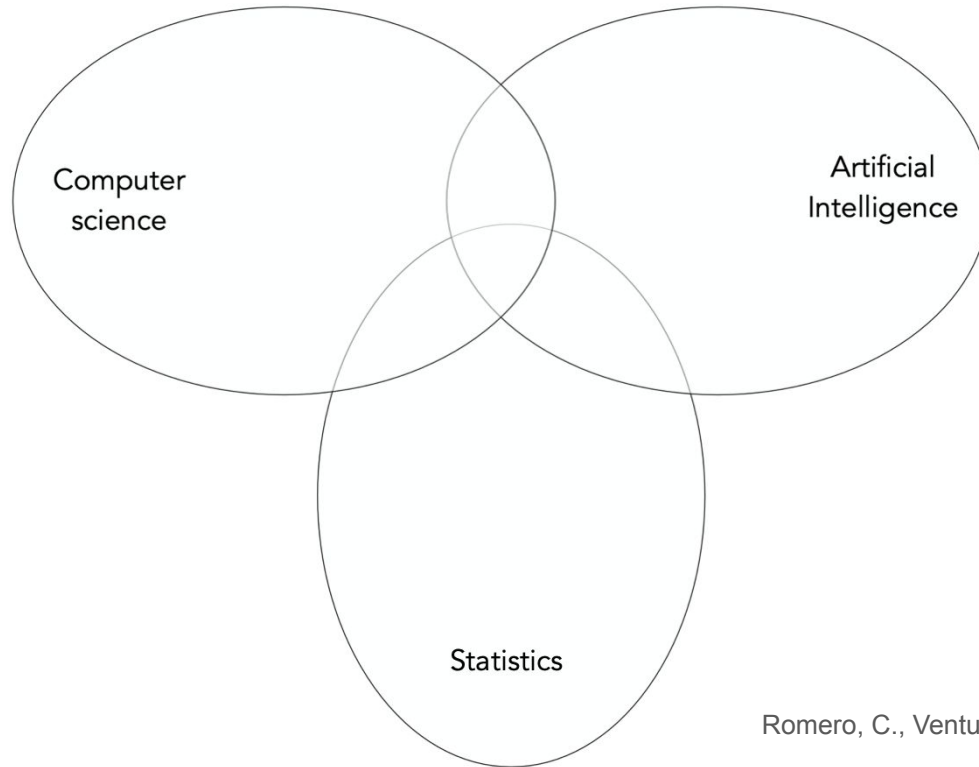
AIEd & Ethique

Des questionnements éthiques se posent, principalement lorsque l'exploitation des données s'éloigne des enjeux pédagogiques initiaux pour répondre à des enjeux financiers ou basculer vers des systèmes de surveillance :

*“L'université australienne Deakin dispose d'un assistant IA nommé Genie qui sait si l'étudiant qui pose une question s'est intéressé à des supports de cours en ligne spécifiques et qui peut **vérifier l'emplacement et les activités des étudiants** pour déterminer s'ils sont allés à la bibliothèque ou leur signaler qu'ils ont passé trop de temps dans le réfectoire et les inviter à passer à autre chose.”*

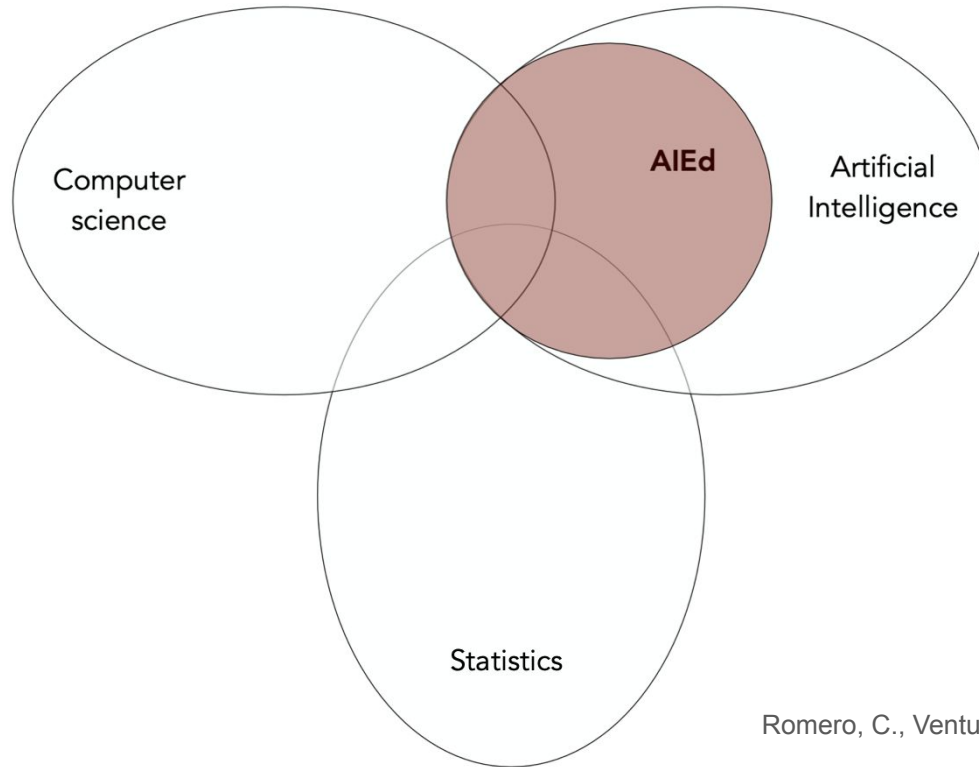
Kopenski M. (2021). From amissions to teaching to grading, AI is infiltrating higher education. *Online publication www.gsra.org.uk*

Champs d'études



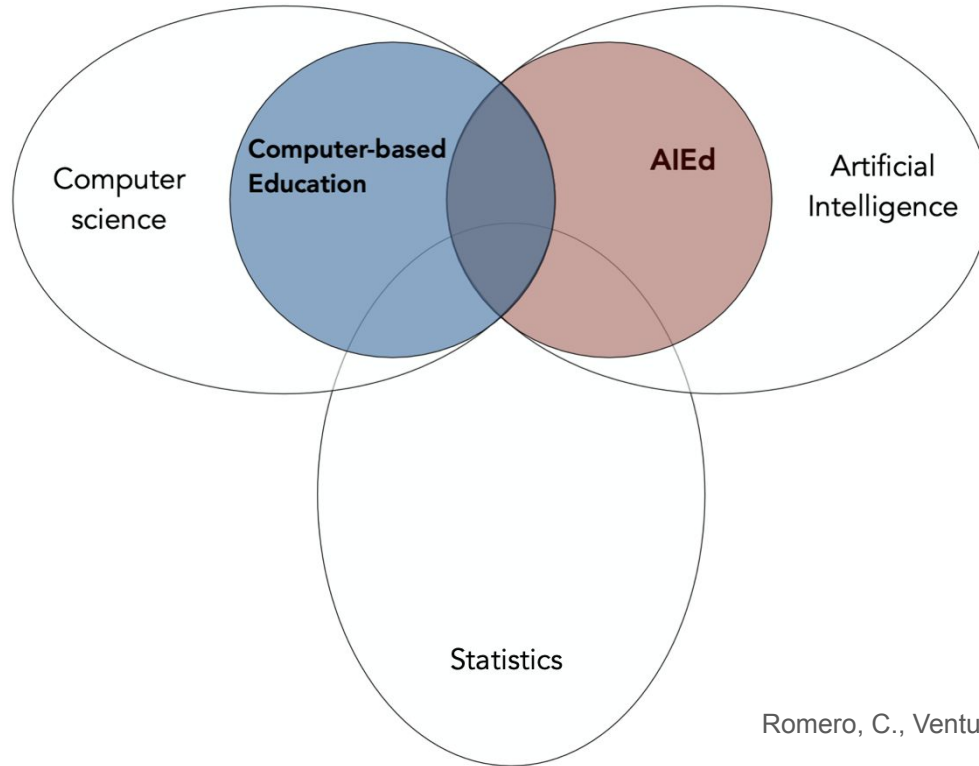
Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & education*, 51(1), 368-384.

Champs d'études



Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & education*, 51(1), 368-384.

Champs d'études

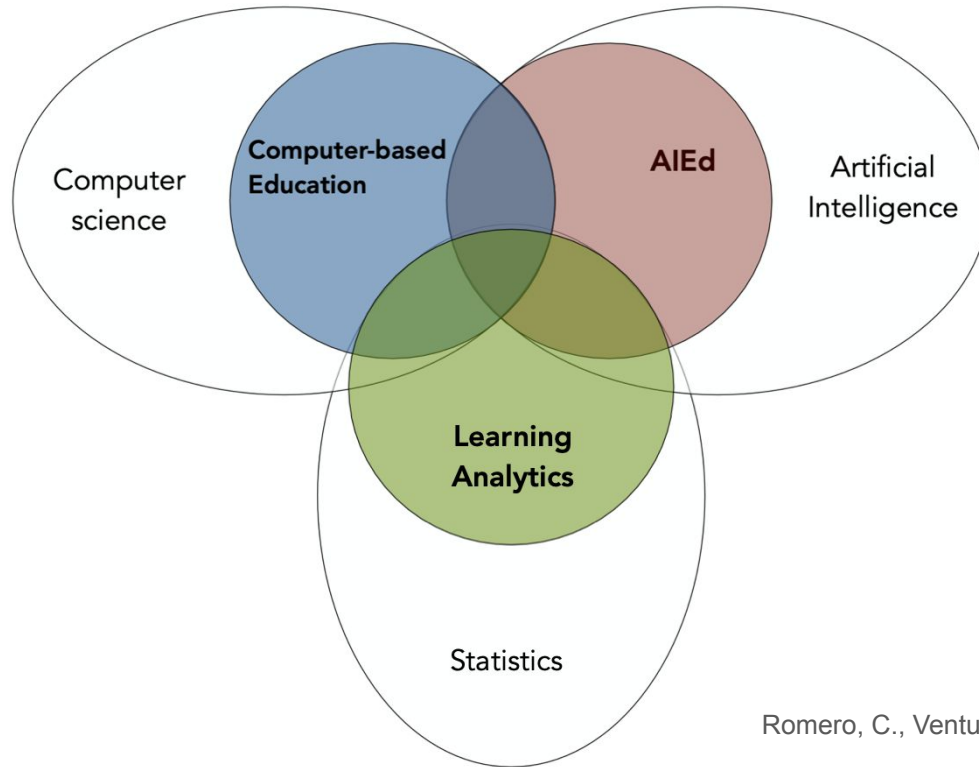


Computer-based Education : usage des ordinateurs pour donner des instructions aux apprenants

- ❑ Tests, Quiz
- ❑ LMS : Learning Management System (ex. Moodle)
- ❑ ITS : Intelligent Tutoring System (ex. projet Voltaire)

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & education*, 51(1), 368-384.

Champs d'études



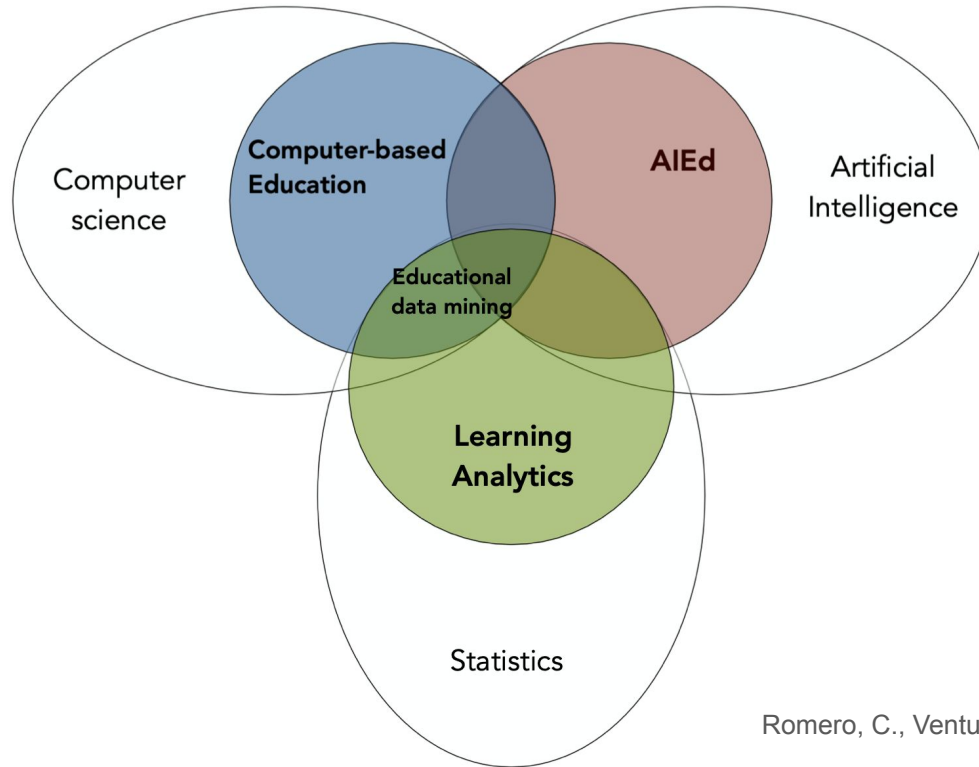
Computer-based Education : usage des ordinateurs pour donner des instructions aux apprenants

Learning Analytics : collecte, analyse, reporting sur les données d'éducation

- ❑ analyse de données (statistiques)
- ❑ visualisations
- ❑ analyse du discours
- ❑ analyse des réseaux sociaux

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & education*, 51(1), 368-384.

Champs d'études



Computer-based Education : usage des ordinateurs pour donner des instructions aux apprenants

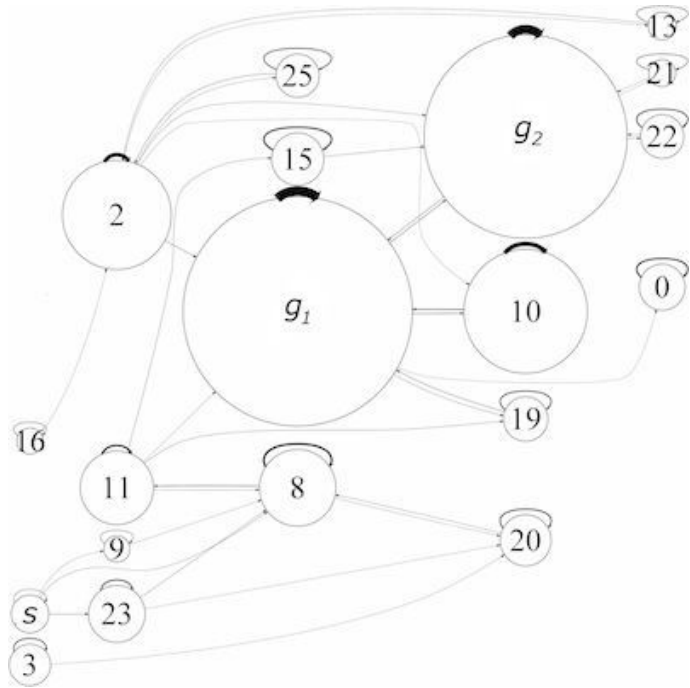
Learning Analytics : collecte, analyse, reporting sur les données d'éducation

Educational Data Mining : extraction de connaissances, détection d'événements, recommandation

- ❑ clustering, classification
- ❑ modèles Bayésiens
- ❑ extraction de relations

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & education*, 51(1), 368-384.

Exemple : Apprentissage de la programmation

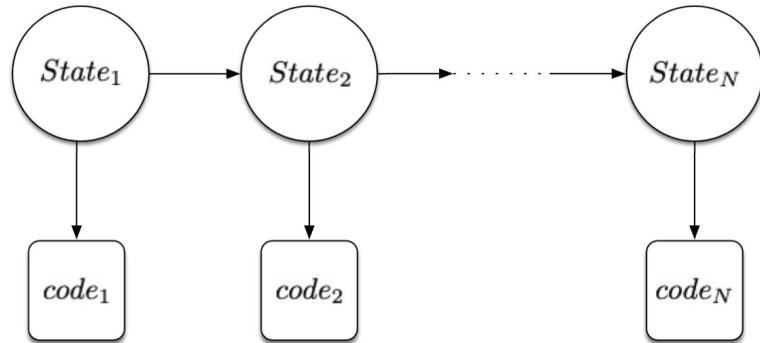


Analyse des trajectoires d'apprenants pour résoudre des problèmes de programmation.

- ❑ Trajectoire = séquences de "commit"
- ❑ **Modélisation** des trajectoires par HMM
- ❑ **Clustering** de trajectoires = principaux chemins de résolution
- ❑ **Détection/prédiction** d'apprenants bloqués
- ❑ Compréhension des mécanismes

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

Exemple : Apprentissage de la programmation

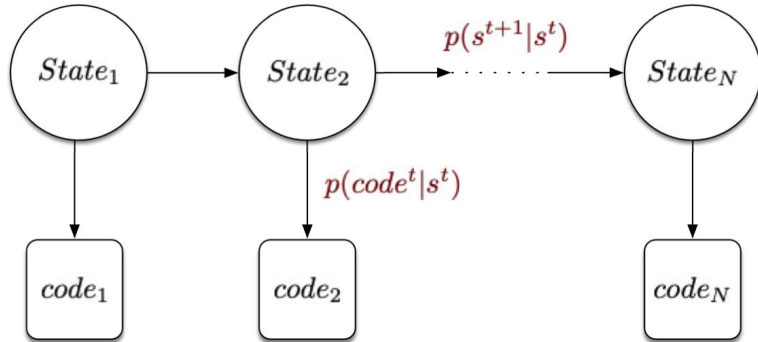


Hypothèses (HMM)

- ❑ programmer = processus incrémental
- ❑ l'étape au temps $t+1$ n'est pas indépendante de l'étape précédente (mais ne dépend que de celle-ci)

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

Exemple : Apprentissage de la programmation



Variables du modèle

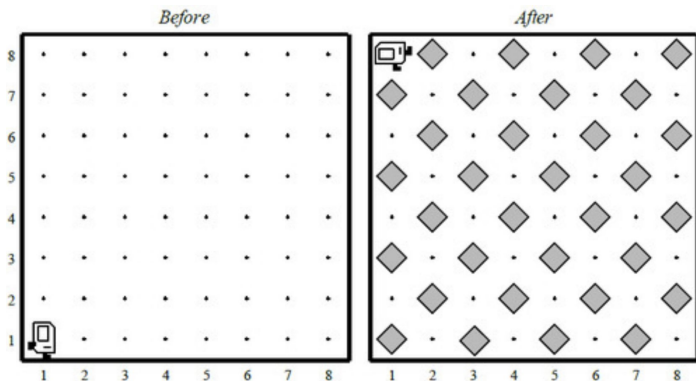
- l'ensemble des états $\{s_1, \dots, s_k\}$
- les probabilités de transitions $\{p(s^{t+1}|s^t)\}_{k \times k}$
- les probabilités d'émissions $\{p(code^t|s^t)\}$

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

Exemple : Apprentissage de la programmation

Exercice : Checkerboard Karel

- ❑ CS1 module à Stanford University
- ❑ Colorier les cases d'un échiquier (taille quelconque)

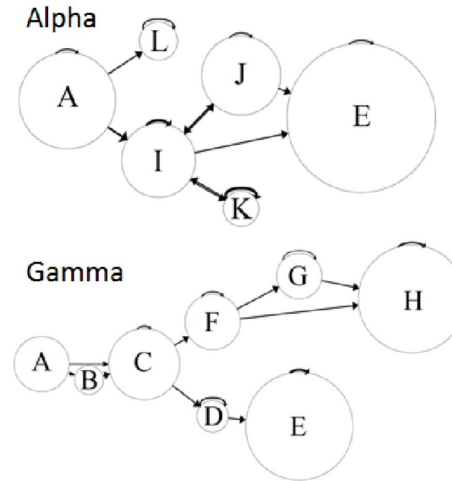
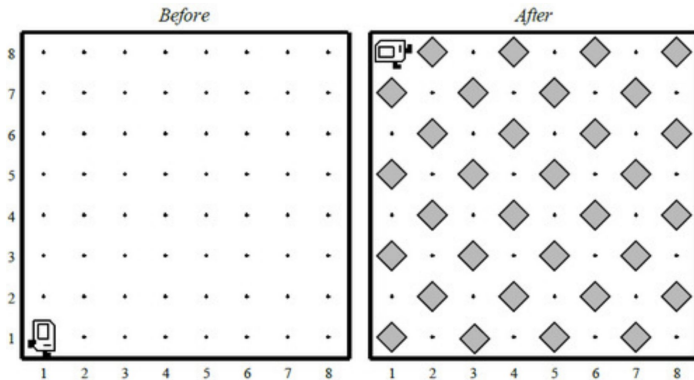


Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

Exemple : Apprentissage de la programmation

Exercice : Checkerboard Karel

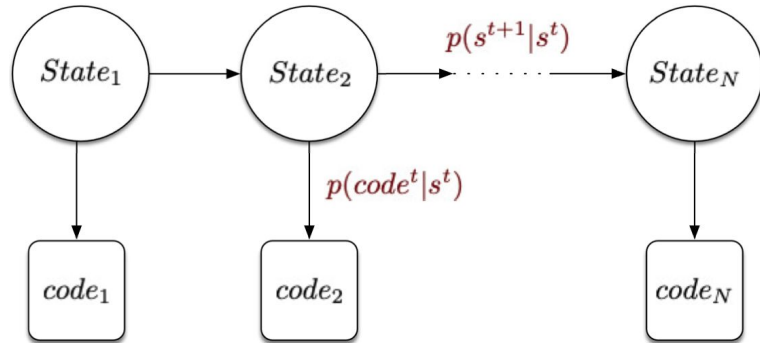
- CS1 module à Stanford University
- Colorier les cases d'un échiquier (taille quelconque)



State	Description
A	Start state.
B	Places a single beeper.
C	Places a perfect row in all worlds.
D	Places two perfect rows of the comb soln.
E	The comb soln. that works on all worlds.
F	The <u>zig zag</u> soln. that works except one col worlds.
G	The <u>zig zag</u> soln. that crashes on odd col worlds.
H	The <u>zig zag</u> soln. that works on all worlds.
I	Places one line. Misses a beeper on odd col worlds.
J	Two rows of the <u>zig zag</u> soln. Second row shifted.
K	Repeats first two rows <i>infinitely</i> .
L	Hardcoded to place first row in the 8x8 world.

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

Exemple : Apprentissage de la programmation

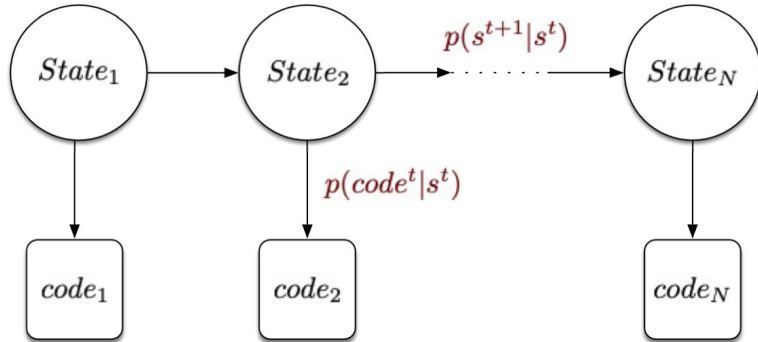


Limites de l'approche

- ❑ nécessité d'un encodage explicite des connaissances (états)
- ❑ dépendances immédiates seulement
- ❑ difficulté à représenter un programme

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

Exemple : Apprentissage de la programmation



Limites de l'approche

- ❑ nécessité d'un encodage explicite des connaissances (états)
- ❑ dépendances immédiates seulement
- ❑ difficulté à représenter un programme

... avant le Deep Learning

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 153-160).

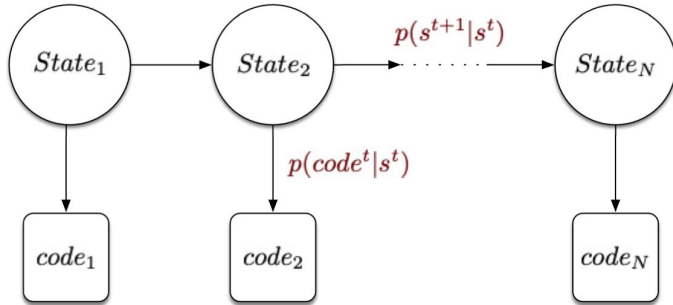
Le Présent

l'AIEd à l'ère du Deep Learning

D(L)KT : Deep (Learning) Knowledge Tracing

Limites du Bayesian Knowledge Tracing

- ❑ nécessite un encodage explicite des connaissances (états)
- ❑ dépendances immédiates seulement
- ❑ difficulté à représenter les entrées

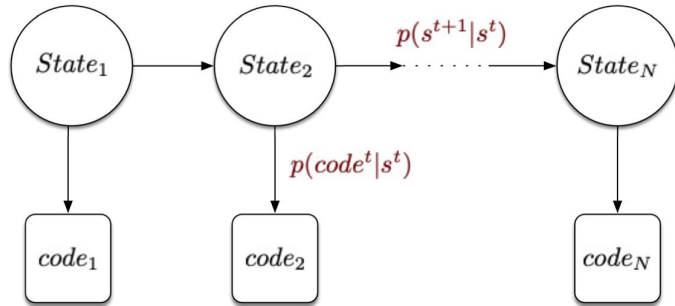


Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, 28.

D(L)KT : Deep (Learning) Knowledge Tracing

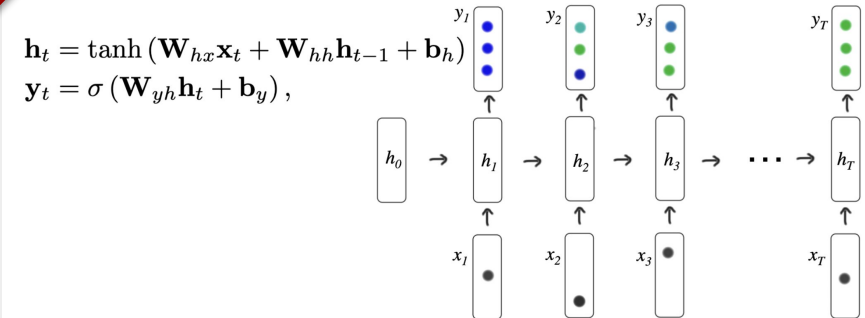
Limites du Bayesian Knowledge Tracing

- ❑ nécessite un encodage explicite des connaissances (états)
- ❑ dépendances immédiates seulement
- ❑ difficulté à représenter les entrées



Avantages des RNNs

- ❑ représentation continue des états cachés
- ❑ dépendances non-uniquement immédiates



$$\mathbf{h}_t = \tanh(\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h)$$

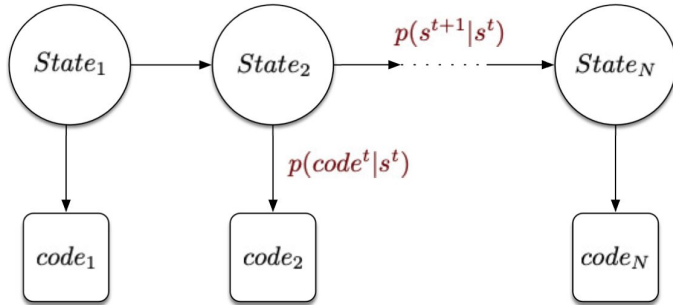
$$\mathbf{y}_t = \sigma(\mathbf{W}_{yh}\mathbf{h}_t + \mathbf{b}_y),$$

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, 28.

D(L)KT : Deep (Learning) Knowledge Tracing

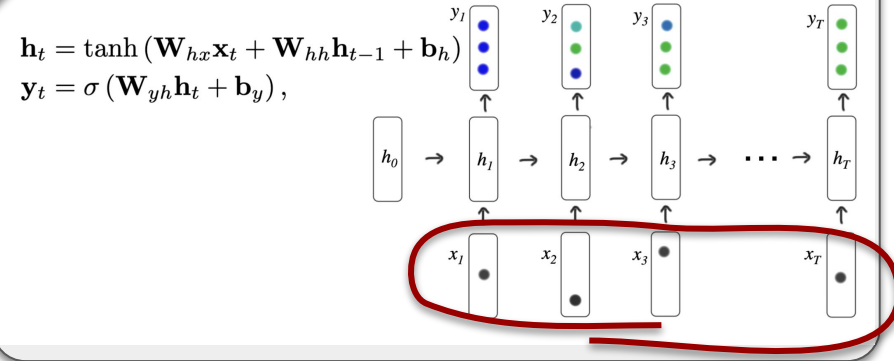
Limites du Bayesian Knowledge Tracing

- ❑ nécessite un encodage explicite des connaissances (états)
- ❑ dépendances immédiates seulement
- ❑ difficulté à représenter les entrées



Avantages des RNNs

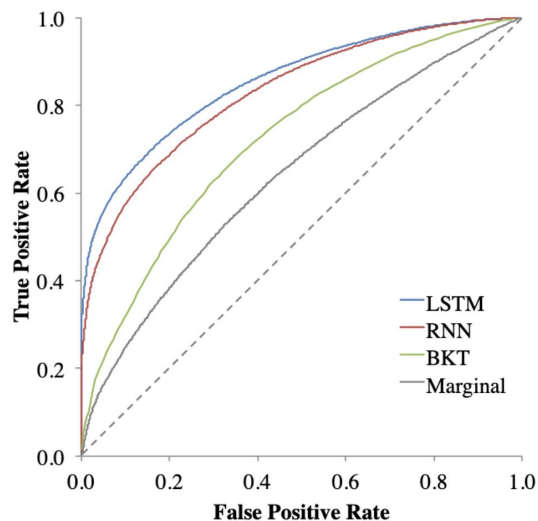
- ❑ représentation continue des états cachés
- ❑ dépendances non-uniquement immédiates
- ❑ difficulté à représenter les entrées (autrement qu'en *one-hot encoding*)



Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, 28.

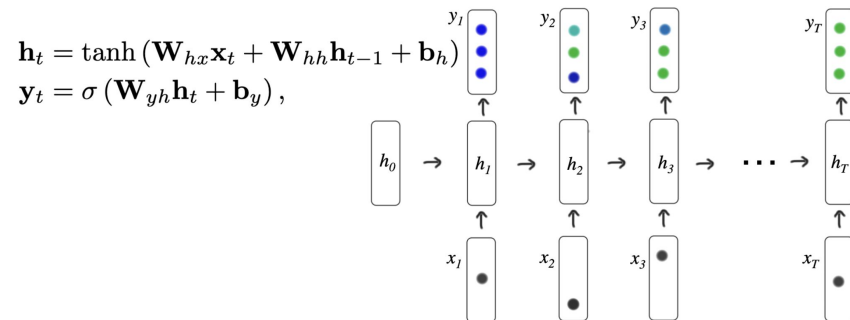
D(L)KT : Deep (Learning) Knowledge Tracing

Tâche : prédire la réussite à un exercice par rapport aux échecs/réussites précédents



Avantages des RNNs

- ❑ représentation continue des états cachés
- ❑ dépendances non-uniquement immédiates
- ❑ difficulté à représenter les entrées (autrement qu'en *one-hot encoding*)



Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, 28.

Apprentissage de représentations de programmes

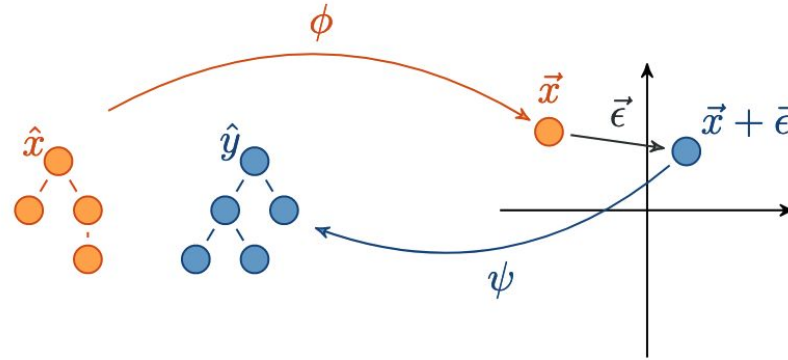
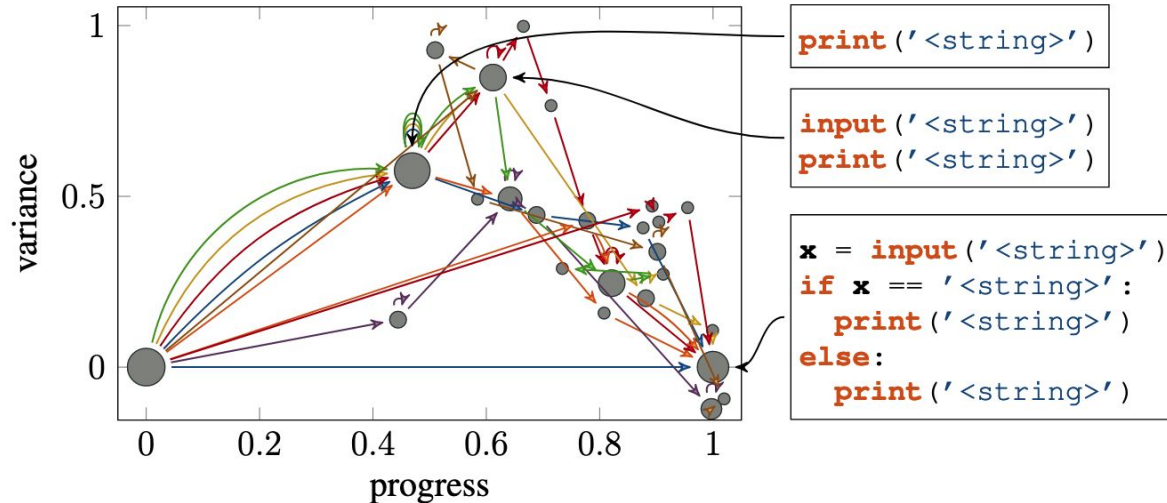


Figure 4: A high-level illustration of the (denoising) autoencoder framework. Syntax tree \hat{x} (left) gets encoded as a vector $\vec{x} = \phi(\hat{x})$ (right). We then add noise $\vec{\epsilon}$ and decode back to a tree $\hat{y} = \psi(\vec{x} + \vec{\epsilon})$. If $\hat{x} \neq \hat{y}$, we adjust the parameters of both encoder and decoder to increase the probability that $\vec{x} + \vec{\epsilon}$ is decoded into \hat{x} instead of \hat{y} .

Paassen, B., McBroom, J., Jeffries, B., Koprinska, I., & Yacef, K. (2021). Mapping python programs to vectors using recursive neural encodings. *Journal of Educational Data Mining*, 13(3), 1-35.

Visualisation des trajectoires d'apprenants

Progress-Variance Projections : Projection 2D originale et "pédagogique" de l'espace des représentations



Paassen, B., McBroom, J., Jeffries, B., Koprinska, I., & Yacef, K. (2021). Mapping python programs to vectors using recursive neural encodings. *Journal of Educational Data Mining*, 13(3), 1-35.

Système de tutorat

Apprentissage d'un système dynamique pour accompagner l'apprenant vers l'étape suivante à partir de l'analyse des trajectoires de résolution observées

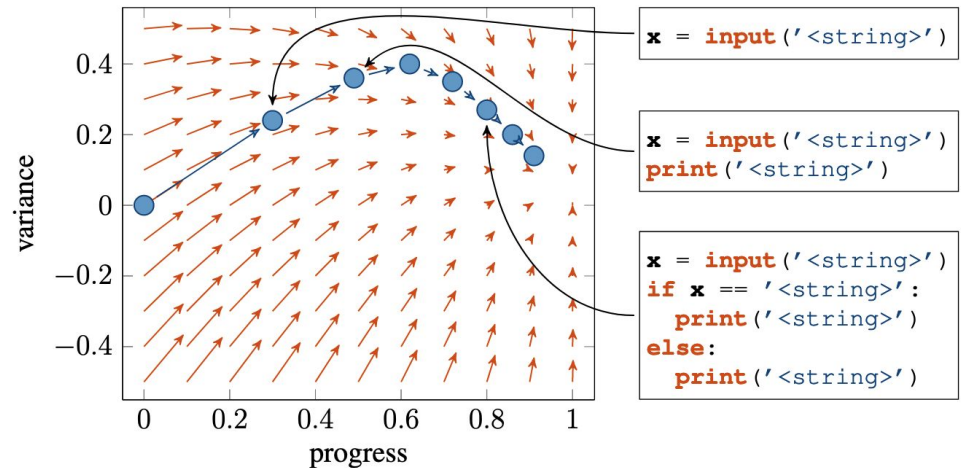
f un système (linéaire) dynamique de la forme :

$$f(\vec{x}) = \vec{x} + \mathbf{W} \cdot (\vec{x}_* - \vec{x})$$

\vec{x}_* le programme solution (*attractor*)

\mathbf{W} la matrice de paramètres (capturant la dynamique) apprise par minimisation de :

$$\min_{\mathbf{W}} \sum_{t=1}^{T-1} \|f(\vec{x}_t) - \vec{x}_{t+1}\|^2 + \lambda \cdot \|\mathbf{W}\|_{\mathcal{F}}^2$$



Paassen, B., McBroom, J., Jeffries, B., Koprinska, I., & Yacef, K. (2021). Mapping python programs to vectors using recursive neural encodings. *Journal of Educational Data Mining*, 13(3), 1-35.

Focus

L'apprentissage d'embeddings de programmes

Embeddings : définition - intuition

Un *embedding* désigne une représentation mathématique $v(x)$ d'une donnée x dans un espace de représentation porteur d'une sémantique.

- ❑ x ne possède pas *a priori* de représentation mathématique naturelle (ex. mot du langage naturel)

Embeddings : définition - intuition

Un *embedding* désigne une représentation mathématique $v(x)$ d'une donnée x dans un espace de représentation porteur d'une sémantique.

- ❑ x ne possède pas *a priori* de représentation mathématique naturelle (ex. mot du langage naturel)
- ❑ *A minima*, la sémantique portée par l'espace d'embeddings sera telle que deux données de sens proches auront des embeddings mathématiquement proches :

$$\text{sens}(x^i) \approx \text{sens}(x^j) \Rightarrow v(x^i) \approx v(x^j)$$

Embeddings : définition - intuition

Un *embedding* désigne une représentation mathématique $v(x)$ d'une donnée x dans un espace de représentation porteur d'une sémantique.

- ❑ x ne possède pas *a priori* de représentation mathématique naturelle (ex. mot du langage naturel)
- ❑ *A minima*, la sémantique portée par l'espace d'embeddings sera telle que deux données de sens proches auront des embeddings mathématiquement proches :

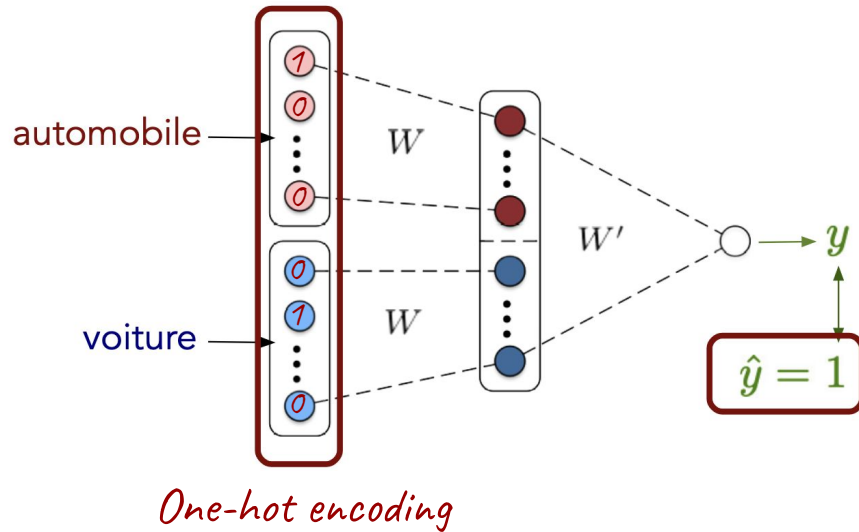
$$\text{sens}(x^i) \approx \text{sens}(x^j) \Rightarrow v(x^i) \approx v(x^j)$$

On parle d'un *plongement* (*embedding*) des données dans un espace généralement de faible dimension (quelques centaines au maximum)

Embeddings : apprentissage d'embeddings de mots

Comment apprendre une représentation "sémantique" des mots de la langue naturelle?

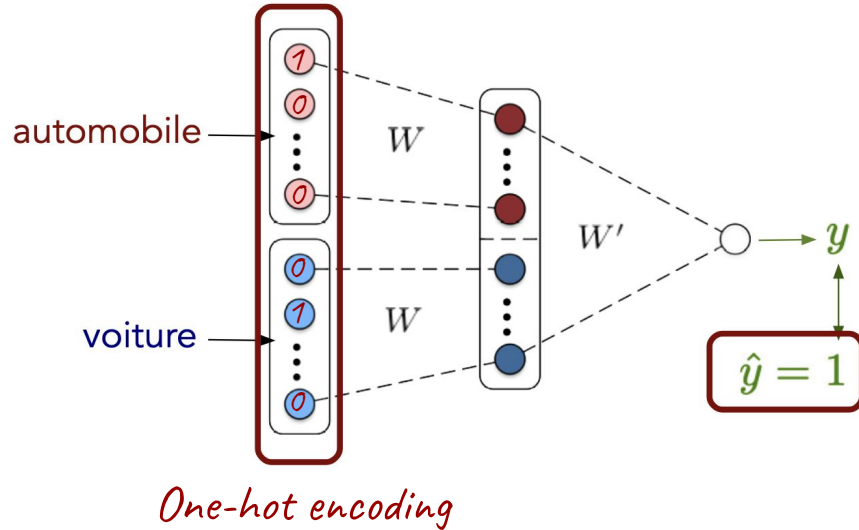
Piste 1 : entraîner un (petit) réseau de neurones à reconnaître des paires de mots synonymes



Embeddings : apprentissage d'embeddings de mots

Comment apprendre une représentation "sémantique" des mots de la langue naturelle?

Piste 1 : entraîner un (petit) réseau de neurones à reconnaître des paires de mots synonymes



OUI...

- Encodage commun aux deux mots de la paire (principe à retenir)
- Couche cachée = embeddings

MAIS...

- Nécessite des données (paires) labellisées

Embeddings : apprentissage d'embeddings de mots

Comment apprendre une représentation "sémantique" des mots de la langue naturelle?

Piste 2 : remplacer l'apprentissage supervisé par un **auto-apprentissage** (non-supervisé)

Hypothèse distributionnelle de Z. Harris (1954) :

Les mots qui apparaissent régulièrement dans les mêmes contextes linguistiques partagent des significations similaires.

Embeddings : apprentissage d'embeddings de mots

Comment apprendre une représentation "sémantique" des mots de la langue naturelle?

Piste 2 : remplacer l'apprentissage supervisé par un **auto-apprentissage** (non-supervisé)

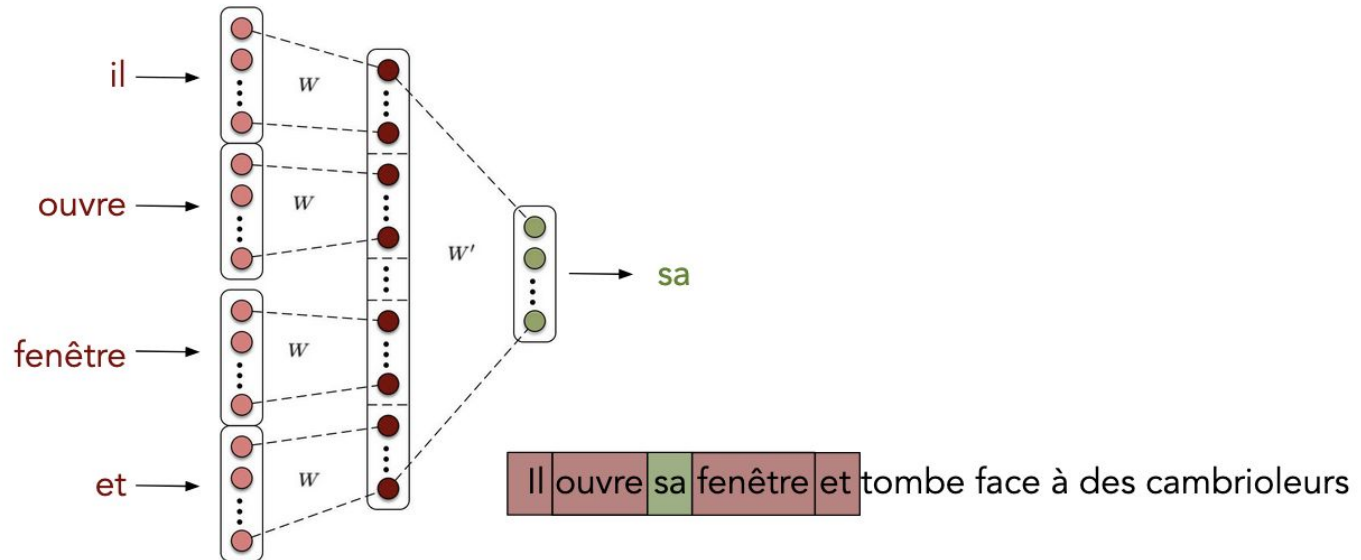
Hypothèse distributionnelle de Z. Harris (1954) :

Les mots qui apparaissent régulièrement dans les mêmes contextes linguistiques partagent des significations similaires.

- ❑ Il existe(raît) une distribution des mots de la langue permettant de prédire un mot masqué à partir des mots qui l'entourent
- ❑ Données d'entraînement : des millions de textes disponibles

Embeddings : l'algorithme word2vec

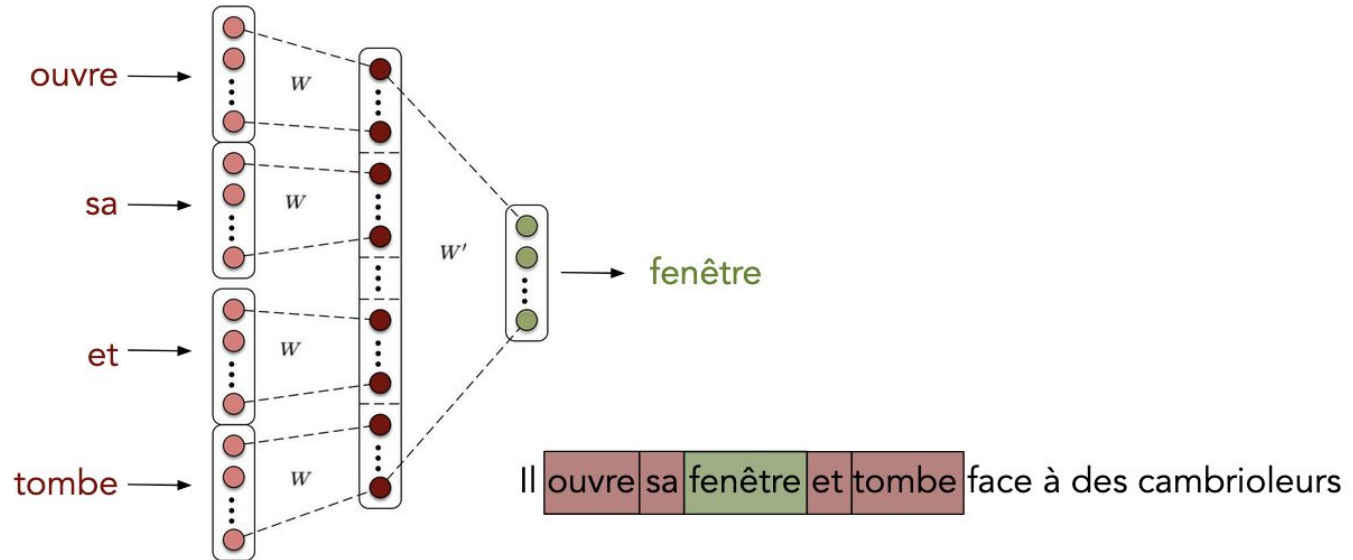
Tâche "prétexte" : entraîner un réseau de neurones à prédire un mot masqué à partir de son contexte



Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Embeddings : l'algorithme word2vec

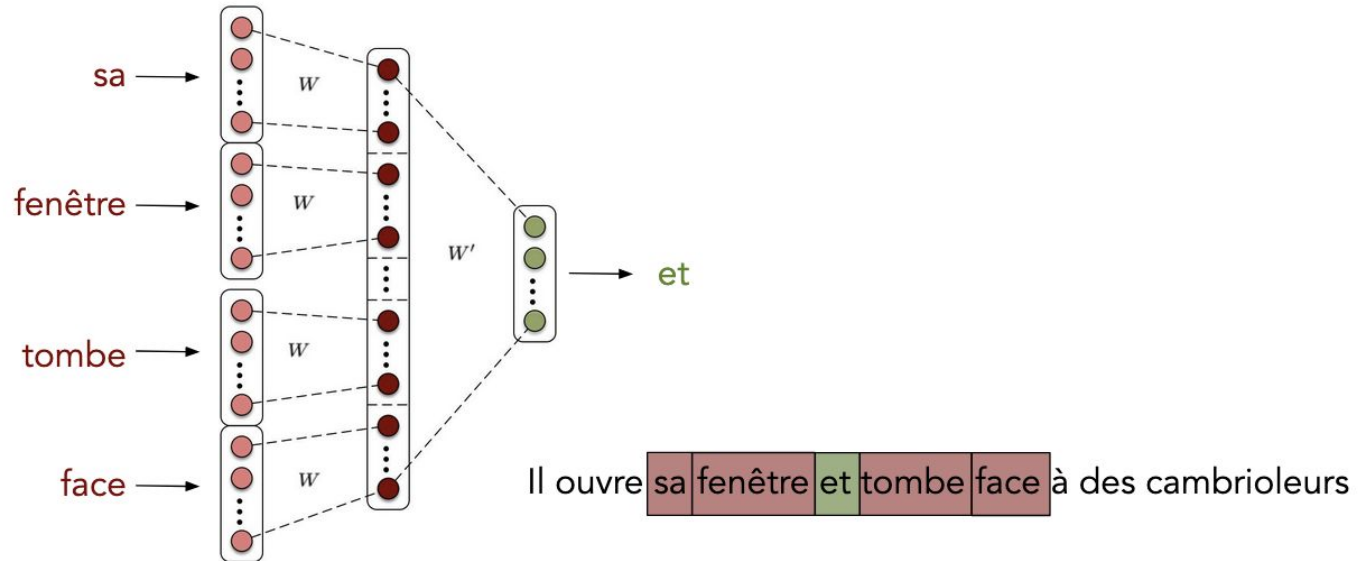
Tâche "prétexte" : entraîner un réseau de neurones à prédire un mot masqué à partir de son contexte



Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Embeddings : l'algorithme word2vec

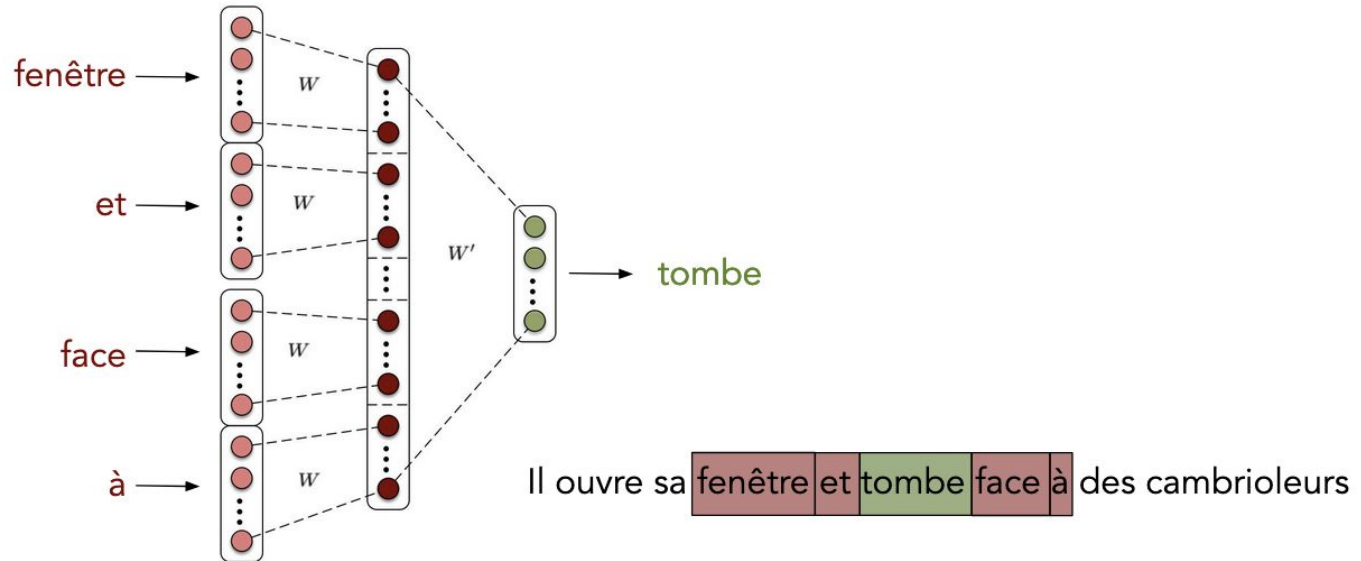
Tâche "prétexte" : entraîner un réseau de neurones à prédire un mot masqué à partir de son contexte



Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Embeddings : l'algorithme word2vec

Tâche "prétexte" : entraîner un réseau de neurones à prédire un mot masqué à partir de son contexte

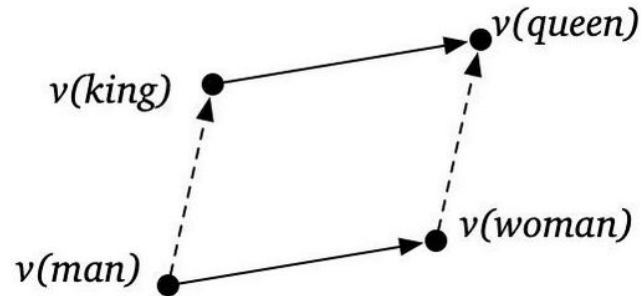


Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Embeddings : les analogies

Les embeddings de mots capture(raie)nt des relations sémantiques et syntaxiques

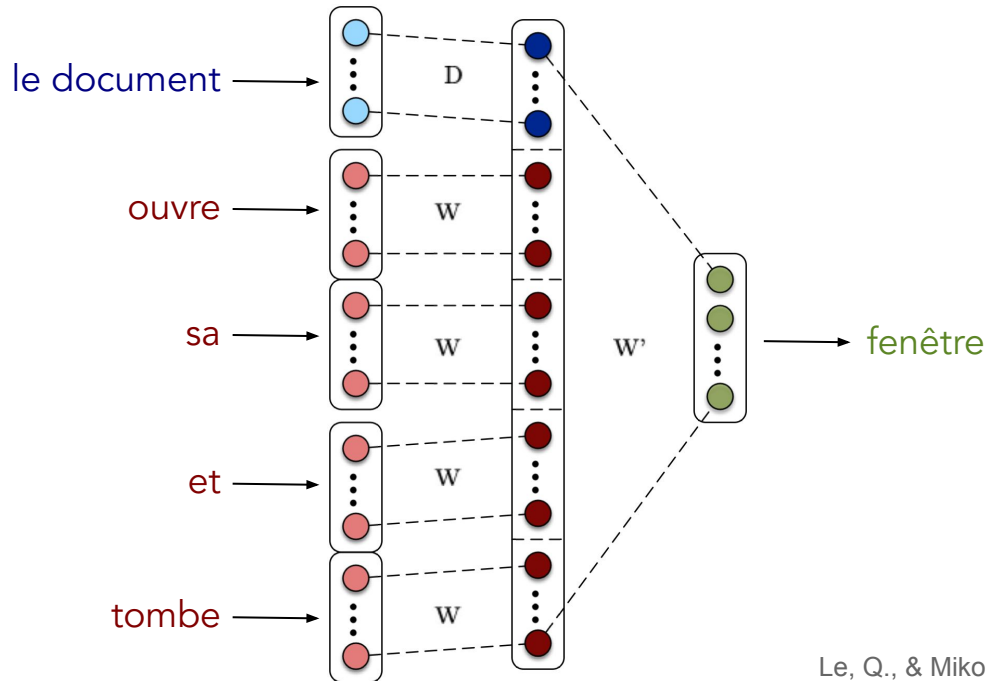
$$v(\text{king}) - v(\text{man}) + v(\text{woman}) \approx v(\text{queen})$$



Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Embeddings de "documents" : l'algorithme doc2vec

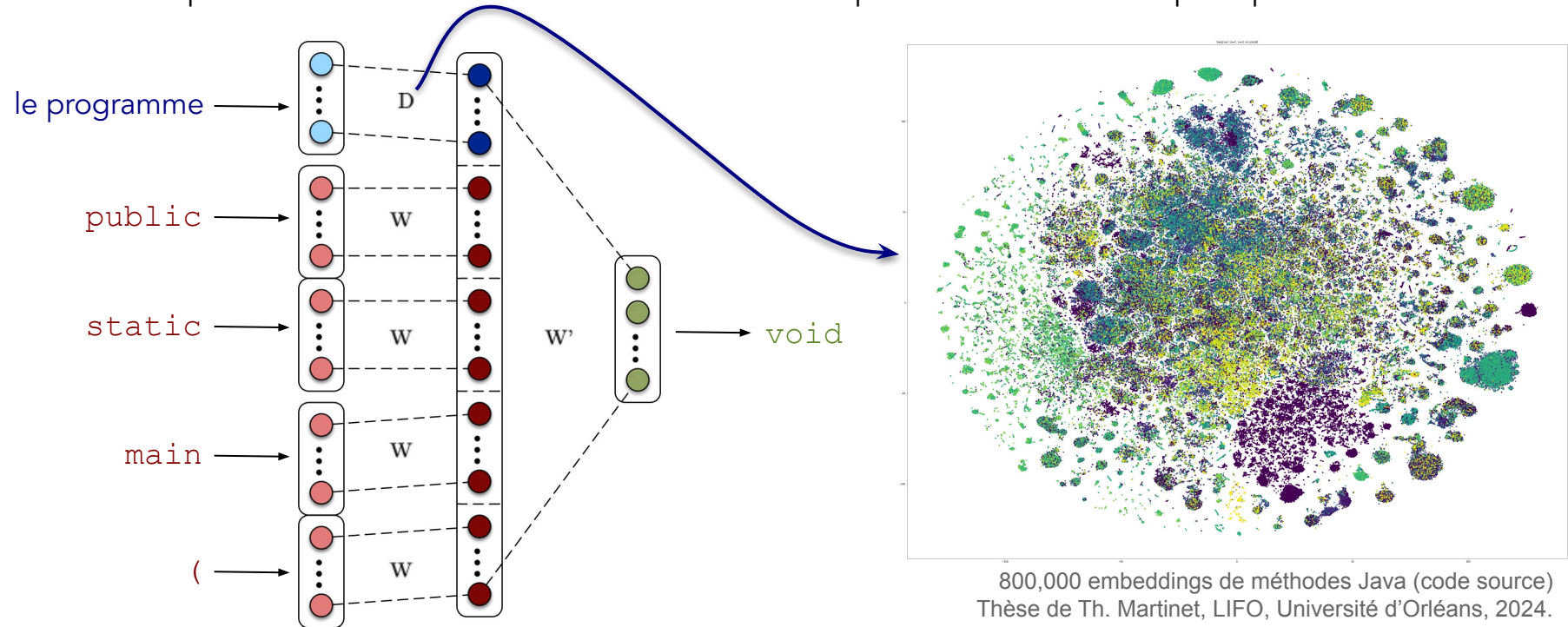
Tâche "prétexte" : entraîner un réseau de neurones à prédire un mot masqué à partir de son **contexte**



Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In International conference on machine learning (pp. 1188-1196). PMLR.

Embeddings de "programmes" : l'algorithme doc2vec

Tâche "prétexte" : entraîner un réseau de neurones à prédire un **token** masqué à partir de son contexte



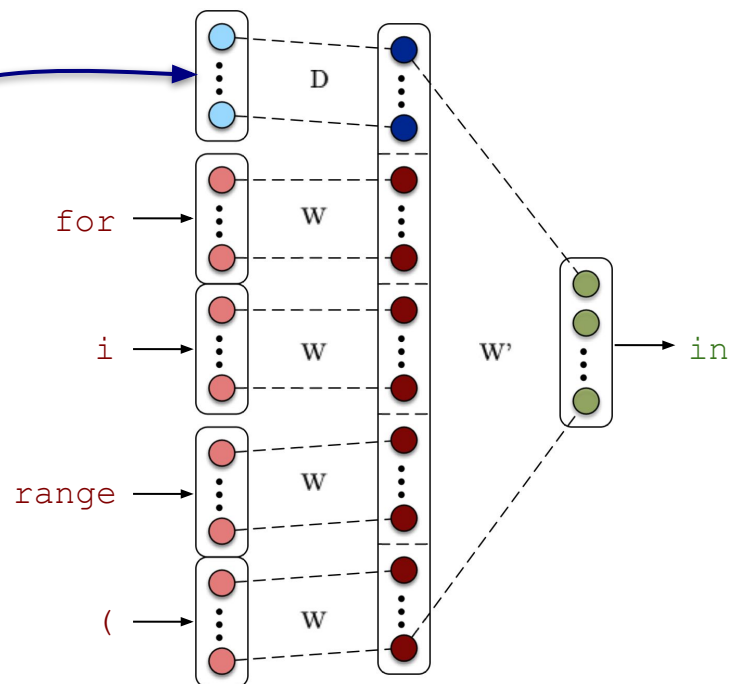
L'hypothèse distributionnelle : LN vs. LP

De (très) nombreux travaux exploitent implicitement une hypothèse distributionnelle pour apprendre des embeddings de programmes.

```

1 def minimum(liste):
2     if len(liste) == 0:
3         res = None
4     else:
5         res = liste[0]
6         for i in range(0, len(liste)):
7             if liste[i] < res :
8                 res = liste[i]
9     return res
  
```

Source : code source
Unités : tokens du code



Feng, Z., Guo, D., Tang, D., Duan, N., Feng, X., Gong, M., ... & Zhou, M. (2020). Codebert: A pre-trained model for programming and natural languages. arXiv preprint arXiv:2002.08155.

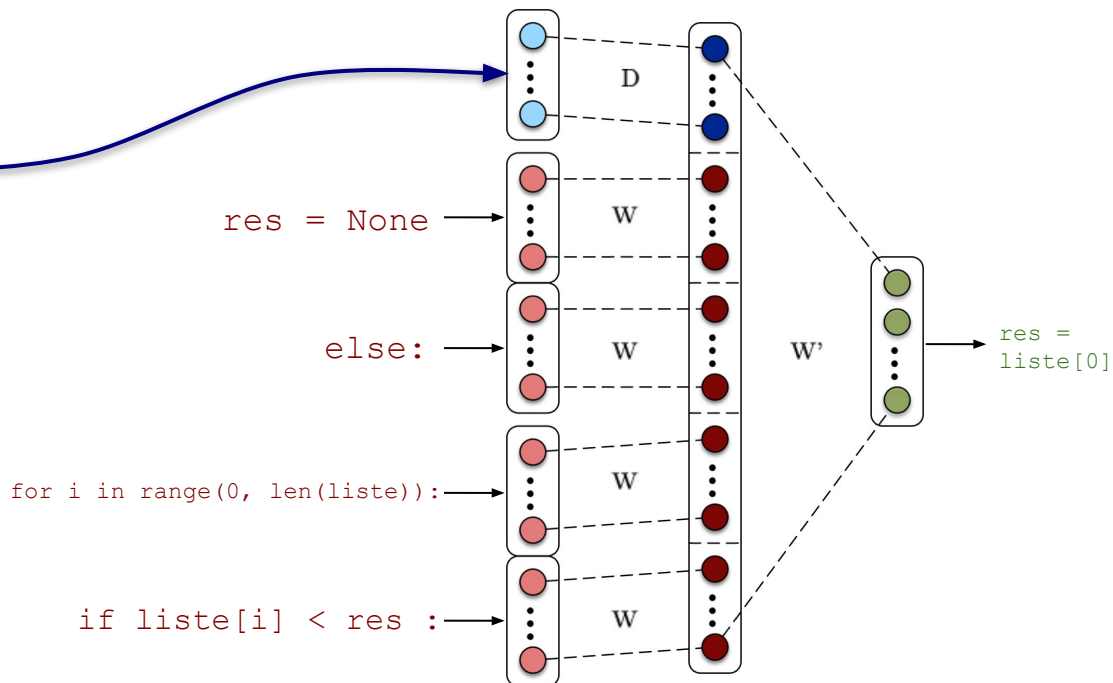
L'hypothèse distributionnelle : LN vs. LP

De (très) nombreux travaux exploitent implicitement une hypothèse distributionnelle pour apprendre des embeddings de programmes.

```

1 | def minimum(liste):
2 |     if len(liste) == 0:
3 |         res = None
4 |     else:
5 |         res = liste[0]
6 |         for i in range(0, len(liste)):
7 |             if liste[i] < res :
8 |                 res = liste[i]
9 |     return res
    
```

Source : code source
Unités : instructions du code



Ben-Nun, T., Jakobovits, A. S., & Hoefler, T. (2018). Neural code comprehension: A learnable representation of code semantics. *Advances in neural information processing systems*, 31.

L'hypothèse distributionnelle : LN vs. LP

De (très) nombreux travaux exploitent implicitement une hypothèse distributionnelle pour apprendre des embeddings de programmes.

```

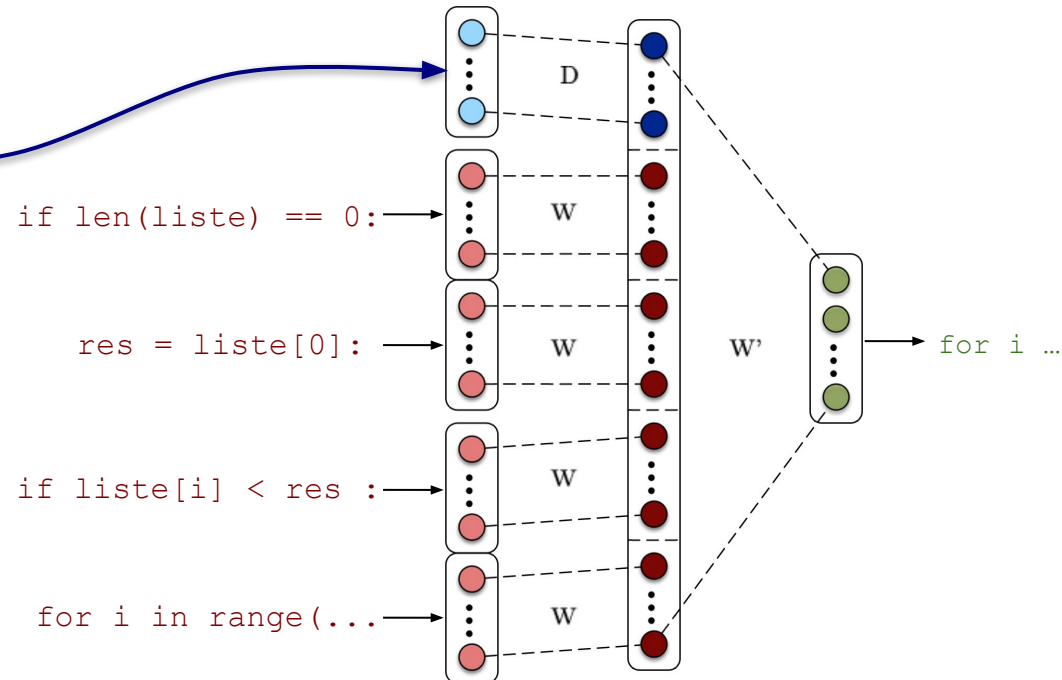
1 def minimum(liste):
2     if len(liste) == 0:
3         res = None
4     else:
5         res = liste[0]
6         for i in range(0, len(liste)):
7             if liste[i] < res :
8                 res = liste[i]
9     return res
    
```

Trace d'exécution

```

2, 5, 6, 7, 6, 7, 8, ..., 9
    
```

Source : trace d'exécution
Unités : instructions du code

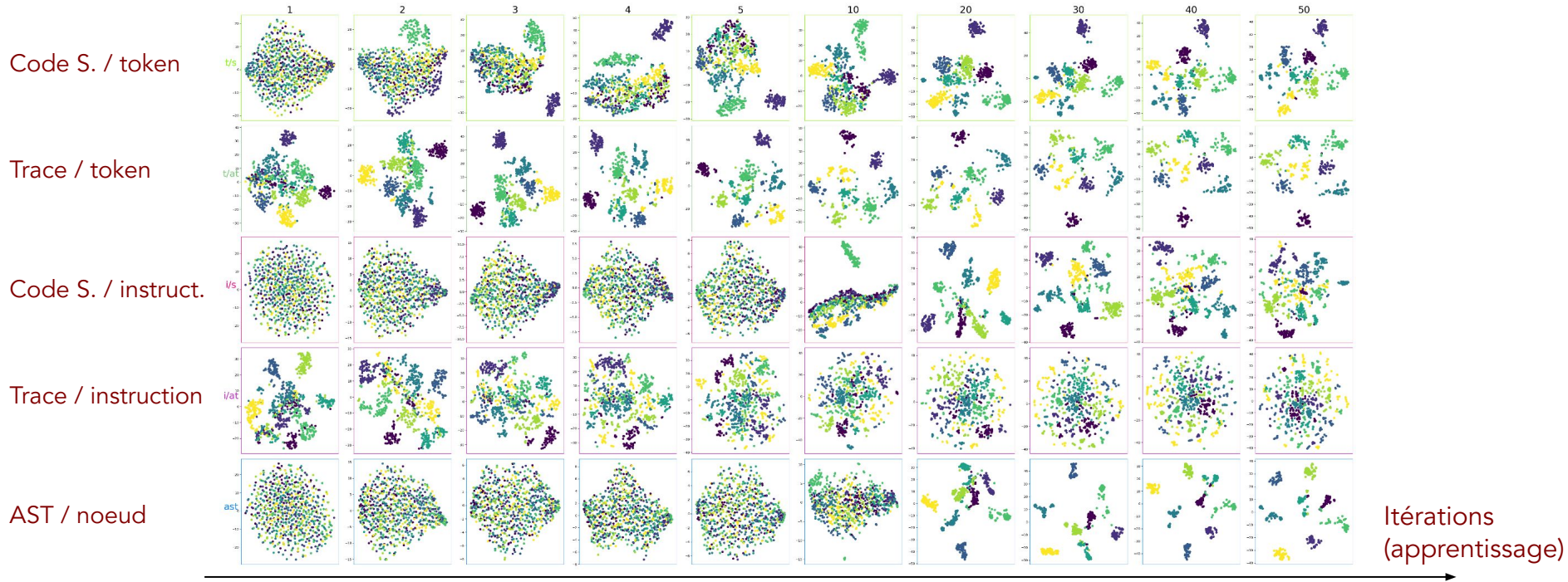


Cleuziou, G., & Flouvat, F. (2021, June).

Learning student program embeddings using abstract execution traces. In 14th International Conference on Educational Data Mining (pp. 252-262).

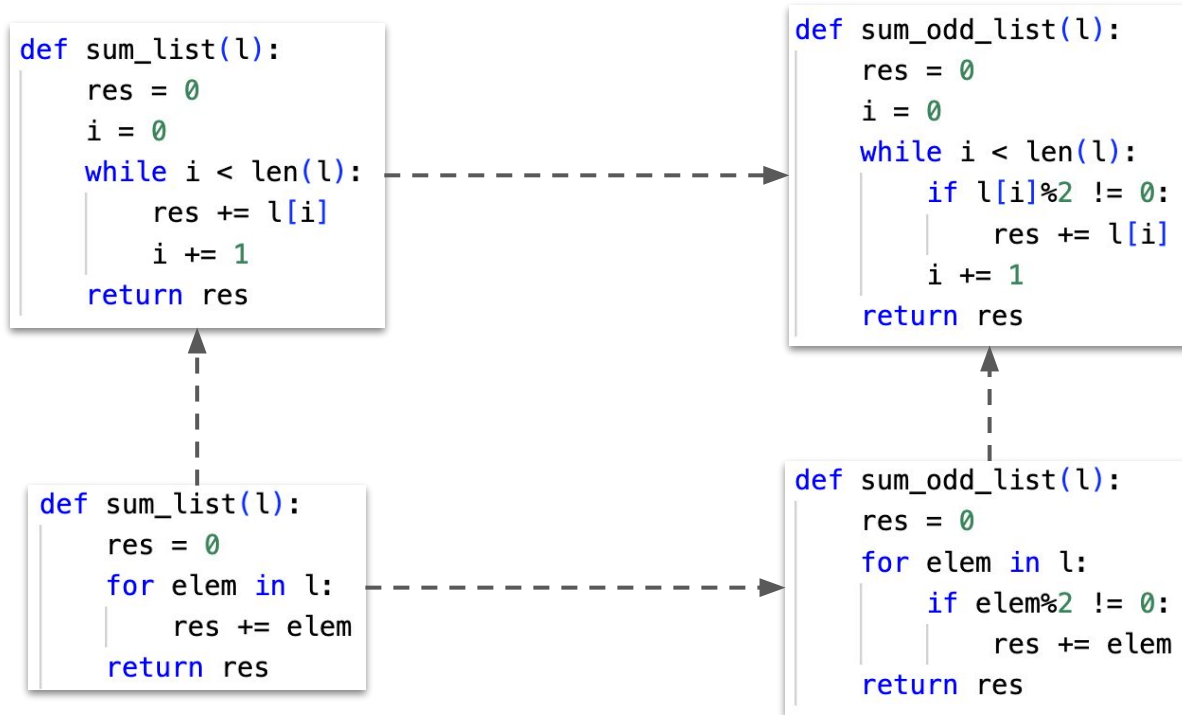
L'hypothèse distributionnelle : LN vs. LP

Quelle hypothèse distributionnelle pour apprendre des embeddings de programmes?



L'hypothèse distributionnelle : LN vs. LP

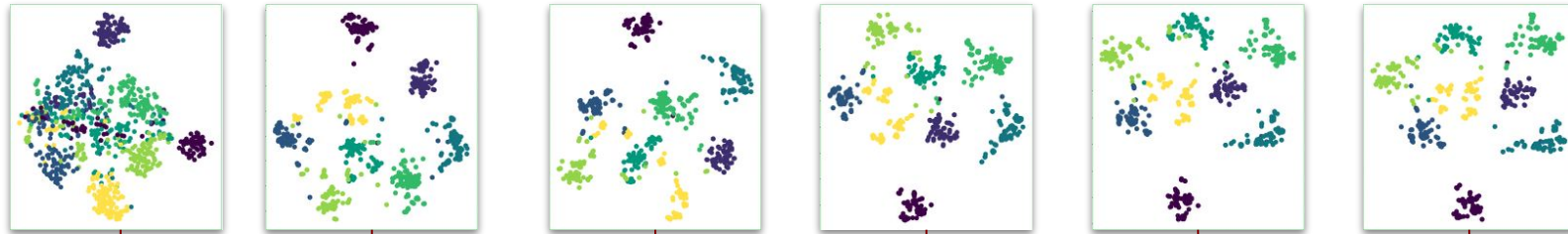
Analogies : les embeddings de programmes capturent-ils des relations syntaxiques?



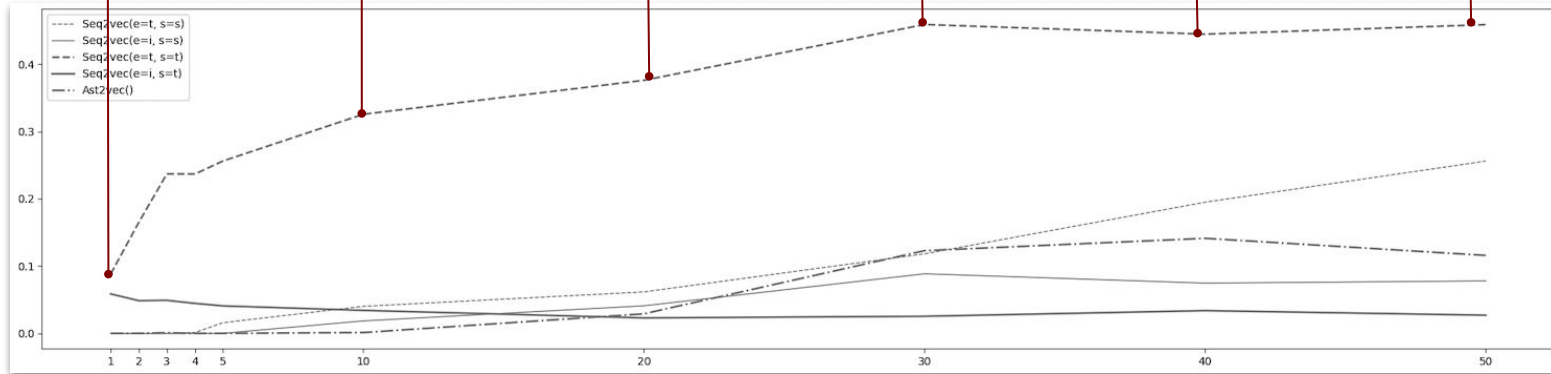
L'hypothèse distributionnelle : LN vs. LP

Quelle hypothèse distributionnelle pour apprendre des embeddings de programmes?

Trace / token



Analogies
syntaxiques entre
programmes
capturées



Le Futur (?)

AEd et IA génératives

ChatGPT : un tsunami dans le monde de l'enseignement

Menace!

 Ouest-France

New York interdit ChatGPT dans ses écoles publiques pour éviter la triche

Les écoles publiques de la ville de New York interdiront aux élèves et aux enseignants d'utiliser ChatGPT, un puissant nouvel outil...

6 janv. 2023



 Le Point

ChatGPT : Sciences Po interdit l'usage de cette intelligence artificielle

Les élèves de Sciences Po qui feront usage de ChatGPT s'exposent à des sanctions pouvant aller jusqu'à l'exclusion de l'établissement.

27 janv. 2023



 Le Figaro Etudiant

ChatGPT: aubaine pour les étudiants, casse-tête pour les enseignants

TÉMOIGNAGES - À l'Insee Lyon, des étudiants ont rendu des copies quasi similaires, sans être pénalisés. Un parfait exemple du phénomène...

27 janv. 2023



ChatGPT : un tsunami dans le monde de l'enseignement

Menace!

Opportunité?



Quest-France

New York interdit ChatGPT dans ses écoles publiques pour éviter la triche



Les écoles publiques de la ville de New York interdiront aux élèves et aux enseignants d'utiliser ChatGPT, un puissant nouvel outil...

6 janv. 2023

Le Point

ChatGPT : Sciences Po interdit l'usage de cette intelligence artificielle



Les élèves de Sciences Po qui feront usage de ChatGPT s'exposent à des sanctions pouvant aller jusqu'à l'exclusion de l'établissement.

27 janv. 2023

Le Figaro Etudiant

ChatGPT: aubaine pour les étudiants, casse-tête pour les enseignants



TÉMOIGNAGES - À l'Insee Lyon, des étudiants ont renoncé à utiliser ChatGPT sans être pénalisés. Un parfait exemple du phénomène.

27 janv. 2023

Toms Guide

ChatGPT : les professeurs ne sauront jamais si les élèves trichent, affirme OpenAI



OpenAI affirme que les professeurs ne pourront jamais détecter l'usage de ChatGPT par les élèves, car les détecteurs d'IA sont inefficaces.

5 sept. 2023

ChatGPT : un tsunami dans le monde de l'enseignement

Menace!

Opportunité?

Ouest-France

New York interdit ChatGPT dans ses écoles publiques pour éviter la triche

Les écoles publiques de la ville de New York interdiront aux élèves et aux enseignants d'utiliser ChatGPT, un puissant nouvel outil...

6 janv. 2023



Le Point

ChatGPT : Sciences Po interdit l'usage de cette intelligence artificielle

Les élèves de Sciences Po qui feront usage de ChatGPT s'exposent à des sanctions pouvant aller jusqu'à l'exclusion de l'établissement.

27 janv. 2023



Le Figaro Etudiant

ChatGPT: aubaine pour les étudiants, casse-tête pour les enseignants

TÉMOIGNAGES - À l'Insee Lyon, des étudiants ont renoncé à utiliser ChatGPT sans être pénalisés. Un parfait exemple du phénomène.

27 janv. 2023



Toms Guide

ChatGPT : les professeurs ne sauront jamais si les élèves trichent, affirme OpenAI

OpenAI affirme que les professeurs ne pourront jamais détecter l'usage de ChatGPT par les élèves, car les détecteurs d'IA sont inefficaces.

5 sept. 2023



LeBigData

Les profs détestaient ChatGPT en 2023, ils vont l'adorer en 2024

Malgré l'interdiction de ChatGPT dans plusieurs établissements, les enseignants vont adopter l'outil dans l'enseignement.

27 août 2023



Bilan.ch

Innovation – De l'interdiction à l'adoption: comment les Universités répondent à ChatGPT

Alors qu'il était surtout question d'interdire ChatGPT dans l'éducation, les universités ont été contraintes de reconsidérer leur position...

12 juil. 2023



Toms Guide

ChatGPT n'est plus interdit dans les écoles de New York, les élèves encouragés à utiliser l'IA

En début d'année, New York avait interdit ChatGPT dans ses écoles, craignant que l'IA nuise à l'apprentissage des élèves.

19 mai 2023



Dompter ChatGPT

G You

Je suis un étudiant qui apprend les bases de la programmation à l'aide du langage Python. **Comporte toi comme un enseignant** et aide-moi à construire une solution étape par étape en me demandant chaque fois une réalisation simple **sans me donner les solutions** ni d'exemple de code et corrige-moi à chaque étape.

Je dois écrire une fonction qui calcule et renvoie la somme des nombres impairs d'une liste passée en paramètre. Donne-moi d'abord la première étape seulement.

Dompter ChatGPT

G You

Je suis un étudiant qui apprend les bases de la programmation à l'aide du langage Python. Comporte toi comme un enseignant et aide-moi à construire une solution étape par étape en me demandant chaque fois une réalisation simple sans me donner les solutions ni d'exemple de code et corrige-moi à chaque étape.
Je dois écrire une fonction qui calcule et renvoie la somme des nombres impairs d'une liste passée en paramètre. Donne-moi d'abord la première étape seulement.



ChatGPT

Bien sûr, c'est une excellente façon d'apprendre! Commençons par la première étape:

Étape 1 : Signature de la fonction

Demande : Peux-tu définir la signature de la fonction? Cela inclut le nom de la fonction, les paramètres qu'elle prend et le type de la valeur renvoyée.

Dompter ChatGPT

G You

Je suis un étudiant qui apprend les bases de la programmation à l'aide du langage Python. Comporte-toi comme un enseignant et aide-moi à construire une solution étape par étape en me demandant chaque fois une réalisation simple sans me donner les solutions ni d'exemple de code et corrige-moi à chaque étape.
Je dois écrire une fonction qui calcule et renvoie la somme des nombres impairs d'une liste passée en paramètre. Donne-moi d'abord la première étape seulement.



ChatGPT

Bien sûr, c'est une excellente façon d'apprendre! Commençons par la première étape:

Étape 1: Signature de la fonction

Demande : Peux-tu définir la signature de la fonction? Cela inclut le nom de la fonction, les paramètres qu'elle prend et le type de la valeur renvoyée.

G

You

je dois utiliser def?

Dompter ChatGPT

G You

Je suis un étudiant qui apprend les bases de la programmation à l'aide du langage Python. Comporte toi comme un enseignant et aide-moi à construire une solution étape par étape en me demandant chaque fois une réalisation simple sans me donner les solutions ni d'exemple de code et corrige-moi à chaque étape.
Je dois écrire une fonction qui calcule et renvoie la somme des nombres impairs d'une liste passée en paramètre. Donne-moi d'abord la première étape seulement.



ChatGPT

Bien sûr, c'est une excellente façon d'apprendre! Commençons par la première étape:

Étape 1: Signature de la fonction

Demande : Peux-tu définir la signature de la fonction? Cela inclut le nom de la fonction, les paramètres qu'elle prend et le type de la valeur renvoyée.

G You

je dois utiliser def?



ChatGPT

Exactement, pour définir une fonction en Python, on utilise le mot-clé `def`. Peux-tu essayer de définir la signature de la fonction en utilisant `def`?

Dompter ChatGPT

G

You

Je suis un étudiant qui apprend les bases de la programmation à l'aide du langage Python. Comporte toi comme un enseignant et aide-moi à construire une solution étape par étape en me demandant chaque fois une réalisation simple sans me donner les solutions ni d'exemple de code et corrige-moi à chaque étape.

Je dois écrire une fonction qui calcule et renvoie la somme des nombres impairs d'une liste passée en paramètre. Donne-moi d'abord la première étape seulement.



ChatGPT

Bien sûr, c'est une excellente façon d'apprendre! Commençons par la première étape:

Étape 1 : Signature de la fonction

Demande : Peux-tu définir la signature de la fonction? Cela inclut le nom de la fonction, les paramètres qu'elle prend et le type de la valeur renvoyée.



ChatGPT

Parfait, tu as ajouté le paramètre "liste" à la signature de la fonction. Maintenant, passons à l'étape suivante.

Étape 2 : Initialisation de la somme

Demande : À l'intérieur de la fonction, commence par initialiser une variable qui va contenir la somme des nombres impairs. Comment pourrais-tu nommer cette variable?



ChatGPT

Bien joué! Tu as initialisé la variable `'somme_impairs'` à zéro. Passons maintenant à la prochaine étape.

Étape 3 : Parcours de la liste

Demande : À l'intérieur de la fonction, comment pourrais-tu parcourir chaque élément de la liste? Utilise une boucle pour cela.



ChatGPT

Exactement, maintenant tu itères sur chaque élément de la liste avec la boucle `'for i in liste'`. Passons à l'étape suivante.

Étape 4 : Condition pour les nombres impairs

Demande : À l'intérieur de la boucle, comment pourrais-tu vérifier si un nombre est impair? Utilise une structure de contrôle pour cela.



ChatGPT

Parfait! Tu as correctement ajusté la condition pour vérifier si `'i'` est impair. Maintenant, passons à l'étape suivante.

Étape 5 : Mise à jour de la somme

Demande : À l'intérieur de la boucle, si `'i'` est impair, comment pourrais-tu mettre à jour la variable `'somme_impairs'` pour ajouter la valeur de `'i'` à la somme actuelle?



ChatGPT

Étape 6 : Retourner le résultat

Demande : À la fin de la fonction, comment pourrais-tu retourner la somme des nombres impairs calculée dans la variable `'somme_impairs'`?

Dompter ChatGPT

G You

Je suis un étudiant qui apprend les bases de la programmation à l'aide du langage Python. Comporte toi comme un enseignant et aide-moi à construire une solution étape par étape en me demandant chaque fois une réalisation simple sans me donner les solutions ni d'exemple de code et corrige-moi à chaque étape.
Je dois écrire une fonction qui calcule et renvoie la somme des nombres impairs d'une liste passée en paramètre. Donne-moi d'abord la première étape seulement.

ChatGPT

Bien sûr, c'est une excellente façon d'apprendre! Commençons par la première étape:

Étape 1 : Signature de la fonction

Demande : Peux-tu définir la signature de la fonction? Cela inclut le nom de la fonction, les paramètres qu'elle prend et le type de la valeur renvoyée.

ChatGPT

Parfait, tu as ajouté le paramètre "liste" à la signature de la fonction. Maintenant, passons à l'étape suivante.

Étape 2 : Initialisation de la somme

Demande : À l'intérieur de la fonction, commence par initialiser une variable qui va contenir la somme des nombres impairs. Comment pourrais-tu nommer cette variable?

ChatGPT

Bien joué! Tu as initialisé la variable `'somme_impairs'` à zéro. Passons maintenant à la prochaine étape.

Étape 3 : Parcours de la liste

Demande : À l'intérieur de la fonction, comment pourrais-tu parcourir chaque élément de la liste? Utilise une boucle pour cela.

ChatGPT

Exactement, maintenant tu itères sur chaque élément de la liste avec la boucle `'for i in liste'`. Passons à l'étape suivante.

Étape 4 : Condition pour les nombres impairs

Demande : À l'intérieur de la boucle, comment pourrais-tu vérifier si un nombre est impair? Utilise une structure de contrôle pour cela.

ChatGPT

Parfait! Tu as correctement ajusté la condition pour vérifier si `'i'` est impair. Maintenant, passons à l'étape suivante.

Étape 5 : Mise à jour de la somme

Demande : À l'intérieur de la boucle, si `'i'` est impair, comment pourrais-tu mettre à jour la variable `'somme_impairs'` pour ajouter la valeur de `'i'` à la somme actuelle?

ChatGPT

Étape 6 : Retourner le résultat

Demande : À la fin de la fonction, comment pourrais-tu retourner la somme des nombres impairs calculée dans la variable `'somme_impairs'`?

Merci pour votre attention

