

|                                                                 |                                         |                                                     |   |
|-----------------------------------------------------------------|-----------------------------------------|-----------------------------------------------------|---|
| Nombre de participant·e·s maximal pour l'atelier                | 18                                      | Nombre de sessions de 45mn que vous pouvez réaliser | 1 |
| Matériel nécessaire (vidéo-projection, grande salle, autre ...) | Vidéo-projection et connexion Internet. |                                                     |   |

# Jeu éducatif sur les biais de l'intelligence artificielle

**Biljana Petreska von Ritter-Zahony**

*Haute École Pédagogique Vaud, 1007 Lausanne, Suisse*

**RÉSUMÉ** L'atelier présente un jeu éducatif sur l'intelligence artificielle, conçu pour introduire les principes fondamentaux de l'apprentissage automatique au lycée, sans prérequis techniques ou mathématiques. Construit autour d'un scénario optimisé pour l'apprentissage, le jeu aborde également des sujets sensibles, tels que les biais raciaux ou de genre. D'une durée d'environ quinze minutes, il est disponible en ligne et jouable sur tablette ou ordinateur. Selon un questionnaire, 56 élèves sur 91 estiment avoir appris « beaucoup » ou « suffisamment ». La note moyenne attribuée par les 131 participant·e·s est de 4,8 sur 6. Le point fort le plus fréquemment cité est : « on apprend en jouant », et le jeu a déjà été joué plusieurs milliers de fois.

**MOTS-CLÉS** • Jeu éducatif – Intelligence artificielle – Biais – Bulles de filtre – Chambres d'écho – Enjeux sociétaux



*Fig. 1. Capture d'écran de la première étape du jeu. Les participant·e·s prennent le rôle d'un algorithme de décision.*

## 1. Déroulement de l'atelier

Dans un contexte où la littératie à l'intelligence artificielle est appelée à prendre une place croissante dans l'enseignement de l'informatique au lycée, les ressources pédagogiques disponibles sont limitées. Le jeu éducatif que nous proposons vise à répondre à ce besoin.

Dans un premier temps, le jeu sera expérimenté collectivement de manière à permettre aux participant·e·s de le découvrir. Le jeu est disponible en ligne et accessible gratuitement.

Dans un deuxième temps, après ce moment de jeu collectif, nous allons échanger sur les choix de conception du jeu, analyser ses apports pédagogiques et rechercher des pistes d'amélioration. En particulier, une mise en commun est prévue autour des objectifs d'apprentissage des quatre chapitres :

1. **Biais des données.** Comprendre comment l'intelligence artificielle fonctionne, comment un algorithme « apprend » à partir de données. Comprendre pourquoi les systèmes d'IA font des erreurs.
2. **Biais sociétal.** Comprendre les mécanismes qui produisent le biais sociétal et les problèmes que cela soulève.
3. **Bulles de filtres.** Comprendre la création et les dangers des bulles de filtres.
4. **Esprit critique.** Questionner les décisions des systèmes d'IA et moduler le degré de confiance accordé.

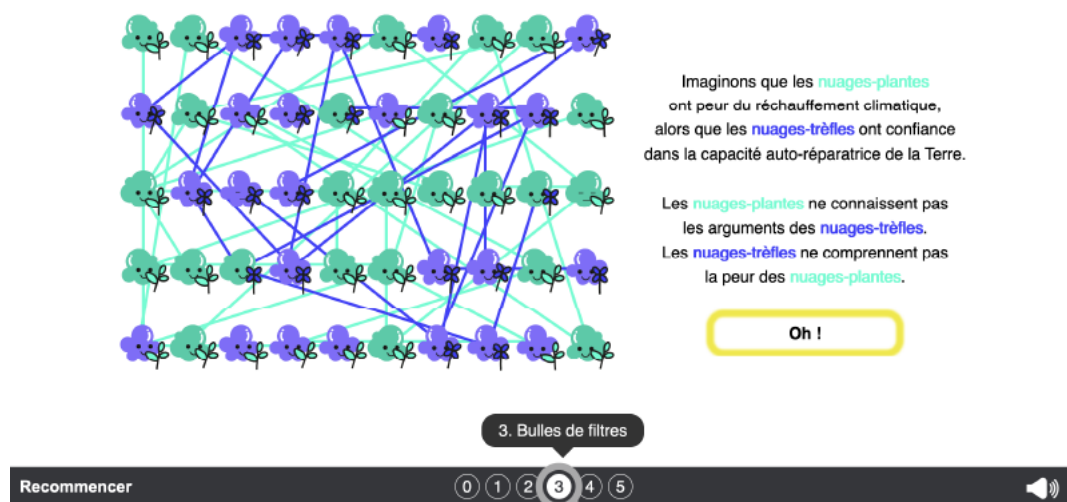


Fig. 2. Capture d'écran de la troisième étape du jeu. Représentation visuelle des bulles de filtre.

## 2. Description de la ressource / activité / projet présenté dans l'atelier

Le jeu s'articule autour de quatre chapitres. Au début du jeu, les participant·e·s sont placé·e·s dans le rôle d'un algorithme de décision (voir Fig. 1). Sur la base d'observations choisies au hasard, elle doit se prononcer sur une question très simple : y a-t-il plus de nuages indigos ou plus de nuages verts à l'écran ? Cette question rend explicite la relation entre les prédictions d'une intelligence artificielle « moderne » et les données qui lui permettent de faire cette prédiction. À des fins didactiques, la prédiction est réduite au simple acte de compter. Les observations sont soigneusement choisies, afin d'induire la personne qui joue en erreur et lui faire vivre le biais des données. Cette manipulation permet de mettre en évidence l'importance de la qualité et de la quantité de données.

Dans le deuxième chapitre, la métaphore des nuages permet d'aborder des sujets plus sensibles, tels que les biais raciaux ou de genre, sans heurter les personnes qui jouent. Les participant·e·s sont amené·e·s à formuler une prédiction sur la plante préférée des petits nuages, prédiction qui se révèle discriminante. Cette situation permet d'aborder le phénomène des biais sociétaux, en montrant comment les données peuvent reproduire – voire amplifier – les biais et préjugés présents dans la société.

Le troisième chapitre s'appuie sur les plantes préférées des petits nuages pour les connecter entre eux, créant ainsi deux bulles de filtre (ou chambres d'écho), telles qu'on peut en observer sur les réseaux sociaux (voir Fig. 2). Les bulles de filtre sont rendues visibles à l'écran et montrent comment une simple préférence peut polariser deux communautés.

Le dernier chapitre montre que, lorsque les données sont peu représentatives, l'algorithme ne parvient pas à produire des prédictions fiables, mais que ces prédictions peuvent néanmoins influencer les comportements humains. Nous concluons le jeu en soulignant la nécessité de questionner les décisions prises par les algorithmes d'intelligence artificielle.

### 3. Expérimentations réalisées ou envisagées

« Les biais de l'IA » est un jeu éducatif conçu pour introduire le fonctionnement de l'apprentissage automatique et sensibiliser aux principaux enjeux associés, de manière à la fois efficace (en environ quinze minutes) et accessible (sans prérequis technique ou mathématique).

Un questionnaire pre-activité indique que le jeu a été joué plusieurs milliers de fois. L'analyse des réponses de 131 élèves de lycée à un questionnaire post-activité suggère que le jeu est adapté au niveau attendu. La majorité des élèves a apprécié le jeu et a indiqué qu'il avait contribué à leur apprentissage de l'intelligence artificielle.

### 4. Liens avec la recherche (si pertinent)

L'intelligence artificielle (IA) est un domaine d'innovation en évolution rapide, dont l'impact sur la société — y compris sur l'éducation et la formation — est désormais majeur. Il apparaît donc essentiel d'aborder l'IA à l'école en tant qu'objet d'apprentissage. Cependant, la dimension mathématique des modèles d'apprentissage automatique constitue un obstacle réel au développement d'une véritable littératie de l'IA (Ng et al., 2021).

Notre démarche s'articule autour de deux enjeux sociétaux : les biais des systèmes d'IA (Mehrabi et al., 2011) et les bulles de filtre, aussi appelées chambres d'écho (Pariser, 2011). Le jeu vise à sensibiliser les élèves du lycée à ces questions, tout en leur permettant de comprendre, de manière ludique et concrète, les mécanismes qui les sous-tendent.

### 5. Liens vers les ressources et point de contact

- <https://mi.hepl.ch/projects/ia/ia.html>
- biljana.petreska-von-ritter-zahony@hepl.ch

### 6. Références

- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., Qiao, M. S. (2021). *Conceptualizing AI literacy: An exploratory review*. Computers and Education: Artificial Intelligence, vol. 2
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. (2011). *A Survey on Bias and Fairness in Machine Learning*. ACM Computing Surveys, vol. 54, 6, pp 1–35
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press, London