

Quelques enjeux éthiques et critiques des systèmes d'intelligence artificielle en enseignement supérieur (SIA-ES)

Simon COLLIN, Université du Québec à Montréal

avec la collaboration de

Alexandre LEPAGE, doctorant, Université de Montréal

Léo NEBEL, analyste des données, EvidenceB



chaire de recherche
du Canada sur l'équité numérique
en éducation



**OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA ET DU NUMÉRIQUE**

Plan

- Aperçu des systèmes d'intelligence artificielle en enseignement supérieur (SIA-ES)
- Quelques enjeux éthiques et critiques des SIA-ES

Aperçu des systèmes d'intelligence artificielle en enseignement supérieur (SIA- ES)

- SIA-ES (Chen et al., 2020; Zawacki-Richter et al., 2019, Zhang et Aslan, 2021)
 - en plein essor
 - Sur le plan scientifique, prédominance des recherches appliquées (recherches-développement; études d'impact)
 - développement et évaluation des effets des SIA-ES en termes d'efficacité

- Ex., ministère de l'Éducation du Québec

Intelligence numérique en éducation

Le ministère de l'Éducation investira plus de 10 millions de dollars afin d'amorcer le virage de l'intelligence numérique dans le réseau scolaire, et ce, dès la rentrée 2022.

L'application des principes d'intelligence numérique dans le milieu scolaire a pour but de mettre à profit les données du réseau afin de soutenir la réussite éducative et d'assurer une gestion plus transparente et efficace. Elle permettra également de mieux comprendre les déterminants de la réussite scolaire et favorisera, tant au Ministère que dans le réseau, la prise de décisions fondées sur des données. L'élève est donc au cœur de ce virage numérique, qui vise à valoriser les données et à utiliser les outils d'intelligence numérique au bénéfice de la réussite de tous les élèves et pour un réseau scolaire plus efficace.

Bénéfices

L'intelligence numérique bénéficiera à tous les acteurs du réseau de l'éducation, soit les élèves, les parents, les enseignantes et enseignants, les directions d'école, les

- Récemment, de plus en plus d'attention aux enjeux éthiques et critiques qui accompagnent les SIA-ES

« un nombre étonnamment faible d'auteurs, seulement deux sur 146 articles (1,4 %), se livrent à une réflexion critique sur les implications éthiques, les défis et les risques liés à l'application de l'IA en éducation » (Zawacki-Richter et al., 2019, p. 10; traduction libre)

- Deux principaux développements éthiques et critiques à propos des SIA-ES
 - développement de cadres éthiques de référence
 - *Consensus de Beijing sur l'intelligence artificielle et l'éducation* (UNESCO, 2019)
 - *Ethical Framework for AI in Education* (Institute for Ethical AI in Education, 2021)
 - *Lignes directrices éthiques sur l'utilisation de l'intelligence artificielle et des données dans l'enseignement et l'apprentissage à l'intention des éducateurs* (Commission européenne, 2022)
 - *Former à l'éthique de l'IA en enseignement supérieur : référentiel de compétence* (Bruneault, Sabourin Laflamme et Mondoux, 2022)

- études théoriques et empiriques des enjeux éthiques et critiques des SIA-ES

Home ▶ All Journals ▶ Learning, Media and Technology ▶ List of Issues ▶ Volume 45, Issue 3

Learning, Media and Technology

Enter keywords, authors, DOI, ORCID etc This Journal

Advanced search Citation search

Publish with us Submit an article About this journal Explore Browse all articles & issues Latest issue Subscribe Alerts & RSS feed + Purchase a subscription



Learning, Media and Technology, Volume 45, Issue 3 (2020)

< **Volume 45, 2020** Vol 44, 2019 Vol 43, 2018 Vol > See all volumes and issues

< Issue 4 **Issue 3** Issue 2 Issue 1 >

Download citations Download PDFs Download issue Browse by section (All) Display order (Default)

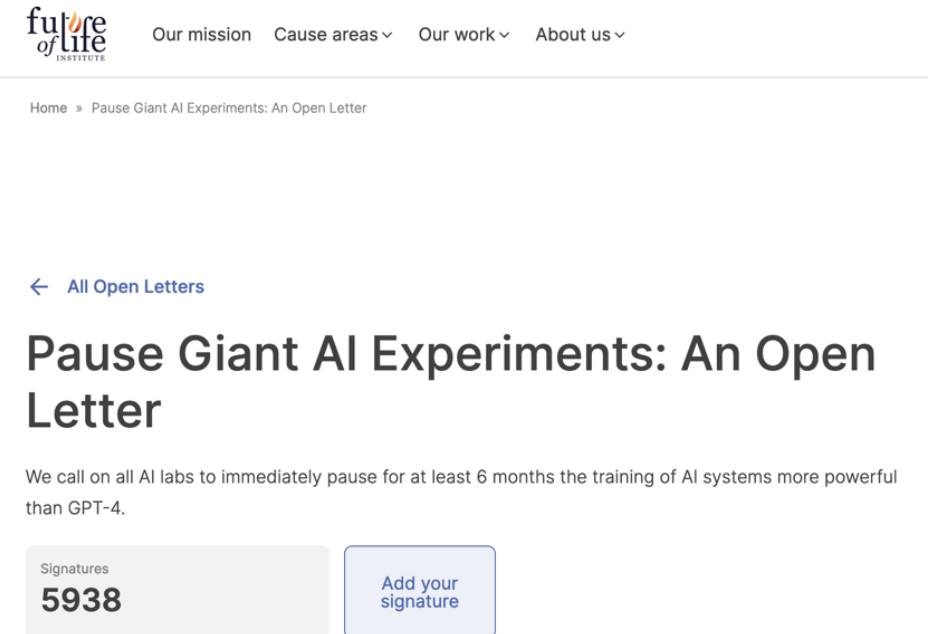
AI and Education: critical perspectives and alternative futures

Browse this journal

- > Latest articles
- > Current issue

- Réflexions éthiques et critiques
 - gagnent à être menées conjointement au développement et à l'évaluation des SIA-ES
 - pour contribuer à les orienter
 - finalité: assurer des effets aussi positifs que possible des SIA-ES et d'en maîtriser les risques

- Pétition initiée par *Future of Life Institute*
 - « Les systèmes d'intelligence artificielle puissants devraient être développés seulement lorsqu'on a l'assurance que leurs effets seront positifs et que leurs risques seront maîtrisables » (traduction libre)



The screenshot shows the website for the Future of Life Institute. At the top left is the logo, which consists of the words "future of life" in a lowercase, sans-serif font, with "INSTITUTE" in a smaller, uppercase font below it. To the right of the logo are four navigation links: "Our mission", "Cause areas" with a downward arrow, "Our work" with a downward arrow, and "About us" with a downward arrow. Below the navigation bar is a breadcrumb trail: "Home » Pause Giant AI Experiments: An Open Letter". A blue link with a left-pointing arrow says "All Open Letters". The main heading is "Pause Giant AI Experiments: An Open Letter" in a large, bold, black font. Below the heading is a paragraph of text: "We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4." At the bottom left, there is a grey box with the text "Signatures" and "5938" in a large, bold font. To the right of this box is a blue button with the text "Add your signature".

<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

Quelques enjeux des SIA-ES

- Revue systématique de la littérature scientifique sur les enjeux éthiques et critiques de l'IA en éducation
 - 58 documents analysés
 - codage par analyse de contenu
 - regroupements par proximité des codes

- Tensions
 - relations entre des aspects d'un SIA et de l'éducation qui peuvent potentiellement devenir problématiques si ces derniers ne sont pas arrimés
 - deviennent discordantes si ne sont pas anticipées et gérées en conséquence
 - porteuses d'enjeux éthiques et critiques potentiels
 - deviennent effectifs que lorsque la relation entre les aspects de l'IA et de l'éducation devient discordante



1. Tension « complexité des situations éducatives – standardisation technique »
 - risque pour un SIA-ES d'être mal adapté à la singularité de certains milieux scolaires ou profils d'actrices éducatives*
 - // « usager universel » (« *user as everybody* », Oudshoorn, 2004)
 - homogénéisation des représentations que les équipes de conception se font à propos des usagers et de leurs contextes
 - perte de vue de la diversité des usagers réels (Oudshoorn, 2016)

| Tension « complexité des situations éducatives – standardisation technique » | | | |
|---|-----------|----------|--|
| <u>Enjeux</u> | Documents | Segments | <u>Principaux articles</u> |
| Diminution de la prise en compte du contexte éducatif | 14 | 23 | Prinsloo & Slade (2017) Romero (2019) Hakimi et al. (2021) |
| Réduction factice de la complexité éducative par les indicateurs retenus | 25 | 40 | Perrota & Selwyn (2020) Corrin et al. (2019) Selwyn & Gašević (2020) Zeide (2019) |
| Réduction factice de la complexité éducative par les acteurs éducatifs lors de leurs interprétations des résultats | 6 | 9 | Holmes et al. (2021) Romero (2019) Perrota & Selwyn (2020) |
| Industrialisation de l'éducation | 24 | 54 | Hakimi et al. (2021) Davies et al. (2021) McStay (2019) |
| Manque d'adaptabilité des technologies au contexte | 7 | 11 | Holstein & Doroudi (2021) |

- Exemple
 - algorithme A-level (Office of Qualifications and Examinations Regulation (Ofqual)) au Royaume-Uni - 2020

The UK exam debacle reminds us that algorithms can't fix broken systems

The problem began when the exam regulator lost sight of the ultimate goal— and pushed for standardization above all else.

By Karen Hao

August 20, 2020

<https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/>

- intention: standardiser les résultats scolaires des élèves pour corriger une inflation ou une discontinuité éventuelles suite aux perturbations de la pandémie
- « Le problème a commencé lorsque les administrateurs en charge de l'examen ont perdu de vue le but ultime – et ont poussé pour une standardisation à tout prix » (Hao, 2020, traduction libre)
- conséquence: scores discriminants à l'égard des élèves des milieux scolaires défavorisés, amplifiant les inégalités scolaires existantes, notamment entre écoles privées et publiques

2. Tension « agentivité des actrices éducatives – automatisation technique »

- risque d'empiéter sur le jugement professionnel et les interventions des actrices éducatives
- depuis la révolution industrielle, automatisation croissante de la production puis de la consommation des biens et des services
 - choix d'usage d'une technologie de plus en plus conditionnés par les équipes de conception en amont de l'usage

| Tension « agentivité des actrices éducatives – <u>automatisation technique</u> » | | | | |
|--|-----------|----------|---|---|
| <u>Enjeux</u> | Documents | Segments | <u>Principaux articles</u> | |
| Cadrage accru du rôle de l'enseignant | 20 | 25 | Buckingham Shum & Luckin (2019) Holmes et al. (2021) | Polonetsky & Tene (2014) Hrastinski et al. (2019) |
| Remplacement de l'enseignant par l'IA | 9 | 9 | Staltman (2020) Holmes et al. (2021) | Lu et al. (2021) |
| Limitation des choix et des actions | 5 | 6 | Hakimi et al. (2021) | |
| Pédagogie et contenus imposés par l'IA | 23 | 42 | Schiff (2021) Smuha (2020) | Holmes et al. (2021) Buckingham Shum & Luckin (2019) |
| Absence de mécanismes d'appel | 14 | 17 | Smuha(2020) | |
| Présence de chambres d'échos | 4 | 6 | Regan & Jesse (2019) Remian (2019) | |
| Perception déterministe des prédictions de l'IA | 15 | 19 | Corrin et al. (2019) Roberts et al. (2017) | Romero (2019) Williamson et al. (2020) |
| Diminution de l'agentivité humaine | 32 | 67 | Holmes et al. (2021) Regan & Jesse (2019) | |

- Exemple:
 - systèmes de recommandation de contenus (tuteurs intelligents): les actrices éducatives peuvent-elles intervenir dans la planification du parcours d'apprentissage des apprenant.es?*

3. Tension « intelligibilité des actrices éducatives – opacité technique »

- risque pour les actrices éducatives d'être entravées dans leur évaluation de la pertinence éducative des SIA-ES
- technologies = boîtes noires (Akrich, 1987)
 - manque de transparence concernant des processus computationnels mis en œuvre
 - manque d'explicabilité des processus computationnels et des résultats obtenus (algorithmes non supervisés)

Tension « intelligibilité des actrices éducatives – opacité technique »

| <u>Enjeux</u> | Documents | Segments | <u>Principaux articles</u> |
|--|-----------|----------|---|
| Manque de transparence quant à la présence de l'IA | 14 | 20 | Nichols & Holmes (2018) <u>Remian (2019)</u> Southgate (2020) |
| Manque de transparence sur les données utilisées | 28 | 71 | Corrin et al. (2019) Lawson et al. (2016) |
| Manque de transparence quant à l'interprétation et l'usage des résultats | 14 | 24 | <u>Corrin et al. (2019)</u> Holmes et al. (2021) Roberts et al. (2017) <u>Zeide (2019)</u> Lawson et al. (2016) |
| Difficulté à expliquer les résultats des techniques d'IA | 22 | 50 | <u>Southgate (2020)</u> |
| Manque de confiance des utilisateurs | 22 | 54 | Qin et al. (2020) |
| Confiance aveugle envers les représentations informatiques du réel | 8 | 8 | Buckingham Shum & Knox (2017) <u>Luckin (2019)</u> |
| Manque de preuves d'efficacité pédagogique | 15 | 28 | Bulger (2016) <u>Corrin et al. (2019)</u> |
| Usage de l'IA sans utilité pédagogique | 7 | 11 | Holmes et al. (2021) Madaio et al. (2021) <u>McStay (2019)</u> <u>Smuha (2020)</u> |

- Exemple:
 - Système de gestion des comportements (Livingstone et al., 2016) :
 - Rempli à la fin du cours par les actrices éducatives sans prévenir les élèves concernés
 - Élèves « flaguées » le système de gestion des comportements et convoqué par la direction d'établissement sans qu'elles soient au courant

4. Tension « dignité des actrices éducatives – exploitation des données »

- risque de manipuler les actrices éducatives et de détourner leurs informations vers d'autres finalités que des finalités éducatives
- développements techniques actuels (p. ex., apprentissage machine) reposent sur la collecte, le traitement et l'interprétation d'un nombre massif de données
 - multiplication des controverses concernant une mauvaise utilisation, une utilisation déficiente ou une usurpation des données

Tension « Dignité des actrices éducatives– exploitation des données »

| <u>Enjeux</u> | <u>Documents</u> | <u>Segments</u> | <u>Principaux articles</u> |
|---|------------------|-----------------|--|
| Absence de demande de consentement | 8 | 12 | <u>Corrin et al. (2019)</u> <u>Madaio et al. (2021)</u> <u>Regan & Jesse (2019)</u> <u>Roberts et al. (2017)</u> |
| Impossibilité de retirer son consentement | 11 | 19 | <u>Berendth et al. (2020)</u> <u>Corrin et al. (2019)</u> <u>Lawson et al. (2016)</u> |
| Absence de mesures particulières pour le consentement des enfants | 9 | 12 | <u>Berendth et al. (2020)</u> <u>Regan & Jesse (2019)</u> <u>Smuha (2020)</u> |
| Consentement mal informé | 20 | 36 | <u>Corrin et al. (2019)</u> <u>Holmes et al. (2021)</u> <u>Jones et al. (2020)</u> <u>Southgate (2020)</u> |
| Consentement factice | 8 | 10 | <u>Berendth et al. (2020)</u> <u>Jones et al. (2020)</u> |
| Collecte de données sensibles ou trop nombreuses | 11 | 20 | <u>Rubel & Jones (2016)</u> |
| Non-respect de l'anonymat | 13 | 28 | <u>Regan & Jesse (2019)</u> <u>Corrin et al. (2019)</u> |
| Réidentification par données proxy | 17 | 23 | <u>Southgate (2020)</u> |
| Ambiguïté sur les obligations relatives à la conservation des données | 4 | 6 | <u>Kitto & Knight (2019)</u> |
| Données détournées de leur usage premier | 22 | 51 | <u>Peraya (2019)</u> |
| Surveillance | 19 | 41 | <u>Berendth et al. (2020)</u> <u>Lawson et al. (2016)</u> <u>Madaio et al. (2021)</u> <u>Regan & Jesse (2019)</u> <u>Remian (2019)</u> |

| Tension « <u>Dignité des actrices éducatives</u> – exploitation des <u>données</u> » (suite) | | | |
|--|-----------|----------|--|
| <u>Enjeux</u> | Documents | Segments | <u>Principaux articles</u> |
| Systèmes non robustes ou peu fiables | 12 | 26 | <u>Remian (2019)</u> <u>Smuha (2020)</u> <u>Corrin et al. (2019)</u> |
| Erreurs de classification | 8 | 9 | Scholes (2016) |
| Présence d'informations inexactes dans les systèmes d'apprentissage | 4 | 10 | Remian (2019) Sjödén (2020) |
| Failles de sécurité | 10 | 13 | Aiken & Epstein (2000) Holmes et al. (2021) <u>Corrin et al. (2019)</u> |
| Conséquences négatives sur l'apprentissage ou le développement | 3 | 4 | Aiken & Epstein (2000) |
| Entrave au développement des enfants | 6 | 11 | <u>Smuha (2020)</u> <u>Remian (2019)</u> |
| Modification induite des comportements des étudiants répondre aux attentes de systèmes d'IA | 15 | 19 | McStay (2019) |
| Génération d'anxiété chez les étudiants | 6 | 10 | Mayfield et al. (2019) |
| Ludification et personnalisation des apprentissages opposée au développement de compétences | 3 | 4 | Davies et al. (2021) |
| Perte de vue de la visée de formation à la citoyenneté | 3 | 4 | <u>Smuha (2020)</u> |
| Diminution des occasions de socialisation | 9 | 13 | Aiken & Epstein (2000) <u>Reiss (2021)</u> |

- Exemple
 - Krutka et al. (2021)

« Nous décrivons comment Google extrait des données personnelles des étudiants, contourne les lois censées les protéger, les cible pour faire du profit, obstrue ses intentions dans ses Termes de service, recommande du contenu dangereux et distord les connaissances des étudiants. » (p. 1, traduction libre)

- Exemple:
 - Human Rights Watch (2022; <https://www.hrw.org/StudentsNotProducts>)



An invisible swarm of tracking technologies surveils our kids throughout their day.

89%

of the EdTech products we looked at monitored or could monitor children, in most cases secretly and without the consent of children or their parents.

Some harvested data on who children are, where they are, and what they do in their virtual classroom. Others took information about who their family and friends are, and what kind of device their families could afford for them to use.

Many online learning platforms installed tracking technologies that followed children across the internet after they left their virtual classroom, outside of school hours and deep into their private lives.

Some invisibly tagged and fingerprinted children in ways that were impossible to get rid of—even if children, their parents, and teachers knew what was happening and wanted to do so—without throwing the device away in the trash.

5. Tension « justice scolaire – rationalité technique »

- risque de naturaliser des situations d'injustice subies par des groupes éducatifs minorisés
- technologies ne sont pas neutres: elles matérialisent et véhiculent des valeurs et des normes sociales
 - or lorsque les équipes de conception se focalisent uniquement sur le fonctionnement technique des SIA (p. ex., efficacité, performance), ont tendance à faire l'impasse sur les valeurs et les normes sociales en jeu
 - valeurs et normes sociales dominantes = étalon de référence par défaut des équipes de conception, aux dépens de valeurs et de normes sociales plus justes (Oudshoorn, 2004)

6. Tension « gouvernance scolaire – conception technique »

- risque d'ingérence d'acteurs, de logiques et de produits incompatibles avec la mission et la gestion de l'éducation, en tant que bien public
- autonomie apparente (bien qu'erronée) des domaines technique et politique

(Law, 1991):

Domaine technique

Domaine politique

Acteurs techniques

Acteurs politiques

Compagnies privées

Institutions publiques

Fonctionnement technocratique

Fonctionnement démocratique



ANGLES MORTS DE LA GOUVERNANCE DE L'IA

SOUS LA DIRECTION DE
Benjamin Prud'homme
Catherine Régis
and Golnoosh Farnadi

Merci

Simon COLLIN, Université du Québec à Montréal

collin.simon@uqam.ca



chaire de recherche
du Canada sur l'équité numérique
en éducation