



Utilisation de LLM pour *l'Adaptive Learning*

Chloé Conrad - chloe.conrad@univ-lyon1.fr

Maëva Somny - maeva.somny@univ-lyon1.fr



Plan

1. Rappels sur l'*Adaptive Learning*
2. Émergence des LLM
3. LLM & *Adaptive Learning*
4. QR actuelles (2025) : LLM & *Adaptive Learning*
5. Limites et enjeux éthiques
6. À vous de jouer



Plan

1. **Rappels sur l'*Adaptive Learning***
 - a. **Qu'est ce que l'*Adaptive Learning* ?**
 - b. **Quelques exemples**
2. Émergence des LLM
3. LLM & *Adaptive Learning*
4. QR actuelles (2025) : LLM & *Adaptive Learning*
5. Limites et enjeux éthiques
6. À vous de jouer

Rappels sur l'Adaptive Learning

Qu'est-ce que l'Adaptive Learning ?

Définitions (cf cours sur l'Adaptive Learning) :

L'Adaptive Learning est un **champ de recherche** actif en EIAH qui s'intéresse à :

1. L'adaptation des ressources pédagogiques **au contexte d'enseignement**
2. L'adaptation **à un apprenant ou un groupe d'apprenants** présentant une même caractéristique

Comment ?

→ Utilisation de **système d'IA** pour mieux identifier **les caractéristiques et besoins** des apprenants

Pourquoi ? [Sari et al., 2024]

- Augmenter l'engagement de l'apprenant
- Favoriser l'apprentissage

Sari, H.E., Tumanggor, B., Efron, D.: Improving Educational Outcomes Through Adaptive Learning Systems using AI. International Transactions on Artificial Intelligence 3(1), 21–31 (Nov 2024). <https://doi.org/10.33050/italic.v3i1.647>, <https://journal.pandawan.id/italic/article/view/647>

Rappels sur l'*Adaptive Learning*

Quelques exemples

- Moteur de recommandations de ressources pédagogiques
 - **ComPer** : Système de recommandation à base de règles qui recommande une liste ordonnée de ressource pédagogique à l'apprenant en fonction de ces compétences [Sablayrolles et al., 2022]
- Tuteur intelligent
 - **ITS avec Multi-Armed Bandit** : Tuteur intelligent basé sur une transformation du problème du bandit manchot adapté au contexte de l'éducation [Clement et al., 2013]
- *Feedbacks* aux apprenants
 - **Pyrates** : Génération de *feedbacks* en utilisant de l'apprentissage par renforcement [Kirouchenassamy et al., 2025]

- Clement, B., Roy, D., Oudeyer, P. Y., & Lopes, M. (2013). Multi-armed bandits for intelligent tutoring systems. arXiv preprint arXiv:1310.3174.
- Kirouchenassamy, B., Yessad, A., Jolivet, S., Branthôme, M., Lallé, S., & Luengo, V. (2025, July). Learning Feedback Policy from Historical Data: An Offline Approach Within Pyrates. In International Conference on Artificial Intelligence in Education (pp. 165-178). Cham: Springer Nature Switzerland.
- Sablayrolles, L., Lefevre, M., Guin, N., Broisin, J.: Design and Evaluation of a Competency-Based Recommendation Process. In: Crossley, S., Popescu, E. (eds.) Intelligent Tutoring Systems. pp. 148-160. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-09680-8_14



Plan

1. Rappels sur l'*Adaptive Learning*
2. **Émergence des LLM**
 - a. **Généralités sur les LLM**
 - b. **LLM en éducation**
3. *LLM & Adaptive Learning*
4. QR actuelles (2025) : *LLM & Adaptive Learning*
5. Limites et enjeux éthiques
6. À vous de jouer

Émergence des *LLM*

Généralités sur les LLM

Définition :

→ *Large Language Model*

2017 : Création du mécanisme d'attention

2020 : Sortie de *chatGPT* basé sur *GPT 3* et utilisation des *LLM* [Vaswani et al., 2020]

Depuis 2020:

- amélioration constante des performances des *LLM*
- diversification des modèles...
- explosion de l'utilisation des *LLM* ... Surtout propriétaires (GPT, Gemini, etc.)

Toutes les facettes de la société sont impactées... *l'éducation n'est pas épargnée*

Émergence des *LLM*

LLM en éducation

Une technologie au cœur des discussions... *

- AIED 2023 : ~0/82 articles (0%)
- AIED 2024 : ~20/76 articles (26%)
- AIED 2025 : 53/249 articles (21%)

... utilisée comme systèmes décisionnaires en EIAH **

- AIED 2023 : ~5/82 articles (6%)
- AIED 2024 : ~20/76 articles (26%)
- AIED 2025 : 86/249 articles (34%)

* articles évoquant les LLM

** articles décrivant l'élaboration d'un système basé sur un LLM

~ car review des articles en cours,
les résultats peuvent évoluer

Données issues d'une revue sur les articles publiés
entre 2023 à 2025 de la conférence AIED (rank A)



Plan

1. Rappels sur l'*Adaptive Learning*
2. Émergence des LLM
3. **LLM & Adaptive Learning**
4. QR actuelles (2025) : LLM & *Adaptive Learning*
5. Limites et enjeux éthiques
6. À vous de jouer

Émergence des LLM

LLM et Adaptive Learning

Pour quels types de tâches ?

- **Génération** : de contenus pédagogiques, de *feedback*, d'images, etc.
- **ChatBot** : Tuteurs intelligents, *Learning by Teaching*, simulation d'application de l'apprentissage, etc.
- **Evaluation** : de l'apprenant, de la difficulté d'une question, de la qualité des réponses de QCM, etc.
Et même des systèmes eux-mêmes (*feedback*, contenu pédagogique,) !

Quels modèles ? *

- Surtout des modèles propriétaires (~80% des modèles utilisés) : GPT (50 %), LLama (10%)
- mais aussi quelques modèles *open source* : Gemma, Qwen, Mistral



Plan

1. Rappels sur l'*Adaptive Learning*
2. Émergence des LLM
3. *LLM & Adaptive Learning*
4. **QR actuelles (2025) : LLM & Adaptive Learning**
5. Limites et enjeux éthiques
6. À vous de jouer

QR actuelles (2025) : *LLM & Adaptive Learning*

Génération : *SlideltRight: Using AI to Find Relevant Slides and Provide Feedback for Open-Ended Questions*

Chloe Qianhui Zhao et al., USA

Contribution : Génération de *feedbacks* sur la réponse d'un apprenant à une question ouverte.

Technique et Modèle : GPT-4o et RAG pour utiliser des informations tirées des slides du cours pertinentes par rapport à la question posée à l'apprenant

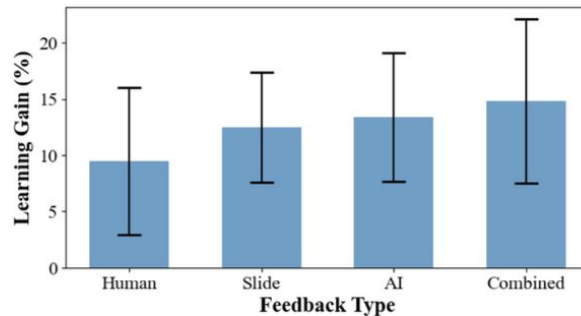
Évaluation :

- A/B testing :
 - 100 apprenants adultes
 - Cours en ligne + 5 questions ouvertes après le cours avec *feedbacks* immédiats à chaque question.
 - 4 types de *feedbacks* : Affichage de la slide pertinente du cours, *feedback* généré par LLM sans RAG, *feedback* généré par LLM avec RAG et *feedback* rédigé par un humain
 - Pré-test et post-test pour évaluer les connaissances des apprenants + questionnaires de satisfaction

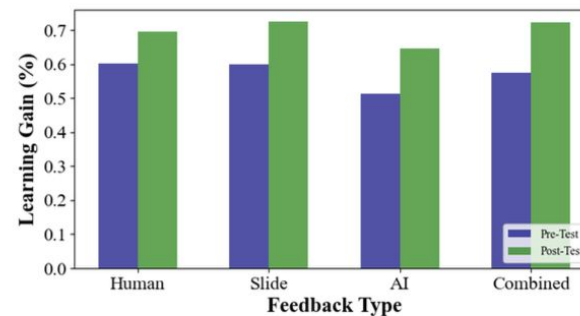
QR actuelles (2025) : LLM & Adaptive Learning

Génération : *SlideltRight: Using AI to Find Relevant Slides and Provide Feedback for Open-Ended Questions*

Chloe Qianhui Zhao et al., USA



(a) Average Learning Gain with 95% Confidence Interval



(b) Pre- v.s. Post-Test Scores

QR actuelles (2025) : LLM & Adaptive Learning

Génération : SlideltRight: Using AI to Find Relevant Slides and Provide Feedback for Open-Ended Questions

Chloe Qianhui Zhao et al., USA

Table 3. Proportion of ‘Agree’ or Higher in the Post-Survey Likert Responses

Question	Human Feedback (%Agree)	Slide (%Agree)	AI Feedback (%Agree)	Combined (%Agree)
Q1. I am satisfied with my overall learning experience.	77.27	86.96	75.00	77.27
Q2. I feel I gained sufficient knowledge and learning outcomes.	81.82	86.96	75.00	77.27
Q3. Feedback was easy to understand.	95.45	39.13	70.83	81.82
Q4. Feedback for learn-by-doing was helpful.	90.91	73.91	75.00	81.82
Q5. Feedback provided actionable suggestions.	68.18	52.17	87.50	59.09
Q6. Feedback encouraged reflection and critical thinking.	81.82	60.87	75.00	68.18
Q7. It was important to me to know whether the feedback was generated by a human or an AI.	36.36	30.43	25.00	22.73
Q8. Trust in feedback.	81.82	65.22	50.00	68.18
Q9. Feedback addressed issues in my responses.	81.82	56.52	79.17	86.36
Q10. Feedback was personalized.	72.73	56.52	79.17	68.18
Q11. Feedback motivated me to engage.	90.91	60.87	62.50	63.64

QR actuelles (2025) : *LLM & Adaptive Learning*

Génération : *StoryLab: Empowering Personalized Learning for Children through Teacher-Guided Multimodal Story Generation* Z. Li et al., USA

Contribution : Génération d'histoires personnalisées pour l'apprentissage de la langue en primaire.

Teacher-in-the-loop : L'enseignant a accès à un dashboard dans lequel il voit les histoires et images générées, le vocabulaire cible, et peut modifier l'histoire générée.

Technique et Modèle : *GPT-4o* (et *DALLE-3*), *role prompting*, et instructions de l'enseignant concernant le vocabulaire cible.

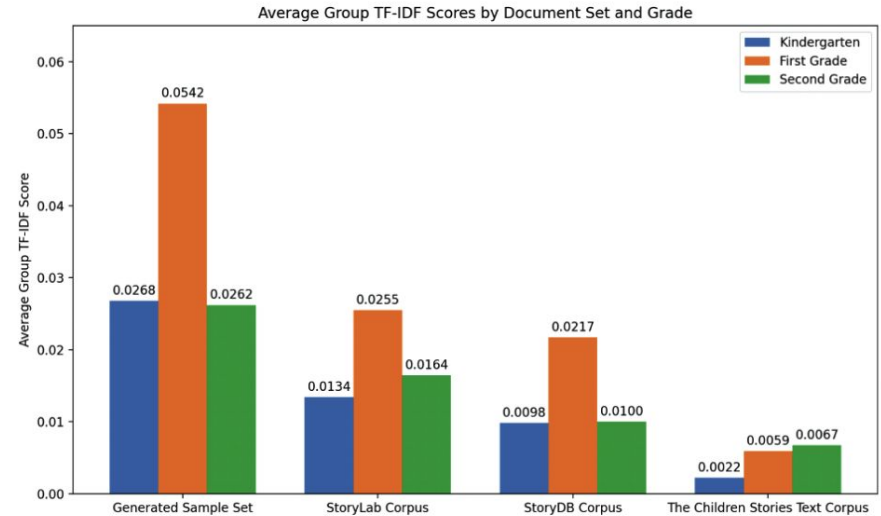
Évaluation :

- Expertise humaine (4 enseignants)
 - Utilisabilité (design)
 - Performance (perception des histoires générées)
- Performance (évaluation quantitative)
 - Métrique TD-IDF pour comparer la richesse des textes générés avec d'autres textes

QR actuelles (2025) : LLM & Adaptive Learning

Génération : *StoryLab: Empowering Personalized Learning for Children through Teacher-Guided Multimodal Story Generation* Z. Li et al., USA

TD-IDF : (term frequency-inverse document frequency) mesure statistique (méthode de pondération) permettant d'évaluer l'importance d'un terme contenu dans un document.



sample size :

10

30

500+

98

QR actuelles (2025) : *LLM & Adaptive Learning*

Chatbot : *Learning by Teaching: Enhancing Music Learning Through LLM-Based Teachable Agents*

L. Jin et al., Corée du Sud

Contribution : Chatbot pour apprendre la musique avec la methode *Learning-by-Teaching*

Technique et Modèle : *GPT* , *role prompting*

Évaluation :

- *A/B Testing* :
 - 28 étudiants chinois ayant déjà des connaissances en théorie musicale (auto-déclaration)
 - Pre-test, sans différence significative de résultats entre le groupe contrôle et le groupe expérimental
 - Phase d'apprentissage, 2 jours de cours en distanciel
 - Post-test, le groupe contrôle à accès aux cours et aux notes, le groupe expérimental à ChatMelody
 - Delayed test, 2 semaines après
- *Situational Motivation Scale* (SIMS) adaptée au contexte de l'apprentissage de la musique



QR actuelles (2025) : *LLM & Adaptive Learning*

Chatbot : *Learning by Teaching: Enhancing Music Learning Through LLM-Based Teachable Agents*

L. Jin et al., Corée du Sud

Bug into feature

“Although LLMs such as GPT-4 occasionally produce inaccurate responses, we do not treat these errors as system flaws. Instead, following, we conceptualize them as learning opportunities, viewing these inaccuracies as cognitive triggers that mirror the reasoning struggles of novice learners. This perspective encourages reflection and supports knowledge construction.”

QR actuelles (2025) : LLM & Adaptive Learning

Chatbot : Learning by Teaching: Enhancing Music Learning Through LLM-Based Teachable Agents

L. Jin et al., Corée du Sud

Impact sur les résultats des apprenants

Analyse de la covariance (ANCOVA) pour déterminer si les différences de résultats entre les groupes sont significatives ou non.

Group		N	Mean	S.D
Pre-test	EG	16	48,58	11,14
	CG	12	51,89	8,33
Post-test	EG	16	57,30	6,72
	CG	12	50,07	8,30
Delayed test	EG	16	55,25	6,88
	CG	12	41,75	9,55

QR actuelles (2025)

Chatbot : *Learning by Teaching: Enhancing Music Learning Through LLM-Based Teachable Agents*

L. Jin et al., Corée du Sud

Impact sur la motivation (SIMS) des apprenants

Analyse de la covariance (ANCOVA) pour déterminer si les différences de résultats entre les groupes sont significatives ou non.

Group		Mean	S.D	S.E
Intrinsic motivation	EG	25,56	3,67	0,92
	CG	21,83	4,04	1,17
Identified regulation	EG	25,38	3,24	0,81
	CG	21,25	4,54	1,31
External regulation	EG	16,13	4,91	1,23
	CG	15,00	4,67	1,35
Amotivation	EG	10,25	5,05	1,26
	CG	11,75	5,12	1,48



QR actuelles (2025) : *LLM & Adaptive Learning*

Chatbot : *Askademia: A Real-Time AI System for Automatic Responses to Student Questions*

Meenakshi Mittal et al., USA

Contribution : Chatbot pour permettre aux apprenants de poser des questions pendant un cours en direct (visioconférence)

Technique et Modèle : GPT-4o + RAG avec les documents de cours et la retranscription de la vidéo du cours découpé en séquences

Évaluation :

- Évaluation des réponses du *chatbot* avec un expert selon 4 dimensions : *Relevance*, *Readability*, *Positive Tone*, et *Factuality* (échelle de Likert à 3 niveaux)

QR actuelles (2025) : LLM & Adaptive Learning

Chatbot : Askademia: A Real-Time AI System for Automatic Responses to Student Questions

Meenakshi Mittal et al., USA

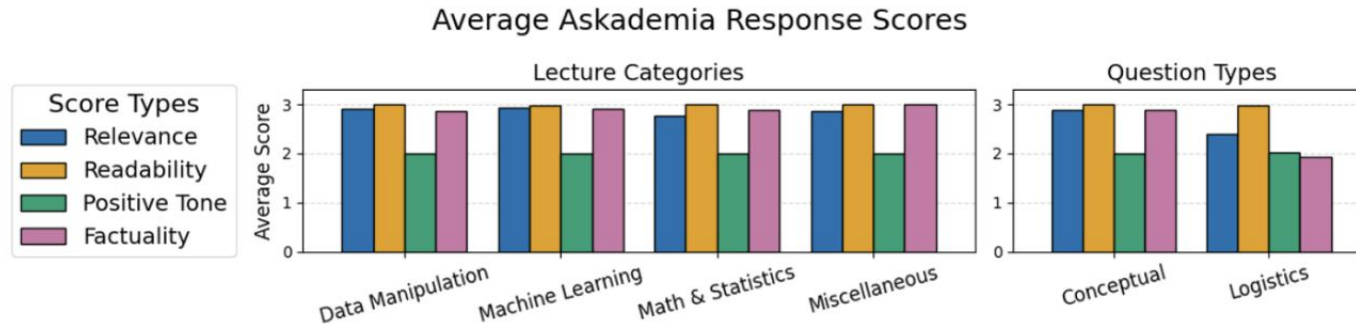


Fig. 3. Average human evaluation scores for Askademia responses, across various lecture categories and question types.



QR actuelles (2025) : *LLM & Adaptive Learning*

Evaluation : *Integrating Large Language Models and Machine Learning to Detect Struggle in Educational Games*

Xiner Liu et al., USA

Contribution : Détecter quand les apprenants rencontrent des difficultés dans un jeu sérieux à partir des logs d'interaction avec le système

Technique et Modèle : *GPT-4o*

Évaluation :

- Labellisation d'un jeu de données par deux experts et utilisation de métriques classique pour évaluer la performance système (*recall*, *précision*, *Kappa* et *F1 score*) + comparaison avec les performances obtenues par un modèle ML classique

QR actuelles (2025) : *LLM & Adaptive Learning*

Evaluation : *Integrating Large Language Models and Machine Learning to Detect Struggle in Educational Games*

Xiner Liu et al., USA

Table 1. Performances metrics for each prediction method

Model	Method	Kappa (κ)	Precision	Recall	F1
GPT	Knowledge Engineered	0.68	0.82	0.98	0.89
	Contextual Information	0.45	0.64	0.97	0.77
	Examples	0.70	0.84	0.97	0.90
XGBoost	Feature Engineering [12]	0.72	0.88	0.91	0.90



QR actuelles (2025) : *LLM & Adaptive Learning*

Evaluation : *Note-Driven RAG for Learner Performance Estimation via Controlling LLM Knowledge*

T. Minematsu and A. Shimada, Japon

Contribution : Estimation des performances d'étudiants en fonction de leur prise de notes (numériques)

Technique et Modèle : *GPT4o, RAG, Fine-Tuning, Reasoning based, Task Segmentation, Role Prompting*

Évaluation :

- Expérimentation :
 - 31 étudiants rémunérés
 - Cours sur l'IA (vidéos), 4x20 min + 1 QCM de 29 questions
- *Ablation study* : chaque modèle répond au QCM
 - *Knowledge restriction* (via persona ou via fine-tuning)

QR actuelles (2025) : *LLM & Adaptive Learning*

Evaluation : *Note-Driven RAG for
Learner Performance Estimation
via Controlling LLM Knowledge*
T. Minematsu and A. Shimada, Japon

Prise de notes Accès au cours LLM qui extrait de la connaissance des notes

		Note	Script	KLLM	P ↑	R ↑	FPR ↓	FNR ↓
Persona	1				0.702	0.868	0.849	0.132
	2	✓			0.718	0.845	0.765	0.155
	3	✓	✓		0.711	0.836	0.783	0.164
	4	✓		✓	0.719	0.850	0.765	0.150
	5	✓	✓	✓	0.721	0.850	0.757	0.150
Finetuning	1				0.000	0.000	0.000	1.000
	2	✓			0.771	0.118	0.081	0.882
	3	✓	✓		0.743	0.166	0.132	0.834
	4	✓		✓	0.792	0.402	0.243	0.598
	5	✓	✓	✓	0.752	0.381	0.290	0.619



QR actuelles (2025) : *LLM & Adaptive Learning*

Autres (non Adaptive Learning) : *From Text to Visuals: Using Large Language Models to Generate Math Diagrams with Vector Graphics*

Jaewook Lee et al., USA

Contribution : Génération de figure de mathématique en SVG (vectoriel) à partir d'un texte

Technique et Modèle : *GPT-4o*

Évaluation :

- Utilisation d'un *LLM* pour évaluer sémantiquement et syntaxiquement les figures générées. Comparaison des résultats obtenus avec ceux de figures de références et de figures générées par un modèle de génération d'image classique (pixel)

QR actuelles (2025) : *LLM & Adaptive Learning*

Autres (non Adaptive Learning) : *From Text to Visuals: Using Large Language Models to Generate Math Diagrams with Vector Graphics*

Jaewook Lee et al., USA

Topic	Ground Truth		Prediction	
	Sem.	Syn.	Sem.	Syn.
Divide by 2	1.00	1.00	1.00	0.88
Multiple by 2 (or 4)	1.00	0.96	0.92	0.84
Multiply by 1 (or 0)	1.00	1.00	0.96	0.92
Comparing Fractions	0.72	0.48	0.56	0.56
Area Formulation Intuition	0.52	0.52	0.56	0.60

QR actuelles (2025) : *LLM & Adaptive Learning*

Autres (non Adaptive Learning) : *Scaling Curriculum Mapping in Higher Education: Evaluating Generative AI's Role in Curriculum Analytics*

V. Jayalath et al., Australie

Contribution : Utilisation de *LLM* pour déterminer (en %) quelle compétence, connaissance et savoir-faire une évaluation évalue

Technique et Modèle : *GPT4, zero-shot, role-prompting*

Évaluation :

- Analyse de 19 cours (descriptif, objectifs, détails des évaluations, ...)
- Labellisation par 2 experts des évaluations (avec accord inter-annotateur)
- Labellisation par *BERT* et *LLM*, comparaison avec experts humains

QR actuelles (2025) : LLM & Adaptive Learning

Autres (non Adaptive Learning) : Scaling Curriculum Mapping in Higher Education: Evaluating Generative AI's Role in Curriculum Analytics V. Jayalath et al., Australie

Skill	Krippendorff's alpha and Interpretation			
	GenAI model		BERT-based NLP model	
Problem-Solving	0.59	Moderate Agreement	0.24	Fair Agreement
Teamwork	0.63	Substantial Agreement	−0.05	Systematic Disagreement
Written Comm	0.37	Fair Agreement	0.04	Slight Agreement
Self-Management	0.50	Moderate Agreement	−0.31	Systematic Disagreement
Ethical Awareness	0.47	Moderate Agreement	−0.12	Systematic Disagreement
Int Perspective	0.26	Fair Agreement	0.10	Slight Agreement
Oral Communication	0.86	Near-Perfect Agreement	−0.07	Systematic Disagreement

Skill	Y1	Y2	Y3
Problem-Solving	0.57	0.44	0.72
Teamwork	0.84	−0.13	0.99
Written Comm	0.41	−0.18	0.74
Self-Management	0.41	−0.11	0.72
Ethical Awareness	0.17	0.80	0.23
Int Perspective	0.38	0.50	−0.08
Oral Communication	0.80	0.77	0.98



Plan

1. Rappels sur l'*Adaptive Learning*
2. Émergence des LLM
3. *LLM & Adaptive Learning*
4. QR actuelles (2025) : *LLM & Adaptive Learning*
5. **Limites et enjeux éthiques**
6. À vous de jouer



Limites et enjeux éthique

Quelles limites ?

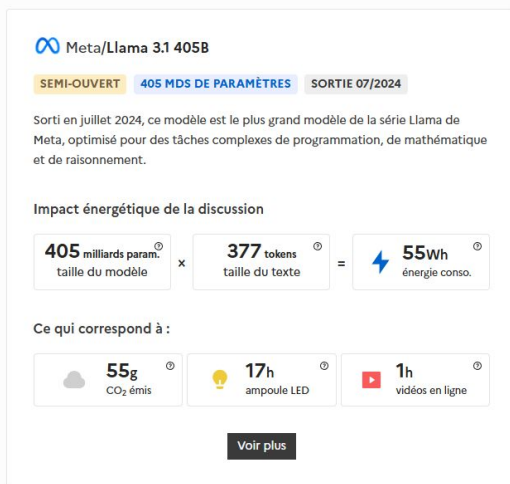
- Évaluation : Exactitude et performance, utilité et utilisabilité
- Balance bénéfices/risques

Quels enjeux éthiques (risque) ?

- Environnementaux
- Exploitation des données : RGPD & AI Act
- Biais (cf cours sur l'équité algorithmique)
- Inégalités numériques

Limites et enjeux éthique

Focus énergie, quelques chiffres...



Requêtes



Pré-entraînement

Estimation du pré-entraînement de GPT3

1 287 000 kWh [2]

Consommation annuelle moyenne électrique d'une personne Française ~ 2 223 kWh [3]

Pré-entraînement de GPT3 ~ 579 habitants

Data Center

En 2022, IA + Cryptomonnaie = 460TWh

Estimation x2 d'ici à 2026



Plan

1. Rappels sur l'*Adaptive Learning*
2. Émergence des LLM
3. *LLM & Adaptive Learning*
4. QR actuelles (2025) : *LLM & Adaptive Learning*
5. Limites et enjeux éthiques
6. **À vous de jouer**



À vous de jouer !

Concevoir un système d'*Adaptive Learning* avec *LLM* qui prend en compte au moins **un enjeu éthique**, et son évaluation.

Constitution de 3 groupes :

- Chatbot
- Génération (*feedback*, contenu pédagogiques)
- Évaluation (classification et *scoring*)

Questionnements éthiques :

- Énergie & Environnement
- Données : au-delà du RGPD et *AI act*
- *Fairness*, un *LLM* non biaisé
- Inégalités numériques

chloe.conrad@univ-lyon1.fr & maeva.somny@univ-lyon1.fr