

# **Proposition de Projet de Recherche Doctoral :**

## **Génération dynamique de contenus pédagogiquement fondés : une architecture hybride articulant IA générative, modélisation de l'apprenant et principes de la psychologie des apprentissages**

**Encadrement :** Amel Yessad (SU, LIP6/MOCAH), François Bouchet (SU, LIP6/MOCAH), Emmanuel Sander (U. Genève, FPSE/IDEA)

### **1 Contexte et problématique**

#### **1.1 Le défi de la création de contenus pédagogiques**

Les systèmes tutoriels intelligents (ITS) constituent depuis plus de trois décennies un champ de recherche majeur à l'intersection de l'intelligence artificielle et des sciences de l'éducation. Le déploiement à grande échelle des ITS reste entravé par un verrou technique bien documenté : **le coût de création des contenus pédagogiques** (Cai et al., 2019) dont la conception nécessite une expertise disciplinaire et didactique considérable.

#### **1.2 Les LLM : une solution potentielle avec un risque réel**

L'émergence des grands modèles de langage (LLM) semble pouvoir offrir une solution à ce problème. Les travaux récents démontrent leur capacité à générer du feedback (Stamper et al., 2024), à produire des contenus (Doughty et al., 2024) et à s'insérer dans des architectures tutorielles (Liu et al., 2024). Les évolutions technologiques récentes (Google Opal, Claude Skills) permettent même de générer dynamiquement des interfaces interactives complètes.

Cependant, **la facilité de génération masque un problème majeure en lien avec la validité et la pertinence pédagogique** des contenus générés. Plusieurs décennies de recherche en psychologie cognitive et en didactique montrent que le choix des contextes, des exemples et de leur séquençement n'est pas anodin. Au contraire, il détermine en grande partie la réussite ou l'échec de l'apprentissage.

#### **1.3 L'apport de la psychologie cognitive : la congruence sémantique**

Les travaux d'Emmanuel Sander et de ses collaborateurs sur le rôle de l'analogie dans la pensée (Hofstadter & Sander, 2013) et sur la **congruence sémantique** dans la résolution de problèmes arithmétiques (Gros, Sander & Thibaut, 2019 ; Gros, Thibaut & Sander, 2020) offrent un cadre théorique essentiel. Le modèle SECO (Semantic Congruence) formalise comment l'interaction entre la sémantique du monde évoqué par l'énoncé et la sémantique mathématique de l'opération influence profondément les stratégies de résolution et les performances des apprenants. Un générateur d'exercices ignorant ces principes produirait des contenus dont la qualité pédagogique serait imprévisible.

#### **1.4 Positionnement : vers une génération pédagogiquement fondée**

Ce projet propose de dépasser l'approche naïve « LLM comme générateur universel » pour développer une **architecture hybride où la génération est contrainte et informée par des principes issus de la psychologie des apprentissages**. L'enjeu n'est pas seulement technique

(« le LLM peut-il générer ? ») mais scientifique (« selon quels principes générer pour que les contenus d'un ITS soient efficaces ? ») et empirique (« quels gains mesurables sur l'apprentissage ? »).

Cette approche nécessite une **collaboration interdisciplinaire** entre : (1) l'intelligence artificielle en éducation, pour les architectures de systèmes adaptatifs et la modélisation multidimensionnelle de l'apprenant (Montserrat, Yessad, Bouchet et al., 2017) ; (2) la psychologie cognitive, pour les principes de conception des exercices et l'évaluation de leur impact sur l'apprentissage (Gros, Thibaut & Sander, 2020 ; Sander, 2025).

## 2 Objectifs scientifiques

L'objectif principal est de concevoir, implémenter et évaluer une **architecture ITS hybride** où un **LLM agentique** génère des contenus interactifs de manière dynamique, en s'assurant de la traçabilité de ceux-ci selon un standard interopérable (xAPI). Il sera alimenté par :

- **Le contenu disciplinaire**, via RAG sur des ressources du domaine.
- **Les contraintes didactiques**, type de congruence sémantique visée, choix des contextes analogiques, progression dans l'abstraction, etc. formalisées à partir du modèle SECO et des travaux sur l'analogie.
- **Le modèle de l'apprenant**, incluant niveau de maîtrise, erreurs récurrentes, avec une modélisation de type BKT (Bayesian Knowledge Tracing) ou DKT (Deep KT).

L'architecture hybride intégrera : (1) un **module de génération** (LLM + RAG disciplinaire + contraintes didactiques) ; (2) un **module de modélisation de l'apprenant** (BKT/DKT) estimant la maîtrise et orientant la sélection des contenus ; (3) un **module de traçabilité** convertissant les interactions en statements xAPI. L'orchestration sera assurée par un contrôleur qui sélectionne les contraintes didactiques appropriées en fonction de l'état de l'apprenant.

## 3 Encadrement et adéquation institutionnelle

Ce projet est **co-direction interdisciplinaire** associant 2 partenaires de 4EU+ :

- **Équipe MOCAH (LIP6, Sorbonne Université)**, avec une expertise en EIAH, IA en éducation, modélisation de l'apprenant, learning analytics, systèmes adaptatifs. Elle dispose d'infrastructures pour la collecte de traces (LRS LATTE financé par le CNRS).
- **Équipe IDEA (FPSE, Université de Genève)**, dont l'expertise en psychologie cognitive des apprentissages, raisonnement analogique et congruence sémantique offre un cadre théorique solide et permet l'évaluation de la pertinence didactique.

Il combine donc des approches d'IA (généraliste, deep) au coeur des thématiques de SCAI, et est solidement ancré dans un cadre SHS éprouvé.

## 4 Références

Disponibles ici : <https://nuage.lip6.fr/index.php/s/YtwzdaYPYkkogYg>