

Avis de Soutenance

Monsieur Sergio RAMÍREZ LUELMO

Sciences de l'éducation et de la formation

Soutiendra publiquement ses travaux de thèse intitulés

Détection automatique du flow (l'expérience optimale d'apprentissage) dans un MOOC via les techniques d'apprentissage automatique

dirigés par Monsieur Jean HEUTTE et Madame Nour EL MAWAS

Soutenance prévue le **jeudi 21 septembre 2023** à 9h00

Lieu : Campus Cité Scientifique, 59655 Villeneuve d'Ascq (2e étage du bâtiment B6).

Salle : B6.206

Composition du jury proposé

M. Jean HEUTTE	Université de Lille	Directeur de thèse
Mme Margarida ROMERO	Université Côte d'Azur	Rapporteuse
Mme ANNELIES RAES	KU Leuven	Examinatrice
M. Bruno POELLHUBER	Université de Montréal	Examinateur
M. Fabien FENOUILLET	Université Paris-Nanterre	Rapporteur
M. Mutlu CKUROVA	University College London	Examinateur
Mme Nour EL MAWAS	Université de Lille	Co-directrice de thèse

Mots-clés : flow, analyse de l'apprentissage, apprentissage automatique, expérience optimale d'apprentissage,

Résumé :

Flow « [...] un état d'épanouissement lié à une profonde implication et au sentiment d'absorption que les personnes ressentent lorsqu'elles sont confrontées à des tâches dont les exigences sont élevées et qu'elles perçoivent que leurs compétences leur permettent de relever ces défis ». L'état de flow est positivement corrélé avec des métriques favorables à l'apprentissage. Ainsi, Csikszentmihályi (2005) affirme que le flow favorise l'apprentissage et le développement personnel parce que les expériences de concentration profonde et totale sont intrinsèquement gratifiantes. Cependant, la recherche associant flow dans les MOOC ne fait que débiter alors que la détection du flow reste particulièrement complexe, car tout artefact tentant de le détecter ou de le mesurer contribue inévitablement à le perturber. L'importance de l'état de flow (en tant qu'état psychologique humain) dans le processus d'apprentissage, en ligne et à distance, nous pousse à proposer un modèle d'apprentissage automatique de détection de flow qui permet une détection de flow fiable, automatique et transparente dans un contexte de MOOC. Nous utilisons des techniques d'apprentissage automatique pour donner du sens aux données multidimensionnelles sans avoir recours à un expert humain en permanence. Ce projet de recherche se différencie des travaux précédents par l'utilisation d'un modèle théorique de flow et de son instrument de mesure, conçus exprès pour détecter flow directement sans passer par des concepts intermédiaires, ainsi que par l'exploitation de deux ensembles de données d'entrée : l'état de flow auto-rapporté des participants ($n \sim 9\,500$), et leurs données de connexion au MOOC ($\sim 80\,000$ Go), pendant deux ans. Aucune donnée de performance n'est collectée ni utilisée. Après des filtrages et des prétraitements rigoureux, nous couplons les états de flow auto-rapportés des utilisateurs ($n \sim 1\,500$) à leurs données de connexion agrégées (23 variables) pour obtenir deux modèles entraînés d'apprentissage automatique. A) Un modèle dit Proof-of-Concept qui corrobore les choix et le bon couplage des méthodes, des instruments de mesure du flow et des techniques d'apprentissage automatique ($F1 = 0,851$, $AUC\,ROC = 0,85$, $Exactitude = 0,797$, $Précision = 0,821$, $Rappel = 0,882$). B) Un modèle dit Prototype qui permet de détecter flow dans un contexte de MOOC de manière abordable ($F1 = 0,689$, $AUC\,PR = 0,87$, $AUC\,ROC = 0,68$, $Exactitude = 0,605$, $Précision = 0,854$, $Rappel = 0,578$), rapide ($>ms/participant$), d'un impact environnemental négligeable ($\sim 0,0000237222222$ g de CO_2eq par exécution), automatique (pas d'interventions supplémentaires une fois installé), et transparente (sans intervention des participants du MOOC). Ces deux modèles identifient mieux le flow que l'absence de flow : Proof-of-Concept : 58% vs. 22% du total, Prototype : $\sim 44\%$ contre $\sim 17\%$ du total. Or, le Prototype présente une proportion plus élevée de Faux Négatifs ($\sim 32\%$ du total) que des Vrais Négatifs ($\sim 17\%$ du total) lorsqu'il est confronté à des données non traitées et jamais vues (le prétraitement des données améliore les métriques mais reconstruit les données d'entrée). Cet écart peut être compris comme un modèle prudent préférant une classification pas-de-flow en cas de doute plutôt qu'une classification flow, ce qui n'est pas nécessairement un comportement indésirable du modèle. D'ailleurs, nos deux modèles souffrent d'un manque de granularité pour la détection du flow, un obstacle insurmontable intrinsèquement lié à la granularité (2 moments) des données d'entraînement. Les perspectives de ce projet de recherche comprennent la mise en œuvre et la commercialisation du modèle Prototype dans un MOOC pour 1) aboutir sur un tableau de bord du flow ; 2) personnaliser le contenu, les activités et le parcours d'apprentissage du MOOC ; et éventuellement 3) évaluer l'incidence de la détection du flow dans la personnalisation du MOOC afin de réduire le taux d'abandon du MOOC.