

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
ECOLE SUPERIEURE EN INFORMATIQUE 08 MAI 1945 SIDI BEL ABBES



Thesis

In partial fulfillment of the requirements for the degree of **State Engineer**

Field: **Computer Science**

Speciality: **Artificial Intelligence and Data Science (IASD)**

Topic

Generation of Texts Simulating Psychological Distress Using Generative Language Models

Presented by :

M^r BAIDAR Samir

Defended on: 07/07/2026

Before the jury composed of:

President:	Dr. SAIDI Imène		ESI SBA
Examiner:	Dr. MEZRAR Samiha		ESI SBA
Supervisor:	Dr. MAHAMMED Nadir		ESI SBA
Co-supervisor:	Dr. ACED Mohamed Reda		ESI SBA

Academic Year 2025/2026

Abstract

Mental health prevention and support professionals must routinely recognise and respond to discourse expressing psychological distress, yet training resources built on authentic crisis narratives raise serious privacy concerns and are difficult to produce at scale. This thesis investigates whether open-source, offline-deployable language models can be fine-tuned to generate synthetic texts that realistically simulate psychological distress at controlled severity levels, offering a privacy-preserving alternative for professional training, detection-system evaluation, and awareness-raising.

To this end, a binary-labelled social media corpus of suicidality-related posts is reformulated into a four-tier severity resource through a dedicated annotation pipeline that combines semantic embedding similarity, clinical keyword scoring, and post length into a single composite score, calibrated against an independently annotated clinical reference corpus and refined with targeted overrides for unambiguous crisis language. Five open-source decoder-only models, spanning three model families and a range of parameter scales, are then fine-tuned under a shared low-rank adaptation regime to generate text conditioned on discrete severity tokens aligned with PHQ-9 criteria, and subsequently quantized and packaged for fully offline deployment.

Comparative evaluation across the five fine-tuned models reveals that base model architecture and tokenizer design influence generation quality at least as strongly as raw parameter count, with the most efficient model of the set outperforming a substantially larger counterpart. Post-quantization analysis further shows that offline packaging preserves the fine-tuned models' behaviour closely enough that deployment introduces no meaningful loss in generation quality or control integrity, supporting the practical viability of a fully local pipeline. The five models are deployed through Spectrum, a browser-based generation console that exposes severity-conditioned inference with full prompt transparency, audit logging, and batch export in a self-contained local environment.

Taken together, these results establish a replicable methodology for severity-controlled distress text generation using exclusively open-source tools, and contribute the annotation pipeline, trained model artifacts, and associated ethical safeguards to the mental health NLP research community. Quantitative validation of severity control through an independent classifier-judge, together with an expert clinical review of generation quality, is planned as a subsequent stage of this work.

Key words: Psychological distress simulation, Severity-conditioned text generation, Large language models, LoRA fine-tuning, Mental health NLP, GGUF quantization.

Résumé

Les professionnels de la prévention et de l'accompagnement en santé mentale doivent régulièrement reconnaître et répondre à des discours de détresse psychologique, mais les formations basées sur des récits de crise authentiques posent de sérieux problèmes de confidentialité et sont difficiles à produire à grande échelle. Ce mémoire examine si des modèles de langage opensource, déployables hors ligne, peuvent être affinés pour générer des textes synthétiques simulant de manière réaliste la détresse psychologique à des niveaux de sévérité contrôlés, offrant une alternative respectueuse de la confidentialité pour la formation, l'évaluation de systèmes de détection et la sensibilisation.

À cette fin, un corpus de publications issues des réseaux sociaux, initialement annoté de manière binaire, est reformulé en une ressource à quatre niveaux de sévérité grâce à un pipeline d'annotation combinant similarité d'embeddings sémantiques, score de mots-clés cliniques et longueur des textes en un score composite, calibré sur un corpus clinique de référence et affiné par des seuils ciblant le langage de crise non ambigu. Cinq modèles décodeurs opensource issus de trois familles et de tailles variées sont ensuite affinés via un même protocole LoRA pour générer du texte conditionné par des jetons de sévérité alignés sur les critères PHQ9, puis quantifiés et packagés pour un déploiement hors ligne.

L'évaluation comparative révèle que l'architecture et la tokenisation influencent la qualité de génération autant que le nombre de paramètres, le modèle le plus efficace surpassant un homologue bien plus volumineux. L'analyse postquantification montre que le packaging hors ligne préserve le comportement sans perte significative de qualité de génération ni d'intégrité du contrôle de sévérité, confirmant la viabilité d'un pipeline entièrement local. Les cinq modèles sont déployés à travers Spectrum, une console de génération en navigateur qui expose l'inférence conditionnée par la sévérité avec transparence complète des prompts, journalisation d'audit et export par lots, dans un environnement local autonome.

Ces résultats établissent une méthodologie reproductible pour la génération de textes de détresse contrôlés en sévérité, s'appuyant exclusivement sur des outils opensource, et apportent à la communauté TALsanté mentale le pipeline d'annotation, les modèles entraînés et les garde-fous éthiques associés. Une validation quantitative par un classifieur indépendant et une revue clinique experte sont prévues comme étape ultérieure.

Mots clés : Simulation de détresse psychologique, Génération de texte contrôlée en sévérité, Grands modèles de langage, Affinage LoRA, TAL en santé mentale, Quantification GGUF.

ملخص

يواجه المهنيون العاملون في مجال الوقاية والدعم النفسي بشكل منتظم حالات تعبر عن ضيق نفسي، إلا أن إعداد موارد تدريبية تعتمد على روايات أزمات حقيقية يطرح مشكلات جدية تتعلق بالخصوصية، فضلاً عن صعوبة إنتاجها على نطاق واسع. تتناول هذه الرسالة إمكانية ضبط نماذج لغوية مفتوحة المصدر، قابلة للنشر دون اتصال بالإنترنت، لتوليد نصوص اصطناعية تحاكي بشكل واقعي الضيق النفسي ضمن مستويات شدة محكمة، بما يوفر بديلاً يحافظ على الخصوصية لأغراض تدريب المختصين، وتقييم أنظمة الكشف الآلي، والتوعية العامة.

ولتحقيق هذا الهدف، أُعيدت صياغة مجموعة بيانات من منشورات شبكات التواصل الاجتماعي، كانت مصنفة في الأصل تصنيفاً ثنائياً، لتتحول إلى مورد موسوم بأربعة مستويات من الشدة، عبر مسار توصيف يجمع بين تشابه التمثيلات الدلالية، ونتيجة الكلمات المفتاحية ذات الدلالة الإكلينيكية، وطول النص، في درجة تركيبية واحدة، تمت معيارتها على مجموعة بيانات إكلينيكية مرجعية، ونُقحت باعتماد عتبات موجهة للغة الأزمات الصريحة. ثم خضعت خمسة نماذج لغوية من نوع المُرمِّز فقط (decoder-only)، تنتمي إلى ثلاث عائلات مختلفة وتغطي أجاماً متنوعة، للضبط الدقيق وفق بروتوكول LoRA نفسه لتوليد نصوص مشروطة برموز شدة متوافقة مع معايير PHQ-9 قبل أن تُكمَّم وتُعبَأ للنشر دون اتصال.

ويُظهر التقييم المقارن أن بنية النموذج الأساسي وتصميم مُرمِّز الكلمات (tokenizer) يؤثران في جودة التوليد بقدر لا يقل عن العدد الخام للعوامل، حيث تجاوز أكثر النماذج كفاءةً نموذجاً آخر أكبر منه حجماً بكثير. كما يُبين تحليل ما بعد التكميم أن التعبئة دون اتصال تحافظ على سلوك النماذج المضبوطة دون تراجع يُذكر في جودة التوليد أو سلامة التحكم في الشدة، مؤكداً الجدوى العملية لمسار عمل محلي بالكامل. وقد نُشرت النماذج الخمسة عبر منصة Spectrum، وهي واجهة توليد في المتصفح تُوفِّر الاستدلال المشروط بالشدة مع شفافية كاملة في المطالبات، وتسجيل تدقيقي، وتصدير الدفعات، في بيئة محلية مستقلة. وإجمالاً، تُؤسّس هذه النتائج منهجية قابلة للتكرار لتوليد نصوص الضيق النفسي محكمة الشدة، باستخدام أدوات مفتوحة المصدر حصراً، وتُقدّم لمجتمع معالجة اللغة الطبيعية في الصحة النفسية مسار التوصيف، والنماذج المدربة، والضوابط الأخلاقية المرافقة. ويُخطّط للتحقق الكمي من دقة التحكم في الشدة بواسطة مُصنِّف مستقل ومراجعة سريرية خبيرة كخطوة لاحقة.

الكلمات المفتاحية: محاكاة الضيق النفسي، توليد النصوص محكوم الشدة، النماذج اللغوية الكبيرة، الضبط الدقيق بطريقة LoRA، معالجة اللغة الطبيعية في الصحة النفسية، التكميم بصيغة GGUF.