

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



Master Thesis

In partial fulfillment of the requirements for the degree of **Master**

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. ACED Mohamed Reda		ESI SBA
Co-Supervisor:	Dr. MAHAMMED Nadir		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 examines whether open-source, offline-deployable language models constitute a viable foundation for generating 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, the theoretical and methodological foundations underlying such a system are established. This covers the Transformer architecture and the pre-training paradigm of large language models, the main families of parameter-efficient fine-tuning techniques (notably LoRA), and the principal approaches to controlled text generation, including control-code conditioning, post-hoc steering, discriminator-guided generation, and prefix-tuning. The intersection of natural language processing with mental health research is then reviewed, covering text classification for depression and suicide-risk assessment, domain-specific text embeddings, psychometric instruments such as the PHQ-9 and BDI-II, and the ethical considerations inherent to this line of research.

A structured literature review is conducted across eight thematic axes: fundamental methods for controlled text generation, existing systems for simulating psychological distress with LLMs, specialized language models for mental health, adaptation strategies for local deployment, offline deployment and confidentiality constraints, reference datasets and benchmarks, evaluation methodologies, and ethical considerations. This review situates the target problem, severity-controlled distress text generation using exclusively open-source, offline-deployable tools, against the current state of the art, and identifies the gaps that motivate a concrete implementation of such a system.

Taken together, this work establishes the conceptual groundwork and a critical synthesis of prior art needed to inform the design of a severity-conditioned distress text generation pipeline, while surfacing the methodological choices, open questions, and ethical safeguards that any such implementation must address.

Key words: Psychological distress simulation, Controlled text generation, Large language models, Parameter-efficient fine-tuning, Mental health NLP, Literature review.

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 open-source, déployables hors ligne, constituent un socle viable 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, les fondements théoriques et méthodologiques sous-tendant un tel système sont établis. Cela couvre l'architecture Transformer et le paradigme de pré-entraînement des grands modèles de langage, les principales familles de techniques d'affinage paramétriquement efficace (notamment LoRA), ainsi que les principales approches de génération de texte contrôlée, incluant le conditionnement par codes de contrôle, le pilotage post-hoc, la génération guidée par discriminateur, et l'affinage de préfixes. L'intersection du traitement automatique du langage naturel et de la recherche en santé mentale est ensuite passée en revue, couvrant la classification de texte pour la détection de la dépression et l'évaluation du risque suicidaire, les représentations vectorielles de texte spécifiques au domaine, les instruments psychométriques tels que le PHQ-9 et le BDI-II, ainsi que les considérations éthiques inhérentes à ce champ de recherche.

Une revue de littérature structurée est menée selon huit axes thématiques : les méthodes fondamentales de génération de texte contrôlée, les systèmes existants de simulation de détresse psychologique par LLM, les modèles de langage spécialisés pour la santé mentale, les stratégies d'adaptation pour un déploiement local, les contraintes de déploiement hors ligne et de confidentialité, les jeux de données et benchmarks de référence, les méthodologies d'évaluation, et les considérations éthiques. Cette revue situe le problème ciblé, la génération de textes de détresse contrôlés en sévérité à l'aide d'outils exclusivement open-source et déployables hors ligne, par rapport à l'état de l'art actuel, et identifie les lacunes qui justifient la mise en uvre concrète d'un tel système.

Dans son ensemble, ce travail établit le socle conceptuel et une synthèse critique des travaux antérieurs nécessaires pour éclairer la conception d'un pipeline de génération de textes de détresse conditionnée par la sévérité, tout en mettant en lumière les choix méthodologiques, les questions ouvertes et les garde-fous éthiques que toute mise en uvre de ce type doit prendre en compte.

Mots clés : Simulation de détresse psychologique, Génération de texte contrôlée, Grands modèles de langage, Affinage paramétriquement efficace, TAL en santé mentale, Revue de littérature.

ملخص

يواجه المهنيون العاملون في مجال الوقاية والدعم النفسي بشكل منتظم حالات تعبر عن ضيق نفسي، إلا أن إعداد موارد تدريبية تعتمد على روايات أزمت حقيقية يطرح مشكلات جدية تتعلق بالخصوصية، فضلاً عن صعوبة إنتاجها على نطاق واسع. تتناول هذه المذكرة مدى صلاحية النماذج اللغوية مفتوحة المصدر، القابلة للنشر دون اتصال بالإنترنت، كأساس لتوليد نصوص اصطناعية تحاكي بشكل واقعي الضيق النفسي ضمن مستويات شدة محكومة، بما يوفر بديلاً يحافظ على الخصوصية لأغراض التدريب، وتقييم أنظمة الكشف الآلي، والتوعية العامة.

ولتحقيق هذا الهدف، تُرسى الأسس النظرية والمنهجية التي يقوم عليها نظام من هذا النوع. ويشمل ذلك بنية المحول (Transformer) ونموذج التدريب المسبق للنماذج اللغوية الكبيرة، والعائلات الرئيسية لتقنيات الضبط الدقيق ذي الكفاءة البارامترية (وفي مقدمتها، LoRA فضلاً عن أبرز مقاربات توليد النصوص المُتحكَّم فيه، بما فيها الاشتراط برموز التحكم، والتوجيه اللاحق، والتوليد الموجه بمُميّز، (discriminator) وضبط البادئات (prefix). tuning) ثم تُستعرض نقطة التقاطع بين معالجة اللغة الطبيعية والبحث في الصحة النفسية، بما يشمل تصنيف النصوص للكشف عن الاكتئاب وتقييم خطر الانتحار، والتمثيلات الدلالية النصية الخاصة بالجال، والأدوات القياسية النفسية مثل PHQ-9 وBDI-II، إضافة إلى الاعتبارات الأخلاقية الملزمة لهذا المجال البحثي.

وتُجرى مراجعة أدبيات منظّمة وفق ثمانية محاور موضوعية: الأساليب الأساسية لتوليد النصوص المُتحكَّم فيه، والأنظمة القائمة لمحاكاة الضيق النفسي باستخدام النماذج اللغوية الكبيرة، والنماذج اللغوية المتخصصة في الصحة النفسية، واستراتيجيات التكيف من أجل النشر المحلي، وقيود النشر دون اتصال والخصوصية، ومجموعات البيانات والمعايير المرجعية، ومنهجيات التقييم، والاعتبارات الأخلاقية. وتُوقع هذه المراجعة المشكلة المستهدفة، أي توليد نصوص ضيق نفسي محكومة الشدة باستخدام أدوات مفتوحة المصدر حصراً وقابلة للنشر دون اتصال، ضمن حالة المعرفة الراهنة، وتحدّد الثغرات التي تُبرّر التنفيذ الفعلي لنظام من هذا النوع.

وإجمالاً، يُؤسّس هذا العمل القاعدة المفاهيمية والتوليف النقدي للأعمال السابقة، اللازمين لتوجيه تصميم مسار توليد نصوص ضيق مشروط بالشدة، مع تسليط الضوء على الخيارات المنهجية، والمسائل المفتوحة، والضوابط الأخلاقية التي ينبغي لأي تنفيذ من هذا النوع أخذها بعين الاعتبار.

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