الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbes

THESIS

To obtain the diploma of **Engineering degree**

Field of Study: **Computer Science**

Specialization: **Artificial Intelligence and Data Science**

Theme

Frugal Analysis of Ski Outings with Language Models

Presented by
Fettache Dina

Submitted on: **July, 2025**
In front of the jury composed of

Mr. Chaib Souleyman	President of the Jury
Mr. Khaldi Miloud	Thesis Supervisor
Ms. Noha Ibrahim	Co-Supervisor
Mr. Bejaoui Mohammed	Examiner

Academic Year: 2024/2025

Abstract

This work explores the intersection of Natural Language Processing NLP and Geospatial Artificial Intelligence GeoAI by introducing a modular framework for structuring narrative reports of ski touring activities. Based on user content from the French platform Skitour, the approach transforms textual unstructured descriptions into structured and actionable insights.

Ski touring trip reports often encode rich information about participants, conditions, intentions, and routes, but in unstructured form. To address this, we propose a system that leverages both small and large language models (SLMs and LLMs) to interpret, extract, and enrich these narratives with semantic and geospatial attributes. Our contributions span four core areas: estimating group size from linguistic cues, assessing outing risk through multi-agent debate system, identifying ski outings type through behaviour classification, and reconstructing itineraries as geographic routes.

A key focus of the work lies in frugal AI practices, relying on lightweight, open-source models to ensure accessibility and sustainability without compromising performance. The resulting outputs can help in avalanche risk forecasting, behavior monitoring, and the enhancement of geographic databases used by outdoor platforms, researchers, and rescue services.

By structuring community data in a reproducible and interpretable way, this research bridges the gap between narrative user content and spatial decision making tools. The framework opens pathways for more informed outdoor planning and environmental understanding, demonstrating how language models can support human safety and exploration in challenging terrains.

Keywords:NLP, SLM, LLM, ski touring, danger analysis, GPX generation, geospatial AI

Abstrait

Ce travail explore l'intersection entre NLP et GeoAI en introduisant un cadre modulaire pour structurer les rapports narratifs des activités de ski de randonnée. Basée sur le contenu des utilisateurs de la plateforme française Skitour, l'approche transforme les descriptions textuelles non structurées en informations structurées et exploitables.

Les rapports d'excursions de ski de randonnée contiennent souvent des informations riches sur les participants, les conditions, les intentions et les itinéraires, mais sous une forme non structurée. Pour y remédier, nous proposons un système qui s'appuie sur des petits et larges modèles de langage SLM LLM pour interpréter, extraire et enrichir ces récits avec des attributs sémantiques et géospatiaux. Nos contributions couvrent quatre domaines principaux : l'estimation de la taille du groupe à partir d'indices linguistiques, l'évaluation du risque de sortie grâce à un système de débat multi-agents, l'identification du type de sortie à ski grâce à la classification des comportements et la reconstitution des itinéraires sous forme de routes géographiques.

L'un des principaux axes de travail réside dans les pratiques d'IA frugales, qui s'appuient sur des modèles légers et open-source pour garantir l'accessibilité et la durabilité sans compromettre les performances. Les résultats obtenus peuvent contribuer à la prévision des risques d'avalanche, à la surveillance des comportements et à l'amélioration des bases de données géographiques utilisées par les plateformes de plein air, les chercheurs et les services de secours.

En structurant les données communautaires de manière reproductible et interprétable, cette recherche comble le fossé entre les données na

Mots clés:NLP, SLM, LLM, randonné de ski, analyse de danger,generation des traces GPX, AI géospatiale

الملخص

يتناول هذا العمل تقاطع تقنيات NLP والذكاء الاصطناعي الجغرافي (GeoAI) من خلال تقديم إطار معياري لتحليل وتنظيم التقارير السردية الخاصة بأنشطة التزلج الجبلي. بالاعتماد على المحتوى الذي ينشره المستخدمون في المنصة الفرنسية Skitour، يقوم هذا الإطار بتحويل الأوصاف النصية غير المنظمة إلى معلومات مهيكلة وقابلة للاستخدام العملي.

عادةً ما تتضمن تقارير التزلج الجبلي معلومات غنية عن المشاركين، والظروف، والنوايا، والمسارات، لكنها تأتي في شكل غير مهيكل. لمواجهة هذا التحدي، نقترح نظامًا يعتمد على نماذج لغوية صغيرة وكبيرة (SLMs وLLMs) من أجل تفسير هذه السرديات، واستخلاص خصائصها الدلالية والجغرافية. تتركز مساهماتنا في أربعة محاور رئيسية: تقدير حجم المجموعة انطلاقاً من المؤشرات اللغوية، تقييم مستوى الخطر عبر نظام مناظرة متعدد العوامل، تصنيف نوع الخروج من خلال تحليل السلوك، وإعادة بناء المسارات كخطوط جغرافية.

يركز العمل على تبني ممارسات الذكاء الاصطناعي المُقتصد، بالاعتماد على نماذج مفتوحة المصدر وخفيفة الوزن لضمان الاستدامة وسهولة الوصول دون التضحية بالكفاءة. ويمكن الاستفادة من المخرجات الناتجة في التنبؤ بخطر الانهيارات الثلجية، ومراقبة سلوك المشاركين، وتعزيز قواعد البيانات الجغرافية المستخدمة من قِبل المنصات المتخصصة، والباحثين، وخدمات الطوارئ.

من خلال تنظيم بيانات المجتمع بطريقة قابلة للتكرار والتفسير، يساهم هذا العمل في سد الفجوة بين المحتوى السردي للمستخدمين وأدوات اتخاذ القرار الجغرافي. يفتح هذا الإطار آفاقًا جديدة لتخطيط أفضل للأنشطة الجبلية وفهم بيئي أعمق، ويُظهر كيف يمكن للنماذج اللغوية أن تعزز السلامة البشرية ودعم الاستكشاف في البيئات الوعرة.

الكلمات المفتاحية: NLP، SLM، LLM، التزلج الجبلي، تحليل المخاطر، توليد GPX، الذكاء الاصطناعي الجغرافي