

Aplicación de modelos de lenguaje para el análisis geoquímico ambiental de suelos en la región de Antofagasta

Language model applications for environmental geochemical analysis of soils in the Antofagasta region

Benjamín Ibarra-Campillay¹  <https://orcid.org/0009-0005-2880-788X>

Elizabeth Lam-Esquenazi²  <https://orcid.org/0000-0002-0388-4660>

Brian Keith-Norambuena^{3*}  <https://orcid.org/0000-0001-5734-8962>

Recibido 27 de agosto de 2025, aceptado 26 de diciembre de 2025

Received: August 27, 2025 Accepted: December 26, 2025

RESUMEN

Este trabajo presenta el desarrollo de un sistema prototipo que utiliza modelos de lenguaje de gran escala (LLMs) para el análisis geoquímico ambiental de suelos en la Región de Antofagasta, Chile. El sistema integra técnicas de Retrieval Augmented Generation (RAG), principios de ingeniería de prompts estructurada inspirados en DSPy, y un mecanismo de autoevaluación mediante un método de LLM-as-a-judge. Se procesaron 94 muestras de suelo con datos de 12 elementos químicos (B, Co, Fe, Zn, Mn, Pb, As, Cu, Cr, Ni, V, Al), implementando transformación de coordenadas UTM a WGS84, análisis de 2 componentes principales (PCA), detección de anomalías basada en percentil 95 para elementos críticos (As, Pb, Cu) y generación automática de informes técnicos. El sistema de autoevaluación emplea 6 criterios específicos con escala 1-5 y temperatura 0.2 para consistencia. El sistema desarrollado demuestra la viabilidad técnica de especializar LLMs genérico en el dominio científico del análisis geoquímico ambiental mediante la implementación de una base de conocimiento contextual y características geológicas regional usando un enfoque local con Ollama. Aunque se logró generar informes estructurados con información relevante y se implementó un prototipo funcional de evaluación automatizada, la efectividad cuantitativa del sistema requiere validación sistemática mediante comparación con análisis de expertos humanos. Este trabajo establece las bases metodológicas y técnicas para adaptar LLMs a dominios científicos especializados sin reentrenamiento costoso, identificando tanto el potencial como las limitaciones actuales de este enfoque.

Palabras clave: Modelos de lenguaje, geoquímica ambiental, análisis de suelos.

ABSTRACT

This work describes the development of a prototype system utilizing large language models (LLMs) for environmental geochemical analysis of soils in the Antofagasta Region, Chile. The system integrates Retrieval Augmented Generation (RAG) techniques, structured prompt engineering principles based on DSPy, and a self-evaluation mechanism employing an LLM-as-a-judge approach. A total of 94 soil samples were processed, each containing data from 12 chemical elements (B, Co, Fe, Zn, Mn, Pb, As, Cu, Cr, Ni, V, Al).

¹ Universidad Católica del Norte. Departamento de Ingeniería Industrial. Antofagasta, Chile. Email: benjamin.ibarra@ucn.cl

² Universidad Católica del Norte. Departamento de Ingeniería Química y Medio Ambiente. Antofagasta, Chile. Email: elam@ucn.cl

³ Universidad Católica del Norte. Departamento de Ingeniería de Sistemas y Computación. Antofagasta, Chile. Email: brian.keith@ucn.cl

* Autor de correspondencia: brian.keith@ucn.cl

The methodology included UTM-to-WGS84 coordinate transformation, 2 principal component analysis (PCA), percentile 95-based anomaly detection for critical elements (As, Pb, Cu), and automatic generation of technical reports. The self-evaluation system applies six specific criteria, each rated on a 1-5 scale, and uses a temperature setting of 0.2 to ensure consistency. The developed system demonstrates the technical feasibility of specializing generic LLMs to the scientific domain of environmental geochemical analysis by incorporating a contextual knowledge base and regional geological characteristics via Ollama. Although the system successfully generates structured reports containing relevant environmental information and implements a functional automated evaluation prototype, its quantitative effectiveness requires systematic validation by comparing its analyses with those conducted by a human expert. This study establishes the methodological and technical foundations for adapting LLMs to specialized scientific domains without costly retraining, while also identifying the approach's potential and current limitations.

Keywords: Language models, environmental geochemistry, soil analysis.

INTRODUCCIÓN

La evaluación de la calidad ambiental de suelos en regiones mineras constituye un desafío técnico y operacional significativo que requiere la integración de múltiples disciplinas científicas [1], [2]. En la Región de Antofagasta, Chile, la intensa actividad extractiva histórica y actual ha generado concentraciones elevadas de elementos potencialmente tóxicos como As, Cu, y Zn que exceden significativamente los límites establecidos en normativas internacionales [3], [4]. La interpretación de estos datos geoquímicos complejos requiere conocimiento especializado y tiempo considerable para generar informes técnicos que apoyen la toma de decisiones ambientales [5].

Los avances recientes en inteligencia artificial, particularmente los modelos de lenguaje de gran escala (LLMs), presentan oportunidades sin precedentes para automatizar parcialmente estos procesos analíticos [6], [7]. Sin embargo, la aplicación directa de LLMs genéricos a dominios científicos especializados enfrenta limitaciones importantes que incluyen la falta de conocimiento específico del dominio, la tendencia a generar información incorrecta pero plausible conocida como *alucinaciones*, y la ausencia de contextualización local necesaria para interpretaciones válidas [8], [9].

La arquitectura *transformer*, introducida por Vaswani *et al.* [10], revolucionó el procesamiento del lenguaje natural al eliminar las conexiones recurrentes tradicionales y depender completamente de mecanismos de atención. Esta innovación permitió el desarrollo de modelos cada vez más grandes y capaces, desde GPT-1 con 117 millones

de parámetros hasta GPT-4 con estimaciones que superan el billón de parámetros [11], [12]. Estos modelos han demostrado capacidades emergentes notables, incluyendo razonamiento complejo, aprendizaje few-shot y seguimiento de instrucciones detalladas [13].

Para abordar las limitaciones de los LLMs en dominios especializados, Lewis *et al.* [14] introdujeron el paradigma de Retrieval Augmented Generation (RAG), que permite a los modelos acceder a conocimiento externo durante la inferencia. Esta técnica ha demostrado reducir significativamente las alucinaciones y mejorar la precisión factual en múltiples dominios [15], [16]. Paralelamente, los avances en ingeniería de prompts han evolucionado desde ajustes manuales hacia enfoques sistemáticos, culminando en frameworks como DSPy que tratan los prompts como módulos programáticos optimizables [17].

El contexto geoquímico de la Región de Antofagasta presenta características únicas que deben considerarse en cualquier sistema de análisis. La región se caracteriza por condiciones hiperáridas con precipitación inferior a 2 mm anuales, lo que limita la lixiviación vertical de contaminantes pero favorece su acumulación superficial [18], [19]. La geología dominada por pórfidos de cobre resulta en un enriquecimiento natural de elementos como Cu, Mo y As que puede confundirse con contaminación antropogénica [20].

Este trabajo aborda estas limitaciones mediante el desarrollo de un sistema que especializa LLMs para el análisis geoquímico ambiental sin requerir

el costoso reentrenamiento del modelo base. La aproximación propuesta combina tres elementos técnicos principales. Primero, la implementación de RAG con una base de conocimiento geoquímico específicamente curada para el contexto de Antofagasta. Segundo, la aplicación de principios de ingeniería de prompts estructurada inspirados en el framework DSPy [17], que permite un control más preciso sobre el comportamiento del modelo. Tercero, el desarrollo de un mecanismo innovador de autoevaluación denominado *LLM-as-a-judge* que proporciona control de calidad automatizado sobre los informes generados.

La relevancia de este trabajo se fundamenta en la creciente necesidad de herramientas que aceleren la evaluación ambiental en regiones mineras, donde el volumen de datos geoquímicos supera la capacidad de análisis manual disponible [21]. Además, la aproximación propuesta tiene implicaciones más amplias para la aplicación de LLMs en otros dominios científicos especializados donde el reentrenamiento no es viable por restricciones de recursos o privacidad de datos [22].

En consideración de estos elementos, las preguntas de investigación que guían este trabajo son:

- **RQ1:** ¿Es técnicamente viable especializar LLMs genéricos para el análisis geoquímico ambiental sin reentrenamiento costoso?
- **RQ2:** ¿Puede la implementación de principios de ingeniería de prompts estructurada mejorar la consistencia y calidad de los informes generados?
- **RQ3:** ¿Es técnicamente viable implementar un sistema de autoevaluación automatizada para informes técnicos geoquímicos?

METODOLOGÍA

Alcance del sistema

El prototipo desarrollado procesa conjuntos de datos geoquímicos de cualquier tamaño en formato Excel con estructura estandarizada, aunque la validación inicial utilizó 94 muestras de la Región de Antofagasta. El tamaño del conjunto de validación estuvo limitado por los datos disponibles en el momento del estudio. El sistema es escalable y puede procesar nuevos datasets que mantengan la misma estructura de columnas y coordenadas UTM Zone 19S y concentraciones de los elementos analizados. La arquitectura modular

permite adaptar el sistema a otras regiones modificando únicamente el polígono de referencia y ajustando el contexto RAG regional.

Arquitectura del sistema

El sistema desarrollado implementa una arquitectura modular de tres componentes principales que interactúan de manera coordinada para procesar datos geoquímicos y generar informes analíticos especializados. El componente de procesamiento de datos, implementado como un backend Flask, gestiona la carga, transformación y análisis preliminar de datos geoquímicos mediante bibliotecas especializadas como Pandas [23] y GeoPandas [24]. Este módulo realiza operaciones críticas incluyendo la transformación de sistemas de coordenadas, normalización de datos y cálculos estadísticos multivariados.

El componente de inteligencia artificial constituye el núcleo analítico del sistema y ejecuta un modelo de lenguaje de gran escala de manera local mediante la plataforma Ollama [25]. Esta implementación local es fundamental para mantener la confidencialidad de datos geoquímicos sensibles que pueden tener implicaciones económicas significativas en el contexto minero. El módulo gestiona la interacción con el LLM a través de una API REST personalizada que encapsula la lógica de generación de prompts y procesamiento de respuestas.

El tercer componente proporciona una interfaz web interactiva basada en la biblioteca Leaflet [26] que permite la exploración espacial de datos geoquímicos. La visualización incluye representación de puntos de muestreo, interpolación espacial mediante *Inverse Distance Weighting* (IDW) y generación de mapas temáticos para diferentes elementos químicos. La arquitectura completa se comunica mediante formatos estándar JSON y GeoJSON, garantizando la interoperabilidad entre componentes. La Figura 1 muestra una visión general de la propuesta.

El sistema implementa una arquitectura distribuida con componentes especializados que se comunican mediante APIs REST y formatos de datos estandarizados. La Tabla 1 detalla los componentes técnicos específicos y sus funciones.

Procesamiento de datos geoquímicos

El conjunto de datos analizado comprende 94 muestras de suelo superficial recolectadas a

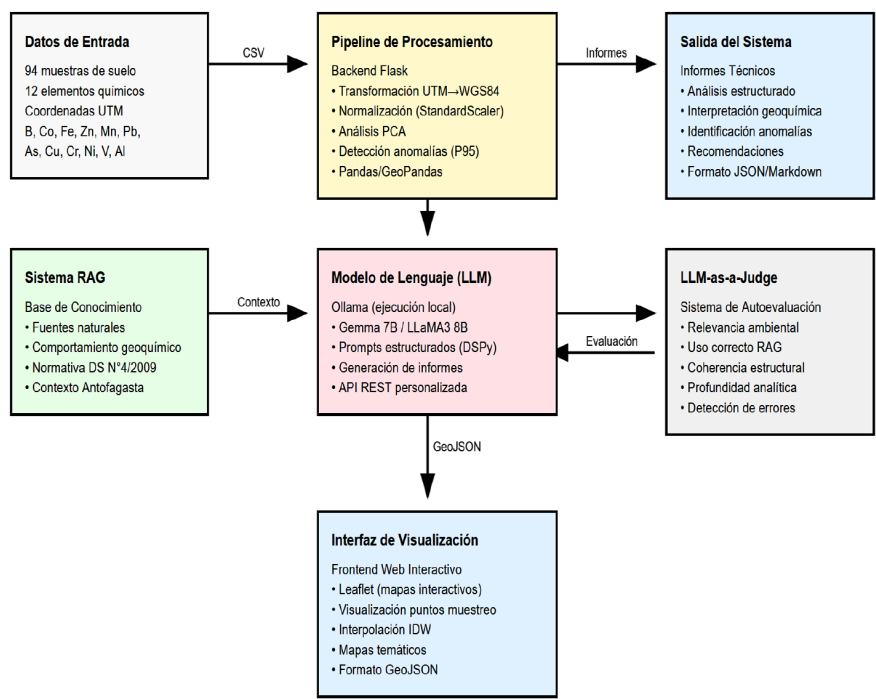


Figura 1. Arquitectura del sistema de análisis geoquímico basado en modelos de lenguaje.

Tabla 1. Componentes técnicos del sistema implementado.

Componente	Implementación	Función específica
Cliente LLM.	OllamaClient.	Interfaz con API local en puerto 11434, modelos Gemma 7B y LLaMA3 8B.
Backend.	Flask.	Servidor en puerto 5000, gestión de rutas / generate_report, /get_interpolated_data.
Procesamiento geoespacial.	GeoPandas con PyProj.	Transformación UTM Zone 19S (EPSG:32719) a WGS84 (EPSG:4326).
Análisis multivariado.	scikit-learn (PCA, StandardScaler).	Reducción dimensional, normalización de 12 elementos.
Interpolación espacial.	IDW personalizado.	Grid 40×40, power = 2, clipping con polígono regional.
Base de conocimiento.	Diccionario Python estructurado.	12 elementos con 8 atributos cada uno.
Visualización.	Leaflet con esquema Magma.	Renderizado de clusters, outliers, interpolaciones.
Evaluador.	LLM-as-Judge con temperatura 0.2.	Evaluación estructurada con 6 criterios.

profundidades de 0-20 cm en la Región de Antofagasta, siguiendo protocolos estandarizados de muestreo geoquímico [27], [28]. Se debe notar que, aunque la validación se realizó con estas 94 muestras de la Región de Antofagasta, el sistema puede procesar cualquier conjunto de datos con estructura similar. Cada muestra fue analizada para determinar las concentraciones totales de 12

elementos químicos (B, Co, Fe, Zn, Mn, Pb, As, Cu, Cr, Ni, V, Al) mediante espectrometría de masas con plasma acoplado inductivamente (ICP-MS) después de digestión ácida completa [29].

Los datos originales, georreferenciados en el sistema de coordenadas proyectadas UTM Zone 19S (EPSG:32719), requirieron transformación al sistema

geográfico WGS84 (EPSG:4326) para su correcta visualización en mapas web. Esta transformación se implementó utilizando la biblioteca PyProj integrada en GeoPandas, aplicando los parámetros de transformación específicos para la zona de estudio [30]. La precisión de la transformación se verificó mediante puntos de control con coordenadas conocidas en ambos sistemas.

El preprocesamiento de datos incluyó múltiples etapas de control de calidad. Los registros con valores faltantes en cualquiera de los elementos químicos analizados fueron eliminados para mantener la integridad del análisis multivariado. Los datos se sometieron a transformación logarítmica para aproximar una distribución normal, requisito fundamental para muchas técnicas estadísticas aplicadas posteriormente [31]. La normalización mediante StandardScaler de scikit-learn [32] se aplicó para eliminar efectos de escala en análisis multivariados, preservando las proporciones relativas entre elementos.

Para el análisis de componentes principales (PCA), se empleó la implementación de scikit-learn configurada para extraer dos componentes que maximizaran la varianza explicada. Este análisis permitió identificar asociaciones geoquímicas características que pueden indicar procesos naturales o antropogénicos [33]. La identificación de valores anómalos se realizó mediante un enfoque estadístico basado en percentiles, específicamente marcando como outliers aquellas muestras que excedían el percentil 95 para elementos indicadores clave (As, Pb, Cu), seleccionados por su relevancia ambiental y asociación con actividades mineras [34].

Las muestras identificadas como outliers reciben un tratamiento diferenciado en el sistema. Estas muestras se marcan con el atributo *outlier_type*=“higher” y se visualizan como triángulos rojos en los mapas interactivos. En los informes generados, estas anomalías se analizan en una sección dedicada que evalúa sus posibles orígenes naturales versus antropogénicos, considerando el contexto geológico y minero de Antofagasta. Aunque el presente trabajo no implementa un tratamiento estadístico diferenciado de estos valores extremos, su identificación permite priorizar estas ubicaciones para investigación adicional y muestreo de confirmación.

Especialización del LLM mediante RAG

La implementación del sistema RAG constituye la estrategia central de este trabajo, diseñada específicamente para inyectar conocimiento geoquímico especializado en el proceso de generación del LLM. A diferencia de implementaciones RAG tradicionales que dependen de bases de datos vectoriales y recuperación dinámica [14], [15], se desarrolló un sistema RAG simulado que mantiene una base de conocimiento estructurada directamente en el código Python.

Esta base de conocimiento contiene información detallada para cada elemento químico analizado, estructurada en categorías específicas que reflejan las necesidades del análisis ambiental. Para cada elemento se documenta información sobre fuentes naturales típicas, incluyendo minerales primarios y procesos geológicos asociados. Las fuentes antropogénicas se detallan considerando las actividades industriales relevantes en la región, como minería, fundición y transporte de minerales [35]. El comportamiento geoquímico en suelos se describe incluyendo información sobre especiación química, movilidad en función del pH y potencial redox, y mecanismos de adsorción a componentes del suelo [36], [37].

La base de conocimiento implementada contiene información estructurada para los 12 elementos analizados (B, Co, Fe, Zn, Mn, Pb, As, Cu, Cr, Ni, V, Al), cada uno documentado con ocho categorías de información: fuentes naturales típicas, fuentes antropogénicas típicas, significado ambiental, comportamiento en el suelo incluyendo movilidad y especiación, implicaciones para la salud, guías genéricas de referencia en suelo, y cuando aplica, se referencia a la normativa chilena. Esta estructura permite al sistema mantener aproximadamente 3.000 tokens de contexto especializado disponible durante la generación, optimizando el uso del límite de contexto del modelo que típicamente es de 8.000 tokens para Gemma 7B.

Las implicaciones ambientales y para la salud humana se documentan basándose en literatura científica actualizada y guías de organizaciones internacionales [38], [39]. Particularmente importante es la inclusión de valores de referencia tanto genéricos como específicos de la normativa chilena, específicamente el Decreto Supremo N° 4/2009

del Ministerio del Medio Ambiente que establece límites para metales en suelos agrícolas [40]. Esta información normativa es crítica para contextualizar los resultados en el marco regulatorio nacional.

Ingeniería de prompts con principios DSPy

La aplicación de principios del framework DSPy [17] representa un avance metodológico significativo en la estructuración y optimización de prompts para tareas analíticas complejas. DSPy propone tratar los prompts no como strings estáticos sino como módulos programáticos con entradas y salidas bien definidas, susceptibles de optimización automática [41].

En este trabajo se implementaron tres principios fundamentales de DSPy. Primero, la definición de “firmas conceptuales” para cada tipo de análisis, especificando explícitamente las entradas esperadas y la estructura de salida deseada. Por ejemplo, para el análisis de un elemento específico, la firma define como entradas el símbolo del elemento, el contexto regional de Antofagasta, y los datos RAG relevantes. Las salidas esperadas incluyen secciones estructuradas sobre comportamiento geoquímico, identificación de fuentes, interpretación de patrones de concentración, evaluación de riesgos, correlaciones geoquímicas esperadas, y referencias a normativa aplicable.

Segundo, se desarrollaron plantillas modulares reutilizables para cada tipo de análisis requerido por el sistema. Estas plantillas incluyen análisis general del conjunto de datos, interpretación de clusters geoquímicos, identificación y caracterización de anomalías, análisis detallado de elementos específicos, y generación de recomendaciones para investigación adicional. Cada plantilla mantiene una estructura consistente, pero permite personalización según el contexto específico del análisis.

Tercero, se implementaron instrucciones estructuradas que obligan al LLM a seguir un formato específico en sus respuestas. Esto incluye el uso de numeración obligatoria para diferentes secciones del análisis, lo que garantiza cobertura completa de todos los aspectos relevantes y facilita la evaluación automatizada posterior. Las instrucciones también especifican el nivel de detalle técnico esperado y la necesidad de fundamentar afirmaciones en el conocimiento RAG proporcionado.

Como ejemplo concreto de la implementación de firmas conceptuales, el análisis de elementos específicos define entradas y salidas estructuradas de la siguiente manera: Las entradas incluyen *element_symbol* como string (por ejemplo “As”), *element_name* como string (por ejemplo “Arsenic”), *environmental_context* como diccionario con 8 categorías de información, y *regional_context* especificando “Antofagasta region, Chile (arid, mining-influenced)”. Las salidas siguen una estructura markdown obligatoria con siete secciones numeradas: “Comportamiento Geoquímico y Movilidad” específico para suelos áridos, “Fuentes Potenciales en Antofagasta” tanto naturales como antropogénicas, “Interpretación de Patrones de Concentración” comparando con referencias, “Riesgos Ambientales y para la Salud” basados en el contexto, “Posibles Correlaciones Geoquímicas” con otros elementos, “Consideraciones Mineras Específicas” para el contexto regional, y “Referencias a Guías y Normativas” aplicables. Esta estructuración garantiza cobertura completa y formato consistente en todos los informes generados.

Sistema de evaluación LLM-as-a-judge

El desarrollo del sistema LLM-as-a-judge permite la evaluación automatizada de la calidad de informes técnicos generados por IA [42], [43]. Este mecanismo utiliza un agente LLM con un prompt especializado para evaluar críticamente los outputs del agente LLM principal según criterios predefinidos. El sistema de evaluación implementa seis criterios específicamente diseñados para informes de análisis geoquímico ambiental, estructurados según se detalla en la Tabla 2.

El sistema genera evaluaciones en formato JSON estructurado con justificaciones textuales para cada puntuación, utilizando temperatura 0.2 para mayor consistencia en las evaluaciones. Cada evaluación incluye además sugerencias específicas de mejora que podrían utilizarse en iteraciones futuras del sistema.

La implementación técnica del evaluador requirió el desarrollo de prompts altamente estructurados que guían al LLM evaluador a proporcionar puntuaciones numéricas (escala 1-5) para criterios cuantitativos y evaluaciones categóricas donde corresponde. Además, el sistema solicita justificaciones textuales para cada puntuación y sugerencias específicas

Tabla 2. Criterios de evaluación implementados en el sistema LLM-as-Judge.

Criterio	Escala	Descripción	Peso en evaluación
C1: Relevancia Ambiental	1-5	Enfoque en calidad ambiental versus aspectos puramente geológicos o de exploración minera.	Crítico
C2: Uso del Contexto RAG	1-5	Precisión en el uso de información proporcionada, evitando omisiones o adiciones no fundamentadas.	Crítico
C3: Coherencia Estructural	1-5	Flujo lógico de ideas, claridad técnica y organización del documento.	Importante
C4: Profundidad Analítica	1-5	Interpretación y comprensión de procesos versus descripción superficial.	Importante
C5: Calidad Recomendaciones	1-5 o N/A	Especificidad, viabilidad práctica y relevancia de sugerencias.	Contextual
C6: Detección de Errores	Categorico	Clasificación como Sin errores, Errores Menores o Errores Graves.	Eliminatorio

de mejora, creando un ciclo de retroalimentación potencialmente automatizable.

Implementación técnica

La implementación del sistema se realizó completamente en Python 3, aprovechando un ecosistema maduro de bibliotecas científicas y de desarrollo web. Flask [44] proporciona el framework web que gestiona las solicitudes HTTP y coordina la interacción entre componentes. Para el procesamiento de datos geoquímicos se utilizaron Pandas [23] para manipulación tabular general y GeoPandas [24] para operaciones espaciales específicas, incluyendo transformaciones de coordenadas y operaciones geométricas.

El análisis estadístico multivariado se implementó mediante Scikit-learn [32], utilizando específicamente sus módulos de preprocesamiento (StandardScaler), descomposición (PCA) y detección de anomalías. Para la ejecución local de modelos de lenguaje se empleó Ollama [25], una plataforma que permite ejecutar LLMs de código abierto como Gemma 7B [45] y LLaMA3 8B [46] sin dependencias de servicios cloud externos.

La interfaz de visualización se construyó sobre Leaflet [26], una biblioteca JavaScript de código abierto para mapas interactivos. Se implementaron capas personalizadas para mostrar puntos de muestreo, superficies interpoladas y polígonos de la región de estudio. La comunicación entre backend y frontend utiliza formatos estándar JSON para datos tabulares y GeoJSON para información geoespacial, garantizando compatibilidad con herramientas GIS externas.

RESULTADOS

Este trabajo presenta el desarrollo de un sistema prototipo que integra modelos de lenguaje de gran escala con análisis geoquímico ambiental. En particular, se logró la integración completa de los componentes en una aplicación web funcional. El backend Flask gestiona exitosamente la carga y procesamiento de datos. El frontend basado en Leaflet visualiza correctamente los 94 puntos de muestreo sobre el mapa de Antofagasta y permite interacción con los datos. El sistema genera informes estructurados que integran el contexto RAG y siguen los formatos definidos mediante los principios DSPy. La Tabla 3 presenta un extracto representativo del análisis generado para un elemento específico, ilustrando la profundidad técnica y contextualización lograda.

Este ejemplo demuestra cómo el sistema integra conocimiento del contexto RAG (fuentes naturales y antropogénicas), considera el contexto regional específico (pórfidos de cobre), y estructura el análisis siguiendo la firma conceptual definida.

Comparación de informes generados

Como parte del desarrollo del sistema, se realizó una comparación exploratoria entre informes generados con y sin la aplicación de principios DSPy. Aunque limitada a un caso de estudio, esta comparación sugiere diferencias potencialmente relevantes que merecen investigación futura. El informe generado aplicando principios DSPy mostró una organización más sistemática, con secciones numeradas y categorizadas. Por ejemplo, mientras el informe control presentó

Tabla 3. Extracto del análisis de Arsénico generado por el sistema con principios DSPy.

Arsenic (As): Highly toxic and carcinogenic.	
Environmental Implications:	
Elevated concentrations of arsenic pose significant threats to soil health and ecosystem integrity. Arsenic can accumulate in the food chain, potentially impacting plant growth and biodiversity. Leaching of arsenic can contaminate groundwater resources, posing risks to human health through drinking water exposure. Dust inhalation and soil ingestion pathways can expose individuals to arsenic, potentially leading to serious health issues including cancer.	
Role of Major Ore Deposits:	
The presence of significant porphyry copper systems in the Antofagasta region highlights the potential for both natural enrichment of arsenic in sulfide minerals and volcanic emissions, as well as anthropogenic introduction through mining activities (particularly gold and copper extraction) and historical use of arsenic-based pesticides.	
Sources:	
Natural: Sulfide minerals, volcanic emissions, geothermal areas. Anthropogenic: Mining (gold, copper), pesticides (historical), wood preservatives.	

recomendaciones como puntos independientes, la versión DSPy las agrupó en categorías funcionales. De manera similar, el tratamiento de elementos individuales fue más detallado en la versión DSPy, incluyendo menciones específicas tanto de fuentes naturales como antropogénicas para cada elemento analizado. Sin embargo, es importante señalar que estas observaciones se basan en una única comparación con el objetivo de verificar la viabilidad del prototipo implementado. La Tabla 4 ilustra diferencias concretas observadas entre ambos enfoques en aspectos clave del análisis.

Prototipo del sistema LLM-as-a-judge

Se desarrolló un mecanismo funcional de autoevaluación donde un LLM evalúa la calidad

de los informes generados según seis criterios predefinidos: relevancia ambiental, uso correcto del contexto RAG, coherencia y estructura, profundidad analítica, calidad de recomendaciones, y detección de errores factuales. El sistema genera evaluaciones en formato JSON estructurado con puntuaciones numéricas (1-5) y comentarios justificativos.

Se realizó una prueba preliminar del mecanismo de evaluación automática aplicándolo a uno de los informes generados. El sistema fue capaz de producir una evaluación estructurada en formato JSON, asignando puntuaciones numéricas y proporcionando comentarios textuales. En esta prueba inicial, el evaluador asignó puntuaciones altas (5/5) en los criterios definidos y sugirió áreas de mejora

Tabla 4. Comparación entre informes generados sin y con principios DSPy.

Aspecto	Sin DSPy	Con DSPy
Estructura del análisis general.	Listado simple de elementos con características básicas.	Secciones organizadas: “Elements of Concern,” “Environmental Implications,” “Role of Major Ore Deposits”.
Contextualización regional.	Mención genérica: “hyper-arid region of northern Chile”.	Específica: “arid climate and presence of significant mineral deposits like porphyry copper systems”.
Tratamiento de cromo.	Descripción breve: “Speciation is key”.	Diferenciación detallada: “Cr (III) is essential, relatively immobile and low toxicity, while Cr (VI) is highly mobile, toxic, and carcinogenic”.
Recomendaciones.	7 puntos no categorizados.	6 categorías funcionales: Site-Specific Investigations, Analytical Enhancements, Risk Assessment, Data Integration, Monitoring, Comparison with Standards.
Mención de normativa.	Referencias genéricas a “Chilean national standards”.	Referencia específica: “European Union Soil Quality Directive (2011/69/EU)” como ejemplo de comparación.

relacionadas con propiedades físico-químicas del suelo. Este resultado demuestra la viabilidad técnica del sistema para generar evaluaciones formateadas y aparentemente coherentes. La Tabla 5 muestra un ejemplo concreto de la evaluación estructurada generada por el sistema.

La evaluación demuestra que el sistema puede generar tanto puntuaciones cuantitativas consistentes como retroalimentación cualitativa específica y accionable, aunque se requiere validación con expertos humanos para confirmar la confiabilidad de estos juicios.

En general, durante las pruebas de desarrollo, el sistema demostró capacidad para seguir el formato solicitado y proporcionar retroalimentación coherente. Sin embargo, se requiere validación sistemática para cuantificar su efectividad real. La ausencia de comparación con evaluaciones humanas expertas y la limitación a un único caso de prueba impiden extraer conclusiones sobre la confiabilidad o utilidad práctica del mecanismo. Estas pruebas iniciales simplemente confirman que el sistema puede ejecutarse y producir outputs en el formato esperado. Se debe notar que el sistema mostró una tendencia a asignar puntuaciones altas, lo que podría indicar un sesgo optimista. En este contexto, el uso de enfoques multiagente con múltiples evaluadores podría mitigar este sesgo en futuras implementaciones.

Procesamiento de datos geoquímicos reales

El prototipo demostró capacidad técnica para procesar un conjunto de 94 muestras de suelo de la

Región de Antofagasta con análisis de 12 elementos químicos. La implementación de la transformación de coordenadas de UTM Zone 19S (EPSG:32719) a WGS84 (EPSG:4326) se realizó con éxito utilizando PyProj integrado en GeoPandas, logrando la correcta georreferenciación y visualización de todos los puntos de muestreo en el mapa interactivo.

El pipeline de procesamiento implementado ejecuta ocho etapas secuenciales con manejo de errores. Primero, realiza la carga de datos desde archivo Excel con validación de columnas requeridas incluyendo CoordX, CoordY y los 12 elementos químicos. Segundo, elimina registros con valores NaN en cualquier elemento químico para mantener integridad del análisis multivariado. Tercero, ejecuta transformación de coordenadas usando GeoPandas con PyProj, incluyendo un *fallback* de escalado simple si la transformación falla. Cuarto, genera clusters basados en los identificadores de muestra originales cuando están disponibles. Quinto, calcula PCA aplicando StandardScaler para normalización y extrayendo 2 componentes principales que capturan la máxima varianza. Sexto, detecta outliers usando el percentil 95 específicamente para As, Pb y Cu por su relevancia ambiental. Séptimo, realiza filtrado espacial usando el polígono de Antofagasta obtenido desde servicio ArcGIS REST con caché local. Octavo, genera el GeoDataFrame final con estructura consistente para visualización y análisis. El sistema incluye logging detallado en cada etapa para facilitar la depuración y verificación del procesamiento.

Tabla 5. Ejemplo de evaluación generada por el sistema LLM-as-a-Judge.

```
{
  "seccion_evaluada": "Análisis General Ambiental,"
  "evaluacion_general": "Buen análisis que proporciona una comprensión profunda del estado ambiental del suelo en Antofagasta. Las recomendaciones son específicas, accionables y relevantes.,"
  "puntuaciones_criterios": {
    "C1_relevancia_ambiental": "5,"
    "C2_precision_y_uso_contexto_rag": "5,"
    "C3_coherencia_claridad_estructura": "5,"
    "C4_profundidad_analisis": "5,"
    "C5_calidad_recomendaciones": "5,"
    "C6_ausencia_alucinaciones": "Sin errores"
  },
  "sugerencias_mejora": [
    "Considerar la posibilidad de un análisis más detallado de las propiedades físicas y químicas del suelo, como la humedad y la composición. También, se podría incluir una evaluación de la contaminación del suelo y las medidas para remediarla."
  ]
}
```

La estructuración de los datos mediante agrupaciones basadas en los identificadores de muestra originales proporciona una organización coherente del conjunto de datos, facilitando el análisis espacial y la interpretación de patrones geoquímicos por parte del sistema de inteligencia artificial. Esta integración exitosa de datos reales representa un avance fundamental respecto al prototipo inicial, proporcionando una base sólida y auténtica para la generación de análisis ambientales contextualizados.

Potencial del enfoque RAG implementado

El sistema RAG simulado desarrollado representa una aproximación práctica para dominios con conocimiento relativamente estable como la geoquímica ambiental. Al mantener la base de conocimiento estructurada directamente en el código, se logra un control preciso sobre la información disponible para el LLM, eliminando incertidumbres asociadas con sistemas de recuperación dinámica [15], [47]. Esta implementación permite asegurar que información crítica como valores normativos y comportamiento geoquímico de elementos esté siempre disponible durante la generación de informes.

Sin embargo, esta aproximación presenta limitaciones inherentes. La escalabilidad está restringida por los límites de contexto del modelo (típicamente 8000 tokens para modelos como Gemma 7B), requiriendo decisiones cuidadosas sobre qué información priorizar. Además, la actualización del conocimiento requiere modificación directa del código, lo que puede ser menos flexible que sistemas basados en bases de datos vectoriales tradicionales.

Implicaciones de la estructuración de prompts con DSPy

La aplicación de principios del framework DSPy, aunque sin su implementación completa, sugiere que el tratamiento sistemático de prompts como componentes programáticos puede mejorar significativamente la consistencia de sistemas basados en LLMs [48]. La definición de firmas conceptuales claras y el uso de plantillas modulares eliminan ambigüedades que típicamente resultan en outputs variables e impredecibles.

Esta aproximación tiene implicaciones importantes para el desarrollo de aplicaciones de IA en dominios científicos. Sugiere que la ingeniería de prompts puede evolucionar de una práctica artesanal hacia

una disciplina más sistemática, donde los prompts son diseñados, versionados y optimizados como cualquier otro componente de software. Para aplicaciones críticas como el análisis ambiental, esta predictibilidad es fundamental.

Consideraciones sobre el sistema LLM-as-a-judge

El prototipo de autoevaluación implementado representa una dirección prometedora para el control de calidad automatizado en sistemas de generación de texto técnico. La capacidad del sistema para generar evaluaciones estructuradas siguiendo criterios predefinidos sugiere que este paradigma podría ser útil para identificar áreas de mejora durante el desarrollo [42], [43].

Sin embargo, es crucial reconocer las limitaciones fundamentales de este enfoque. Un LLM evaluando outputs de otro LLM (o del mismo modelo) puede perpetuar sesgos sistemáticos y puntos ciegos compartidos. La validación real de la calidad técnica requiere inevitablemente la participación de expertos humanos del dominio, particularmente para verificar la precisión factual de interpretaciones geoquímicas complejas.

Relevancia de la implementación local

La decisión de utilizar Ollama para ejecutar modelos localmente aborda una preocupación crítica en el sector minero-ambiental: la confidencialidad de datos geoquímicos que pueden tener implicaciones económicas significativas [25]. La viabilidad técnica demostrada sugiere que es posible desarrollar aplicaciones de IA sofisticadas sin comprometer la seguridad de datos sensibles.

El rendimiento observado con modelos de 7-8B parámetros en hardware accesible (GPU de gama media) indica que la barrera tecnológica para implementar estos sistemas no es prohibitiva. Esto es particularmente relevante para organizaciones en países en desarrollo o instituciones con recursos limitados que podrían beneficiarse de herramientas de análisis avanzadas.

Implicaciones para la práctica de evaluación ambiental

Aunque el sistema requiere validación extensiva antes de su aplicación práctica, el prototipo desarrollado sugiere varias aplicaciones potenciales en la gestión ambiental. La capacidad de generar

análisis estructurados que integran conocimiento específico del dominio podría acelerar las fases preliminares de evaluación de sitios, permitiendo a los profesionales enfocarse en aspectos que requieren juicio experto [49].

La integración de normativa local relacionada (DS N°4/2009) [40] en el sistema demuestra cómo estas herramientas pueden adaptarse a contextos regulatorios específicos, abordando una limitación común de herramientas genéricas que aplican estándares inapropiados. Esto es particularmente relevante en regiones como Antofagasta donde los valores de fondo natural pueden exceder estándares internacionales debido al enriquecimiento geológico.

Desafíos, limitaciones y direcciones futuras

El trabajo realizado identifica varios desafíos que deben abordarse para el desarrollo de un sistema completamente funcional. Primero, se requiere un diseño experimental que permita cuantificar las mejoras reales en calidad de análisis comparado con métodos tradicionales. Esto incluye la definición de métricas objetivas y la realización de estudios comparativos con evaluadores humanos expertos.

Segundo, la base de conocimiento RAG debe expandirse significativamente para incluir información más detallada sobre procesos de transporte de contaminantes en ambientes áridos, factores de biodisponibilidad, y métodos de remediación específicos para diferentes tipos de contaminación.

Tercero, la integración con flujos de trabajo existentes en organizaciones ambientales requiere consideración cuidadosa. El sistema debe diseñarse para complementar, no reemplazar, las prácticas establecidas de evaluación y los sistemas de información geográfica utilizados actualmente.

Además de los desafíos metodológicos mencionados, el sistema presenta limitaciones técnicas específicas que deben considerarse antes de su aplicación práctica. Específicamente, existe una fuerte dependencia entre el modelo de lenguaje utilizado y los recursos computacionales disponibles. La calidad de los informes varía significativamente según el modelo LLM utilizado (Gemma 2B vs Gemma 7B vs Llama3 8B) y los recursos computacionales disponibles. Los modelos más pequeños mostraron dificultades para seguir instrucciones complejas

y para integrar profundamente el contexto RAG, requiriendo al menos modelos de 7B parámetros para resultados satisfactorios. Por otro lado, para propósitos del desarrollo de este prototipo se han realizado simplificaciones en el análisis de datos. La detección actual de outliers está basada en percentiles y la definición de clusters usando la información muestral son implementaciones funcionales pero simplificadas, pues un análisis más adecuado requeriría técnicas como *Robust PCA* para detección de anomalías geoquímicas y *UMAP/HDBSCAN* para clustering no supervisado.

Finalmente, la validación del componente LLM-as-a-judge constituye una prioridad para trabajo futuro. Se requiere diseñar un protocolo de validación que incluya comparación sistemática con evaluaciones de expertos humanos en geoquímica ambiental, cálculo de métricas de concordancia inter-evaluador, y análisis de sesgo potencial en las evaluaciones automatizadas. Adicionalmente, la expansión de la base de conocimiento RAG debe incluir información específica sobre procesos de transporte de contaminantes en ambientes hiperáridos, factores de biodisponibilidad en suelos salinos, y métodos de remediación adaptados a las condiciones de Antofagasta. La integración con la futura Norma Primaria de Calidad de Suelos de Chile, cuando sea publicada, requerirá actualización del contexto normativo en el sistema RAG.

CONCLUSIONES

Este trabajo ha presentado el desarrollo de un sistema prototipo que demuestra la viabilidad técnica de integrar modelos de lenguaje de gran escala con análisis geoquímico ambiental. A través de la implementación exitosa de RAG con una base de conocimiento estructurada, la aplicación de principios de ingeniería de prompts del framework DSPy, y el desarrollo de un mecanismo experimental de autoevaluación, se ha creado un sistema funcional capaz de procesar 94 muestras de datos geoquímicos reales y generar informes estructurados. Sin embargo, es importante reconocer que estos resultados representan una prueba de concepto que requiere validación adicional antes de su aplicación práctica en contextos operacionales.

La implementación exitosa del sistema RAG simulado demuestra que es posible inyectar conocimiento

especializado en LLMs sin reentrenamiento [50], manteniendo información crítica sobre comportamiento geoquímico de elementos, fuentes de contaminación y normativa aplicable disponible durante la generación de texto. La aplicación de principios DSPy resultó en una estructura más consistente de los informes generados, con plantillas modulares que aseguran la cobertura de aspectos críticos del análisis ambiental.

El procesamiento exitoso de 94 muestras reales de la Región de Antofagasta, incluyendo transformación de coordenadas, análisis multivariado y visualización interactiva, valida la capacidad técnica del sistema para manejar datos geoquímicos complejos. La identificación de asociaciones elementales mediante PCA y la detección de muestras anómalas demuestran la integración efectiva de técnicas analíticas tradicionales con capacidades de generación de lenguaje natural.

Sin embargo, es fundamental reconocer que este trabajo representa una etapa inicial de desarrollo. Se requiere validación sistemática mediante comparación con análisis realizados por expertos humanos, evaluación de la precisión técnica de las interpretaciones generadas, y estudios de usabilidad con profesionales del área ambiental. Las limitaciones identificadas definen una agenda clara para investigación futura. El trabajo pendiente incluye expansión de la base de conocimiento RAG con información más detallada sobre procesos ambientales específicos de zonas áridas, integración de la futura normativa chilena de calidad de suelos cuando sea publicada, implementación completa del framework DSPy para optimización automática de prompts, y validación del sistema LLM-as-a-judge mediante comparación con evaluaciones humanas expertas.

La implementación local mediante Ollama ha demostrado ser una solución viable para mantener la confidencialidad de datos sensibles mientras se aprovechan las capacidades de LLMs modernos. Esto es particularmente relevante para el sector minero-ambiental donde la información geoquímica puede tener implicaciones económicas significativas.

En síntesis, este trabajo establece las bases metodológicas y técnicas para el desarrollo de sistemas inteligentes de análisis geoquímico ambiental

basado en LLMs. Aunque se requiere trabajo adicional significativo antes de su aplicación práctica, el prototipo desarrollado demuestra el potencial de estas tecnologías para asistir en la evaluación ambiental de suelos, siempre como herramienta de apoyo al juicio experto humano, no como su reemplazo.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el proyecto ANID de la Subdirección de Investigación Aplicada FONDEF IT24I0030 “Plataforma inteligente para la gestión ambiental y monitoreo de suelos en la región de Antofagasta, mediante el modelamiento de línea base utilizando inteligencia artificial”.

REFERENCIAS

- [1] G. Müller, “Index of geoaccumulation in sediments of the Rhine River,” *GeoJournal*, vol. 2, no. 3, pp. 108-118, 1969.
- [2] L. Håkanson, “An ecological risk index for aquatic pollution control: a sedimentological approach,” *Water Research*, vol. 14, no. 8, pp. 975-1001, 1980, doi: 10.1016/0043-1354(80)90143-8.
- [3] B. Keith, E. Lam, I. Montofré, V. Zetola, J. Urrutia, C. Herrera, and J. Bech, “Evaluation of the geochemical background of soil in a hyper-arid zone using a multivariate statistical methodology: The case of the city of Antofagasta in the Atacama Desert,” *Chemosphere*, vol. 366, Art. no. 143472, 2024, doi: 10.1016/j.chemosphere.2024.143472
- [4] J. Tapia, S. Audry, B. Townley, and J. Duprey, “Geochemical background, baseline and origin of contaminants from sediments in the mining-impacted Altiplano and Eastern Cordillera of Oruro, Bolivia,” *Geochemistry: Exploration, Environment, Analysis*, vol. 12, no. 1, pp. 3-20, 2012, doi: 10.1144/1467-7873/10-RA-049.
- [5] R. Oyarzún, J. Oyarzún, J. Lillo, H. Maturana, and P. Higuera, “Mineral deposits and Cu-Zn-As dispersion-contamination in stream sediments from the semiarid Coquimbo Region, Chile,” *Environmental Geology*, vol. 53, no. 2, pp. 283-294, 2007, doi: 10.1007/s00254-007-0643-8.
- [6] T. Brown, *et al.*, “Language models are few-shot learners,” 2020, arXiv: 2005.14165.

- [7] Y. Chang *et al.*, "A survey on evaluation of large language models," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 3, Art. no. 39, pp. 1-45, 2024, doi: 10.1145/3641289.
- [8] Z. Ji *et al.*, "Survey of hallucination in natural language generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1-38, 2023, doi: 10.1145/3571730.
- [9] S. Bubeck *et al.*, "Sparks of artificial general intelligence: early experiments with GPT-4," 2023, arXiv: 2303.12712.
- [10] A. Vaswani *et al.*, "Attention is All You Need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017.
- [11] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI Technical Report, 2018. [Online]. Available: <https://openai.com/index/language-unsupervised/>
- [12] J. Achiam *et al.*, "GPT-4 Technical Report," 2023, arXiv: 2303.08774.
- [13] J. Wei *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," in *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, New Orleans, Louisiana, USA, 2022, pp. 24824-24837, doi: 10.52202/068431-1800.
- [14] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proceedings of the 34th Conference on Neural Information Processing Systems*, Vancouver, BC, Canada, Art. no. 793, 2020, pp. 9459-9474. [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/3495724.3496517>
- [15] Y. Gao *et al.*, "Retrieval-augmented generation for large language models: A survey," 2023, arXiv: 2312.10997.
- [16] S. Borgeaud *et al.*, "Improving language models by retrieving from trillions of tokens," 2022, arXiv:2112.04426v3.
- [17] O. Khattab *et al.*, "DSPy: Compiling declarative language model calls into self-improving pipelines," 2023, arXiv: 2310.03714.
- [18] J. Houston and A. Hartley, "The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert," *International Journal of Climatology*, vol. 23, no. 12, pp. 1453-1464, 2003, doi: 10.1002/joc.938.
- [19] J. Rech, J. Quade, and J. Betancourt, "Late Quaternary paleohydrology of the central Atacama Desert (lat 22°-24°S), Chile," *Geological Society of America Bulletin*, vol. 114, no. 3, pp. 334-348, 2002, doi: 10.1130/0016-7606(2002)114<0334:LQPOTC>2.0.CO;2.
- [20] S. Kojima, A. Astudillo, J. Rojo, D. Tristá, and H. Hayashi, "Ore mineralogy, fluid inclusion, and stable isotopic characteristics of stratiform copper deposits in the coastal cordillera of northern Chile," *Mineralium Deposita*, vol. 38, no. 2, pp. 208-216, 2003, doi: 10.1007/s00126-002-0304-5.
- [21] R. Zuo, Y. Xiong, J. Wang, and E. Carranza, "Deep learning and its application in geochemical mapping," *Earth-Science Reviews*, vol. 192, pp. 1-14, 2019, doi: 10.1016/j.earscirev.2019.02.023.
- [22] R. Xu, N. Baracaldo, and J. Joshi, "Privacy-preserving machine learning: Methods, challenges and directions," 2021, arXiv: 2108.04417.
- [23] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference (SCIPY 2010)*, Austin, Texas, USA, 2010, pp. 56-61, doi: 10.25080/Majora-92bf1922-00a.
- [24] Geopandas, "GeoPandas: Python tools for geographic data," GitHub, 2020. [Online]. Available: <https://github.com/geopandas/geopandas>
- [25] Ollama, "Ollama," GitHub, 2023. [Online]. Available: <https://github.com/ollama/ollama>
- [26] V. Agafonkin, "Leaflet - an open-source JavaScript library for interactive maps," leafletjs.com. Accessed: Oct. 16, 2025. [Online]. Available: <https://leafletjs.com>
- [27] ISO 18400-104:2018, "Soil quality – Sampling – Part 104: Strategies," *International Organization for Standardization*, Geneva, Switzerland, 2018. [Online]. Available: <https://www.iso.org/standard/70912.html>
- [28] U.S. Environmental Protection Agency, "Guidance on systematic planning using the data quality objectives process," EPA QA/G-4, Washington, DC, USA, 2006. [Online]. Available: <https://www.epa.gov/sites/default/files/2015-06/documents/g4-final.pdf>

- [29] M. Chen and L. Ma, "Comparison of three aqua regia digestion methods for twenty Florida soils," *Soil Science Society of America Journal*, vol. 65, no. 2, pp. 491-499, 2001, doi: 10.2136/sssaj2001.652491x.
- [30] J. Evenden, "Cartographic projection procedures for the UNIX environment: A user's manual," U.S. Geological Survey Open-File Report 90-284, 1990. [Online]. Available: <https://pubs.usgs.gov/of/1990/0284/report.pdf>
- [31] C. Reimann, P. Filzmoser, R. Garrett, and R. Dutter, *Statistical Data Analysis Explained: Applied Environmental Statistics with R*, Chichester, UK: John Wiley & Sons, 2008.
- [32] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [33] I. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.
- [34] C. Reimann and P. de Caritat, "Establishing geochemical background variation and threshold values for 59 elements in Australian surface soil," *Science of the Total Environment*, vol. 578, pp. 633-648, 2017, doi: 10.1016/j.scitotenv.2016.11.010.
- [35] A. Demetriades, C. Reimann, and P. Filzmoser, "Evaluation of GEMAS project quality control results," in *Chemistry of Europe's Agricultural Soils*, C. Reimann, M. Birke, A. Demetriades, P. Filzmoser, and P. O'Connor, Eds., Stuttgart, Germany: Schweizerbart, 2014, pp. 47-60.
- [36] D. Sparks, *Environmental Soil Chemistry*, 2nd ed. San Diego, USA: Academic Press, 2003.
- [37] G. Sposito, *The Chemistry of Soils*, 3rd ed. New York, USA: Oxford University Press, 2016.
- [38] *Guidelines for Drinking-water Quality*, 4th ed. World Health Organization, Geneva, Switzerland, 2017. [Online]. Available: <https://www.who.int/publications/i/item/9789241549950>
- [39] Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health, Canadian Council of Ministers of the Environment, Winnipeg, Canada, 2021.
- [40] *Decreto Supremo N°4/2009: Reglamento para el manejo de lodos generados en plantas de tratamiento de aguas servidas*, MINSEGPRES, Santiago, Chile, 2009. [En línea]. Disponible: <https://www.bcn.cl/leychile/navegar?idNorma=1007456>
- [41] O. Khattab *et al.*, "DSPy Assertions: Computational constraints for self-refining language model pipelines," 2023, *arXiv*: 2312.13382.
- [42] M. Suzgun and A. Kalai, "Meta-prompting: Enhancing language models with task-agnostic scaffolding," 2024, *arXiv*: 2401.12954.
- [43] L. Chiang and C. Lee, "Can Large Language Models Be an Alternative to Human Evaluations?," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds., Toronto, Canada, Jul. 9-14, 2023, pp. 15607-15631, doi: 10.18653/v1/2023.acl-long.870.
- [44] M. Grinberg, *Flask Web Development: Developing Web Applications with Python*, 2nd ed. Sebastopol, USA: O'Reilly, 2018.
- [45] Team Gemma and Google DeepMind, "Gemma: Open models based on Gemini research and technology," 2024, *arXiv*: 2403.08295v4.
- [46] H. Touvron *et al.*, "Llama 2: open foundation and fine-tuned chat models," 2023, *arXiv*: 2307.09288.
- [47] X. Wang *et al.*, "Self-Consistency Improves Chain of Thought Reasoning in Language Models," 2022, *arXiv*: 2203.11171.
- [48] L. Reynolds and K. McDonell, "Prompt programming for large language models: beyond the few-shot paradigm," 2021, *arXiv*: 2102.07350.
- [49] K. Smith and P. Malin, *Environmental Site Assessment: Phase I - A Basic Guide*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2014.
- [50] E. Perez, D. Kiela, and K. Cho, "True few-shot learning with language models," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, Eds., vol. 14, Red Hook, NY, USA: Curran Associates, 2021, pp. 11054-11070.