

Procesamiento de Lenguaje Natural Avanzado

Introducción



iimas

Dra. Helena Gómez Adorno
helena.gomez@iimas.unam.mx

Dr. Fazlourrahman Balouchzahi
fbalouc@iimas.unam.mx

Correo del curso:
pln.cienciadedatos@gmail.com



iimas

OBJETIVOS

General:

- Analizar, diseñar e implementar soluciones modernas de procesamiento del lenguaje natural basadas en modelos de lenguaje a gran escala, técnicas de prompting, recuperación aumentada con generación y agentes, partiendo de una revisión sintética de los fundamentos y enfatizando la evaluación crítica, las limitaciones y los aspectos éticos de estos enfoques.

Específicos:

- Consolidar de manera sintética los fundamentos del procesamiento del lenguaje natural necesarios para el estudio de enfoques modernos.
- Analizar la transición de representaciones clásicas del lenguaje hacia representaciones contextualizadas y modelos de lenguaje a gran escala.
- Diseñar y aplicar técnicas modernas de PLN, incluyendo prompting, recuperación aumentada con generación y agentes basados en modelos de lenguaje.
- Evaluar de forma crítica sistemas de PLN modernos mediante métricas, análisis de errores y consideraciones éticas.
- Desarrollar un proyecto aplicado o de investigación que integre de manera coherente técnicas avanzadas de procesamiento del lenguaje natural.

TEMAS



iimas

1. Revisión de fundamentos de PLN
2. Representaciones contextualizadas del lenguaje, Transformers y modelos preentrenados
3. Modelos de lenguaje a gran escala (LLMs)
4. Prompting y aprendizaje basado en instrucciones
5. Recuperación aumentada con generación (RAG)
6. Agentes basados en modelos de lenguaje
7. Evaluación, ética y aplicaciones avanzadas



iimas

RECURSOS VIRTUALES

- **Aula virtual**

<https://tuaulavirtual.educatic.unam.mx/course/edit.php?id=16134>

Contraseña:

- **ZOOM (contingencia)**

<https://cuaieed-unam.zoom.us/j/82575588095>

ID de reunión: 825 7558 8095

PROGRAMA



iimas

Semana	Fechas	Temas
1	3 y 5 feb	Teoría: Introducción general del curso: reglas, evaluación y proyecto. Revisión sintética de los fundamentos del PLN.
2	10 y 12 feb	Teoría: Limitaciones <u>de</u> embeddings estáticos y motivación <u>de</u> representaciones contextualizadas. Práctica: Uso de embeddings contextualizados (BERT-like) como extractores de características. Tema: Representaciones contextualizadas del lenguaje. Tarea: Comparación Word2Vec <u>vs</u> embeddings contextualizados en similitud semántica.
3	17 y 19 feb	Teoría: Fine-tuning vs feature extraction. Sesgos y limitaciones <u>de</u> representaciones contextualizadas. Práctica: Fine-tuning ligero para una tarea de clasificación/etiquetado (dataset pequeño). Tema: Representaciones contextualizadas: ajuste y análisis. Tarea: Reporte breve de resultados + análisis de errores.
4	24 y 26 feb	Teoría: Modelos de lenguaje a gran escala (LLMs): evolución, capacidades y límites (alucinaciones, costo, contexto). Práctica: <u>Exploración</u> guiada de comportamiento de un LLM (casos controlados). Tema: LLMs: fundamentos y comportamiento.

		Tarea: Ensayo técnico corto: "qué puede y qué no puede hacer un LLM" con ejemplos.
5	3 y 5 mar	<p>Teoría: Transformer (visión funcional) aplicado a LLMs: atención, contexto y escalamiento (sin matemáticas pesadas).</p> <p>Práctica: Uso de un modelo preentrenado en tareas <u>tipo</u>: clasificación, QA o resumen (pipeline).</p> <p>Tema: Transformers y modelos preentrenados.</p> <p>Tarea: Experimento reproducible con un modelo preentrenado <u>+</u> <u>discusión</u> <u>de</u> limitaciones.</p>
6	10 y 12 mar	<p>Teoría: Prompting: zero-shot, few-shot <u>e</u> instruction prompting. Estructura y buenas prácticas.</p> <p>Práctica: Diseño y evaluación de prompts para una tarea (p.ej., clasificación semántica/emocional).</p> <p>Tema: Prompting e instrucciones.</p> <p>Tarea: <u>Portafolio</u> de prompts con métricas y análisis de fallos.</p>
7	17 y 19 mar	<p>Teoría: Prompting avanzado: Chain-of-Thought (concepto), self-consistency y estrategias de robustez.</p> <p>Práctica: Comparación de variantes de prompting en el mismo problema (ablation simple).</p> <p>Tema: Prompting avanzado y razonamiento guiado.</p> <p>Tarea: Tabla comparativa de prompts + conclusiones (qué funcionó y por qué).</p>
8	24 y 26 mar	<p>Teoría: RAG: motivación, arquitectura y flujo (retrieval → context → generation).</p> <p>Práctica: Diseño de pipeline RAG (a nivel conceptual + prototipo base).</p> <p>Tema: Recuperación aumentada con generación (RAG).</p> <p>Actividad: Presentación y definición del proyecto (elección: LLM / Prompting / RAG / Agents).</p> <p>Tarea: Propuesta de proyecto (1–2 páginas) + dataset/fuente + criterios de evaluación.</p>
9	7 y 9 abr	<p>Teoría: Embeddings para retrieval, indexación y estrategias de búsqueda (top-k, reranking básico).</p> <p>Práctica: Implementación de RAG (indexación + recuperación + generación) con corpus pequeño/mediano.</p> <p>Tema: Implementación de sistemas RAG.</p>

		<p>Tarea: Demo funcional + reporte de diseño (decisiones y supuestos).</p> <p>Evaluación: <u>Primer examen parcial</u>.</p>
10	14 y 16 abr	<p>Teoría: Evaluación de LLMs/RAG: calidad, factualidad, cobertura, citación <u>y</u> análisis de errores.</p> <p>Práctica: Evaluación del sistema (métricas automáticas + evaluación manual con rúbrica).</p> <p>Tema: Evaluación y confiabilidad en LLM/RAG.</p> <p>Tarea: Informe de evaluación del sistema (con errores típicos).</p>
11	21 y 23 abr	<p>Teoría: Agentes con LLM: concepto, herramientas, planificación y memoria (visión general).</p> <p>Práctica: Diseño de un flujo agente-herramientas (tool calling / pipeline).</p> <p>Tema: Agentes basados en modelos de lenguaje.</p> <p>Tarea: Especificación de un agente (objetivo, herramientas, límites, criterios de éxito).</p>
13	5 y 7 may	<p>Teoría: Ética, sesgos, privacidad, seguridad y confiabilidad en sistemas <u>LLM/RAG/Agentes</u>.</p> <p>Práctica: Auditoría rápida de sesgos/errores + mitigaciones básicas (prompts, filtros, trazabilidad).</p> <p>Tema: Consideraciones éticas y sociales del PLN moderno.</p> <p>Tarea: <u>Checklist</u> de riesgos + plan de mitigación para el proyecto.</p>
14	12 y 14 may	<p>Práctica: Integración del proyecto final: mejoras (<u>prompting/RAG/agentes</u>) y preparación de resultados.</p> <p>Tema: Integración de métodos avanzados en un sistema completo.</p> <p>Tarea: Avance <u>de</u> proyecto (resultados preliminares <u>+ análisis</u>).</p>
15	19 y 21 may	<p>Actividad: Revisión y asesoría de proyectos finales (orientación a resultados y claridad experimental).</p> <p>Tarea: Versión casi final del reporte + <u>diapositivas</u> borrador.</p>
16	26 y 28 may	<p>Evaluación: <u>Segundo examen parcial</u>.</p>
17	2 y 4 jun	<p>Actividad: Presentación y entrega de proyectos finales.</p>

EVALUACIÓN



iimas

Actividad	Porcentaje
Exámenes Parciales	30%
Prácticas OBLIGATORIAS	40%
Proyecto OBLIGATORIO	30%
Total	100%
Examen Final	100%



iimas

BIBLIOGRAFÍA BÁSICA

- Practical Natural Language Processing. [[Vajjala et al., 2020](#)]
- Speech and Language Processing [[Jurafky & Martin](#)]
- Efficient Estimation of Word Representations in Vector Space [[Mikolov et al., 2013](#)]
- GloVe: Global Vectors for Word Representation [[Pennington et al., 2014](#)][[Sitio Oficial](#)]
- Enriching word vectors with subword information [[Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T., 2017](#)][[Sitio Oficial](#)]
- Attention Is All You Need [[Vaswani et al., 2017](#)]
- The Illustrated Transformer [[Jay Alammar](#)]
- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [[Devlin et al., 2019](#)]
- Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks [[Lewis et al., 2020](#)]