

Línea de TFM. NeuroSimb: Sistemas neurosimbólicos, integración de redes neuronales con modelos simbólicos, ontologías, grafos de conocimiento, reglas y otros modelos de representación del conocimiento explícito.

Descripción y justificación

Los sistemas basados en aprendizaje profundo, incluidos los grandes modelos de lenguaje y los modelos fundacionales, han alcanzado una gran capacidad para percibir, generalizar a partir de datos y generar contenido. Sin embargo, siguen teniendo limitaciones importantes: necesitan grandes volúmenes de datos, les cuesta razonar de forma sistemática y verificable, su comportamiento es poco interpretable, producen errores e invenciones difíciles de detectar y no incorporan de forma explícita el conocimiento experto de un dominio. Los sistemas simbólicos (ontologías, grafos de conocimiento, lógicas, reglas, modelos causales y probabilísticos estructurados) ofrecen justamente lo contrario: precisión en la representación y en la inferencia, conocimiento explícito, razonamiento trazable, capacidad de explicación y reutilización del conocimiento. A cambio, son frágiles ante datos ruidosos o no estructurados y costosos de construir y mantener.

Esta línea propone TFM cuyo objetivo es estudiar y desarrollar sistemas neurosimbólicos que combinen la potencia de aprendizaje y percepción de las redes neuronales con la precisión y la capacidad de razonamiento explícito de los modelos simbólicos, para conseguir sistemas inteligentes más fiables y explicativos. Esa fiabilidad se entiende en tres sentidos:

- **Mayor precisión y capacidad de razonamiento:** obtener resultados exactos y consistentes en problemas que exigen inferencia estructurada, composicionalidad o respeto de restricciones del dominio, y generalizar mejor fuera de la distribución de entrenamiento.
- **Mayor interpretabilidad y verificabilidad:** que las decisiones del sistema puedan explicarse en términos de conceptos y relaciones del dominio, contrastarse con el conocimiento experto y auditarse.
- **Menor requerimiento de datos y mejor mantenimiento:** aprovechar el conocimiento previo para aprender con menos ejemplos, y facilitar la actualización y reutilización del conocimiento sin reentrenar el sistema completo.

La combinación puede darse en ambas direcciones. El conocimiento simbólico puede guiar, restringir o enriquecer el aprendizaje neuronal. A su vez, las técnicas neuronales pueden servir para construir, completar, poblar o consultar estructuras simbólicas. Ambos enfoques se pueden integrar además en arquitecturas híbridas en las que cada componente asume la parte del problema para la que está mejor preparado.

Ámbitos de aplicación

- **Aprendizaje guiado por conocimiento:** incorporación de ontologías, reglas o restricciones lógicas en el entrenamiento o la inferencia de redes neuronales.
- **Grafos de conocimiento y aprendizaje de representaciones:** embeddings de grafos, redes neuronales sobre grafos, completado de enlaces y razonamiento sobre grafos.

- **Modelos de lenguaje y conocimiento estructurado:** generación aumentada con grafos de conocimiento (GraphRAG), anclaje de respuestas en ontologías, verificación de hechos y reducción de invenciones.
- **Construcción y mantenimiento de conocimiento:** extracción de entidades y relaciones, población y alineamiento de ontologías, y aprendizaje de ontologías a partir de texto con apoyo de modelos neuronales.
- **Percepción y razonamiento:** interpretación de escenas, imágenes o señales en términos de conceptos simbólicos, y razonamiento visual y espacial.
- **Agentes y planificación:** combinación de modelos de lenguaje con planificadores, sistemas de reglas o representaciones formales del entorno.
- **Dominios de aplicación:** salud y biomedicina, educación, ingeniería, ámbito jurídico, patrimonio y otros donde exista conocimiento experto formalizable y se requiera explicabilidad.

Técnicas de IA implicadas

Representación del conocimiento y ontologías (OWL, RDF, lógicas descriptivas); grafos de conocimiento y sus lenguajes de consulta; redes neuronales sobre grafos y embeddings de conocimiento; programación lógica diferenciable y lógicas continuas (Logic Tensor Networks, redes neuronales lógicas y enfoques afines); programación lógica probabilística y modelos gráficos; aprendizaje de reglas e inducción de programas; grandes modelos de lenguaje, recuperación aumentada y uso de herramientas; modelos causales; e IA explicable.

Posibles tipos de TFM

- Integración de una ontología o grafo de conocimiento de un dominio concreto con un modelo neuronal, y evaluación de la mejora frente al modelo puramente neuronal.
- Sistemas de pregunta-respuesta o generación aumentada con grafos de conocimiento, con análisis de fiabilidad y trazabilidad de las respuestas.
- Construcción semiautomática de ontologías o grafos de conocimiento a partir de textos mediante modelos de lenguaje, con validación experta.
- Incorporación de restricciones lógicas o reglas del dominio en el entrenamiento de redes neuronales para mejorar la coherencia de sus predicciones.
- Explicación de modelos neuronales en términos de conceptos de una ontología.
- Comparación experimental de arquitecturas neurosimbólicas en tareas de razonamiento o de aprendizaje con pocos datos.
- Revisiones sistemáticas del estado del arte en alguna de estas familias de técnicas.
- Diseño de marcos de evaluación del razonamiento, la consistencia y la explicabilidad de sistemas híbridos.

Aspectos metodológicos y éticos

Los trabajos podrán usar ontologías, grafos de conocimiento y conjuntos de datos públicos, o recursos disponibles en proyectos del grupo, respetando sus licencias y, en su caso, la protección de datos personales. Se prestará especial atención a la calidad y la cobertura del conocimiento simbólico empleado, a la definición de métricas que valoren no solo la precisión sino también la consistencia lógica, la explicabilidad y la robustez,

a la comparación con líneas base neuronales y simbólicas bien establecidas y a la reproducibilidad de los experimentos. En los trabajos que usen grandes modelos de lenguaje se cuidará la evaluación de la veracidad de sus salidas y la documentación de los recursos de cómputo empleados.

Perfil del estudiante

Interés tanto por el aprendizaje automático como por la representación del conocimiento y el razonamiento, y disposición para trabajar en la frontera entre ambos enfoques. Es recomendable tener conocimientos de aprendizaje profundo y, según el TFM, de lógica, ontologías y web semántica, grafos, representación del conocimiento o procesamiento del lenguaje natural.