

Plataforma de ejecución para chatbots multimodales

Director: Luis Fernando D'Haro

Correo electrónico: luisfernando.dharo@upm.es

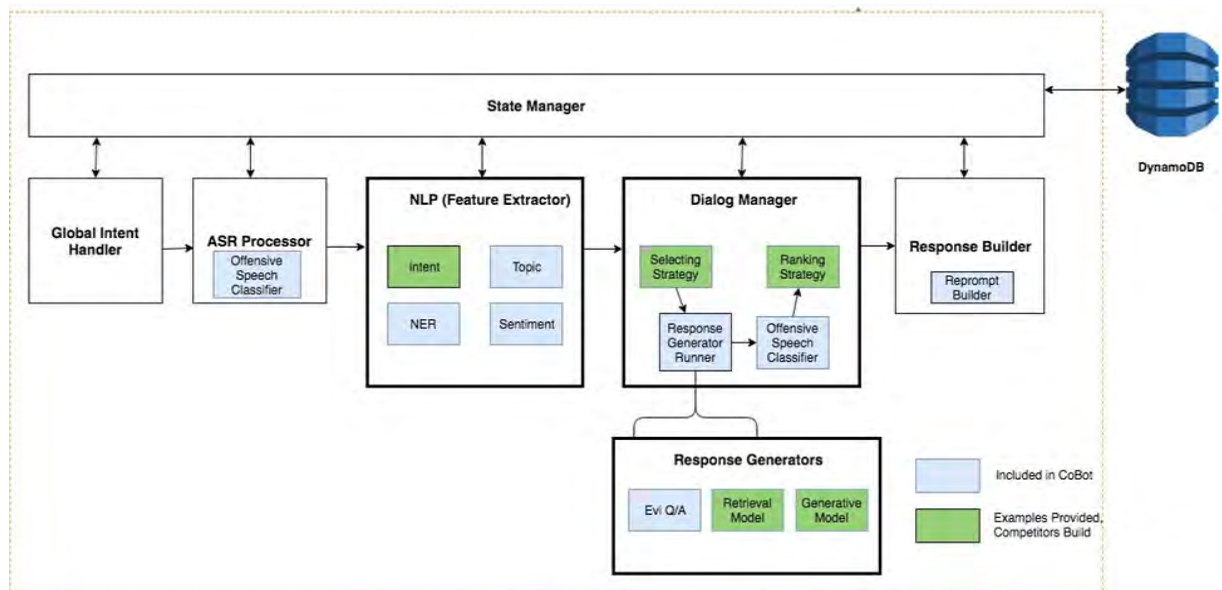
Despacho: B-108

Número de TFM's ofertados: 1 (cuatro horas diarias; posibilidad de trabajo a tiempo completo)

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: Los asistentes conversacionales son una de las tecnologías más utilizadas hoy en día en millones de hogares, sitios web y aplicaciones en dispositivos móviles. Agentes como Siri, Alexa, Cortana o Google proveen diferentes mecanismos de interacción tanto por voz, texto e imágenes.

En este proyecto, el alumno trabajará en desarrollar una plataforma software que permita la comunicación entre diferentes módulos (reconocedor y sintetizador de voz, chatbots de diferentes tipos, detector de eventos acústicos, etc), y el desarrollo del diálogo. El alumno partirá de un código base que permita la incorporación de nuevos módulos.



Condiciones de los candidatos: Buenas notas promedio (+7.0), conocimientos de programación en C++/Python, dominio de inglés, iniciativa e interés genuino por el tema propuesto. Se considerará un plus importante la experiencia en el uso de Streamlit o Gradio, Comunicaciones Asíncronas, bases de datos como MongoDB, Elasticsearch o similares. Para valorar al candidato enviar CV y listado de notas actualizados.

Chatbots emocionales y personalizados multimodales

Director: Luis Fernando D'Haro

Correo electrónico: luisfernando.dharo@upm.es

Despacho: B-108

Número de TFM's ofertados: 1

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: Los asistentes conversacionales son una de las tecnologías más utilizadas hoy en día en millones de hogares, sitios web y aplicaciones en dispositivos móviles. Agentes como Siri, Alexa, Cortana o Google proveen un innumerable tipo de servicios, desde aplicaciones específicas hasta entretenimiento. Sin embargo, estas aplicaciones aún requieren generar frases más emocionales y consistentes con características personalizadas tanto al usuario como al dominio.

En este proyecto, el alumno trabajará en fine-tunear un modelos de lenguaje multimodal de más de 7B de parámetros usando datos conversacionales y técnicas de adaptación rápidas (PEFT/LORA) y de cuantización (8/4 bits) para la rápida inferencia de respuestas. El modelo a utilizar será multimodal permitiendo la incorporación de imágenes en las conversaciones.



Condiciones de los candidatos: Buenas notas promedio (+7.0), conocimientos de programación en C++/Python, dominio de inglés, iniciativa e interés genuino por el tema propuesto. Se considerara un plus importante la experiencia con frameworks de desarrollo de aplicaciones conversacionales(HuggingFace, Pytorch, LangChain) y/o haber realizado cursos demostrables sobre Deep Learning o machine learning. Para valorar al candidato enviar CV y listado de notas actualizados.

Sistema de identificación automática de emociones en la voz

Director: Luis Fernando D'Haro

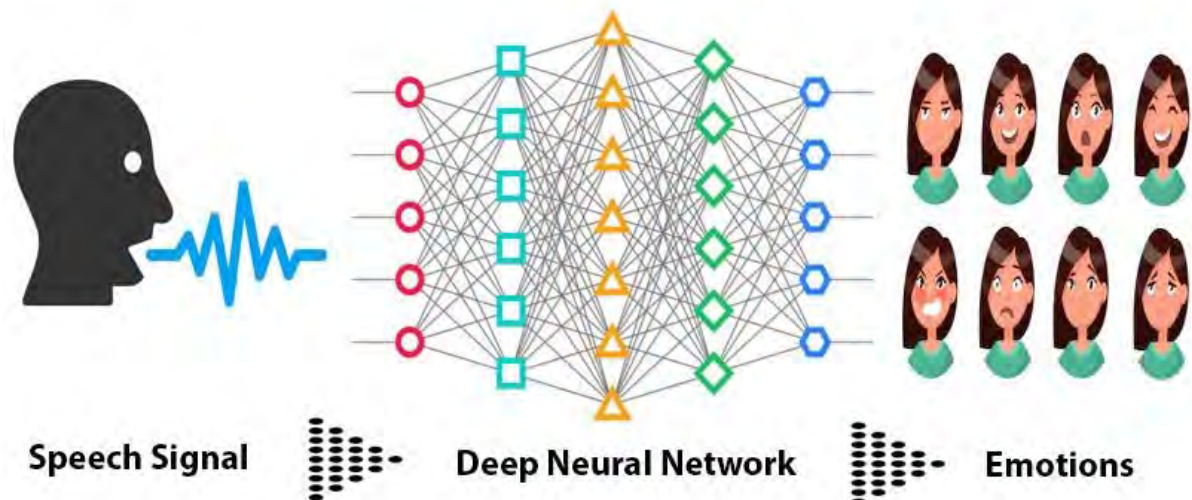
Correo electrónico: luisfernando.dharo@upm.es

Despacho: B-108

Número de TFM's ofertados: 1 (cuatro horas diarias; posibilidad de trabajo a tiempo completo)

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: El reconocimiento de emociones es el proceso de determinar diversas emociones en una conversación de audio grabada. Este proceso se hace mediante el modelado de la señal acústica, y opcionalmente del texto reconocido, usando un clasificador a la salida. Todo ello empleando redes neuronales profundas que modelan diversos aspectos de la voz y que permiten la fusión de las diversas fuentes de información. En este proyecto se partirá de un modelo pre-entrenado que deberá ser adaptado para permitir la incorporación de información textual y acústica.



Condiciones de los candidatos: Buenas notas promedio (+7.0), conocimientos de programación en C++/Python, dominio de inglés (escrito y leído), iniciativa e interés genuino por el tema propuesto. Se considerará un plus importante la experiencia con frameworks como PyTorch/Tensorflow y/o haber realizado cursos demostrables sobre Deep Learning o machine learning. Para valorar al candidato enviar CV y listado de notas actualizados.

Sistema de clasificación de toxicidad & sesgos

Director: Luis Fernando D'Haro

Correo electrónico: luisfernando.dharo@upm.es

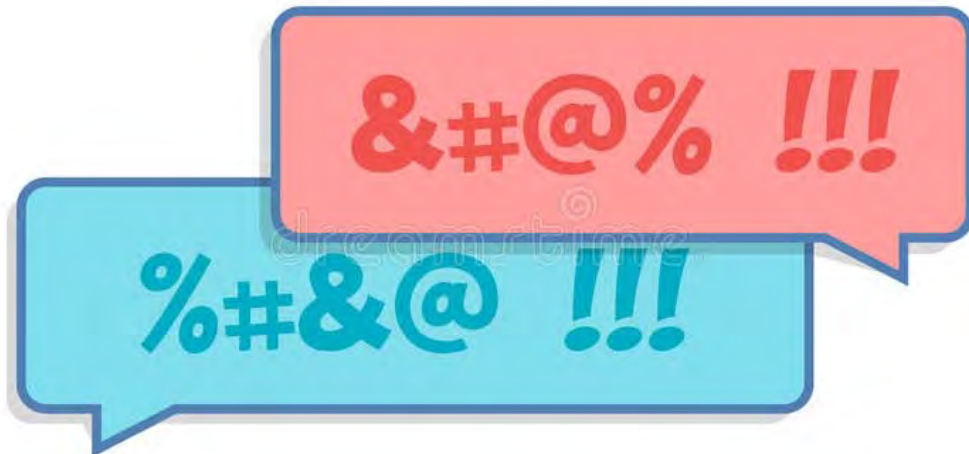
Despacho: B-108

Número de TFM's ofertados: 1 (cuatro horas diarias; posibilidad de trabajo a tiempo completo)

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: Uno de los problemas que deben evitar los sistemas conversacionales es detectar y evitar caer en la provocación causada por las frases dichas por usuarios tóxicos. Estos usuarios realizan comentarios sexistas, xenófobos, racistas, homofóbicos, entre otros, que deben poder ser detectados y evitados por el sistema.

El objetivo de este proyecto es entrenar y probar diferentes modelos de clasificación de toxicidad y sesgos, así como entrenar/finetunar un modelo que permita dar respuestas apropiadas ante este tipo de provocaciones. Para ello, se utilizarán bases de datos existentes y benchmarks para entrenar y probar las soluciones desarrolladas. Además, ayudará en la creación de un nuevo recurso para disponer de respuestas adecuadas.



Condiciones de los candidatos: Buenas notas, conocimientos de programación en C++/Python, dominio de inglés (escrito y leído), iniciativa e interés genuino por el tema propuesto. Se considerará un plus importante la experiencia con frameworks tales como Tensorflow/Keras/Pytorch o haber realizado cursos demostrables sobre Deep Learning o machine learning. Para valorar al candidato enviar CV y listado de notas actualizados.

BBDD multimodal para diálogo

Director: Luis Fernando D'Haro

Correo electrónico: luisfernando.dharo@upm.es







Despacho: B-108

Número de TFM's ofertados: 1 (cuatro horas diarias; posibilidad de trabajo a tiempo completo)

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: Actualmente existe una gran variedad de bases de datos textuales en el contexto de diálogo que permiten entrenar potentes sistemas como ChatGPT or GPT-4. Sin embargo, la posibilidad de interactuar con información multimodal del mismo tipo no está tan disponible.

El objetivo de este proyecto es crear una base de datos multimodal utilizando para ello datos textuales de diálogo y extendiéndolos para disponer de un recurso más completo. El alumno deberá utilizar diferentes modelos generativos y scripts que permitan generar el nuevo recurso paralelo. Posteriormente se buscará fine-tunar un modelo multimodal de red neuronal profunda utilizando el nuevo recurso generado.

 <p>Explain the advantages of this product.</p> <p>The audi e-tron quattro concept is a plug-in hybrid electric sports car that has a range of 310 miles and can accelerate from 0 to 60 mph in just 4 seconds.</p>	 <p>Tell me something about the history of this place.</p> <p>The great wall of china was built in 221 bc by emperor qin shihuang to protect the emperor's capital from invaders from the north.</p>	 <p>Write down the facts that you know about this flower.</p> <p>It is a genus of flowering plants in the orchid family, orchidaceae. It is native to eastern asia and south east asia. It is a perennial flowering plant that grows up to a height of 1 m.</p>
 <p>Is this photo unusual?</p> <p>Yes, it's a house that looks like it's upside down.</p> <p>How could someone get out of the house?</p> <p>It has a slide on the side of the house.</p>	 <p>What are shown in the photo?</p> <p>A man and a chicken.</p> <p>What does the man feel and why?</p> <p>He is scared of the chicken because it is flying at him.</p>	 <p>What are the ingredients I need to make this?</p> <p>Pepperoni, mozzarella cheese, pizza sauce, olive oil, salt, pepper, basil.</p> <p>What is the first step?</p> <p>Place the pizza dough on a baking sheet, brush with olive oil, sprinkle with salt, pepper, and basil.</p>

Condiciones de los candidatos: Buenas notas, conocimientos de programación en C++/Python, dominio de inglés (escrito y leído), iniciativa e interés genuino por el tema propuesto. Se considerará un plus importante la experiencia con frameworks tales como Tensorflow/Keras/Pytorch o haber realizado cursos demostrables sobre Deep Learning o machine learning. Para valorar al candidato enviar CV y listado de notas actualizados.

Chatbots intuitivos y generalizables

Tutor: Marcos Estecha Garitagoitia

Ponente: Luis Fernando D'Haro

Correo electrónico: marcos.estecha.garitagoitia@upm.es (luisfernando.dharo@upm.es)

Despacho: B-108

Número de TFM's ofertados: 1 (tres a cinco horas diarias)

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: Quizás hayas escuchado de chatbots tan inteligentes que parece que tuvieran consciencia (ej. el caso de LaMDA¹, un chatbot desarrollado por Google que según uno de sus ingenieros parecía que tenía consciencia). Sin embargo, por muy sofisticados que sean, este tipo de chatbots aún encuentran difícil el poder mantener conversaciones en el que lleven la contraria a los usuarios o provean información para la cual no cuentan con una referencia concreta ("perfil de persona") que deban usar.

Los objetivos de este proyecto es abordar el problema de las personas-no-predefinidas (del término en inglés "*Out-of-predefined (OOP) personas*"), el cual sucede cuando el chatbot tiene que responder a una consulta que requiera de información que no esté incluida entre las descripciones de su "*persona*" previamente definidas. Así mismo, el desarrollar un mecanismo que permita controlar la generación de frases que puedan llevar la contraria al usuario (ej. el usuario es hincha de un equipo de futbol, en tanto que el chatbot lo es del equipo rival). Para este fin se estudiará implementar un sistema neuronal capaz de predecir nuevas respuestas automáticas y consistentes condicionadas a perfiles de usuario más dinámicos e incluso contrarios a las expectativas del usuario.



Condiciones de los candidatos: Buenas notas, conocimientos de programación en Python, dominio de inglés, iniciativa e interés genuino por el tema propuesto. Para valorar al candidato enviar CV y listado de notas actualizados. Se considerará un plus importante el conocer frameworks tales como Tensorflow/Keras/Pytorch, haber realizado cursos demostrables sobre Deep Learning, Natural Language Processing o machine learning.

¹ <https://theconversation.com/is-googles-lamda-conscious-a-philosophers-view-184987>

Detector de información personal en sistemas de diálogo

Director: Luis Fernando D'Haro

Correo electrónico: luisfernando.dharo@upm.es

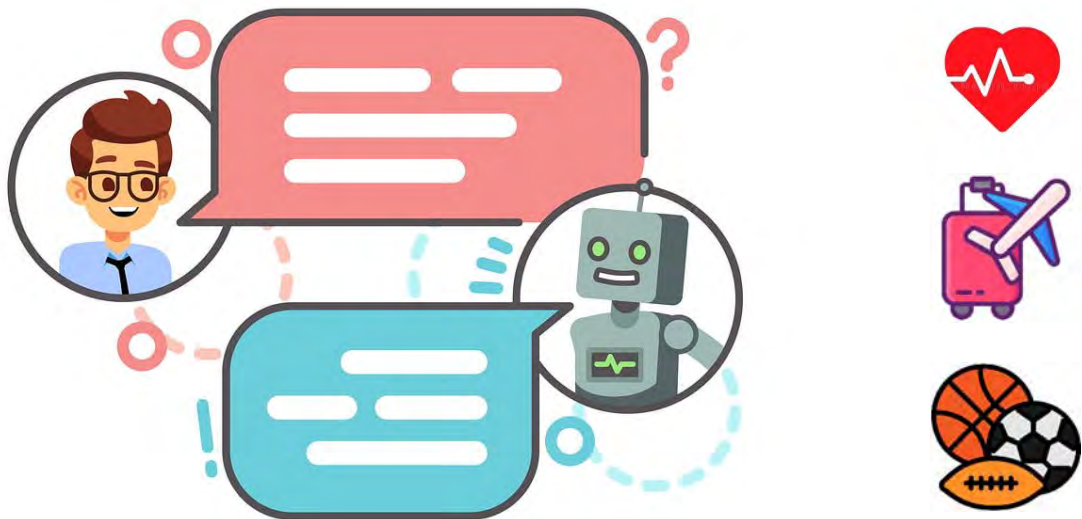
Despacho: B-108

Número de TFM's ofertados: 1 (tres a cinco horas diarias)

Competencias relacionadas: Innovación tecnológica, diseño de sistemas inteligentes (Titulaciones: MUIT/MUISE/MUTSC/MUIRST).

Descripción del TFM: Los agentes conversacionales o chatbots pueden ser de gran utilidad a la hora de hablar de temas tan diversos como el entretenimiento, recomendaciones turísticas o incluso la salud. Sin embargo, durante estas conversaciones los usuarios suelen compartir una gran variedad de información personal explícita o implícita que va desde gustos ("me gusta el helado"), relaciones personales ("mi hermana es Ingeniera"), ocupaciones ("hablé con el director de la orquesta para que me pase la partitura"), rasgos de carácter ("trabajé duro para entregar mi tarea"), entre otros.

Este proyecto tiene como objetivo desarrollar un sistema que permita automáticamente detectar este tipo de información a lo largo de un diálogo. Se hará uso de diversas bbdd de diálogo con y sin este tipo de información, además de modelos neuronales pre-entrenados.



Condiciones de los candidatos: Muy buenas notas promedio (+7.5), conocimientos de programación en C++/Python, dominio del inglés (escrito y leído), iniciativa e interés genuino por el tema propuesto. Se considerará un plus importante la experiencia con frameworks tales como Pytorch, TensorFlow, HuggingFace o haber realizado cursos demostrables sobre Deep Learning o machine learning.

Estudio de herramientas para la separación de un audio en habla/no habla y en diferentes voces/hablantes

Director: José Manuel Pardo Muñoz

Correo electrónico: josemanuel.pardom@upm.es

Despacho: C-224

Número de TFM's ofertados: 1

Competencias Relacionadas: Aprendizaje automático, Tecnología del Habla, Interpretación de audios, reconocimiento de locutores. (Titulaciones: MUIT/MUISE)

Descripción del TFM: Para procesar grabaciones de audio de reuniones o telefónicas y extraer la información contenida en las mismas, es necesaria una serie de pasos en el procesamiento.

- El primero es separar en el audio los contenidos de habla de otros contenidos como pueden ser ruidos de fondo, música etc. Este paso no está exento de dificultad pues los sonidos que pueden acompañar a las grabaciones pueden ser muy variados y difíciles de caracterizar.
- El segundo paso es separar el habla temporalmente en las distintas voces que intervienen.
- Finalmente hay que aplicar un reconocedor de habla para transcribir el contenido de la grabación y poder asignar a cada locutor su intervención, almacenarla y acceder a la misma de forma ágil.



Se dispone de herramientas, tanto internas como públicas para realizar las fases una y dos. Estas herramientas han sido evaluadas de forma preliminar con algunos experimentos y grabaciones limitadas.

El objetivo primero y principal de este Trabajo fin de Máster es un estudio en profundidad de la tecnología de las distintas herramientas y una evaluación rigurosa de las mismas utilizando bases de datos estandar o privadas.

Condiciones de los candidatos: Conocimiento de lenguaje python, buen conocimiento de inglés, iniciativa e interés genuino por el tema propuesto. Se valorará los conocimientos de procesado de señal.

Sistema de identificación de idioma y eventos acústicos por voz

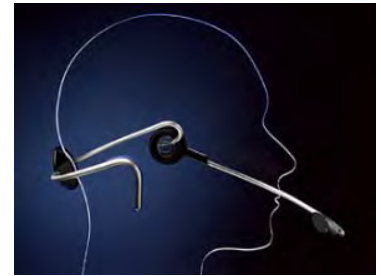
Director: Ricardo de Córdoba Herralde

Correo electrónico: cordoba@die.upm.es

Despacho: B-108

Número de TFG ofertados: 1

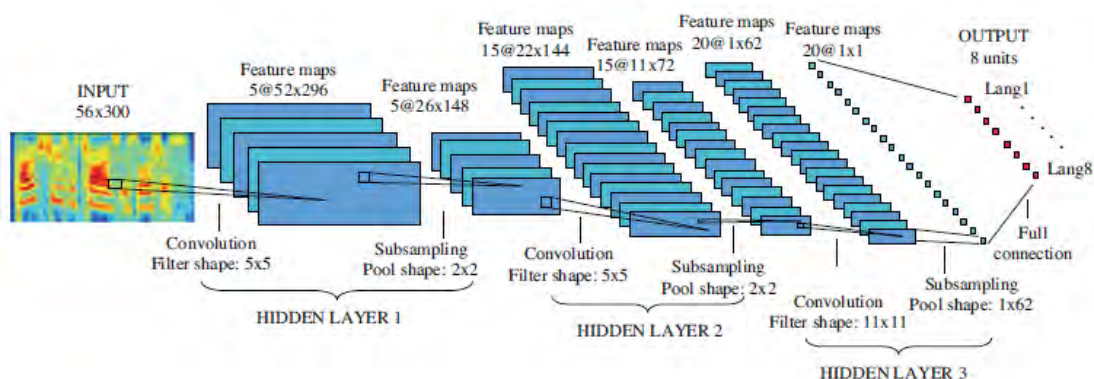
Descripción del TFG: No es un secreto que vivimos en un mundo cada vez más globalizado, en el que personas de diferentes partes del mundo hablando idiomas muy distintos se comunican entre sí. Los sistemas de reconocimiento de idioma sirven como un paso fundamental para la realización de tareas más complejas como son hablar con un agente que hable el mismo idioma, un traductor de voz automático, o un sistema de etiquetado automático de vídeo.



El objetivo primero y principal de este Trabajo Fin de Grado (TFG) es mejorar uno de los sistemas de identificación multilingüe más avanzados que hay actualmente a nivel mundial. Para ello el estudiante aprenderá los conceptos más importantes de este tipo de tecnologías de voz en la que también trabajan empresas como Google o Microsoft, o universidades como MIT y Berkeley. Así mismo, profundizará en el conocimiento de herramientas como redes neuronales profundas (DNNs), gestión de grandes cantidades de datos, los algoritmos basados en "big data", utilización de GPUs para la aceleración de la ejecución, soluciones de código abierto, etc. Así mismo, se va a trabajar con datos reales de una conocida empresa española para el desarrollo de un sistema.

Este objetivo se compone, a su vez, de los siguientes subobjetivos diferenciados:

- Utilización de una de las herramientas de código abierto más potentes disponibles, llamada Kaldi, para la utilización de DNNs para el reconocimiento de idioma. Las redes neuronales han revolucionado el mundo actual del procesamiento de datos, siendo de aplicación en múltiples tareas, desde la voz, al procesamiento de textos, big data, e infinidad de tareas. En concreto, se utilizará la tecnología puntera de x-vectors.
- Aplicación de estas mismas tecnologías a la identificación de eventos acústicos presentes en la señal de habla, para poder utilizarlos en sistemas de interacción como chatbots.
- Mejoras de los scripts o guiones de ejecución de los experimentos en GPUs.
- Utilización de datos de una empresa española de tecnologías de voz para desarrollar un nuevo sistema.



Desarrollo de sistemas de algoritmos de aprendizaje profundo para la generación de lenguaje de signos

Tutor: Rubén San Segundo Hernández

Correo Electrónico: ruben.sansegundo@upm.es

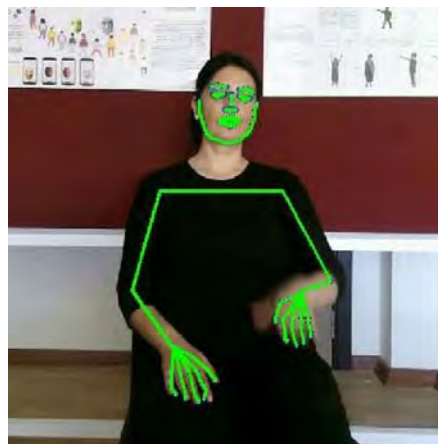
Despacho: B-109

Número de TFM's ofertados: 1

Competencias relacionadas: Programación en Python, Aprendizaje Automático e Inteligencia Artificial.

Descripción del TFM: El uso de avatares para la representación de la lengua de signos ofrece una flexibilidad importante porque evita la necesidad de grabar a una persona signando cada contenido específico. Pero, por otro lado, controlar un avatar puede resultar complicado porque es necesario controlar todos sus elementos (esqueleto y malla). Esta dificultad puede producir movimientos con poca naturalidad. La naturalidad es, sin duda, una de las principales limitaciones de este tipo de sistemas.

En este sentido, se propone generar información de movimiento en lengua de signos (en 2D) a partir de características de alto nivel de los signos (forma de la mano, orientación, localización de la mano, etc.). El objetivo técnico principal sería desarrollar un algoritmo de aprendizaje profundo capaz de asociar características de signos de alto nivel con el movimiento del esqueleto. Una limitación importante es la disponibilidad de suficientes datos para desarrollar y entrenar un buen sistema DL. Este objetivo de investigación está planificado en dos fases: Fase 1: Generar un buen conjunto de datos que contenga un número relevante de descripciones de señales con información de movimiento de varias representaciones de la misma señal. Fase 2: desarrollo de un algoritmo de aprendizaje profundo capaz de asociar características de signos de alto nivel con el movimiento del esqueleto.



Condiciones de los candidatos: Interés en aspectos de sensado humano e interacción persona-máquina. Vocación por el desarrollo de tecnologías para ayudar a las personas.

Supervisión de pacientes con enfermedades neurodegenerativas mediante sensores de movimiento

Tutor: Rubén San Segundo Hernández

Correo Electrónico: ruben.sansegundo@upm.es

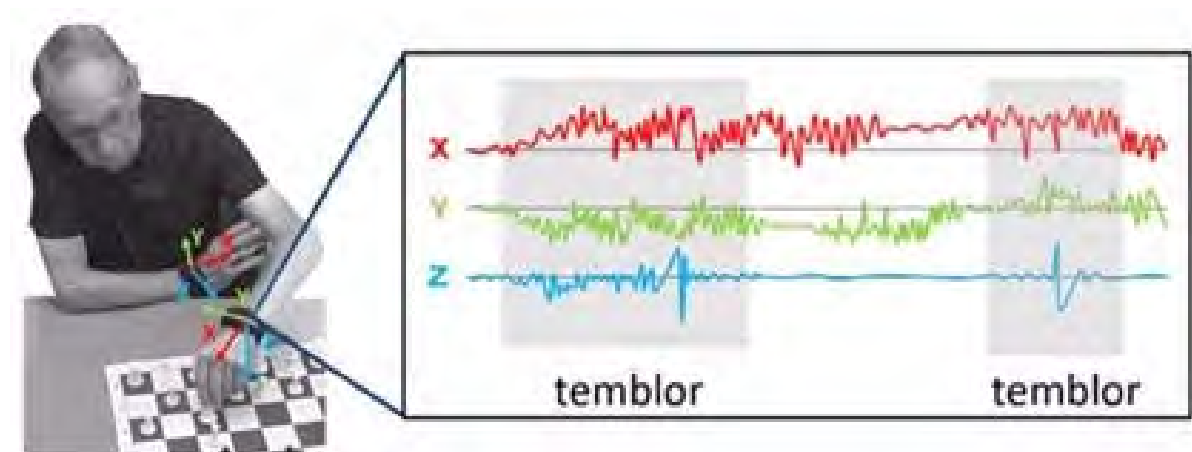
Despacho: B-109

Número de TFM's ofertados: 1

Competencias Relacionadas: Programación en Python, Aprendizaje Automático e Inteligencia Artificial.

Descripción del TFM: Los pacientes de enfermedades neurodegenerativas presentan una importante variedad en el tipo e intensidad de los síntomas de dicha enfermedad. Esta variedad supone un reto para los médicos que deben detectar los casos de dicha enfermedad, y posteriormente, supervisar la evolución de la enfermedad para ajustar la medicación necesaria.

En este trabajo se propone el estudio de técnicas de Deep Learning para el procesamiento de señales inerciales recogidas de varios sensores (acelerómetros, giróscopos, etc.) localizados en la ropa, con el fin de detectar los síntomas de la enfermedad y supervisar la evolución de la dicha enfermedad. Con este TFM se pretende ofrecer medidas objetivas que ayuden a los médicos en su diagnóstico.



Condiciones de los candidatos: Interés en aspectos de sensado humano e interacción persona-máquina. Vocación por el desarrollo de tecnologías para ayudar a las personas.

Modelado de comportamiento humano mediante el procesamiento de señales obtenidas mediante sensores de movimiento

Tutor: Rubén San Segundo Hernández

Correo Electrónico: ruben.sansegundo@upm.es

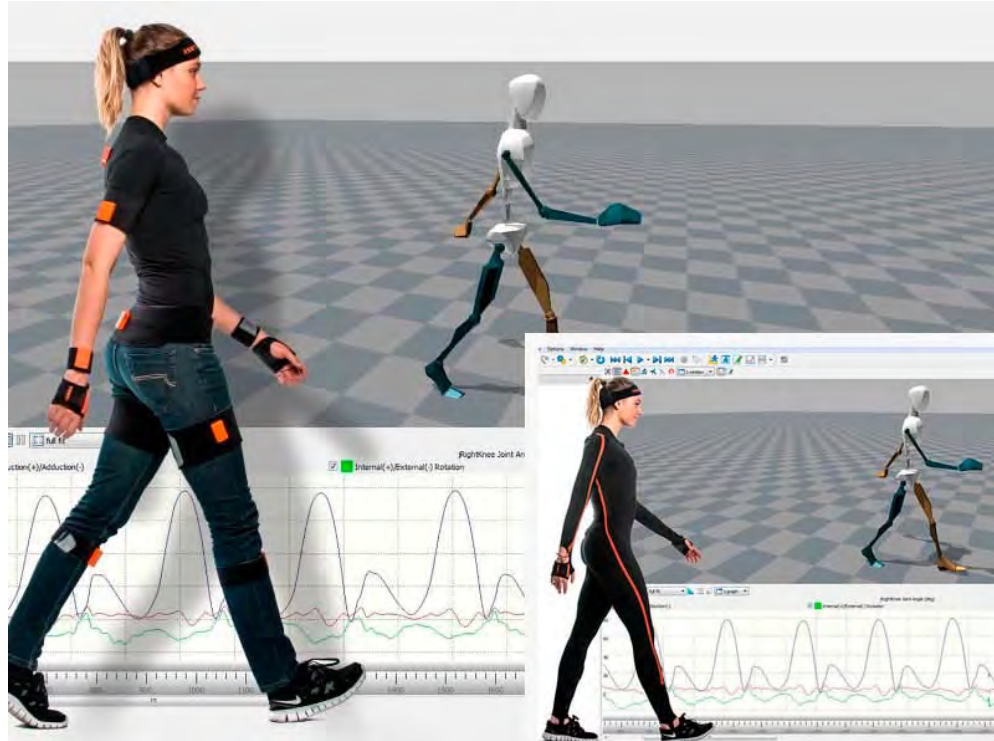
Despacho: B-109

Número de TFM's ofertados: 1

Competencias relacionadas: Programación en Python, Aprendizaje Automático e Inteligencia Artificial.

Descripción del TFM: En la actualidad el desarrollo tecnológico de los sensores está permitiendo su incorporación en objetos de la vida cotidiana como ropa, teléfonos, relojes, etc. Este alto nivel de integración está permitiendo el desarrollo de aplicaciones de supervisión y modelado del comportamiento humano.

En este trabajo se propone el estudio de técnicas de *Deep Learning* para el procesamiento de señales inerciales recogidas de varios sensores (acelerómetros, giróscopos, etc.) localizados en objetos de la vida cotidiana. Mediante el estudio de las señales de los sensores es posible detectar el comportamiento de una persona, su identidad o sus intenciones.



Condiciones de los candidatos: Interés en aspectos de sensado humano e interacción persona-máquina. Vocación por el desarrollo de tecnologías para ayudar a las personas.

Desarrollo de servicios y sistemas con inteligencia artificial

Director: Rubén San Segundo Hernández

Correo Electrónico: ruben.sansegundo@upm.es

Despacho: B-109

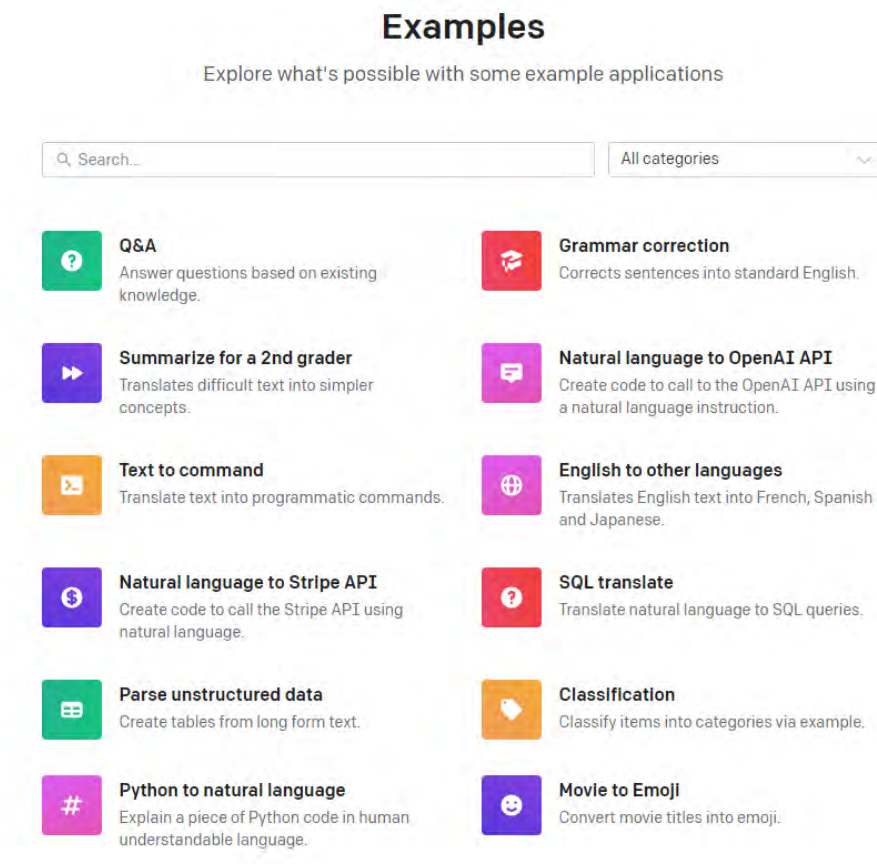
Número de TFM's ofertados: 1

Competencias Relacionadas: Programación en Python, Aprendizaje Automático e Inteligencia Artificial.

Descripción del TFM: Actualmente, están habiendo una gran cantidad de iniciativas para el desarrollo de tecnologías que permitan desarrollar sistemas y servicios con inteligencia artificial. Estas tecnologías están abriendo una gran variedad de posibles aplicaciones.

El objetivo de este trabajo es analizar las tecnologías disponibles (como por ejemplo las desarrolladas por OpenAI), evaluar sus posibilidades y proponer servicios basados en estas tecnologías que permitan mejorar la vida de las personas.

<https://platform.openai.com/examples>



Condiciones de los candidatos: Interés en aspectos de inteligencia artificial. Vocación por desarrollar sistemas novedosos.