



Anexo 1. Formato de protocolo de investigación

1. Datos generales de la propuesta

- **Nombre de la persona proponente:** Daniel Alejandro Cervantes Cabrera
- **Título de la propuesta:** Cómputo de Alto Rendimiento para Aprendizaje Profundo y sus Aplicaciones
- **Línea de Generación y Aplicación del Conocimiento (LGAC) de INFOTEC en la que incide la propuesta:** Inteligencia computacional en la Ciencia de Datos
- **Periodo de ejecución:** Del 01/04/2022 al 30/03/2024

2. Descripción de la propuesta:

- **Resumen (ejecutivo):** Esta propuesta de investigación tiene el objetivo de realizar investigación aplicada en el área de Aprendizaje Profundo para la definición de nuevos funcionales de costo a través de un proceso de minimización o “entrenamiento” de Redes Neuronales, que permitan realizar “predicciones” en diferentes problemática de interés, como por ejemplo aproximación de campos vectoriales sujetos a condiciones de físicas y de frontera (como velocidades de viento), detección de intentos de intrusión en sistemas de cómputo en particular para sistemas de IOT, manejo de correo no deseado o también llamado SPAM, en sistemas de correo institucionales, etc.. Para lo anterior además se propone realizar un desarrollo de una biblioteca computacional optimizada Orientada a Objetos en donde se implementen éstos métodos la cual será de acceso libre bajo una licencia de software LGPL.
- **Antecedentes:** El Aprendizaje Profundo es una área de la Inteligencia Artificial que en los últimos años ha atraído la atención no sólo de la comunidad de las ciencias exactas sino también de las ciencias sociales como por ejemplo Economía, Arqueología o Derecho etc. Esto se debe principalmente a su amplio rango de aplicación por su flexibilidad y robustez, ya que con estos métodos es posible la construcción de las llamadas Redes Neuronales Computacionales, las



Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental

cuales permiten realizar aproximación y/o clasificación de un gran y variedad de tipos de información. Otra característica importante, es que los algoritmos involucrados en estas metodologías son altamente paralelizables, permitiendo así utilizar tecnologías modernas de cómputo de alto rendimiento como los llamados GPUs y así poder abordar problemas a gran escala.

- **Justificación:** Desarrollo de investigación en el área de Aprendizaje Profundo para su aplicación en la solución de problemas de interés para INFOTEC como ejemplo, dispersión de contaminantes, Seguridad en Cómputo, Detección Automática de SPAM, etc.
- **Objetivo general:** Investigación de nuevas formulaciones de funcionales de costo en Aprendizaje Profundo, así como su implantación para la solución de problemas de aproximación y/o clasificación en diferentes áreas de aplicación.
- **Objetivos específicos:**
 - Desarrollo de investigación básica que permita definir nuevos funcionales de costo para entrenamiento de Redes Neuronales Computacionales.
 - Implantación de biblioteca computacional Orientada a Objetos y Optimizada con GPUs de los funcionales establecidos en la investigación.
 - Aplicación de los puntos anteriores para la solución de problemas particulares como:
 - Predicción de dispersión de contaminantes en proyectos de colaboración con la Dra. Magali Arellano INFOTEC.
 - Procesamiento de Lenguaje Natural para detección de SPAM en correo institucional.
 - Seguridad en cómputo para proyecto de investigación doctoral del Mtro. Alfredo Munguia INFOTEC.
- **Metas:**
 - a. Publicación de al menos un artículo de investigación en revista de arbitraje internacional JCR.
 - b. Liberación de biblioteca computacional con licencia LGPL de acceso libre con los métodos implementados.



c. Uso de biblioteca computacional en proyectos de aplicación de desarrollo tecnológico de interés para el INFOTEC.

● **Metodología:**

- a. Definición de nuevos funcionales de costo para la minimización de modelos de aproximación o clasificación de datos a través de la incorporación de términos relevantes al problema abordado, i.e. irrotacionalidad del campo vectorial aproximado etc..
- b. Preprocesamiento, vectorización y etiquetado de información de entrenamiento de Red Neuronal Computacional.
- c. Implantación Orientada a Objetos para GPUs de métodos de entrenamiento de Red Neuronal Computacional.
- d. Predicción de aproximación y/o clasificación.
- e. Validación.

● **Beneficios esperados:**

- a. Aplicación de metodologías de Aprendizaje Profundo en problemáticas de interés de INFOTEC, como en proyecto de predicción de dispersión de contaminantes, seguridad en cómputo, detección automática de SPAM.
- b. Liberación de biblioteca computacional optimizada de acceso libre con licencia LGPL.

● **Resultados esperados:**

- a. Publicación de al menos un artículo de investigación en revista internacional JCR.
- b. Aplicación de biblioteca computacional en la solución de problemas de interés para INFOTEC como son:
 - Aproximación de velocidades de viento de la CDMX para modelos de predicción de dispersión de contaminantes.
 - Detección automática de intentos de intrusión en desarrollados basados en IOT.
 - Detección de SPAM en correos institucionales de gobierno.



3. Plan de actividades

- **Descripción de las actividades**

Etapa núm.:		
#	Actividad	Descripción de la actividad
1	Definición de nuevos funcionales de costo.	Se establecerán los términos correspondientes para la definición de funcionales de costo de acuerdo al problema que se esté abordando.
2	Análisis y diseño de biblioteca computacional.	Definición de arquitectura de software a través de análisis y diseño de módulo requeridos para su implantación.
3	Implantación de biblioteca computacional.	Desarrollo de biblioteca computacional Orientada a Objetos con el lenguaje de programación Python
4	Aplicación I de biblioteca computacional.	Se abordarán problemas de aproximación de campos vectoriales en particular de velocidades de viento enfocado en problemas de dispersión de contaminantes.
5	Aplicación II de biblioteca computacional.	Se abordarán problemas para proyectos en seguridad en cómputo de sistemas de IOT. Del proyecto de investigación doctoral del Mtr. Alfredo Munguia.
6	Aplicación III de biblioteca computacional.	Se abordarán problemas para proyectos de detección automática de SPAM en correo institucional.

- **Descripción de las metas**

Etapa núm.:		
#	Actividad	Meta
1	Definición de nuevos funcionales de costo.	Publicación de al menos un artículo de investigación en revista internacional JCR.
2	Análisis y diseño de biblioteca computacional.	Generación de documentación técnica de biblioteca computacional.
3	Implantación de biblioteca computacional.	Liberación de biblioteca en repositorio institucional con acceso libre con licencia LGPL.



**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

4	Aplicación I de biblioteca computacional.	Solución de problemas en el área de aproximación de campos vectoriales enfocado a velocidades de viento, para problemas de dispersión de contaminantes.
5	Aplicación II de biblioteca computacional	Solución de problemas en el área de seguridad en cómputo enfocado en sistemas de IOT, como parte del trabajo de investigación doctoral del Mtr. Alfredo Munguia.
6	Aplicación III de biblioteca computacional.	Solución de problemas en el área de detección automática de SPAM en correo institucional.

● **Productos (entregables)**

Etapa núm.:		
#	Actividad	Producto
1	Definición de nuevos funcionales de costo.	Al menos un artículo de investigación en revista internacional JCR.
2	Análisis y diseño de biblioteca computacional.	Documentación técnica y manual de usuario.
3	Implantación de biblioteca computacional.	Código fuente de biblioteca computacional.
4	Aplicación I de biblioteca computacional.	Código fuente de implantación de solución en el problema de aproximación de velocidades de viento.
5	Aplicación II de biblioteca computacional	Código fuente de implantación de solución en seguridad de cómputo para sistemas IOT.
6	Aplicación III de biblioteca computacional.	Código fuente de implantación de solución en detección automática de SPAM.

● **Especificaciones de los productos**

Etapa núm.:		
#	Actividad	Especificación del producto (entregable)
2	Definición de nuevos funcionales de costo.	Artículo de investigación en revista internacional JCR.



	Análisis y diseño de biblioteca computacional.	Documentación técnica de biblioteca y manual de usuario.
	Implantación de biblioteca computacional.	Código fuente de biblioteca computacional en repositorio institucional de acceso libre con licencia LGPL.
	Aplicación I de biblioteca computacional.	Reporte técnico de implantación de solución de aplicación I.
	Aplicación II de biblioteca computacional	Reporte técnico de implantación de solución de aplicación II.
	Aplicación III de biblioteca computacional.	Reporte técnico de implantación de solución de aplicación III.

• **Mecanismos de transferencia**

Etapa núm.:		
#	Actividad	Mecanismo de transferencia
2	Definición de nuevos funcionales de costo.	Publicación en revista internacional arbitrada.
	Análisis y diseño de biblioteca computacional.	Documentación técnica y manuales de usuario.
	Implantación de biblioteca computacional.	Reporte técnico y tutorial de uso.
	Aplicación I de biblioteca computacional.	Reporte técnico y manual de uso.
	Aplicación II de biblioteca computacional	Reporte técnico y manual de uso.
	Aplicación III de biblioteca computacional.	Reporte técnico y manual de uso.

4. Cronograma de actividades

Etapa núm.:				
#	Actividad	Productos (entregables)	Fecha de inicio	Duración (núm. de semanas)
2	Definición de nuevos	Artículo de investigación en	1 de Abril 2022	52





	funcionales de costo.	revista internacional JCR.		
	Análisis y diseño de biblioteca computacional.	Documentación técnica de biblioteca y manual de usuario.	1 de Abril 2022	52
	Implantación de biblioteca computacional.	Código fuente de biblioteca computacional en repositorio institucional de acceso libre con licencia LGPL.	1 de Abril 2022	52
	Aplicación I de biblioteca computacional.	Reporte técnico de implantación de solución de aplicación I.	1 de Abril 2023	18
	Aplicación II de biblioteca computacional	Reporte técnico de implantación de solución de aplicación II.	1 de Agosto 2023	17
	Aplicación III de biblioteca computacional.	Reporte técnico de implantación de solución de aplicación III.	1 de Enero 2023	17

5. Referencias

1. Cervantes, D.A., González Casanova, P., Gout, C. et al. A line search algorithm for wind field adjustment with incomplete data and RBF approximation. *Comp. Appl. Math.* **37**, 2519–2532 (2018). <https://doi.org/10.1007/s40314-017-0461-x>
2. Daniel Cervantes, Miguel Angel Moreles, Joaquin Peña, Alonso Ramirez-Manzanares. A computational method for the covariance matrix associated with extracellular diffusivity on disordered models of cylindrical brain axons[J]. *Mathematical Biosciences and Engineering*, 2021, 18(5): 4961-4970. doi: 10.3934/mbe.2021252





GOBIERNO DE
MÉXICO



Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental

3. E, W., Yu, B. The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems. *Commun. Math. Stat.* **6**, 1–12 (2018). <https://doi.org/10.1007/s40304-018-0127-z>
4. Deep Learning-Based Security Behavior Analysis in IoT Environments: A Survey, Security and Communication Networks Volume 2021, Article ID 8873195, 13 pages <https://doi.org/10.1155/2021/8873195>
5. Isra'a AbdulNabi, Qussai Yaseen, Spam Email Detection Using Deep Learning Techniques, Procedia Computer Science, Volume 184, 2021, Pages 853-858, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2021.03.107>.

ATENTAMENTE

Dr. Daniel Alejandro Cervantes Cabrera
Profesor Investigador Ingeniero Tecnólogo Titular "A" ITA

C.c.p. **Mtro. Carlos Josué Lavandeira Portillo**, Director Adjunto de Innovación y Conocimiento. Presente.
Dr. Juan Antonio Vega Garfias, Subgerente de Innovación Gubernamental. Presente.