

Informe final publicable de proyecto

Verificación de Sistemas Inteligentes con Componentes con Capacidad de Aprendizaje

Código de proyecto ANII: IA_1_2022_1_173516

Fecha de cierre de proyecto: 01/09/2024

YOVINE, Sergio (Responsable Técnico - Científico)

BRABERMAN, Victor (Co-Responsable Técnico-Científico)

CARRASCO PIAGGIO, Matías (Investigador)

DA SILVA BARLOCO, Juan Pedro (Investigador)

DANDOIS, Santiago José (Investigador)

DELGADO, Tomás (Investigador)

GAGLIARDI, Hernán (Investigador)

GARAT, Alejo (Investigador)

ITURBIDE NORIA, Martín (Investigador)

KIDD, Johny (Investigador)

MARTÍNEZ VARSÍ, Nicolás (Investigador)

MAYR, Franz (Investigador)

MOLINOLO, Matías (Investigador)

PAN SUAREZ, Federico (Investigador)

SANCHEZ, Marco (Investigador)

UCHITEL, Sebastián (Investigador)

VILENSKY, Federico (Investigador)

WÜRTH, Mateus (Investigador)

UNIVERSIDAD ORT. FACULTAD DE INGENIERÍA (Institución Proponente) \\ FACULTAD DE CIENCIAS EXACTAS Y NATURALES DE LA UNIVERSIDAD DE BUENOS AIRES \\
UNIVERSIDAD ORT

Resumen del proyecto

Los sistemas de información actuales incluyen cada vez más componentes de inteligencia artificial con capacidad de aprendizaje, en particular las redes neuronales artificiales. Esto no se limita a sistemas de bajo riesgo, como por ejemplo, organizar un álbum de fotos, sino que el área está evolucionando con gran rapidez hacia su aplicación en sistemas críticos de alto riesgo, como el control de vehículos autónomos o el diagnóstico médico, cuyo mal funcionamiento puede causar daños irreparables a personas o al medioambiente. Para hacer frente a esta realidad, este proyecto se ubica dentro del área general de la "IA verificada" cuyo objetivo es diseñar sistemas de IA que tengan garantías sólidas de corrección con respecto a requisitos especificados matemáticamente. Su objetivo contribuir al desarrollo de métodos formales para la construcción de sistemas críticos con componentes de IA verificados. La dificultad de esta iniciativa es que si bien los componentes capaces de aprender implementan intrínsecamente algún modelo de cómputo, esto no significa que su verificación formal sea tratable con las técnicas existentes dado que son difíciles de formalizar. Un área de especial interés en el ámbito de sistemas críticos es garantizar que la evolución del estado de un sistema en el tiempo, consecuencia de las acciones que hace durante su funcionamiento, cumple ciertas propiedades. Este problema se expresa naturalmente en términos de lenguajes formales y lógicas temporales. En el caso de las redes neuronales artificiales, se trata de aquellas que manipulan secuencias, como las recurrentes y los Transformers. Concretamente, el objetivo del proyecto es desarrollar técnicas que contribuyan a garantizar el correcto comportamiento de sistemas de inteligencia artificial basados en redes neuronales sobre secuencias mediante la extracción de modelos formales que sirvan de base para la verificación de propiedades y la síntesis de controladores que guíen su comportamiento.

Ciencias Naturales y Exactas / Ciencias de la Computación e Información / Ciencias de la Computación / Inteligencia Artificial

Palabras clave: Verificación Formal / Inteligencia Artificial / Aprendizaje Automático /

Antecedentes, problema de investigación, objetivos y justificación.

Hoy en día, el desafío pasa por la construcción de sistemas que se caracterizan por contar con componentes capaces de aprender (learning-enabled components) [1]. Dicho aprendizaje se puede entender como la capacidad del componente de ir mejorando su rendimiento en una tarea dada mediante la experiencia. Para esto, el componente tiene que representar su conocimiento del universo mediante un modelo que es "entrenado" a partir de conjuntos de datos. En este contexto, las redes neuronales artificiales sobresalen como uno de los modelos más exitosos de la IA en varios campos de aplicación, debido a su capacidad de aprendizaje a partir de grandes volúmenes de datos y su consecuente eficacia predictiva.

La rápida adopción de este enfoque, conocida como aprendizaje profundo, no se limita a sistemas de bajo riesgo, como por ejemplo, organizar un álbum de fotos, recomendar una película, etc., sino que el área está evolucionando con gran rapidez a la integración de componentes de IA con capacidad de aprendizaje en sistemas críticos de alto riesgo. Estos sistemas son aquellos cuya falla puede resultar en consecuencias graves, como causar pérdida de vidas o daños irreparables a personas o al medioambiente. Como ejemplos recientes de la utilización de componentes con capacidad de aprendizaje en aplicaciones críticas, podemos mencionar el control de procesos de fusión nuclear, el control de vehículos autónomos, la detección de intrusos en ciberseguridad y el diagnóstico médico.

Surge entonces, el interés en garantizar el correcto funcionamiento de sistemas críticos que integran componentes de IA capaces de aprender, en particular aquellos que usan redes neuronales artificiales. Esto se ubica dentro del área general de la "IA verificada" [2] cuyo objetivo es diseñar sistemas de IA que tengan garantías sólidas, idealmente demostrables, de corrección con respecto a requisitos especificados matemáticamente. Tradicionalmente, la validación de la corrección de éstas con respecto a la tarea para la cual han sido entrenadas se centra en la evaluación empírica de métricas de eficacia en conjuntos de datos de prueba, siendo estos datos que no han sido utilizados durante el régimen de entrenamiento. Sin embargo, este método de validación presenta dos problemas principales. En primer lugar, resulta inviable evaluar exhaustivamente la respuesta del componente que embebe el modelo para todos los valores posibles de entrada. En segundo lugar, aún en los casos en que la validación arroje resultados de eficacia que podrían considerarse aceptables en grandes volúmenes de datos de prueba, los modelos pueden, de todas maneras, no generalizar correctamente a situaciones nuevas o ataques adversarios. En suma, validar experimentalmente los modelos con datos de prueba no asegura su correcto funcionamiento. Esta afirmación cobra especial relevancia en el ámbito de la IA responsable [9] en el contexto de sistemas críticos para los cuales es necesario garantizar la ausencia de fallas y de posibles usos indebidos. El correcto diseño, implementación y funcionamiento de sistemas críticos, ha sido abordado históricamente por el área de teoría de la computación llamada métodos formales, que se centra en la aplicación de técnicas fundadas en teorías matemáticas y lógicas que habilitan la especificación del comportamiento del sistema y la demostración rigurosa (deductiva o algorítmica) de propiedades. In fine, este proceso conduce al desarrollo de sistemas correctos.

Dentro de esta área se pueden identificar dos enfoques que han tenido gran impacto: verificación basada en modelos (model checking) y síntesis de controladores (controller synthesis). Model checking [3] es una familia de técnicas automáticas que, dado un modelo de un sistema, expresado como algún tipo de máquina de estados (que describe el comportamiento operacional del mismo), y una propiedad formal, expresada en alguna lógica temporal (que especifica el conjunto de secuencias válidas), buscan probar o refutar la propiedad en el modelo. El enfoque clave de esta técnica es la construcción de un modelo a un nivel de abstracción adecuado a fin de poder realizar una inspección sistemática y exhaustiva del modelo que garantice que todo comportamiento del sistema es correcto con respecto a la propiedad o, en su defecto, que encuentre evidencia de la existencia de comportamientos inválidos. Controller synthesis [4] agrupa técnicas que se encargan de construir automáticamente estrategias de control que induzcan a un sistema a exhibir un comportamiento deseado. Dicho de otra manera, estas técnicas se centran en limitar el comportamiento de un sistema a fin de que este cumpla con una implementación dada.

Así, el objetivo del proyecto es contribuir al desarrollo de métodos formales para la construcción de sistemas críticos con componentes de IA verificados, integrando model-checking y controller synthesis. La dificultad de esta iniciativa es que si bien los componentes capaces de aprender implementan

intrínsecamente algún modelo de cómputo, esto no significa que su verificación formal sea tratable con las técnicas existentes dado que son difíciles de formalizar.

La aplicación de métodos formales de verificación necesita una abstracción del componente bajo análisis. Dicho modelo abstracto es una especificación formal que puede ser realizada durante la etapa de diseño, o extraída de la implementación. En el caso de los componentes de IA con capacidad de aprendizaje, como las redes neuronales, el comportamiento es justamente aprendido durante la etapa de entrenamiento, por lo cual no se dispone de una especificación, siendo la extracción la única alternativa posible. La extracción de modelos de redes neuronales puede dividirse en dos grandes líneas: 1) el enfoque de "caja blanca" se basa en la inspección de la arquitectura (capas, funciones de activación, operaciones matriciales, regularizadores, etc.); y el enfoque de "caja negra" se apoya solamente en la observación de las salidas en respuesta a las entradas. La extracción "caja blanca", por ejemplo, sufre de un inconveniente importante que es la baja portabilidad dado su estrecha dependencia con la arquitectura y operaciones internas de la red. En cambio, las técnicas de extracción de tipo "caja negra" son más portables dado que no están atadas a características específicas de la clase de red neuronal. Por esta razón, este proyecto se centra principalmente en este último enfoque.

Por otro lado, la probabilidad juega un rol importante en los sistemas críticos donde se utiliza, por ejemplo, para modelar incertidumbre en el ambiente, para romper simetrías, elegir líderes y obtener consenso en algoritmos distribuidos, analizar rendimiento en sistemas de tiempo real, modelar comportamientos poco fiables o imprevisibles en sistemas tolerantes a fallas, etc. La aplicación de métodos formales en estos casos requiere de modelos estocásticos y lógicas temporales probabilísticas, lo que ha dado lugar al desarrollo de técnicas específicas. Por eso, el proyecto se centra en la extracción de modelos formales probabilísticos.

Concretamente, el proyecto extiende la técnica de extracción de modelos formales probabilísticos de las redes neuronales sobre secuencias propuesta en [5] y la técnica de síntesis de controladores propuesta [6] con el objetivo de contribuir a la construcción de sistemas críticos basados en componentes de IA verificada.

Además de los objetivos científicos planteados, el proyecto consolida la colaboración entre los equipos y genera una sinergia positiva para el desarrollo conjunto de técnicas formales de construcción de IA verificada, a la vez de contribuir a la formación y capacitación de jóvenes investigadores y estudiantes de grado y posgrado de ambas instituciones en un área de gran relevancia.

Metodología/Diseño del estudio

El proyecto se organiza en torno a dos áreas principales: A) Investigación y desarrollo y B) Capacitación.

A) Investigación y desarrollo. Esta actividad se descompone en cuatro subactividades.

A.1) Extracción de modelos.

El propósito de esta actividad es desarrollar técnicas de extracción de modelos, en particular diferentes clases de autómatas, a partir de redes neuronales que procesan secuencias, siguiendo la línea de investigación del equipo de ORT [5] en el marco general llamado inferencia gramatical dentro del aprendizaje computacional. Una de las dificultades mayores de la extracción es hacer frente al desafío que plantea la alta dimensionalidad del espacio de parámetros y entradas de la redes neuronales, dado que ésta es susceptible de generar espacios de estados muy grandes y eventualmente infinitos. Por esta razón es necesario desarrollar abstracciones adecuadas y técnicas eficientes de extracción a fin de obtener modelos tratables por métodos formales de verificación y síntesis. Para abordar esta problemática, se busca continuar con el enfoque planteado en [5], a través del desarrollo de algoritmos más eficientes y el uso de relaciones de equivalencia relevantes, al mismo tiempo que hacen el problema tratable computacionalmente.

A.2) Verificación.

El objetivo de esta tarea es habilitar la verificación de propiedades estocásticas sobre las redes neuronales. Una manera de hacerlo es, teniendo un modelo del componente de IA, recurrir a una herramienta de verificación estándar externa, como PRISM [7]. Otra consiste en construir abstracciones guiadas por la propiedad a verificar [8].

A.3) Síntesis.

Se busca extender las técnicas de síntesis de controladores [6] para integrar los modelos extraídos usando los métodos desarrollados en A.1 con el objetivo de construir sistemas de control basados en IA correctos. Por otro lado, se propone estudiar cómo integrar la síntesis de controladores y el proceso de extracción de modelos de componentes de IA. En efecto, el proceso de síntesis y el de extracción son dos caras de la misma moneda. La síntesis responde la pregunta "¿cómo debería comportarse el componente de IA?" y la extracción realiza ingeniería reversa para responder la pregunta "¿cómo se comporta la componente de IA?". El proyecto busca dar los primeros pasos de exploración de los desafíos conceptuales y técnicos que involucra la integración de estas dos áreas en el contexto de la IA verificada.

A.4) Desarrollo y Validación experimental.

Transversalmente se desarrollan y ejecutan experimentos con el objetivo de evaluar empíricamente sobre casos de estudio pertinentes las técnicas desarrolladas. El esquema de trabajo planteado se apoya en la paralelización de las actividades de extracción de modelos, verificación y síntesis, y en la interacción continua entre los equipos.

B) Capacitación.

El proyecto pone énfasis en la formación de estudiantes de grado y posgrado. Para esto se plantean dos acciones. Por un lado, la participación en eventos específicos de capacitación, en particular la Escuela de Ciencias Informáticas, organizada por el Departamento de Computación de la Universidad de Buenos Aires, y KHIPU (Latinamerican meeting in AI). Por otro, la integración de estudiantes en las publicaciones de los equipos, así como su asistencia a

conferencias internacionales.

Resultados, análisis y discusión

Los resultados obtenidos en las líneas de trabajo identificadas previamente son los siguientes.

A.1) Extracción de modelos.

En primer lugar, se desarrolló un marco teórico para la extracción de autómatas probabilísticos a partir de modelos neuronales del lenguaje basado en relaciones de equivalencia sobre distribuciones de probabilidad [10, 11, 12]. Las motivaciones prácticas del enfoque propuesto son dos. Por un lado, se puede ajustar la relación de equivalencia en virtud de la métrica definida para evaluar el comportamiento de los modelos de lenguaje (por ejemplo, el error con respecto a un "gold standard", esto es un conjunto de datos que han sido preparados manualmente y que representan las respuestas consideradas "correctas" lo más fielmente posible) o, incluso, a la estrategia de muestreo utilizada durante su utilización como IA generativa (por ejemplo, es habitual muestrear de las "k" palabras más probables o de aquellas cuya probabilidad es mayor a un cierto umbral). Por otro lado, las equivalencias permiten reducir significativamente el espacio de estados del autómata extraído, y hasta eventualmente asegurar su finitud, siendo entonces una herramienta muy eficaz para mitigar el problema conocido como "explosión de estados" en el ámbito de la verificación formal basada en modelos. En segundo lugar, se implementaron diversas optimizaciones al algoritmo de aprendizaje que condujeron a la obtención de ganancias significativas en memoria y tiempo de ejecución [12, 13, 14]. Este trabajo sentó las bases de una tesis de Maestría en curso que busca generalizar el algoritmo de extracción de modelos con el objetivo de capturar lenguajes caracterizables por gramáticas libres de contexto, más expresivas que los autómatas.

A.2) Verificación.

La herramienta desarrollada para la extracción de modelos fue evaluada experimentalmente para la verificación de casos de estudio de diferentes áreas críticas, como por ejemplo bioinformática, ciberseguridad e identificación de sesgos en generación de lenguaje natural [10, 11, 12]. Además, se propuso una técnica basada en la sincronización de la red neuronal con un controlador que permite la verificación on-the-fly de propiedades durante la extracción de modelos, así como también la obtención de modelos más reducidos. Los resultados experimentales obtenidos en el análisis de propiedades de grandes modelos de lenguaje como GPT-2 fueron muy prometedores. Por una lado, en términos de performance computacional, se consiguieron tiempos de ejecución hasta 90 veces más rápidos que con el algoritmo existente previo a la realización del proyecto [5]. Por otro lado, se extendió el alcance de la herramienta, habilitando la verificación de propiedades estocásticas. Esta línea de investigación está siendo continuada en el marco de una tesis de Maestría. Además, se trabajó en otro aspecto de la IA verificada, concretamente la IA explicable. En particular se abordó el estudio del comportamiento de las arquitecturas de tipo Transformer, que constituyen la base de los grandes modelos de lenguaje como GPT, desde un punto de vista de la interpretabilidad mecanística [15]. Los resultados experimentales obtenidos en este proyecto fueron el punto de partida de a una tesis de Maestría que se está desarrollando actualmente.

A.3) Síntesis.

En este proyecto desarrollamos un método de síntesis on-the-fly que se apoya en el uso de técnicas de inteligencia artificial con el objetivo de paliar el problema de la explosión de estados de los algoritmos de síntesis basados en autómatas], explorando solamente el espacio necesario para garantizar que el sistema controlado cumple con la tarea especificada [16, 17]. El trabajo experimental realizado sobre un conjunto de benchmarks mostró que es posible alcanzar ganancias de tiempo sustanciales y en general superar la mejor heurística obtenida con el algoritmo existente previamente [6].

Conclusiones y recomendaciones

Los resultados obtenidos durante este proyecto de investigación contribuyeron al estado del arte en el contexto de la inteligencia artificial verificada y a la formación de recursos humanos altamente calificados. Además, fueron el disparador de ideas que se terminaron plasmando en el marco del proyecto ANII FMV 1_2023_1_175864 "Herramientas para la verificación de aplicaciones industriales de grandes modelos de lenguaje" que comenzó en abril 2024, con la participación de las empresas De Larrobla y Asociados Uruguay SA (Bantotal) como contraparte y Rootstrap como participante y co-responsable. Finalmente, se definió una nueva línea de investigación conjunta entre los equipos que participaron en este proyecto. La propuesta consiste en mejorar la especialización de aplicaciones basadas en grandes modelos de lenguaje a través de un proceso iterativo que combine la extracción de controladores obtenidos a partir de prompts mediante el aprendizaje de modelos subrogados y la verificación de propiedades metamórficas [18].

Productos derivados del proyecto

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
Artículo científico	A Congruence-based Approach to Active Automata Learning from Neural Language Models	Mayr, Franz - Yovine, Sergio - Carrasco, Matías - Pan, Federico - Vilensky, Federico	https://proceedings.mlr.press/v217/mayr23a.html	https://hdl.handle.net/20.500.12381/3419	Finalizado
Artículo científico	Results of Neural-Checker Toolbox in Taysir 2023 Competition	Mayr, Franz - Yovine, Sergio - Carrasco, Matías - Garat, Alejo - Iturbide, Martín - da Silva, Juan - Vilensky, Federico	https://proceedings.mlr.press/v217/mayr23b.html	https://hdl.handle.net/20.500.12381/3420	Finalizado
Artículo científico	Exploration Policies for On-the-Fly Controller Synthesis: A Reinforcement Learning Approach	Delgado, Tomás - Sánchez Sorondo, Marco - Braberman, Víctor - Uchitel, Sebastián	https://doi.org/10.48550/arXiv.2210.05393	https://hdl.handle.net/20.500.12381/3418	Finalizado
Tesis de grado/monografías	Aprendiendo políticas de exploración generales para escalar la síntesis de controladores discretos	Delgado, Tomás		https://hdl.handle.net/20.500.12381/3417	Finalizado
Tesis de doctorado	Active Learning of Regular Languages as an Approach to Neural Language Models Verification	Franz Mayr		https://hdl.handle.net/20.500.12381/3622	Finalizado

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
Presentación en evento	Analyzing constrained LLM through PDFAs-learning	Carrasco, Matías - Mayr, Franz - Yovine, Sergio - Kidd, Johny - Iturbide, Martín - da Silva, Juan - Garat, Alejo		https://hdl.handle.net/20.500.12381/3624	Finalizado
Tesis de grado/monografías	Analysis, Evaluation and Improvement of Active Regular Inference Algorithms for Neural Sequence Acceptors	Juan Pedro da Silva, Martín Iturbide, Alejo Garat		https://hdl.handle.net/20.500.12381/3626	Finalizado
Tesis de grado/monografías	Exploring Attention Patterns and Neural Activations in Transformer Architectures for Sequence Classification in Context Free Grammars	Molinolo, Matías		https://hdl.handle.net/20.500.12381/3656	Finalizado
Software	Transformer Checker	Molinolo, Matías		https://redata.anii.org.uy/dataset.xhtml?persistentId=doi:10.60895/redata/Z8QDEZ	Finalizado
Software	Analyzing constrained LLM through PDFAs-learning experiments repository	Carrasco, Matías; Mayr, Franz; Yovine, Sergio; Kidd, Johny; Iturbide, Martín; da Silva, Juan; Garat, Alejo		https://redata.anii.org.uy/dataset.xhtml?persistentId=doi:10.60895/redata/KNERSJ	Finalizado
Software	A	Mayr,		https://redata.anii.org.uy/dataset.xhtml?persistentId=doi:10.60895/redata/KNERSJ	Finalizado

Tipo de producto	Título	Autores	Identificadores	URI en repositorio de Silo	Estado
	Congruence-based Approach to Active Automata Learning from Neural Language Models code repository	Franz; Yovine, Sergio; Carrasco, Matías; Pan, Federico; Vilensky, Federico		persistentId=doi:10.60895/redata/NDHQQQ	
Software	Results of Neural-Checker Toolbox in Taysir 2023 Competition code repository	Mayr, Franz; Yovine, Sergio; Carrasco, Matías; Garat, Alejo; Iturbide, Martín; da Silva, Juan; Vilensky, Federico		https://redata.anii.org.uy/dataset.xhtml?persistentId=doi:10.60895/redata/JY5DUS	Finalizado
Documento de trabajo	Un enfoque para mejorar el proceso de prompt engineering en aplicaciones basadas en LLM	Víctor Braberman, Sebastián Uchitel, Matías Carrasco, Franz Mayr, Sergio Yovine		https://hdl.handle.net/20.500.12381/3730	Finalizado

Referencias bibliográficas

- [1] Shengduo Chen, Yaowei Sun, Dachuan Li, Qiang Wang, Qi Hao, Joseph Sifakis. Runtime Safety Assurance for Learning-enabled Control of Autonomous Driving Vehicles. ICRA 2022: 8978-8984
- [2] Sanjit A. Seshia, Dorsa Sadigh, S. Shankar Sastry. Toward Verified Artificial Intelligence. Communications of the ACM, July 2022, Vol. 65 No. 7, Pages 46-55.
- [3] Edmund Clarke, Orna Grumberg, Daniel Kroening, Doron Peled, Helmut Veith (2018). Model Checking. Cyber Physical Systems Series. The MIT Press.
- [4] Christel Baier, Marcus Größer, Martin Leucker, Benedikt Bollig, Frank Ciesinski (2004). Controller Synthesis for Probabilistic Systems. Exploring New Frontiers of Theoretical Informatics. International Federation for Information Processing, vol 155. Springer, Boston, MA.
- [5] Franz Mayr, Sergio Yovine, Federico Pan, Nicolas Basset, Thao Dang (2022). Towards Efficient Active Learning of PDFAs. LearnAut 2022, Paris, France.
- [6] Daniel Ciolek, Matias Durán, Florencia Zanollo, Nicolás Pazos, Julián Braier, Víctor Braberman, Nicolás D'Ippolito, Sebastián Uchitel (2023). On-the-fly informed search of non-blocking directed controllers. Automatica, 147: 110731.
- [7] Marta Kwiatkowska, Gethin Norman, David Parker (2007). Stochastic model checking. In International School on Formal Methods for the Design of Computer, Communication and Software Systems (pp. 220-270). Springer, Berlin, Heidelberg.

- [8] Igor Khmel'nitsky, Daniel Neider, Rajarshi Roy, Xuan Xie, Benoît Barbot, Benedikt Bollig, Alain Finkel, Serge Haddad, Martin Leucker, Lina Ye (2021). Property-directed verification and robustness certification of recurrent neural networks. SpringerLink.
- [9] Virginia Dignum (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer Nature.
- [10] Franz Mayr, Sergio Yovine, Matías Carrasco, Federico Pan, Federico Vilensky (2023). Proceedings of 16th edition of the International Conference on Grammatical Inference, PMLR 217:250-264, 2023.
- [11] Franz Mayr (2024). Active Learning of Regular Languages as an Approach to Neural Language Models Verification. Tesis de Doctorado, PEDECIBA.
- [12] Matías Carrasco, Franz Mayr, Sergio Yovine, Johnny Kidd, Martín Iturbide, Juan Pedro da Silva, Alejo Garat (2023). Analyzing constrained LLM through PDFa-learning. LearnAut 2024. CoRR abs/2406.08269 (2024)
- [13] Franz Mayr, Sergio Yovine, Matías Carrasco, Alejo Garat, Martín Iturbide, Juan da Silva, Federico Vilensky (2023). Proceedings of 16th edition of the International Conference on Grammatical Inference, PMLR 217:295-298.
- [14] Alejo Garat, Martín Iturbide, Juan da Silva (2024). Analysis, Evaluation and Improvement of Active Regular Inference Algorithms for Neural Sequence Acceptors. Trabajo final de Ingeniería en Sistemas, Universidad ORT Uruguay.
- [15] Matías Molinolo (2024). Exploring Attention Patterns and Neural Activations in Transformer Architectures for Sequence Classification in Context Free Grammars. Trabajo final de Ingeniería en Sistemas, Universidad ORT Uruguay.
- [16] Tomás Delgado, Marco Sánchez Sorondo, Víctor A. Braberman, Sebastián Uchitel (2023). Exploration Policies for On-the-Fly Controller Synthesis: A Reinforcement Learning Approach. ICAPS 2023: 569-577.
- [17] Tomás Delgado (2023). Aprendiendo políticas de exploración generales para escalar la síntesis de controladores discretos. Tesis de Lic. en Computación, DC, UBA.
- [18] Víctor Braberman, Sebastián Uchitel, Matías Carrasco, Franz Mayr, Sergio Yovine (2024). Un enfoque para mejorar el proceso de prompt engineering en aplicaciones basadas en LLM. Documento de trabajo.

Licenciamiento

Reconocimiento 4.0 Internacional. (CC BY)

