Con el apoyo de

PROGRAMA DEL CURSO

Herramientas para explorar sesgos y estereotipos de la inteligencia artificial en las aulas (HESEIA).

Profesores que dictarán el curso
Dra. Luciana Benotti (FAMAF)/ Dra. Laura Alonso Alemany (FAMAF) /Dr. Marcos Gomez (FAMAF) /
Dra. Emilia Echeveste (CONICET y FCEFYN) / Dra. Beatriz Busaniche (Fundación Vía Libre)
Lic. Guido Ivetta (FAMAF y Fundación Vía Libre) /Lic. Hernán J. Maina (CONICET)
Lic. Nair Carolina Mazzeo (Fundación Vía Libre)

Objetivos: Los objetivos de este proyecto se desprenden de una herramienta construida por el equipo de ética en IA de la Fundación Vía Libre, llamada EDIA (Estereotipos y Discriminación en Inteligencia Artificial), la cual permite identificar y registrar discriminación en modelos de Inteligencia Artificial. El objetivo principal será que los docentes puedan reflexionar sobre los sesgos sociales negativos y los estereotipos embebidos en los modelos de Inteligencia Artificial basados en datos. También es un objetivo explicar cómo funcionan y cómo aprenden estos modelos y desarrollar junto a los docentes insumos para evaluar estos modelos. La realización de estos talleres permitirá que este sea un contenido a desarrollar en las escuelas y se prevé construir conjuntamente materiales didácticos adaptados.

CONTENIDOS: El curso contempla el cursado de 4 módulos con los siguientes contenidos:

Módulo 1: Inteligencia artificial como parte de la ciudadanía digital

Introducción de conceptos básicos de inteligencia artificial (IA) generativa. Exploración de aplicaciones prácticas de los modelos de lenguaje generativos. Reflexión sobre las respuestas de los modelos generativos y sus sesgos.

Módulo 2: Introducción a modelos generativos, primeros acercamientos

Exploración de herramientas para analizar sesgos en modelos de IA. Comprender cómo se generan los estereotipos y sesgos en las predicciones de IA. Introducción a conjuntos de datos de evaluación colaborativos y su importancia. Interacción con los datos generados.

Módulo 3: Sesgos sociales y estereotipos en modelos generativos

Presentación sobre el extractivismo de datos y sus implicaciones. Taller sobre estrategias de recolección de datos en el aula, respetando la privacidad y ética. Discusión sobre la integración de estos datos con las currículas específicas.

Módulo 4: Analizar sesgos en las aulas

Actividad para analizar y discutir los datos recolectados.

Taller para formalizar la experiencia en un documento de propuesta educativa para futuras intervenciones. Presentación de las experiencias y retroalimentación entre pares.

CERTIFICACIÓN

El mismo cuenta con la certificación de la Junta de Clasificación Secundaria de la Provincia de Córdoba de 46 horas (Resolución en trámite) y la certificación de Curso de Extensión de la Facultad de Matemática, Astronomía, Física y Computación (FaMAF) de la Universidad Nacional de Córdoba (RHCD-2024-205-E-UNC-DEC#FAMAF).

Bibliografía de Referencia (No obligatoria en el curso):

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI and Ethics, 2(3), 431-440.

Alonso Alemany, L., Benotti, L., Maina, H., Gonzalez, L., Martínez, L., Busaniche, B., … Rajngewerc, M. (2023, May). Bias assessment for experts in discrimination, not in computer science. In S. Dev, V. Prabhakaran, D. Adelani, D. Hovy, & L. Benotti (Eds.), Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP) (pp. 91–106). doi:10.18653/v1/2023.c3nlp-1.10

Antoniak, M., & Mimno, D. (2021, August). Bad Seeds: Evaluating Lexical Methods for Bias Measurement. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1889–1904). doi:10.18653/v1/2021.acl-long.148

Blodgett, S. L., Barocas, S., Daumé, H., III, & Wallach, H. (2020, July). Language (Technology) is Power: A Critical Survey of ``Bias" in NLP. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5454–5476). doi:10.18653/v1/2020.acl-main.485

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Proceedings of the 30th International Conference on Neural Information Processing Systems, 4356–4364. Presented at the Barcelona, Spain. Red Hook, NY, USA: Curran Associates Inc.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183–186. doi:10.1126/science.aal4230

Fesakis, G., & Prantsoudi, S. (2021, November). Raising Artificial Intelligence Bias Awareness in Secondary Education: The Design of an Educational Intervention. In ECIAIR 2021 3rd European Conference on the Impact of Artificial Intelligence and Robotics (p. 35). Academic Conferences and publishing limited.

Jurafsky, D and Martin, J. (2000). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (1st. ed.). Prentice Hall PTR, USA.

Manning, C. D., Schütze, H. (1999). Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. ISBN: 0262133601

Maina, H., Alonso Alemany, L., Ivetta G., Rajngewerc, M., Busaniche, B., and Benotti, L. (2024). Exploring stereotypes and biases in language technologies in Latin America. Communications of the Association for Computing Machinery (ACM). ISSN:0001-0782.

Ng, D. T. K., Su, J., Leung, J. K. L., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy education in secondary schools: a review. Interactive Learning Environments, 1-21.

Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2023). A review of AI teaching and learning from 2000 to 2020. Education and Information Technologies, 28(7), 8445-8501.

Payne, B. H. (2019). An ethics of artificial intelligence curriculum for middle school students. MIT Media Lab Personal Robots Group. Retrieved Oct, 10, 2019.