

Conceptos		Desglose de conceptos
1.	Título del programa académico	Inteligencia Artificial para detectar y moderar el discurso de odio en plataformas digitales
2.	Fecha de inicio	09 de febrero de 2027
3.	Fecha límite de inscripción	05 de febrero del 2027
4.	Modalidad:	A distancia (Zoom)
5.	Costo por persona:	\$3,480.00 IVA incluido
6.	Datos de contacto:	Sofía Rodríguez Correo: profesionalizacion@inap.org.mx Tel. 55 50 81 26 00 ext. 2655 Comercialización Teléfono: 55 5081 2600 ext. 2003 y 2006 WhatsApp: 55 64257912 Correo Electrónico: comercializacion@inap.org.mx
7.	Duración	20 horas
8.	Sesiones y Horario	09, 11, 16, 18, 23,25 de febrero, 02 y 04 de marzo de 2027 martes y jueves de las 18:30 a 21:00 horas
9.	Ventajas de estudiar el programa académico.	El programa académico permitirá a los participantes desarrollar capacidades para interpretar el fenómeno del discurso de odio en entornos digitales desde una perspectiva sociotécnica; distinguir entre categorías de riesgo discursivo; usar criterios básicos de análisis de datos, monitoreo digital e inteligencia artificial aplicada; reconocer señales tempranas de escalamiento; leer tableros e indicadores de conversación; valorar el alcance y los límites de modelos de clasificación automática; y diseñar respuestas institucionales que articulen legalidad, proporcionalidad, protección de derechos y eficacia operativa. • Delimitar conceptualmente el discurso de odio y diferenciarlo de otras formas de violencia o conflicto discursivo. • Explicar cómo operan los algoritmos de amplificación y recomendación en la circulación del odio digital.

		• Identificar variables clave para la detección automática: contexto, intención, grupo objetivo, severidad, alcance, probabilidad de daño y temporalidad. • Comprender la lógica básica de un sistema de clasificación de odio basado en reglas, corpus anotados y modelos supervisados. • Reconocer los sesgos más frecuentes de los sistemas automáticos y los retos de explicabilidad, auditabilidad y justicia algorítmica. • Diseñar un esquema institucional mínimo de monitoreo, respuesta y evaluación frente a episodios de odio digital.
10.	Presentación del programa	La expansión del discurso de odio en plataformas digitales constituye uno de los desafíos más complejos para la administración pública contemporánea. No se trata únicamente de mensajes ofensivos aislados, sino de dinámicas sociotécnicas que combinan polarización, desinformación, acoso coordinado, discriminación y amplificación algorítmica. En este contexto, la inteligencia artificial ofrece herramientas para detectar patrones lingüísticos de hostilidad, identificar narrativas de riesgo, monitorear comunidades digitales, establecer alertas tempranas y diseñar respuestas institucionales proporcionales, compatibles con la libertad de expresión y los estándares de derechos humanos. Este curso propone una formación aplicada para que las y los participantes comprendan el ecosistema del discurso de odio en línea, conozcan los fundamentos conceptuales, jurídicos y tecnológicos de su detección y moderación, y desarrollen capacidades para utilizar herramientas de inteligencia artificial en tareas de monitoreo, análisis, clasificación, evaluación de riesgos y diseño de protocolos institucionales de respuesta.
11.	Objetivo general del programa	Proporcionar una visión amplia, actualizada y aplicada del uso de la inteligencia artificial para detectar, clasificar, monitorear y moderar el discurso de odio en plataformas digitales, fortaleciendo las capacidades de la administración pública para prevenir riesgos, proteger derechos, diseñar respuestas institucionales y construir políticas públicas basadas en evidencia.
12.	Objetivos específicos	1. Comprender qué es el discurso de odio, cómo se diferencia del discurso ofensivo, del acoso y de la crítica política dura, y por qué representa un problema estratégico para el Estado.

		2. Analizar la ecología digital en la que circula el odio: algoritmos de recomendación, cámaras de eco, viralización, anonimato, coordinación y circulación multiplataforma. 3. Identificar los principales enfoques de inteligencia artificial y procesamiento de lenguaje natural aplicados a la detección y moderación del discurso de odio. 4. Reconocer sesgos, límites y riesgos de los modelos automáticos, especialmente en español de México, lenguajes locales, ironía, multimodalidad y contextos políticos sensibles. 5. Diseñar criterios institucionales para monitoreo, alerta temprana, priorización de casos, protección de grupos vulnerables y evaluación de intervenciones públicas. 6. Traducir evidencia digital en herramientas concretas de política pública, prevención, respuesta y coordinación interinstitucional.
13.	Requisitos de admisión	Llenar el formato de registro y hacer su pago.
14.	Contenido temático	DEL INSULTO AL ALGORITMO: QUÉ ES EL DISCURSO DE ODIO Y POR QUÉ IMPORTA HOY Objetivo de la sesión: Comprender el concepto de discurso de odio, su delimitación frente a otras expresiones y su relevancia para la administración pública. Subtemas: 1.1. Definiciones, umbrales y daños. 1.2. Diferencia entre discurso ofensivo, acoso, discriminación e incitación. 1.3. Derechos humanos, libertad de expresión y test de contexto. 1.4. Grupos objetivo y daños individuales, sociales y democráticos. 1.5. Particularidades del caso mexicano: diversidad lingüística, polarización y violencia digital. ECOLOGÍA DIGITAL, PLATAFORMAS Y CIRCULACIÓN DEL ODIO Objetivo de la sesión: Entender cómo la arquitectura de plataformas favorece la amplificación, persistencia y normalización del odio. Subtemas:

- 2.1. Economía de la atención, recomendación algorítmica y viralización.
- 2.2. Cámaras de eco, burbujas de filtro y polarización.
- 2.3. Multiplataforma, anonimato, bots y coordinación.
- 2.4. Eventos gatillo: elecciones, crisis de seguridad, género y movilizaciones sociales.
- 2.5. Riesgos específicos en X, Facebook, TikTok, YouTube, WhatsApp y blogs.

INTELIGENCIA ARTIFICIAL Y PROCESAMIENTO DE LENGUAJE NATURAL PARA DETECTAR DISCURSO DE ODIO

Objetivo de la sesión: Conocer los fundamentos técnicos y operativos de la detección automática del odio digital.

Subtemas:

- 3.1. Corpus, anotación y categorías de clasificación.
- 3.2. Reglas, modelos supervisados y grandes modelos de lenguaje.
- 3.3. Variables de análisis: severidad, target, contexto, temporalidad y modalidad.
- 3.4. Retos de detección en español de México: ironía, resignificación, sarcasmo, lenguaje multimodal y code-switching.
- 3.5. Indicadores de desempeño: precisión, recall, falsos positivos y falsos negativos.
- 3.6. Explicabilidad y auditoría de modelos.

MODERACIÓN, RESPUESTA INSTITUCIONAL Y COORDINACIÓN GUBERNAMENTAL

Objetivo de la sesión: Traducir el análisis del odio digital en protocolos y capacidades institucionales de respuesta.

Subtemas:

- 4.1. Monitoreo continuo y sistemas de alerta temprana.
- 4.2. Moderación automática, moderación humana y contradiscurso.
- 4.3. Umbrales de intervención y rutas de priorización.
- 4.4. Protección de mujeres en política, periodistas, personas defensoras y grupos vulnerables.
- 4.5. Coordinación entre áreas de seguridad, derechos humanos, comunicación, justicia y órganos reguladores.

		4.6. Evidencia digital, trazabilidad y cadena institucional de actuación. DE LOS DATOS A LA POLÍTICA PÚBLICA: TABLERO, INDICADORES Y PLAN DE ACCIÓN Objetivo de la sesión: Diseñar instrumentos de gestión pública para prevenir, responder y evaluar el odio digital. Subtemas: 5.1. Tableros de monitoreo e indicadores clave. 5.2. Señales, umbrales y tiempos de respuesta. 5.3. Evaluación de intervenciones y aprendizaje institucional. 5.4. Auditoría cívica, transparencia y rendición de cuentas algorítmica. 5.5. Agenda pública para México: prevención, regulación, alfabetización digital y gobernanza democrática. 5.6. Presentación de ejercicios finales.
15.	Al finalizar el curso de capacitación recibirá	Constancias a cada participante, siempre y cuando cumplan con al menos el 80% de asistencia al curso y obtengan una calificación mínima aprobatoria de 70 en una escala de 0 al 100 en la evaluación final. La constancia es digital y se enviará a su correo electrónico a más tardar 20 días hábiles, contando a partir de la fecha de la última sesión. Constancia con valor curricular
16.	Profesores	Dr. Javier Sánchez Galicia Doctor en Investigación de la Comunicación Dr. Ulises Cruz Valencia Doctor en Ciencias de Gobierno y Política