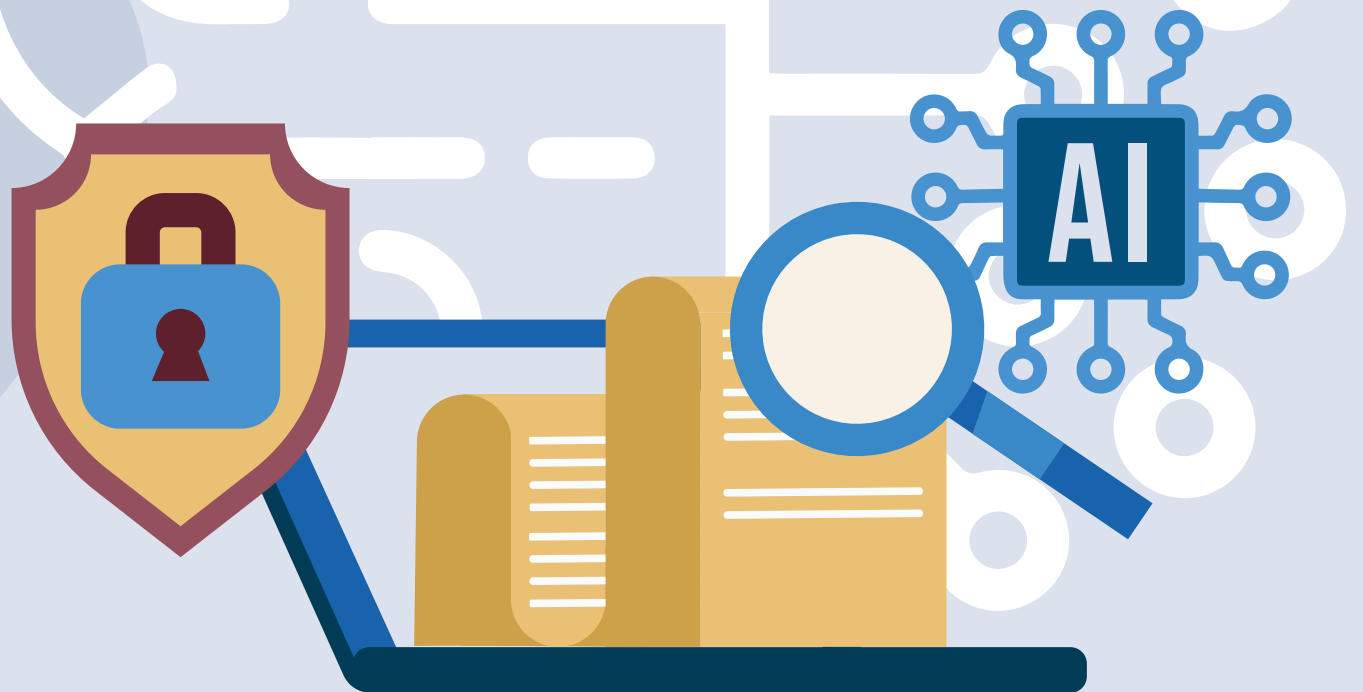




MÓDULO DE
**TÉCNICAS DE INVESTIGACIÓN
ADMINISTRATIVA Y PREVENTIVA**

UNIDAD VII

USO RESPONSABLE DE IA EN LA INVESTIGACIÓN ADMINISTRATIVA Y PREVENTIVA



7.1 Uso responsable de IA en la investigación administrativa y preventiva

Debe tenerse en cuenta, que un módulo de capacitación no es norma vinculante, sino más bien una guía de buenas prácticas. Por ello, los criterios y recomendaciones se plantean como buenas prácticas, que deben dialogar con el marco normativo vigente: Constitución, Ley del Organismo Judicial, Ley de la Carrera Judicial y su Reglamento, Ley de Servicio Civil y su Reglamento, Normas de Comportamiento Ético, LAIP, Pacto Colectivo y Manual de Procedimientos de la SGT. El objetivo es ofrecer un marco prudente y alineado con estándares internacionales (OCDE, UNESCO, UE, CEPEJ, CNJ de Brasil, Consejo Superior de la Judicatura de Colombia, entre otros), que evite tanto el rechazo tecnófobo como el uso acrítico o riesgoso de sistemas de inteligencia artificial (IA).

7.2 ¿Qué es la IA (incluida la IA generativa) y cuáles son sus riesgos en justicia?

La discusión sobre qué es exactamente la “inteligencia artificial” está lejos de estar cerrada. Diferentes organismos internacionales han propuesto definiciones funcionales que, más que pretender ser “esenciales”, buscan describir familias de sistemas relevantes para la regulación y la gobernanza.

La OCDE, en su Recomendación sobre IA, define los sistemas de IA como sistemas basados en máquinas que, para objetivos explícitos o implícitos, procesan datos y producen resultados (predicciones, recomendaciones o decisiones) que afectan entornos físicos o virtuales. Esta definición subraya tres

elementos: la base algorítmica/máquina, la orientación a un objetivo y la capacidad de generar resultados que influyen en la realidad (por ejemplo, priorización de casos, sugerencias de redacción, clasificadores de riesgo).



El AI Act europeo –Reglamento de la Unión Europea sobre inteligencia artificial– retoma y refina esta lógica al entender por sistema de IA un sistema basado en máquinas, diseñado para operar con diversos grados de autonomía, que puede mostrar adaptabilidad tras su despliegue y que, para objetivos explícitos o implícitos, infiere a partir de las entradas que recibe cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. El énfasis está en la capacidad de inferir (no solo seguir reglas explícitas), en la autonomía variable y en el hecho de que los outputs pueden condicionar decisiones humanas o automatizadas.

En este marco general, la IA generativa es un subconjunto de sistemas de IA capaces de producir contenido nuevo (texto, imágenes, audio, video, código, resúmenes, borradores de informes) a partir de instrucciones o “prompts” en lenguaje natural. Modelos como los grandes modelos de lenguaje

(LLMs) son ejemplos paradigmáticos: no “recuperan” simplemente fragmentos de texto, sino que generan respuestas probabilísticas en función de patrones aprendidos en grandes volúmenes de datos.

Para la administración de justicia, y específicamente para funciones de supervisión e investigación administrativa y preventiva, estos sistemas presentan un potencial importante (borradores de actas, síntesis documentales, analítica descriptiva de cargas), pero también riesgos significativos:

1. Distorsiones y errores fácticos.

Los modelos generativos pueden producir información factualmente errónea, pero redactada con alto grado de fluidez y aparente seguridad. En un contexto disciplinario esto puede implicar la construcción de resúmenes o argumentos basados en datos inexistentes, citas normativas falsas o caracterizaciones inexactas de actuaciones procesales. El riesgo es mayor cuando el usuario asume que “si lo escribe la máquina, debe ser cierto” y relaja el control de verificación.

2. Sesgos y discriminación.

Muchos sistemas de IA aprenden a partir de datos históricos que reflejan desigualdades estructurales (por ejemplo, distribución desigual de sanciones disciplinarias, prácticas discriminatorias, sesgos de género o étnico). Sin supervisión adecuada, estos sistemas pueden reproducir o amplificar dichos sesgos, por ejemplo, al sugerir prioridades de supervisión o al describir perfiles de riesgo que afecten de forma desproporcionada a mujeres, pueblos indígenas, personas con discapacidad o colectivos sindicalizados. Esto entra en tensión con los principios constitucionales de igualdad y no discriminación, así como con las obligaciones internacionales en materia de derechos humanos.

3. Privacidad, confidencialidad y protección de datos personales.

La SGT trabaja con expedientes que contienen datos personales, datos sensibles, información laboral y disciplinaria, e incluso información clasificada como confidencial o reservada conforme a la LAIP. Cargar esta información en servicios externos de IA (especialmente en plataformas públicas en la nube que almacenan o reutilizan los prompts y los documentos) puede constituir una violación de la reserva legal, de las normas de protección de datos personales y de la propia cadena de custodia administrativa y digital. La propia LAIP exige clasificación formal, prueba de daño y plazos de reserva; cualquier práctica de IA que vulnere este esquema podría constituir una falta disciplinaria (por violación de reserva, negligencia grave o abuso de autoridad), sin que sea necesario crear nuevos tipos.

4. Opacidad y explicabilidad limitada.

Algunos sistemas de IA –particularmente los modelos “caja negra”– no permiten explicar de manera clara cómo se llegó a un determinado resultado. En materia de justicia, esto es especialmente problemático: la motivación, la transparencia y la posibilidad de control judicial son requisitos esenciales del debido proceso. Si una recomendación disciplinaria, o una priorización de casos, se apoya de facto en un algoritmo opaco, la SGT corre el riesgo de no poder justificar racionalmente sus decisiones, lo que erosiona la confianza y expone a nulidades o reproches de la Corte de Constitucionalidad o de la propia carrera judicial.

5. Fuga de datos y dependencia tecnológica.

La integración de herramientas de IA que no estén alojadas o gobernadas por el propio Organismo Judicial puede generar dependencias tecnológicas y riesgo de fuga de datos (por ejemplo, proveedores que retienen logs, metadatos o fragmentos de expedientes). Además, la falta de control de versiones y de gobernanza institucional (como advierten la OCDE y la UNESCO para el sector público) puede conducir a una fragmentación de prácticas, donde cada unidad adopta herramientas distintas, sin análisis de impacto ni base legal clara.

Por estas razones, el uso de IA en la SGT debe entenderse siempre como herramienta de apoyo bajo control humano significativo, nunca como sustituto de la deliberación jurídica, de la valoración probatoria conforme a sana crítica, ni de la responsabilidad personal de quienes investigan, recomiendan y deciden.

7.3 Principios y lineamientos: control humano significativo y garantías básicas

A partir de los principios de la OCDE y de la Recomendación de la UNESCO sobre la ética de la IA –que enfatizan el carácter centrado en el ser humano, el respeto de derechos, la transparencia, la robustez y la responsabilidad–, así como de la Carta Ética Europea del CEPEJ para el uso de IA en sistemas judiciales, pueden formularse algunos lineamientos de buenas prácticas para la SGT.

Primero, debe asegurarse un control humano significativo en todas las etapas del uso de IA.

Esto implica que las herramientas de IA solo pueden operar como apoyo a la actividad humana, y que tanto la apertura de una investigación disciplinaria, como la calificación jurídica de los hechos, la propuesta de sanción o archivo y la redacción final del informe deben ser decisiones tomadas por personas investidas de competencia, con capacidad de comprender, aceptar o rechazar las sugerencias de la herramienta. Ningún modelo de IA debe configurar una “decisión automática” en el sentido de la última palabra sobre derechos o sobre la situación disciplinaria de una persona.

En segundo lugar, se impone el principio de no delegar decisiones de fondo a sistemas de IA.

De acuerdo con el enfoque de riesgo del AI Act, los usos de IA en el ámbito de la justicia se consideran de alto riesgo y están sujetos a exigencias reforzadas de gobernanza, transparencia y control humano. En el contexto guatemalteco, esto se traduce en que los sistemas de IA pueden asistir en tareas instrumentales (búsqueda de información, organización de datos, redacción de borradores), pero no pueden decidir por sí mismos sobre la existencia de una falta, la gravedad de la conducta, la sanción aplicable ni el archivo de actuaciones.

Un tercer principio es la gestión responsable de datos y confidencialidad.

alineada con la LAIP, la Ley de Servicio Civil, las Normas Éticas y el Pacto Colectivo. Ello supone:

- Limitar el uso de IA, especialmente generativa, a entornos institucionales seguros (preferiblemente soluciones internas o proveedores con garantías legales claras y contratos específicos de protección de datos);
- Evitar la introducción de datos personales, sensibles o información clasificada en servicios externos de IA que no ofrezcan garantías equivalentes a las exigidas por la LAIP;
- Documentar por escrito las decisiones de clasificación, anonimización y reserva cuando la IA se utilice en la elaboración de matrices, informes o versiones públicas.

El cuarto principio es la trazabilidad y registro.

Toda utilización relevante de IA en la investigación administrativa y preventiva debe dejar huella en los expedientes de la SGT: qué herramienta se usó, con qué propósito, qué datos se le suministraron, qué resultados produjo y cómo fueron validados o modificados por los funcionarios responsables. Este registro permite reconstruir el proceso y evita que, en un futuro, se alegue que la decisión fue “de la máquina” sin posibilidad de reconstruir la cadena de razonamiento.

Finalmente, debe existir una validación humana sistemática.

Los borradores, resúmenes o análisis generados por IA deben ser revisados críticamente, comparados con los documentos originales y corregidos o descartados cuando contengan errores, sesgos o formulaciones incompatibles con el lenguaje claro y la motivación exigida en sede administrativa y judicial. La validación humana no es una formalidad; implica asumir un deber reforzado de lectura crítica, consciente de los límites de la herramienta.

7.4 Casos de uso permitidos (bajo supervisión): buenas prácticas

Partiendo de estos principios, y siguiendo la lógica de lineamientos como los del Consejo Superior de la Judicatura de Colombia (Acuerdo PCSJA24-12243, que adopta un plan especial de IA generativa para la Rama Judicial) y de la Resolução CNJ n. 332/2020 y normas posteriores del Consejo Nacional de Justicia de Brasil sobre IA en el Poder Judicial, es posible identificar un conjunto de usos que pueden considerarse prudencialmente aceptables, siempre bajo supervisión humana y sin tratamiento de datos sensibles ni información reservada en servicios externos.

En primer lugar, la búsqueda y organización de información normativa y jurisprudencial interna.

Herramientas de IA (no necesariamente generativa) pueden apoyar a la SGT en la recuperación de leyes, reglamentos, acuerdos de la CSJ y del Consejo de la Carrera Judicial, precedentes de la Corte de Constitucionalidad y resoluciones disciplinarias de las Juntas de Disciplina y de la URD, siempre que operen sobre bases de datos institucionales y sin extraer información a servicios externos. En este ámbito, la IA se asemeja a un motor de búsqueda avanzado que ayuda a localizar normas, clasificar documentos y detectar vínculos relevantes, pero la selección y la interpretación siguen siendo responsabilidad de la persona investigadora.

En segundo lugar, la elaboración de borradores y síntesis documentales.

La IA generativa puede ser un apoyo útil para producir primeros borradores de:

- Resúmenes ejecutivos de informes.
- Sistematizaciones de hechos.
- Síntesis de testimonios o de antecedentes disciplinarios.
- Esquemas de motivación fáctica (por ejemplo, organizar un relato cronológico complejo).

No obstante, por prudencia y conforme a las Normas Éticas y a la LAIP, es recomendable que:

- Cuando se utilicen servicios externos, se trabaje exclusivamente con textos previamente anonimizados o “maquetas” sin datos personales ni detalles identificables.
- Los borradores generados se sometan a una revisión exhaustiva para asegurar exactitud, neutralidad y ausencia de lenguaje estigmatizante o sesgado.
- Se cite en el informe, de manera transparente (cuando sea pertinente), que se ha utilizado una herramienta de apoyo tecnológico, sin trasladar a la IA la autoría ni la responsabilidad del documento.

En tercer lugar, la analítica descriptiva de cargas y tiempos de respuesta.

Inspirándose en experiencias internacionales donde la IA se utiliza para mapear tiempos de tramitación y cargas de trabajo, la SGT podría valerse de herramientas analíticas (no necesariamente generativas) para:

- Describir el volumen de quejas.
- Identificar cuellos de botella en visitas preventivas o en investigaciones.
- Localizar áreas de riesgo por acumulación de señalamientos o retrasos.

Siempre que los datos se trabajen de forma agregada y anonimizada, y que no se utilice la IA para calificar individualmente a jueces o auxiliares como “buenos” o “malos”, esta analítica puede ayudar a orientar estrategias preventivas, planificar visitas y priorizar recursos, en armonía con la gestión por riesgos que propone el Manual de Procedimientos de la SGT.

En cuarto lugar, la asistencia en la elaboración de actas y formatos estándar. La IA puede apoyar en la redacción de plantillas de actas de visita, matrices de verificación o cuadros comparativos, siempre que:

- Los contenidos sensibles se consignen posteriormente de forma manual o se trabajen en entornos controlados; y
- se mantenga el respeto a la cadena de custodia documental y a las exigencias de firma y autenticidad previstas en la Ley del Organismo Judicial y en la normativa administrativa.

En todos estos usos, es clave insistir en que se trata de buenas prácticas sugeridas, coherentes con la doctrina internacional según la cual la IA debe reforzar –no debilitar– los principios de independencia, imparcialidad, transparencia y motivación de las decisiones judiciales y administrativas.

7.5 prohibidos o fuertemente restringidos (como buenas prácticas)

De manera complementaria, la experiencia comparada muestra una serie de usos que deben considerarse altamente problemáticos o directamente contrarios a los estándares internacionales y al marco normativo interno, de modo que el módulo de capacitación debería desaconsejarlos expresamente como parte de las buenas prácticas de la SGT.

En primer lugar, la automatización de decisiones disciplinarias. Ni el diseño de la Ley de la Carrera Judicial ni el de la Ley de Servicio Civil ni las garantías del debido proceso disciplinario son compatibles con sistemas en los que una herramienta de IA decida, por sí misma, la existencia de una falta, la graduación de su gravedad o la sanción correspondiente. La línea que adoptan el AI Act, la Carta Ética del CEPEJ y las directrices de CNJ y CSJ de Colombia es clara: la IA puede aportar insumos, pero la decisión debe ser humana, motivada y controlable.



En segundo lugar, el acceso no autorizado de herramientas de IA a expedientes disciplinarios o de supervisión. Cualquier implementación que suponga conectar un servicio externo a los sistemas informáticos del Organismo Judicial, de modo que éste pueda leer expedientes completos, metadatos o información de personal, debe considerarse fuertemente desaconsejada sin un marco de autorización institucional, análisis de impacto y garantías contractuales equivalentes a las exigidas por la LAIP y la normativa de archivo. De lo contrario, podría configurarse una violación de reserva, de protección de datos personales o incluso un supuesto de abuso de autoridad, según las tipologías de faltas graves o gravísimas de la Ley de la Carrera Judicial y de la Ley de Servicio Civil.

En tercer lugar, la carga de datos personales o información reservada en servicios externos de IA generativa. Ingresar nombres de jueces, detalles de quejas disciplinarias, diagnósticos médicos, datos sindicales o información financiera de personal judicial en plataformas públicas de IA implica, por regla general, una exposición innecesaria y contraria al principio de mínima divulgación y a las obligaciones de confidencialidad y reserva de la LAIP. Incluso si el proveedor promete no usar los datos para entrenar modelos, el solo hecho de introducir información reservada fuera del perímetro institucional puede considerarse negligencia grave.

En cuarto lugar, el uso de IA sin citación, revisión ni atribución clara de responsabilidades. Presentar como propios textos generados por IA sin revisión crítica, o sin ajustar el lenguaje a los estándares de motivación y claridad exigidos en resoluciones e informes, no solo puede comprometer la calidad argumentativa, sino también ocultar la cadena de elaboración. Como buena práctica, siempre que el uso de IA sea relevante, debería dejarse constancia interna (incluso si no se hace una mención explícita en el texto final) y garantizar que quien firma asume plenamente el contenido.

Finalmente, debe desaconsejarse el uso de IA para perfiles automatizados de “riesgo disciplinario” de jueces, magistrados o auxiliares, basados en variables sensibles (género, pertenencia étnica, afiliación sindical, etc.). La experiencia internacional ha mostrado que este tipo de herramientas puede derivar en discriminación estructural y en presiones indirectas sobre la independencia judicial.

En todos estos casos, conviene subrayar que no se está “creando” un nuevo ilícito disciplinario por el uso de IA, sino recordando que el uso negligente o abusivo de herramientas tecnológicas se subsume en faltas ya existentes: violación de reserva, negligencia grave, abuso de autoridad, incumplimiento de deberes, entre otras.

7.6 Análisis de cumplimiento: evaluación de impacto, base legal y seguridad

Una cultura de uso responsable de IA en la SGT debería incluir, como parte de sus buenas prácticas, una metodología básica de análisis de cumplimiento, alineada con los estándares emergentes en materia de gobernanza de IA.

En primer lugar, es recomendable realizar, antes de la adopción de herramientas de IA de cierto impacto, una evaluación de impacto (similar, en espíritu, a las “AI impact assessments” que proponen la OCDE y el AI Act). Esta evaluación no requiere una ley nueva, pero sí un análisis interno que considere:

- Finalidad del uso (¿qué problema pretende resolver la IA?),
- categorías de datos involucrados (¿hay datos personales, sensibles, información disciplinaria o clasificada?),
- posibles riesgos para derechos (debido proceso, igualdad, integridad personal, libertad de expresión),
- medidas técnicas y organizativas para mitigarlos (anonimización, límites de acceso, registro de acciones, auditorías periódicas),
- y un plan de revisión y discontinuación en caso de detectar efectos no deseados.

En segundo lugar, toda práctica de IA debe tener una base legal de tratamiento de datos clara. En el contexto guatemalteco, esta base se encuentra en:

- la Constitución y la Ley del Organismo Judicial, que asignan al OJ y a la SGT funciones de supervisión, disciplina y administración;
- la Ley de la Carrera Judicial y la Ley de Servicio Civil, que prevén evaluación del desempeño, investigación de faltas y aplicación de sanciones con respeto al debido proceso;
- la LAIP, que regula acceso, confidencialidad, información reservada y protección de datos personales.

Cualquier uso de IA que suponga tratamiento de datos personales debe encajar en estas finalidades y respetar los límites que la LAIP impone para la clasificación, el plazo de reserva, el hábeas data y la prohibición de comercialización de datos sensibles.

En tercer lugar, es necesario establecer medidas de seguridad y control de versiones. Esto implica:

- Identificar qué versiones de modelos de IA se utilizan (incluyendo actualizaciones o cambios de proveedor);
- documentar quién puede acceder a las herramientas y bajo qué credenciales;
- y garantizar que los registros (logs) se conserven conforme a las normas de archivo del Manual de Procedimientos de la SGT y a la legislación aplicable.

La seguridad no es solo un asunto tecnológico; es también organizativo: perfiles de acceso, capacitación específica, protocolos claros sobre qué información puede o no introducirse en sistemas de IA y mecanismos de reporte cuando se identifique un uso indebido o una brecha.

Por último, el análisis de cumplimiento debe dialogar con el régimen disciplinario existente, no para amenazar, sino para clarificar: el mensaje pedagógico del módulo debería ser que el uso diligente de herramientas de IA puede mejorar la calidad y eficiencia de la supervisión; el uso imprudente o abusivo, en cambio, puede acarrear responsabilidad conforme a las normas vigentes.

7.7 Mapeo comparado de lineamientos y buenas prácticas

Para enriquecer el módulo con perspectivas comparadas, conviene referirse a algunos lineamientos de poderes judiciales y organismos internacionales que ya han comenzado a regular o guiar el uso de IA, incluida la IA generativa, en la justicia.

→ Referentes internacionales y regionales para orientar buenas prácticas

7.8 Carta Ética Europea de la CEPEJ (2018)

La Carta Ética Europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno, adoptada en el marco de la Comisión Europea para la Eficacia de la Justicia (CEPEJ) del Consejo de Europa, se concibe como un instrumento no vinculante orientado a guiar el diseño, adquisición e implementación de herramientas de IA en contextos judiciales y parajudiciales. Su propósito es eminentemente normativo y prudencial: parte del supuesto de que la introducción de IA en justicia puede aportar eficiencia y consistencia, pero también puede tensionar garantías básicas del Estado de derecho si se adopta sin límites, sin gobernanza y sin control humano robusto. Por ello, la Carta opera como un “marco de orientación” que traduce preocupaciones constitucionales y de derechos humanos en exigencias de calidad institucional y de diligencia técnica.

En términos sustantivos, la Carta sintetiza cinco principios rectores: (i) respeto de los derechos fundamentales, como parámetro de compatibilidad material (debido proceso, privacidad, libertad y dignidad); (ii) no discriminación, para evitar la reproducción o amplificación de sesgos; (iii) calidad y seguridad, como exigencia de fiabilidad, robustez y control de riesgos; (iv) transparencia, imparcialidad y equidad, en el sentido de que los sistemas deben ser comprensibles, auditables y no distorsionar la imparcialidad; y (v) control por el usuario, que preserva la capacidad efectiva de los operadores humanos para comprender, cuestionar y apartarse del resultado del sistema. La fuerza de la Carta radica en que estos principios no se presentan como meras declaraciones, sino como criterios para evaluar si un uso concreto de IA es institucionalmente aceptable.

7.9 Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021)

La Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO (2021) es un instrumento no vinculante de soft law adoptado por los Estados Miembros de la organización, y es ampliamente reconocida como el primer marco normativo global de referencia específicamente dedicado a la ética de la IA. Su punto de partida es reconocer que la gobernanza de la IA debe estructurarse desde un enfoque centrado en derechos humanos y dignidad, y no como un mero apéndice técnico de innovación. En esa lógica, la Recomendación articula valores, principios y orientaciones de política pública para minimizar daños previsibles (por ejemplo, riesgos

de discriminación, opacidad o vulneración de privacidad) y para asegurar que el despliegue de IA sea socialmente beneficioso y compatible con obligaciones estatales en materia de derechos.

Un rasgo distintivo del instrumento es su insistencia en el enfoque de ciclo de vida: la responsabilidad ética no se agota en el desempeño del sistema “ya funcionando”, sino que abarca decisiones previas y posteriores (diseño, selección de datos, entrenamiento, validación, despliegue, monitoreo y retiro). En coherencia con ello, la UNESCO promueve herramientas de gobernanza como evaluaciones ex ante, mecanismos de rendición de cuentas y arreglos institucionales que permitan supervisión pública y control social; además, enfatiza dimensiones frecuentemente subestimadas en la regulación sectorial, como educación y alfabetización, inclusión, impacto sobre grupos en situación de vulnerabilidad, y condiciones de seguridad y protección de datos a lo largo de toda la operación del sistema.

7.10 Principios de la OCDE sobre IA y Recomendación del Consejo (2019)

Los Principios de la OCDE sobre inteligencia artificial, formalizados en la Recomendación del Consejo de 2019, constituyen uno de los estándares intergubernamentales más influyentes en la construcción del paradigma de “IA confiable”. Su relevancia no proviene únicamente del contenido, sino del papel que han jugado como lenguaje común para políticas públicas: establecen un núcleo de criterios compartibles entre Estados,

reguladores y actores privados respecto de qué debe considerarse un despliegue aceptable de IA. En su dimensión sustantiva, la OCDE consolida cinco ejes ampliamente citados: promoción del bienestar y crecimiento inclusivo; respeto de valores centrados en el ser humano y la equidad; transparencia y explicabilidad; robustez, seguridad y resiliencia; y responsabilidad de actores humanos a lo largo de la cadena de decisiones.



De forma complementaria, la OCDE acompaña estos principios con recomendaciones operativas dirigidas a los gobiernos, orientadas a crear condiciones institucionales para que la IA sea útil sin degradar garantías. Entre ellas destacan: inversión en innovación y desarrollo, y medición de impactos; construcción de un ecosistema de datos e infraestructura; marcos regulatorios y de gobernanza que reduzcan riesgos (incluida la gestión de calidad, documentación y trazabilidad); fortalecimiento de capacidades humanas (educación, habilidades y preparación del mercado laboral); y cooperación internacional para reducir asimetrías y evitar “carreras hacia el abismo” regulatorio. Este diseño, de principios y orientaciones de política, explica por qué el estándar de la OCDE suele funcionar como puente entre declaraciones éticas generales y aproximaciones de gobernanza concretas.

7.11 Reglamento de IA de la Unión Europea (AI Act)

El Reglamento de Inteligencia Artificial de la Unión Europea (AI Act) se configura como un marco jurídicamente vinculante para dicha región que introduce una arquitectura regulatoria integral basada en un enfoque de riesgo. Su premisa es que no todos los usos de IA exigen la misma intensidad de control: por ello, diferencia prácticas prohibidas (por considerarse incompatibles con derechos y valores), obligaciones de transparencia para ciertos casos, y un régimen reforzado para sistemas de “alto riesgo”. En términos de técnica regulatoria, el AI Act se centra en la gobernanza verificable: gestión de riesgos, documentación técnica, trazabilidad, control humano y mecanismos de supervisión y sanción, todo ello bajo una lógica de cumplimiento ex ante y monitoreo ex post.

En el componente de alto riesgo, el Reglamento exige protecciones como sistemas de gestión de riesgos, calidad y gobernanza de datos, documentación técnica, registro y conservación de logs, información clara para usuarios/operadores, medidas de supervisión humana, y estándares de precisión, robustez y ciberseguridad; además, incorpora un esquema institucional europeo para coordinación y aplicación. Junto a ello, el AI Act aborda de manera específica el fenómeno de la IA generativa y de propósito general mediante obligaciones de transparencia y deberes de información que buscan reducir asimetrías y riesgos sistémicos. El resultado es un modelo regulatorio que trata la IA como infraestructura sociotécnica con impactos sobre derechos, y no como simple herramienta neutral, lo que explica su influencia extraterritorial como referencia comparada.

Resolução CNJ n. 332/2020 (IA en el Poder Judicial)

La Resolução n. 332/2020 del Conselho Nacional de Justiça (CNJ) fija directrices para el desarrollo, adopción y uso de modelos de inteligencia artificial en el Poder Judicial brasileño, con un acento marcado en compatibilidad con derechos fundamentales y en parámetros éticos de transparencia, previsibilidad, auditabilidad, imparcialidad y justicia sustancial. En su parte considerativa, la norma explicita preocupaciones típicas de gobernanza pública: la necesidad de minimizar errores y prejuicios derivados de datos históricos, de asegurar que las tecnologías operen con fuentes seguras y rastreables, y de preservar la privacidad mediante control informado sobre datos personales. También incorpora un diálogo explícito con referentes europeos al mencionar la Carta ética del CEPEJ como antecedente orientador.

En su dimensión operativa, la Resolução despliega reglas de control que son relevantes por su concreción: antes de poner un modelo en producción exige homologación para identificar sesgos o generalizaciones y ordena medidas correctivas, incluyendo la discontinuación si el sesgo no puede eliminarse; además, define transparencia como un paquete de deberes, incluyendo documentación de riesgos, mecanismos de auditoría y, de forma especialmente exigente, la obligación de contar con una explicación satisfactoria y auditable por autoridad humana respecto de cualquier propuesta de decisión generada por IA. En paralelo, refuerza el control por el usuario ya que la propuesta del sistema no debe vincular, debe poder

revisarse, y los usuarios externos deben ser informados de manera clara sobre el carácter no vinculante de la solución sugerida y su sometimiento a autoridad competente.

Acuerdo PCSJA24-12243 (Plan Especial de IA Generativa para la Rama Judicial)

El Acuerdo PCSJA24-12243 se inserta en la discusión institucional colombiana sobre cómo introducir IA generativa en la Rama Judicial sin comprometer estándares de responsabilidad pública, seguridad de la información y garantías procesales. Su relevancia comparada radica en que no parte de la idea de automatizar decisiones jurisdiccionales, sino de encauzar el fenómeno mediante un plan especial, es decir, un marco administrativo de orientación que reconoce tanto el potencial de apoyo (por ejemplo, asistencia en tareas de trabajo intelectual rutinarias) como los riesgos asociados a errores, sesgos y tratamientos indebidos de información. En esa medida, el Acuerdo se presenta como una respuesta de gobernanza. En ese sentido busca, primero, ordenar, delimitar y supervisar; luego escalar usos conforme se consoliden capacidades institucionales y controles.

En términos de contenido, el plan se describe públicamente como orientado a promover un uso responsable, seguro y ético de la IA generativa, con especial énfasis en reglas de prudencia sobre información, responsabilidades humanas y adopción gradual. En particular, se ha destacado que el enfoque prioriza usos

de apoyo (por ejemplo, elaboración de borradores y asistencias instrumentales) bajo supervisión humana, evitando comprender la IA como sustituto del juicio institucional. Esta aproximación es consistente con tendencias comparadas: la legitimidad del uso de IA en justicia depende menos de la novedad tecnológica y más de la calidad de sus salvaguardas (gobernanza, trazabilidad, control, y reglas claras de responsabilidad). El énfasis está en el carácter experimental y supervisado de la IA generativa, y en la necesidad de capacitar a jueces y funcionarios antes de un despliegue amplio.

Para Guatemala, estos referentes internacionales y regionales no son jurídicamente vinculantes, pero constituyen un canon persuasivo que puede orientar la elaboración de protocolos internos, guías de buenas prácticas y contenidos de capacitación. El módulo puede aprovecharlos para:

- Mostrar que la preocupación por el uso responsable de IA en justicia es global;
- subrayar que el objetivo no es “prohibir la tecnología”, sino usarla de manera compatible con la independencia, la imparcialidad y la integridad judicial;
- y ofrecer a la SGT criterios comparados para evaluar propuestas de herramientas, contratos tecnológicos o proyectos piloto que involucren IA.

Al articular estos estándares con el bloque interno guatemalteco (Constitución, LOJ, LCJ y Reglamento, Normas Éticas, LAIP, Manual de la SGT y Pacto Colectivo), el TEMA 7 del módulo puede transmitir un mensaje claro: la IA es una herramienta potente que, si se utiliza bajo control humano significativo, con respeto a la confidencialidad y con trazabilidad, puede mejorar el trabajo de supervisión y prevención; pero, si se usa de forma acrítica o no gobernada, puede agravar problemas ya identificados por la CIDH y la ONU en materia de captura, presiones internas y erosión de la confianza pública en la justicia guatemalteca.