Una experiencia real de uso de la IA en el análisis de fuentes abiertas: la factoría de análisis del Centro de Análisis y Prospectiva de la Guardia Civil

Enrique Ávila Gómez
Lorenzo Luna Zambrana

Resumen

El documento examina el uso de la inteligencia artificial (IA) en el análisis de inteligencia de fuentes abiertas, enfocándose en la experiencia del Centro de Análisis y Prospectiva (CAP) de la Guardia Civil.

Aborda cómo la IA generativa transforma la detección, selección y análisis de información, permitiendo un acceso más rápido y amplio a fuentes diversas.

Sin embargo, se señalan los desafíos, en especial, las «alucinaciones» de la IA, y se exploran estrategias para mitigar estos riesgos, como el uso de grafos de conocimiento y la validación humana.

El texto destaca la importancia del talento humano en la revisión y verificación de la información, así como la necesidad de adaptarse al cambio tecnológico para mejorar la precisión y eficiencia en la toma de decisiones estratégicas.

Además, enfatiza la necesidad de una visión holística y multidisciplinaria en el análisis de inteligencia, incorporando diversos modelos culturales para reducir sesgos.

Palabras clave

Prospectiva, Análisis, IA, Alucinaciones, Sesgos.

A real experience of the use of ia in open source analysis: the analysis factory of the Civil Guard's analysis and forecasting centre

Abstract

The document examines the use of Artificial Intelligence (AI) in open source intelligence analysis, focusing on the experience of the Center for Analysis and Foresight (CAP) of the Guardia Civil.

It addresses how generative AI transforms the detection, selection and analysis of information, allowing faster and broader access to diverse sources.

However, challenges are pointed out, especially AI «hallucinations», and strategies to mitigate these risks are explored, such as the use of knowledge graphs and human validation.

The text highlights the importance of human talent in the review and verification of information, as well as the need to adapt to technological change to improve accuracy and efficiency in strategic decision-making.

In addition, it emphasizes the need for a holistic and multidisciplinary vision in intelligence analysis, incorporating various cultural models to reduce biases.

Keywords

Prospective, Analysis, AI, Hallucinations, Biases.

1 Introducción

En este artículo se va a descender a territorios táctico-operativos en el mundo de la IA aplicada a los procesos de análisis de inteligencia orientada a la toma de decisión estratégica, abandonando la perenne tendencia hacia la descripción de niveles estratégicos que no permiten acercarse a este nuevo reto tecnológico de forma holística y con un objetivo de aplicabilidad práctica en sectores diversos entre los que, por supuesto, se encuentra el sector de la Defensa, ocupando, además, un lugar cada vez más central, en un entorno geopolítico complejo, cambiante y con la necesidad de herramientas que permitan su rápida adaptación al escenario.

Las estructuras sociales, políticas, económicas, de nuestras sociedades son cada vez más complejas. Cada avance tecnológico ha introducido, en un modelo de capas superpuestas, una nueva capa de complejidad. La superestructura generada ha supuesto, a nuestro juicio, una barrera en la comprensión de un sistema que, además, requiere de una atención continua, cuya interconexión provoca efectos desconocidos y, a menudo, indeseado entre las diversas capas y en el que, más grave aún, se ha perdido una parte importante de la comprensión del funcionamiento de esta superestructura por imposibilidad intelectual del ser humano para el manejo de este nivel de complejidad, tanto a nivel individual como colectiva.

La eclosión de la inteligencia artificial (no solo la generativa) ha abierto un conjunto de posibilidades, a menudo inaccesibles hasta hace bien poco, a los equipos de análisis de información de fuentes abiertas que elaboran los informes que habrán de ayudar a la toma de decisión en los niveles de decisión estratégica, es decir, a interpretar esa realidad compleja de una forma adecuada para que el decisor pueda ejercer su principal misión.

Las limitaciones existentes, hasta hace bien poco, tanto de índole temporal como de acceso al talento especializado disponible en las fases de detección de activos de información de interés y, más aún, en las fases de análisis y diseminación de información eran cada vez más evidentes para cualquier unidad de análisis de inteligencia de fuente abierta. La enorme dinamicidad y complejidad del mundo actual determinaban, sin duda, una limitación estructural a la cantidad y calidad de los análisis que debían de servir de base para la decisión estratégica.

No menos importante resultaba la progresiva barrera que los sistemas de diseminación automatizada introducían por, a menudo, inundación de información supuestamente relevante (y probablemente lo es) que provocaba un progresivo «desenganche» del decisor del acceso a la información elaborada.

Este artículo describe la experiencia práctica que se ha estado desarrollando desde el Centro de Análisis y Prospectiva de la Guardia Civil para

configurar un ecosistema de detección, selección, análisis y diseminación de información de interés para la Dirección Estratégica, en materia de Seguridad Interior, prácticamente, en tiempo real, con intervención humana mínima y orientada esta, en específico, a la validación de los contenidos en las fases de selección y verificación final de contenido que, se considera, será la principal tarea de futuro de muchas profesiones, entre ellas, la de analista.

Se ha intentado, además, generar un modelo de acceso y diseminación de la información generada que permita segmentar al receptor de la información, limitando (que no eliminando) la decisión de acceso sobre los activos de información recibida a través de las herramientas tecnológicas habituales en este modelo y que, en demasiadas ocasiones, provocaban un colapso por inundación.

2 El Centro de Análisis y Prospectiva de la Guardia Civil (CAP)

Hace ahora 31 años, en un acuerdo innovador entre el mundo de la Academia y la Guardia Civil, se procedió a la creación del Centro de Análisis y Prospectiva de la Guardia Civil (García López, 2010: 45-60). Un hecho sin precedentes, en la historia de nuestro país, el que fuese una Fuerza de Seguridad del Estado la que detectase la necesidad de anticiparse a las amenazas emergentes haciendo uso de la prospectiva y, desde un posicionamiento proactivo, intentase desarrollar el diseño de futuros (probables, plausibles o posibles) a partir de un conocimiento exhaustivo del presente, apoyándose en las herramientas tecnológicas y conceptuales disponibles en cada instante.

Desde aquel momento de génesis hasta el momento actual, el Centro ha pasado por varios modelos de Dirección Estratégica, más o menos orientados a la generación de productos de corte científico-técnico, siguiendo modelos clásicos de detección de información de interés, clasificación, análisis, generación de base documental y diseminación de esta, por los circuitos definidos interna y externamente, por parte de la Institución.

La eclosión de Internet y el acceso global a fuentes de interés para la elaboración de Inteligencia de fuentes abiertas generó la necesidad de dotarse de un conjunto de herramientas tecnológicas que configurasen un ecosistema de tratamiento de la información que permitiese, fundamentalmente, adaptarse a los tiempos, cada vez más inmediatos, en los que es preciso contar con información relevante para su tratamiento y conversión en conocimiento interno y orientado, en el caso del CAP a la toma de decisión estratégica.

De forma paralela, el CAP ha sido centro de detección (y algunas veces de posterior reclutamiento vía oposición) de talentos especiales. Configurado

como un elemento de conexión con la educación superior de nuestro país, sus programas de prácticas, considerados de un nivel de excelencia por las universidades colaboradoras y por el estudiantado que ha pasado por esa experiencia, ha permitido incorporar visiones, experiencias y capacidades desarrolladas por las nuevas cohortes de ciudadanía de nivel universitario.

La incorporación temporal de talento joven, con ideas exógenas a la Institución, brillante, a menudo con experiencia internacional, se ha configurado como un excelente marco de trabajo y de generación de capacidades de análisis que, entre otras externalidades positivas generadas, nos permiten orientar el futuro de las personas que acaban sus prácticas hacia nichos de inteligencia que no se limitan al Estado, sino que se amplían hacia nuestros sectores productivos que, hace tiempo ya, hacen un uso intensivo de estas capacidades en sus procesos de Inteligencia Competitiva e Inteligencia Económica.

Algunos de aquellos y aquellas primeras personas que realizaron parte de sus prácticas profesionales en el CAP son, por tanto, parte de las unidades de Inteligencia que se han ido desplegando, sobre todo, en el nicho de las grandes estructuras corporativas nacionales.

Por último, el CAP tiene encomendada la misión de generar «cultura de la seguridad». Orientada su actividad a mejorar la resiliencia social a través de la interacción con la ciudadanía, utilizando diversos canales, tanto analógicos como digitales, y adaptando los contenidos para impactar sobre grupos poblacionales concretos, a lo largo de los últimos cinco años, se ha participado en actividades con universidades, generado documento y creado y mantenido canales de interacción con la Academia, el sector privado, otras Instituciones y la propia ciudadanía, incorporando tecnologías tales como metaversos, metahumanos o Inteligencia Artificial para la interacción directa con los grupos específicamente seleccionados como de interés (Soliman, 2024).

Conocida la misión del CAP y descritas las herramientas de que dispone para llevarla a cabo, nos introducimos en el proceso de configuración del ecosistema que en la actualidad se está desplegando para mejorar la calidad y los tiempos de respuesta de generación de productos de Inteligencia adaptados a las necesidades del decisor estratégico.

3 Escenario previo: idioma, detección, validación y tiempo

La misión actual de una unidad dedicada al análisis de fuentes abiertas es un auténtico infierno. Detectar, valorar, analizar y diseminar la infinita cantidad de información generada a cada instante no está al alcance de las capacidades humanas, con independencia del número o la calidad de los recursos humanos dedicados a esta misión.

Hace no tanto, las capacidades a desarrollar por parte de los analistas en la primera fase del proceso de detección de información relevante o de interés para la toma de decisión eran, fundamentalmente, la de conocer los repositorios de fuentes de interés (un puñado de revistas científicas, unas cuantas bibliotecas seleccionadas, algún informe de fuente más o menos validada y conocida...).

Este tipo de capacidad se transformó, con el tiempo y con el apoyo de tecnología básica de etiquetado y clasificación de la información y, con suerte, el uso de algún gestor documental con capacidad de indexación de contenidos (se habla de hace no mucho más de dos años el límite temporal en que todas capacidades del analista se podían considerar las más valiosas para el análisis de documento de fuente abierta), en un proceso más o menos automatizado de clasificación y recuperación de documentos de interés por parte de la persona o grupo interesado en algún aspecto relacionado con la Seguridad Interior de nuestro país.

Básicamente, se generaban productos de inteligencia desde fuentes acotadas, consideradas, por su inserción en determinados índices, como de validez absoluta y sin necesidad de análisis en cuanto a su calidad. Esta metodología, de validez contrastada, no impedía, en todo caso la aparición de sesgos o de información incompleta dado que se perdía el acceso a otras corrientes de pensamiento desestimadas por la propia Academia.

Los resultados, como decimos, a menudo probablemente sesgados (sí, ahora se habla de sesgos en la IA, pero no se piensa en los sesgos inducidos por equipos de análisis pequeños, con capacidades pluridisciplinares muy limitadas y con bajo acceso a expertos y expertas de conocimientos no directamente conexos con el asunto de interés que se estuviera analizando (Mosqueira-Rey, 2023). Eso por no hablar de las limitaciones de enfoque inducidas por las limitaciones de acceso a otras áreas geopolíticas y culturales por problemas de idioma), se podrían considerar como óptimos para los recursos disponibles en ese momento, pero claramente insuficientes para entender y, menos aún operar, sobre la complejidad e inmediatez en los procesos de toma de decisión que requiere el mundo actual.

En 2013, a través de un proyecto de investigación del IUISI (Instituto Universitario de Investigación de Seguridad Interior, lamentablemente desaparecido), se detecta este inicial problema en el acceso temprano a información de interés, dentro y fuera de las corrientes de pensamiento principales y se procedió a estudiar el viaje de ida que se había producido desde el mundo de la Inteligencia Estratégica al de la Inteligencia Competitiva y la Económica y que había permitido el desarrollo de nuevas metodología, así como la aplicación de herramientas tecnológicas avanzadas por parte de corporaciones con capacidades económicas comparables a las de los Estados.

De ese modo y con el objetivo de planificar un viaje de vuelta hacia los modelos de Inteligencia Estratégica, de ámbito estatal, de las metodologías y herramientas que las grandes corporaciones habían desarrollado para la generación y uso de su propia Inteligencia, con objetivos sectoriales y eminentemente económicos, incorporamos a nuestro propio modelo de conocimiento los procesos de Vigilancia Tecnológica[1].

Durante el desarrollo de ese proceso de investigación, no nos centramos tanto en la generación de conocimiento documental ni, menos aún, académico, sino que se trabaja en el diseño y despliegue de un ecosistema tecnológico de adquisición, validación y análisis de información de interés a partir de la definición y evolución de un sistema de fuentes orientado a la detección temprana de amenazas emergentes relacionadas con tecnologías de doble uso (en realidad, se podría decir que todas aunque, en aquel momento nos centramos en drónica y biotecnología instrumental desarrollada por grupos de *biohacking*)

Se descubre la existencia de una plataforma tecnológica de *software* libre, desarrollada con el apoyo de las administraciones públicas que permitía automatizar la adquisición de fuentes de interés desde Internet, almacenar una parte de la información de forma local, validar su interés de forma sencilla por parte de los equipos de análisis, ampliarla gracias un trabajo posterior de análisis en profundidad y diseminarla, de forma automatizada, haciendo uso de sistemas de boletinado mediante suscripción por parte de los receptores interesados en un determinado área de conocimiento.

La plataforma HONTZA[2] estaba diseñada para la ejecución de un ciclo completo de Inteligencia Competitiva que abarcaba desde el diseño inicial de las estrategias de la organización y de detección de fuentes relevantes para la misma, hasta la diseminación final automatizada de la información, haciendo uso de boletinado mediante suscripción, pasando por los procesos de transformación de la información en Inteligencia, con la actividad de los equipos de analistas desarrollándose en un ecosistema de red social interna y que incluía la coordinación y gestión de redes de expertos externos a la organización, así como un sistema de propuestas y gestión de proyectos asociados a una determinada detección de interés durante el proceso de vigilancia tecnológica.

[1] La Vigilancia Tecnológica es un proceso sistemático que permite a las organizaciones detectar, seleccionar, analizar y diseminar información relevante sobre avances tecnológicos y científicos. Este proceso es fundamental para anticiparse a las amenazas emergentes y tomar decisiones estratégicas informadas.

[2] La plataforma HONTZA permite a los equipos de trabajo colaborar de manera eficiente, automatizando procesos de vigilancia tecnológica y facilitando la gestión de información relevante. Entre sus características destacan la capacidad de importar canales de información, generar documentos colaborativos, y analizar la evolución de las noticias. La idea de HONTZA surgió para ofrecer una alternativa a los *software* de vigilancia tecnológica de código cerrado, que suelen ser costosos y limitan la personalización y colaboración entre usuarios. El proyecto fue impulsado por asociaciones empresariales del País Vasco y financiado por la Sociedad para la Transformación Competitiva del Gobierno Vasco (SPRI).

Con un equipo de investigación formado por al teniente coronel Arturo Crespo Arranz, al teniente coronel Santiago Alonso Pradillo y la investigadora y experta en Servicios de Inteligencia María Cuiñas Insúa, se opta por el desarrollo de un ecosistema orientado a la formación de analistas de inteligencia y explotación intensiva de la información de fuente abierta de interés, apoyándonos en esta plataforma.

Esa es la base de un moderno ecosistema de generación de Inteligencia: ser capaces de detectar y adquirir, de forma automatizada y en tiempo real, información de interés para la elaboración de conocimiento e inteligencia sobre asuntos de interés estratégico (o táctico) de la organización.

Se hablaba de 2013-2014 y, por supuesto, durante los siguientes diez años, las herramientas orientadas a este tipo de procesos mejoraron en todo lo referente a las capacidades de acceso, detección, indexación y multimodalidad. Ya no se trabajaba solo con elementos textuales, sino que la tecnología permitió el acceso a conocimiento multimodal de todo tipo: información desestructurada, audio, vídeo, etc. (AENOR, 2018).

Es a partir de finales de 2023 cuando entramos en un nuevo escenario: se empiezan a liberar servicios de inteligencia artificial generativa que, al principio sobre base documental textual, pero, poco a poco, con capacidades de multimodalidad, empieza a hacer posible la automatización de varios de los procesos de análisis provocando la evolución del modelo preexistente de forma radical (Storey, 2025).

¿Qué problemas se encontraban en el modelo anterior? Básicamente, se pueden citar cuatro: el tiempo, la enorme cantidad de fuentes disponibles, el idioma y el talento pluridisciplinar al que se puede acceder por parte de cualquier organización. Y el primero, el del tiempo, además, fuertemente mediatizado por todos los demás.

El idioma siempre ha sido un cuello de botella de cualquier organismo que genere informes para un decisor estratégico. En el caso de la elaboración de Inteligencia, además, la no incorporación en el análisis de diversos enfoques pluridisciplinares, multiidioma y multiculturales, induce un conjunto de sesgos que impiden disponer del conocimiento adecuado para el manejo de situaciones de complejidad creciente. Más aún en una sociedad fuertemente globalizada e interdependiente en el que el impacto de diferentes modelos culturales en los procesos de toma de decisión puede llegar a tener que dar por completo la vuelta al resultado de un determinado análisis. Un ejemplo práctico de esta aseveración se puede encontrarlo en el Intelligence Practitioner's HANDBOOK (New Zealand Institute of Intelligence Professionals, 2024), en el que se introducen elementos de la cultura mahorí en los procesos de análisis de inteligencia que aparecen en este interesante libro[3].

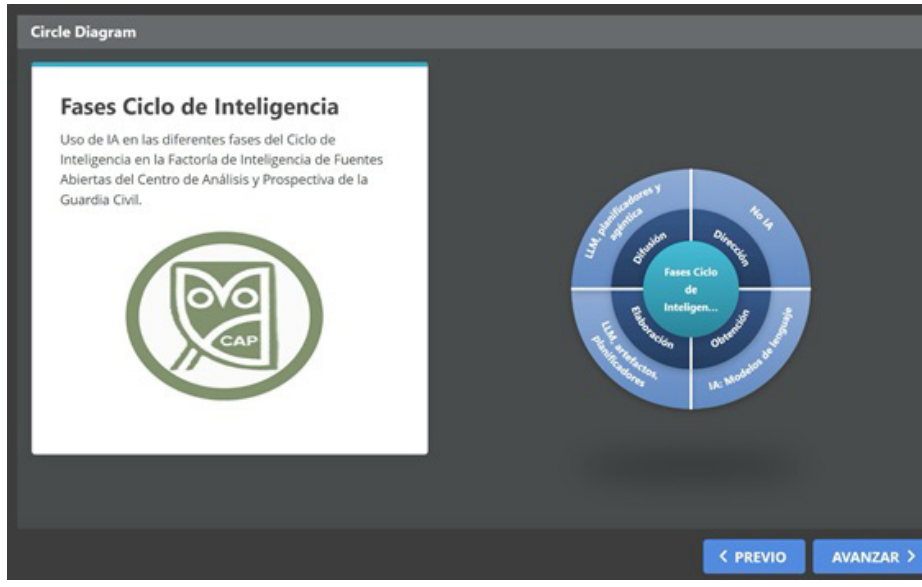
[3] Para una acceso analítico al contenido de este se puede utilizar esta URL: https://www.ciberinteligencia.eu/new-zealand-intelligence-principles-and-practice/ que, además,

A menudo, esta disfunción limita fuertemente la posibilidad de realizar recomendaciones acertadas al decisor. Intentar no caer en este error significa, en un modelo clásico, pasar por farragosos procesos de traducción e interpretación cultural, usando capacidades humanas, muchas veces indisponibles, que, gracias a herramientas de traducción automática han permitido mitigar, en parte, esta primera barrera.

Pero es el momento en que la IA generativa hace su aparición cuando se produce la apertura del foco de acceso de un equipo de Inteligencia hacia un modelo que podría denominarse como 360.

Desaparecen, por ensalmo, las barreras idiomáticas y se reducen, drásticamente, las temporales, a la hora de analizar múltiples fuentes, en diferentes idiomas, en un nuevo proceso dialogado con las fuentes documentales a las que diferentes expertos, en diferentes materias, pueden acceder de forma sencilla y usando lenguaje natural para su «diálogo» con ese conjunto de fuentes.

Es este el momento en el que se empieza a trabajar en la introducción de las herramientas de IA que se van desarrollando en las diferentes fases del Ciclo de Inteligencia clásico. A continuación, se presenta un gráfico que intenta exponer esta implementación. Para un acceso directo al gráfico interactivo se puede usar la URL: https://ciberinteligencia.eu/interacciones/ciclo/.



constituye un ejemplo práctico del modelo de análisis que se está desplegando en el CAP, incluyendo los productos generados para decisor estratégico.

Más adelante, se tratará el problema del talento disponible y de las nuevas capacidades que han de desarrollarse para su progresiva adaptación al nuevo escenario tecnológico porque se trata de un problema de difícil solución inmediata y que requiere modificaciones sustanciales en asuntos relacionados con formación especializada, cultura o rotaciones acotadas.

Nuestro Centro ha ido estudiando y probando durante los dos últimos años, los servicios de inteligencia artificial que se han considerado de interés para mejorar la eficiencia de nuestra misión, con el objetivo de, en la medida de lo posible, transformar el proceso de Ciclo de Inteligencia en un modelo lo más automatizado e industrializado posible, siempre orientado a ofrecer al decisor estratégico la inteligencia que necesita, en el momento preciso, para que sea incorporada en sus procesos de toma de decisión y, por supuesto, preservando el necesario espacio de intervención humana en el último paso de validación de contenidos.

Entre los servicios que se han utilizado, se pueden incluir los siguientes:

— NotebookLM, de Google. Ampliamente utilizado en el mundo académico, permite generar materiales de interés para los procesos de análisis tales como

- Documentos generados desde fuentes documentales desestructuradas de difícil análisis por métodos tradicionales:

  fi Fuentes de audio y vídeo: ya no es necesario un operador humano visualizando/escuchando durante horas una o varias fuentes. Se realiza una transcripción automática y la generación de los productos básicos de este sistema:

  a) Documento resumen.

  b) Preguntas frecuentes.

  c) Fichero resumen de audio en formato podcast.

- Los materiales generados sirven como punto de partida para el «diálogo» con el sistema de fuentes, de tal manera que, una vez adquirido el conocimiento necesario sobre el asunto de interés, el analista, a través de este proceso de «diálogo» con la base documental, es capaz de orientar de manera aedcuada el *prompting* a través de una adecuada contextualización y orientación hacia un público objetivo específico.

- El producto final es una base documental textual, de no más de tres páginas, que delimita el problema e informa al decisor de las cuestiones de interés que deberían ser integradas en el proceso de decisión.

— Google AI Studio. Esta plataforma incorpora los nuevos modelos de razonamiento profundo de esta empresa. Sus capacidades de planificación, razonamiento y búsqueda de información, señalando las fuentes y permitiendo al analista definir cambios en los procesos de planificación de la investigación convierten a esta herramienta en invaluable para un equipo de analistas. La posibilidad de generación de productos de inteligencia tanto documentales clásicos como en otros formatos, más accesibles y comprensibles, incluso, interactivos son, a nuestro juicio un anticipo ya muy avanzado, de un conjunto de agentes de IA que van a desplazar al ser humano del centro de los procesos de análisis[4].

— Copilot de Microsoft. Sus capacidades de integración con procesos a través de la inclusión de agentes y, por supuesto, con herramientas ofimáticas estándar han sido, sin duda, de enorme valor en la integración y confección de productos de interés para el análisis de Inteligencia. Alto impacto en los procesos formativos. Generar los materiales para la realización de una actividad de formación desplegable en una plataforma de teleformación se ha convertido en una actividad de horas y no de meses. Se ha conseguido generar dos cursos de teleformación interna para analistas, en materia de análisis de inteligencia de fuente abierta y en materia de diseño de futuros en menos de una semana. Partiendo de un esquema generado desde un modelo de pensamiento profundo, el uso de Copilot para la generación de presentaciones y otros productos de contenido formativo, para ser desplegados, en formato SCORM, en la plataforma de teleformación corporativa ha resultado una experiencia altamente eficiente. Cualquier conocimiento se encuentra accesible. Cualquier conocimiento puede ser rápidamente transformado en una acción formativa estructurada.

— Deepseek. Se ha incorporado este modelo para disponer de respuestas generadas a partir de otros conjuntos de datos provenientes de otras áreas geopolíticas. Poder comparar los enfoques, calidad de la respuesta y espectro ideológico del modelo ha permitido entender un contexto complejo, así como detectar ciertas amenazas emergentes relacionas con sesgos y determinación de entidades. Este modelo es computacionalmente utilizable con recursos limitados, de tal manera que es posible aislar contenedores de información que haya de ser protegida y no expuesta a sistemas en la nube sobre los

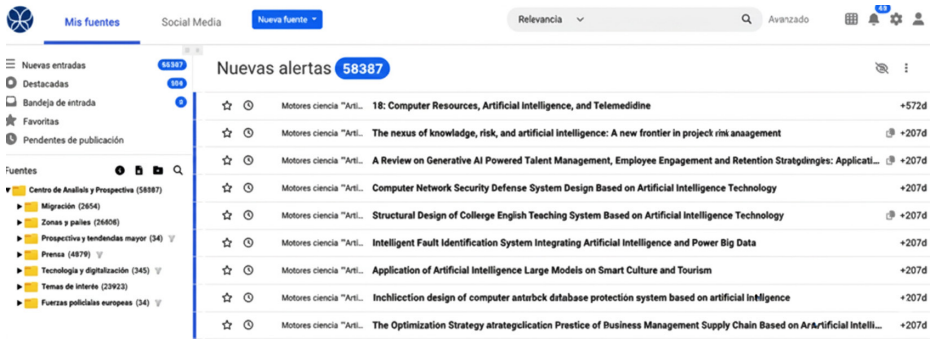
[4] Para ver una «prueba de vida» de esta controvertida aseveración se puede acceder a este ejemplo de análisis, en bruto, que trae razón de la noticia: https://edition.cnn.com/2025/06/03/business/trump-tariffs-china-pharmaceuticals-intl-hnk. A partir de esta amenaza detectada por un medio global, se procede a usar el modelo Gemini 2.5 Pro y su utilidad de generación de informe en profundidad para obtener el siguiente resultado: https://www.ciberinteligencia.eu/informe-impacto-amoxicilina/.

que se pueda perder el control. Su código abierto permite la realización de actividades de *fine tunning* que adapte el campo semántico a las necesidades de la organización, minimizando la posibilidad de que se produzcan las indeseadas alucinaciones.

— Qwen3. Este modelo se caracteriza por su poderosa multicanalidad y sus capacidades de generación de productos diversos incluyendo audio, vídeo, imágenes, artefactos o código, entre otras funcionalidades. El análisis de información desestructurada multimedia ha sido nuestro principal uso al determinar que su calidad era superior a la de los otros modelos.

— Manus. Esto no es un modelo de lenguaje sino un sistema multiagéntico con capacidad de interacción con el sistema operativo. Esto es: puede no solo generar un plan, sino, además, ejecutarlo en forma de tareas complejas. Nos ha interesado, sobre todo, el concepto. La capacidad agéntica de ejecución de tareas complejas constituye, a nuestro juicio, un punto de inflexión. Más aún en entornos que requieran modos de funcionamiento de detección, análisis y respuesta en, prácticamente, tiempo real. Por ejemplo, los de teatro de operaciones. Se trata de un modelo que se define como aún experimental pero que está evolucionando de forma rápida.

La fase temprana de detección de contenidos de interés ya utiliza servicios de IA para facilitar la generación de un sistema de fuentes coherente y orientado a los objetivos descritos por el analista. Esta actividad constituía, históricamente, una de las principales fuentes de error y de discusión por parte del equipo de analistas. En la actualidad, se transforma en un proceso semiguiado por un agente planificador al que, a través del lenguaje natural, informas de cuáles son tus necesidades de información para, a partir de ese nivel, ir profundizando en la selección de las fuentes de interés. El analista pasa a ser un validador de las decisiones del planificador y, para ello, ha de armarse de herramientas culturales que le permitan tomar ese tipo de decisiones.



La fase de análisis a través del «diálogo» con las fuentes objetivo contextualizando los objetivos y el público objetivo al que irá dirigido el análisis, haciendo uso de varias herramientas de inteligencia artificial para, en una última fase, analizar el resultado de los productos obtenidos como nuevas fuentes de información sobre las que, en caso de ser necesario volver a trabajar, ya de manera directa, por parte de los equipos de analistas humanos. (Bosso y Ruberto, 2025).

El analista profesional realizará una primera crítica radical a este modelo apoyándose, a buen seguro, en argumentos conexos con la experiencia profesional y la necesidad de precisión en la respuesta. Argumentará, del mismo modo, sobre la posibilidad de error que toda tecnología introduce en los procesos que se consideran críticos y fuertemente intelectualizados en los que, a menudo, la experiencia profesional es la que permite disponer de ese «momento de inspiración» que parece caracterizar al ser humano y colocarle en un nivel aún superior al de la tecnología.

Sobre consecuencias nefastas de esos «momentos de inspiración humana» a lo largo de la historia, así como su falibilidad se podría mantener un interesante debate que trasciende los objetivos de este documento, ello sin entrar en los «intereses» que pueden mediatizar las recomendaciones de un grupo de análisis.

No obstante, se puede señalar que la falibilidad de la inspiración humana, muy conectada con nuestros sesgos (Kahneman, 2012), es un elemento fundamental a la hora de configurar los equipos de análisis, así como a la hora de desarrollar sus nuevas capacidades. Veamos qué nos dice al respecto Gemini (IA, 2025):

— «**Sesgos cognitivos**: La inspiración, a menudo, surge de intuiciones y corazonadas, que pueden estar influenciadas por sesgos cognitivos. Estos sesgos, como el sesgo de confirmación (buscar información

que confirme nuestras creencias preexistentes) o el efecto halo (dejar que una sola característica positiva influya en nuestra percepción general), pueden distorsionar nuestra evaluación de la situación y llevarnos a tomar decisiones basadas en información incompleta o engañosa.

— **Falta de objetividad**: La inspiración puede ser subjetiva y emocional, lo que puede dificultar la objetividad en el análisis. Cuando nos dejamos llevar por la pasión o el entusiasmo, podemos pasar por alto detalles importantes o ignorar riesgos potenciales.

— **Exceso de confianza**: La inspiración puede generar un exceso de confianza en nuestras ideas y juicios. Cuando estamos convencidos de que tenemos la razón, podemos ser menos propensos a escuchar a los demás o a considerar diferentes perspectivas. Este exceso de confianza puede llevarnos a tomar decisiones arriesgadas o imprudentes».

No podemos, en todo caso, sino señalar nuestro acuerdo parcial con ese tipo de críticas, por parte de los profesionales de la Inteligencia, a la hora de hacer uso de las herramientas descritas, desde el plano estrictamente intelectual clásico. No obstante, el enorme cambio de escenario que impone la inteligencia artificial aplicada al análisis de inteligencia induce la necesidad de posicionarnos de manera crítica (constructiva) frente a las consideraciones de un modelo de trabajo «clásico».

Estos serían los argumentos que se consideran esenciales para defender nuestro posicionamiento a favor del uso de la IA en las fases del ciclo de inteligencia:

— ¿Acaso los seres humanos no cometemos errores de percepción evaluación respuesta al realizar análisis? Creo que es indiscutible nuestro sesgo individual y colectivo a la hora de evaluar situaciones de interés. Los sesgos individuales (educacionales, ideológicos, de género) impactan sobre nuestros análisis. Pero también lo hacen los de grupo a través de liderazgos y relaciones de poder en los equipos.

— ¿La comprensión de la complejidad actual de cualquier sistema humano se encuentra al alcance de una persona (o de un equipo), con limitaciones en el acceso a todo el conocimiento necesario para realizar una evaluación correcta de una amenaza compleja? No se puede a aspirar a entender estructuras tan complejas como las que configuran las realidades actuales. No disponemos de capacidades intelectuales ni individuales ni colectivas para lograrlo. La IA ofrece una oportunidad de simplificación para la comprensión de la complejidad.

— ¿El tiempo necesario para la comprensión y evaluación de una amenaza en el mundo actual se adapta a los tiempos de análisis de un equipo humano y, sobre todo, a la necesidad temporal de respuesta que requiere el decisor estratégico? Sin duda, no. Ya se ha adaptado gran parte de los procesos de toma de decisión actuales a la realidad algorítmica. Y se va a continuar por ese sendero porque no hay una posibilidad alternativa. No se va a parar en el camino de la complejidad y el uso de la tecnología para intentar tomar decisiones, en tiempo real, sobre esa complejidad.

Consideramos los autores que la respuesta a las tres preguntas planteadas es negativa y por ello hemos operado de forma diferente desde nuestro Centro de Análisis y Prospectiva.

Por los motivos señalados, las capacidades humanas, limitadas en cantidad y tiempo, son usadas solo en dos de las fases de elaboración del producto de inteligencia:

— Detección, validación y mantenimiento del sistema de adquisición automatizada de fuentes.

— En la fase de validación del producto obtenido, justo antes de ser diseminado.

Hemos entrenado a nuestros analistas para realizar comprobaciones de estructura argumental y verificación de fuentes con el fin de detectar, de forma temprana, lo que conocemos como «alucinaciones» de la inteligencia artificial, por cierto, nada que no sufra cualquier analista, a lo largo de su trayectoria profesional, en múltiples ocasiones.

No se puede dejar de señalar, en todo caso, que el riesgo se ha detectado más alto en función de ciertos parámetros relacionados con el uso de Inteligencias Artificiales generativas de propósito general como pueden ser:

— Todo lo relacionado con entidades de diversa tipología (personas, lugares, tiempos)

— Información actualizada sobre un determinado asunto de nicho.

Siendo lo antedicho cierto, nos hemos encontrado durante la última fase de desarrollo de nuestra Factoría de Inteligencia, en el segundo semestre de 2024 y primer trimestre de 2025, con un inesperado aliado: la competencia geopolítica (o geoeconómica).

Hemos sido testigos, desde finales de 2025 y, sobre todo, a lo largo de lo que llevamos de 2025, de la competición global por ser referentes en materia de inteligencia artificial. Hasta hace no muchas fechas, al menos en servicios de inteligencia artificial generativa, nadie tenía ninguna duda de que era EE. UU. la ganadora… Hasta que China se ha posicionado. Y

sus modelos (dos, con capacidades de razonamiento profundo, hoy en día, DeepSeek y Qwen) ofrecen respuestas con una orientación diferente cuando se les plantea una cuestión de interés para la decisión estratégica (Khan, 2024).

Nuestros equipos de análisis tienen como misión operar con diferentes tipos de IA con el fin de disponer de una visión más holística sobre cualquier problema que se someta a su análisis, así como de disponer de un sistema paralelo de verificación de la información que levante alertas sobre la existencia de una determinada alucinación.

El uso en paralelo de varios servicios de IA generativa, según nuestra experiencia, mejora y enriquece el contenido de los análisis y, sobre todo, permite disponer de una visión muchas veces radicalmente contrapuesta a la que nuestros equipos de analistas manejan por modelo cultural y geopolítico.

Esta última reflexión nos conecta con el talento. Su disponibilidad. Su calidad. Su pluridisciplinaridad.

4 El problema del talento

El CAP, configurado en la actualidad como una unidad civil dentro de la Guardia Civil, tiene como una de sus misiones principales la relación con las universidades.

Nuestro programa de prácticas está considerado dentro de un nivel de excelencia por las universidades y las especialidades formativas con las que trabajamos habitualmente.

Históricamente, la gran mayoría de nuestro alumnado en prácticas procedía de los grados de Traducción e Interpretación, Relaciones Internacionales, Derecho, Criminología o Comunicación. Hace ahora cuatro años, se detectó la necesidad de nutrir nuestros programas con capacidades de Ciencias Exactas, Ciencias Químicas, Ciencias Físicas, Ingenierías y especialidades conexas que nos permitiesen desarrollar capacidades propias relacionadas con las tecnologías más avanzadas que, en ese momento, comenzaban a eclosionar en el mercado.

La formación que se ofrece a este alumnado es el de analista de inteligencia de nivel básico y, durante su estancia, se les proponen trabajos y problemas que nos ayuden en nuestro proceso de mejora continua de capacidades.

En el nuevo escenario de desarrollo de sus competencias se apuesta por incidir en su capacidad crítica y el desarrollo de habilidades culturales de índole transversal, con frecuentes encuentros para el debate que sean generadores de una necesaria «inteligencia colectiva» orientada a objetivo (Rey, 2022). Estas sesiones son registradas y transcritas para, haciendo uso de sistemas

de inteligencia artificial, incorporar el conocimiento generado a las ontologías semánticas propias del ecosistema desarrollado.

Configurar equipos estables y pluridisciplinares de análisis de Inteligencia de Fuentes Abiertas requiere capacidades de acceso a talento y organizativas comparables a las que despliegan las unidades de Inteligencia de grandes Corporaciones, en nuestro caso, con el problema añadido de la altísima rotación y de la necesidad de un control exhaustivo de la información de nuestros equipos en prácticas.

El proceso de entrenamiento del equipo comienza por la superación de dos cursos en línea que permiten, por un lado, conocer el entorno y el contexto en el que se van a desarrollar las labores de análisis: la Guardia Civil. Una Institución de Seguridad Pública. Ello tiene una importancia estratégica dado que el estudiantado, proveniente, fundamentalmente, de un entorno «urbanita», tiene un conocimiento limitado no ya de la seguridad del Estado y de la seguridad ciudadana, sino de la propia Guardia Civil, a la que conecta de forma inmediata con sus competencias de tráfico, desconociendo todo de sus misiones y capacidades reales.

Este hecho induce la necesidad previa de una capacitación en el conocimiento de las antedichas misiones y del «modelo semántico» adecuado para que los equipos puedan comunicarse entre ellos y con el decisor estratégico a través de sus informes.

El entrenamiento se dirige a configurar un modelo de «cultura general compartida» en las dos áreas de conocimiento y competencia para el CAP: el análisis de inteligencia y el diseño de escenarios de futuro. Dos herramientas complementarias y fundamentales para configurar una estructura de apoyo a la toma de decisión orientada a las necesidades de un entorno VUCA como el actual.

Una vez superada esta formación inicial en la que, como reiteramos, se ha construido el campo semántico adecuado, se procede a aplicar las herramientas disponibles a través de ejemplos específicos, a través de los cuales se entrenan las capacidades de generación del *prompting* adecuado para definir un contexto, unos objetivos y una determinación de fuentes y planificación de investigación adecuadas al objetivo marcado.

De forma recursiva y haciendo uso constante de metodología DELPHI para la mejora del producto final, ya sea a partir de la optimización en la fase de generación como en la fase final de aprobación del producto que pase a la fase de diseminación.

La experiencia, una vez generadas las estructuras de conocimiento y flujos de información, puede calificarse como óptima, toda vez que todas estas cohortes de estudiantado se caracterizan por un conocimiento y uso de las capacidades tecnológicas que se puden calificar de sobresaliente.

Veamos una representación gráfica del proceso que se puede seguir, de forma interactiva, a partir de la URL: https://ciberinteligencia.eu/interacciones/analistas/.

Por todo lo antedicho, se cierra el círculo y, siendo nuestra misión entrenar a este estudiantado en capacidades básicas de análisis y de diseño de futuros, se dedica su talento a la revisión final de los productos generados de manera casi automática por las herramientas de inteligencia artificial que ponemos a su disposición para la generación de informes para decisor estratégico. Además, sus propias aportaciones, como se ha reseñado, pasan a ser parte del conocimiento compartido tanto dentro de cada cohorte como para cohortes posteriores lo que, al final, constituye una de las mejores externalidades positivas conseguidas: La persistencia del conocimiento, independientemente de las personas. Todo el conocimiento generado pasa, de forma automática, a formar parte de la ontología de conocimiento del CAP, de tal manera que podrá ser utilizada, en una fase posterior, para la realización de un tuneado fino recursivo del modelo, de tal forma que se minimice la posibilidad de que se produzcan «alucinaciones» en el proceso de análisis.

Se comprueba, por tanto, que la diversidad (que no solo las capacidades pluridisciplinarias) adquiere todo su sentido en el caso que nos ocupa. Edad, sexo, ideología, conocimiento diverso..., sin duda, enriquecen la calidad del producto del análisis y, por ende, la capacidad de detección de posibles alucinaciones en el producto generado por la IA.

5 Alucinaciones y otras cuestiones: una aproximación práctica

Las IA generalistas, aunque avanzan rápido en materia de pensamiento profundo, aún tienen problemas graves para el manejo de cierta información relacionada con entidades específicas. La mayoría de nosotros hemos hecho la prueba de preguntarle a la IA qué sabe de nosotros mismos y hemos sido conscientes del desastroso resultado que ha presentado, aunque, es necesario reconocerlo, la precisión avanza a una velocidad incontestable.

Imaginemos un resultado parecido al valorar una entidad persona, física o jurídica, una geolocalización o un vehículo, por poner ejemplo de interés para la Seguridad Interior. Esas «alucinaciones» harían inviable el uso de procesos de automatización en investigación policial o, peor aún en un escenario de teatro de operaciones de cualquier tipo.

Los modelos de lenguaje entrenados con objetivos generales tienen, entre otras, esta fundamental, podría denominarse, tara. Eso las transformaría en herramientas prácticamente inútiles para la generación de Inteligencia por la imprecisión o, aún peor, por la evidente tendencia a contestar inventándose las respuestas.

Afortunadamente, además de la mejora exponencial de sus capacidades y precisión que ya se ha enunciado con anterioridad, existen posibles soluciones, hasta el momento parciales y sujetas a revisión, para la mitigación de este problema.

¿Cómo se podría mejorar la calidad y actualización de los datos que han de nutrir las respuestas de los modelos de lenguaje para que resulten confiables?

La generación de modelos específicos, de agentes especializados, de ontologías semánticas adaptadas o de grafos de conocimiento, permiten mejorar la precisión de las respuestas de la IA generativa, así como mitigar las posibles alucinaciones de esta, de forma sectorial.

Vamos a entrar en el detalle en este conjunto de tecnologías, a nuestro juicio, justo estos son los elementos fundamentales para poder hacer un uso intensivo y preciso de los sistemas de inteligencia artificial generativa en los procesos de generación de Inteligencia, en el caso del CAP, a partir de fuentes abiertas.

El documento Ontology Learning for Hybrid Threats (Bosso y Ruberto, 2025), en el que se recomienda el uso de la herramienta HYBOLT que se describe como:

> «una herramienta para el aprendizaje de ontologías aplicada al análisis de amenazas híbridas. Se propone usar modelos de lenguaje extenso (LLM) y grafos de conocimiento (KG) para extraer información fiable de fuentes abiertas, mitigando problemas como las alucinaciones de los LLM. La metodología se

centra en la extracción de tripletas sujeto-acción-objeto para construir un KG, simplificando la información para mejorar la precisión. Se evalúa el rendimiento de HYBOLT en tareas como la extracción de ubicaciones y cantidades de individuos afectados, mostrando una alta precisión y resistencia a las alucinaciones».

Este documento ofrece, a nuestro juicio, las líneas maestras básicas a explorar para conseguir los resultados deseados.

Tras «dialogar» con un sistema de fuentes relacionadas con conocimiento sobre ontologías semánticas y otras estrategias asociadas, aplicadas al análisis de inteligencia, se ha obtenido el conocimiento suficiente para intentar desarrollar un conjunto de pautas con el objetivo de generar un ecosistema consistente de mitigación de riesgos asociados a posibles alucinaciones de la IA generativa durante el proceso de generación de informes de Inteligencia dentro de nuestro Centro de Análisis y Prospectiva, orientados a la Seguridad Interior.

Con el fin de mejorar la precisión de una IA generativa en un sector específico a partir de servicios de IA generativa generalistas, se ha considerado implementar las siguientes estrategias:

— **Utilizar bases de conocimiento formales (Knowledge Graphs – KG) para corregir las limitaciones de los Modelos de Lenguaje Grandes (LLM)** (Dieter Fensel, 2020). Los LLM pueden sufrir de alucinaciones, sesgos y falta de conocimiento, en especial, en campos específicos como el análisis de amenazas híbridas. Las bases de conocimiento, como los grafos de conocimiento (KG), proporcionan información estructurada y verificada para mejorar la calidad de las respuestas de los LLM. Estas bases de conocimiento se nutrirían de fuentes abiertas consolidadas y, de manera recurrente, con productos propios validados y correctamente etiquetados. Se entraría, por tanto, en un proceso de mejora continua de la calidad de la información manejada y utilizada para generar nuevo conocimiento.

— **Implementar un sistema que transforme texto no estructurado en datos estructurados**. Este sistema incluye una herramienta de aprendizaje de ontologías, diseñada para inferir un KG específico del dominio Seguridad Interior a partir del texto y poblarlo con datos de alta calidad. En la actualidad, desde el CAP se está evaluando el servicio Protegé[5], de la Universidad de Stanford, para la generación de una ontología semántica específica de Seguridad Interior. Se ha iniciado la generación a partir de subdominios específicos dado que la tarea de generar una ontología semántica de Seguridad Interior requeriría recursos y equipos, ahora mismo, inalcanzables por nuestro equipo.

[5] Véase: https://protege.stanford.edu/

— **Aprovechar las capacidades básicas de los LLM**. En lugar de depender de conocimientos específicos del campo o de razonamiento lógico complejo, se pueden usar las capacidades básicas de los LLM, como el reconocimiento de patrones gramaticales (sujeto, acción, objeto), para extraer información y crear KG. Esto minimiza los sesgos y alucinaciones, ya que solo se requieren habilidades lingüísticas básicas. Apoyarse en las propias capacidades de los LLM para la generación de los KG constituye la herramienta básica para asegurar la eficiencia del proceso. La generación de KG es un proceso complejo y costoso que minora los recursos necesarios para llevarlo a efecto, haciendo uso de las propias herramientas de Inteligencia Artificia generativa disponibles[6].

— **Atomizar entidades para crear grafos de conocimiento más simples e informativos**. Este proceso implica simplificar los nombres de las entidades y agregar información adicional como atributos, lo cual hace que los grafos sean más legibles y fáciles de analizar. Etiquetar correctamente las entidades y hacer un diseño coherente de sus atributos constituye un elemento fundamental a la hora de tener éxito en la implementación del ecosistema. Para llevar a cabo esta acción se precisa de la incorporación de talento específico. Se ha procedido a ampliar las ofertas de incorporación de personal en prácticas a las áreas de conocimiento necesarias para trabajar en el sentido expuesto.

Pongamos un ejemplo para entender mejor el concepto.

Imaginemos una noticia de fuente abierta de interés para una investigación.

«El empresario Juan Pérez, CEO de la empresa tecnológica Neuronix, se reunió en Madrid con representantes del Ministerio de Ciencia para discutir una posible colaboración en inteligencia artificial».

Se procedería a ejecutar un proceso de extracción de entidades, atomización de estas y generación de los grafos de conocimiento simples y representativos de la siguiente forma:

Paso 1: extracción de entidades (sin atomizar)

En un inicio, se extraerán las entidades detectadas de la siguiente forma:

— Juan Pérez.
— Neuronix.
— Ministerio de Ciencia.

[6] Se puede acceder a un ejemplo visualmente muy atractivo en la sede del Tribunal Constitucional. Dominio Derecho Constitucional desde la URL: https://hj.tribunalconstitucional.es/es/Descriptor/Index?vista=1.

— Madrid.

— Inteligencia artificial.

Pero esto puede generar un grafo poco claro si las relaciones entre entidades no están bien definidas.

Paso 2: atomización de entidades

Atomizar significa dividir entidades complejas en componentes más simples y específicos. Así, se pueden representar las relaciones entre las mismas de forma mucho más clara.

Entidades atomizadas:

- Persona: Juan Pérez.
- Rol: CEO.
- Organización: Neuronix.
- Ubicación: Madrid.
- Institución gubernamental: Ministerio de Ciencia.
- Tema: Inteligencia artificial.
- Evento: Reunión.
- Relación: Colaboración potencial.

Paso 3: construcción del grafo de conocimiento

Con las entidades atomizadas, se construirá el grafo a partir de estas relaciones entre las mismas:

- Juan Pérez → ocupa el rol de → CEO.
- CEO → de → Neuronix.
- Juan Pérez → participó en → reunión.
- Reunión → ubicación → Madrid.
- Reunión → con → Ministerio de Ciencia.
- Reunión → tema → Inteligencia artificial.
- Ministerio de Ciencia → interesado en → colaboración.
- Neuronix → interesado en → colaboración.

De esta forma, se estructuraría el conocimiento de tal forma que se podría alimentar nuestra ontología de forma mucho más segura y firme, con conocimiento estructurado que minimizaría la posibilidad de error por parte del LLM.

— **Utilizar un enfoque de *human-in-the-loop***. Aunque se puede automatizar parte del proceso de aprendizaje de ontologías, es importante que un experto revise y ajuste la salida generada por la IA. Nuestro circuito de validación actual debería poder implementar la revisión por pares del producto final. De momento, esta revisión se asigna a un único analista que comprueba coherencia del documento sin entrar en elementos de precisión salvo detección de alucinación lo que induce un conjunto indeseable de riesgos. Valoramos el riesgo residual subyacente como adecuado, en función del tiempo.

— **Construir ontologías de manera semiautomática**. Se pueden usar un LLM para generar preguntas de competencia para el desarrollo de ontologías y convertir oraciones de lenguaje natural en axiomas OWL (W3C, 2025), reduciendo el esfuerzo asociado con la creación manual. El apoyo del propio LLM en esta tarea es fundamental para la obtención de resultados y, sobre todo, el mantenimiento de la precisión del sistema. El ya referenciado Protegé constituye una valiosa herramienta para acometer este proceso de automatización.

— **Implementar estrategias de validación de conocimiento**. Se pueden comparar las respuestas del LLM con la información de fondo de los grafos de conocimiento, verificando la coherencia y la veracidad de la salida. También se pueden utilizar grafos de conocimiento como entrada en procesos de inferencia de múltiples pasos, limitando aún más el riesgo de alucinación. Esta actividad la se integrará en el proceso de auditoría de la calidad de las respuestas. En un proceso de mejora continua. Probablemente, podrá ser automatizada en el futuro. En este momento, en el CAP se está estudiando la posibilidad de dedicar parte de los recursos disponibles a entrenar estas capacidades con el fin de validar, de forma más ajustada, la calidad de los análisis obtenidos.

— **Usar un modelo de lenguaje más pequeño, con menos parámetros**. Modelos más pequeños, pueden ejecutarse de forma local sin necesidad de recursos computacionales externos, ahorrando energía y permitiendo la posibilidad de usar la complejidad del modelo para futuras necesidades. Esta acción ha de complementarse con bases de conocimiento específicas, con alta calidad de sus datos, que, de alguna manera, permitan «sectorizar» el modelo. Se estudia la posibilidad de creación de agentes específicos que minimicen los tiempos de respuesta, orientados a tareas específicas relacionadas con determinados procesos de toma de decisión o de obtención de información acotada sobre un determinado asunto o entidad.

— **Utilizar la estructura gramatical del texto para la extracción de conocimiento**. El sistema puede identificar y extraer elementos

como sujeto, acción y objeto utilizando una estrategia de *few-shot prompting*[7].

— **Descomponer los nombres de las entidades en componentes más pequeños**. El sistema puede usar LLM para dividir los nombres de las entidades en nombre principal y atributos, lo que permite la creación de entidades simplificadas y enriquecidas, más manejables para su incorporación en otros procesos más complejos. En el CAP, hemos hecho pruebas de extracción y atomización de entidades en determinados documentos para, a partir de los mismos, generar KG sobre los que trabajar en la mejora de la eficiencia y precisión de la respuesta. Aún en proceso de refinado, los resultados son muy prometedores.

— **Asegurar la consistencia de las entidades**. El sistema puede comparar las entidades atomizadas con las originales para garantizar que la misma entidad no sea convertida en diferentes entidades con nombres similares. La consistencia ha de asegurarse para que la precisión no se resienta.

— **Implementar diferentes estrategias para construir una base de conocimiento coherente**. Esto puede ser por medio de estrategias de alta o baja resolución en la validación de entidades de texto con entidades de una base de conocimiento existente. En nuestro sectorial específico, de Seguridad Interior, esta acción es fundamental para mitigar el riesgo de alucinaciones. La coherencia de la base de conocimiento y una ontología semántica precisa son las herramientas fundamentales de mitigación de alucinaciones.

— **Mejorar la calidad de las instrucciones que se dan al LLM en los *prompts***. Se pueden comparar los diferentes grafos de conocimiento generados por diferentes «prompts» para determinar cuál es el más prometedor. Forma parte de las capacidades que han de desarrollar los y las nuevas analistas: ser capaces de contextualizar desde diferentes ángulos una determinada necesidad y, por ende, ser capaces de expresar esa contextualización en «prompts» que generen respuestas de calidad diversa para poder escoger la más adecuada para el escenario propuesto. Entrenar al equipo de analistas para desarrollar estas capacidades específicas y, más aún, que el conocimiento adquirido en esta materia goce de persistencia constituye el «core de negocio» de estos nuevos equipos de analistas. No registrar el ecosistema de *prompts* de manera adecuada ha sido un error que

[7] Se trata de una técnica de aprendizaje automático que permite a los modelos de lenguaje aprender a realizar una tarea nueva con solo unos pocos ejemplos. En lugar de entrenar el modelo en un conjunto de datos masivo, se le proporcionan algunos ejemplos de la tarea y se espera que generalice a partir de ellos para realizar la tarea en nuevas entradas.

se ha cometido durante el desarrollo de las nuevas metodologías automatizadas de análisis.

— **Implementar el método de *Retrieval Augmented Generation* (*RAG*)**. Se puede usar una base de datos vectorial para realizar el seguimiento de los grafos de conocimiento importantes cargados por el usuario (memoria declarativa) o las interacciones previas usuario-LLM (memoria episódica). La implementación de un RAG es un proyecto complejo y dependiente de acciones previas que, aunque estudiada, pasaría a formar parte de una evolución de nuestras capacidades actuales. Fundamentalmente orientada al mantenimiento proactivo del dominio semántico.

Todo lo antedicho puede representarse gráficamente de la siguiente forma:

Pueden acceder al contenido del gráfico interactivo desde: https://ciberinteligencia.eu/interacciones/mitigacion/.

6 Conclusiones

El análisis de inteligencia de fuentes abiertas dispone de una herramienta, la inteligencia artificial generativa, que ha abierto un nuevo escenario en este sector.

Las capacidades multiidioma, la multimodalidad y el multimodelo LLM constituyen una oportunidad para los equipos de análisis de inteligencia en

su objetivo de detección temprana y acceso a información de fuente abierta de interés para la decisión estratégica pudiendo realizar un tratamiento desde diversos modelos culturales proporcionados por los entrenamientos de los LLM, con lo que es posible disminuir los posibles sesgos y, gracias a ello, mejorar la precisión de los análisis automatizados.

El problema de las alucinaciones, sobre todo, las asociadas a entidades específicas, constituye un riesgo de confiabilidad de los análisis. En este artículo, se han glosado posibles metodologías y acciones orientadas a la mitigación del antedicho problema. Algunas de las mismas ya están siendo estudiadas para implementación en nuestra factoría de inteligencia de fuente abierta, de tal forma que se substituya parte del trabajo manual de los equipos por herramientas y procesos automatizados que, además, mejoren de forma recursiva la calidad de los análisis.

El continuo desarrollo de nuevas capacidades de tipo agéntico, planificación y ejecución automatizada de procesos de investigación de fuente abierta, por parte de los nuevos modelos de razonamiento profundo sorprenden a diario por su precisión y calidad de los productos generados que, sin duda, pasan a ser base de trabajo para el inicio de la acción de validación final de contenidos por parte del analista humano mejorando, exponencialmente, la eficiencia del proceso y los tiempos de respuesta.

El conocimiento generado, además, es incorporado al corpus de conocimiento especializado de la Institución, lo que permite el desarrollo de una ontología semántica especifica con la que trabajar, de forma recursiva, en el reentrenamiento de modelos adaptados a las necesidades de la Institución. Reduciendo, al tiempo, la posibilidad de errores o alucinaciones durante los procesos de análisis.

Al aplicar estas técnicas, es posible construir sistemas de IA generativa más precisos, confiables y adaptados a dominios específicos, lo que mejora la calidad de la información extraída y reduce la posibilidad de errores o alucinaciones.

Parte de estas técnicas y metodologías de mejora recursiva de los modelos ya están en estudio para ser aplicadas en el proceso de evolución de las capacidades de análisis del CAP. Requerirán del desarrollo de proyectos específicos asociados a una visión estratégica que ha de determinar cuáles son los objetivos y misiones que habrá de llevar a cabo nuestra unidad especializada en análisis de inteligencia de fuentes abiertas para que el desarrollo y despliegue sea capaz de cumplir con la misión encomendada.

Por supuesto, la asignación de recursos económicos y la captación del talento para una explotación intensiva y eficiente de estas herramientas constituye, ahora mismo, el principal de los objetivos del CAP en el corto y medio plazos.

Estamos aprendiendo, a veces, a través del error y, desde luego, gracias al mismo. No hay que temer al error. Estamos en un proceso de exploración metodológica en el que la recompensa promete ser de muy alto nivel. El proceso de mejora en la precisión de los análisis haciendo uso de herramientas de inteligencia artificial está en evolución y sometido a la posibilidad de fracasos parciales pero los resultados que estamos obteniendo, sin duda, nos anticipan la disponibilidad de un conjunto de conocimientos y herramientas que harán que manejar la complejidad no sea un sueño inalcanzable.

No nos resistamos al cambio porque, precisamente, ese cambio es el que nos abre alguna oportunidad de éxito en el futuro.

Bibliografía

AENOR. (2018). UNE 166006:2018. Gestión de la I+D+i: Sistema de vigilancia e inteligencia competitiva. Obtenido de https://en.tienda.aenor.com/norma-une-166006-2018-n0059973

Dieter Fensel, U. Ş. (2020). Knowledge Graphs: Methodology, Tools and Selected Use Cases. Springer.

Francesco BOSSO, Stefano RUBERTO. (2025). EU Science Hub. Comisión Europea. Obtenido de https://publications.jrc.ec.europa.eu/repository/bitstream/JRC140863/JRC140863_01.pdf

Guldenmund, F. W. (2000). The nature of safety culture: A review of theory and research. Safety Science, 215-257.

IA, G. (2025). Fuentes de error en la inspiración humana.

Kahneman, D. (2012). Pensar rápido, pensar despacio. Debate.

Khan, M. A. (2024). Impact of Artificial Intelligence on the Global Economy and Technology Advancements. Artificial General Intelligence (AGI) Security, 147–180.

Mosqueira-Rey, E. H.-P.-R.-B.-L. (2023). Human-in-the-loop machine learning: A state of the art. Artificial Intelligence Review 56, 3005–3054.

Rey, A. A. (2022). El libro de la Inteligencia colectiva. Almuzara.

Soliman, M. M. (2024). Artificial intelligence powered Metaverse: Analysis, challenges and future perspectives. Artificial Intelligence Review, 57, Art. 36.

Storey, V. C. (2025). Generative Artificial Intelligence: Evolving Technology, Growing Societal Impact, and Opportunities for Information Systems Research. Information Systems Frontiers.

W3C. (2025). Estándar OWL. Obtenido de W3C: https://www.w3.org/2007/09/OWL-Overview-es.html