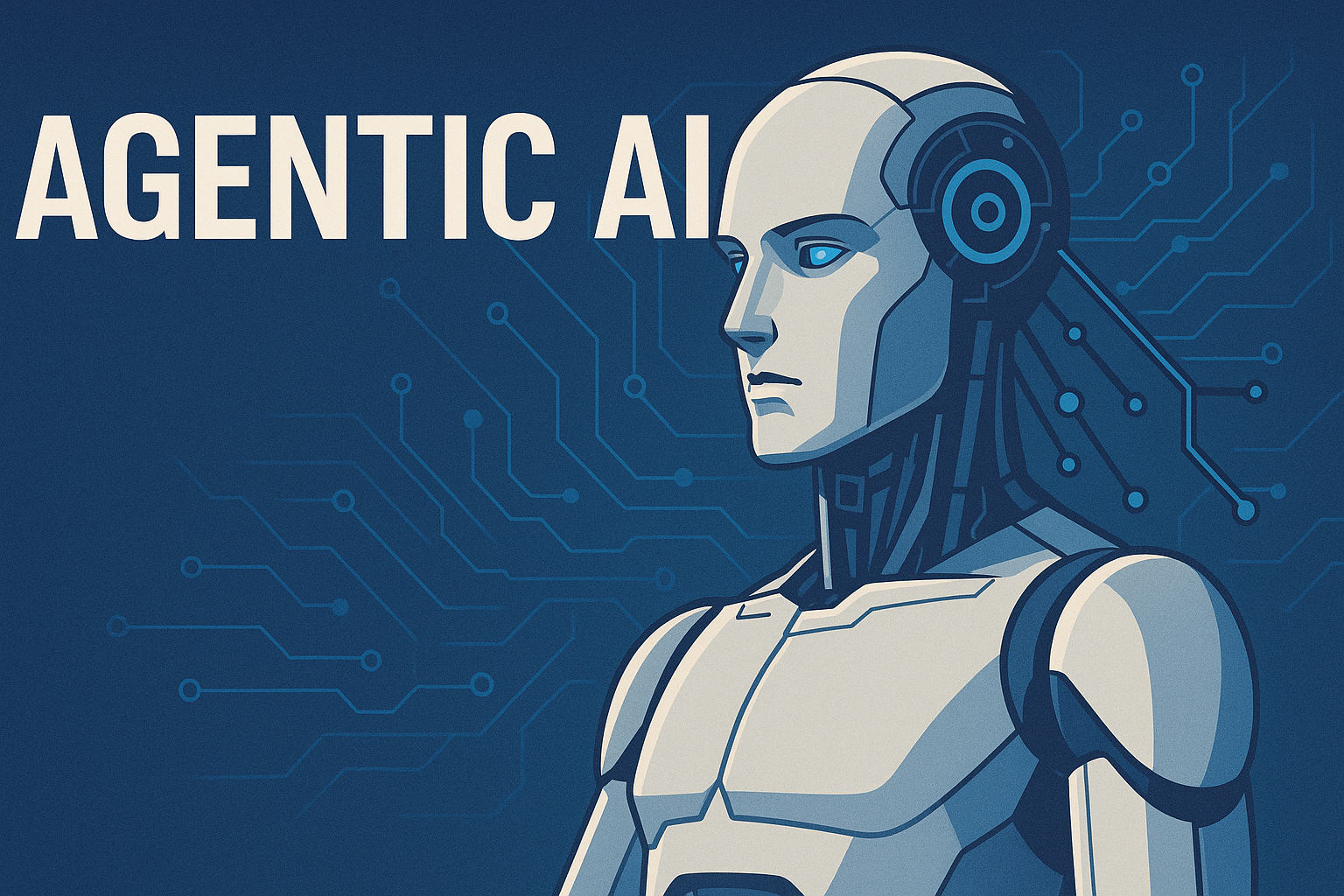
Trabajo de Fin de Máster - BusinessAnalistGPT - Juan Miron

Juan Manuel Miron Lubieniecki  
Máster de Formación Permanente en Inteligencia Artificial  
Curso académico: 2024 - 2025  
Modalidad de TFG: TFM Profesionalizante  
Área de estudio: Agentic AI



# Resumen

El presente trabajo describe el desarrollo de un asistente basado en inteligencia artificial capaz de realizar un cuestionario inicial a clientes internos o externos que soliciten un nuevo proyecto, una modificación o una mejora de procesos. Este asistente automatiza la primera iteración de la recopilación de requisitos mediante interacción en lenguaje natural, optimizando la generación de documentos de especificación funcional. A través de la implementación de este Agente se estudiara los sistemas Multi Agentes su coordinación e implementación.

Palabras clave: RAG, Chatbot, Human-Computer Interaction, Large Language Models, Topic-Based Group Formation, Multi-Agent Systems, Natural Language Processing, Adaptive Interaction Protocols, Agentic AI

# Índice de contenidos

1. Introducción

2. Marco teórico

3. Metodología

4. Resultados

5. Discusión

6. Conclusiones

7. Limitaciones y futuras líneas de investigación

8. Referencias bibliográficas

9. Índice de figuras

10. Índice de tablas

11. Acrónimos

12. Anexos

# 1. Introducción

El presente Trabajo Fin de Máster (TFM, en adelante) se enmarca dentro del área de estudio de la Agentic AI, abordando la necesidad de optimizar la fase de recopilación de requisitos mediante el desarrollo de un asistente inteligente.

Justificación: Desde el punto de vista económico y practico, se observa un creciente interés en la automatización de procesos iniciales de captura de información en proyectos.

Desde el punto de vista académico este proyecto se utilizará con el foco de profundizar el conocimiento en la utilización de sistemas multi agente utilizando y su organización experimentando con lo ultimo que ofrece esta tecnología al día de hoy incluyendo sistemas experimentales.

Problema y finalidad: desde el punto de vista del proyecto la problemática principal radica en la falta de automatización en la recopilación de requisitos iniciales para proyectos de mejora o desarrollo.

Desde el punto de vista de sistemas Multi Agentes la problemática principal es lograr la coordinación más eficiente para diferentes tipos de Agentes

Objetivos del TFM: A modo practico desarrollar un asistente basado en IA que realice la primera iteración de la recopilación de requisitos.

A modo de investigación, entender las diferentes propuestas actuales y experimentales de coordinación de sistemas Multi Agentes.

# 2. Marco teórico

Marco contextual y alcance

Antes de exponer el marco formal propuesto, resulta fundamental establecer sus bases conceptuales y posicionarlo dentro del contexto más amplio de los sistemas multiagente (MAS) y los enfoques de coordinación semántica. En esta sección se presentan los conceptos esenciales y las tecnologías que respaldan nuestra propuesta, así como la terminología que se empleará a lo largo del documento. Nos centraremos particularmente en la teoría de **TB-CSPN** (*Topic-Based Conditional Structured Process Networks*) y en el uso de **redes de Petri** como soporte formal, sin ahondar en sus demostraciones matemáticas, ya que estas exceden el alcance y los objetivos de este trabajo.

**TB-CSPN (Topic-Based Conditional Structured Process Networks)** es un marco arquitectónico diseñado para modelar y coordinar interacciones entre múltiples agentes —humanos, LLMs y agentes especializados— dentro de sistemas distribuidos y orientados a procesos. Se basa en la idea de representar las dinámicas organizativas como redes estructuradas de procesos, donde cada transición está condicionada por el estado de tópicos semánticos compartidos.

En este marco, los **"tópicos"** actúan como unidades conceptuales que encapsulan el significado y contexto de una conversación o flujo de trabajo, y su evolución se regula mediante **umbrales de activación (threshold guards)** y mecanismos de supervisión jerárquica. Las transiciones entre estados del sistema están organizadas en tres **capas funcionales**:

* **Capa de Superficie (Surface Layer):** donde ocurren interacciones de alto nivel entre agentes supervisores y humanos.
* **Capa de Observación (Observation Layer):** dedicada al monitoreo, análisis semántico y síntesis de información, comúnmente gestionada por agentes basados en LLMs.
* **Capa de Cómputo (Computation Layer):** donde se ejecutan tareas específicas y deterministas mediante agentes especializados.

El modelo TB-CSPN no solo permite una separación clara de funciones entre tipos de agentes, sino que también facilita una **coordinación transversal**, al permitir que agentes con diferentes roles puedan interactuar y modificar transiciones distribuidas en múltiples capas. Este diseño favorece la trazabilidad, la auditabilidad y la preservación de la intención semántica humana en todo el ciclo de vida del sistema.

En suma, TB-CSPN es una abstracción poderosa para diseñar arquitecturas agenticas híbridas, garantizando una integración coherente entre la lógica humana y la ejecución automática distribuida.

**Redes de Petri para el modelado de procesos**

Las **redes de Petri** constituyen la base formal del marco TB-CSPN, ofreciendo un enfoque matemáticamente riguroso para modelar sistemas concurrentes y distribuidos. Su estructura se basa en grafos bipartitos compuestos por **lugares** (que representan estados o condiciones) y **transiciones** (que representan eventos o acciones), conectados mediante arcos dirigidos.

La variante utilizada en TB-CSPN es la de las **redes de Petri coloreadas**, que extienden el modelo básico permitiendo que los tokens transporten datos (colores). En este contexto, los tokens pueden codificar distribuciones complejas de tópicos, influenciando el comportamiento de las transiciones en función del contenido semántico.

A diferencia de otros marcos más abstractos o de difícil interpretación, las redes de Petri combinan **claridad estructural** con **capacidad formal de análisis**, facilitando tanto la implementación como la verificación del sistema. Esta propiedad es especialmente relevante en entornos críticos —como salud, finanzas o seguridad— donde es necesario garantizar propiedades como *accesibilidad*, *vivacidad* y *acotación*.

Frente a modelos tradicionales de MAS que se enfocan en grafos de comunicación o diagramas de estados, las redes de Petri permiten representar la **concurrencia**, el **flujo basado en tokens** y la **coordinación semántica** de forma explícita y verificable. También superan a formalismos como las *álgebras de procesos* (p. ej., π-cálculo), que aunque precisos a nivel lógico, son menos intuitivos para capturar estructuras organizacionales y flujos de datos reales.

Una ventaja adicional clave es su capacidad para representar la **colaboración humano-IA** mediante la asignación de transiciones a decisiones de alto nivel (estratégicas, éticas o contextuales), permitiendo modelar el juicio humano sin necesidad de describir su cognición interna. Esto convierte a las redes de Petri en una herramienta ideal para entornos híbridos donde agentes humanos, LLMs y agentes centáuricos deben coordinarse de forma coherente y trazable.

Finalmente, el marco TB-CSPN puede beneficiarse de extensiones avanzadas como las **redes de Petri jerárquicas**, **con arcos inhibidores** o **contextuales**, que permiten modularidad, activación condicional o dependencia semántica, y representan líneas prometedoras para su evolución futura.

**Los LLMs como agentes**

La consolidación de los modelos de lenguaje de gran tamaño (LLMs) como **agentes funcionales** representa un avance notable en el campo de la inteligencia artificial. Investigaciones recientes, como las de Singh et al., han documentado una transición desde interacciones lineales hacia flujos de trabajo más complejos, iterativos y dinámicos, en los que los LLMs participan activamente mediante procesos de planificación, reflexión y colaboración multiagente.

En este nuevo rol, los LLMs actúan como **mediadores semánticos**, facilitando la integración de marcos estratégicos con heurísticas de decisión, como lo demuestra el trabajo de Ghisellini et al. A través del uso de representaciones en espacios vectoriales y cálculos de similitud semántica, estos modelos son capaces de conectar conocimientos provenientes de distintas disciplinas, desempeñando un papel muy similar al de nuestros **agentes consultores**, cuya función es sintetizar información transversalmente entre dominios.

En este contexto, los LLMs se configuran como puentes entre estructuras analíticas formales y conocimiento experiencial, enlazando la **intención estratégica** de los agentes supervisores con la **ejecución operativa** de los agentes especializados.

Esta línea de trabajo se conecta directamente con investigaciones sobre el uso de LLMs para tareas de **planificación y razonamiento**, como las de Capitanelli y Mastrogiovanni, quienes demuestran la capacidad de estos modelos para orquestar acciones en entornos de interacción humano-robot. Esta función corresponde al rol que asignamos a los agentes consultores en TB-CSPN: facilitar la traducción del propósito estratégico en acciones específicas y viables.

Finalmente, para abordar el reto de integrar razonamiento simbólico y subsimbólico, nuestro enfoque se inspira en el trabajo de Tenenbaum, que propone mecanismos de conciliación entre ambos tipos de procesamiento cognitivo. En este marco, el modelado de tópicos se convierte en una capa de mediación clave que permite a los distintos tipos de agentes comunicarse de manera coherente y efectiva.

**Herramientas en sistemas de agentes LLM**

Las herramientas representan un componente esencial en la arquitectura de los agentes basados en modelos de lenguaje (LLM). Cuando los agentes enfrentan tareas complejas o que exceden sus capacidades básicas —como cálculos precisos, acceso a datos en tiempo real o manipulación estructurada de información— recurren a **herramientas externas** para complementar sus respuestas y ampliar su margen de acción.

Además de utilizar herramientas preexistentes, algunos agentes pueden **diseñar o construir herramientas propias**, adaptadas a los requerimientos específicos del entorno en el que operan. Esta capacidad de autogeneración se activa especialmente en tareas creativas o de resolución de problemas no estructurados.

Por otro lado, para que un agente LLM funcione de forma confiable en un entorno real, también necesita un conjunto de **herramientas de soporte para su despliegue, mantenimiento y adquisición de datos**. Estas herramientas aseguran que el agente no solo sea competente desde el punto de vista funcional, sino también **sostenible y operativamente robusto**.

**3.2.1 Herramientas utilizadas por los agentes LLM**

Dado que los LLMs presentan limitaciones notables en ciertas áreas —como el razonamiento cuantitativo exacto, la ejecución de tareas con dependencia temporal o el acceso a fuentes de datos en tiempo real— se integran herramientas externas especializadas para ampliar su funcionalidad.

Estas herramientas se pueden agrupar en tres grandes categorías, según el tipo de apoyo que ofrecen:

1. **Herramientas para acceso a datos externos**: permiten a los agentes consultar APIs, bases de datos en vivo o motores de búsqueda para obtener información actualizada o específica.
2. **Herramientas de cálculo y procesamiento simbólico**: utilizadas para resolver operaciones matemáticas avanzadas, evaluar expresiones lógicas o realizar transformaciones precisas de datos.
3. **Herramientas de interacción estructurada**: como generadores de código, visualizadores de diagramas, editores de documentos o conectores con plataformas externas (por ejemplo, Slack, Notion, Jira), que permiten a los agentes producir resultados tangibles y funcionales en formatos útiles para el usuario final.

Estas herramientas actúan como **extensiones funcionales** del agente, permitiendo que su razonamiento sea más preciso, verificable y adaptado a tareas del mundo real.

**Protocolo MCP: el “USB-C” para conectar LLMs con el mundo real**

El **Model Context Protocol (MCP)** es un estándar abierto, presentado por Anthropic a finales del 2024, diseñado para **conectar de manera automática, segura y estandarizada los modelos de lenguaje (LLMs) con datos y herramientas externas** [modelml.com+15Wikipedia+15Label Studio+15](https://en.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com).

**¿Por qué es relevante?**

Antes de MCP, cada integración con datos o APIs requería un desarrollo específico para el agente, lo que complicaba el mantenimiento y la escalabilidad (problema *M×N*). MCP lo resuelve mediante una interfaz única, independiente del modelo o fuente, facilitando:

* Conectar el LLM con múltiples herramientas sin escribir conectores repetitivos.
* Cambiar entre distintos proveedores de LLM sin reprogramar las integraciones.
* Mantener un control seguro del acceso a datos, sin exponer claves sensibles [Wikipedia](https://de.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com)[Wikipedia](https://en.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com)[Model Context Protocol](https://modelcontextprotocol.io/introduction?utm_source=chatgpt.com).

**Arquitectura cliente‑servidor**

MCP sigue un modelo clásico **cliente‑servidor** basado en JSON‑RPC 2.0:

* **Hosts**: aplicaciones LLM que requieren contexto (como IDEs o chatbots).
* **Clientes MCP**: componentes que gestionan conexiones a servidores.
* **Servidores MCP**: exponen datos, recursos y herramientas accesibles al modelo.
* **Fuentes de datos**: pueden ser locales (archivos, bases) o remotas (APIs) [Google GitHub+15Wikipedia+15Model Context Protocol+15](https://en.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com)[Wikipedia+1Model Context Protocol+1](https://es.wikipedia.org/wiki/Protocolo_de_Contexto_de_Modelo?utm_source=chatgpt.com).

**Componentes clave del protocolo**

* **Resources**: datos consultables como documentos, bases de datos, estados.
* **Tools**: funciones ejecutables (API calls, cálculos, etc.).
* **Prompts**: plantillas específicas que el servidor ofrece al agente para estandarizar el acceso [Wikipedia](https://de.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com)[Wikipedia](https://es.wikipedia.org/wiki/Protocolo_de_Contexto_de_Modelo?utm_source=chatgpt.com).

**Ecosistema y adopción**

En 2025, MCP ha sido adoptado por actores como OpenAI, Google DeepMind, Replit, Sourcegraph y Microsoft Copilot Studio, consolidándose como la interfaz estándar entre LLMs y el ecosistema de herramientas [ThirdEye Data+12Wikipedia+12fr.wikipedia.org+12](https://en.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com).

Además, existen múltiples **servicios MCP open-source**, por ejemplo para Slack, GitHub, Google Drive, bases SQL y automatización de navegador [Model Context Protocol+2Wikipedia+2fr.wikipedia.org+2](https://en.wikipedia.org/wiki/Model_Context_Protocol?utm_source=chatgpt.com).

En resumen, **MCP es un estándar crucial que convierte a los LLMs en “agentes conectables”**, capaces de interactuar con el mundo real sin complejos desarrollos por integración. Su enfoque seguro, modular y escalable lo posiciona como una herramienta estratégica para arquitecturas humanas‑IA organizadas y seguras, como BusinessAnalistGPT.

**Inteligencia Artificial Agéntica y coordinación en sistemas multiagente**

Aunque el término *Inteligencia Artificial Agéntica* (Agentic AI) ha ganado popularidad recientemente, su fundamento teórico se encuentra profundamente arraigado en el campo de los sistemas multiagente (MAS). A pesar de que en el discurso popular los modelos de lenguaje de gran tamaño (LLMs) suelen ser representados como inteligencias generales autónomas, en la práctica, las aplicaciones reales dependen en mayor medida de redes coordinadas de agentes especializados.

En este contexto, la coordinación multiagente resulta esencial para combinar capacidades cognitivas y operativas diferenciadas, regular dinámicas emergentes en entornos distribuidos y garantizar que las decisiones que se tomen sean rastreables, verificables y auditables. Además, la interacción entre la IA y la cognición humana está desdibujando los límites entre los agentes puramente humanos y aquellos aumentados por IA.

Particularmente relevante es el concepto de *agentes centáuricos monotónicos*, que surge cuando la IA complementa el juicio humano sin alterar la integridad semántica de sus decisiones. En estos casos, la IA puede refinar, aclarar o reforzar las intenciones humanas, pero nunca contradecirlas. Este principio de *monotonía semántica* resulta crítico en escenarios donde la precisión y la seguridad son prioritarias, como en protocolos clínicos o procedimientos de emergencia.

Dentro del marco TB-CSPN (Topic-Based Conditional Structured Process Networks), esta propiedad se preserva mediante mecanismos como *umbrales de control (threshold guards)* y supervisión directa por parte de agentes humanos o centáuricos. De este modo, se asegura que los valores, intenciones y semántica original definidos por humanos se mantengan intactos, mientras que los agentes se benefician de mejoras computacionales en velocidad, precisión y consistencia.

La arquitectura propuesta en el TB-CSPN contempla dos dimensiones organizativas principales. En primer lugar, se encuentra la categorización funcional de los agentes:

* **Agentes Supervisores**: incluyen tanto humanos como centáuricos monotónicos; son responsables de decisiones de alto nivel y supervisión general.
* **Agentes Consultores**: implementados típicamente con LLMs, se encargan de tareas como interpretación semántica, abstracción y síntesis lingüística.
* **Agentes Ejecutores (Workers)**: desempeñan tareas específicas dentro de dominios concretos, con comportamientos deterministas y bien definidos.

En segundo lugar, se definen tres capas de comunicación o interacción que estructuran el sistema:

* **Capa de Superficie (Surface Layer)**: donde interactúan los agentes supervisores.
* **Capa de Observación (Observation Layer)**: asociada a los agentes consultores y al monitoreo del sistema.
* **Capa de Cómputo (Computation Layer)**: donde operan los agentes ejecutores.

Cada transición entre estados del sistema pertenece exclusivamente a una de estas capas, pero los agentes no están limitados a una única dimensión. Si bien tienen responsabilidades principales en capas específicas (por ejemplo, los supervisores en la superficie, los consultores en la observación y los ejecutores en el cómputo), pueden participar en transiciones ubicadas en otras capas cuando sea necesario.

Esta flexibilidad estructural permite a los agentes operar predominantemente en su dominio de especialización, sin perder la capacidad de interactuar transversalmente en todo el sistema. El concepto de *Espacio de Eventos* (Event Space) define además el punto de contacto entre los estímulos externos y las dinámicas internas del sistema, funcionando como interfaz de entrada sensorial e interpretativa.

En conjunto, esta arquitectura basada en capas y tipos de agentes favorece una coordinación coherente, modular y trazable, al tiempo que permite mantener funciones especializadas. Así, se potencia la colaboración efectiva entre agentes humanos e IA, respetando la semántica y los valores originales de la organización.

**ReAct: Razonamiento y Acción en Sistemas Multiagente**

**ReAct (Reasoning and Acting)** es un enfoque emergente que combina razonamiento deliberativo con ejecución de acciones en entornos interactivos, y ha sido adoptado recientemente como estrategia eficaz dentro de arquitecturas de agentes basados en modelos de lenguaje. Introducido por Yao et al. (2022), ReAct permite que un agente —típicamente un LLM— **razone en lenguaje natural mientras interactúa con su entorno** mediante acciones observables, creando así un ciclo iterativo de pensamiento y actuación.

En el contexto de sistemas multiagente, ReAct representa un **paradigma híbrido** que supera las limitaciones de agentes puramente reactivos o puramente planificadores. A través de este modelo, un agente puede:

* Formular hipótesis sobre el entorno (razonamiento),
* Ejecutar acciones exploratorias o informativas (actuación),
* Evaluar los resultados de sus acciones,
* Y ajustar su razonamiento de forma dinámica en función del nuevo contexto.

Esta capacidad es especialmente útil en entornos distribuidos y con alta incertidumbre, como aquellos donde interactúan humanos, LLMs y agentes especializados. Integrar ReAct en una arquitectura como TB-CSPN, por ejemplo, permite que los **agentes consultores** basados en LLM no solo generen conocimiento pasivamente, sino que participen activamente en la resolución de tareas, ajustando sus inferencias con base en retroalimentación y efectos causales.

Además, el enfoque ReAct facilita una **colaboración más coherente entre agentes**, ya que el razonamiento intermedio se vuelve transparente y trazable para otros componentes del sistema. Esto permite coordinar flujos de información entre capas (superficie, observación, cómputo) y mantener la integridad semántica de los objetivos del sistema.

En suma, ReAct ofrece una solución flexible y eficaz para la implementación de agentes autónomos que requieren adaptabilidad, interpretación contextual y capacidad de intervención en tiempo real, posicionándose como una herramienta clave en arquitecturas agenticas modernas.

**MAD: Debate Multiagente como Mecanismo de Razonamiento Colectivo**

**MAD (Multi-Agent Debate)** es un enfoque de coordinación basado en deliberación colaborativa, en el cual múltiples agentes —habitualmente implementados mediante LLMs— participan en una estructura dialógica de discusión para **resolver problemas complejos, validar decisiones o generar respuestas optimizadas**. Inspirado en principios de argumentación lógica y deliberación humana, MAD busca aprovechar la diversidad de perspectivas entre agentes autónomos para alcanzar conclusiones más robustas y justificadas.

En lugar de delegar una tarea a un solo agente, el enfoque MAD propone que varios agentes **expongan argumentos, contraargumentos y justificaciones** en un entorno controlado, simulando un debate estructurado. A lo largo de este proceso, un moderador (que puede ser otro agente o un humano) evalúa la validez, coherencia y relevancia de las aportaciones, seleccionando la mejor propuesta o integrando múltiples perspectivas en una solución compuesta.

Dentro de un marco como **TB-CSPN**, MAD puede integrarse como **mecanismo semántico de validación cruzada** en las capas de supervisión y observación. Por ejemplo:

* Agentes consultores pueden sostener un debate sobre la interpretación de requisitos de usuario ambiguos.
* Agentes especializados pueden presentar distintas estrategias de implementación.
* Agentes centáuricos pueden intervenir con criterios éticos o estratégicos desde la supervisión humana.

El valor de MAD radica en su capacidad para **detectar contradicciones**, **fortalecer la trazabilidad semántica**, y **reducir el sesgo de decisión**, especialmente en contextos donde la ambigüedad, la subjetividad o la incertidumbre son elevadas.

Además, el debate multiagente fomenta la transparencia y la explicabilidad en los sistemas, ya que cada decisión final puede estar respaldada por un registro estructurado de razonamientos y justificaciones intercambiadas entre los agentes.

En resumen, **MAD proporciona un marco eficaz para el razonamiento colectivo en sistemas multiagente**, reforzando la calidad de las decisiones mediante colaboración argumentativa, y resulta especialmente relevante en arquitecturas híbridas donde la coordinación entre humanos e IA requiere un alto nivel de interpretación y deliberación.

**Integración y colaboración entre humanos e inteligencia artificial**

La integración efectiva entre sistemas de IA y agentes humanos constituye un eje central del marco TB-CSPN. Este enfoque se apoya en el concepto de **agentes centáuricos**, es decir, híbridos formados por humanos asistidos por inteligencia artificial. Esta idea se inspira en los desarrollos en **inteligencia aumentada** y en el creciente cuerpo de trabajo sobre **equipos humano-IA**, que promueven una colaboración fluida y complementaria entre ambas partes.

Para comprender cómo los humanos interactúan con los sistemas de IA en contextos reales, se consideran los aportes de Giudici et al., quienes analizan las respuestas cognitivas y conductuales de los usuarios frente a sistemas interactivos basados en IA. Asimismo, estudios como los de Kim e Im abordan cómo los usuarios atribuyen características humanas a los sistemas inteligentes, lo que influye directamente en la forma en que los **agentes supervisores** humanos se relacionan con el resto de la arquitectura del sistema.

Además, los aspectos **emocionales y sociales** de la interacción humano-máquina cobran especial relevancia. En este sentido, el trabajo de Ahmadi y Haddadi sobre **Inteligencia Emocional Artificial** ofrece un marco útil para diseñar canales de comunicación más empáticos y eficaces entre agentes de distinta naturaleza.

A diferencia de muchos enfoques tradicionales, en los que el ser humano actúa como controlador externo o mero supervisor de sistemas automáticos, nuestro marco conceptual considera a los humanos como **agentes de primer orden** dentro del modelo formal. Esta elección permite una integración más natural del juicio humano en la dinámica general del sistema, al tiempo que preserva la distribución adecuada de roles, responsabilidades y autoridad entre los distintos tipos de agentes.

Mientras que TB-CSPN se centra en la **mediación semántica** y en la **orquestación dinámica de tareas**, otras propuestas —como la de Zhang et al.— abordan la colaboración humano-IA desde perspectivas complementarias, como la **fusión de características multidimensionales** y el **aprendizaje por ensamblado**, orientadas a mejorar la detección robusta de anomalías y decisiones complejas en entornos inciertos.

**Modelado de tópicos como puente semántico**

Para lograr una integración efectiva entre formas simbólicas y subsimbólicas de inteligencia, se propone el uso del **modelado de tópicos** como capa intermedia de mediación semántica. Aunque esta técnica no es profunda en términos de aprendizaje automático, desempeña un papel fundamental como interfaz semántica entre humanos e inteligencia artificial.

El modelado de tópicos permite condensar grandes volúmenes de datos textuales en vectores de tópicos fácilmente interpretables. Esto facilita que agentes humanos o centáuricos puedan inspeccionar, modificar o seleccionar tópicos de manera explícita. Por su parte, los **agentes consultores**, generalmente basados en LLMs, son capaces de desarrollar, enriquecer o recombinar estos tópicos para generar nuevas interpretaciones o descubrimientos. En paralelo, los **agentes ejecutores** pueden suscribirse dinámicamente a los tópicos más relevantes, filtrando la información que necesitan para realizar tareas específicas de forma más eficiente.

A diferencia de los sistemas opacos de aprendizaje profundo de extremo a extremo, el modelado de tópicos introduce un nivel de **alineación explícita entre la comprensión humana y la inferencia automática**, lo que resulta clave para la trazabilidad y control semántico. Entre sus enfoques más representativos destaca el algoritmo **Latent Dirichlet Allocation (LDA)** desarrollado por Blei et al., el cual modela los documentos como combinaciones probabilísticas de tópicos latentes, y los tópicos como distribuciones de palabras, generando así una estructura semántica interpretable.

Este enfoque ha sido extendido en múltiples direcciones, incluyendo modelos de tópicos neuronales que combinan las ventajas de las representaciones profundas con la necesidad de mantener interpretabilidad. Las aplicaciones del modelado de tópicos abarcan desde el descubrimiento de hipertextos latentes hasta la construcción de índices financieros a partir de análisis de redes sociales.

El marco propuesto en este trabajo amplía dichos usos, aplicándolos específicamente al dominio de la coordinación multiagente, donde los tópicos actúan como vectores compartidos de significado y contexto entre agentes con funciones diferenciadas.

**Sistemas de interacción guiados por tópicos**

La orientación temática del modelo TB-CSPN se alinea con investigaciones recientes que abordan la coordinación basada en tópicos en distintos dominios de aplicación. Este enfoque pone el énfasis en utilizar representaciones semánticas compartidas como mecanismo de estructuración de la interacción entre agentes.

Por ejemplo, Lu et al. han integrado el reconocimiento emocional multimodal con modelos conscientes del contexto temático en el ámbito de la robótica social, lo que permite generar interacciones más sensibles al entorno y al estado emocional del usuario. En una línea similar, Wang et al. desarrollaron mecanismos de selección de conocimiento impulsados por tópicos para la generación de diálogos, logrando así respuestas más relevantes y alineadas con el contexto conversacional.

El concepto de interacción implícita basada en intereses temáticos también ha sido analizado por Hanteer y Rossi, quienes estudiaron cómo los usuarios en redes sociales tienden a agruparse en torno a intereses compartidos. El marco TB-CSPN amplía este principio, aplicándolo a la coordinación explícita entre agentes heterogéneos —humanos, LLMs y agentes de IA especializados— utilizando **distribuciones de tópicos** como mecanismo central para la **formación dinámica de grupos** y la organización de tareas colaborativas.

2.1. Introducción a la teoría organizacional para sistemas multiagente  
  
Los sistemas multiagente actuales enfrentan desafíos crecientes debido a la necesidad de integrar agentes humanos, modelos de lenguaje de gran tamaño (LLMs) y sistemas de inteligencia artificial especializados. Borghoff et al. (2025) sostienen que, en este nuevo paradigma, no basta con optimizar las capacidades individuales de cada agente; es imprescindible establecer una **estructura organizativa** que facilite su colaboración de forma eficiente y coherente.

En este sentido, la teoría organizacional propuesta por estos autores introduce una arquitectura cognitiva y operativa que traslada principios del diseño organizacional humano —como jerarquías funcionales, asignación de roles y normativas compartidas— al ámbito de los sistemas distribuidos compuestos por agentes híbridos. Este enfoque ofrece una base conceptual sólida para analizar y diseñar entornos colaborativos complejos, como el que se persigue con el desarrollo de **BusinessAnalistGPT**.

Dentro de las distintas alternativas existentes, se ha seleccionado el marco **TB-CSPN (Topic-Based Conditional Structured Process Networks)** como referencia teórica principal, por su alineación con los objetivos de este proyecto. Si bien se trata de una propuesta experimental, su carácter innovador responde al propósito de este trabajo: explorar y aplicar las tecnologías más recientes en la coordinación semántica de sistemas multiagente.

2.2. La organización como marco para la interacción  
  
En lugar de tratar a los agentes como entidades aisladas, la teoría los enmarca como actores en una organización distribuida, donde el conocimiento está fragmentado y requiere ser coordinado. Las decisiones emergen no de un solo agente omnisciente, sino del proceso interactivo entre múltiples entidades con información parcial, motivaciones específicas y capacidades heterogéneas.  
  
Este enfoque permite abordar los problemas de coordinación, alineación y validación que suelen afectar los entornos de agentes autónomos. Para ello, se propone una ontología organizacional basada en cinco componentes clave: agentes, tareas, roles, normas y artefactos. Cada elemento está explícitamente representado y vinculado mediante estructuras de control y colaboración que regulan el comportamiento grupal.

La propuesta parte de considerar a los sistemas multiagente no como simples conjuntos de entidades colaborativas, sino como organizaciones sociotécnicas distribuidas, donde cada agente (humano o artificial) posee capacidades, intenciones y restricciones diferenciadas. Esta perspectiva reconoce que la eficacia de los sistemas depende menos de la sofisticación individual de los agentes, y más de los mecanismos de coordinación, regulación y alineación de objetivos entre ellos.  
  
En particular, se destaca que la incorporación de agentes humanos y LLMs en un mismo ecosistema operativo requiere marcos estructurados que guíen las interacciones, minimicen errores y aseguren coherencia semántica. Para ello, la teoría propone adoptar principios del diseño organizacional clásico, como los roles funcionales, la jerarquía de control, y las reglas operativas compartidas, adaptadas a un entorno digital y automatizado.  
  
2.3. Modelos de interacción y arquitectura de referencia  
  
El artículo introduce una arquitectura referencial compuesta por capas funcionales:  
- Capa de percepción y comunicación, donde los agentes reciben información contextual o instrucciones.  
- Capa deliberativa, donde los agentes interpretan sus tareas y contexto.  
- Capa organizacional, que regula la asignación de roles, negociación y resolución de conflictos.  
- Capa de memoria compartida, que almacena estados, decisiones y conocimiento generado colectivamente.  
  
Este modelo se inspira en estructuras organizacionales humanas como los equipos de proyecto, pero adaptadas a la lógica computacional distribuida.  
2.4. Tipología y diseño de agentes  
  
El marco conceptual distingue tres tipos de agentes:  
  
1. Agentes Humanos: aportan juicio contextual, interpretación pragmática y toma de decisiones en condiciones de ambigüedad.  
2. Agentes LLM: procesan información textual compleja, proponen inferencias y sintetizan conocimiento en lenguaje natural.  
3. Agentes Especializados: ejecutan tareas técnicas concretas (p. ej., búsqueda, validación, ejecución de código) y son típicamente sistemas simbólicos o basados en herramientas específicas.  
  
Cada agente puede asumir diferentes roles organizativos dentro de una jerarquía de producción de conocimiento, desde generadores de contenido hasta revisores, integradores o coordinadores.

2.5. Mecanismos de alineación y control  
  
Uno de los aportes más relevantes es la identificación de mecanismos para mitigar el desalineamiento entre agentes, particularmente en la interacción entre humanos y LLMs. Entre los mecanismos propuestos se destacan:  
- Control distribuido con puntos de validación: se permite cierta autonomía, pero con nodos de supervisión o evaluación.

- Objetivos compartidos y planes jerárquicos: cada agente debe actuar en función de un objetivo organizacional superior, estructurado en subtareas que se asignan según competencia y rol.  
- Mediación semántica: protocolos que aseguran que diferentes agentes entiendan los conceptos clave de forma coherente, incluso si sus arquitecturas son distintas (p. ej., humanos vs. LLMs) y protocolos de interacción normativos que definen formas estandarizadas de comunicación, validación, escalado de conflictos y revisión de tareas. Esto evita malentendidos y bucles improductivos típicos en arquitecturas agenticas con LLMs.  
- Mecanismos de autoevaluación y retroalimentación: los agentes deben ser capaces de juzgar la calidad de sus contribuciones, proponer mejoras y coordinar con otros agentes para la resolución de errores.

2.6. Memoria compartida y representaciones externas  
  
Un aspecto innovador del marco es la introducción de una memoria organizacional compartida que almacena todas las interacciones, decisiones, artefactos y planes generados por los agentes. Esta memoria permite:  
  
- Recuperar información contextual.  
- Verificar coherencia entre decisiones pasadas y acciones presentes.  
- Facilitar la incorporación de nuevos agentes a tareas en curso.  
  
Además, se promueve el uso de representaciones estructuradas (diagramas, tablas, flujos) como medio de comunicación preferente entre agentes, frente al uso exclusivo de lenguaje natural, debido a su mayor robustez semántica.

**Una tecnología clave que complementa la idea de memoria organizacional compartida es el uso de técnicas de Retrieval-Augmented Generation (RAG).** Este enfoque combina la generación basada en modelos de lenguaje con la recuperación de información relevante desde una base de conocimientos, funcionando como una memoria extendida. En el contexto de BusinessAnalistGPT, RAG permite que los agentes accedan a documentos históricos, especificaciones funcionales previas o notas de entrevistas anteriores para contextualizar y enriquecer sus respuestas.

Esto no solo mejora la precisión y relevancia de las interacciones, sino que también refuerza la trazabilidad y consistencia organizacional. Integrar RAG en la arquitectura del sistema proporciona a los agentes una capacidad de recuperación semántica sobre representaciones estructuradas y no estructuradas, potenciando el valor de la memoria compartida y promoviendo una toma de decisiones más informada.

2.7. Representaciones y trazabilidad  
  
El marco enfatiza la necesidad de representaciones externas compartidas (documentos, mapas de tareas, logs, planes) como medio de coordinación entre agentes. Estas representaciones permiten:  
- Trazabilidad de decisiones.  
- Incorporación de nuevos agentes a sistemas en ejecución.  
- Validación de resultados por humanos supervisores.  
- Desambiguación en tareas complejas, especialmente si se requiere intervención humana posterior.  
  
2.8. Implicaciones para la construcción de agentes inteligentes como BusinessAnalistGPT  
  
Aplicar esta teoría al caso del asistente BusinessAnalistGPT implica varios beneficios directos:  
- Permite definir roles claros para los distintos componentes (generador de requerimientos, verificador, sintetizador, validador de flujos BPMN).  
- Facilita el escalado modular mediante la incorporación de agentes especializados, manteniendo una estructura organizacional robusta.  
- Mejora la interacción con el usuario final al contar con mecanismos de trazabilidad, interpretación contextual, memoria organizativa y control semántico.  
- Introducir mecanismos de verificación iterativa con feedback ejecutable, tal como ocurre en sistemas de desarrollo colaborativo como MetaGPT o ChatDev.

Además, al tratar la interacción como un proceso sociotécnico más que técnico, se promueve una interfaz más natural con los usuarios humanos, al tiempo que se optimiza la eficiencia operativa del sistema. También se logra una mayor coherencia en la ejecución de tareas complejas, se minimiza la ambigüedad en la interacción humano-IA y se mejora la adaptabilidad del sistema ante cambios o intervenciones humanas

# 3. Metodología

Objetivos e hipótesis: Hipótesis principal: La implementación de un asistente de IA mejora la eficiencia de la recopilación de requisitos en proyectos.

Diseño: Se utilizará un diseño descriptivo, no experimental.

Participantes: Business Analysts (BAs) de una organización de servicios globales.

Instrumentos: Cuestionarios de satisfacción, análisis de tiempos de respuesta, entrevistas estructuradas.

Procedimiento: Fases: diseño, desarrollo, prueba piloto, validación.

**3.1. Enfoque metodológico general**

El diseño metodológico de este trabajo adopta el enfoque propuesto por el marco "Build–Collaborate–Evolve", identificado como una de las estructuras más completas para analizar sistemas de agentes basados en LLMs. Esta metodología permite descomponer el ciclo de vida de un agente en tres dimensiones fundamentales: **construcción**, **colaboración** y **evolución**, abarcando desde el diseño de roles y planificación hasta la evaluación y mejora continua del sistema.

**3.2. Construcción del agente**

Durante la fase de construcción, se definieron perfiles funcionales para los distintos tipos de agentes: *Consultores* (basados en LLM), *Especializados* (orientados a tareas concretas) y *Supervisores* (humanos o centáuricos). Este proceso se estructuró en torno a cuatro pilares interdependientes, que guían el diseño completo del agente:

1. Definición del perfil (2.1.1): identificación de las capacidades, restricciones y tipo de interacción esperado para cada agente, asignando roles específicos dentro del sistema.
2. Mecanismo de memoria (2.1.2): integración de memoria episódica y a largo plazo, incorporando técnicas como RAG para recuperación semántica contextualizada.
3. Capacidad de planificación (2.1.3): desarrollo de módulos que permiten descomponer tareas complejas en subtareas, estableciendo secuencias lógicas y dependencias.
4. Ejecución de acciones (2.1.4): implementación de interfaces que permiten a los agentes llevar a cabo acciones concretas, ya sea a través de generación de texto, invocación de herramientas o interacción con usuarios.

Estos componentes se diseñaron para operar de manera coordinada y modular, permitiendo la extensión y reconfiguración del sistema en función de nuevos requerimientos o contextos de uso.

**3.3. Colaboración multiagente**

Para facilitar la colaboración entre agentes, se adoptó una arquitectura descentralizada, semánticamente mediada y orientada a tópicos, inspirada en el marco TB-CSPN (Topic-Based Conditional Structured Process Networks). Este enfoque permite organizar la interacción entre agentes según capas funcionales (superficie, observación, cómputo) y distribuir las transiciones de tareas en función del contexto temático, manteniendo así la trazabilidad y la alineación semántica entre intenciones humanas y acciones automáticas.

Cada agente, ya sea humano, consultor basado en LLM o especializado, actúa sobre transiciones específicas del sistema, suscribiéndose dinámicamente a tópicos relevantes. Esta estructura favorece una orquestación distribuida, donde los agentes colaboran sin perder su autonomía funcional ni su foco operativo.

Además, se incorporaron elementos del enfoque MAD (Multi-Agent Debate) como mecanismo de validación semántica. Los agentes consultores pueden participar en debates estructurados, donde presentan argumentos y contraargumentos para evaluar propuestas o decisiones generadas. Un agente supervisor o moderador humano puede intervenir para seleccionar la solución más coherente, promoviendo así transparencia, explicabilidad y mejora colectiva del sistema.

La combinación de TB-CSPN y MAD permite una colaboración robusta y dinámica, integrando razonamiento, discusión y ejecución de tareas dentro de un entorno controlado, verificable y alineado con los principios de diseño de sistemas multiagente avanzados.

**3.4. Evolución del sistema**

La capacidad evolutiva del sistema se implementó mediante ciclos de retroalimentación continua, donde los agentes son evaluados y ajustados a partir de los resultados de sus propias decisiones. Se emplearon agentes evaluadores encargados de analizar la calidad de las respuestas generadas, proponiendo ajustes con base en métricas internas de desempeño como coherencia semántica, completitud de requisitos y alineación con la intención del usuario.

Además, se adoptó un enfoque basado en Competitive Co-Evolution, donde múltiples agentes (por ejemplo, consultores) compiten entre sí para proponer soluciones a una misma tarea. Este proceso fomenta una dinámica de mejora continua, ya que cada agente aprende de los errores y aciertos de sus competidores. La comparación entre salidas permite identificar patrones de razonamiento más eficaces, filtrar respuestas subóptimas y ajustar estrategias de generación y verificación.

Esta dinámica se complementa con técnicas de *trial-and-error*, *self-debugging* y *external feedback loops*, permitiendo una evolución adaptativa tanto desde el punto de vista técnico como semántico. El objetivo final es garantizar que el sistema se mantenga flexible, autorregulado y alineado con las expectativas humanas a medida que se enfrenta a contextos cada vez más complejos.

**3.5. Evaluación**

La evaluación del sistema se llevó a cabo mediante herramientas propias desarrolladas para registro (logging), trazabilidad (tracing) y observabilidad, permitiendo monitorear el comportamiento del agente en cada etapa del flujo de interacción. Estas herramientas fueron diseñadas específicamente para capturar eventos clave, tiempos de respuesta, transiciones entre estados y decisiones tomadas por los agentes en cada capa del sistema TB-CSPN.

Además de los datos cuantitativos obtenidos a través de métricas internas, se aplicaron cuestionarios de satisfacción y entrevistas estructuradas a analistas de negocio, con el fin de evaluar la utilidad percibida, la coherencia de los resultados y la facilidad de interpretación de las salidas generadas por el agente.

Si bien en esta primera fase no se incorporaron marcos de evaluación estandarizados como AgentBench, se contempla su utilización en futuras iteraciones del proyecto, con el objetivo de comparar el rendimiento del sistema frente a benchmarks establecidos en tareas de planificación, razonamiento y colaboración multiagente.

**3.6. Proceso experimental e implementacion**

El desarrollo del agente se realizó en cuatro fases:

1. **Diseño conceptual** del sistema y definición de roles de agentes.

El flujo de trabajo se gestiona mediante un mecanismo de selección dinámica de agentes, que permite activar en cada momento a los participantes más adecuados según el contexto de la conversación y el estado del proceso. A continuación se describen los agentes implementados:

* **planning\_agent:** primer agente en activarse; descompone la tarea global en subtareas y coordina la ejecución secuencial del equipo.
* **interviewer\_agent:** actúa como analista de negocio; entrevista al usuario para recopilar los requisitos del proyecto. Utiliza memoria vectorial RAG para adaptar las preguntas al contexto y alimentar una plantilla de especificación funcional.
* **summarizer\_agent:** sintetiza la información obtenida en la entrevista. Entrega un resumen al usuario y valida si desea ampliar la información. En caso afirmativo, reactiva al interviewer\_agent.
* **doc\_writter\_agent:** utiliza el resumen generado para redactar un documento de especificación funcional en formato PDF, siguiendo una estructura estandarizada y exhaustiva.
* **funct\_spec\_checker\_agent:** revisa si el documento generado está completo. Evalúa la cobertura de todas las secciones obligatorias y señala cualquier omisión para su corrección.
* **diagram\_creator\_agent:** genera un diagrama del proceso descrito en el documento funcional, utilizando el lenguaje Mermaid. Escribe el código generado en un archivo .md en canvas.
* **mermaid\_code\_reviewer\_agent:** revisa el código generado, identifica posibles mejoras o faltantes, y valida su completitud. Si no hay observaciones, genera la versión final en .mmd.
* **user\_proxy:** representa la interacción humana dentro del sistema, permitiendo que el flujo conversacional se mantenga controlado y rastreable.

Este diseño, alineado con TB-CSPN y enriquecido por dinámicas de deliberación inspiradas en MAD, garantiza una coordinación semántica efectiva, modularidad organizativa y trazabilidad completa de la información generada por los agentes en cada fase del proceso.

1. **Implementación técnica** del agente BusinessAnalistGPT, utilizando una arquitectura modular.

La implementación técnica del sistema BusinessAnalistGPT se estructuró sobre una arquitectura multiagente modular, seleccionada tras una evaluación comparativa de plataformas líderes en el desarrollo y despliegue de agentes basados en modelos de lenguaje (LLMs). Se analizaron propuestas como **LangChain**, **LlamaIndex**, **MetaGPT**, **ChatDev** y **AutoGen**, cada una con distintos niveles de madurez, flexibilidad y alineación con paradigmas organizativos avanzados.

Además de considerar la arquitectura y filosofía de coordinación de cada framework, se evaluaron también sus capacidades desde el punto de vista del **despliegue, operación y mantenimiento**, según el rol que desempeñan en el ciclo de vida de un agente LLM:

**Herramientas de producción (*Productionization*):** AutoGen se distingue como una plataforma open-source extensible y personalizable que permite a los desarrolladores construir agentes conversacionales multiagente, adaptando flujos de trabajo y módulos a necesidades específicas. Por su parte, **LlamaIndex** ofrece herramientas orientadas al acceso, indexación y recuperación de datos locales, actuando como una infraestructura centrada en la gestión semántica de datos para aplicaciones LLM. **LangChain**, en particular, es un framework open-source ampliamente adoptado que permite construir aplicaciones basadas en LLMs con un alto grado de **extensibilidad y personalización**, facilitando la creación de módulos y flujos de trabajo adaptados a necesidades específicas. Esta flexibilidad ha hecho de LangChain una herramienta popular en prototipado rápido y orquestación de componentes modulares en aplicaciones de IA generativa. **MetaGPT** simula un entorno industrial clásico, donde cada fase del flujo está asignada a un agente con un rol funcional fijo (p. ej., project manager, developer, QA), controlado por un planificador central. Este diseño, aunque eficaz en contextos deterministas, restringe la autonomía local de los agentes y su capacidad para adaptarse a entornos cambiantes. MetaGPT forma parte de una línea de investigación de **sistemas con controlador explícito**, junto con modelos como *The Coscientist* o *LLM-Blender*, donde la toma de decisiones está centralizada o automatizada por un ente superior que controla el flujo completo.

* **Operación y mantenimiento:** Ollama destacan como herramientas que permiten **observabilidad y monitoreo** en tiempo real del comportamiento de los agentes, facilitando el ajuste de prompts, datasets y modelos en base a datos de uso reales y anotaciones. Esta capacidad resulta clave para garantizar robustez y mejora continua en entornos productivos.

**Paradigmas comparados: ReAct vs. TB-CSPN**

El análisis reveló dos enfoques metodológicos claramente diferenciados:

1. **Paradigma ReAct / Reflexión iterativa**, como el que implementan **MetaGPT**, **ChatDev**, **LangChain** donde los agentes ejecutan acciones, reflexionan sobre ellas, y ajustan su razonamiento en ciclos secuenciales. Estos sistemas operan con **roles estáticos y protocolos deterministas**, útiles para tareas estructuradas como desarrollo de software o tareas académicas.
2. **Paradigma temático y deliberativo (TB-CSPN)**, representado por **AutoGen** y **OpenAgents**, donde los agentes colaboran mediante **tópicos compartidos**, roles dinámicos y capas de interacción (superficie, observación, ejecución). Este enfoque facilita la adaptación semántica, la trazabilidad y la orquestación distribuida en entornos con alta incertidumbre.

**AutoGen como sistema deliberativo y adaptable**

**AutoGen**, en cambio, representa un sistema **basado en comunicación estructurada** e interacción deliberativa, más cercano a las necesidades del presente proyecto. Permite la creación de agentes con roles definidos pero reconfigurables, que colaboran mediante controladores de flujo que intercambian argumentos y refinan decisiones colectivamente. Esta arquitectura no sólo está alineada con **TB-CSPN**, sino que también integra principios de **MAD (Multi-Agent Debate)** y **coevolución competitiva**.

AutoGen permite la participación de múltiples agentes en **debates iterativos** estructurados, en los que se contrastan ideas, se corrigen errores, y se construyen soluciones más sólidas y verificables. Esta dinámica ha sido reconocida como fundamental en trabajos recientes como **MADR** y **MDebate**, donde la confrontación estratégica entre agentes estimula el pensamiento divergente, mejora la **alineación semántica** y reduce problemas como la alucinación de LLMs o la convergencia prematura.

Además, el soporte de AutoGen para **coevolución competitiva y adversarial** —por ejemplo, mediante simulaciones de red-teaming entre agentes LLM— refuerza su capacidad para generar agentes más resilientes, adaptativos y estratégicamente inteligentes. Esta propiedad es especialmente relevante para sistemas como BusinessAnalistGPT, que deben operar de forma segura, trazable y alineada con intenciones humanas explícitas.

**Otro punto distintivo es que AutoGen ofrece soporte nativo para el protocolo MCP (Model Context Protocol)**, lo que permite a los agentes integrarse directamente con fuentes de datos externas, herramientas especializadas y flujos de trabajo empresariales mediante enlaces seguros y estandarizados.

**AutoGen como controlador de flujo**

En sistemas multiagente modernos, la capacidad de coordinar agentes sin depender de un orquestador central rígido es clave para lograr escalabilidad, flexibilidad y adaptabilidad semántica. En este contexto, **AutoGen** incorpora varios mecanismos avanzados de **control de flujo conversacional**, entre los que destacan **SelectorGroupChat** y el recientemente adoptado patrón **Swarm**, ambos inspirados por el diseño de **OpenAgents**, un sistema introducido por OpenAI.

**SelectorGroupChat: selección basada en modelos generativos**

**SelectorGroupChat** es una arquitectura de chat grupal que permite a múltiples agentes participar en una conversación estructurada. A diferencia de esquemas secuenciales como RoundRobin, en SelectorGroupChat, **el siguiente agente en hablar se selecciona dinámicamente usando un modelo generativo (por ejemplo, un LLM)**. Este modelo analiza el contexto compartido (historial de conversación, roles y descripciones de los agentes) y determina cuál agente está mejor posicionado para intervenir a continuación.

Las características clave de SelectorGroupChat incluyen:

* **Selección del hablante basada en modelos** (model-based speaker selection).
* **Configuración de roles y descripciones** de cada agente participante.
* **Prevención de turnos repetidos** por el mismo agente (a menos que se permita explícitamente).
* **Funciones de selección personalizables**, que permiten definir cómo se elige el próximo hablante o restringir el conjunto de candidatos.

Este enfoque favorece una colaboración **dinámica y sensible al contexto**, en la que los agentes actúan de forma autónoma pero coherente con el objetivo general de la conversación. Todo el historial conversacional se mantiene como **contexto compartido**, lo que facilita la trazabilidad y permite que nuevas tareas se conecten con sesiones anteriores.

**Swarm: delegación distribuida con control local**

**Swarm** representa un paso más allá en la descentralización del control, al implementar un patrón en el que los agentes pueden **delegar tareas entre sí de manera directa**, sin que exista un coordinador central. Esta estrategia, introducida por OpenAI dentro del sistema **Swarm de OpenAgents**, simula dinámicas de colaboración similares a equipos humanos distribuidos.

En Swarm:

* Cada agente tiene la capacidad de generar un **HandoffMessage**, señalando a qué otro agente transfiere el control de la tarea.
* Todos los agentes **comparten el mismo contexto de conversación**, lo que permite que las decisiones locales tengan impacto global.
* El flujo se organiza en turnos, donde el **agente hablante es seleccionado en función del último mensaje de traspaso (HandoffMessage)**.
* El sistema continúa hasta cumplir una condición de término explícita, como completar la tarea o alcanzar un límite de iteraciones.

Este patrón refleja un sistema donde **la lógica de planificación está distribuida entre los agentes**, promoviendo la autonomía, la colaboración directa y la adaptación en tiempo real, características fundamentales para entornos complejos y no deterministas.

**AutoGen como controlador conversacional**

Dentro del sistema **BusinessAnalistGPT**, AutoGen no actúa como un controlador centralizado tradicional, sino como un **entorno que habilita y media la interacción entre agentes autónomos**, siguiendo una lógica más cercana a **TB-CSPN** que a arquitecturas de planificación secuencial.

Mediante **SelectorGroupChat**, AutoGen permite que los agentes respondan según su competencia contextual, mientras que con **Swarm** se habilita la **delegación horizontal de tareas**. Esta dualidad convierte a AutoGen en una plataforma híbrida, capaz de ejecutar procesos colaborativos que requieren:

* Coordinación semántica a través de tópicos y objetivos compartidos.
* Transferencia de control entre agentes en función del estado del diálogo.
* Colaboración contextualizada y controlada por modelos generativos.

Por tanto, AutoGen puede considerarse un **controlador de flujo conversacional distribuido**, en el que la gestión del proceso no se impone desde un nodo superior, sino que emerge desde las decisiones locales de cada agente dentro de un marco de reglas y objetivos compartidos. Esta propiedad lo convierte en la elección ideal para sistemas alineados con paradigmas como **TB-CSPN** y **MAD**, en los que la semántica, la trazabilidad y la deliberación entre agentes son componentes esenciales.

**Prueba piloto con casos de uso reales en entornos controlados**

La validación del sistema se desarrolló bajo dos enfoques principales: una evaluación comparativa de frameworks LLM, y pruebas piloto con casos de uso representativos, utilizando distintas herramientas y entornos para garantizar funcionalidad, trazabilidad y adecuación al entorno de producción.

**Evaluación práctica de frameworks**

Para comparar frameworks, se llevaron a cabo pruebas utilizando **LangChain**, **LlamaIndex** y **AutoGen**, configurando workflows simples de chatbot que integraban **RAG** y un esquema de **interacción human-in-the-loop**. No se incluyen pruebas con **ChatDev**, **MetaGPT** u **OpenAgents**, ya que su enfoque no se alineaba con el objetivo exploratorio de esta etapa.

**Principales resultados**:

* **LangChain** demostró gran extensibilidad para crear módulos personalizados.
* **LlamaIndex** destacó en manejo de datos locales y consultas semánticas.
* **AutoGen** resultó ser el framework más ágil y sencillo de configurar, con integración directa de agentes conversacionales, memoria contextual y soporte para flujos iterativos con SelectorGroupChat y Swarm.

**Entornos de prueba y herramientas**

Para las pruebas piloto se empleó un **Jupyter Notebook**, aprovechando su flexibilidad, capacidad interactiva para *debugging* y corrección en tiempo real. Simultáneamente se desarrolló una **interfaz web usando Chainlit**, simulando un entorno más cercano a producción, interactivo y representable.

En cuanto a los modelos LLM, se evaluaron tanto soluciones locales como en nube:

* **Modelos locales** desplegados con **Ollama** y **LM Studio**: se probaron modelos de 8B y 12B parámetros (como Llama, DeepSeek R1, Quen, Mistral, Gemma y Phi 4). Aunque alcanzaron resultados aceptables, mostraron tendencia a confundirse o no completar tareas, debido a su naturaleza SLM y las restricciones de hardware. Por razones técnicas no fue posible probar versiones de 32B, las cuales, según benchmarks, se acercan en desempeño a los LLM más avanzados [blog.threatresearcher.com+3collabnix.com+3github.com+3](https://collabnix.com/lm-studio-vs-ollama-picking-the-right-tool-for-local-llm-use/?utm_source=chatgpt.com).
* En la nube, se utilizó exclusivamente **OpenAI GPT‑4o**, con más de 100 billones de parámetros. Este modelo ofreció:
  + Respuestas casi instantáneas sin retardos perceptibles.
  + Tasa de éxito superior al 99 % en tareas de especificación funcional, consultas RAG y workflows colaborativos.

**Pruebas piloto con casos reales**

Se diseñaron tres casos reales que utilizaron agentes orquestados por **AutoGen**, SelectorGroupChat y **Swarm**:

1. **Ai Dialer**: un proyecto de chatbot orientado a ventas telefónicas.
2. **Human Resource Management software**: sistema de gestión de recursos humanos.
3. **Online Apparels Shopping Website**: plataforma de comercio electrónico para prendas.

En cada caso, los agentes entrevistaron usuarios, sintetizaron requisitos, generaron documentos funcionales, verificaron contenidos y crearon diagramas de proceso. Se validaron resultados ante usuarios, confirmándose:

* La **viabilidad funcional** del sistema en escenarios reales de negocio.
* La efectividad de la **coordinación semántica** y la trazabilidad en cada etapa.
* La **flexibilidad** en el uso de flujos de tratamiento y herramientas de debugging.

1. **Evaluación y ajuste**, con base en indicadores de eficiencia, satisfacción y coherencia semántica.

# 4. Resultados

Se espera una reducción en los tiempos de entrega de documentos iniciales y un incremento en la satisfacción de los usuarios.

# 5. Discusión

Los resultados serán comparados con proyectos tradicionales, considerando discrepancias y oportunidades de mejora.

# 6. Conclusiones

(Se incluirá posteriormente)

# 7. Limitaciones y futuras líneas de investigación

Limitaciones: muestras limitadas, dependencia de hardware GPU para la ejecución de modelos avanzados o servicios pagos de cloud .

Futuras líneas: integración completa de procesamiento de voz y aprendizaje continuo.

# 8. Referencias bibliográficas

Borghoff, U. M., Bottoni, P., & Pareschi, R. (2025). An organizational theory for multi-agent interactions integrating human agents, LLMs, and specialized AI. Discover Computing, 28(138). https://doi.org/10.1007/s10791-025-09667-2

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. Advances in Neural Information Processing Systems, 33, 9459–9474. https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

# 9. Índice de figuras

(Se incluirá posteriormente)

# 10. Índice de tablas

(Se incluirá posteriormente)

# 11. Acrónimos

LLM: Large Language Model

SLM: Small Language Model

RAG: Retrieval-Augmented Generation

# 12. Anexos

(Se incluirán documentos de ejemplo, flujos de interacción, etc.)