

# Informe Técnico Detallado: El Modelo de Razonamiento Jerárquico (HRM)

Jordi Blasco Lozano  
Universidad de Alicante

10 de septiembre de 2025

## Resumen

## Índice

<b>1. Introducción al Razonamiento Jerárquico: Un Nuevo Paradigma Frente a las Arquitecturas Actuales</b>	<b>2</b>
1.1. El Desafío del Razonamiento Complejo en la IA	2
1.2. Análisis de las Limitaciones: La Profundidad Computacional Fija de los Transformers y la Fragilidad del Chain-of-Thought (CoT)	2
1.2.1. La Arquitectura Transformer	2
1.2.2. La Técnica Chain-of-Thought (CoT)	2
1.3. Principios Fundamentales del HRM: Inspiración en la Computación Cortical	2
1.4. Tabla Comparativa: HRM vs. Transformer vs. LLM con CoT	3
<b>2. Deconstrucción de la Arquitectura del Hierarchical Reasoning Model</b>	<b>3</b>
2.1. La Estructura Dual: El Módulo de Alto Nivel (H) como Planificador Abstracto	3
2.2. El Módulo de Bajo Nivel (L): Ejecutor de Cómputos Rápidos y Detallados	3
2.3. Interdependencia y Dinámica Temporal: Cómo la Separación de Escalas Temporales Permite la Estabilidad	4
2.4. Convergencia Jerárquica: El Proceso Iterativo Hacia un Equilibrio Computacional	4
<b>3. Mecanismos de Cómputo y “Razonamiento Latente”</b>	<b>4</b>
3.1. El Flujo de Procesamiento en un Único “Forward Pass”	4
3.2. “Pensar en Silencio”: El Proceso de Refinamiento Interno vs. la Externalización Lingüística del CoT	4
3.3. El “Outer Loop Refinement Process”: Un Análisis Profundo del Bucle de Refinamiento Iterativo	4
3.4. Visualización de los Pasos Intermedios del Razonamiento	4
<b>4. Metodología de Entrenamiento y Optimización</b>	<b>5</b>
4.1. Eficiencia de Datos Extrema: Aprendizaje con Muestras Mínimas y Sin Pre-entrenamiento	5
4.2. Supervisión Profunda (Deep Supervision) para la Estabilidad del Entrenamiento	5
4.3. Aproximación de Gradiente de 1 Paso: Evitando la Complejidad de BPTT	5
4.4. Análisis Detallado del Tiempo de Cómputo Adaptativo (ACT)	5
<b>5. Evaluación de Rendimiento y Análisis Comparativo en Benchmarks</b>	<b>5</b>
5.1. Rendimiento en Tareas de Razonamiento Inductivo: Abstraction and Reasoning Corpus (ARC-AGI)	5
5.2. Dominio en Puzzles de Búsqueda Simbólica: Sudoku-Extreme y Maze-Hard	5
5.3. Tabla de Resultados: Comparativa Cuantitativa del HRM	5

<b>6. Análisis Crítico y Discusión de la Comunidad Científica</b>	<b>6</b>
6.1. El Estudio de Ablación de ARC Prize: Cuestionando los Pilares del Rendimiento del HRM	6
6.2. ¿Arquitectura o Proceso de Entrenamiento? El Impacto Real del “Outer Loop” y la Aumentación de Datos . . . . .	6
6.3. Tabla de Ablación: Rendimiento del HRM vs. un Transformer de Parámetros Equivalentes	6
6.4. Recontextualización de las Aportaciones del HRM a la Luz de Nuevas Evidencias . . . . .	6
<b>7. Aplicaciones Potenciales y el Futuro de las Arquitecturas Inspiradas en el Cerebro</b>	<b>6</b>
7.1. Oportunidades en Dominios con Datos Escasos . . . . .	6
7.2. El HRM como Alternativa a la Ley de Escalado: Hacia una IA más Eficiente y Accesible .	7
7.3. El “Renacimiento Recurrente”: El Papel del HRM en las Futuras Tendencias Arquitectónicas	7
<b>8. Conclusión y Vías de Exploración para un TFG</b>	<b>7</b>
8.1. Síntesis de las Contribuciones, Fortalezas y Debilidades del HRM . . . . .	7
8.2. Recomendaciones para el Proyecto de Fin de Grado (TFG) . . . . .	7

# 1. Introducción al Razonamiento Jerárquico: Un Nuevo Paradigma Frente a las Arquitecturas Actuales

## 1.1. El Desafío del Razonamiento Complejo en la IA

El razonamiento, definido como el proceso de concebir y ejecutar secuencias de acciones complejas orientadas a la consecución de un objetivo, sigue siendo uno de los desafíos más críticos y persistentes en el campo de la inteligencia artificial.<sup>1</sup> A pesar de los notables avances logrados por los modelos de lenguaje de gran tamaño (LLMs), las tareas que exigen una planificación deliberada, una manipulación simbólica precisa o una búsqueda exhaustiva con capacidad de retroceso (backtracking) continúan representando una frontera formidable para las arquitecturas actuales.<sup>3</sup>

## 1.2. Análisis de las Limitaciones: La Profundidad Computacional Fija de los Transformers y la Fragilidad del Chain-of-Thought (CoT)

El estado del arte actual se apoya fundamentalmente en dos pilares: la arquitectura Transformer y la técnica de prompting de Cadena de Pensamiento (Chain-of-Thought o CoT). Sin embargo, ambos presentan limitaciones intrínsecas que el Modelo de Razonamiento Jerárquico (HRM) busca superar.

### 1.2.1. La Arquitectura Transformer

A pesar de su éxito revolucionario, la arquitectura Transformer es, paradójicamente, “superficial” en su capacidad computacional.<sup>3</sup> Su profundidad está fijada por el número de capas, lo que la sitúa en clases de complejidad computacional como AC0 o TC0. Esta característica estructural le impide, en teoría, resolver problemas que requieren un tiempo de ejecución polinomial y, por consiguiente, la arquitectura no es Turing-completa.<sup>3</sup> Esta limitación se manifiesta empíricamente en tareas complejas como la resolución de Sudokus de alta dificultad, donde el rendimiento de los modelos Transformer se satura y no mejora sustancialmente, incluso al aumentar drásticamente el número de capas.<sup>3</sup>

### 1.2.2. La Técnica Chain-of-Thought (CoT)

Para mitigar la limitación de profundidad, los LLMs recurren al CoT, una técnica que externaliza el proceso de razonamiento en una secuencia de pasos intermedios expresados en lenguaje natural.<sup>4</sup> Aunque ha demostrado ser efectiva para mejorar ciertas capacidades de razonamiento, esta estrategia adolece de debilidades fundamentales:

- **Fragilidad:** La descomposición de la tarea en pasos lingüísticos es inherentemente frágil. Un único error en la cadena puede desviar por completo el proceso y conducir a una conclusión incorrecta.<sup>1</sup>
- **Dependencia de la Escala:** El CoT es una habilidad emergente que solo funciona de manera fiable en modelos de escala masiva, típicamente con más de 100 mil millones de parámetros, y requiere vastos conjuntos de datos para su entrenamiento.<sup>1</sup>
- **Alta Latencia:** La generación secuencial de tokens para cada paso del “pensamiento” es un proceso lento, lo que resulta en una alta latencia de inferencia, un inconveniente significativo para aplicaciones en tiempo real.<sup>1</sup>
- **Razonamiento Engañoso:** Se ha observado que el razonamiento verbalizado por el modelo no siempre refleja el proceso computacional interno que llevó a la respuesta. Esto genera un problema de confianza e interpretabilidad, ya que una justificación aparentemente lógica puede enmascarar un proceso de inferencia defectuoso.<sup>7</sup>

## 1.3. Principios Fundamentales del HRM: Inspiración en la Computación Cortical

Frente al paradigma dominante de “escalar es todo lo que necesitas”, Sapiient Intelligence, la organización detrás del HRM, propone un cambio de enfoque: en lugar de depender de la fuerza bruta computacional, se debe innovar en la arquitectura, inspirándose en los principios de eficiencia y profundidad computacional perfeccionados por el cerebro humano a lo largo de miles de millones de años de evolución.<sup>11</sup> El HRM se fundamenta en tres principios clave observados en la computación cortical<sup>3</sup>:

- **Procesamiento Jerárquico:** El cerebro organiza la información en una jerarquía de áreas corticales. Las áreas de nivel superior, como la corteza prefrontal, integran información a lo largo de escalas temporales extensas para formar representaciones abstractas y planes estratégicos. En contraste, las áreas de nivel inferior se especializan en el procesamiento sensorial y motor inmediato y detallado.
- **Separación Temporal:** Estos niveles jerárquicos operan en escalas de tiempo intrínsecas distintas, un fenómeno reflejado en los ritmos neuronales (por ejemplo, las ondas lentas theta, 4–8 Hz, y las rápidas gamma, 30–100 Hz). Esta separación temporal es crucial, ya que permite que la planificación abstracta de alto nivel guíe de manera estable los cálculos rápidos y detallados de bajo nivel, sin que estos últimos desestabilicen el proceso global.
- **Conectividad Recurrente:** El cerebro se caracteriza por sus extensas conexiones recurrentes y bucles de retroalimentación. Estos circuitos permiten un refinamiento iterativo de las representaciones neuronales, lo que da lugar a soluciones más precisas y sensibles al contexto a cambio de un mayor tiempo de procesamiento.

#### 1.4. Tabla Comparativa: HRM vs. Transformer vs. LLM con CoT

La siguiente tabla resume las diferencias conceptuales entre los tres enfoques.

Cuadro 1: Comparativa de Paradigmas de Razonamiento en IA.

Característica	HRM	Transformer (Base)	LLM con CoT
Paradigma	Interno, iterativo	Feed-forward, paralelo	Externo, secuencial
Profundidad	Dinámica, ilimitada	Fija, limitada	Extendida artificialmente
Eficiencia Datos	Muy alta	Baja	Muy baja
Pre-entrenamiento	No requerido	Requerido	Requerido
Latencia	Potencialmente baja	Baja	Alta
Base Teórica	Neurociencia	Procesamiento secuencias	Habilidad emergente

La emergencia del HRM no es solo la presentación de una nueva arquitectura, sino una crítica fundamental al paradigma de escalado que ha dominado la investigación en IA. El hecho de que un modelo de solo 27 millones de parámetros, entrenado con un millar de ejemplos, pueda superar a gigantes de cientos de miles de millones de parámetros en tareas específicas de razonamiento<sup>1</sup>, sugiere que la estructura computacional puede ser tan importante, o incluso más, que el tamaño bruto.

## 2. Deconstrucción de la Arquitectura del Hierarchical Reasoning Model

### 2.1. La Estructura Dual: El Módulo de Alto Nivel (H) como Planificador Abstracto

El Módulo de Alto Nivel (H) es una red recurrente diseñada para la planificación lenta y abstracta.<sup>1</sup> Su función principal es establecer la estrategia general o el contexto para la resolución de un problema. Opera en una escala de tiempo más lenta, análoga a lo que la psicología cognitiva denomina “Sistema 2”: el pensamiento deliberado, lento y analítico.<sup>11</sup>

### 2.2. El Módulo de Bajo Nivel (L): Ejecutor de Cálculos Rápidos y Detallados

Complementando al módulo H, el Módulo de Bajo Nivel (L) es otra red recurrente, pero optimizada para ejecutar cálculos rápidos y detallados.<sup>1</sup> Este módulo opera bajo la dirección del estado actual del módulo H. Por cada paso lento que da el módulo H, el módulo L ejecuta múltiples pasos computacionales a alta velocidad.<sup>3</sup> Su funcionamiento es análogo al “Sistema 1” del pensamiento: rápido, automático e intuitivo.<sup>11</sup>

### 2.3. Interdependencia y Dinámica Temporal: Cómo la Separación de Escalas Temporales Permite la Estabilidad

La clave de la arquitectura no reside solo en la existencia de los dos módulos, sino en su interacción dinámica y la estricta separación de sus escalas temporales. La dinámica del modelo se desarrolla a lo largo de  $N$  ciclos de alto nivel. Cada uno de estos ciclos se compone, a su vez, de  $T$  pasos de tiempo de bajo nivel.<sup>3</sup> Durante los  $T$  pasos de un ciclo, el estado oculto del módulo  $H$ ,  $z_H$ , permanece constante, guiando las operaciones del módulo  $L$ . Mientras tanto, el módulo  $L$  actualiza su propio estado oculto,  $z_L$ , en cada uno de sus  $T$  pasos.<sup>13</sup>

### 2.4. Convergencia Jerárquica: El Proceso Iterativo Hacia un Equilibrio Computacional

La interacción entre los módulos  $H$  y  $L$  da lugar a un proceso denominado “convergencia jerárquica”.<sup>4</sup> Dentro de un único ciclo de alto nivel, el módulo  $L$  itera hasta alcanzar un “equilibrio local”.<sup>21</sup> Una vez que  $L$  ha convergido,  $H$  ejecuta un único paso de actualización para revisar su plan estratégico. Este bucle iterativo permite al modelo realizar de forma implícita procesos complejos como la búsqueda en árbol y el backtracking.<sup>3</sup>

## 3. Mecanismos de Cómputo y “Razonamiento Latente”

### 3.1. El Flujo de Procesamiento en un Único “Forward Pass”

Una de las características más notables del HRM es su capacidad para resolver tareas de razonamiento complejas en una única pasada hacia adelante (forward pass).<sup>1</sup> El modelo recibe una entrada completa y, a través de la dinámica interna de sus ciclos recurrentes, su estado converge hacia una representación que codifica la solución.

### 3.2. “Pensar en Silencio”: El Proceso de Refinamiento Interno vs. la Externalización Lingüística del CoT

Este modelo operativo introduce el concepto de “razonamiento latente” o “pensar en silencio”.<sup>21</sup> A diferencia de los modelos con CoT, que necesitan “pensar en voz alta” generando una cadena de tokens<sup>4</sup>, todo el proceso computacional del HRM ocurre en el espacio latente de alta dimensión de sus estados ocultos ( $z_H$  y  $z_L$ ).<sup>22</sup>

### 3.3. El “Outer Loop Refinement Process”: Un Análisis Profundo del Bucle de Refinamiento Iterativo

Un análisis crítico independiente ha revelado que un mecanismo de refinamiento iterativo externo es un componente crucial.<sup>23</sup> Este mecanismo, denominado “outer loop refinement process”, funciona de la siguiente manera:

1. Se ejecuta una pasada completa del modelo HRM, que produce una predicción de salida provisional.
2. Esta predicción se retroalimenta como parte de la entrada para la siguiente pasada del modelo.
3. El proceso se repite varias veces, permitiendo que el modelo refine iterativamente su propia salida.<sup>23</sup>

### 3.4. Visualización de los Pasos Intermedios del Razonamiento

A pesar de que el razonamiento del HRM es latente, no es una caja negra impenetrable. El repositorio de código fuente incluye herramientas que permiten decodificar y observar las salidas intermedias generadas al final de cada segmento del bucle de refinamiento externo, proporcionando una ventana a la interpretabilidad del modelo.<sup>25</sup>

## 4. Metodología de Entrenamiento y Optimización

### 4.1. Eficiencia de Datos Extrema: Aprendizaje con Muestras Mínimas y Sin Pre-entrenamiento

El HRM puede alcanzar un rendimiento de vanguardia con conjuntos de datos de entrenamiento de solo ~1000 ejemplos originales por tarea, sin necesidad de pre-entrenamiento.<sup>1</sup> Sin embargo, esto depende críticamente de una aumentación de datos masiva, donde cada puzle original puede ser aumentado hasta 1000 veces.<sup>23</sup>

### 4.2. Supervisión Profunda (Deep Supervision) para la Estabilidad del Entrenamiento

El HRM emplea “supervisión profunda”.<sup>3</sup> El entrenamiento se estructura en “segmentos”. En cada segmento  $m$ , se calcula una pérdida  $L_m$  y los parámetros se actualizan. El estado oculto final del segmento  $m$ ,  $z_m$ , se “desconecta” del grafo computacional antes de ser usado en el segmento  $m + 1$ , impidiendo que los gradientes retropropaguen entre segmentos.<sup>3</sup>

### 4.3. Aproximación de Gradiente de 1 Paso: Evitando la Complejidad de BPTT

La desconexión de gradientes crea una “aproximación de gradiente de 1 paso”, que evita la necesidad de usar Backpropagation Through Time (BPTT) a lo largo de toda la secuencia.<sup>21</sup> Esto mantiene la huella de memoria constante,  $O(1)$ , independientemente del número de segmentos.<sup>3</sup>

### 4.4. Análisis Detallado del Tiempo de Cómputo Adaptativo (ACT)

El HRM incorpora un mecanismo de Tiempo de Cómputo Adaptativo (ACT), que permite al modelo ajustar dinámicamente el número de segmentos de refinamiento usando Q-learning. Al final de cada segmento, una “Q-head” predice los valores Q para las acciones “continuar” o “detener”.<sup>13</sup>

## 5. Evaluación de Rendimiento y Análisis Comparativo en Benchmarks

### 5.1. Rendimiento en Tareas de Razonamiento Inductivo: Abstraction and Reasoning Corpus (ARC-AGI)

En el benchmark ARC-AGI, el HRM ha demostrado un rendimiento sobresaliente, alcanzando una puntuación del 40.3 % en la versión 1 y 5 % en la versión 2, superando a modelos mucho más grandes.<sup>1,11</sup>

### 5.2. Dominio en Puzzles de Búsqueda Simbólica: Sudoku-Extreme y Maze-Hard

En tareas que requieren búsqueda algorítmica como Sudoku-Extreme y Maze-Hard, el HRM alcanza una precisión casi perfecta, mientras que los LLMs de vanguardia con CoT fallan por completo (0 % de precisión).<sup>1,3</sup>

### 5.3. Tabla de Resultados: Comparativa Cuantitativa del HRM

Cuadro 2: Rendimiento del HRM vs. otros modelos.

Modelo	Parámetros	ARC-AGI-1	Sudoku-Extreme	Maze-Hard
HRM	~27M	40.3 %	~100 %	~100 %
OpenAI o3-mini-high	>100B	34.5 %	0 %	0 %
Claude 3.7 8K	>100B	21.2 %	0 %	0 %
DeepSeek R1	>100B	<34.5 %	0 %	0 %

## 6. Análisis Crítico y Discusión de la Comunidad Científica

### 6.1. El Estudio de Ablación de ARC Prize: Cuestionando los Pilares del Rendimiento del HRM

El equipo de ARC Prize replicó los resultados del HRM, confirmando una puntuación impresionante del 32 % en su conjunto de datos semi-privado, pero sus estudios de ablación recontextualizaron las contribuciones del modelo.<sup>23</sup>

### 6.2. ¿Arquitectura o Proceso de Entrenamiento? El Impacto Real del “Outer Loop” y la Aumentación de Datos

El análisis de ARC Prize arrojó tres hallazgos principales<sup>23</sup>:

- **Impacto Mínimo de la Arquitectura Jerárquica:** Reemplazar el núcleo H-L por un Transformer estándar de tamaño similar, manteniendo el resto del pipeline, resultó en un rendimiento comparable.
- **El “Outer Loop” como Motor Principal:** El “outer loop refinement process” fue el principal responsable de las ganancias de rendimiento.
- **Rol Crítico de la Aumentación y la Memorización:** La aumentación de datos fue crítica, y gran parte del rendimiento en ARC provenía de la memorización efectiva de las tareas de entrenamiento y sus variantes.

### 6.3. Tabla de Ablación: Rendimiento del HRM vs. un Transformer de Parámetros Equivalentes

Cuadro 3: Rendimiento conceptual según arquitectura y bucles de refinamiento.

Arquitectura	Nº Bucles Externos	Puntuación ARC-AGI-1 (Aprox.)
HRM	1	Base
Transformer Baseline	1	Ligeramente inferior a HRM
HRM	>1	Aumento significativo
Transformer Baseline	>1	Aumento significativo (comparable a HRM)

### 6.4. Recontextualización de las Aportaciones del HRM a la Luz de Nuevas Evidencias

Estos hallazgos no invalidan la eficacia del sistema HRM, pero sí desplazan el foco de su innovación. La contribución principal podría ser el pipeline de entrenamiento de refinamiento iterativo, que parece ser agnóstico a la arquitectura del núcleo computacional que utiliza.<sup>24</sup>

## 7. Aplicaciones Potenciales y el Futuro de las Arquitecturas Inspiradas en el Cerebro

### 7.1. Oportunidades en Dominios con Datos Escasos

El HRM es adecuado para dominios donde los datos son escasos, como<sup>11</sup>:

- **Salud:** Apoyo al diagnóstico de enfermedades raras.
- **Pronóstico Climático:** Precisión del 97 % en pronósticos subestacionales a estacionales (S2S).
- **Robótica:** Su tamaño compacto lo hace ideal para ser un “cerebro de decisión” embarcado (on-device).

## 7.2. El HRM como Alternativa a la Ley de Escalado: Hacia una IA más Eficiente y Accesible

El HRM ofrece un camino hacia sistemas de razonamiento potentes que priorizan la “elegancia arquitectónica” sobre la “fuerza bruta computacional”, lo que podría democratizar el acceso a la IA de vanguardia y reducir su huella ambiental y económica.<sup>11,20</sup>

## 7.3. El “Renacimiento Recurrente”: El Papel del HRM en las Futuras Tendencias Arquitectónicas

El HRM es parte de una tendencia emergente que reevalúa arquitecturas recurrentes y basadas en estado como alternativas a los Transformers, como el Universal Transformer, Mamba, RWKV y RecurrentGemma.<sup>14,15</sup>

# 8. Conclusión y Vías de Exploración para un TFG

## 8.1. Síntesis de las Contribuciones, Fortalezas y Debilidades del HRM

- **Contribuciones Clave:** Demostración de que un modelo pequeño puede alcanzar un rendimiento de vanguardia en tareas de razonamiento, desafiando el paradigma de escalado.
- **Fortalezas:** Eficiencia en el uso de datos, rendimiento excepcional en búsqueda simbólica y potencial para aplicaciones de baja latencia.
- **Debilidades y Controversias:** La atribución de su éxito (arquitectura vs. pipeline de entrenamiento) es el principal punto de controversia.

## 8.2. Recomendaciones para el Proyecto de Fin de Grado (TFG)

Se proponen varias líneas de investigación viables para un TFG:

- **Línea 1 (Replicación y Verificación Empírica):** Replicar los hallazgos del estudio de ablación de ARC Prize para verificar empíricamente si el rendimiento de la arquitectura HRM y un Transformer es comparable bajo el mismo pipeline de entrenamiento.
- **Línea 2 (Exploración Arquitectónica del “Núcleo”):** Experimentar reemplazando el núcleo computacional con otras arquitecturas recurrentes (Mamba, RWKV, LSTM) para investigar la universalidad del enfoque de entrenamiento de refinamiento.
- **Línea 3 (Aplicación y Generalización a Nuevos Dominios):** Aplicar el modelo a nuevos dominios de razonamiento algorítmico, como la resolución de problemas de programación de nivel introductorio o el análisis de finales de partidas de ajedrez.
- **Línea 4 (Análisis de Interpretabilidad y Visualización):** Desarrollar nuevas técnicas para analizar y visualizar los estados ocultos de los módulos a lo largo de los bucles de refinamiento para responder preguntas sobre qué conceptos aprende el modelo en cada paso.