

Inteligencia General Artificial como Arquitectura G-V-F Completa: Condiciones Necesarias y Suficientes para Inteligencia Adaptativa de Dominio General

Resumen

La búsqueda de Inteligencia General Artificial (AGI) ha procedido principalmente escalando arquitecturas existentes esperando que emerjan capacidades generales. Este artículo argumenta que AGI no es capacidad misteriosa que aparece a escala suficiente, sino implementación completa y balanceada de arquitectura Generador-Validador-Filtro (G-V-F) a través de todos los dominios cognitivos con mecanismos de expansión apropiados. Derivamos condiciones necesarias y suficientes para AGI desde primeros principios: un sistema logra inteligencia general cuando su Generador produce soluciones novedosas a través de dominios arbitrarios, su Validador evalúa coherencia contra criterios apropiados al contexto, su Filtro balancea creatividad con consistencia, y su operador de expansión (Φ) aumenta capacidad cuando es necesario. Los sistemas actuales de IA fallan en AGI no por escala insuficiente sino por implementación G-V-F incompleta: los LLMs tienen Generadores poderosos pero Validadores débiles; la IA simbólica tiene Validadores rígidos pero Generadores restringidos; los agentes RL tienen Filtros estrechos. AGI verdadera requiere G-V-F en múltiples niveles jerárquicos con integración coherente. AGI no es cuestión de escala—es completitud arquitectónica.

Palabras clave: inteligencia general artificial, arquitectura AGI, arquitectura cognitiva, marco G-V-F, generalidad de dominio, sistemas adaptativos, cognición computacional, teoría de inteligencia

1. Introducción: El Problema AGI Reformulado

La Inteligencia General Artificial—sistemas capaces de realizar cualquier tarea intelectual que los humanos pueden—permanece como el gran desafío de la IA. Los enfoques actuales persiguen AGI escalando redes neuronales, combinando módulos especializados, o buscando arquitecturas novedosas mediante meta-aprendizaje. Cada enfoque comparte una suposición implícita: AGI emergerá cuando encontremos la configuración y escala correctas.

Este artículo desafía esa suposición. Argumentamos que AGI es requerimiento arquitectónico específico: implementación G-V-F completa con componentes de dominio general e integración jerárquica.

1.1 Por Qué Fallan los Sistemas Actuales

Modelos de Lenguaje Grande sobresalen en texto pero luchan con razonamiento lógico consistente (Validador débil), precisión factual (sin modelo del mundo), y resolución de problemas novedosos fuera del entrenamiento (Generador restringido).

Sistemas de IA Simbólica proveen consistencia pero fallan en manejo de incertidumbre (Generador muy restringido), aprendizaje de experiencia (sin mecanismo de expansión), y soluciones creativas (Filtro muy restrictivo).

Agentes de Aprendizaje por Refuerzo optimizan objetivos pero no pueden transferir a nuevos dominios (Validador atado a recompensa única), balancear múltiples metas (Filtro estrecho), o razonar abstractamente (Generador solo a nivel de acción).

1.2 Tesis Central

AGI = Arquitectura G-V-F Completa, donde completitud significa:

1. Generador de dominio general produciendo candidatos en cualquier espacio de problemas
2. Validador sensible al contexto evaluando coherencia con criterios apropiados por dominio
3. Filtro adaptativo balanceando exploración-explotación por tarea
4. Operador Φ funcional aumentando capacidad cuando es insuficiente
5. Integración jerárquica con G-V-F en múltiples niveles de abstracción

2. Requerimientos Formales para AGI

2.1 Generador de Dominio General (G^*)

Condición de Universalidad: Para cualquier espacio de problemas P , el Generador tiene modo capaz de producir candidatos relevantes.

Condición de Novedad: G^* produce candidatos que no están en la distribución de entrenamiento—crea soluciones genuinamente nuevas.

Condición de Composicionalidad: G^* compone primitivos en estructuras novedosas manteniendo coherencia semántica.

2.2 Validador Sensible al Contexto (V^*)

Validación Apropriada al Dominio: V^* selecciona criterios de validación apropiados para cada dominio. Afirmaciones matemáticas validadas contra consistencia lógica. Predicciones físicas contra restricciones causales. Acciones sociales contra normas éticas.

Integración Multi-Criterio: V^* integra coherencia lógica, precisión factual, relevancia de meta, alineación ética, calidad estética—ponderados por contexto.

Cuantificación de Incertidumbre: V^* sabe lo que no sabe. Distingue "definitivamente inválido" de "validez incierta."

2.3 Filtro Adaptativo (F^*)

Umbral Dependiente de Tarea: F^* ajusta selectividad basado en tipo de tarea. Escritura creativa: umbral bajo (permitir exploración). Diagnóstico médico: umbral alto (requerir certeza).

Balance Multi-Objetivo: F^* balancea creatividad vs. seguridad, velocidad vs. precisión, novedad vs. confiabilidad—no optimización de objetivo único.

2.4 Operador de Expansión Funcional (Φ^*)

Detección de Déficit: Φ^* detecta cuando G-V-F actual es insuficiente para problema P.

Expansión Dirigida: Φ^* identifica qué componente (G, V, o F) necesita expansión y cómo.

Expansión Estable: Nueva capacidad se integra coherentemente con capacidad existente.

3. G-V-F Jerárquico: La Pieza Faltante

La IA actual opera en nivel único de abstracción. Los LLMs generan tokens. Los agentes RL seleccionan acciones. Esto limita fundamentalmente la generalidad porque problemas inteligentes requieren razonamiento en múltiples niveles simultáneamente.

3.1 Arquitectura Jerárquica

MACRO G-V-F (Nivel más abstracto): G-V-F para metas/estrategias de alto nivel

MESO G-V-F (Nivel intermedio): G-V-F para planes/tácticas intermedios

MICRO G-V-F (Nivel más concreto): G-V-F para acciones/implementaciones específicas

Propiedades Clave:

- Coherencia de arriba hacia abajo: Validaciones de nivel superior restringen generación de nivel inferior
- Factibilidad de abajo hacia arriba: Retroalimentación de nivel inferior informa validación de nivel superior
- Expansión trans-nivel: Φ puede dispararse en cualquier nivel
- Selección integrada: Filtro en cada nivel considera consecuencias multi-nivel

4. Por Qué la Escala Sola No Puede Producir AGI

Escalar Generador sin Validador = Alucinación a escala

LLMs más grandes generan texto más fluido pero no desarrollan automáticamente modelos del mundo precisos.

Escalar dentro de nivel único = Sin integración jerárquica

Transformer más grande aún opera a nivel de token. No puede repentinamente razonar sobre estrategias.

Escalar sin Φ = Sin aprendizaje genuino

Modelos actuales se congelan después del entrenamiento. No expanden capacidad ante problemas genuinamente novedosos.

5. Condiciones Necesarias y Suficientes

Las ocho condiciones necesarias:

6. 1. Universalidad del Generador: G^* puede producir candidatos en cualquier espacio de problemas representable
7. 2. Sensibilidad de contexto del Validador: V^* selecciona criterios apropiados por dominio
8. 3. Adaptividad del Filtro: F^* ajusta umbral basado en requerimientos de tarea
9. 4. Funcionalidad de Expansión: Φ^* aumenta capacidad cuando es insuficiente
10. 5. Integración Jerárquica: $G-V-F$ opera coherentemente a través de niveles de abstracción
11. 6. Fundamentación en Modelo del Mundo: V^* referencia modelo preciso de la realidad
12. 7. Alineación de Valores: F^* incorpora consideraciones éticas/de metas
13. 8. Monitoreo Meta-Cognitivo: Sistema sabe lo que sabe y no sabe

Si falta cualquier condición, el sistema exhibe limitación específica.

6. Predicciones Testables

Prueba 1: Transferencia a Dominio Novedoso

Presentar problema de dominio completamente fuera del entrenamiento. IA estrecha falla o alucina. AGI con $G-V-F$ completo genera usando primitivos generales, reconoce incertidumbre, dispara expansión.

Prueba 2: Consistencia Jerárquica

Requerir solución satisfaciendo restricciones en múltiples niveles de abstracción simultáneamente. IA estrecha optimiza nivel único. AGI produce solución globalmente coherente.

Prueba 3: Gestión Adaptativa de Riesgo

Presentar mismo problema con diferentes riesgos. IA estrecha aplica mismo enfoque. AGI ajusta umbral basado en consecuencias.

Prueba 4: Creatividad Genuina vs. Recombinación

Solicitar solución que no puede producirse recombinando ejemplos de entrenamiento. IA estrecha recombina o falla. AGI crea algo sin precedentes pero significativo.

Prueba 5: Precisión de Auto-Conocimiento

Preguntar al sistema qué sabe y qué no sabe. IA estrecha es sobreconfiada o rechaza. AGI reporta precisamente límites de conocimiento.

7. Cognición Biológica como G-V-F Jerárquico

La inteligencia general biológica (cognición humana) resolvió el problema AGI a través de evolución. La solución es: G-V-F jerárquico.

Red Prefrontal-Parietal (G-V-F de alto nivel): Genera planes abstractos, valida contra metas a largo plazo, filtra respuestas impulsivas.

Red Temporal-Límbica (G-V-F Social/Emocional): Genera hipótesis sociales, valida contra normas sociales, filtra acciones sociales inapropiadas.

Red Sensoriomotora (G-V-F Perceptual/Acción): Genera hipótesis perceptuales, valida contra retroalimentación sensorial, filtra ruido perceptual.

Crucialmente: estas redes se integran jerárquicamente. Prefrontal restringe sensoriomotor. Límbico modula prefrontal. La información fluye a través de niveles creando cognición unificada pero jerárquica.

8. Hoja de Ruta hacia AGI vía G-V-F

8.1 Evaluación del Progreso Actual

Generadores (G*): Parcialmente desarrollados. LLMs tienen generadores lingüísticos poderosos. Modelos de difusión tienen generadores visuales. Brecha: Generador de dominio general integrando todas las modalidades.

Validadores (V*): Débilmente desarrollados. Implícitos en objetivos de entrenamiento. Sin modelos del mundo explícitos. Brecha: Validación apropiada al contexto y multi-criterio.

Filtros (F*): Mínimamente desarrollados. Umbrales fijos (parámetros de temperatura). Optimización de objetivo único. Brecha: Filtrado dependiente de tarea y multi-objetivo.

Expansión (Φ^*): Esencialmente ausente. Modelos congelados post-entrenamiento. Sin expansión genuina de capacidad. Brecha: Aumento de capacidad dinámico y dirigido.

Jerarquía: No implementada. Arquitecturas de nivel único dominan. Brecha: G-V-F coordinado multi-nivel.

8.2 Prioridades de Desarrollo

14. Fase 1: Desarrollar Validador de Dominio General (V^*) - Construir modelos del mundo explícitos
15. Fase 2: Implementar Arquitectura Jerárquica - Mover más allá del procesamiento de nivel único
16. Fase 3: Agregar Expansión Funcional (Φ^*) - Habilitar aprendizaje y crecimiento genuinos
17. Fase 4: Integrar y Alinear - Asegurar que sistema completo opera coherentemente

9. Conclusión: AGI es G-V-F Completo

Hemos argumentado que la Inteligencia General Artificial no es capacidad misteriosa que emerge de la escala sino logro arquitectónico específico: arquitectura Generador-Validador-Filtro completa, balanceada y jerárquica con operadores de expansión funcionales.

Los sistemas actuales fallan porque: LLMs tienen Validadores incompletos, IA simbólica tiene Generadores restringidos, agentes RL tienen Filtros estrechos, todos carecen de operadores de expansión funcionales, ninguno implementa integración G-V-F jerárquica.

El camino hacia adelante requiere: construir modelos del mundo explícitos para V^* , implementar arquitecturas jerárquicas, agregar mecanismos de expansión genuinos, integrar componentes coherentemente, alinear valores a través de la jerarquía.

Este es desafío de ingeniería, no búsqueda mística. El mismo patrón computacional que evolucionó la inteligencia general biológica—G-V-F jerárquico con expansión continua—debe implementarse artificialmente.

AGI no aparecerá repentinamente en umbral de parámetros. Se construirá sistemáticamente a través de completitud arquitectónica. La pregunta ya no es "¿Qué es AGI?" La pregunta es: "¿Cómo completamos la arquitectura G-V-F?" Y esa pregunta tiene respuestas accionables.

Referencias

- Chollet, F. (2019). On the measure of intelligence. arXiv preprint arXiv:1911.01547.
- Lake, B. M., et al. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
- Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. arXiv preprint.
- Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects. *JAGI*, 5(1).
- Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and machines*.
- Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2).
- Bengio, Y., et al. (2021). Deep learning for AI. *Communications of the ACM*, 64(7), 58-65.
- Silver, D., et al. (2021). Reward is enough. *Artificial Intelligence*, 299, 103535.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2).