

Documento de Fundamentos del Sistema de Dirección IA

Una derivación desde primeros principios de la dirección como variable determinante de la salida en sistemas generativos

Dani Collada

Investigador independiente · Dani Collada IA

ORCID 0009-0000-5026-6673

DOI 10.5281/zenodo.20765017 · <https://doi.org/10.5281/zenodo.20765017>

Dominio Dirección de sistemas generativos · Especificación e infraespecificación · Economía de la información

Tipo Working paper — capa de fundamento. No revisado por pares.

Versión v1.2 · Junio de 2026

Capa Fundamento. Autoridad operativa: cero. Subordinación: Biblia v2.5 > Sistema Operativo v1.3 > Manual v4.

Licencia CC BY-NC-ND 4.0

Collada, D. (2026). Documento de Fundamentos del Sistema de Dirección IA (v1.2). Working paper. Zenodo. <https://doi.org/10.5281/zenodo.20765017>

0. Nota de capa y autoridad

Este documento no decide nada. No fija invariantes, no emite instrucciones ejecutables, no se consulta en chats de trabajo y no altera ninguna decisión de producción. Su autoridad operativa es nula por diseño y no por omisión.

Su lugar en la jerarquía es preciso y no admite lectura alternativa. La Biblia decide el universo: qué se construye y qué no. El Sistema Operativo decide el comportamiento que ejecuta esas decisiones. El Manual es la referencia de uso diario. El Documento de Fundamentos hace una sola cosa, situada por debajo de las tres anteriores y nunca a su lado: justifica, desde primeros principios, por qué el universo está bien decidido. Es la cantera de la que la Biblia extrae argumentos, no una instancia que pueda discutirlos.

De ahí una regla dura que rige la lectura de todo lo que sigue. Cuando este documento y la Biblia parezcan entrar en conflicto, gana la Biblia de forma automática, y el conflicto se trata como defecto de este documento. La razón es estructural: un texto de fundamento solo puede derivar lo que la autoridad ya ha decidido; no puede proponer una alternativa a esa decisión sin dejar de ser fundamento y convertirse en disputa. Si el lector encuentra aquí una afirmación que la Biblia contradice, la afirmación es errónea, no la Biblia.

El documento tiene, sin embargo, una función que la Biblia no puede cumplir, y esa función es la única razón por la que existe como pieza separada. La Biblia convence

por autoridad a quien ya está dentro del proyecto. Este documento debe convencer por principio a quien está fuera de él: un evaluador escéptico, un interlocutor técnico, un cliente futuro. Está escrito, por tanto, para sostenerse sin acceso a la Biblia. Todo lo que sigue debe poder ser juzgado por su propia coherencia interna.

1. El vacío que este documento llena

El método de dirección que gobierna esta práctica está hoy justificado de dos maneras, y ambas son insuficientes para lo que el método necesita ser.

La primera justificación es la autoridad. La Biblia se declara a sí misma un sistema de decisiones y afirma, por mandato, qué hay que construir. Esto basta para producir: un equipo que acepta la autoridad ejecuta sin necesitar más. No basta para un interlocutor externo. A un evaluador de subvención, a un revisor escéptico o a un cliente potencial no se le puede ofrecer como argumento que la Biblia lo diga. La autoridad es interna y no transfiere credibilidad fuera de su propio perímetro.

La segunda justificación es la demostración: el método funciona, y los resultados lo respaldan. Esto es más fuerte que la autoridad, pero adolece de un defecto distinto, el de la portabilidad. Un método justificado solo por sus resultados y por la voluntad de quien lo decide se lee como una lista de control atada a un dominio concreto —en este caso, la producción de vídeo—. Nada en esa justificación explica por qué el método seguiría siendo válido si cambiara la herramienta o el dominio. Queda mudo precisamente donde más se le exigiría hablar.

El vacío exacto es este: el método está justificado por autoridad y por demostración, pero no por principio. La autoridad convence al equipo; la demostración convence al observador presente; solo el principio convence a un escéptico que no ha visto los resultados y sobrevive al cambio de dominio. Este documento se propone convertir el enunciado «funciona porque lo decido y porque lo demuestro» en el enunciado «funciona porque, dado cómo se comporta cualquier sistema generativo entrenado sobre un corpus amplio y compartido, no puede no funcionar». Un método sin esta capa es frágil ante el cambio de herramienta y mudo ante quien no está ya dentro.

2. Contribución

Esta sección declara, de forma explícita, qué afirma este documento que no estaba dicho así antes. La declaración es necesaria porque un lector académico tiene derecho a saber qué se reclama como aportación y qué se toma prestado de terceros.

El documento no reclama haber descubierto que los sistemas generativos tienden, sin restricciones, hacia la salida más probable de su corpus. Esa propiedad es conocida. Tampoco reclama originalidad en sus tres linajes intelectuales —la economía de la abundancia, la augmentación humana y el principio informacional de especificación—, que son préstamos atribuidos en su lugar.

Lo que el documento reclama como aportación propia es una articulación que, hasta donde alcanza el conocimiento de quien escribe, no había sido formulada en estos

términos. Primero, tratar la tendencia central de un sistema generativo compartido no como un defecto a corregir, sino como el comportamiento correcto frente al cual la dirección es la única excepción posible, y derivar de ahí, y no de un mandato, que dirigir precede a ejecutar. Segundo, distinguir con precisión la tendencia estadística de un sistema individual de la convergencia social entre actores que consultan la misma capacidad con peticiones igualmente poco especificadas —lo que aquí se nombra promedio automatizado—, y sostener que es esta segunda convergencia, y no la primera, la que erosiona la diferenciación en un mercado. Tercero, y centralmente, derivar las nueve piezas del sistema como dimensiones de subespecificación —ejes de decisión que el sistema, si no se le especifican, rellena con el valor por defecto del corpus— en lugar de presentarlas como pasos de un procedimiento, y someter su número a una prueba de falsación explícita en vez de afirmarlo.

Ninguna de estas tres afirmaciones es un teorema. La tercera, en particular, se ofrece como conjetura falsable y no como prueba, según se detalla en su sección y se recoge en los límites. La contribución del documento es articular, derivar y exponer a refutación; no es demostrar de forma concluyente.

Relación con el trabajo previo

Las dos afirmaciones empíricas sobre las que se apoya este documento —que un sistema generativo subespecificado tiende al centro de su distribución, y que el uso compartido de la misma capacidad homogeneiza las salidas de actores distintos— no son observaciones nuevas. Existe un cuerpo de trabajo que las ha documentado, y delimitar la contribución frente a él es parte de declararla.

La tendencia a perder los extremos de la distribución y a concentrarse en su centro ha sido documentada en la línea del colapso de modelos: Shumailov y colaboradores (2024) muestran que entrenar de forma recurrente un sistema generativo sobre datos generados por él mismo hace desaparecer las colas de la distribución original. Es un fenómeno adyacente pero distinto del que aquí se deriva —aquel describe una dinámica de entrenamiento a lo largo de generaciones; este, la convergencia bajo especificación insuficiente en una sola generación—, y el documento no afirma que compartan mecanismo: ambos señalan al centro, por vías diferentes.

La convergencia entre actores —lo que aquí se nombra promedio automatizado— está documentada de forma más directa. Doshi y Hauser (2024) hallan que el acceso a ideas de un mismo modelo aumenta la creatividad individual pero reduce la diversidad colectiva: las salidas de autores distintos se parecen más entre sí. Padmakumar y He (2024) miden el mismo efecto en escritura asistida: el uso compartido de un modelo ajustado por retroalimentación reduce la diversidad de contenido entre autores. Y la línea de la monocultura algorítmica —Kleinberg y Raghavan (2021), y Bommasani y colaboradores (2022)— formaliza, en contextos de decisión, que muchos agentes que dependen del mismo sistema correlacionan sus resultados y producen homogeneización.

Frente a este cuerpo de trabajo, la contribución de este documento es precisa y de otra naturaleza. La literatura mide los fenómenos empíricamente y, en la línea de la

monocultura, analiza sus consecuencias; lo hace, además, sobre modelos y tareas concretos. Este documento no aporta una medición más: aporta una derivación desde un único principio, agnóstico a la herramienta, del que ambos fenómenos se siguen; reencuadra la convergencia entre actores como el mecanismo que erosiona la diferenciación en un mercado; y organiza la palanca correctora —la especificación, la dirección— en un marco de nueve piezas sometido a una conjetura de completitud falsable. La literatura establece que el problema existe y lo cuantifica; no ofrece ni su derivación desde principio ni una estructura falsable de la palanca que lo revierte. Ahí se sitúa lo propio.

3. El principio de convergencia

Todo el argumento descansa sobre un único principio, enunciado a un nivel que no depende de ninguna herramienta concreta.

Principio. Cualquier sistema que genere su salida muestreando la distribución de un corpus amplio y compartido, en ausencia de restricciones que aíslen una región específica de esa distribución, devuelve la tendencia central de la distribución.

Conviene subrayar de inmediato lo que este principio no dice. No dice que el sistema falle. La tendencia central es el comportamiento correcto del sistema, no su error: ante una petición que no excluye nada, devolver lo más representativo del corpus es exactamente lo que el sistema debe hacer. La herramienta no se equivoca cuando entrega el promedio; trabaja como debe. Esta precisión importa porque ordena el resto del documento: si el promedio fuera un defecto, la solución sería arreglar la herramienta; como es el comportamiento correcto, la única palanca disponible está antes de la herramienta, en lo que se especifica.

El principio se desdobra en dos afirmaciones que el documento mantiene separadas para no confundirlas.

La primera es estadística y se refiere a un sistema individual. Sin especificación suficiente, la salida de un sistema generativo tiende a la moda de su distribución. Es una propiedad de qué se devuelve cuando no se ha excluido nada.

La segunda es social y de mercado, y se refiere a muchos actores a la vez. Cuando numerosos actores consultan la misma capacidad compartida con peticiones igualmente poco especificadas, sus salidas convergen en la misma región central de la distribución. Esa convergencia entre actores —no la tendencia de un solo sistema— es lo que aquí se denomina promedio automatizado: no es que cada salida sea mediocre, es que las salidas de actores distintos se vuelven intercambiables porque ninguno ha excluido lo mismo que los demás conservan. La diferenciación se colapsa no por mala ejecución, sino por especificación insuficiente compartida.

Queda por salvar un paso que el principio, tal como se ha enunciado, no cubre todavía, y que conviene derivar en lugar de afirmar. El principio habla de la moda de un corpus amplio y compartido; la tesis que se construye después habla del promedio de un sector. No son lo mismo: la moda incondicionada del corpus global —lo más probable sin más— no es el promedio de un nicho concreto, y deslizarse de una a otro

sin premisa sería un salto injustificado. El puente es el condicionamiento. Una petición casi nunca es, en la práctica, del todo incondicionada: incluso poco especificada, llega formulada en el vocabulario, el encuadre y las convenciones de un dominio. Pedir «un vídeo de recetas», «un documental de naturaleza» o «un post de marca» no excluye nada dentro de ese dominio, pero sí selecciona el dominio. Esa selección condiciona la distribución de la que el sistema muestrea: ya no devuelve la moda global, sino la moda de la distribución condicionada a ese sector. Y la moda condicionada a un sector, bajo una petición típica y poco especificada de ese sector, es precisamente lo que ese sector produce por defecto: su convención.

De ahí se deriva —y no se afirma— que la tendencia central relevante para la diferenciación no es el promedio de todo el corpus, sino la convención del nicho, lo que en clave de mercado equivale al promedio del sector. La versión social de este mismo hecho es el promedio automatizado ya descrito: cuando muchos actores de un nicho consultan la misma capacidad con peticiones igualmente típicas de ese nicho, todos caen en esa misma moda condicionada, y sus salidas se vuelven intercambiables. La subespecificación dentro de un dominio ya seleccionado devuelve la convención de ese dominio.

Lo único que mueve una salida fuera de esa región central es la especificación que excluye regiones de la distribución. Especificar es, en este marco, prohibir: cada decisión que retira una parte del espacio de salidas posibles desplaza el resultado lejos del centro. Las decisiones que hacen ese trabajo —intención, contexto, identidad, referencias y restricciones— son la dirección.

El principio es independiente de qué modelo o herramienta implementa la distribución. Se cumplió en sistemas generativos anteriores y se cumplirá en los siguientes, porque es una propiedad de «muestrear un corpus compartido bajo especificación insuficiente» y no de ninguna arquitectura concreta. Por esta misma razón, el documento se prohíbe explicar el mecanismo por la maquinaria interna de ningún sistema. No se apela aquí a la arquitectura de los modelos ni a sus procedimientos de entrenamiento: hacerlo ataría el argumento a una generación de herramientas, caducaría con ella y rebajaría el texto a divulgación técnica. El mecanismo se enuncia, y se enunciará siempre, como convergencia a la moda de un corpus compartido bajo especificación insuficiente.

4. La dirección como única variable que mueve la salida fuera del centro

Tesis 1: la IA no empieza en el prompt, empieza en la dirección.

La tesis se deriva directamente del principio anterior y no se sostiene por sí sola. Si un sistema sin dirección especificada devuelve la salida más probable, y la salida más probable —bajo una petición típica del dominio, según la premisa puente de la sección anterior— es la convención del sector, entonces lo único capaz de producir una salida no promedio es aquello que se especifica antes de ejecutar. A esa especificación previa la llamamos dirección. La consecuencia es que la dirección no es un refinamiento de la

ejecución: es la condición sin la cual la ejecución solo puede devolver el centro.

Esto obliga a una distinción que la tesis necesita para no ser malinterpretada. El prompt pertenece a la producción, no a la dirección. Pedir un mejor resultado a la herramienta sin haber especificado intención, contexto, identidad, referencias y restricciones es pedir una variante del promedio, no una salida dirigida. Por eso la tesis no debe leerse como «escribe mejores prompts». Escribir mejores prompts opera dentro de la capa de ejecución y, en ausencia de dirección, solo recorre la vecindad del centro con mayor destreza. La tesis afirma algo anterior y más fuerte: que el lugar donde se gana o se pierde la diferenciación está antes de abrir la herramienta, y que ninguna habilidad de ejecución compensa una dirección ausente.

5. La economía del criterio

Tesis 2: la ejecución se abarata, el criterio sube de precio.

Esta tesis es descripción del presente, no predicción del futuro, y el documento la ancla así de forma deliberada. Se deriva de un principio económico establecido: cuando una capacidad productiva se vuelve abundante y su coste marginal tiende a cero, su valor económico tiende también a ese coste marginal, y el valor migra hacia el insumo complementario que permanece escaso.

Esta derivación no es incondicional, y el supuesto que la sostiene debe nombrarse. Que el valor de mercado de una capacidad tienda a su coste marginal presupone un mercado competitivo: muchos oferentes con acceso a la misma capacidad y sin barreras que permitan sostener un precio por encima del coste. Bajo poder de mercado —una capacidad patentada, un acceso restringido, un efecto de red que encierre a la demanda— el precio puede mantenerse muy por encima del coste marginal de forma duradera, y la tesis no se cumpliría. La tesis se sostiene precisamente porque el acceso a la ejecución generativa es hoy lo contrario de eso: ampliamente disponible, con múltiples proveedores sustituibles y barreras de entrada bajas. Es ese carácter competitivo del acceso a la ejecución, y no una ley económica incondicional, lo que hace que el valor se desplace hacia el complemento escaso. Si ese acceso se concentrara, el argumento debería reexaminarse.

Aplicado al caso: la ejecución —generar imagen, texto, voz, vídeo— ha pasado a ser abundante y de coste marginal próximo a cero. Por el principio anterior, su valor económico tiende a su coste, es decir, a poco. El valor no desaparece; se desplaza al complemento escaso. El complemento escaso de la ejecución abundante es el criterio que decide qué merece ejecutarse: qué intención, qué exclusiones, qué se mide como bueno. Ese criterio es, de nuevo, la dirección. La tesis 2 es, por tanto, la lectura económica de la tesis 1: lo que la tesis 1 establece como mecanismo —solo la dirección mueve la salida fuera del centro— la tesis 2 lo traduce en precio: solo la dirección conserva valor cuando la ejecución se vuelve gratuita.

Conviene no sobreinterpretar el alcance temporal de la afirmación. No se predice un estado futuro del mercado; se describe un desplazamiento que ya ha ocurrido y cuya dirección el principio económico explica. El documento se limita a nombrar lo que es

observable y a darle su fundamento.

6. Las nueve piezas como dimensiones de especificación

Tesis 3: el Sistema de Dirección IA consta de nueve piezas, ninguna opcional.

El marco de la derivación es el siguiente. Una salida queda determinada solo cuando toda dimensión de decisión que el sistema rellenaría, en su ausencia, con el valor por defecto del corpus ha sido especificada. Cada pieza del sistema es una de esas dimensiones. Las piezas no son pasos en secuencia ni fases de un procedimiento: son ejes de especificación. Derivarlas como un flujo de trabajo sería un error, porque sugeriría un orden temporal donde lo que hay es una lista de dimensiones que deben quedar cerradas, en el orden que sea, antes de que la salida deje de tender al centro.

Las nueve piezas, en el orden canónico, y la dimensión de ambigüedad que cada una cierra:

1. Intención cierra la dimensión de propósito. Sin ella, el sistema optimiza una competencia genérica sobre el tema, que es el promedio del tema. Reduce la ambigüedad de para qué existe la pieza y qué debe cambiar en quien la recibe.
2. Contexto cierra la dimensión de mundo. Sin él, el sistema trabaja desde la distribución global y no desde la realidad específica en la que la pieza opera. Reduce la ambigüedad del mundo concreto de la pieza, en sus capas de marca, audiencia, plataforma, pieza, histórico y competencia.
3. Identidad cierra la dimensión de lo propio. Sin ella, el sistema construye con elementos genéricos. Reduce la ambigüedad de qué hace esto reconociblemente de quien lo dirige y qué lo haría de cualquiera.
4. Referencias cierra la dimensión de ejemplar. Sin ellas, el sistema toma como modelo el ejemplar más común, que es el ejemplar medio. Reduce la ambigüedad de contra qué ejemplos concretos se miden el estilo y la calidad.
5. Restricciones cierra la dimensión de exclusión. Sin ellas no se excluye nada: queda disponible toda la distribución y gana el centro. Reduce la ambigüedad de qué regiones de lo posible están prohibidas. Es la pieza que hace el trabajo más directo de empujar la salida lejos del promedio, porque especificar es, en el límite, excluir.

Las cinco primeras constituyen la Dirección: el conjunto de decisiones anteriores a abrir la herramienta. La frontera entre estas cinco y las siguientes es la frontera entre dirección y ejecución.

6. Producción cierra la dimensión de ejecución bajo las cinco decisiones anteriores. Es la única pieza dependiente de la herramienta, y por eso es la pieza que hace portable al método: cuando la herramienta cambia, solo cambia Producción, y las otras ocho permanecen intactas.
7. Auditoría cierra la dimensión de verificación contra la dirección definida. Sin ella no hay forma de detectar el regreso a la moda: una salida puede haber recaído en el centro sin que nadie lo advierta. Reduce la ambigüedad de cómo se sabe que la

salida cumplió la dirección y no el promedio. El documento enuncia el principio de la verificación, no sus instrumentos.

8. Publicación cierra la dimensión de recepción. Sin ella, una pieza dirigida se suelta en un canal cuyo contrato de recepción puede romperla, y la diferenciación se colapsa en el punto de contacto con la audiencia. Reduce la ambigüedad de bajo qué condiciones la pieza se encuentra con quien la recibe.
9. Aprendizaje cierra la dimensión de iteración. Sin él, el sistema no actualiza su propia dirección: cada pieza reinicia desde cero y el método no compone en el tiempo. Reduce la ambigüedad de cómo la dirección se corrige con los resultados.

Por qué nueve y por qué ninguna sustituible

La justificación de que sean exactamente estas nueve descansa en dos propiedades que deben sostenerse conjuntamente.

La primera es la exhaustividad: las nueve cubren el trayecto completo de una pieza dirigida, desde su propósito hasta su composición en el tiempo. Propósito (1), mundo (2), lo propio (3), ejemplar (4) y exclusión (5) determinan qué se va a producir; ejecución (6) lo produce; verificación (7), recepción (8) e iteración (9) lo cierran, lo entregan y lo realimentan. No queda, según esta derivación, un tramo del trayecto sin cubrir.

La segunda es la no sustituibilidad: ninguna pieza cierra la ambigüedad de otra. La intención no fija el mundo; el contexto no fija lo propio; las referencias no excluyen; las restricciones no verifican; la auditoría no publica; la publicación no aprende. Omitir cualquiera de las nueve reintroduce el promedio exactamente en su eje, y ninguna de las restantes puede tapar ese hueco, porque opera sobre otra dimensión. De aquí se deriva «ninguna es opcional» desde principio, y no desde mandato: una pieza ausente no es una pieza que falta en una lista, es una dimensión de ambigüedad que el sistema vuelve a rellenar con el valor por defecto del corpus.

Conviene separar aquí dos argumentos que el texto ha venido usando entrelazados, porque no pesan igual. El primero es el argumento de cobertura de trayecto: las nueve piezas recorren el camino completo de una pieza dirigida, de su propósito a su realimentación, sin dejar tramo descubierto. El segundo es el marco informacional: la lectura de cada pieza como cierre de una «dimensión de subespecificación», donde especificar equivale a excluir regiones del espacio de salidas. El peso de la afirmación de completitud lo lleva el primero, la cobertura de trayecto, que es estructural y no depende de ninguna analogía. El marco informacional es apoyo y lenguaje unificador, no demostración: da una forma coherente de nombrar qué hace cada pieza, pero —como se declara en los límites— el mapeo entre entropía y ambigüedad es analogía rigurosa, no mecanismo medible. Si el marco informacional se retirara por completo, la completitud seguiría sostenida por la cobertura de trayecto; lo contrario no se cumple. Esta declaración de pesos importa para la falsación que sigue: el contraejemplo que la refuta debe ser un tramo del trayecto sin cubrir, no una objeción al lenguaje de la analogía.

Núcleo falsable

La afirmación de que las nueve son completas no es una prueba, es una conjetura falsable, y el documento la declara como tal sin atenuar el término.

La prueba propuesta es la siguiente, y se ofrece abiertamente al lector escéptico como invitación a refutarla. Nómbrase una dimensión de subespecificación —un eje de decisión que el sistema, si no se le especifica, rellenaría con el valor por defecto del corpus— que no quede cubierta por ninguna de las nueve piezas. Si tal dimensión puede nombrarse, el sistema es incompleto y la conjetura queda refutada. Si, tras el intento, no puede nombrarse, el número nueve queda justificado provisionalmente, hasta el próximo intento.

El documento no afirma haber demostrado que la lista es completa. Afirma haberla sometido a una prueba de falsación y no haberla visto caer todavía. Esta distinción es deliberada y se mantiene también en la sección de límites: confundir «no he encontrado un contraejemplo» con «he demostrado que no existe» sería un error que invalidaría el rigor del resto. La conjetura vive precisamente porque está expuesta a ser refutada por quien encuentre el eje que aquí no se ha visto. La prueba solo conserva su valor bajo una condición que se detalla en los límites: que las definiciones de cada pieza permanezcan fijas durante el intento de refutación, sin redescibir el contraejemplo dentro de una pieza existente para neutralizarlo.

Una segunda conjetura, de otro orden: exclusión frente a acumulación

La conjetura anterior es de completitud: afirma que las nueve dimensiones cubren el trayecto. Conviene aislar una segunda conjetura, de orden distinto y alcance más estrecho, que no se refiere al número de piezas sino al mecanismo de una sola de ellas: la pieza de Restricciones, donde se afirmó que especificar es, en el límite, excluir. Si esa afirmación es correcta, tiene una consecuencia contrastable.

Conjetura (H1). A igualdad de coste de especificación —un mismo presupuesto de instrucción—, dirigir por exclusión, nombrando qué regiones del espacio de salidas quedan prohibidas, aleja la salida de la tendencia central más que dirigir por acumulación, nombrando atributos deseados. La razón, derivada del principio, es que una exclusión retira una región completa de la distribución, mientras que un atributo deseado solo reordena preferencias dentro de una región que sigue disponible al muestreo.

La conjetura es falsable, y conviene declarar con precisión qué mide y qué no. No mide la entropía interna del sistema, que este documento trata como analogía y no como magnitud; mide una distancia en el espacio de salidas, observable sin acceso a los pesos: la proximidad de las salidas a un referente identitario bajo presupuesto fijo, comparando una condición de exclusión con una condición de acumulación. Un resultado en el que la acumulación iguale o supere a la exclusión, a igualdad de presupuesto, refutaría la forma fuerte de «especificar es excluir» y la rebajaría a una preferencia de estilo.

Resultado del piloto exploratorio (junio de 2026)

La conjetura H1 se sometió a una primera medición, de carácter exploratorio y declarada como tal. Un generador de caja negra produjo, para ocho encargos neutros, quince muestras en cada una de tres condiciones a presupuesto de instrucción igualado —sin dirección, dirección por acumulación de atributos y dirección por exclusión de anti-patrones—, y cada salida se proyectó a un espacio de embeddings para medir su distancia coseno a un vector identitario construido desde piezas canónicas de la marca.

En su forma fuerte, H1 no se sostiene. En la métrica de precisión —cercanía al vector identitario—, la exclusión no quedó más cerca que la acumulación; la diferencia agregada fue pequeña y de signo contrario a lo predicho (medias de distancia en torno a 0,093 sin dirección, 0,091 por acumulación y 0,094 por exclusión; contraste pareado no significativo y favorable, si acaso, a la acumulación). Tampoco ninguna de las dos direcciones acercó la salida al objetivo de forma claramente mayor que la ausencia de dirección. Por la regla declarada arriba, la afirmación de que «excluir acerca a la identidad más que acumular, a igual presupuesto» se retira.

Lo que sí se observó, de forma consistente, es de otra naturaleza: la exclusión produjo salidas más alejadas del centro genérico y más diversas entre sí que la acumulación. Es decir, excluir descentra y mantiene variedad, pero no apunta: aleja del promedio sin acercar a la identidad concreta. Acumular atributos, en cambio, dio salidas algo más próximas al objetivo. Exclusión y enumeración hacen, por tanto, trabajos distintos —descontaminar frente a dirigir— y no son intercambiables.

El piloto sugiere, además, una hipótesis más fina que conviene enunciar con cautela y no como resultado: la ventaja relativa de una polaridad sobre la otra parece condicionada al modo de fallo por defecto del contexto. En encargos técnicos, donde la salida no dirigida cae en la convención genérica, la enumeración positiva ayudó más, porque aporta la forma que falta. En un encargo emocional, donde la salida no dirigida cae no en lo genérico sino en el sentimentalismo, la exclusión ayudó más, porque la identidad de la marca en ese registro se define por la contención, y prohibir el exceso es precisamente lo que la alcanza. La exclusión domina, así, allí donde la voz se constituye por aquello que rechaza; la enumeración, allí donde se constituye por un gesto positivo. La mayoría de las identidades fuertes son ambas cosas, lo que explica que el método operativo combine las dos. Esta condicionalidad es conjetura abierta, pendiente de una prueba diseñada para contrastarla.

El alcance de esta evidencia es deliberadamente modesto: un solo generador, un solo modelo de embeddings, pocas tareas y pocas muestras por celda. No concluye; orienta. La forma fuerte de H1 queda retirada; la versión condicional queda abierta, y su contraste limpio depende de resolver la limitación de medida que se declara en los límites.

7. Genealogía intelectual

El método no es una invención sin antecedentes. Se sitúa dentro de tres linajes de pensamiento establecidos, que le dan respaldo y lo conectan con un trabajo previo más

amplio. Las atribuciones que siguen han sido verificadas contra fuentes primarias; sus referencias completas figuran en el aparato final.

Economía de la abundancia y migración del valor

Este linaje sostiene la tesis 2. La idea de que la abundancia de un recurso reorganiza el valor en torno a lo que permanece escaso tiene varios apoyos.

Herbert Simon formuló que, en un mundo rico en información, la abundancia de información produce una escasez de aquello que la información consume: la atención. William Stanley Jevons observó en *The Coal Question* que el abaratamiento de un recurso aumenta su consumo total en lugar de reducirlo —la paradoja de Jevons—; trasladada al caso, implica que abaratar la ejecución dispara la demanda de ejecución y, con ella, el valor de lo que la dirige. Carl Shapiro y Hal Varian sistematizaron la economía de la información y la lógica de los complementos en *Information Rules*. La formulación popular «comoditiza tu complemento» se atribuye a Joel Spolsky, si bien la microeconomía de complementos que la sustenta es muy anterior. Tyler Cowen, en *Average Is Over*, describió un mercado que se polariza entre quienes complementan a la máquina y quienes no; su obra se invoca aquí por ese marco de polarización del valor, y no como fuente de la formulación «el criterio sube de precio», que es propia de este documento.

Augmentación humana

Este linaje sostiene que la herramienta aumenta la capacidad de ejecución pero no provee ni la intención ni el criterio: el humano aporta la dirección. Es el puente exacto entre el método y la lectura de la IA en clave de augmentación.

El soporte principal del linaje es Douglas Engelbart, cuyo programa de aumentar el intelecto humano mediante herramientas describe con precisión la relación que este documento defiende: la máquina amplía la capacidad de abordar problemas complejos, mientras la formulación del propósito y el criterio permanecen en el humano. Es el antecedente que da fundamento riguroso a la afirmación de que dirigir es la parte que la herramienta no puede asumir, y el que conecta con mayor exactitud con la lectura de la IA como augmentación. J. C. R. Licklider, en su visión de simbiosis hombre-computadora, ofrece un apoyo más matizado: aporta el marco de una relación complementaria entre humano y máquina, pero su horizonte de largo plazo contemplaba que el razonamiento mismo acabara cedido a la máquina, lo que diverge de la tesis de que la dirección sea irreductiblemente humana. Se invoca, por tanto, por el marco de simbiosis y no como soporte de la tesis fuerte. Vannevar Bush, en un texto anterior y de uso más tangencial, anticipó la idea de la máquina como extensión de la memoria y el pensamiento.

A esta genealogía conviene añadir una formulación más reciente y más directamente aplicable al caso, procedente no de los fundadores del campo sino de su crítica contemporánea. Kentaro Toyama, en *Geek Heresy* (2015), enuncia la Ley de Amplificación: la tecnología amplifica las fuerzas humanas subyacentes —intención y capacidad— en lugar de sustituirlas; mejora los contextos donde esas fuerzas ya son

competentes y deja intactos, o agrava, aquellos donde son indiferentes o disfuncionales. Es el antecedente más exacto del núcleo de este documento: la herramienta no provee la dirección, la amplifica, y por eso una ejecución abundante sin dirección no nivela, sino que multiplica una diferencia que ya existía. Toyama añade un corolario que este documento no toma en su tesis pero sí recoge entre sus límites: si la tecnología amplifica las fuerzas previas, amplifica con ellas su desigualdad. Ese corolario se trata en la sección 10 y acota cualquier lectura del criterio como mérito puramente voluntario.

El principio informacional de especificación

Este linaje da respaldo a la afirmación, central en la sección 6, de que cada pieza reduce ambigüedad.

Claude Shannon estableció que la información es reducción de incertidumbre: un mensaje que no excluye posibilidades no transporta información. El documento usa este resultado exclusivamente como principio sobre la especificación, y no como descripción de cómo computa ningún sistema. Especificar es reducir el espacio de salidas posibles; cada pieza retira grados de libertad y, al hacerlo, aleja la salida del resultado de máxima generalidad. Esta es la línea que el documento no cruza: el principio informacional se invoca como epistemología de la especificación —qué significa especificar— y jamás como mecánica de un modelo.

El mapeo entre entropía y ambigüedad se presenta, por tanto, como analogía rigurosa y declarada como tal, no como un cálculo literal. La afirmación «cada pieza reduce la entropía de la salida» es una metáfora precisa de «cada pieza retira regiones del espacio de salidas posibles»; no se reclama que exista una magnitud entrópica medible en el sentido técnico. Del mismo modo, la tendencia de una salida no especificada a recaer en el centro se ilustra a veces con la regresión a la media asociada a Galton; esto se ofrece como analogía y no como mecanismo: nombra el fenómeno de retorno al centro de modo evocador, sin afirmar que la convergencia de un sistema generativo y la regresión estadística clásica compartan causa.

8. Portabilidad

El fundamento es lo que hace al método independiente de la herramienta, y esta sección lo demuestra a partir de lo ya establecido.

El método se deriva del principio de convergencia, no del funcionamiento de ninguna herramienta concreta. Por consiguiente, se aplica a cualquier dominio en el que un sistema generativo o agéntico subespecificado tendería al promedio. La estructura de las nueve piezas lo refleja: de las nueve, ocho son independientes de la herramienta. Intención, contexto, identidad, referencias, restricciones, auditoría, publicación y aprendizaje cierran dimensiones de ambigüedad que existen con independencia de cómo se ejecute la pieza. Solo Producción depende de la herramienta. Cuando la herramienta cambia, cambia Producción y las otras ocho permanecen intactas.

De aquí se sigue la aplicabilidad del método a la dirección de agentes. Dirigir agentes es el dominio donde la capa de producción es más volátil —las herramientas se suceden con rapidez— y donde la capa de dirección es, por la misma razón, más necesaria. Un sistema agéntico subespecificado tiende al promedio igual que cualquier otro sistema generativo, y las cinco piezas de dirección y las tres de cierre se trasladan sin alteración: solo se reescribe la pieza de ejecución. El documento es, en este sentido, la prueba escrita de que el método no es una lista de control atada a la producción de vídeo, sino una estructura que sobrevive al cambio de capa de ejecución.

9. Qué absorbe la Biblia y qué permanece como activo citable

Esta sección cierra el plan de absorción, mirando hacia dentro del sistema documental.

La regla de decisión es estricta: una conclusión de este documento se absorbe en la Biblia solo si altera una decisión de producción. En caso contrario permanece aquí. Como este documento es justificación y no decisión, la mayor parte no se absorbe, y eso es deliberado y no un defecto.

Lo que se absorbe se limita a las conclusiones que cambian una decisión de producción: un enunciado canónico breve del principio de convergencia y la derivación de las nueve piezas como justificación que la Biblia pueda citar. Su lugar natural es una extensión del bloque de canon absorbido. Lo que permanece fuera, como activo de credibilidad externa, son las derivaciones completas, la genealogía intelectual, el argumento económico y el criterio de falsación: viven en este documento, al servicio de la subvención, del contexto de avales externos y de futuros clientes, y no en la Biblia.

La disciplina de proceso es parte de la regla. La redacción de este documento ocurre en un espacio de trabajo limpio; cualquier absorción en la Biblia ocurre en un espacio de sistema distinto, también limpio, y nunca como reacción al lanzamiento de una herramienta ni dentro de un espacio de producción.

10. Límites

Esta sección declara qué no demuestra el documento, dónde el argumento es analogía y qué queda abierto a refutación. Su inclusión es deliberada: un texto que reclama rigor debe enunciar sus propios límites antes de que lo haga su crítico.

Lo que no se demuestra. El documento no demuestra que las nueve piezas sean completas. Ofrece una conjetura falsable y declara no haber encontrado todavía un contraejemplo; esto no equivale a una prueba de inexistencia de contraejemplos. La completitud permanece, por construcción, abierta a refutación por quien nombre una dimensión de subespecificación no cubierta.

La condición que mantiene viva la falsación. La prueba de falsación de la sección 6 corre un riesgo que debe declararse abiertamente. Las nueve dimensiones están enunciadas con amplitud suficiente para que casi cualquier contraejemplo

propuesto pueda, por redesccripción, reabsorberse dentro de una pieza existente: un eje nuevo siempre puede presentarse como «una capa más de contexto» o «un caso particular de restricción». Si esa redesccripción se permite, la conjetura deja de ser falsable en la práctica, porque ningún contraejemplo podría sobrevivir al reetiquetado y la prueba se vuelve retórica. Por eso la falsación solo es válida bajo una condición estricta: las definiciones de cada pieza, tal como quedan fijadas en la sección 6, permanecen fijas durante el intento de refutación, y el contraejemplo no se redescrive dentro de una pieza existente para neutralizarlo. Un eje de subespecificación que solo encaja en una pieza ensanchando la definición de esa pieza cuenta como contraejemplo, no como confirmación. La conjetura es genuinamente falsable únicamente mientras se respete esta condición.

Dónde el argumento es analogía y no mecanismo. Dos piezas del razonamiento son analogías declaradas, no descripciones mecánicas. El mapeo entre entropía y ambigüedad se usa como principio sobre la especificación y como metáfora rigurosa, no como un cálculo de una magnitud medible; la afirmación de que cada pieza «reduce la entropía» de la salida es figurada. La regresión a la media se invoca para nombrar el retorno al centro, sin afirmar identidad causal entre la convergencia de un sistema generativo y la regresión estadística clásica. Tratar cualquiera de las dos como mecanismo, y no como analogía, excedería lo que el documento sostiene.

Lo que el principio de convergencia no prueba. El principio sostiene que un sistema subespecificado tiende al centro de su distribución; no sostiene que toda salida no especificada sea idéntica, ni que la especificación garantice calidad. Especificar desplaza la salida fuera del centro; no asegura que la región a la que la desplaza sea mejor. La calidad de la dirección queda fuera del alcance de este documento, que solo establece que sin dirección la salida tiende al promedio, no que con dirección la salida sea necesariamente buena.

El criterio no es mérito puramente voluntario. La tesis 2 sostiene que, al abaratare la ejecución, el valor migra al criterio que la dirige. De ahí no se sigue, y este documento no lo afirma, que quien no dirige carezca de criterio por elección. La capacidad de imponer una desviación intencional a un sistema es, en buena medida, capital cultural acumulado —formación, entorno, tiempo disponible para la reflexión—, en el sentido de Bourdieu, y no un acto de mera voluntad. La consecuencia es doble y debe declararse: la amplificación amplifica también la desigualdad previa de ese capital, según el corolario de Toyama; y, por tanto, el criterio se trata aquí como competencia cultivable y desigualmente distribuida, nunca como rasgo de carácter. Cualquier lectura de la tesis 2 que culpe al promediado de no aspirar excede lo que el documento sostiene.

Qué es el centro y qué no es. El documento llama centro a la moda de la distribución condicionada a un sector y la identifica con la convención de ese sector. Conviene no confundir esa moda con la tendencia central del mundo. El corpus sobre el que un sistema muestrea no es una muestra neutral de la realidad, sino un agregado ya conformado por qué se escribió, qué se recogió y qué se conservó. Su moda es, por tanto, el centro de una distribución ya sesgada, no un

espejo del mundo. Esto refuerza el argumento —el centro que la dirección abandona es un artefacto del corpus, no una verdad—, pero impide leer la salida no dirigida como «lo que la realidad diría». El documento no apela, para sostener esto, a ningún procedimiento de entrenamiento: la afirmación se mantiene en el plano del corpus, coherente con la prohibición de explicar el mecanismo por la maquinaria interna.

El confusor entre tema y voz en la medición de H1. La prueba de H1 mide la distancia de una salida a un vector identitario construido a partir de piezas canónicas de la marca. Como esas piezas comparten no solo una voz sino también un universo temático, el vector mezcla ambos: la distancia resultante premia en parte la coincidencia de tema, no solo la de voz. En el piloto esto se hizo visible —en un encargo cuyo tema coincidía con el de las referencias, incluso la salida sin dirección puntuó cerca del objetivo—. La métrica es, por tanto, un detector parcial de tema, y una prueba limpia de la conjetura exige separar tema de voz en la construcción del vector. Mientras esa separación no se haga, los valores absolutos de distancia deben leerse con esta reserva.

El límite agéntico. La sección 8 extiende el método a la dirección de agentes, y esa extensión se sostiene mientras el agente, por autónomo que sea en la ejecución, siga recibiendo su dirección de un humano. El límite aparece en el horizonte que la propia genealogía ya anticipó al matizar a Licklider: un sistema con razonamiento prolongado, capaz de generar sus propias restricciones y de simular desviación, dejaría de depender de una dirección externa. Si ese supuesto se cumpliera, la premisa de que la dirección es irreductiblemente humana —no la portabilidad del método, sino su sujeto— exigiría reexamen. El alcance de este documento se acota, en consecuencia, al régimen en que la dirección la aporta un humano; nada de lo afirmado se extiende por defecto a sistemas que dirijan su propia generación.

Estado de las atribuciones. Las fuentes de la genealogía han sido verificadas contra fuentes primarias: autoría, título, año y publicación están confirmados. El cuerpo del argumento —el principio de convergencia, la derivación de las nueve piezas, las tres tesis— no depende lógicamente de ninguna atribución concreta: las fuentes respaldan y sitúan, no fundan. Lo que permanece abierto en este documento no es bibliográfico, sino intelectual: la conjetura de completitud, bajo la condición de falsación declarada, y las analogías declaradas como tales (entropía-ambigüedad, regresión a la media).

Alcance del agnosticismo. El documento afirma que el principio es independiente de la herramienta y se cumplirá en sistemas futuros. Esto es una conjetura sobre la naturaleza de «muestrear un corpus compartido bajo especificación insuficiente», no una garantía sobre tecnologías que aún no existen. Se sostiene mientras los sistemas generativos sigan operando por muestreo de distribuciones de corpus amplios; un cambio radical en ese supuesto exigiría reexaminar el principio.

Cierre

El método de dirección no es arbitrario ni es una preferencia de quien lo decide. Es la consecuencia necesaria de cómo se comporta una capacidad generativa compartida: como sin especificación la salida tiende al centro, y como el centro es el promedio del sector, dirigir antes de ejecutar es la única forma de no entregar el promedio. Las nueve piezas son las dimensiones de esa especificación; su completitud es una conjetura ofrecida a la refutación, no un dogma; su portabilidad se sigue de que ocho de las nueve no dependen de la herramienta. El documento no manda nada. Solo muestra por qué lo que la Biblia decidió no podía decidirse de otro modo.

Referencias

Atribuciones verificadas contra fuentes primarias en junio de 2026: autoría, título, año y publicación confirmados en cada caso.

Economía de la abundancia y migración del valor

- Simon, Herbert A. (1971). «Designing Organizations for an Information-Rich World». En Martin Greenberger (ed.), *Computers, Communications, and the Public Interest*. Baltimore: The Johns Hopkins Press, pp. 37-52. La formulación «a wealth of information creates a poverty of attention» en pp. 40-41.
- Jevons, William Stanley (1865). *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*. Londres: Macmillan. Paradoja de Jevons: las mejoras de eficiencia aumentan el consumo total del recurso.
- Shapiro, Carl; Varian, Hal R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press. Economía de la información y lógica de complementos.
- Spolsky, Joel (2002). «Strategy Letter V». Joel on Software, 12 de junio de 2002. Popularización de «comoditiza tu complemento»; la microeconomía de complementos que la sustenta es anterior.
- Cowen, Tyler (2013). *Average Is Over: Powering America Beyond the Age of the Great Stagnation*. Nueva York: Dutton. Invocado por su marco sobre la polarización entre quienes complementan a la máquina y el resto; no se le atribuye la formulación «el criterio sube de precio».

Augmentación humana

- Engelbart, Douglas C. (1962). *Augmenting Human Intellect: A Conceptual Framework*. Informe del Stanford Research Institute para el Director of Information Sciences, Air Force Office of Scientific Research. Menlo Park, CA. Soporte principal del linaje.
- Licklider, J. C. R. (1960). «Man-Computer Symbiosis». *IRE Transactions on Human Factors in Electronics*, HFE-1, pp. 4-11. Soporte de apoyo, con reserva: su horizonte de largo plazo contemplaba la cesión del razonamiento a la máquina.
- Bush, Vannevar (1945). «As We May Think». *The Atlantic Monthly*, julio de 1945. Antecedente tangencial: la máquina como extensión de la memoria y el pensamiento.
- Toyama, Kentaro (2015). *Geek Heresy: Rescuing Social Change from the Cult of Technology*. Nueva York: PublicAffairs. Ley de Amplificación: la tecnología amplifica las fuerzas humanas subyacentes en lugar de sustituirlas, y amplifica con ellas su desigualdad. Antecedente más directo del núcleo del documento; el corolario de desigualdad se recoge en Límites.

El principio informacional de especificación

- Shannon, Claude E. (1948). «A Mathematical Theory of Communication». *Bell System Technical Journal*, 27, pp. 379-423 y 623-656. La información como reducción de incertidumbre; usado solo como principio de especificación, no como descripción de cómputo.

Galton, Francis (1886). «Regression towards Mediocrity in Hereditary Stature». *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, pp. 246-263. Usado como analogía, no como mecanismo.

Sociología del criterio (citado en Límites)

Bourdieu, Pierre (1979). *La Distinction. Critique sociale du jugement*. París: Éditions de Minuit. El criterio como capital cultural acumulado, no como voluntad; sostiene el límite meritocrático de la sección 10.

Trabajo previo: homogeneización, monocultura y colapso de modelos

Shumailov, Ilia; Shumaylov, Zakhar; Zhao, Yiren; Papernot, Nicolas; Anderson, Ross; Gal, Yarin (2024). «AI models collapse when trained on recursively generated data». *Nature*, 631, pp. 755-759.

Doshi, Anil R.; Hauser, Oliver P. (2024). «Generative AI enhances individual creativity but reduces the collective diversity of novel content». *Science Advances*, 10(28), eadn5290.

Padmakumar, Vishakh; He, He (2024). «Does Writing with Language Models Reduce Content Diversity?». *International Conference on Learning Representations (ICLR 2024)*.

Kleinberg, Jon; Raghavan, Manish (2021). «Algorithmic Monoculture and Social Welfare». *Proceedings of the National Academy of Sciences (PNAS)*, 118(22), e2018340118.

Bommasani, Rishi; Creel, Kathleen A.; Kumar, Ananya; Jurafsky, Dan; Liang, Percy (2022). «Picking on the Same Person: Does Algorithmic Monoculture Lead to Outcome Homogenization?». *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pp. 3663-3678.

Depósitos previos de la serie

Collada, D. (2026). Andamiaje de contexto en inferencia sobre modelos de lenguaje de caja negra: mitigación de la regresión a la media y la complacencia (sycophancy) sin modificación de pesos. Working paper. Zenodo. <https://doi.org/10.5281/zenodo.20705077>

Collada, D. (2026). El valor informacional de la disensión: validación condicionada, regímenes de reflejo-resistencia-dirección y una definición de aumento cognitivo para modelos de lenguaje de caja negra. Working paper. Zenodo. <https://doi.org/10.5281/zenodo.20714304>

Cómo citar. Collada, D. (2026). Documento de Fundamentos del Sistema de Dirección IA (v1.2). Working paper. Zenodo. <https://doi.org/10.5281/zenodo.20765017>

Historial de versiones. Versión 1.2 (junio de 2026) — se incorpora a la sección 6 el resultado de un piloto exploratorio de la conjetura H1: en su forma fuerte (la exclusión acerca a la identidad más que la acumulación a igual presupuesto) no se sostiene y se retira; se observa que la exclusión descentra y diversifica pero no dirige, y se abre una hipótesis condicional —la polaridad ventajosa depende del modo de fallo del contexto— enunciada como conjetura, no como resultado. Se añade a la sección 10 un límite de medición: el confusor entre tema y voz en el vector identitario. Las tres tesis, la conjetura de completitud y la genealogía no cambian. Versión 1.1 (junio de 2026) — adiciones de encuadre y de exposición a refutación, sin cambio de tesis ni de las nueve piezas. Se incorpora a la genealogía la Ley de Amplificación (Toyama, 2015) como antecedente más directo del núcleo. Se aísla en la sección 6 una segunda conjetura falsable (H1) sobre la pieza de Restricciones, distinta de la conjetura de completitud. Se añaden tres límites en la sección 10: el criterio como capital cultural (Bourdieu), la naturaleza no neutral del centro a nivel de corpus, y el límite agéntico. Las tres tesis y la conjetura de completitud no cambian respecto a la versión 1.0.

Working paper · Dani Collada IA · Serie Sistema de Dirección IA, depósito 1. No revisado por pares. Fuentes verificadas contra fuentes primarias. Licencia: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Internacional (CC BY-NC-ND 4.0).