Cuantificando la equidad de proveedores más allá de la exposición uniforme en sistemas de recomendación

F. J. Martínez[1 [0000−0001−7969−1737]], Raciel Yera[1 [0000−0001−9759−261X]], and L. Martínez[1 [0000−0003−4245−8813]]

Universidad de Jaén, Jaén, Spain
fjmimbre@ujaen.es, ryera@ujaen.es, martin@ujaen.es

Resumen Los sistemas de recomendación deciden, en buena medida, qué productos y servicios reciben visibilidad en las plataformas digitales, y dos algoritmos distintos pueden generar diferencias muy notables en la exposición de los proveedores. La literatura sobre *fairness* (entendido como equidad en el reparto de exposición y resultados) en sistemas de recomendación ha crecido rápidamente, pero la mayoría de propuestas o bien trata por igual todas las posiciones del ranking, o bien se concentra en el lado del consumidor. Para abordar esta carencia, proponemos una formulación común de visibilidad desigual (*disparate visibility*) para cuantificar el *provider fairness*, parametrizada por una función de atención posicional $w(k)$. Este marco recupera como casos particulares métricas existentes y da lugar a cuatro variantes (uniforme, ponderada por posición, basada en zonas y binaria a nivel de usuario), complementadas con una medida agregada $D(w)$ que resume el comportamiento global del sistema. Las métricas se han diseñado para integrarse en entornos reales de *provider fairness* a gran escala, capturando además su variabilidad temporal a lo largo de sucesivos ejercicios. Acompañamos la propuesta con un estudio experimental sobre el dominio musical, con cortes anuales independientes y cinco algoritmos. Este análisis ilustra el comportamiento diferenciado de cada variante y revela una tensión estructural entre volumen y cobertura de exposición que ningún algoritmo resuelve completamente.

Keywords: Sistemas de recomendación · *Provider fairness* · Visibilidad desigual · *fairness a largo plazo* · Exposición en rankings

1. Introducción

En el contexto de la inteligencia artificial (IA), suele entenderse por sistema justo aquel que reparte de forma equitativa los beneficios y los costes entre las partes implicadas [7]. Esta noción (el *fairness*, entendido como equidad en el reparto entre los distintos actores afectados) es uno de los pilares de una IA fiable y se alinea con los marcos regulatorios que están entrando en vigor, como la Ley de IA de la Unión Europea [1]. Los sistemas de recomendación, pensados

para ayudar a los usuarios a moverse en entornos saturados de información [17], tienen una fuerte tendencia a heredar y amplificar los sesgos que arrastran los datos históricos [4] en escenarios de alto riesgo tales como una búsqueda de empleo o una evaluación crediticia [15].

En estos sistemas conviven actores con intereses muy distintos y a veces enfrentados [3]. El *consumer fairness* (equidad hacia el usuario) busca que las recomendaciones no perjudiquen sistemáticamente a unos grupos de usuarios frente a otros [16]. El *provider fairness* (equidad hacia el proveedor), de especial relevancia en mercados electrónicos, servicios de *streaming* (difusión bajo demanda) o portales de empleo, persigue evitar dinámicas monopolísticas asegurando oportunidades comparables de exposición [8,13,12,5]. Buscar únicamente la precisión predictiva limita la visibilidad de los proveedores con menor presencia en los datos [19], y asimismo el sesgo posicional (*position bias*) agrava el problema, puesto que los usuarios interactúan sobre todo con los primeros elementos del ranking [10], lo que retroalimenta la exposición futura de los proveedores ya favorecidos [15].

En contraposición con esto, los grandes despliegues comerciales casi nunca contemplan criterios de *fairness* a largo plazo [14]. Este trabajo aborda esta limitación, proponiendo una familia de métricas de visibilidad desigual, que sea capaz de capturar la dinámica no estacionaria del *fairness* a largo plazo.

La principal contribución de este trabajo es una formulación común de visibilidad desigual para proveedores en rankings de recomendación, parametrizada por una función de atención posicional $w(k)$. Este marco unificado permite recuperar como casos particulares métricas previamente propuestas en la literatura (tales como las basadas en paridad demográfica de la exposición) y, al mismo tiempo, derivar nuevas variantes que capturan distintos niveles de granularidad en la exposición: uniforme, ponderada por posición, basada en zonas y, como caso límite con agregación no lineal, binaria a nivel de usuario. Sobre esta base, el trabajo aporta además la formalización de su integración en un escenario de *fairness* a largo plazo, siguiendo su evolución temporal, junto con un estudio experimental sobre distintos conjuntos de datos y algoritmos de recomendación que ilustra el comportamiento diferenciado de cada variante.

2. *Fairness* en sistemas de recomendación

El estudio del *provider fairness*, situado dentro del panorama más amplio de *fairness* en sistemas de recomendación [5], ha seguido varias direcciones. Una línea clásica adapta nociones de paridad demográfica para comparar la visibilidad entre grupos de proveedores [19]. Otras propuestas atacan el problema desde la optimización: P-MMF [18] busca maximizar la exposición del proveedor peor tratado, mientras que Gómez et al. [8] aplican re-ranking según la distribución de preferencias. Más recientemente, Wu et al. [11] y Smith et al. [6] han propuesto clasificaciones y métricas co-diseñadas con proveedores reales.

Nuestra propuesta se apoya en el *fairness* de exposición en rankings. Singh y Joachims [15] fueron los primeros en modelar la exposición como un recurso

escaso que se reparte según la posición, y Biega et al. [2] extendieron la idea proponiendo amortizar la equidad a lo largo del tiempo. Ambos trabajos, sin embargo, razonan a nivel de elemento individual y dejan abiertas dos cuestiones que aquí nos interesan: cómo agregar el *fairness* por proveedor y cómo trasladarlo a un escenario de *fairness* a largo plazo, siguiendo su evolución temporal. En adición, es válido señalar que las métricas habituales o bien tratan todas las posiciones por igual, o bien las reducen a una decisión binaria. Por otro lado, la familia de métricas se sitúa entre ambos extremos al permitir un descuento posicional graduado y parametrizable.

3. Métricas de visibilidad desigual para el *provider fairness*

Las métricas que se proponen extienden el marco de trabajo de Singh y Joachims [15], que entiende la exposición como un recurso escaso repartido según la posición, llevándolo al nivel de proveedor y dando cuenta del sesgo posicional (*position bias*) [10] de manera graduada.

3.1. Notación y definiciones preliminares

Sea U el conjunto de usuarios, I el de elementos y $C = \{c_1, \ldots, c_M\}$ el de proveedores, con $I_c \subseteq I$ el catálogo de cada proveedor c. Denotamos por $R = \{r_{ui}\}$ el conjunto de interacciones, por $\hat{R}_u = (i_{u,1}, \ldots, i_{u,K})$ la lista de K recomendaciones generada para el usuario u y por rank_{ui} la posición que ocupa el elemento i dentro de $\hat{R}_u$.

La función de atención $w : \{1, \ldots, K\} \to \mathbb{R}_{\geq 0}$ se asume no creciente:

$$w(1) \geq w(2) \geq \cdots \geq w(K) \geq 0, \quad \sum_{k=1}^{K} w(k) > 0. \tag{1}$$

La exposición que recibe un elemento i dentro de la lista del usuario u se define como:

$$\text{Exp}_u(i) = \begin{cases} w(\text{rank}_{ui}) & \text{si } i \in \hat{R}_u, \\ 0 & \text{en caso contrario.} \end{cases} \tag{2}$$

La exposición agregada del proveedor c para el usuario u se obtiene sumando la exposición de todos sus elementos presentes en la recomendación:

$$\text{Exp}_u(c) = \sum_{i \in I_c} \text{Exp}_u(i) = \sum_{i \in \hat{R}_u \cap I_c} w(\text{rank}_{ui}). \tag{3}$$

La representación del proveedor c en los datos de entrada se define como:

$$R_c = \frac{|\{r_{ui} : u \in U, i \in I_c\}|}{|R|}. \tag{4}$$

El valor R_c funciona como referencia y recoge la proporción de interacciones atribuidas al proveedor c en los datos observados. Este valor se tomaría por tanto como referencia para la formulación de la paridad demográfica aplicada a la exposición [19]. Así, la visibilidad de un grupo debería guardar proporción con su representación en los datos. No obstante, es importante tener en cuenta que al calcularse tomando como base los datos históricos, el valor R_c puede arrastrar sesgos previos. Con vistas a realizar una modelación más avanzada, esto daría lugar a alternativas tales como una distribución uniforme de los diferentes atributos ($R_c = 1/|C|$), o una basada en criterios de mérito.

Con estas definiciones, la visibilidad desigual del proveedor c adopta la forma general:

$$\Delta V_c(w) = \frac{1}{|U|} \sum_{u \in U} \frac{\mathrm{Exp}_u(c)}{\sum_{j \in \hat{R}_u} w(\mathrm{rank}_{uj})} - R_c. \tag{5}$$

3.2. Visibilidad desigual uniforme

Como punto de partida consideramos el caso más simple, en el que todas las posiciones reciben la misma atención: $w(k) = 1$ para todo k. Sustituyendo en la Ec. (5) y siguiendo a Gómez et al. [8], queda:

$$\Delta V_c^{\mathrm{uni}} = \Delta V_c(w \equiv 1) = \frac{1}{|U|} \sum_{u \in U} \frac{\left|\{i \in \hat{R}_u : i \in I_c\}\right|}{|\hat{R}_u|} - R_c. \tag{6}$$

Esta métrica compara la fracción media de recomendaciones que recae sobre el proveedor c con la representación que cabría esperar en los datos. Es, en esencia, equivalente a las métricas de *exposure fairness* basadas en paridad demográfica que aparecen en la literatura [19,11], pero formulada a nivel de proveedor. Un valor $\Delta V_c^{\mathrm{uni}} = 0$ indica representación perfectamente proporcional; los valores negativos delatan infrarrepresentación del proveedor, y los positivos, sobrerrepresentación.

3.3. Visibilidad desigual ponderada por posición

Suponer que todas las posiciones pesan igual ignora cómo se comporta realmente el usuario. Los estudios con seguimiento ocular y los análisis de tasas de clic han concluido que la atención cae rápidamente conforme bajamos en el ranking [10]. De ahí, precisamente, viene el descuento logarítmico que usa el clásico DCG (*Discounted Cumulative Gain*) en recuperación de información [9]. Singh y Joachims [15] formalizan esta intuición con vectores de pesos de exposición no crecientes para definir la exposición posicional dentro de un ranking.

Aplicando directamente la formulación general de la Ec. (5) con una función de atención w arbitraria (no creciente, según la Ec. (1)), se obtiene la visibilidad desigual ponderada por posición:

$$\Delta V_c^{\mathrm{pos}} = \Delta V_c(w) = \frac{1}{|U|} \sum_{u \in U} \frac{\sum_{i \in \hat{R}_u \cap I_c} w(\mathrm{rank}_{ui})}{\sum_{j \in \hat{R}_u} w(\mathrm{rank}_{uj})} - R_c. \tag{7}$$

Cuando $w(k) = 1$ para todo k, se recupera ΔV_c^{uni} como caso particular. En la práctica, w puede instanciarse, por ejemplo, con el descuento logarítmico del DCG, $w(k) = 1/\log_2(k+1)$ [9], o con los pesos del modelo basado en posición que se emplea con frecuencia para analizar clics [10]. Esa flexibilidad permite ajustar la métrica al modelo de atención que mejor encaje en cada dominio.

3.4. Visibilidad desigual basada en zonas

Partiendo de lo anterior, una versión más sencilla y más interpretable de la atención del usuario vendría dada por usar una función de pesos constante a trozos definida sobre zonas del ranking. La idea se inspira en cómo presentan la información muchas interfaces: los elementos se agrupan en bloques visibles sin hacer *scroll*, accesibles con un *scroll* moderado y, finalmente, fuera de la vista [2]:

$$w_{\text{zone}}(k) = \begin{cases} \alpha_1, & 1 \le k \le K_1, \\ \alpha_2, & K_1 < k \le K_2, \\ 0, & k > K_2, \end{cases} \quad \text{con } \alpha_1 > \alpha_2 > 0. \tag{8}$$

La métrica de visibilidad desigual correspondiente es:

$$\Delta V_c^{\text{zone}} = \frac{1}{|U|} \sum_{u \in U} \frac{\sum_{i \in \hat{R}_u \cap I_c} w_{\text{zone}}(\text{rank}_{ui})}{\sum_{j \in \hat{R}_u} w_{\text{zone}}(\text{rank}_{uj})} - R_c. \tag{9}$$

En este caso, este enfoque sería un caso particular de la formulación ponderada por posición que gana en interpretabilidad, considerando que los parámetros K_1, K_2, α_1 y α_2 pueden fijarse a partir de las características concretas de la interfaz.

3.5. Visibilidad desigual binaria a nivel de usuario

Finalmente, en el contexto del presente trabajo se propone una noción binaria de visibilidad que pregunta simplemente si el proveedor c aparece, de forma visible, ante cada usuario. La idea entronca con la diversidad agregada [19], que valora la presencia o ausencia de un elemento en las recomendaciones con independencia de cuántas veces aparezca, y también con la cobertura de catálogo discutida en el contexto del *provider fairness* por Xu et al. [18].

Dado un umbral de visibilidad K_v, la visibilidad desigual binaria se define como:

$$\Delta V_c^{\text{bin}} = \frac{1}{|U|} \sum_{u \in U} \mathbb{I}\left(\exists\, i \in \hat{R}_u \cap I_c \text{ s.t. } \text{rank}_{ui} \le K_v\right) - R_c. \tag{10}$$

Esta métrica captura la cobertura de la exposición a nivel de usuario, no el volumen total. Este enfoque minimalista estaría enfocado en identificar en qué fracción de usuarios aparece el proveedor c dentro de las K_v primeras posiciones, en contraposición con los enfoques anteriores que dan lugar a un mayor nivel de granularidad. Es útil cuando lo crítico es que el proveedor tenga, al menos,

una oportunidad de ser visto, antes que la cantidad de elementos suyos que se muestren.

A diferencia de las tres anteriores, ΔV_c^{bin} no es una instancia directa de la Ec. (5), ya que el operador de agregación deja de ser lineal y pasa a ser una indicadora de presencia. Más bien, constituye una extensión de cobertura derivada del mismo principio de atención posicional (con un umbral K_v que delimita la región visible), pero sustituyendo la suma ponderada por una indicadora de presencia. De este modo, se complementan las métricas volumétricas con una perspectiva de cobertura, especialmente útil en escenarios de *fairness* a largo plazo en los que se requiere garantizar presencias mínimas de determinados proveedores.

3.6. Unificación y aplicación a *fairness* a largo plazo

Las tres primeras métricas son instancias directas de la formulación general $\Delta V_c(w)$ (Ec. (5)), diferenciándose únicamente en la elección de la función de atención w: $w(k) = 1$ (uniforme), $w(k)$ no creciente arbitraria (ponderada) y constante a trozos (zonas). La binaria se sitúa como extensión de cobertura del mismo marco, según se ha discutido en la Sec. 3.5. La estructura parametrizada deja la puerta abierta a incorporar otros modelos de atención sin tocar el armazón formal. La Tabla 1 resume sus características.

Tabla 1. Resumen comparativo de las métricas de visibilidad desigual propuestas.

Métrica	Modelo de atención	Parámetros	Caso de uso
ΔV_c^{uni}	Uniforme	Ninguno	Evaluación general
ΔV_c^{pos}	Decaimiento continuo	$w(\cdot)$	Análisis detallado
ΔV_c^{zone}	Constante a trozos	$K_1, K_2, \alpha_1, \alpha_2$	Definición de *fairness* a largo plazo
ΔV_c^{bin}	Binaria (presencia)	K_v	Cobertura mínima

El rango teórico de todas las métricas es $[-R_c, 1 - R_c]$. Los valores negativos delatan infraexposición del proveedor c; los positivos, sobreexposición; el cero, representación proporcional. La magnitud permite, además, comparar proveedores directamente: $\Delta V_{c_1} = -0{,}15$ frente a $\Delta V_{c_2} = -0{,}03$ indica una infraexposición cinco veces mayor para c_1.

Esta propiedad es la que vuelve las métricas directamente integrables en un escenario de *fairness* a largo plazo. Formalmente, el modelo quedaría definido por la tupla $\langle \Delta V, \epsilon, C' \rangle$, donde ΔV es la métrica elegida, $\epsilon > 0$ el umbral de tolerancia y $C' \subseteq C$ el conjunto de proveedores cubiertos. El desempeño esperado del sistema vendría dado por la verificación de que $|\Delta V_c| \leq \epsilon$ para todo $c \in C'$.

3.7. Agregación global del sistema

Las métricas anteriores se definen a nivel de proveedor y son útiles para diagnosticar incumplimientos individuales. Sin embargo, para comparar algoritmos o auditar el sistema en su conjunto resulta conveniente disponer también de

una medida agregada que resuma, con un único valor, el comportamiento global del recomendador respecto a un conjunto de proveedores auditados. Entre las múltiples familias de medidas de desigualdad disponibles [26], optamos por la media de desviaciones absolutas por su interpretabilidad directa en el contexto de auditoría. Definimos así la desviación media de visibilidad como:

$$D(w) = \frac{1}{|C'|} \sum_{c \in C'} |\Delta V_c(w)|, \tag{11}$$

donde $C' \subseteq C$ es el conjunto de proveedores sujetos a auditoría. Esta medida resume el nivel medio de desviación de visibilidad entre los proveedores auditados y facilita la comparación entre algoritmos: $D(w) = 0$ caracteriza la equidad perfecta a nivel agregado, y valores crecientes indican mayor brecha media respecto a la representación de referencia.

A nivel operativo, $D(w)$ admite una lectura directa en el escenario de *fairness* a largo plazo introducido previamente: la condición $D(w) \leq \epsilon$ constituye un criterio de auditoría agregada complementario al criterio individual $|\Delta V_c| \leq \epsilon$.

4. Evaluación experimental

La motivación de esta evaluación es eminentemente operativa. En un modelo de *fairness* a largo plazo, para un proveedor (un sello discográfico, un artista) el sistema debe auditarse para detectar incumplimientos. En ese marco, nuestra evaluación gira en torno a tres preguntas de investigación:

RQ1 ¿Las métricas propuestas permiten detectar desviaciones anuales claras del *provider fairness* ?

RQ2 ¿Las cuatro variantes capturan aspectos complementarios del *fairness*, o producen diagnósticos equivalentes?

RQ3 ¿Existe una tensión entre repartir el volumen de exposición y garantizar cobertura de presencia entre proveedores, y cómo evoluciona en el tiempo?

Se han evaluado cinco algoritmos de recomendación: Mult-VAE [25], SVD [24], LightGCN [20], ALS [22] y LMF [23], todos generando listas de $K = 20$ recomendaciones, con inferencia mediante FAISS sobre GPU. Sobre cada lista se calculan las cuatro métricas de visibilidad desigual propuestas: la uniforme ΔV_c^{uni}, la ponderada por posición ΔV_c^{pos}, la basada en zonas ΔV_c^{zone} y la binaria ΔV_c^{bin}.

4.1. Conjunto de datos

Hemos usado el subconjunto *CDs & Vinyl* de Amazon Reviews 2023 [21], tratado como *feedback* implícito. El protocolo reproduce una auditoría anual de *fairness* a largo plazo: para cada ejercicio $t \in \{2017, \ldots, 2022\}$ se aíslan las interacciones de ese año, se reparten al azar en un 80 % de entrenamiento y un 20 % de evaluación (con semilla fija), y cada modelo se entrena y evalúa por separado. Este diseño emula el caso real en el que la plataforma opera un

sistema de recomendación a lo largo del año t y, al cierre del ejercicio, calcula las métricas ΔV_c para cada proveedor. Comparar el valor observado $\Delta V_c^{(t)}$ con el umbral ϵ es lo que determina si procede compensación.

El dominio se presta especialmente a este ejercicio porque la popularidad de los sellos discográficos sigue una distribución muy sesgada. Los cinco proveedores analizados —*Legacy Recordings*, *Unknown*, *Capitol*, *CD Baby* y *Universal Music Group* (UMG)— concentran en conjunto en torno al 18 % de las interacciones, con *Legacy Recordings* y *Unknown* aportando, cada uno, alrededor de un 7 %. Conviene precisar que *Unknown* agrupa ítems sin sello identificado en los metadatos y se incluye como referencia del comportamiento del recomendador sobre catálogo no atribuido; su lectura no es por tanto equivalente a la de un sello real, pero su volumen e impacto en la exposición justifican monitorizarlo.

4.2. Evolución de la calidad predictiva

La Tabla 2 recoge la evolución de Precision@20, Recall@20 y NDCG@20 para cada modelo entre 2017 y 2022 (en negrita, el mejor valor anual de cada métrica). Mult-VAE se sitúa por encima del resto con un margen muy amplio: su Recall@20 se mueve entre 0.106 y 0.131, frente a valores que no superan 0.040 en los demás modelos. A partir de 2019–2020 se aprecia, además, una tendencia general a la baja, probablemente asociada a la creciente fragmentación del catálogo y al menor solapamiento entre los patrones de consumo de cada ejercicio.

Tabla 2. Evolución de las métricas de calidad ($K = 20$) por modelo y año. En negrita, el mejor valor anual de cada métrica.

Métrica	Modelo	2017	2018	2019	2020	2021	2022
P@20	Mult-VAE	**.0079**	.0077	.0078	.0069	.0074	.0061
	LightGCN	**.0024**	.0021	.0021	.0014	.0016	.0015
	ALS	**.0023**	.0020	.0019	.0012	.0014	.0012
	LMF	.0017	.0015	**.0021**	.0014	.0013	.0012
	SVD	**.0010**	.0009	.0009	.0008	.0008	.0008
R@20	Mult-VAE	.1273	.1262	**.1310**	.1232	.1251	.1062
	LightGCN	**.0391**	.0345	.0356	.0249	.0266	.0260
	ALS	**.0335**	.0299	.0281	.0195	.0213	.0176
	LMF	.0285	.0264	**.0368**	.0253	.0212	.0207
	SVD	**.0151**	.0145	.0146	.0122	.0127	.0123
NDCG@20	Mult-VAE	.0714	.0728	.0785	**.0839**	.0806	.0764
	LightGCN	**.0203**	.0179	.0187	.0131	.0143	.0137
	ALS	**.0181**	.0158	.0156	.0110	.0110	.0097
	LMF	.0156	.0146	**.0197**	.0173	.0144	.0169
	SVD	**.0094**	.0088	.0090	.0074	.0071	.0074

4.3. Evolución del *fairness* de proveedores

La Fig. 1 muestra la evolución temporal de la desviación media agregada $D(w)$ definida en la Ec. (11), por modelo y año, para cada una de las cuatro variantes propuestas. La línea discontinua horizontal en $D = 0$ marca la equidad perfecta a nivel agregado. Para no sobrecargar la figura principal, las trayectorias

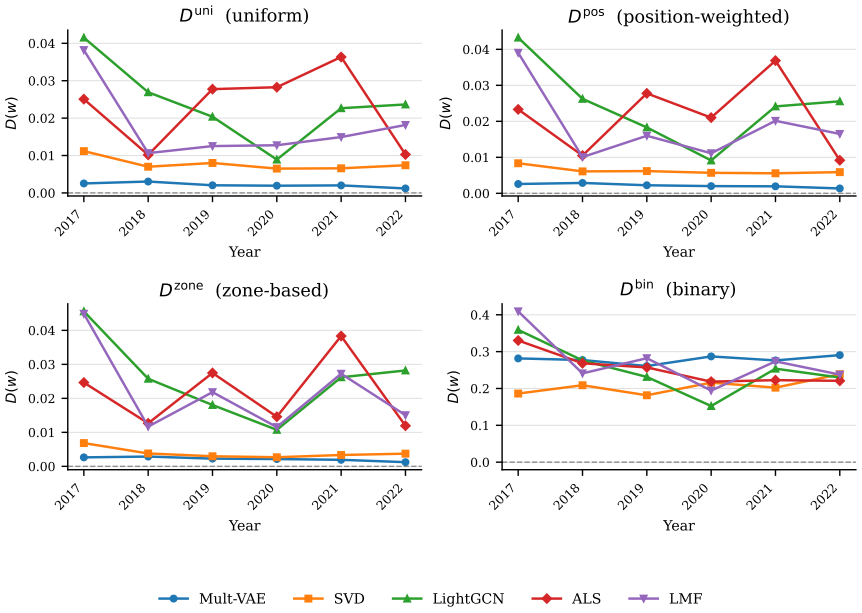


Figura 1. Desviación media agregada $D(w)$ por modelo y año, para las cuatro variantes propuestas: uniforme (D^{uni}), ponderada por posición (D^{pos}), basada en zonas (D^{zone}) y binaria (D^{bin}). La línea discontinua indica $D = 0$ (equidad perfecta a nivel agregado).

individuales por proveedor se omiten; los valores numéricos relevantes a nivel de proveedor se discuten en el texto.

RQ1: Detección de desviaciones. Las métricas revelan diferencias notables entre modelos y una dinámica temporal claramente no estacionaria, lo que sostiene su uso como instrumento de auditoría anual. En las métricas volumétricas (D^{uni}, D^{pos}, D^{zone}), Mult-VAE registra los valores más próximos a cero en todos los ejercicios ($D^{\text{uni}} \approx 0{,}002$ de media), seguido por SVD (en torno a 0,004–0,008), mientras que LightGCN, ALS y LMF se sitúan claramente por encima ($D \approx 0{,}02$–$0{,}04$). ALS, en particular, alcanza su máximo en 2021 ($D^{\text{uni}} \approx 0{,}036$), mostrando desviaciones importantes desde el punto de vista del nivel de *fairness* contratado. Las tres variantes volumétricas dibujan trayectorias muy parecidas, mientras que la binaria (Fig. 1, panel inferior derecho) opera en un rango distinto y con un orden de modelos diferente, fenómeno que se analiza con detalle en RQ2.

En cuanto a los proveedores, *Legacy Recordings* y *Unknown* aparecen sobrerrepresentados de forma sistemática (brecha media de +0,108 y +0,130, respectivamente). Bajo un modelo de *Fairness-as-a-Service*, estos sellos habrían recibido *exposición por encima de lo contratado*, y la plataforma podría retener la diferencia o redistribuirla. *CD Baby* y *UMG*, en cambio, se sitúan cerca de lo proporcional (+0,007 y +0,017), sin indicios de incumplimiento sistemático.

Capitol presenta una sobrerrepresentación moderada (+0,075). Esta variabilidad año a año pone de manifiesto algo importante en relación con el hecho de que un nivel de *fairness* contratado calculado sobre métricas estáticas puede acabar sistemáticamente incumplido por ciertos algoritmos, y subraya la necesidad de auditar el *fairness* de manera continua.

RQ2: Complementariedad de las variantes. Las cuatro métricas dan lugar a perfiles de *fairness* distintos. Las tres volumétricas (D^{uni}, D^{pos}, D^{zone}) coinciden de manera muy estrecha, lo cual sería esperado puesto que modelan el mismo fenómeno con esquemas de ponderación diferentes. La binaria, D^{bin}, en cambio, redibuja por completo el panorama (Fig. 1, panel inferior derecho). Mult-VAE, que dominaba en las volumétricas, registra ahora los valores más altos ($D^{\text{bin}} \approx 0{,}279$ de media) y mantiene esa trayectoria plana a lo largo del periodo: sus recomendaciones colocan a los grandes proveedores en el *top-K* de una fracción muy elevada de usuarios. La lectura diagnóstica es interesante: Mult-VAE distribuye bien el *volumen* de exposición entre proveedores, pero no garantiza diversidad de *presencia* en cada lista individual. SVD, por su parte, ofrece los mejores valores globales de D promediando las cuatro variantes ($\approx$ 0,056) y es además el modelo con menor D^{bin} ($\approx 0{,}206$), aunque al precio de una calidad predictiva claramente inferior. En definitiva, las cuatro variantes son necesarias para un diagnóstico completo.

RQ3: Tensión entre volumen y cobertura, y su evolución temporal. Los resultados revelan una tensión sistemática entre repartir el *volumen* de exposición entre proveedores y garantizar *cobertura* de presencia en las listas individuales. Mult-VAE ejemplifica esta tensión con nitidez: ofrece el mejor reparto volumétrico ($D^{\text{uni}}, D^{\text{pos}}, D^{\text{zone}}$ casi nulos y planos en todo el periodo) pero a costa del peor comportamiento en cobertura ($D^{\text{bin}} \approx 0{,}279$, también estable), porque sus listas saturan el *top-K* con los mismos grandes proveedores. SVD describe el patrón inverso, con la mejor cobertura agregada y un reparto volumétrico claramente peor, mientras que LightGCN, ALS y LMF se sitúan en posiciones intermedias en ambas dimensiones, sin destacar en ninguna. La tensión, además, es ortogonal a la calidad predictiva (Tabla 2): Mult-VAE lidera ampliamente en P@20, R@20 y NDCG@20, y aun así no escapa al compromiso volumen–cobertura. Esta tensión se mantiene a lo largo de los seis ejercicios analizados, lo que indica que no es un fenómeno coyuntural sino una propiedad estructural de los algoritmos. En la práctica, una plataforma que opere bajo un esquema de *Fairness-as-a-Service* no puede comprometerse a un único umbral ϵ sobre una única métrica: el contrato debe especificar qué dimensión (volumen, cobertura, o una combinación) se garantiza, y la elección de algoritmo se condiciona en consecuencia.

5. Conclusiones y trabajo futuro

En este trabajo se ha presentado una familia unificada de métricas de visibilidad desigual para cuantificar el *provider fairness* en sistemas de recomendación. Las cuatro variantes (uniforme, ponderada por posición, basada en zonas y bina-

ria) nacen de un mismo marco formal parametrizado por una función de atención posicional $w(k)$ y se han diseñado para ser utilizadas en escenarios de *fairness* a largo plazo.

La evaluación sobre Amazon *CDs & Vinyl*, con cortes anuales independientes entre 2017 y 2022 y cinco algoritmos distintos, deja tres conclusiones claras. Primera: las métricas discriminan eficazmente entre modelos y dejan ver dinámicas temporales no estacionarias. Segunda: las cuatro variantes son complementarias; en particular, ΔV^{bin} invierte el ranking de modelos respecto a las volumétricas, de modo que un diagnóstico riguroso requiere las cuatro. Y tercera: existe una tensión estructural entre volumen y cobertura de exposición que ningún algoritmo resuelve completamente (Mult-VAE optimiza el volumen y SVD la cobertura) y que se mantiene estable a lo largo del periodo analizado, lo que justifica especificar la dimensión de *fairness* garantizada en cualquier acuerdo operativo.

Como líneas futuras, se contemplan tres frentes. El primero, extender el estudio a otros dominios y conjuntos de datos para comprobar hasta qué punto los resultados se generalizan. El segundo, diseñar algoritmos de *re-ranking* que optimicen directamente las métricas propuestas. Y el tercero, llevar la propuesta a una plataforma completa de evaluación de *fairness*, ya desplegada y en tiempo real.

Referencias

1. Parlamento Europeo y Consejo de la Unión Europea: Reglamento (UE) 2024/1689 por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial). Diario Oficial de la Unión Europea (2024)
2. Biega, A.J., Gummadi, K.P., Weikum, G.: Equity of attention: amortizing individual fairness in rankings. In: Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'18), pp. 405–414. ACM (2018)
3. Burke, R.: Multisided fairness for recommendation. In: Workshop on Fairness, Accountability and Transparency in Machine Learning (FATML). Halifax, NS, Canada (2017)
4. Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X.: Bias and debias in recommender system: a survey and future directions. ACM Trans. Inf. Syst. **41**(3), 1–39 (2023)
5. Deldjoo, Y., Jannach, D., Bellogín, A., Difonzo, A., Zanzonelli, D.: Fairness in recommender systems: research landscape and future directions. User Modeling and User-Adapted Interaction. **34**(1), 59–108 (2024)
6. Smith, J.J., Satwani, A., Burke, R., Fiesler, C.: Recommend me? Designing fairness metrics with providers. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 2389–2399)
7. Ekstrand, M.D., Das, A., Burke, R., Diaz, F.: Fairness in information access systems. Found. Trends Inf. Retr. **16**(1–2), 1–177 (2022)
8. Gómez, E., Contreras, D., Boratto, L., Salamó, M.: Enhancing recommender systems with provider fairness through preference distribution awareness. Int. J. Inf. Manag. Data Insights **5**(1), 100311 (2025)

9. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Trans. Inf. Syst. **20**(4), 422–446 (2002)
10. Joachims, T., Swaminathan, A., Schnabel, T.: Unbiased learning-to-rank with biased feedback. In: Proceedings of the 10th ACM International Conference on Web Search and Data Mining (WSDM'17), pp. 781–789. ACM (2017)
11. Wu, Y., Cao, J., Xu, G.: Fairness in recommender systems: evaluation approaches and assurance strategies. ACM Trans. Knowl. Discov. Data **18**(1),1-37
12. Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., Diaz, F.: Towards a fair marketplace: counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM'18), pp. 2243–2251. ACM (2018)
13. Shafiloo, R., Stratigi, M., Peltonen, J., Olsson, T., Stefanidis, K.: The many facets of fairness in recommender systems: consumers, providers and items. Inf. Syst. **137**, 102643 (2026)
14. Patro, G.K., Biswas, A., Ganguly, N., Gummadi, K.P., Chakraborty, A.: FairRec: two-sided fairness for personalized recommendations in two-sided platforms. In: Proceedings of The Web Conference 2020 (WWW'20), pp. 1194–1204. ACM (2020)
15. Singh, A., Joachims, T.: Fairness of exposure in rankings. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'18), pp. 2219–2228. ACM (2018)
16. Vassøy, B., Langseth, H.: Consumer-side fairness in recommender systems: a systematic survey of methods and evaluation. Artif. Intell. Rev. **57**, Article 101 (2024)
17. Wang, Y., Ma, W., Zhang, M., Liu, Y., Ma, S.: A survey on the fairness of recommender systems. ACM Trans. Inf. Syst. **41**(3), 1–43 (2023)
18. Xu, C., Chen, S., Xu, J., Shen, W., Zhang, X., Wang, G., Dong, Z.: P-MMF: provider max-min fairness re-ranking in recommender system. In: Proceedings of the ACM Web Conference 2023 (pp. 3701–3711. ACM (2023)
19. Zhao, Y., Wang, Y., Liu, Y., Cheng, X., Aggarwal, C., Derr, T.: Fairness and diversity in recommender systems: a survey. ACM Trans. Intell. Syst. Technol. **16**(1), 1–28 (2025)
20. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN: simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20), pp. 639–648. ACM (2020)
21. Hou, Y., Li, J., He, Z., Yan, A., Chen, X., McAuley, J.: Bridging language and items for retrieval and recommendation. arXiv preprint arXiv:2403.03952 (2024)
22. Hu, Y., Koren, Y., Volinsky, C.: Collaborative filtering for implicit feedback datasets. In: Proceedings of the 8th IEEE International Conference on Data Mining (ICDM'08), pp. 263–272. IEEE (2008)
23. Johnson, C.C.: Logistic matrix factorization for implicit feedback data. In: Advances in Neural Information Processing Systems (NeurIPS'14) Workshop (2014)
24. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. IEEE Computer **42**(8), 30–37 (2009)
25. Liang, D., Krishnan, R.G., Hoffman, M.D., Jebara, T.: Variational autoencoders for collaborative filtering. In: Proceedings of The Web Conference 2018 (WWW'18), pp. 689–698. ACM (2018)
26. Shorrocks, A.F.: The class of additively decomposable inequality measures. Econometrica **48**(3), 613–625 (1980)