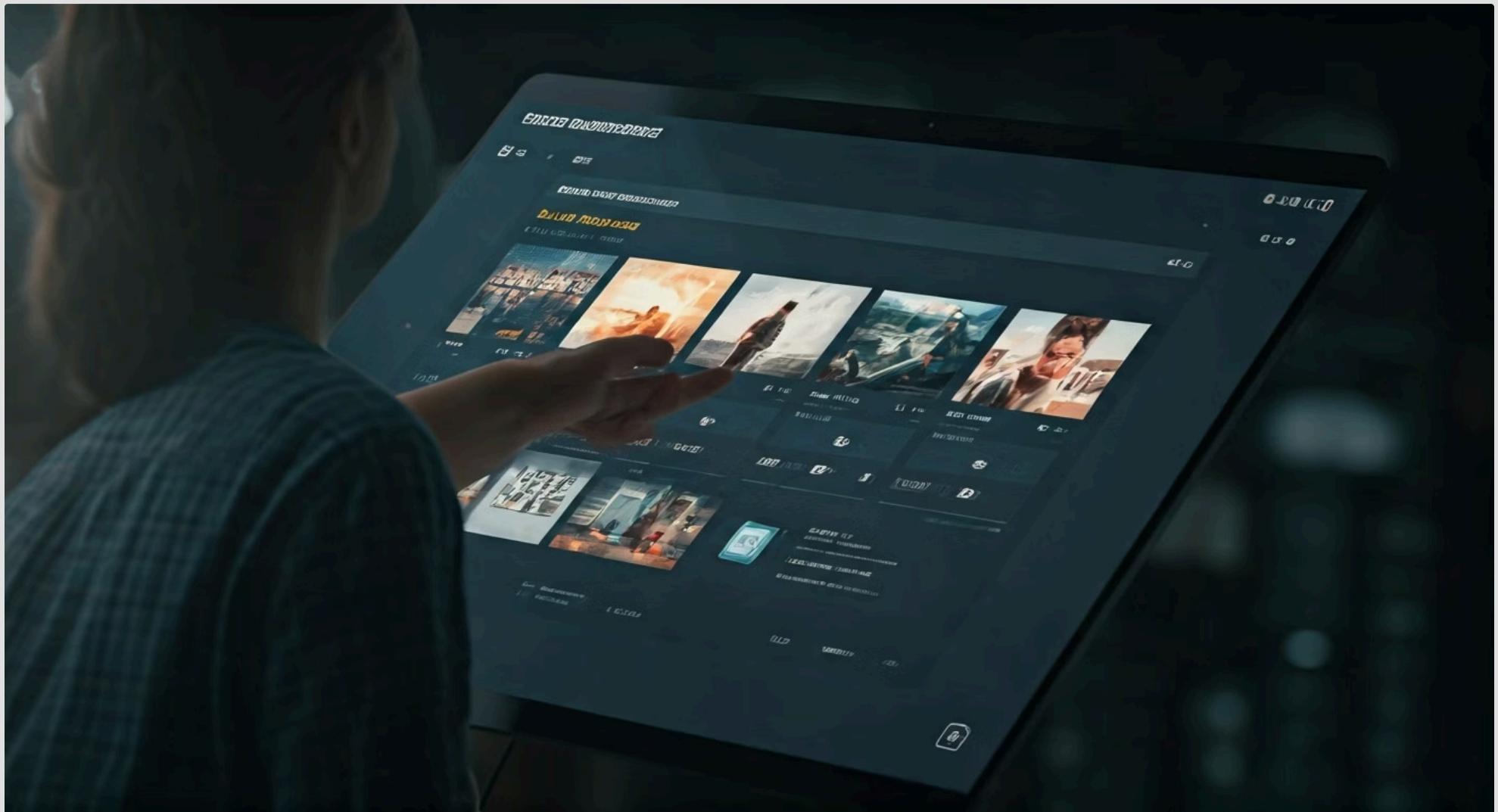


# Aula 13 – Métricas de Ranking e Relevância (Parte 2)



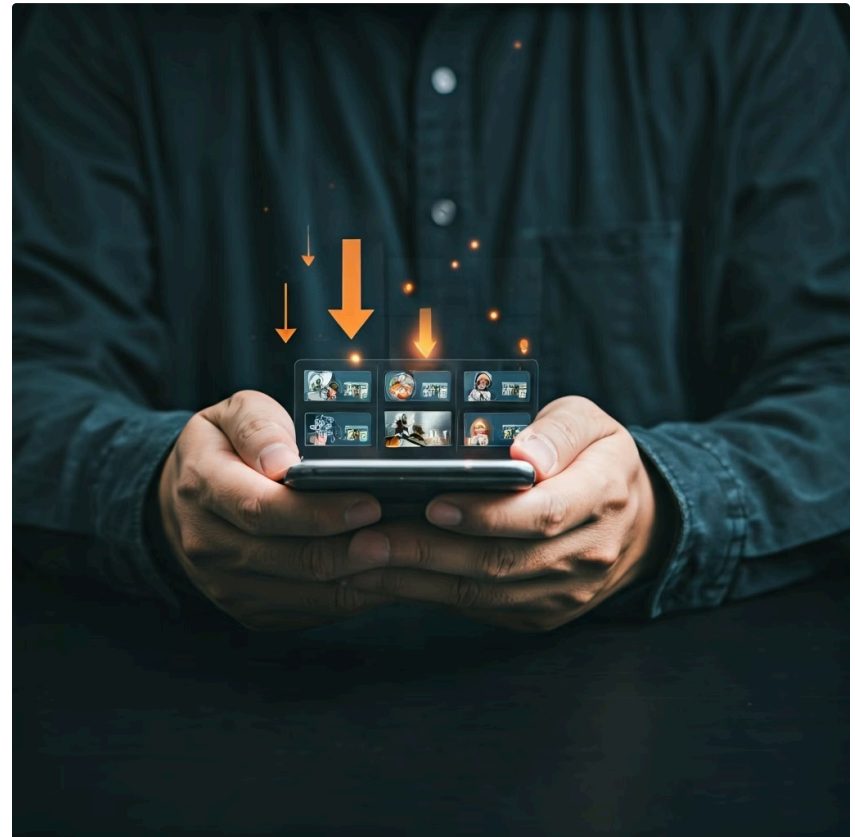
No universo dos sistemas de recomendação, entregar o item certo é apenas metade da batalha. A outra metade, igualmente crucial, reside em como e onde esse item é apresentado ao usuário. Imagine-se navegando por um e-commerce ou um serviço de streaming: a probabilidade de você clicar em um item na primeira linha é muito maior do que em um que aparece após rolar a página várias vezes. Essa percepção intuitiva é a base para entendermos as métricas de ranking e relevância que consideram a posição.

Nesta aula, mergulharemos em métricas sofisticadas que vão além da simples contagem de acertos, incorporando a importância da posição do item na lista de recomendação. Você aprenderá a calcular e interpretar o Ganho Cumulativo Descontado (DCG) e o Ganho Cumulativo Descontado Normalizado (NDCG), ferramentas essenciais para avaliar a qualidade de um sistema de recomendação de forma mais granular e alinhada à experiência do usuário. Nosso objetivo é que, ao final, você seja capaz de escolher a métrica mais adequada para diferentes problemas de negócio, considerando não apenas a relevância, mas também o impacto da apresentação. Prepare-se para refinar sua compreensão sobre como medir o sucesso em um dos campos mais dinâmicos da inteligência artificial.

# A Importância da Posição na Lista de Recomendação

Ao avaliar a eficácia de um sistema de recomendação, é tentador focar apenas na quantidade de itens relevantes que ele consegue identificar. No entanto, a realidade da interação humana com essas listas nos mostra que a **posição** de um item é tão crítica quanto sua relevância. Pense na sua própria experiência: você tende a prestar mais atenção aos primeiros resultados de uma busca ou aos primeiros produtos sugeridos em uma página. Itens relevantes que aparecem muito abaixo na lista têm uma chance mínima de serem vistos ou clicados, perdendo seu valor potencial.

Essa dinâmica nos leva a um problema fundamental: como podemos quantificar o impacto da posição na avaliação de um sistema? Métricas simples, como Precisão@K ou Recall@K, tratam todos os itens dentro de um "top K" como igualmente importantes, independentemente de estarem na primeira ou na décima posição. Isso ignora um comportamento humano básico: a fadiga de rolagem e a atenção decrescente. Precisamos de ferramentas que reflitam essa realidade, atribuindo maior peso aos itens relevantes que aparecem nas posições de destaque.



- ☐ **Analogia do Supermercado:** Os produtos mais caros ou com maior margem de lucro geralmente são colocados na altura dos olhos, nas pontas de gôndola ou em locais de grande fluxo. Da mesma forma, em um sistema de recomendação, os itens que o algoritmo considera mais relevantes para o usuário devem ocupar as posições mais privilegiadas.

Se um sistema recomenda um filme perfeito para você, mas o coloca na 50ª posição da lista, é provável que você nunca o encontre. É essa nuance que as métricas de ranking e relevância buscam capturar, garantindo que a avaliação do sistema espelhe a experiência real do usuário.

# Entendendo o Ganho Cumulativo (CG) e a Relevância Ponderada

Antes de mergulharmos nas complexidades do desconto posicional, é fundamental compreender o conceito de **ganho cumulativo** e como atribuímos relevância aos itens. Em muitos cenários, a relevância de um item não é uma simples questão de "sim" ou "não". Um filme pode ser "muito relevante", "moderadamente relevante" ou "pouco relevante" para um usuário específico. Essa gradação é crucial para uma avaliação mais rica dos sistemas de recomendação.

## Score 0

Não relevante

## Score 1

Relevância baixa

## Score 2

Relevância média

## Score 3

Relevância alta

Para capturar essa nuance, atribuímos um **score de relevância** a cada item. Por exemplo, em uma escala de 0 a 3, onde 0 significa "não relevante", 1 "relevância baixa", 2 "relevância média" e 3 "relevância alta". O Ganho Cumulativo (CG) de uma lista de recomendação é simplesmente a soma desses scores de relevância para todos os itens na lista, até uma determinada posição K. Ele nos dá uma ideia do "valor total" de relevância que o sistema entregou.

**Limitação do CG:** Ele trata todos os itens relevantes como se tivessem o mesmo impacto, independentemente de sua posição. Um item com relevância 3 na primeira posição contribui o mesmo para o CG que um item com relevância 3 na décima posição.

Isso, como discutimos, não reflete a realidade da interação do usuário. A atenção do usuário diminui drasticamente à medida que ele percorre a lista, e, portanto, a contribuição de um item relevante para a satisfação geral deveria ser maior se ele estiver em uma posição mais alta.

Essa lacuna nos leva à necessidade de uma métrica que não apenas some os ganhos de relevância, mas também os **desconte** com base na posição. É como encontrar um tesouro: a alegria é maior se você o encontra logo no início da sua busca, em vez de depois de horas de escavação. O valor percebido do tesouro é influenciado pela facilidade e rapidez com que ele foi descoberto. Essa é a intuição por trás do Ganho Cumulativo Descontado, que veremos a seguir.

# Desvendando o Ganho Cumulativo Descontado (DCG)

A limitação do Ganho Cumulativo (CG) em não considerar a posição nos leva diretamente ao **Ganho Cumulativo Descontado (DCG - Discounted Cumulative Gain)**. Esta métrica foi criada para resolver exatamente esse problema, atribuindo pesos decrescentes aos itens relevantes à medida que sua posição na lista de recomendação diminui. Em outras palavras, um item altamente relevante na primeira posição contribui muito mais para o DCG do que o mesmo item na décima posição.

A ideia central por trás do "desconto" é que a atenção do usuário e a probabilidade de interação com um item diminuem logaritmicamente com sua posição. É como a lei dos rendimentos decrescentes: quanto mais você precisa "trabalhar" (rolar a página) para encontrar algo, menor o valor percebido daquele item, mesmo que ele seja intrinsecamente bom. O DCG formaliza essa intuição, aplicando um fator de desconto baseado no logaritmo da posição.



## 📄 Fórmula do DCG

$$DCG_K = \sum_{i=1}^K \frac{rel_i}{\log_2(i+1)}$$

### Onde:

- $rel_i$  é o score de relevância do item na posição  $i$
- $\log_2(i+1)$  é o fator de desconto

01

### Posição 1

Sem desconto:  $\log_2(2) = 1$

02

### Posições seguintes

Desconto crescente: denominador aumenta

03

### Resultado

Itens no topo têm maior contribuição

Essa abordagem garante que sistemas que colocam itens altamente relevantes no topo da lista sejam recompensados com um DCG maior. É uma métrica que espelha de forma mais fiel a experiência do usuário, pois reconhece que a "descoberta" de um item relevante é mais valiosa quando ocorre cedo. O DCG nos permite comparar a qualidade de diferentes listas de recomendação, levando em conta não apenas o que é recomendado, mas também onde.

# DCG em Ação: Um Exemplo Prático

Para solidificar a compreensão do DCG, vamos aplicar a fórmula a um cenário hipotético. Imagine que um sistema de recomendação gerou a seguinte lista de 5 itens para um usuário, e nós atribuímos scores de relevância (0-3) a cada um:

Posição (i)	Item Recomendado	Score de Relevância
1	Filme A	3 (Muito Relevante)
2	Filme B	2 (Relevância Média)
3	Filme C	0 (Não Relevante)
4	Filme D	3 (Muito Relevante)
5	Filme E	1 (Relevância Baixa)

## Cálculo do DCG<sub>5</sub>:

### Posição 1 (Filme A)

$rel_1 = 3$

$Fator: \log_2(2) = 1$

$Contribuição: 3/1 = 3$

### Posição 2 (Filme B)

$rel_2 = 2$

$Fator: \log_2(3) \approx 1.58$

$Contribuição: 2/1.58 \approx 1.26$

### Posição 3 (Filme C)

$rel_3 = 0$

$Fator: \log_2(4) = 2$

$Contribuição: 0/2 = 0$

### Posição 4 (Filme D)

$rel_4 = 3$

$Fator: \log_2(5) \approx 2.32$

$Contribuição: 3/2.32 \approx 1.29$

### Posição 5 (Filme E)

$rel_5 = 1$

$Fator: \log_2(6) \approx 2.58$

$Contribuição: 1/2.58 \approx 0.39$

## Resultado Final

$DCG_5 = 5.94$

## Cenário Otimizado

Agora, imagine que movemos o Filme D (relevância 3) para a posição 2 e o Filme B (relevância 2) para a posição 4:

### Nova Lista

- Posição 1: Filme A (3)
- Posição 2: Filme D (3)
- Posição 3: Filme C (0)
- Posição 4: Filme B (2)
- Posição 5: Filme E (1)

### Novo Cálculo

$Posição 2 (Filme D): 3/1.58 \approx 1.90$

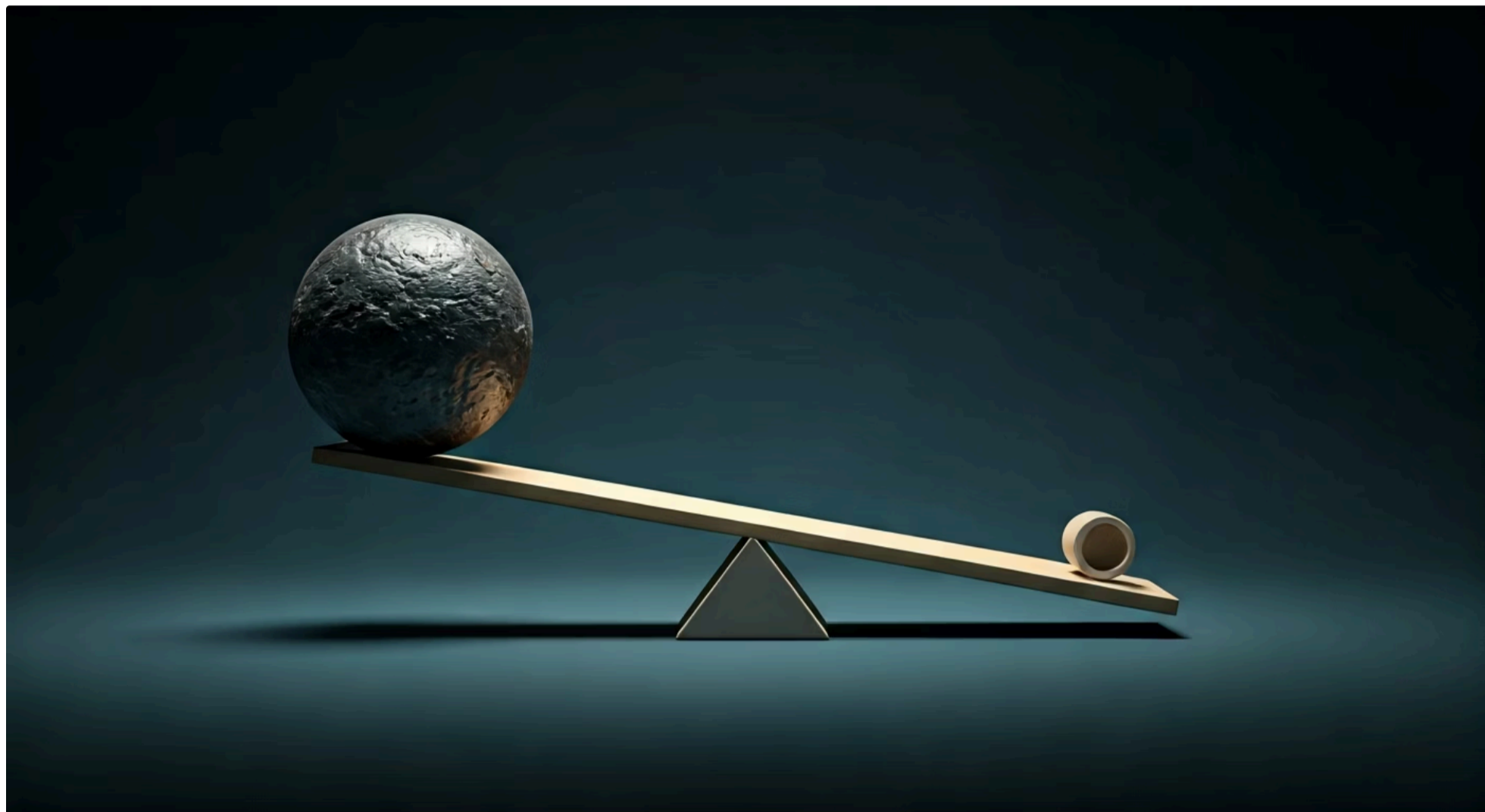
$Posição 4 (Filme B): 2/2.32 \approx 0.86$

$Novo DCG_5 = 6.15$

**Insight:** Ao mover um item altamente relevante para uma posição mais alta (Filme D da posição 4 para 2), o DCG aumentou de 5.94 para 6.15, mesmo mantendo os mesmos itens relevantes na lista. Isso demonstra como o DCG recompensa a colocação estratégica de itens relevantes, refletindo melhor a qualidade percebida pelo usuário.

# A Necessidade de Normalização: Por que o NDCG?

O DCG é uma métrica poderosa para avaliar a qualidade de uma única lista de recomendação, considerando a relevância e a posição. No entanto, ele apresenta uma limitação crucial quando precisamos **comparar** o desempenho de diferentes sistemas ou o mesmo sistema em diferentes cenários (por exemplo, para diferentes usuários ou diferentes consultas). O valor absoluto do DCG pode variar muito dependendo do número total de itens relevantes disponíveis e da extensão da lista de recomendação.



## Usuário A

Muitos itens relevantes disponíveis

**DCG potencialmente alto**



## Usuário B

Poucos itens relevantes disponíveis

**DCG potencialmente baixo**

### O Problema da Comparação

Um DCG de 10 para o Usuário A e um DCG de 5 para o Usuário B não nos diz qual sistema está performando melhor em relação ao seu potencial máximo. O DCG não é **normalizado**, o que significa que seu valor máximo não é fixo e depende do cenário específico.

Isso torna difícil comparar diretamente o desempenho entre diferentes consultas ou usuários, ou até mesmo entre diferentes modelos que podem gerar listas de tamanhos variados.

**Analogia:** É como comparar a pontuação de um aluno em um teste: saber que ele tirou 80% é mais informativo do que saber que ele tirou 80 pontos, pois o percentual nos dá uma referência em relação ao máximo possível.

Para superar essa limitação, precisamos de uma forma de padronizar o DCG. Precisamos de uma métrica que nos diga não apenas o "ganho descontado" que obtivemos, mas também o quão perto estamos do "ganho descontado ideal" que poderíamos ter obtido. Essa referência é o que o **Ganho Cumulativo Descontado Normalizado (NDCG)** nos oferece, permitindo comparações justas e significativas.

# O Poder do Ganho Cumulativo Descontado Normalizado (NDCG)

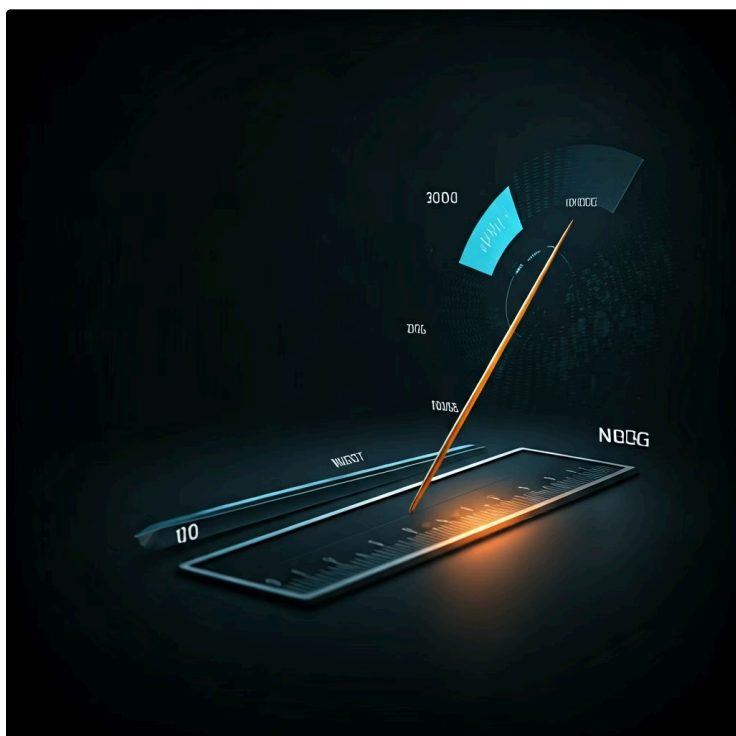
Para tornar o DCG comparável entre diferentes cenários, introduzimos o conceito de **Ganho Cumulativo Descontado Normalizado (NDCG - Normalized Discounted Cumulative Gain)**. O NDCG resolve o problema de comparabilidade do DCG ao normalizá-lo pelo **DCG Ideal (IDCG)**. O IDCG representa o DCG máximo possível que poderia ser alcançado para uma dada consulta, ou seja, o DCG de uma lista de recomendação perfeitamente ordenada, onde todos os itens relevantes são classificados nas posições mais altas, em ordem decrescente de relevância.

## Fórmula do NDCG

$$NDCG_K = \frac{DCG_K}{IDCG_K}$$

Onde:

- $DCG_K$  é o Ganho Cumulativo Descontado da lista gerada
- $IDCG_K$  é o Ganho Cumulativo Descontado da lista ideal



## Interpretação do NDCG

- **NDCG = 1:** Lista de recomendação perfeita, idêntica à lista ideal
- **NDCG = 0:** Nenhum item relevante recomendado ou em posições muito baixas
- **0 < NDCG < 1:** Quanto mais próximo de 1, melhor o desempenho



### Lista Gerada

Calcular DCG da recomendação atual



### Lista Ideal

Calcular IDCG da ordenação perfeita



### Normalização

Dividir DCG por IDCG



### Resultado

Score entre 0 e 1 comparável

**Vantagem Principal:** Pense no NDCG como uma pontuação padronizada. Assim como em um teste padronizado, onde sua pontuação é comparada à pontuação máxima possível, o NDCG compara o desempenho do seu sistema ao melhor desempenho que ele poderia ter. Isso permite que você compare a eficácia de diferentes algoritmos, mesmo que eles operem em conjuntos de dados com diferentes distribuições de relevância ou gerem listas de tamanhos variados.

É uma métrica robusta e amplamente utilizada na pesquisa e na indústria para avaliar a qualidade de ranking de sistemas de recomendação e busca.

# NDCG na Prática: Comparando Cenários

Vamos revisitar nosso exemplo anterior para calcular o NDCG e entender como ele facilita a comparação.

## Cenário 1: Lista Original

### Lista Gerada

- Filme A (relevância 3)
- Filme B (relevância 2)
- Filme C (relevância 0)
- Filme D (relevância 3)
- Filme E (relevância 1)

$DCG_5 = 5.94$

### Lista Ideal (IDCG)

- Filme A ou D (3)  $\rightarrow 3/1 = 3$
- Filme D ou A (3)  $\rightarrow 3/1.58 \approx 1.90$
- Filme B (2)  $\rightarrow 2/2 = 1$
- Filme E (1)  $\rightarrow 1/2.32 \approx 0.43$
- Filme C (0)  $\rightarrow 0/2.58 = 0$

$IDCG_5 = 6.33$

#### 📄 Cálculo do NDCG - Cenário 1

$$NDCG_5 = \frac{DCG_5}{IDCG_5} = \frac{5.94}{6.33} \approx 0.938$$

## Cenário 2: Lista Otimizada

### Lista Gerada Otimizada

- Filme A (relevância 3)
- Filme D (relevância 3)
- Filme C (relevância 0)
- Filme B (relevância 2)
- Filme E (relevância 1)

$DCG_5 = 6.15$

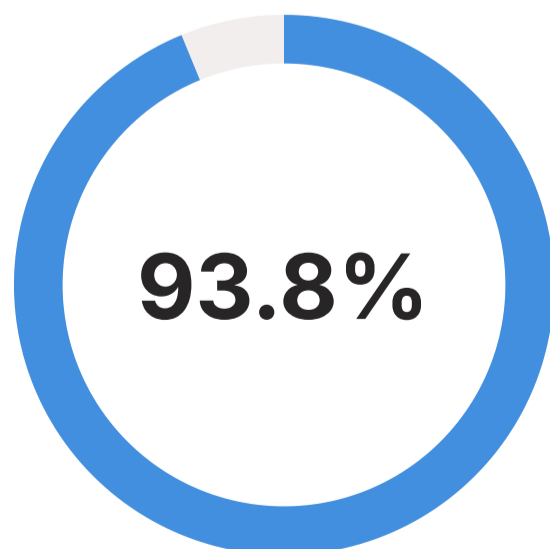
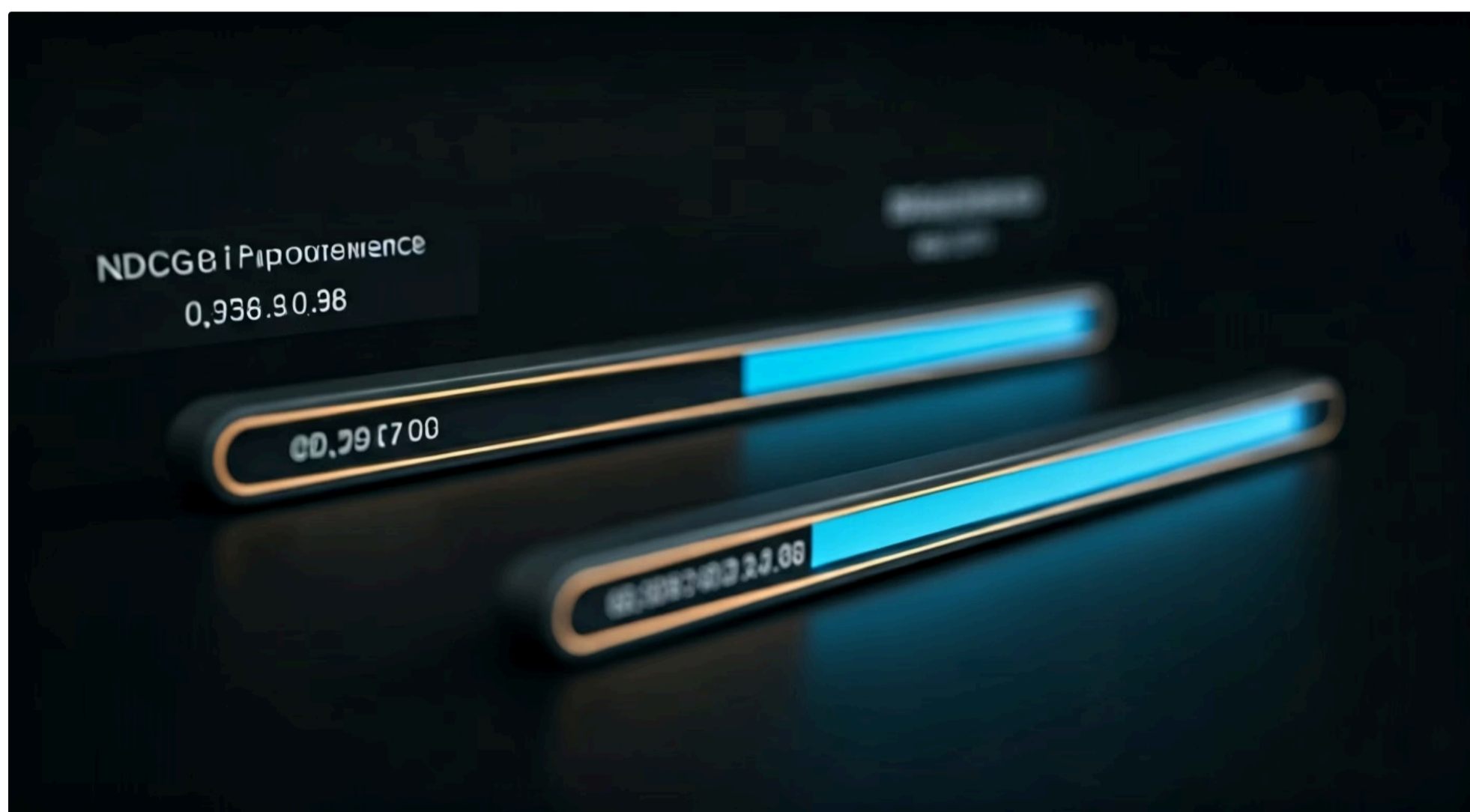
### IDCG (mesmo do Cenário 1)

O IDCG permanece o mesmo, pois é baseado em todos os itens relevantes disponíveis, não na lista gerada.

$IDCG_5 = 6.33$

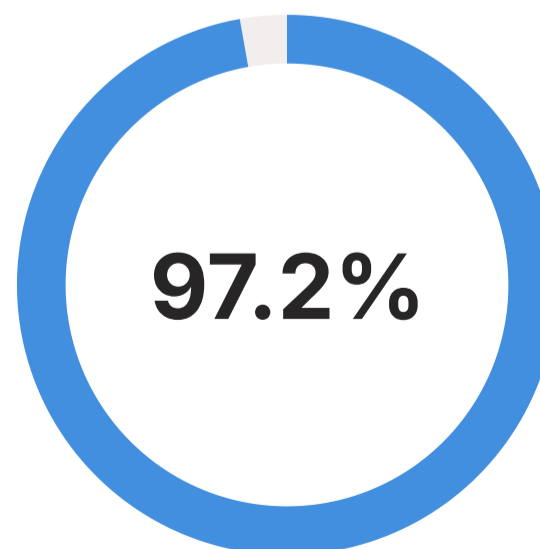
#### 📄 Cálculo do NDCG - Cenário 2

$$NDCG_5 = \frac{DCG_5}{IDCG_5} = \frac{6.15}{6.33} \approx 0.972$$



Cenário 1

Lista Original



Cenário 2

Lista Otimizada

**Conclusão:** A comparação dos NDCG (0.938 vs. 0.972) nos mostra claramente que a lista otimizada é superior, pois está mais próxima do ideal. O NDCG nos fornece uma métrica relativa, permitindo-nos avaliar o quão bem um sistema está performando em relação ao seu potencial máximo, o que é inestimável para benchmarking e otimização contínua.

# DCG vs. NDCG: Quando Usar Cada Um?

A escolha entre DCG e NDCG, ou mesmo outras métricas como Precision@K e Recall@K, depende fundamentalmente do objetivo da sua avaliação e do contexto do problema. Embora ambos o DCG e o NDCG considerem a relevância e a posição, suas aplicações ideais diferem devido à normalização.

## DCG

### Quando usar

Ganho absoluto de relevância descontada para uma consulta específica

### Vantagem

Mais direto, entende o "valor bruto" de uma lista

### Desvantagem

Difícil comparação entre diferentes consultas ou usuários

### Cenário ideal

Avaliação de uma única lista ou listas com mesmo conjunto de itens relevantes

## NDCG

### Quando usar

Comparação e benchmarking entre sistemas ou cenários

### Vantagem

Normalização permite comparações justas e robustas

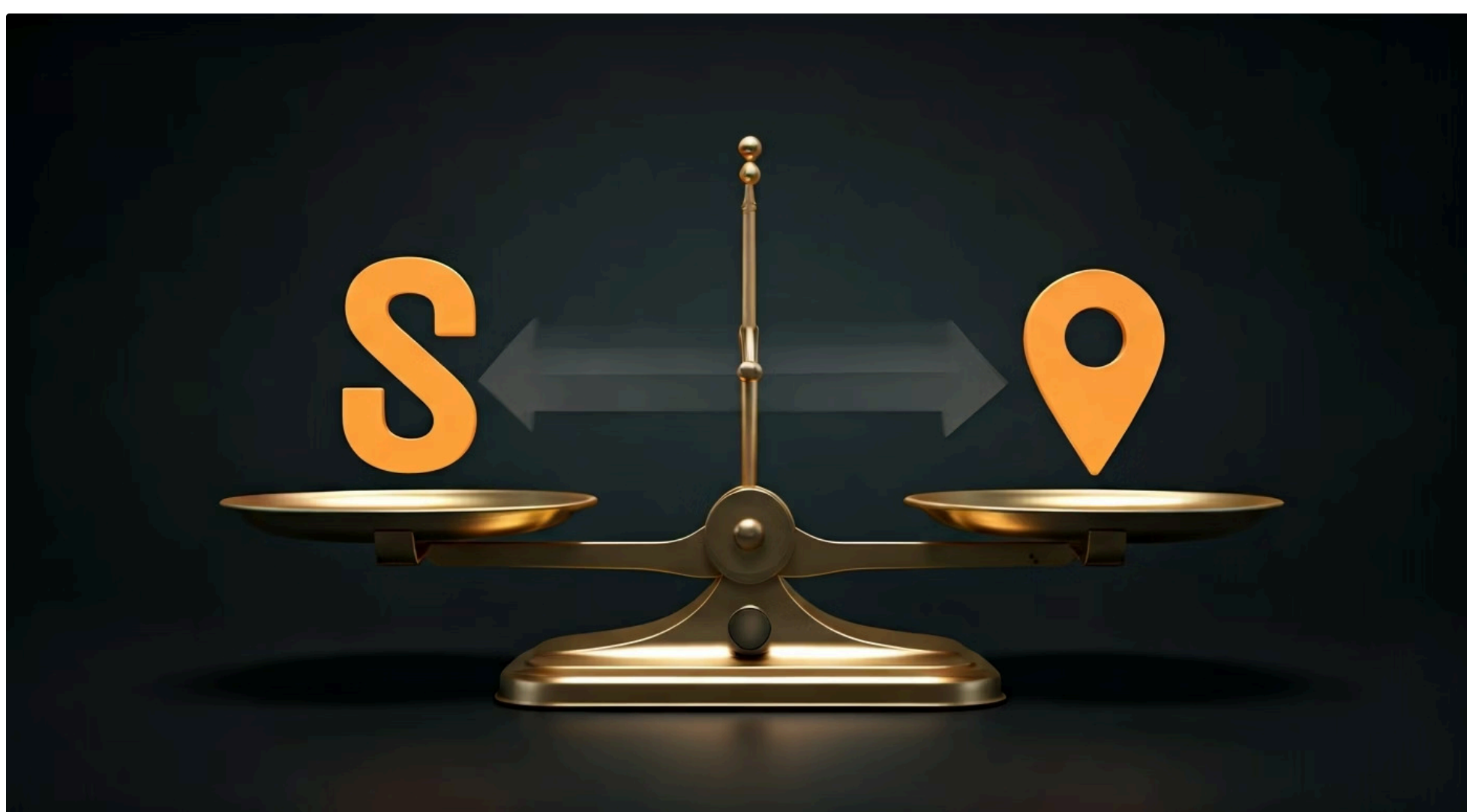
### Desvantagem

Requer cálculo adicional do IDCG

### Cenário ideal

Otimização de modelos, testes A/B, reportes padronizados

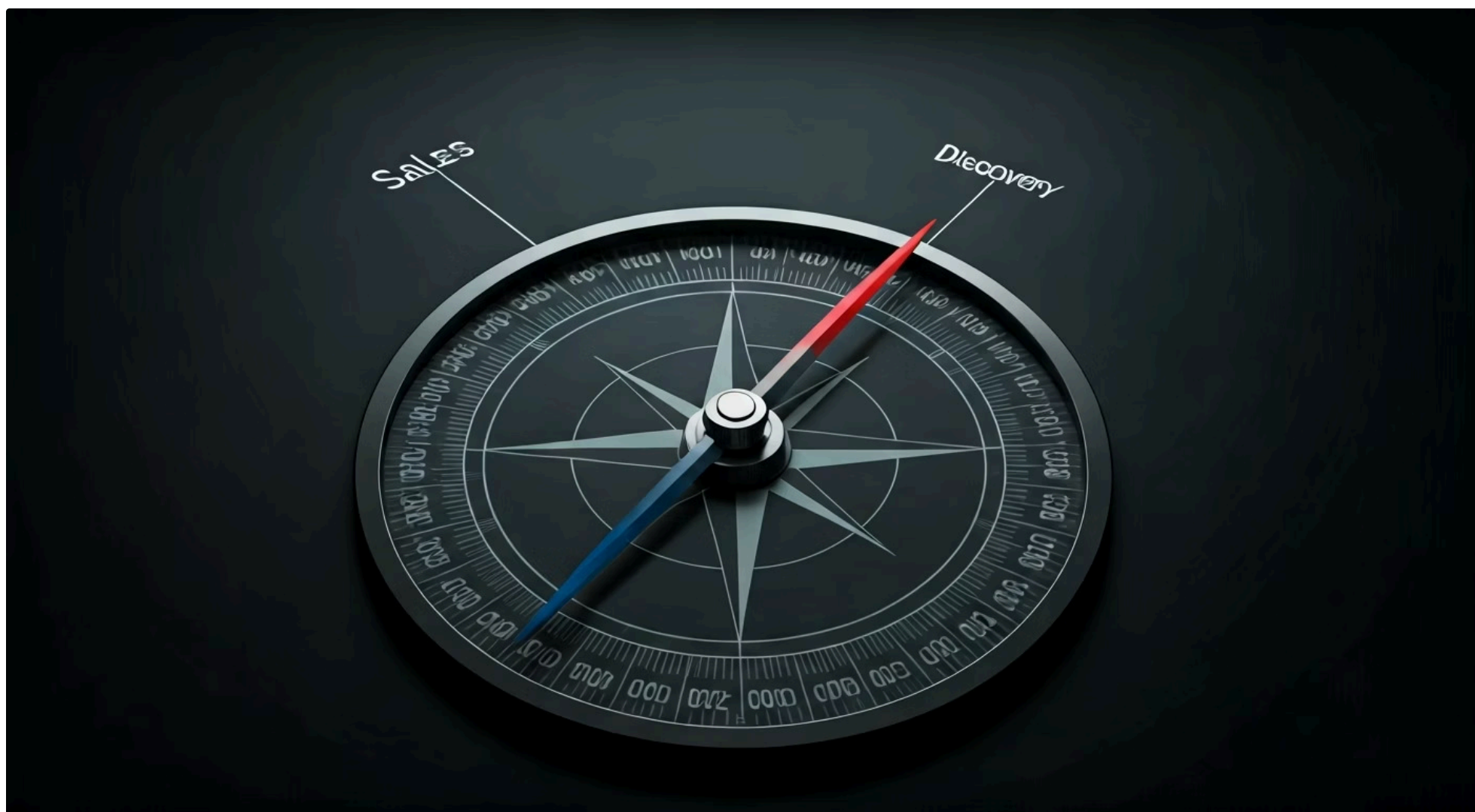
Conceito	Âmbito/Aplicação	Exemplo
DCG	Medir ganho absoluto de uma lista específica	Avaliar a qualidade de uma lista de 10 filmes para um usuário específico
NDCG	Comparar performance entre sistemas/cenários	Benchmarking de dois algoritmos de recomendação em um e-commerce



**Resumo:** Se você quer saber o "quão bom" é uma lista em si, o DCG pode ser informativo. Mas se você quer saber o "quão bom" é um sistema em relação ao que ele *poderia ser*, e compará-lo com outros, o NDCG é indispensável. A maioria das aplicações práticas e pesquisas em sistemas de recomendação e busca utiliza o NDCG devido à sua capacidade de fornecer uma medida de qualidade de ranking comparável e robusta.

# Além do Básico: Escolhendo a Métrica Certa para o Negócio

A seleção da métrica de avaliação não é apenas uma decisão técnica; é uma **decisão de negócio**. Cada métrica reflete uma perspectiva diferente sobre o que significa "sucesso" para um sistema de recomendação. Um erro comum é adotar uma métrica padrão sem antes alinhar com os objetivos estratégicos da empresa. Por exemplo, um e-commerce pode priorizar a conversão de vendas, enquanto um serviço de notícias pode focar no engajamento e tempo de leitura.



## Perguntas-Chave para Escolher a Métrica

❏ "O que queremos que o usuário faça e qual comportamento queremos incentivar?"



### Satisfação Imediata

**Objetivo:** Garantir que o usuário encontre rapidamente o que procura

**Métrica ideal:** NDCG@K (com K pequeno)

**Exemplo:** Sistema de busca, feed de notícias



### Diversidade de Itens

**Objetivo:** Maximizar a variedade de itens relevantes que o usuário vê

**Métrica ideal:** Recall@K (com K maior) + métricas de diversidade

**Exemplo:** Plataformas de descoberta de conteúdo



### Relevância Granular

**Objetivo:** Considerar níveis de relevância e posição

**Métrica ideal:** DCG e NDCG

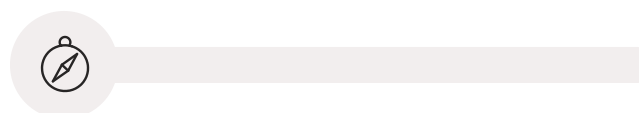
**Exemplo:** Sistemas com múltiplos níveis de relevância

**Analogia:** A escolha da métrica é como selecionar a ferramenta certa para um trabalho. Você não usaria um martelo para apertar um parafuso. Da mesma forma, usar Precision@K para um problema onde a relevância gradual e a posição são cruciais pode levar a conclusões erradas sobre o desempenho do seu sistema.

É fundamental que a métrica escolhida reflita diretamente o valor que o sistema de recomendação está gerando para o negócio, seja ele engajamento, conversão, retenção ou satisfação do usuário.

# Métricas e o Ciclo de Vida do Produto: Da Descoberta à Retenção

A jornada do usuário com um produto ou serviço é dinâmica, e as métricas de recomendação devem se adaptar a cada estágio desse ciclo. Um sistema de recomendação eficaz não apenas ajuda o usuário a encontrar o que ele já sabe que quer, mas também o guia através de novas descobertas e o mantém engajado ao longo do tempo.



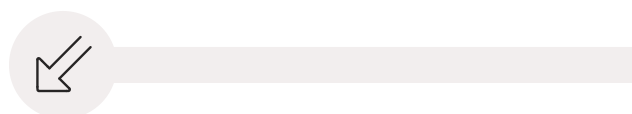
## 1. Descoberta (Novos Usuários)

**Foco:** Boa primeira impressão e exploração do catálogo

### Métricas importantes:

- Diversidade de recomendações
- Serendipidade (descoberta inesperada)
- Cobertura do catálogo

**Objetivo:** Evitar a "tela em branco" e construir perfil inicial



## 2. Engajamento (Usuários Ativos)

**Foco:** Precisão nas recomendações baseadas em histórico

### Métricas importantes:

- NDCG@K e DCG@K
- Taxa de cliques (CTR)
- Tempo de sessão
- Taxa de retorno

**Objetivo:** Maximizar interações e satisfação contínua



## 3. Retenção e Fidelidade (Usuários Leais)

**Foco:** Manter relevância e introduzir novidades

### Métricas importantes:

- Novidade (itens não vistos antes)
- Cobertura do catálogo
- NDCG + métricas de diversidade
- Lifetime Value (LTV)

**Objetivo:** Evitar "bolha de filtro" e manter interesse a longo prazo

### Insight Importante

A escolha da métrica não é estática; ela deve ser revisada e ajustada conforme o usuário avança em sua jornada e os objetivos de negócio se transformam. Um sistema maduro deve ter um conjunto de métricas que evolui com o ciclo de vida do usuário.

# Tendências 2025: Deep Learning e Métricas Adaptativas

O campo dos sistemas de recomendação está em constante evolução, impulsionado por avanços em inteligência artificial. Para 2025, a influência do **Deep Learning** é inegável, transformando a forma como os modelos aprendem e, conseqüentemente, como avaliamos seu desempenho. A adoção massiva de redes neurais, especialmente **Embeddings**, permite capturar relações complexas entre usuários e itens que modelos tradicionais não conseguiam.



## O Poder dos Embeddings

Os embeddings representam usuários e itens como vetores em um espaço de alta dimensão, permitindo que os sistemas de recomendação aprendam preferências sutis e contextualizadas.

Isso significa que a relevância não é mais uma medida estática, mas uma relação dinâmica influenciada por:

- Humor do usuário
- Hora do dia
- Eventos externos
- Contexto situacional

## Métricas Adaptativas: O Futuro da Avaliação



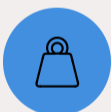
### Relevância Contextual

Métricas que consideram o contexto temporal e situacional do usuário



### Janelas de Tempo

NDCG calculado em períodos específicos para capturar mudanças de comportamento



### Pesos Contextuais

Ajuste dinâmico de pesos baseado em fatores externos



### Diversidade e Novidade

Avaliação de capacidade de surpreender e expandir horizontes

## Complexidade Crescente

A capacidade dos modelos de Deep Learning de gerar recomendações mais personalizadas e diversas exige que as métricas também avaliem esses aspectos. A avaliação se torna um processo contínuo e multifacetado, espelhando a complexidade dos modelos que estão sendo avaliados.

**Tendência-chave:** A avaliação deixa de ser um snapshot único e se torna um monitoramento contínuo que se adapta ao contexto, ao comportamento do usuário e às mudanças no ambiente. As métricas tradicionais como NDCG são complementadas por métricas que capturam nuances contextuais e temporais.

# MLOps e Recommendation as a Service (RaaS): Métricas em Escala

A operacionalização de modelos de Machine Learning, ou **MLOps**, tornou-se um pilar fundamental para a entrega de sistemas de recomendação robustos e escaláveis. Em 2025, a integração de métricas de avaliação no pipeline de MLOps é mais crítica do que nunca. Não basta apenas calcular o NDCG uma vez; é preciso monitorá-lo continuamente em produção para garantir que o sistema mantenha sua performance e se adapte às mudanças nas preferências dos usuários e no catálogo de itens.



## Recommendation as a Service (RaaS)

### AWS

#### Personalize

Plataforma gerenciada para sistemas de recomendação com monitoramento integrado

### Google Cloud

#### Recommendations AI

Soluções prontas com ferramentas de A/B testing nativas

### Azure

#### Personalizer

Serviço de personalização com métricas em tempo real

A ascensão do **Recommendation as a Service (RaaS)**, onde plataformas de nuvem oferecem soluções prontas para sistemas de recomendação, simplifica a infraestrutura, mas intensifica a necessidade de monitoramento de métricas. Essas plataformas fornecem ferramentas para **testes A/B** e **testes multivariados**, permitindo que as equipes experimentem diferentes algoritmos e avaliem seu impacto em métricas como NDCG, cliques e conversões em tempo real.

01

#### Cálculo Automático

Métricas calculadas e reportadas automaticamente em produção

03

#### Identificação Rápida

Equipes identificam problemas em tempo real

02

#### Alertas Inteligentes

Sistema aciona alertas quando performance cai abaixo do limiar

04

#### Correção Ágil

Ajustes rápidos garantem valor contínuo ao negócio

### Impacto em Escala

A capacidade de escalar modelos e métricas de avaliação é essencial para empresas que operam em grande volume, onde cada ponto percentual de melhoria no NDCG pode se traduzir em **milhões de dólares** em receita ou engajamento. A MLOps não é apenas sobre implantar modelos, mas sobre garantir que eles funcionem de forma otimizada e sejam continuamente avaliados em escala.

# 24/7

## Monitoramento

Contínuo de métricas em produção

# <5min

## Tempo de Resposta

Para identificação de anomalias

# 99.9%

## Disponibilidade

Garantida pelas plataformas RaaS

# Ética e Responsabilidade (Responsible AI): O Viés nas Métricas

À medida que os sistemas de recomendação se tornam mais poderosos e onipresentes, a preocupação com a **ética e responsabilidade (Responsible AI)** cresce exponencialmente. Em 2025, a avaliação de métricas não pode mais ignorar questões como **viés (bias)** e **justiça (fairness)**. Um sistema pode ter um NDCG excelente, mas ainda assim ser injusto, por exemplo, ao recomendar predominantemente itens para um grupo demográfico em detrimento de outro, ou ao perpetuar estereótipos.



## Tipos de Viés em Sistemas de Recomendação

	<b>Viés de Popularidade</b> Recomendar apenas itens populares, ignorando o "long tail" e dificultando a descoberta de novos criadores ou produtos. <b>Impacto:</b> Criadores menores nunca ganham visibilidade
	<b>Viés de Gênero/Raça</b> Recomendações que reforçam estereótipos ou excluem grupos minoritários. <b>Impacto:</b> Perpetuação de desigualdades sociais
	<b>Viés de Feedback</b> O sistema aprende com o que os usuários clicam, mas se certos itens nunca são mostrados, eles nunca receberão feedback positivo, criando um ciclo vicioso. <b>Impacto:</b> Bolha de filtro auto-reforçada

## Métricas de Justiça e Responsabilidade

Para combater o viés, as métricas de avaliação precisam ser complementadas por **métricas de justiça**:

<b>NDCG por Grupo Demográfico</b> Avaliar o NDCG separadamente para diferentes grupos de usuários para garantir que nenhum grupo seja desfavorecido.	<b>Diversidade de Provedores</b> Medir se as recomendações estão distribuindo a visibilidade de forma equitativa entre diferentes fontes de conteúdo.
<b>Novidade e Serendipidade</b> Garantir que o sistema não apenas reforce as preferências existentes, mas também introduza novas opções para todos os usuários.	<b>Cobertura do Catálogo</b> Verificar se diferentes segmentos do catálogo recebem exposição adequada.

**Abordagem Holística**  
A construção de sistemas de recomendação responsáveis exige uma abordagem holística, onde a otimização de métricas de desempenho como o NDCG é balanceada com a garantia de que o sistema seja justo, transparente e benéfico para todos os usuários.

**Risco de Negócio:** Ignorar o viés nas métricas é um risco não apenas ético, mas também de negócio, podendo levar à alienação de usuários e danos à reputação da marca. Em 2025, empresas líderes integram métricas de justiça como parte essencial de seus KPIs de recomendação.



# Consolidação e Próximos Passos

Chegamos ao fim de nossa jornada pelas métricas de ranking e relevância, aprofundando-nos no Ganho Cumulativo Descontado (DCG) e no Ganho Cumulativo Descontado Normalizado (NDCG). Compreendemos que a posição de um item na lista de recomendação é um fator crítico para a experiência do usuário e que essas métricas são essenciais para capturar essa nuance. O DCG nos oferece uma medida do ganho absoluto de relevância descontada, enquanto o NDCG, ao normalizar pelo ideal, permite comparações justas e robustas entre diferentes sistemas e cenários.



## Em Prática

- ❑ A escolha da métrica é uma decisão estratégica que deve alinhar-se aos objetivos de negócio, seja maximizar o engajamento, a conversão ou a satisfação do usuário. Acompanhar as tendências de Deep Learning, MLOps e Responsible AI é fundamental para construir e avaliar sistemas de recomendação que não apenas performem bem, mas também sejam escaláveis, adaptativos e éticos.

## Autoavaliação

1

### Questão 1

Qual a principal limitação do Ganho Cumulativo (CG) que o DCG busca resolver?

- a) Não considera a relevância dos itens.
- b) Não permite a comparação entre diferentes listas.
- c) Não atribui pesos diferentes aos itens com base em sua posição.
- d) É muito complexo de calcular para listas grandes.

2

### Questão 2

Para que serve a normalização no NDCG?

- a) Para simplificar o cálculo do DCG.
- b) Para garantir que o valor do DCG seja sempre 1.
- c) Para permitir a comparação justa do desempenho entre diferentes sistemas ou cenários.
- d) Para remover itens irrelevantes da lista de recomendação.

3

### Questão 3

Um sistema de recomendação obteve um NDCG de 0.95. O que isso significa?

- a) O sistema é perfeito e não pode ser melhorado.
- b) O sistema está performando 95% melhor que o sistema ideal.
- c) O sistema está entregando 95% do ganho descontado que uma lista ideal entregaria.
- d) O sistema tem 95% de precisão em suas recomendações.

4

### Questão 4

Qual das seguintes tendências de 2025 impacta diretamente a necessidade de considerar o viés e a justiça nas métricas de avaliação?

- a) Recommendation as a Service (RaaS).
- b) MLOps.
- c) Deep Learning e Embeddings.
- d) Ética e Responsabilidade (Responsible AI).

5

### Questão 5

Explique como a escolha de uma métrica de avaliação pode ser considerada uma decisão de negócio e não apenas técnica, fornecendo um exemplo.

*(Questão dissertativa)*

## Gabarito:

1. c)

2. c)

3. c)

4. d)

## Próxima Aula

### Aula 14: Metodologias e Protocolos de Avaliação

Exploraremos as Metodologias e Protocolos de Avaliação, aprendendo a desenhar experimentos, realizar testes A/B e interpretar resultados de forma estatisticamente significativa para validar e otimizar seus sistemas de recomendação.

## Recursos Adicionais

### • Artigos de Pesquisa

Para aprofundar nos fundamentos matemáticos e variações de DCG/NDCG.

### • Documentação de Bibliotecas

Surprise, LightFM - Para ver a implementação prática dessas métricas em Python.

### • Cursos Online

Coursera, edX - Para exemplos interativos e estudos de caso sobre avaliação de sistemas de recomendação.

- ❑ **NOTA IMPORTANTE:** As informações regulatórias/legais/técnicas desta aula estão atualizadas até 2025. Consulte sempre fontes oficiais para verificar alterações.