

# Accelerating iteration speed: ranker evaluation with debiased interleaving

MICES 2025

**„Full assortment“ two-sided marketplace**

**GMV 7 billion EUR**

Up to **35** orders per second

**12.2** million active customers

**18** million products



**Dr. Leonie Brinkmann**  
**Senior Digital Analyst - Ranking**

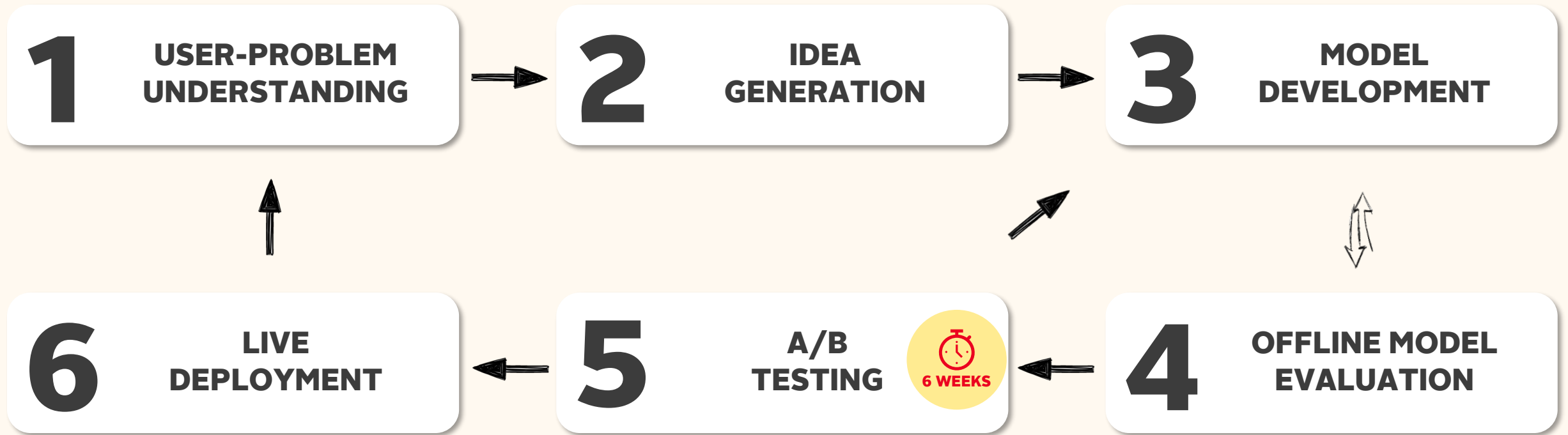
**Mika Wolle**  
**Product Manager - Ranking**

***“Product development is a risky business. If you don’t evolve your product quickly enough, competitors will outrun you.”<sup>1</sup>***



<sup>1</sup> Risk-Aware Product Decisions in A/B Tests with Multiple Metrics, Spotify R&D Blog

# A/B test runtime is a problem for our iteration speed



# In the literature we found a promising approach

## Debiased Balanced Interleaving at Amazon Search

Nan Bi

Amazon, Palo Alto, CA, USA

Pablo Castells

Amazon, Madrid, Spain

Daniel Gilbert

Amazon, Palo Alto, CA, USA

## Beyond A/B Test : Speeding up Airbnb Search Ranking Experimentation through Interleaving



Qing Zhang · Follow

Published in The Airbnb Tech Blog · 10 min read · Oct 6, 2022

## Innovating Faster on Personalization Algorithms at Netflix Using Interleaving

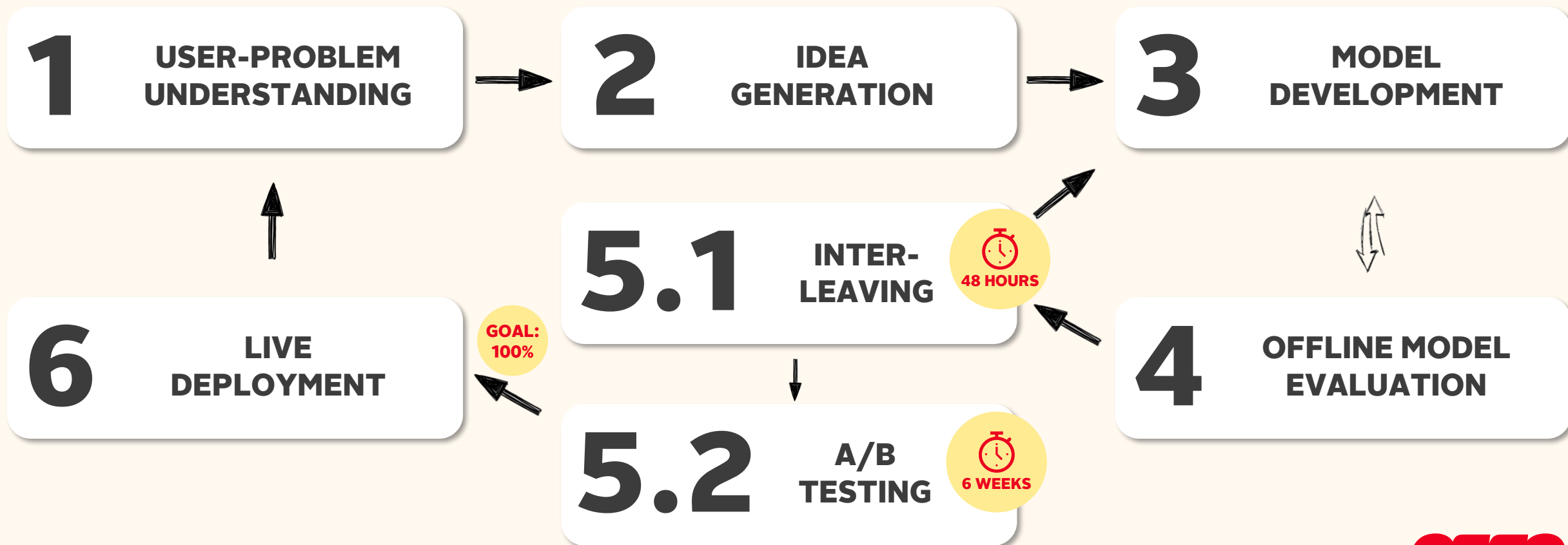


Netflix Technology Blog · Follow

Published in Netflix TechBlog · 8 min read · Nov 29, 2017

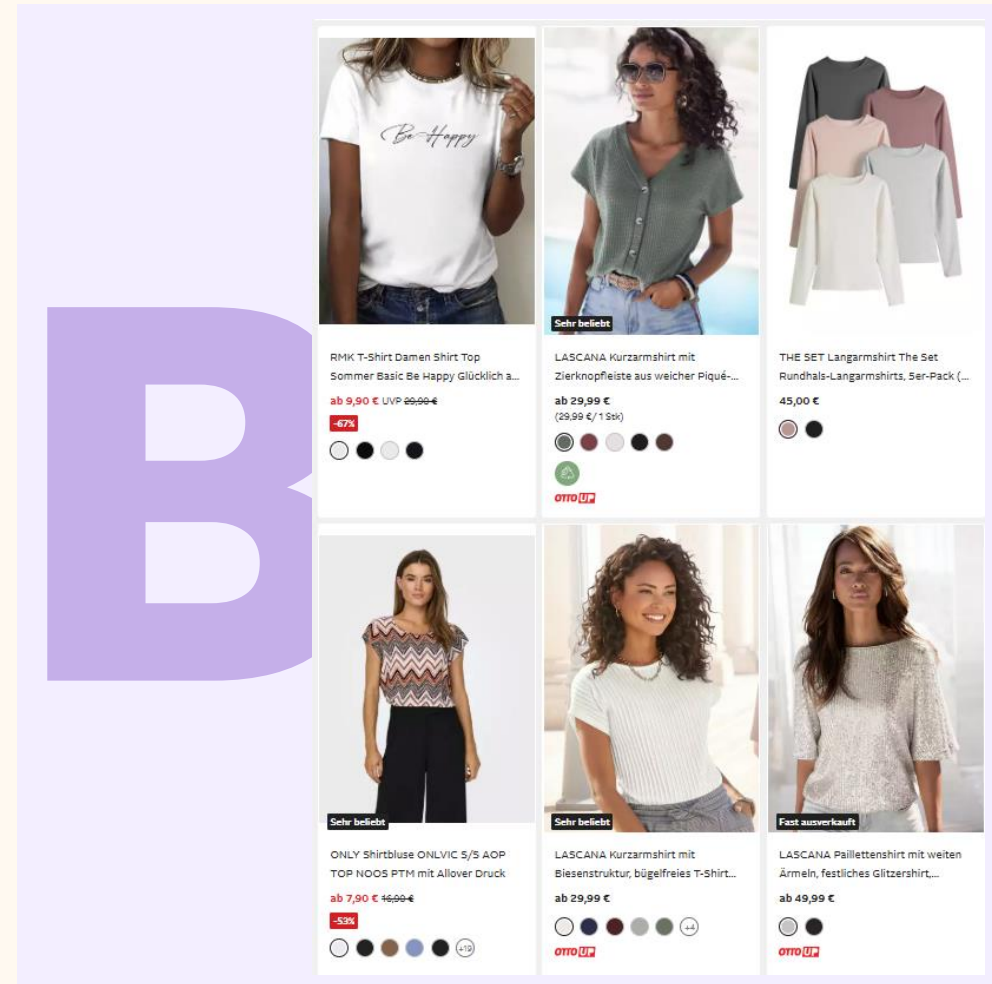
**OTTO**

# Interleaving enables us to collect meaningful feedback within 48 hours





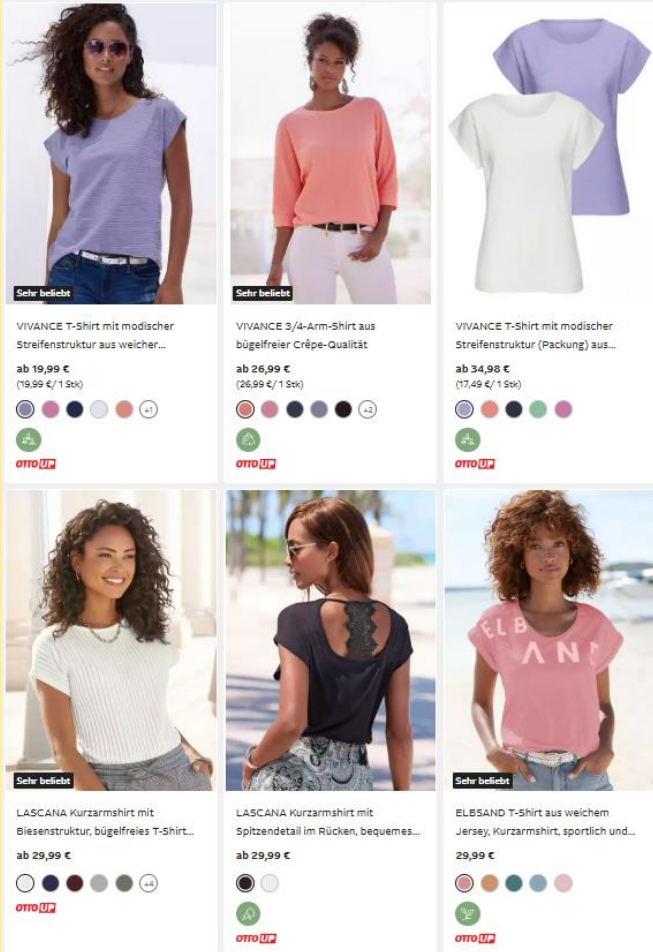
# In a classic A/B test users either see ranking A or B



**OTTO**

# With interleaving both rankers are presented to the same user simultaneously

**A**



Sehr beliebt

VIVANCE T-Shirt mit modischer Streifenstruktur aus weicher...

ab 19,99 €  
(19,99 €/1 Stk)

Sehr beliebt

VIVANCE 3/4-Arm-Shirt aus bügelfreier Crêpe-Qualität

ab 26,99 €  
(26,99 €/1 Stk)

Sehr beliebt

VIVANCE T-Shirt mit modischer Streifenstruktur (Packung) aus...

ab 34,98 €  
(17,49 €/1 Stk)

Sehr beliebt

LASCANA Kurzarmshirt mit Biesenstruktur, bügelfreies T-Shirt...

ab 29,99 €

Sehr beliebt

LASCANA Kurzarmshirt mit Spitzendetail im Rücken, bequemes...

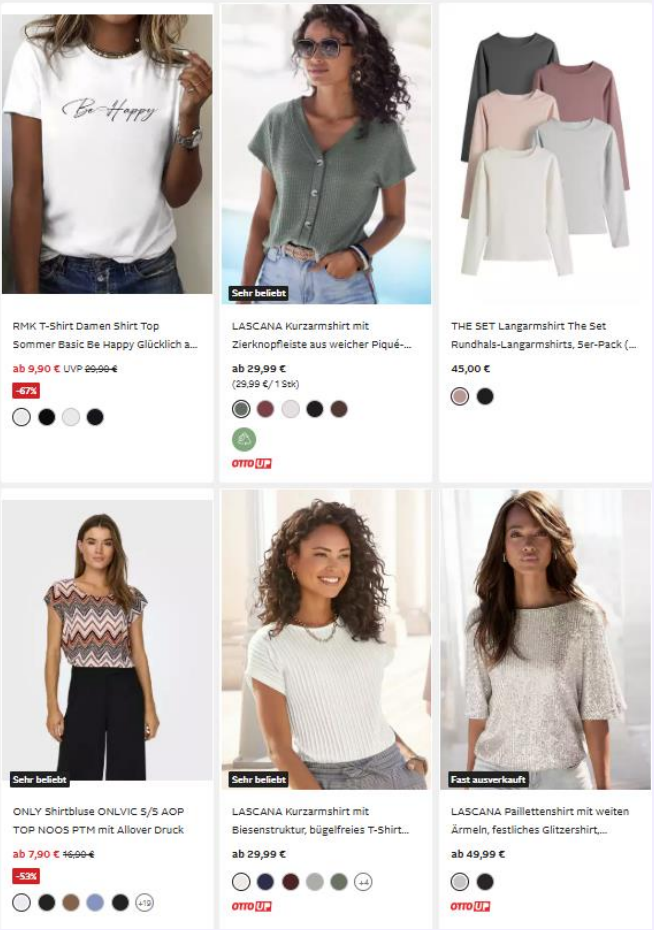
ab 29,99 €

Sehr beliebt

ELBSAND T-Shirt aus weichem Jersey, Kurzarmshirt, sportlich und...

29,99 €

**B**



Sehr beliebt

RMK T-Shirt Damen Shirt Top Sommer Basic Be Happy Glücklic a...

ab 9,90 € UVP 20,00 €  
-67%

Sehr beliebt

LASCANA Kurzarmshirt mit Zierknopfleiste aus weicher Piqué...

ab 29,99 €  
(29,99 €/1 Stk)

THE SET Langarmshirt The Set Rundhals-Langarmshirts, Ser-Pack (...)

45,00 €

Sehr beliebt

ONLY Shirtbluse ONLYVIC S/S AOP TOP NOOS PTM mit Allover Druck

ab 7,90 € 16,90 €  
-53%

Sehr beliebt

LASCANA Kurzarmshirt mit Biesenstruktur, bügelfreies T-Shirt...

ab 29,99 €

Fast ausverkauft

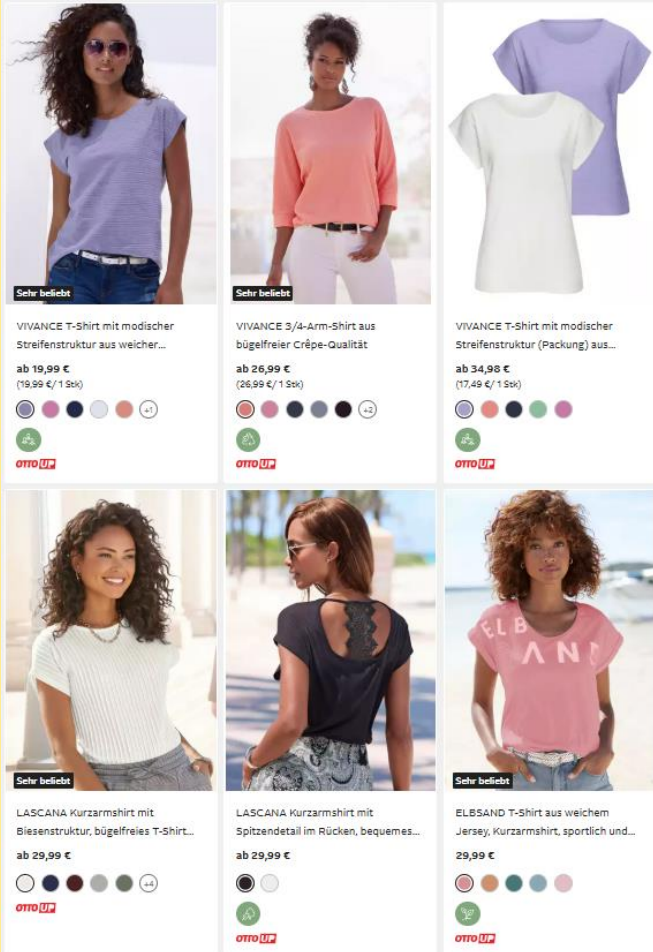
LASCANA Paillettenshirt mit weiten Ärmeln, festliches Glitzershirt...

ab 49,99 €



# With interleaving both rankers are presented to the same user simultaneously

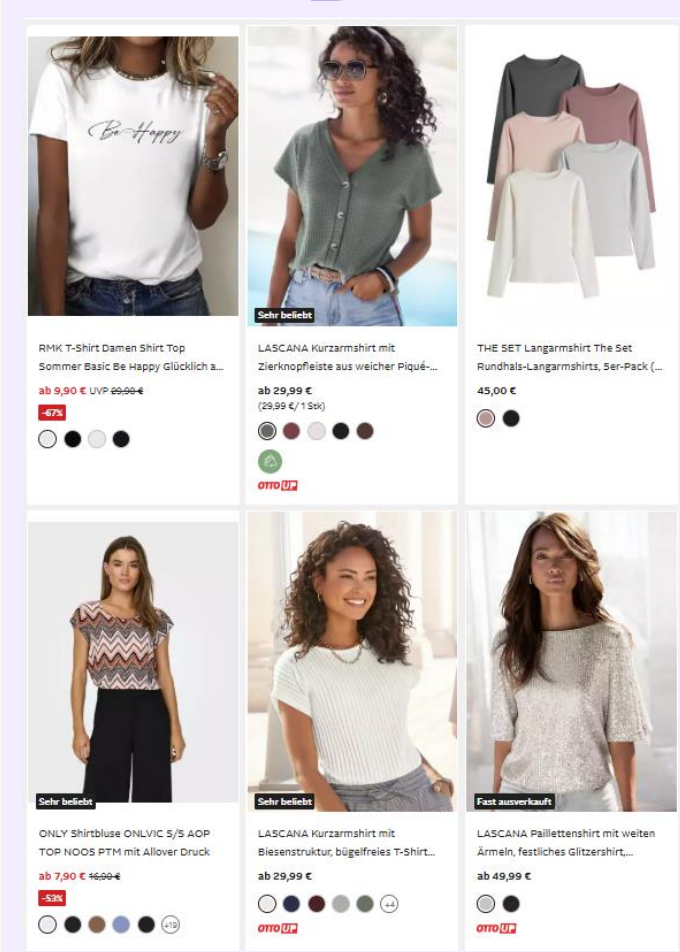
**A**



Grid A displays six clothing items in a 2x3 layout. Each item includes a model photo, a 'Sehr beliebt' (Very popular) badge, a description, a price, and color selection options.

- Item 1:** VIVANCE T-Shirt mit modischer Streifenstruktur aus weicher...  
ab 19,99 € (19,99 €/1 Stk)  
Colors: 4 options
- Item 2:** VIVANCE 3/4-Arm-Shirt aus bügelfreier Crêpe-Qualität  
ab 26,99 € (26,99 €/1 Stk)  
Colors: 4 options
- Item 3:** VIVANCE T-Shirt mit modischer Streifenstruktur (Packung) aus...  
ab 34,98 € (17,49 €/1 Stk)  
Colors: 4 options
- Item 4:** LASCANA Kurzarmshirt mit Biesenstruktur, bügelfreies T-Shirt...  
ab 29,99 €  
Colors: 4 options
- Item 5:** LASCANA Kurzarmshirt mit Spitzendetail im Rücken, bequemes...  
ab 29,99 €  
Colors: 2 options
- Item 6:** ELBSAND T-Shirt aus weichem Jersey, Kurzarmshirt, sportlich und...  
29,99 €  
Colors: 4 options

**B**



Grid B displays six clothing items in a 2x3 layout. Each item includes a model photo, a 'Sehr beliebt' (Very popular) badge, a description, a price, and color selection options.

- Item 1:** RMK T-Shirt Damen Shirt Top Sommer Basic Be Happy Glücklic...  
ab 9,90 € UVP 20,00 €  
-67%  
Colors: 4 options
- Item 2:** LASCANA Kurzarmshirt mit Zierknopfleiste aus weicher Piqué...  
ab 29,99 € (29,99 €/1 Stk)  
Colors: 4 options
- Item 3:** THE SET Langarmshirt The Set Rundhals-Langarmshirts, Ser-Pack (...)  
45,00 €  
Colors: 2 options
- Item 4:** ONLY Shirtbluse ONLYVIC S/S AOP TOP NOOS PTM mit Alllover Druck  
ab 7,90 € 16,90 €  
-53%  
Colors: 4 options
- Item 5:** LASCANA Kurzarmshirt mit Biesenstruktur, bügelfreies T-Shirt...  
ab 29,99 €  
Colors: 4 options
- Item 6:** LASCANA Pailletenshirt mit weiten Ärmeln, festliches Glitzershirt...  
ab 49,99 €  
Colors: 2 options

# With interleaving both rankers are presented to the same user simultaneously

## A

Grid A displays 6 items in a 2x3 layout on a yellow background. Each item includes a model photo, a 'Sehr beliebt' badge, a description, a price, and color selection options.

- VIVANCE T-Shirt mit modischer Streifenstruktur aus weicher...**  
ab 19,99 € (19,99 €/1 Stk)
- VIVANCE 3/4-Arm-Shirt aus bügelfreier Crêpe-Qualität**  
ab 26,99 € (26,99 €/1 Stk)
- VIVANCE T-Shirt mit modischer Streifenstruktur (Packung) aus...**  
ab 34,98 € (17,49 €/1 Stk)
- LASCANA Kurzarmshirt mit Biesenstruktur, bügelfreies T-Shirt...**  
ab 29,99 €
- LASCANA Kurzarmshirt mit Spitzendetail im Rücken, bequemes...**  
ab 29,99 €
- ELBSAND T-Shirt aus weichem Jersey, Kurzarmshirt, sportlich und...**  
29,99 €

## INTERLEAVED LIST

Grid B displays 6 items in a 2x3 layout on a light purple background. Each item includes a model photo, a 'Sehr beliebt' badge, a description, a price, and color selection options.

- VIVANCE T-Shirt mit modischer Streifenstruktur aus weicher...**  
ab 19,99 € (19,99 €/1 Stk)
- RMK T-Shirt Damen Shirt Top Sommer Basic Be Happy Glücklic a...**  
ab 9,90 € UVP 20,00 € -67%
- VIVANCE 3/4-Arm-Shirt aus bügelfreier Crêpe-Qualität**  
ab 26,99 € (26,99 €/1 Stk)
- LASCANA Kurzarmshirt mit Zierknopfleiste aus weicher Piqué...**  
ab 34,98 € (17,49 €/1 Stk)
- VIVANCE T-Shirt mit modischer Streifenstruktur (Packung) aus...**  
ab 34,98 € (17,49 €/1 Stk)
- THE SET Langarmshirt The Set Rundhals-Langarmshirts, Ser-Pack (...)**  
45,00 €

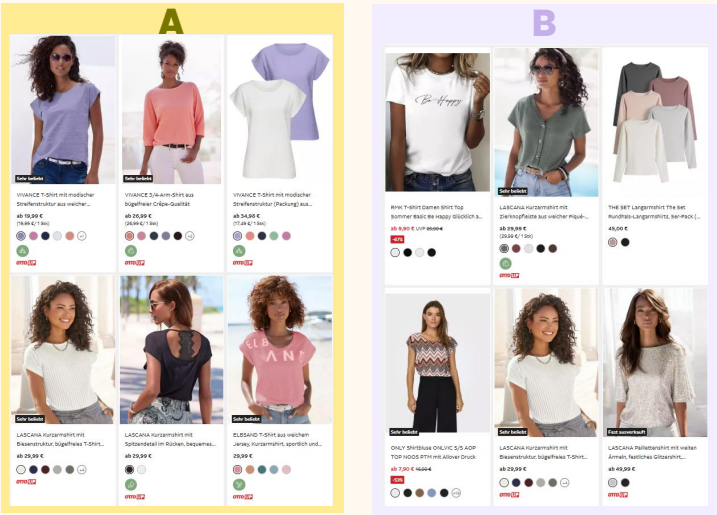
## B

Grid C displays 6 items in a 2x3 layout on a light purple background. Each item includes a model photo, a 'Sehr beliebt' badge, a description, a price, and color selection options.

- RMK T-Shirt Damen Shirt Top Sommer Basic Be Happy Glücklic a...**  
ab 9,90 € UVP 20,00 € -67%
- LASCANA Kurzarmshirt mit Zierknopfleiste aus weicher Piqué...**  
ab 29,99 € (29,99 €/1 Stk)
- THE SET Langarmshirt The Set Rundhals-Langarmshirts, Ser-Pack (...)**  
45,00 €
- ONLY Shirtbluse ONLYVIC S/S AOP TOP NOOS PTM mit Allover Druck**  
ab 7,90 € 16,90 € -53%
- LASCANA Kurzarmshirt mit Biesenstruktur, bügelfreies T-Shirt...**  
ab 29,99 €
- LASCANA Pailletenshirt mit weiten Ärmeln, festliches Glitzershirt...**  
ab 49,99 €

# Interleaving is much more sensitive in detecting effects

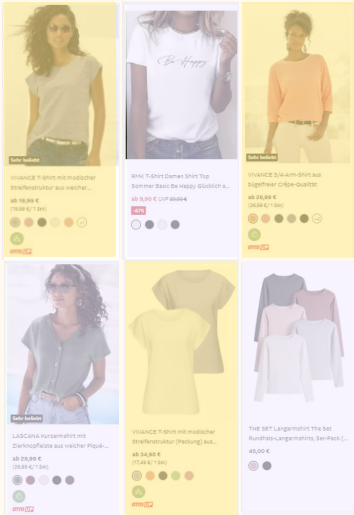
## BETWEEN-SUBJECT DESIGN



100% TRAFFIC



## WITHIN-SUBJECT DESIGN



10% TRAFFIC



OTTO

# Products are placed on the list following a specific logic

A



B



RESULTING LIST



**OTTO**



# Products are placed on the list following a specific logic

A



B



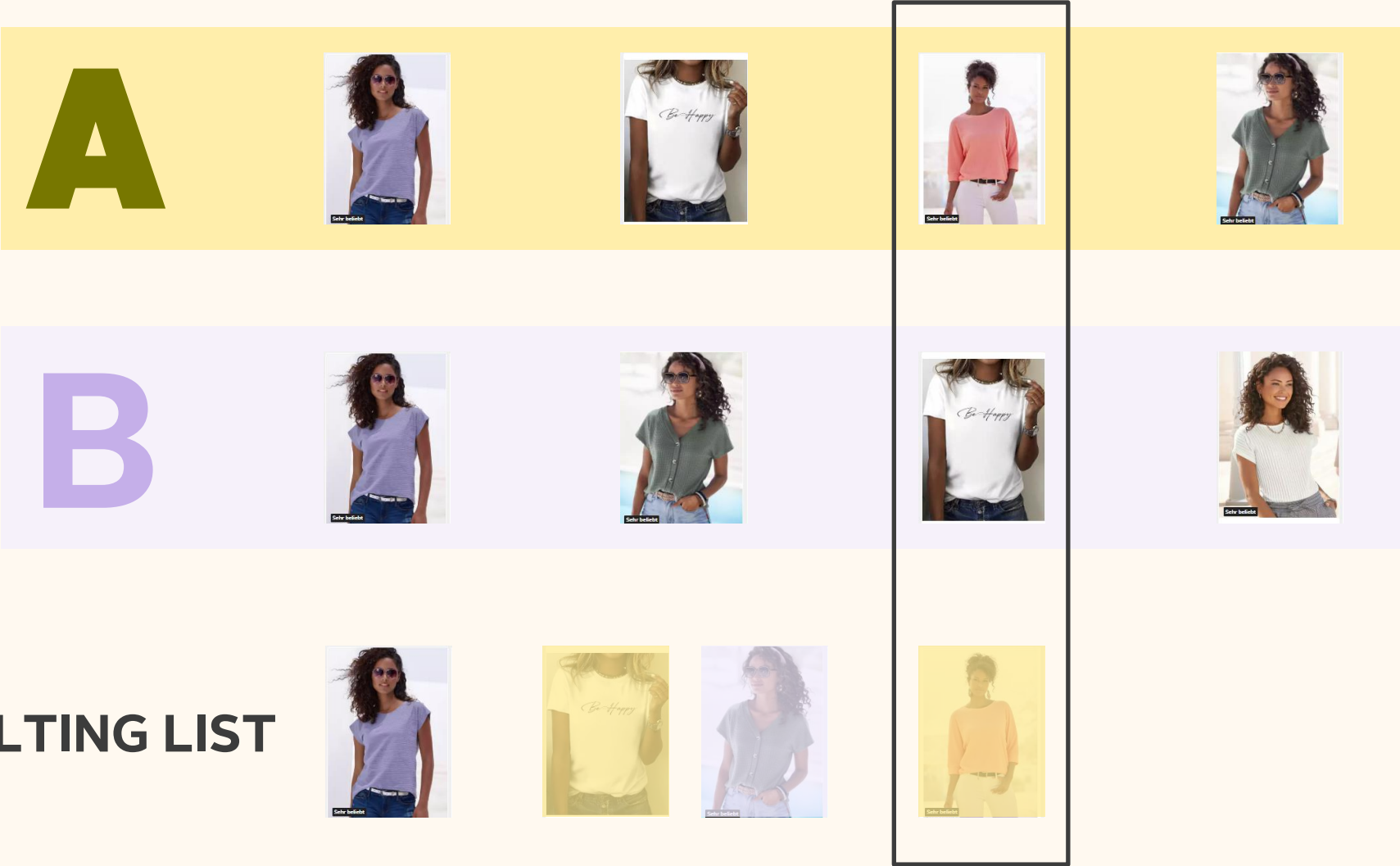
RESULTING LIST



**OTTO**



# Products are placed on the list following a specific logic



# Products are placed on the list following a specific logic

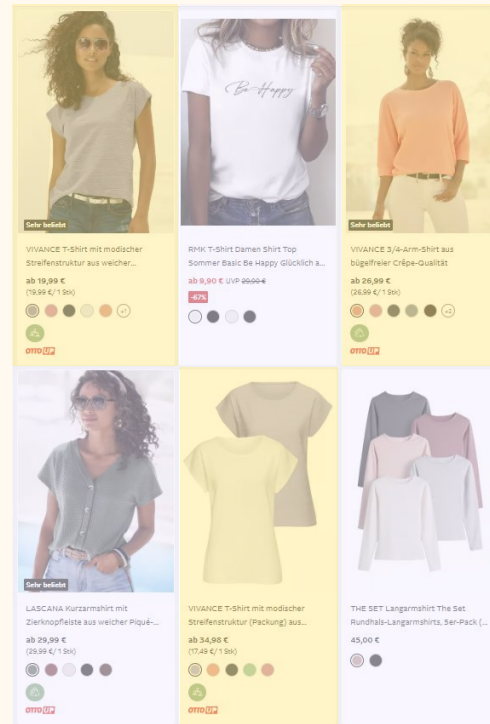


RESULTING LIST

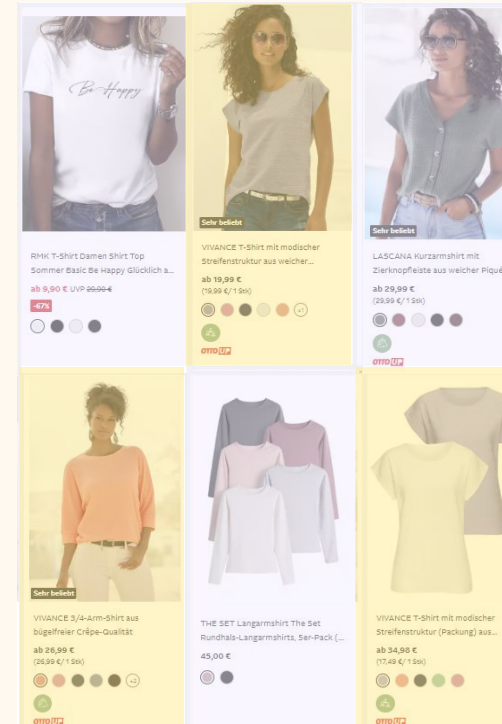
OTTO

# Bias #1: The ranker that starts has an advantage

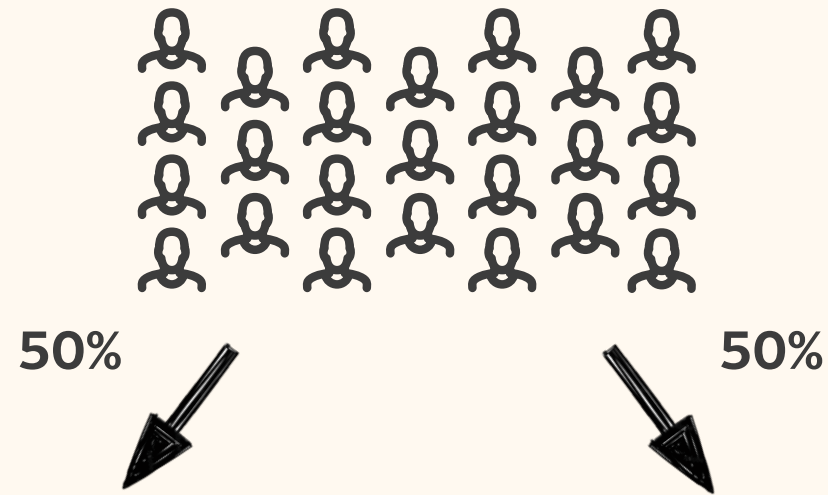
## RANKER A STARTING



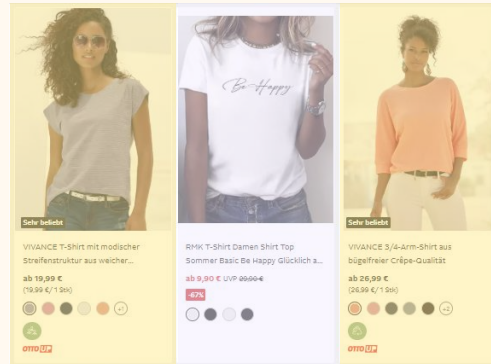
## RANKER B STARTING



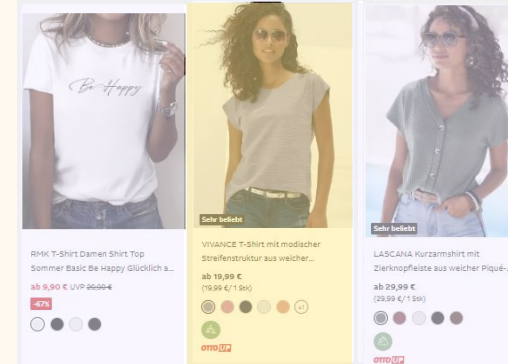
# We distribute the advantage evenly among rankers



RANKER A STARTING



RANKER B STARTING



# Bias #2: Ranker similarity can lead to an imbalanced list

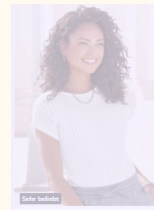
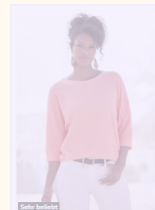
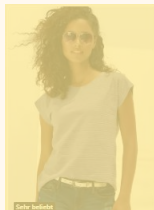
A



B



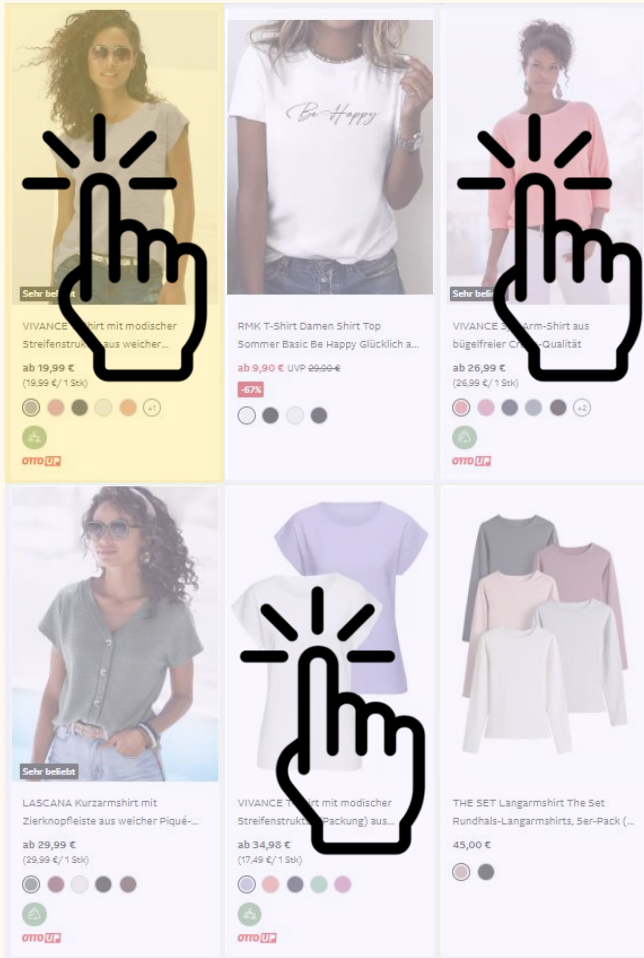
RESULTING LIST



**OTTO**



# In the evaluation of ranker preference we account for imbalanced lists



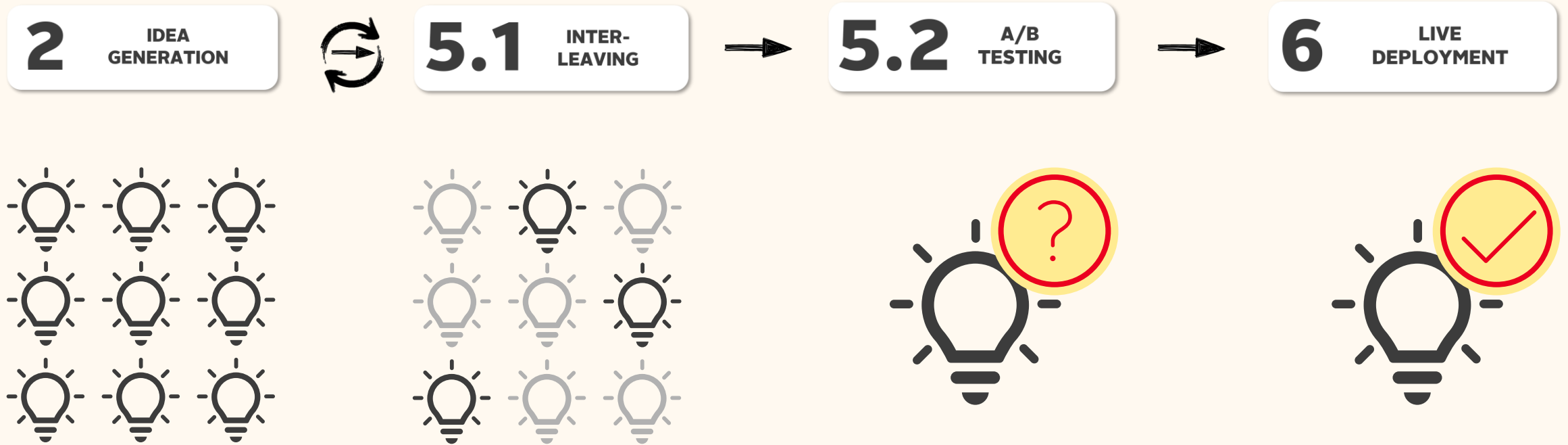
**CREDIT = WINS \* VISIBILITY RATIO**

$$\text{CREDIT RANKER A} = 1 * \frac{6}{1} = 6$$

$$\text{CREDIT RANKER B} = 2 * \frac{6}{5} = 2,4$$

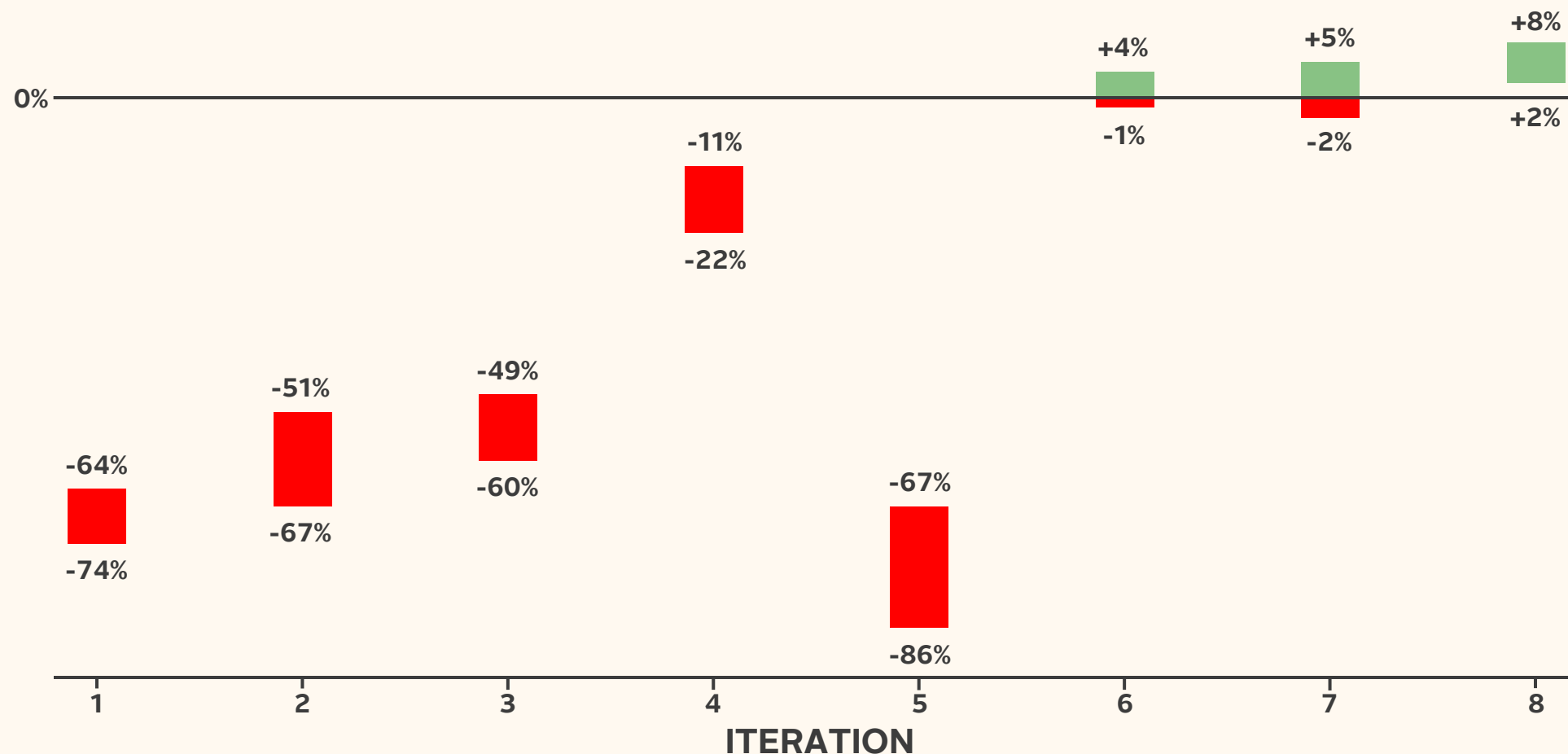
**OTTO**

# With a two phase procedure we speed up our iteration cycles



# Within 8 weeks we developed a successful deep learning ranking model

95% CONFIDENCE INTERVALS FOR PERCENTAGE DIFFERENCE BETWEEN GROUPS IN SOLD UNITS



**OTTO**

# We still see potential for improvements

## CORRELATION OF INTERLEAVING AND A/B TEST RESULTS

ADD  
FOR  
LATER

**0.68**

UNITS

**0.47**

ORDERS

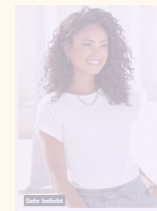
**0.60**

**OTTO**

# Hypothesis: using position differences will improve the correlation



INTERLEAVED LIST



**OTTO**

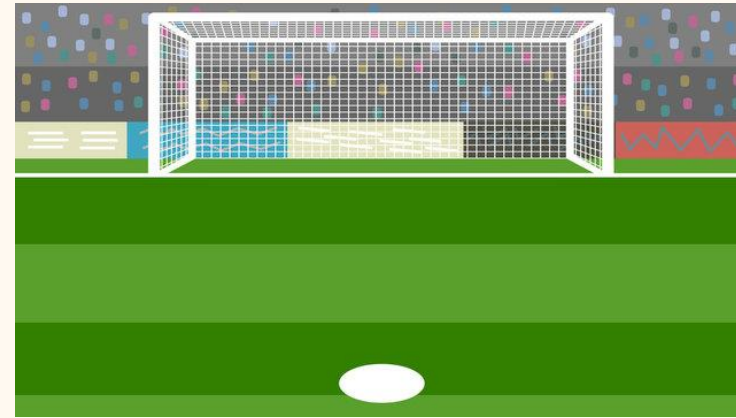


# Interleaving relates to A/B testing as penalty shootout relates to a football match

**A/B TEST**



**INTERLEAVING**

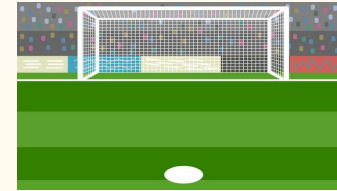


# Interleaving relates to A/B testing as penalty shootout relates to a football match

**A/B TEST**



**INTERLEAVING**



**The player by itself decides the outcome**



**The score doesn't represent the „classic“ outcome**



**More efficient to get results**

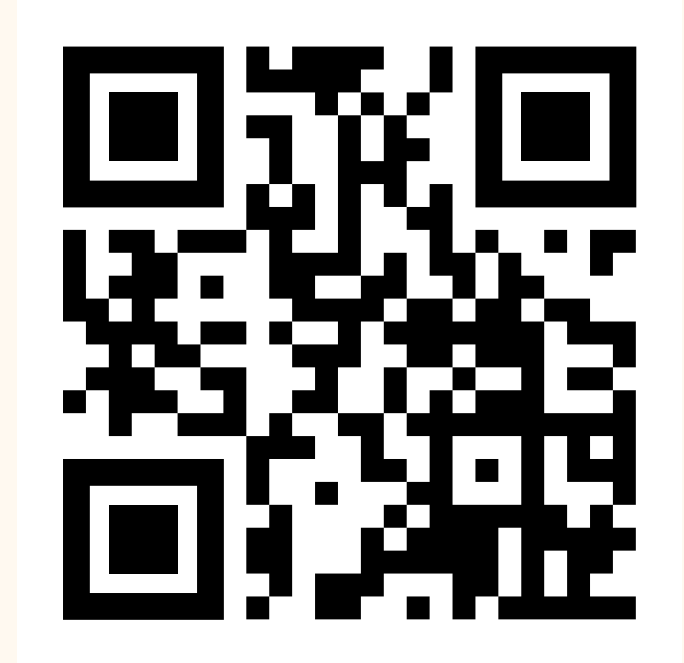


**Classic match is still closer to reality**

P  
R  
O  
B  
E  
L  
M  
S

S  
U  
M  
M  
A  
R  
Y

# Big shoutout to Andreas Wagenmann



**Freelance Senior Data Scientist  
& ML / Data / Search Engineer**

**OTTO**

anke

***OTTO***