

Scaling AI Entertainment in the Car.

Making generative content deliverable,
trustworthy and relevant.

PRESENTED BY

Artem Melnytskyi



A MORE HUMAN JOURNEY
THROUGH AI

DELIVERY

REAL-TIME
AT SCALE
IN THE CAR

GENERATIVE
INTELLIGENCE
FOR A MORE
HUMAN JOURNEY

QUALITY

SAFE
TRUSTWORTHY
HIGH VALUE

RELEVANCE

PERSONAL
CONTEXT-AWARE
MEANINGFUL

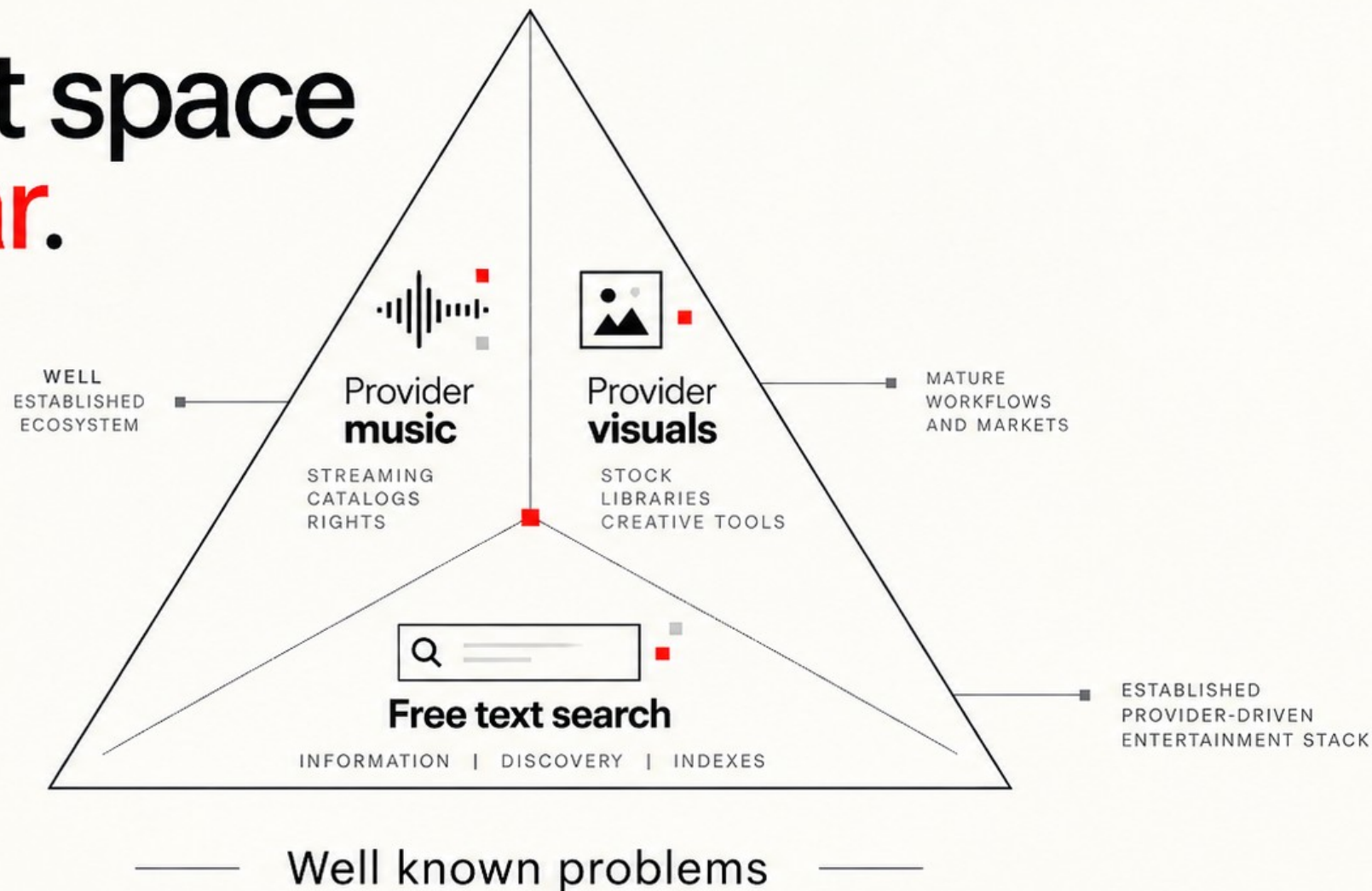


“When you invent the ship,
you also invent the shipwreck.”

— Paul Virilio, French philosopher

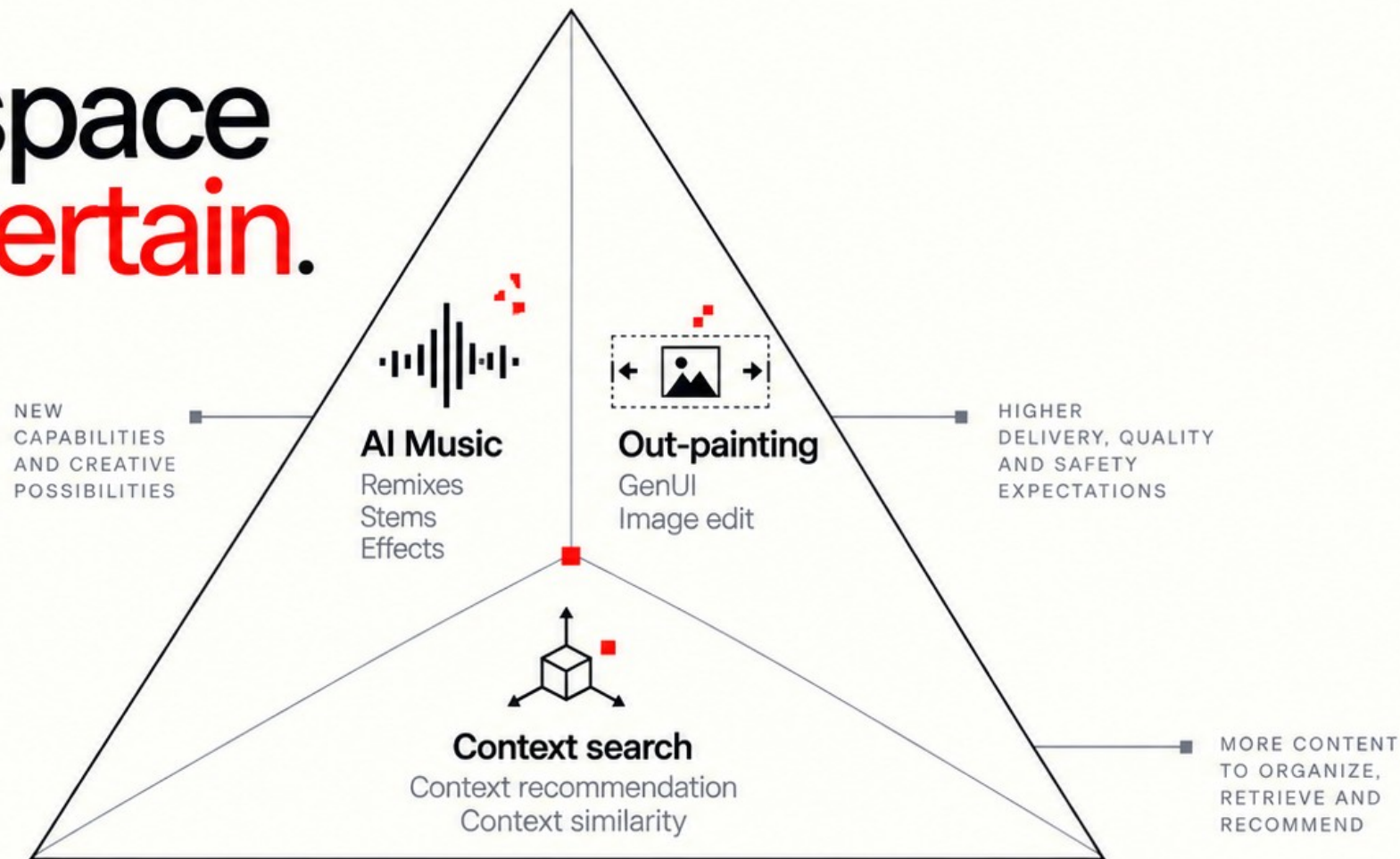
Before AI, the content space was familiar.

A provider-driven world of music, visuals, and text search – with well understood challenges.



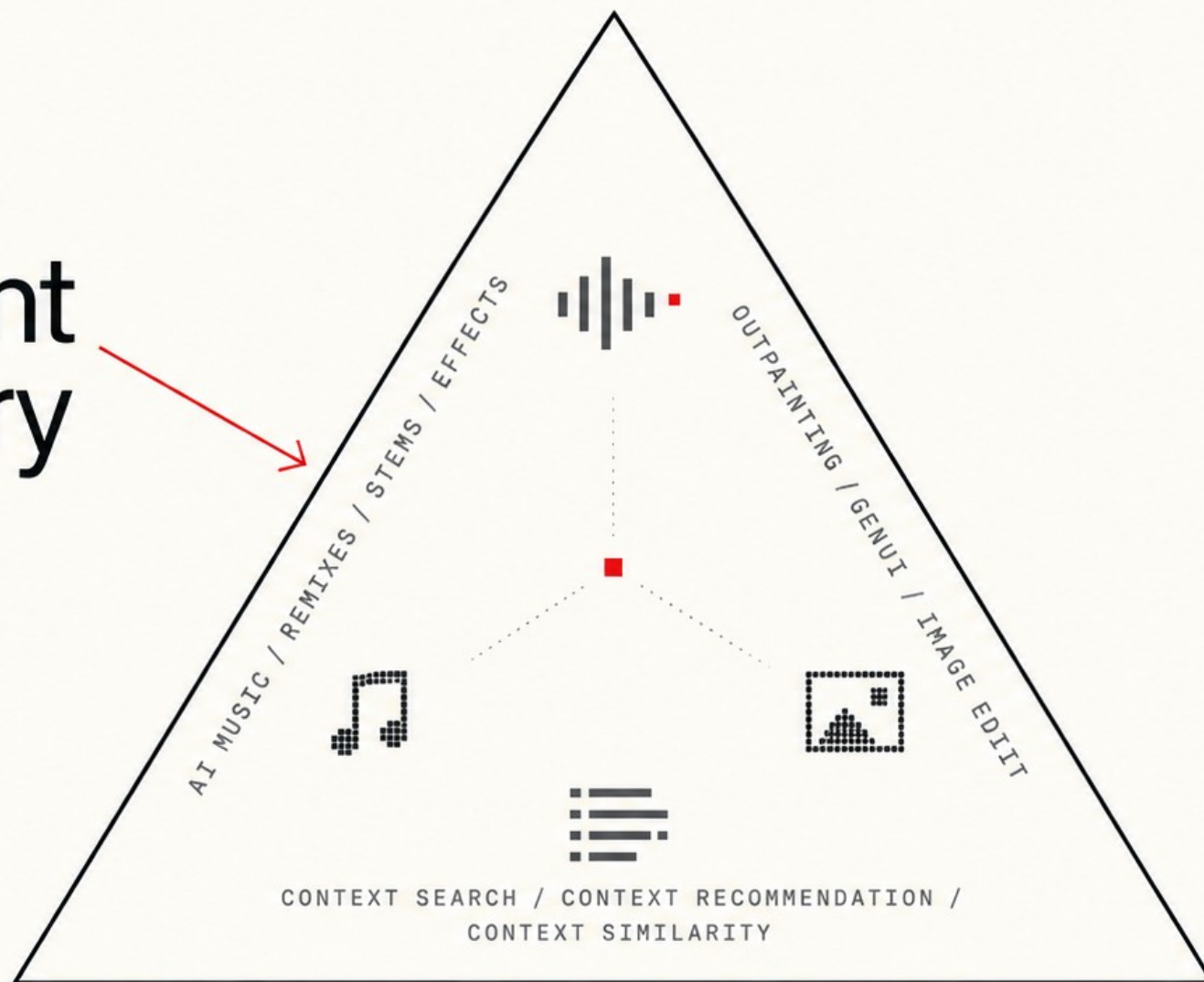
With AI, the content space became uncertain.

New entertainment features expand the experience space — but they also introduce delivery, quality and retrieval challenges we did not have to solve before.



— Unknown problem area —

Content Delivery



Regular entertainment package.

1 track and 1 cover art.

AUDIO TRACK



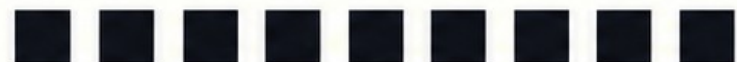
4 MB

COVER ART



0.5 MB

4.5 MB total



■ 1 block = 0.5 MB

AI-boosted data unit.

More experience. More data to deliver.

AI TRACK (STEMS)



Drums	4 MB
Bass	4 MB
Vocal	4 MB
Other	4 MB

+ REMIX



2 versions

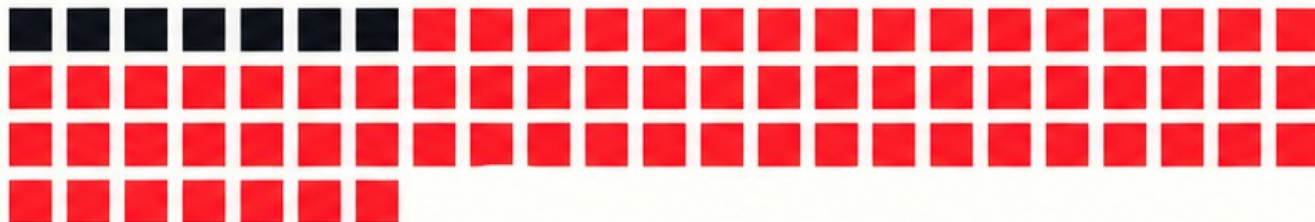
4 MB × 4 stems
× 2 versions
= **32 MB**



OUT-PAINTING COVER ART

≈ 6 MB

38 MB total



76 blocks (38 MB)

AUDIO

SpectroStream.

Multi-channel neural audio codec designed for high-quality 48 kHz stereo music.



ON-DEVICE PERFORMANCE

F25 FP32
Exynos 2400
2 threads
RTF 0.79

RESULT

**No audible
quality loss observed.**

SIZE (PER TRACK)

1 track:
4 MB → 300 KB

BEFORE COMPRESSION

38 MB total



■ 1 block = 0.5 MB

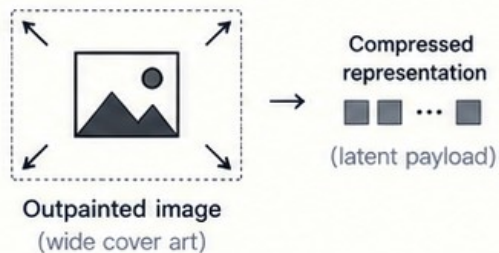
76 blocks (38 MB)

VISUAL

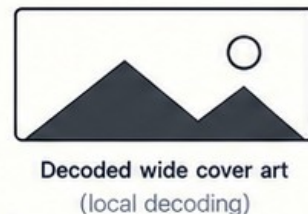
Cool-Chic.

Server builds a compact representation and sends the thin decoder plus learned weights to the vehicle.

SERVER-SIDE ENCODING



IN-VEHICLE DECODING



PAYLOAD SIZE

Wide cover art:
~6 MB → ~30 KB

SENT TO VEHICLE

Decoder + learned weights
(lightweight module)

WITH NEURAL DELIVERY

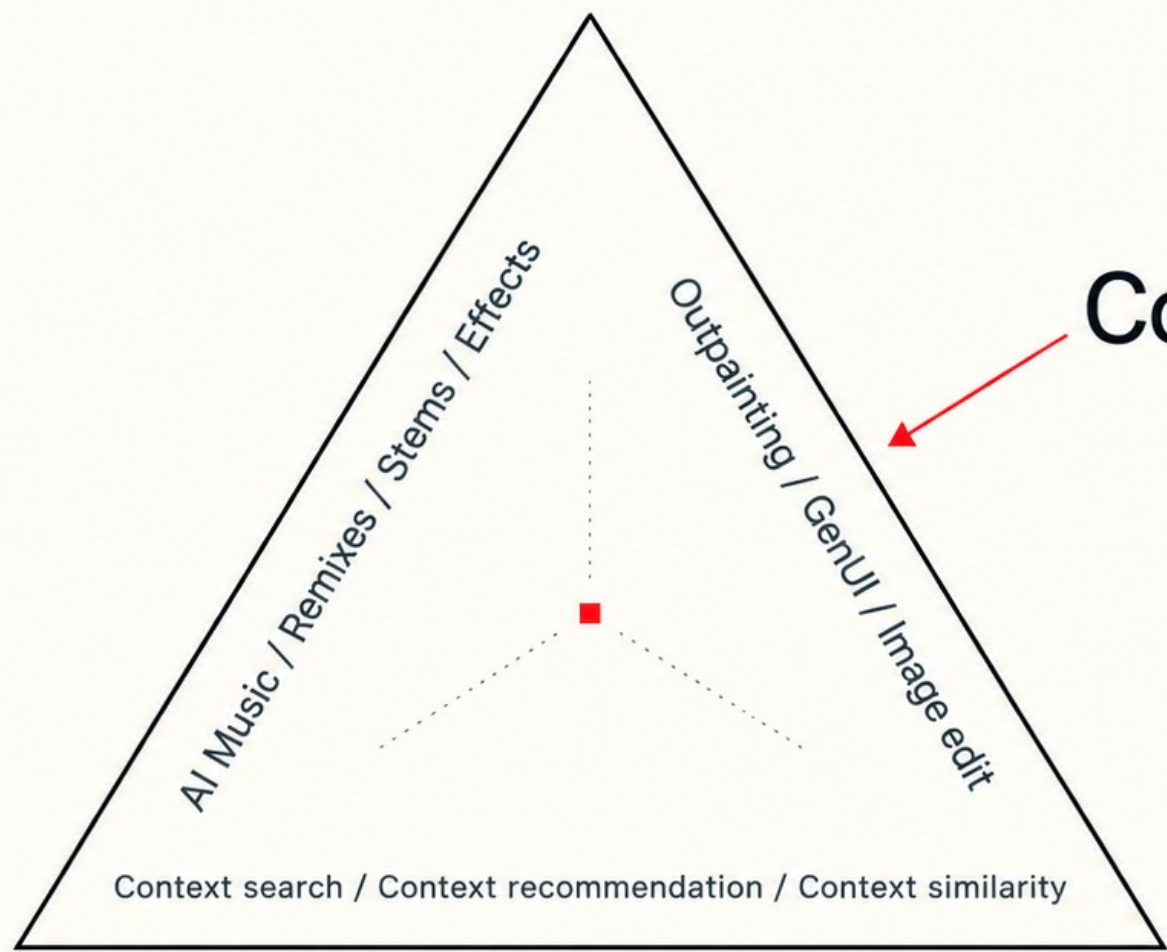
~2.4 MB total



$300 \text{ KB} \times 4 \text{ stems} \times 2 \text{ versions} = 2.4 \text{ MB}$
+ ~30 KB image payload

**From 38 MB
to ~2.4 MB.**

Compression is solved.
Real-time performance still needs work.



Content Quality

Generation is probabilistic. Production cannot be.

- No provider guarantees output quality.
- There is no bulletproof prompt.
- There is no universally safe seed.

SAME INPUT.
DIFFERENT OUTCOMES.
REAL-WORLD IMPACT.

PROMPTS
SEEDS
REALITY

01

SEED A



WIDER OUTPAINTING RESULT

02

SEED B



WIDER OUTPAINTING RESULT

03

SEED C



WIDER OUTPAINTING RESULT

Same API. Different seeds.

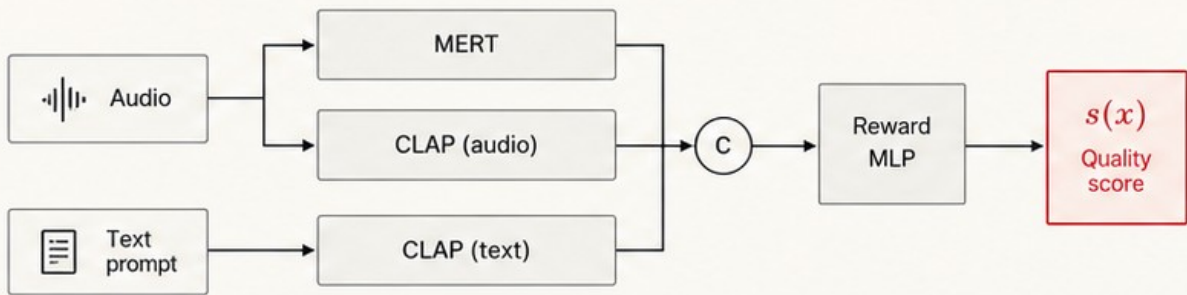
MORE
POSSIBILITIES
MORE VARIANCE

AI GENERATION
IN THE REAL WORLD

AUDIO

TuneJury.

How good is a track relative to its description?



Three ways to use it.

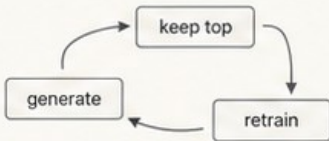
01
Best-of-N selection
Generate N candidates, keep the highest-scoring one.



02
Latent optimization
Backpropagate the score through the sample into noise. The backbone stays frozen.



03
Expert iteration
Fine-tune on top-scoring outputs, balancing reward and distribution fidelity.



■ Core idea: score alignment between audio and prompt.

RESULTS

Prompt changes what “best” means.

Prompt: arena rock power ballad

RANK	TRACK	SCORE
1	Queen — We Are The Champions	1.4120
2	Instrumental cover — We Are The Champions	0.7439
3	Adam Lambert — We Are The Champions	0.3707

Prompt: instrumental violin

RANK	TRACK	SCORE
1	Instrumental cover — We Are The Champions	1.9847
2	Queen — We Are The Champions	1.3709
3	Adam Lambert — We Are The Champions	-0.0688



Queen
WE ARE THE CHAMPIONS



Instrumental cover
WE ARE THE CHAMPIONS

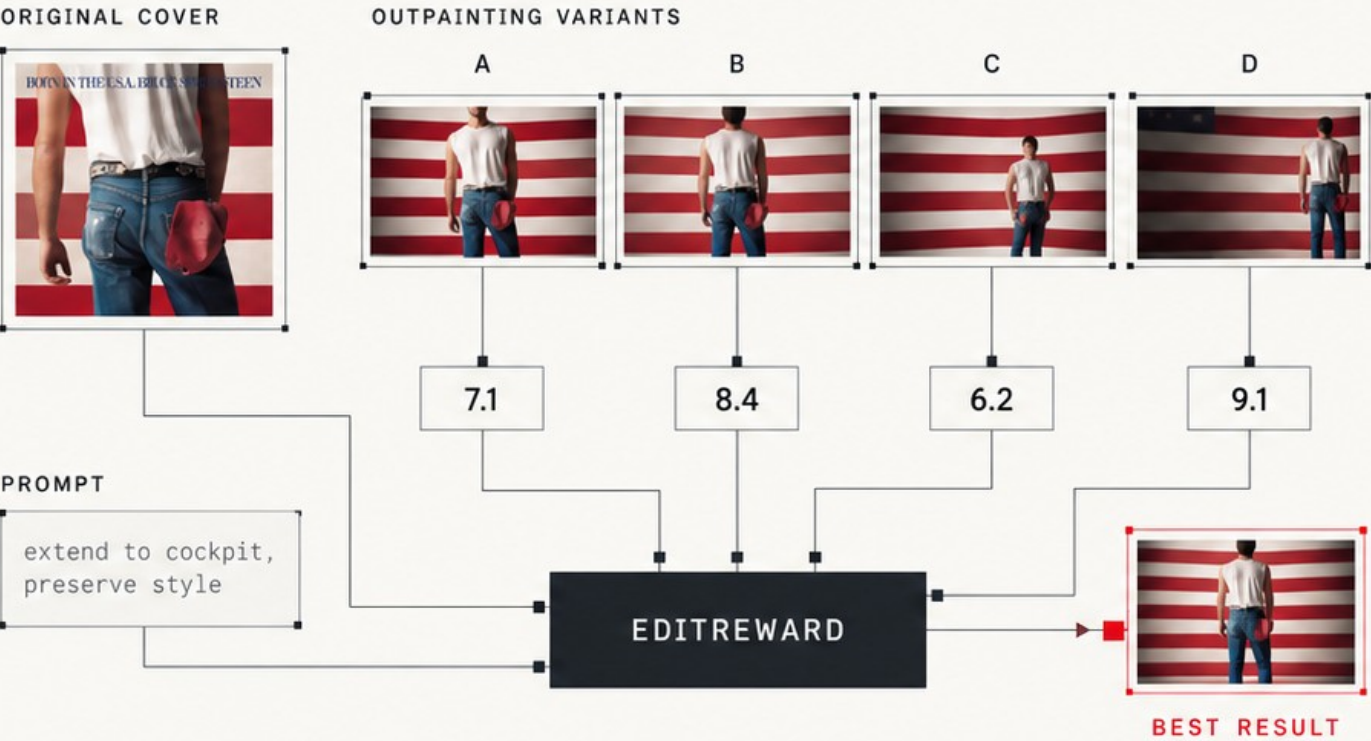


Adam Lambert
WE ARE THE CHAMPIONS

■ Same tracks. Different prompt. Different ranking.

EditReward.

How good is an outpainting relative to the original cover and its edit instruction?



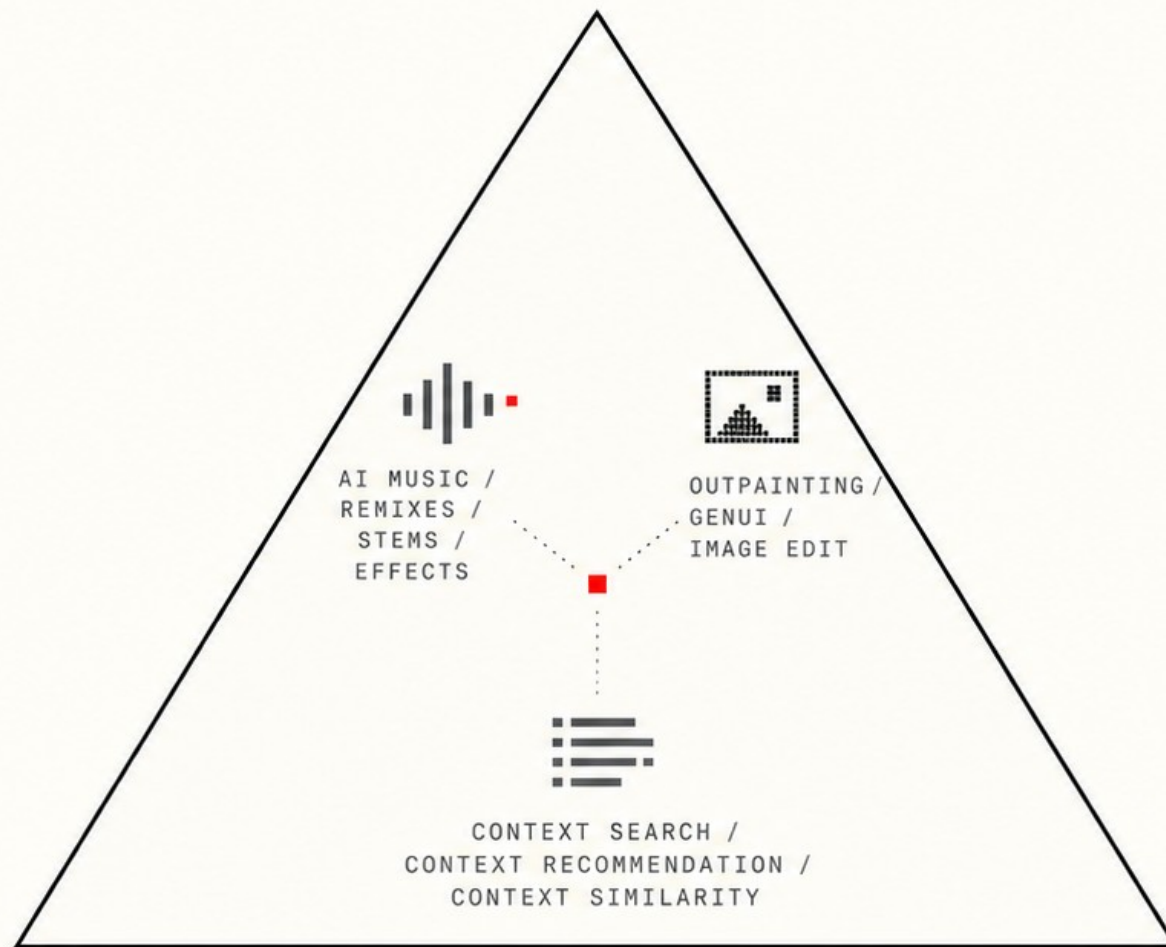
Core idea: score alignment between the original image, edit instruction, and generated result.

RESULTS

The model helps select the more natural extension.

VARIANT 1	VARIANT 2
0.20	0.38
pairwise 0.2024 / single 0.2145	pairwise 0.3774 / single 0.3804
unnatural pose / weak extension	more natural extension

Same prompt. Different results. The score helps choose.



Content Relevance

Context close $\neq$ user intent.

Query: "image for kid"

SAME
VECTORS.
DIFFERENT
INTENT.

01

RANK #1



Cosine similarity: **0.0396**

adult handing a photo print to a child

02

RANK #2



Cosine similarity: **0.0323**

actual children's book illustration

03

RANK #3



Cosine similarity: **-0.0020**

oil painting of a young girl

RANK	IMAGE	COSINE SIMILARITY	CONTENT
1	3.png	0.0396	adult handing a photo print to a child
2	1.png	0.0323	actual children's book illustration
3	2.png	-0.0020	oil painting of a young girl

Similarity across meanings is not the same as **relevance for intent.**

BETTER
CONTEXT.
BRIGHTER
OUTCOMES.

Search retrieves. Context decides.

A practical retrieval stack for relevant AI entertainment.

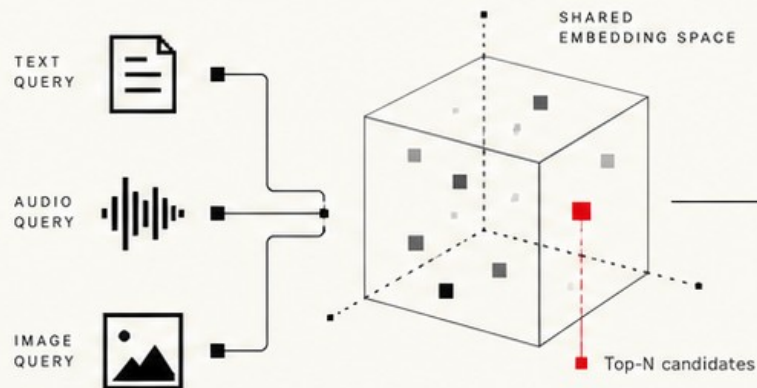
DIFFERENT
CONTENT.
A MORE
HUMAN TOMORROW.



DATA
CULTURE
BETTER
ENTERTAINMENT

01 Multimodal search

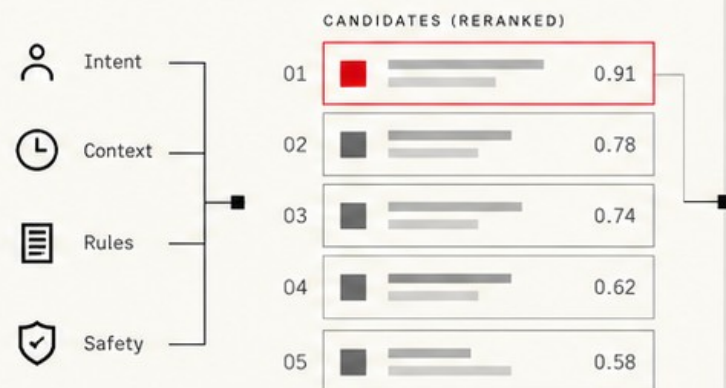
Retrieve text, audio and visual candidates in a shared embedding space.



+ SAME SPACE.
MORE POSSIBILITIES.

02 Contextual reranking

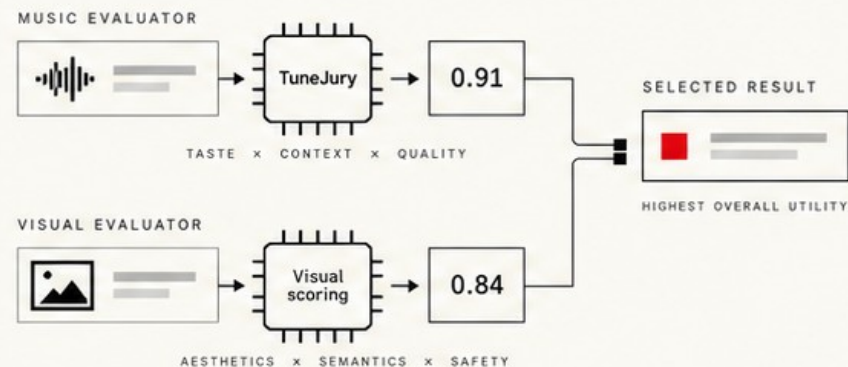
Re-score candidates against real user intent, current context and product constraints.



+ SAME SIGNALS.
A CLEARER ORDER.

03 Domain evaluators

Use models such as TuneJury for music and visual scoring models for images to select the most suitable result.



+ SPECIALIZED MODELS.
A BETTER FINAL CHOICE.

Relevance = similarity + context + quality

Not the nearest vector. The most useful result.

AI FOR
A RICHER
ENTERTAINMENT WORLD.

PEOPLE
CONTENT
CONTEXT
A BRIGHTER TOMORROW.

