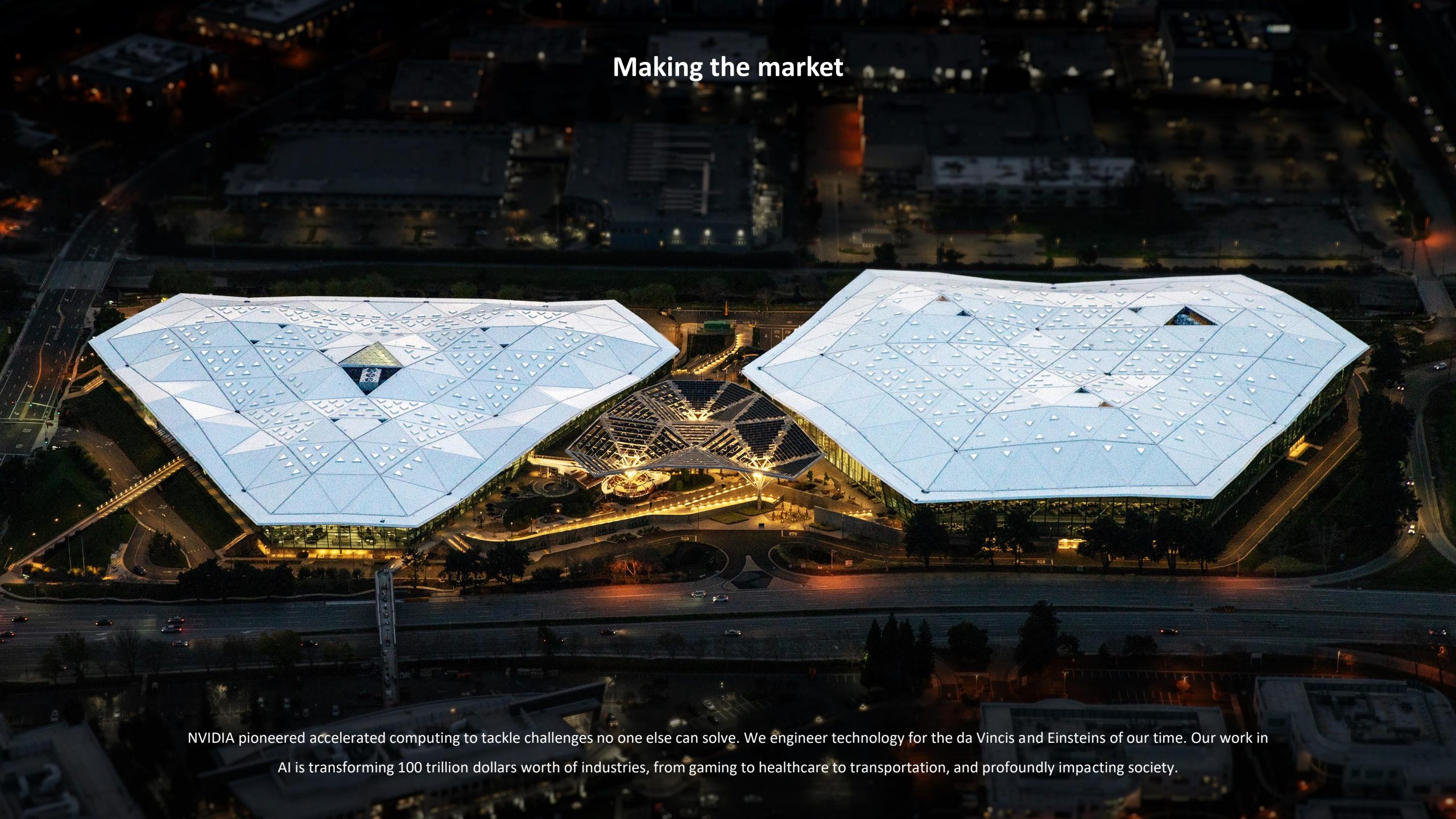
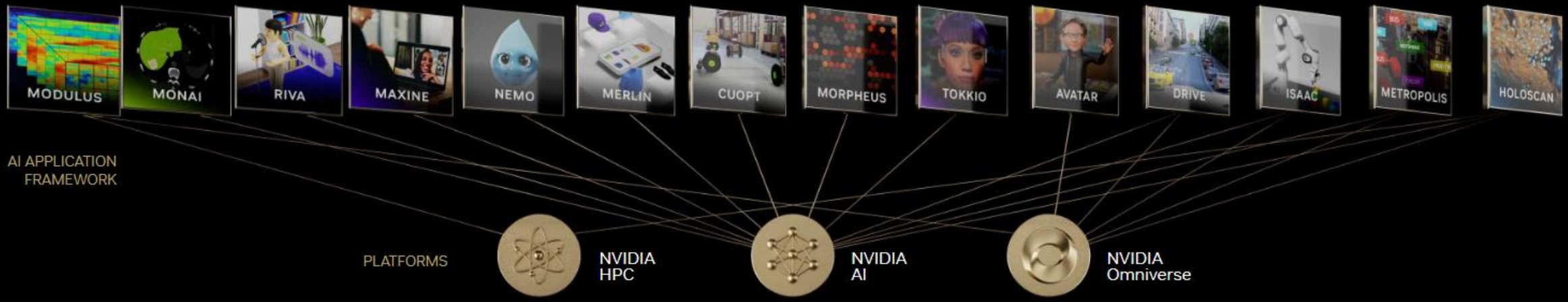
Clarifai
0:12

ROCKSTAR GAMES PRESENTS
RED DEAD REDEMPTION II
GHOSTWIRE
TOKYO
DYING LIGHT 2
STAY HUMAN
GOD OF WAR
CYBERPUNK
MINECRAFT
FORT
MARVEL
GUARDIANS OF THE GALAXY
© 2022 MARVEL
RTX 4090

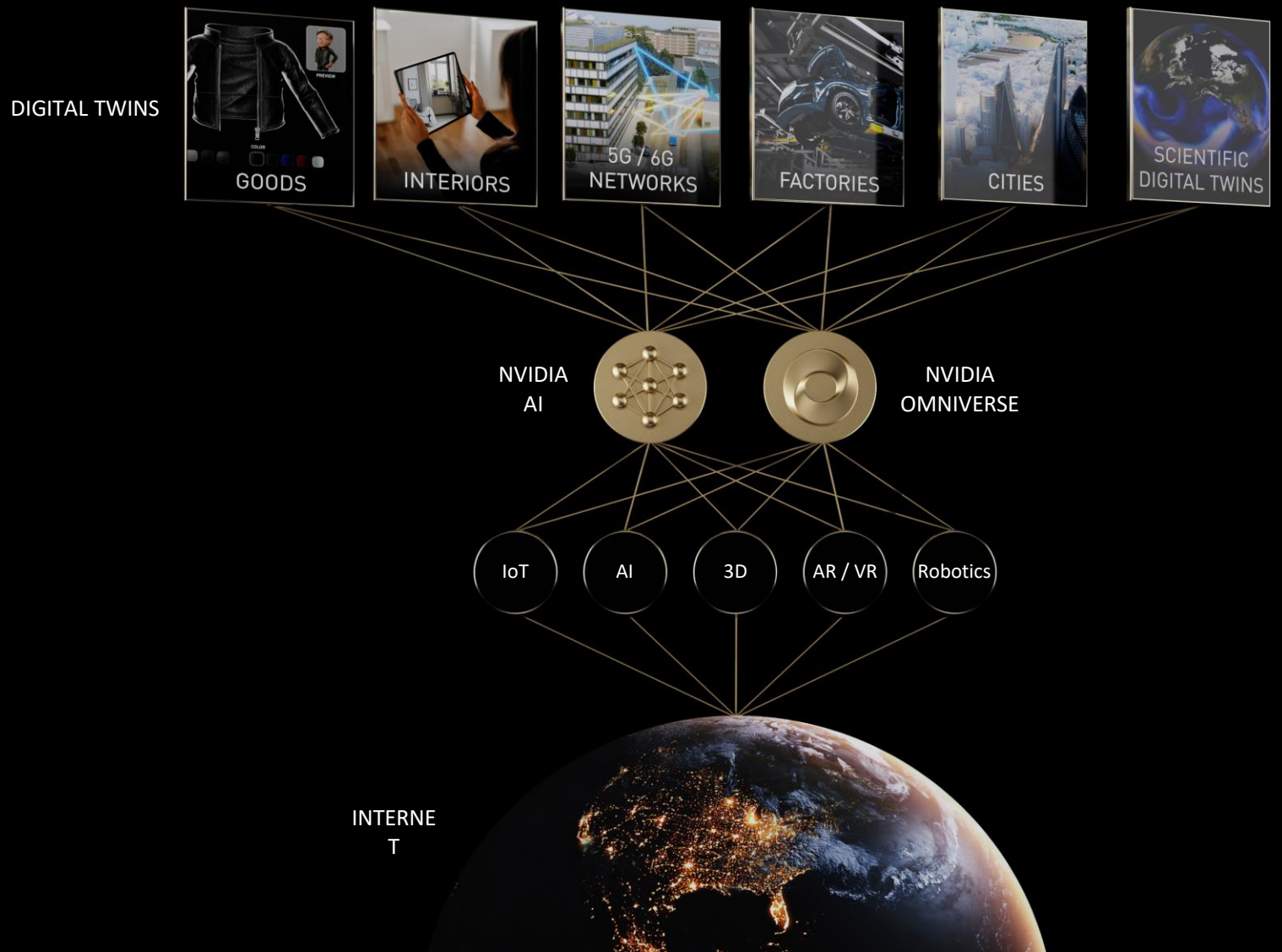
Making the market



NVIDIA pioneered accelerated computing to tackle challenges no one else can solve. We engineer technology for the da Vincis and Einsteins of our time. Our work in AI is transforming 100 trillion dollars worth of industries, from gaming to healthcare to transportation, and profoundly impacting society.

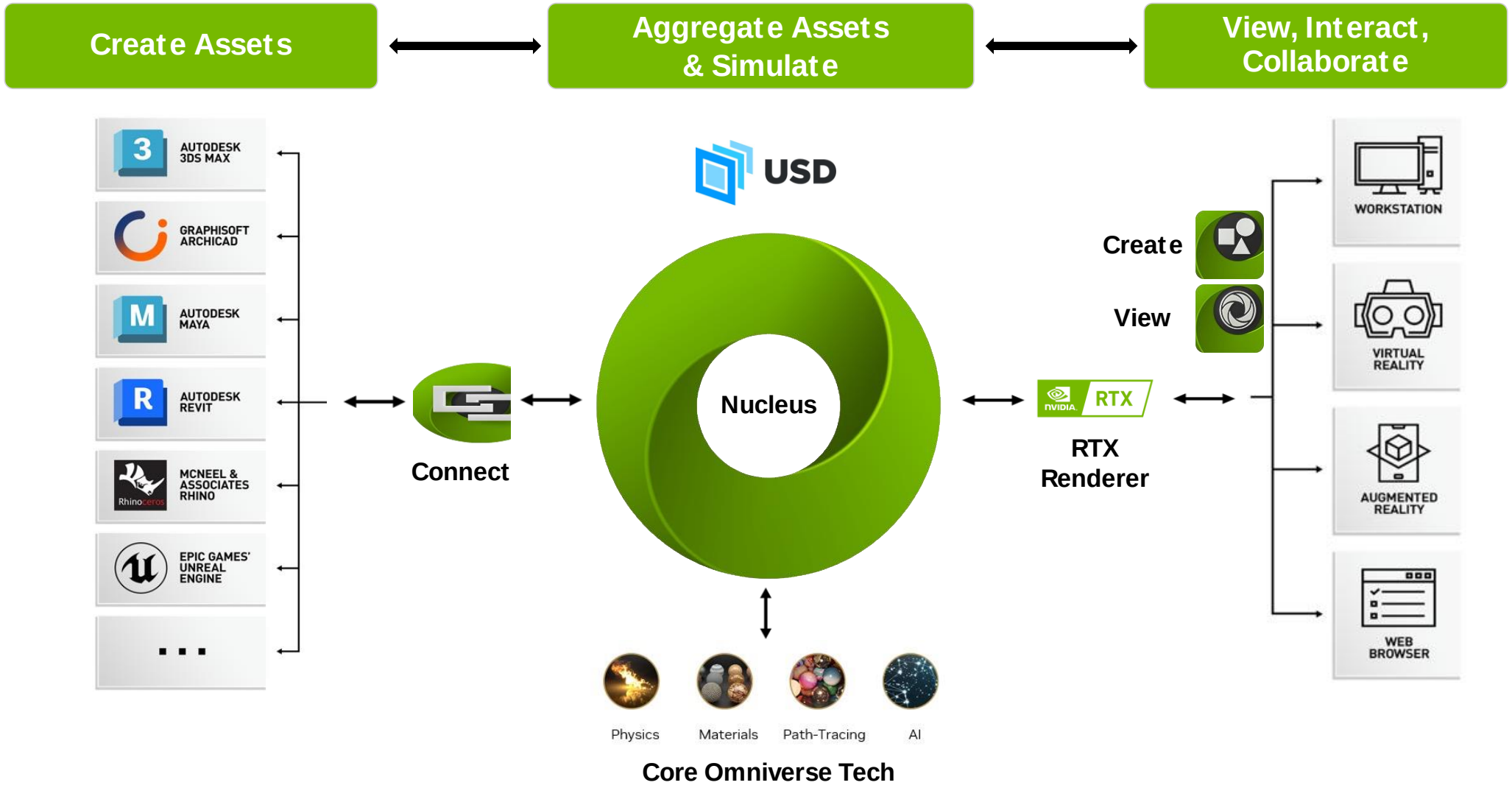

MODULUS
MONAI
RIVA
MAXINE
NEMO
MERLIN
CUOPT
MORPHEUS
TOKKIO
AVATAR
DRIVE
ISAAC
METROPOLIS
HOLOSCAN
AI APPLICATION FRAMEWORK
PLATFORMS
NVIDIA HPC
NVIDIA AI
NVIDIA Omniverse
ACCELERATION LIBRARIES
cuNumeric
CV-CUDA
cuQuantum
Parabricks
Sionna
JetPack
RAPIDS
Spark
cuDNN
cuGraph
TensorRT
Triton
DeepStream
Flare
DOCA
Mag IO
Aerial
CLOUD-TO-EDGE DATACENTER-TO-ROBOTIC SYSTEMS
RTX
DGX
HGX
EGX
OVX
Super POD
AGX
3 CHIPS
GPU
CPU
DPU

THE METAVERSE IS THE 3D EVOLUTION OF THE INTERNET



How Omniverse Enterprise Works

Create, Aggregate, Collaborate



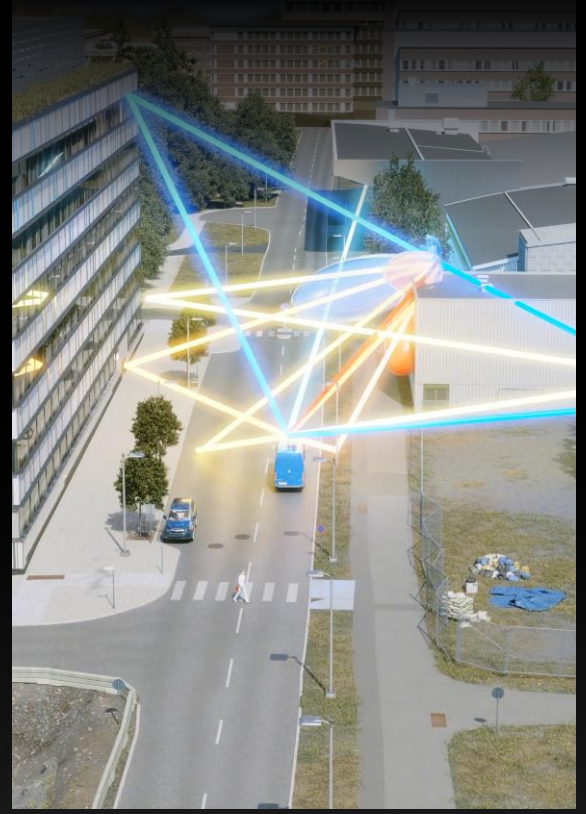
DIGITAL TWINS WILL ONE DAY EXIST AT EVERY SCALE

Design build operate

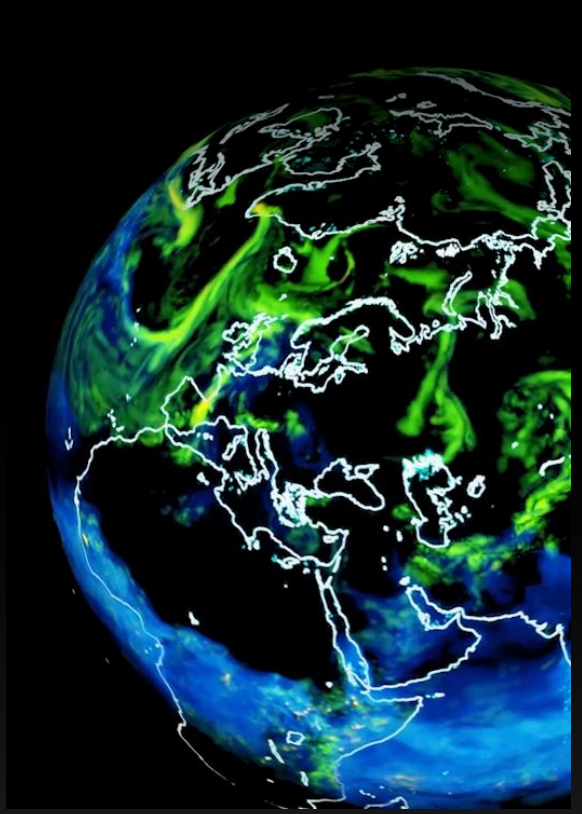
PRODUCT

FACTORY

CITY

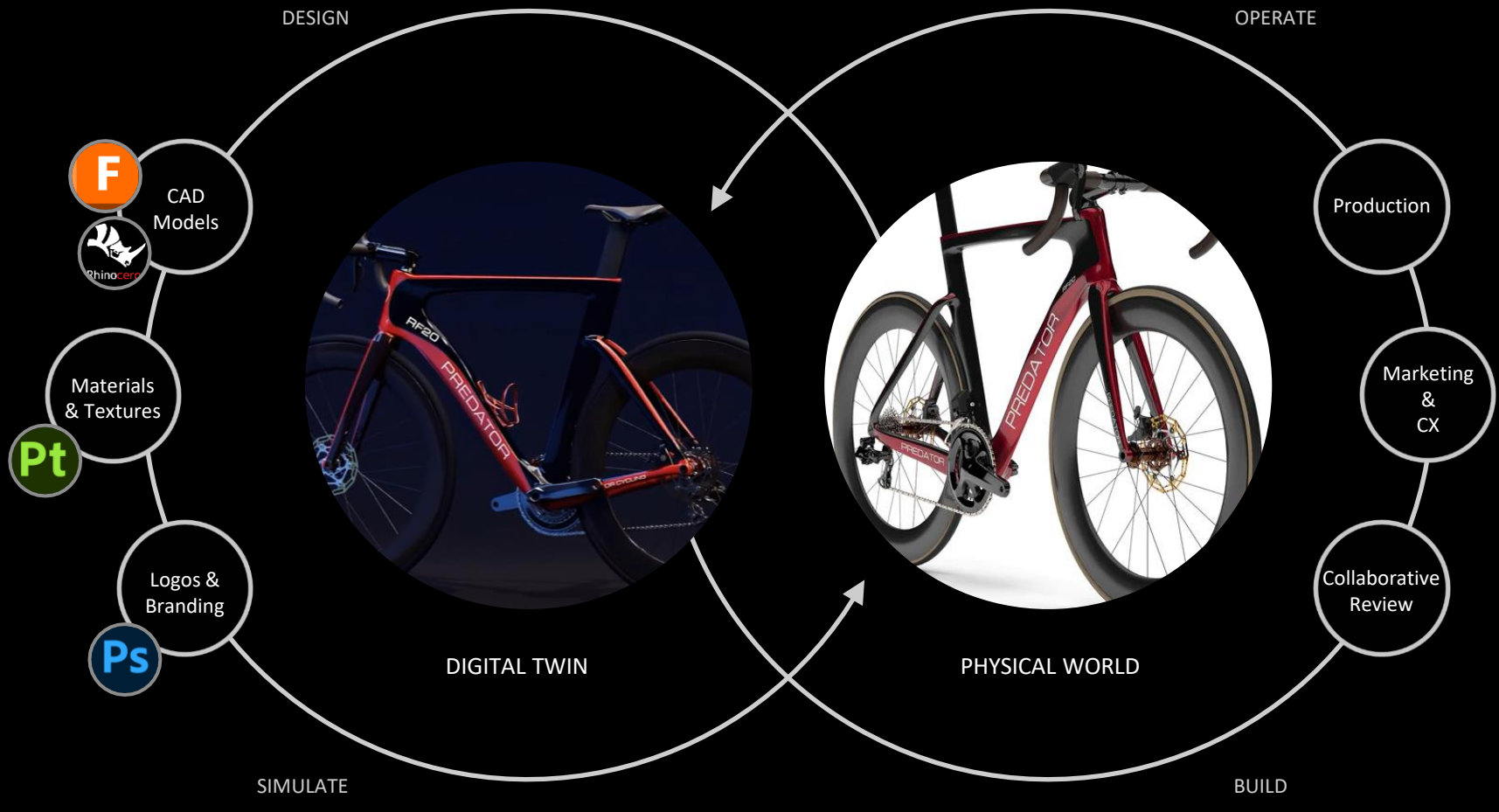


PLANETARY



NVIDIA CONFIDENTIAL. DO NOT DISTRIBUTE.

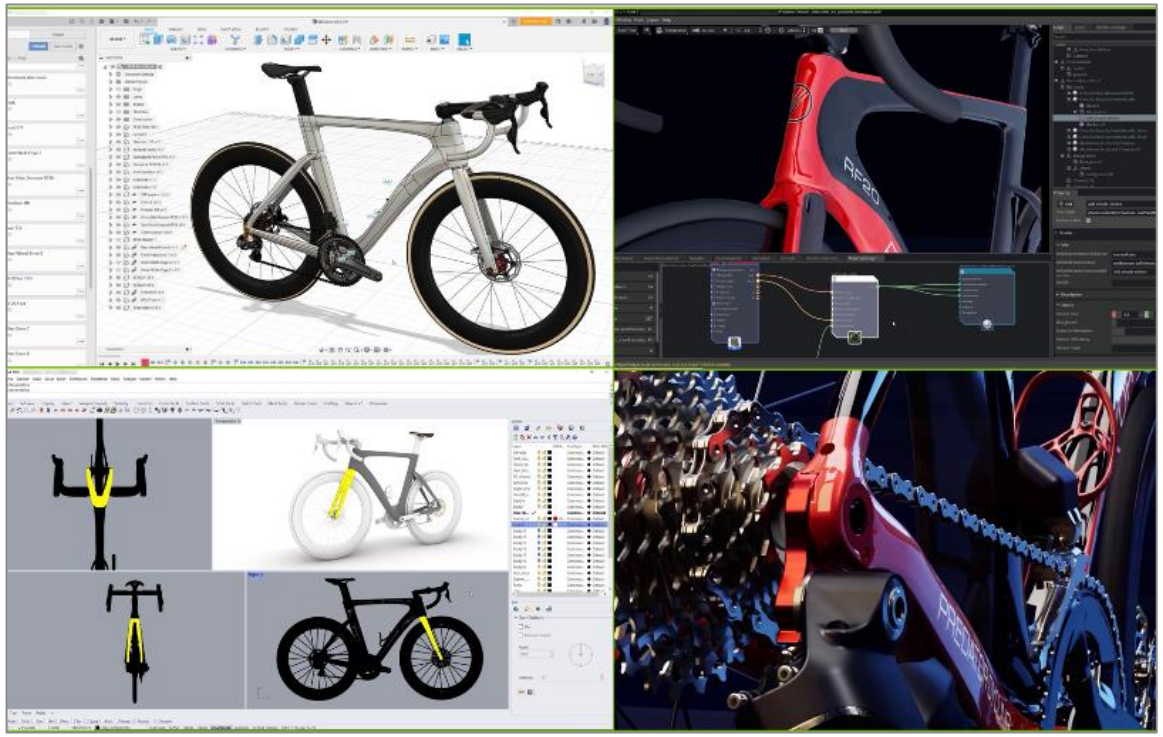
Transforming Product Design and Manufacturing



PREDATOR CYCLING

More Seamless Multi-App Workflows

Rhino, Fusion360, Omniverse Create



CHALLENGES

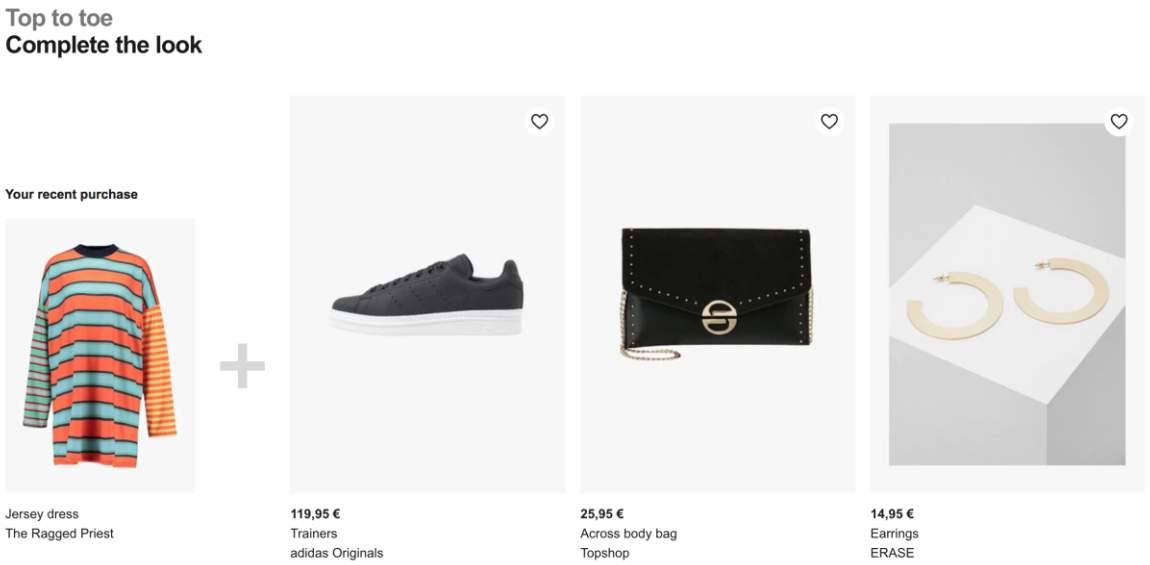
- ✗ Complex design workflows leading to delays in product launch
- ✗ Incompatible data formats from multiple engineering applications hamper creativity
- ✗ Lag in synchronizing data between design, engineering, manufacturing and marketing

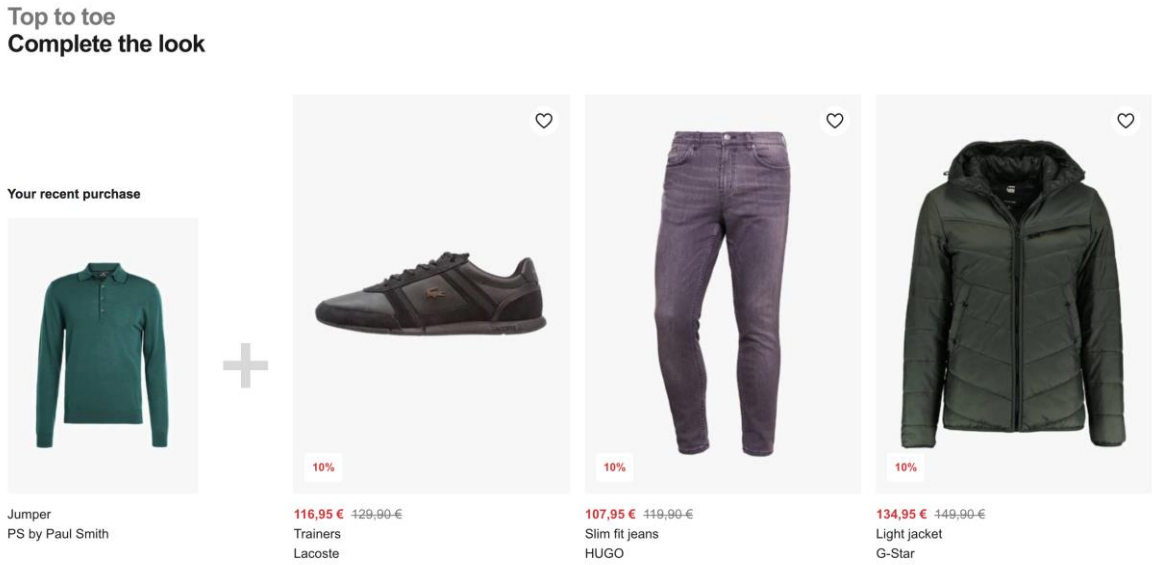
OMNIVERSE VALUE

- ☑ Better communication, feedback, and decision-making, dramatically reducing production time from 12-18 weeks to 4 weeks
- ☑ Efficient and modern manufacturing workflow that aggregates full-fidelity 3D data across multiple apps without any data loss
- ☑ Interactive and vivid environment for design reviews; photorealistic models for marketing and customer engagements

Transforming Product Design Workflows

Predator Cycling – Design and Manufacturing Company





THE RIGHT PRODUCT AT THE RIGHT PRICE… EVERY TIME

Fashion is a fast-moving business. Customers' needs and desires change quickly to keep up with the latest trends. Zalando uses consumer data and sophisticated algorithms, powered by NVIDIA GPUs in Amazon EC2 P3 instances, to improve product recommendations.

When shoppers place items in their carts, the AI responds with personalized suggestions for additional items — giving each customer a hyper-personalized shopping experience.

NVIDIA zalando aws

ANNOUNCING
TENCENT IN PRODUCTION
WITH NVIDIA MERLIN

Over 1 Billion Monthly Active Users

Recommender Training Time Reduced from 20 to 3 Hours

Rapid Iteration Improved Model Accuracy and CTR

NVIDIA Tencent

SNAP IMPROVES
CONTENT RANKING RECOMMENDATIONS
WITH NVIDIA MERLIN

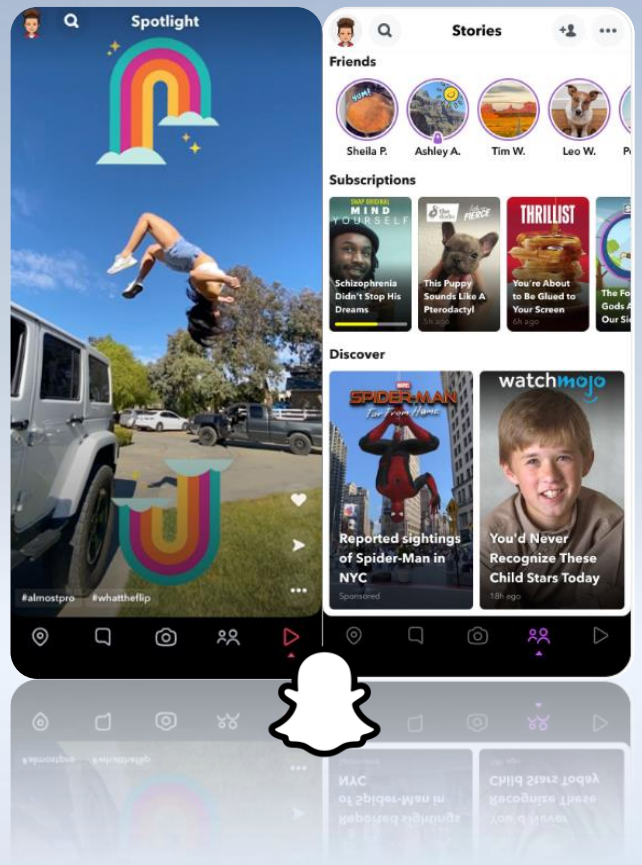
265M People | 5 Billion SNAPS | 30X Engage | 100s Content Ranking Models

Recommenders Curating Real Time Content Feeds, Training Models Daily, Seeking Hourly, and More Rapid Experimentation

Merlin Software Improved ML Inference Cost Efficiency by 50%

Decreased Serving Latency by Over 60%

Providing Compute Headroom to Experiment and Deploy Heavier, More Accurate ad and Content Ranking Models



SPOTIFY IMPROVES PLAYLIST RECOMMENDATIONS WITH NVIDIA MERLIN

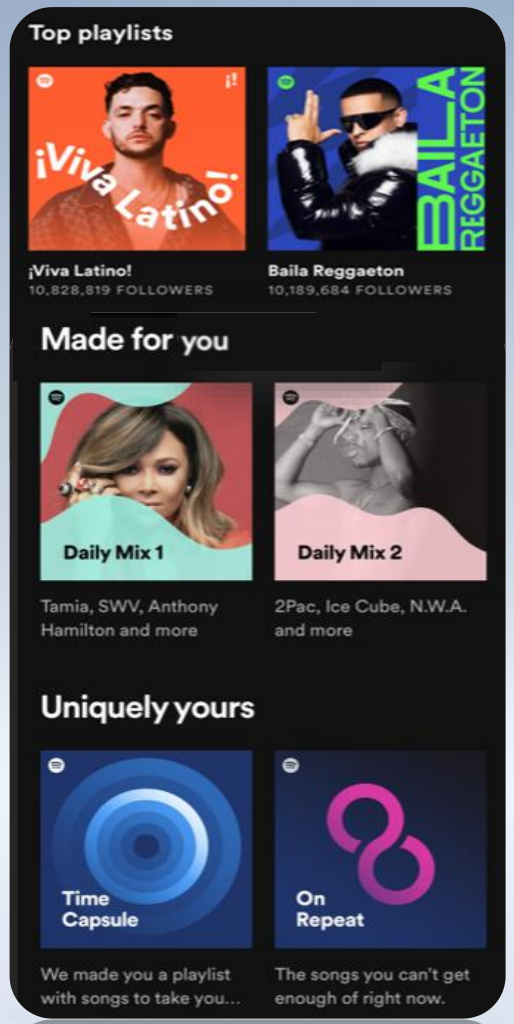
350M Listening Monthly| 70M Songs | 4B Playlists

Transform How Millions Access, Share, And Listen To Music

Keeping Recommenders Fresh With Daily Training, Targeting Hourly

Model Validation Time Reduced From 3.5 Hours To 10 Seconds

Faster Time To Production, Enables Better Recommenders



Spotify®

WALMART IMPROVES DYNAMIC PRICING FOR CUSTOMER SAVINGS WITH NVIDIA MERLIN

200M Customers Per Week | 500M Items Per Year |4600 Stores | 2M Employees

Recommenders Connect Right Item, Location, Person At A Low Price

10X ETL Operations Speedup With Merlin NVTabular

Optimizing And Scaling Recommenders Provides Additional Savings To Customers

More Information: 2021 GTC Session

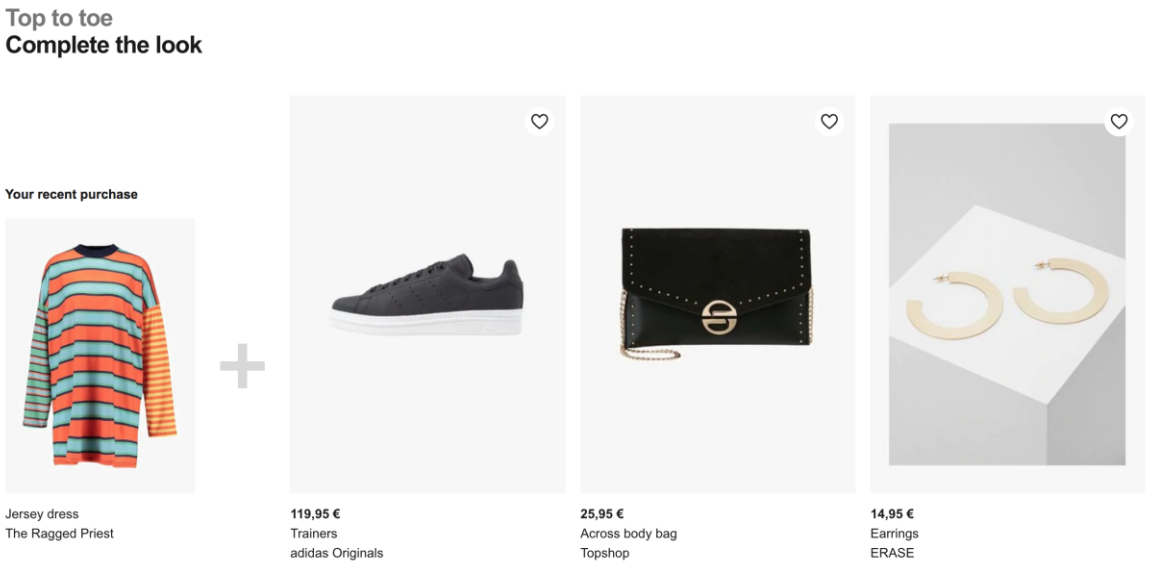
PINTEREST USES TRITON TO IDENTIFY TRENDS IN 300B PINS FOR BETTER SEARCH RESULTS

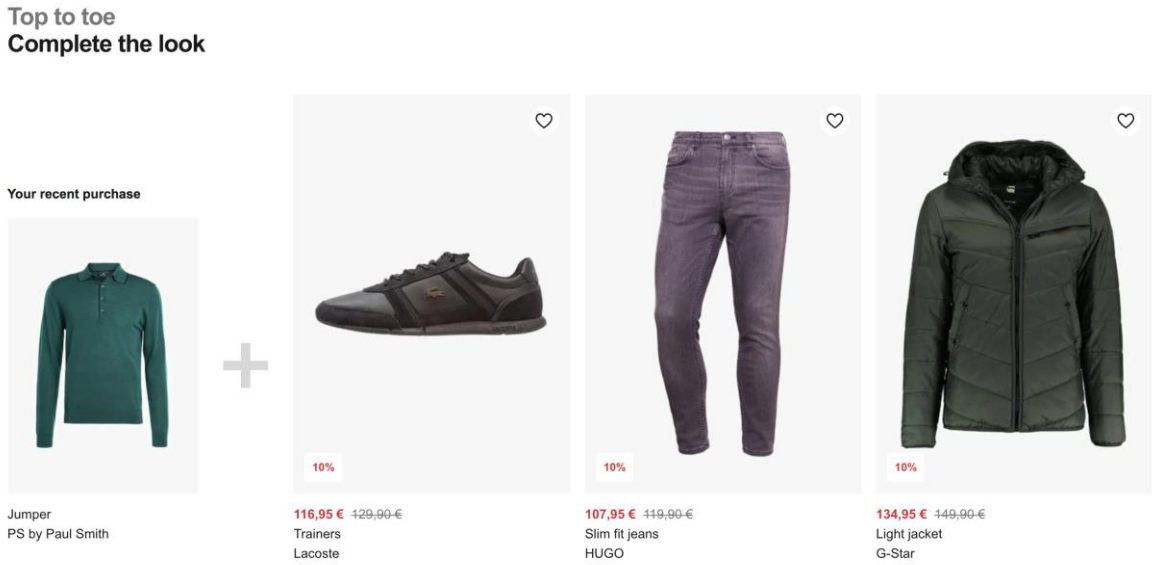
259M MAU worldwide

Provide a delightful customer experience through highly relevant search results and recommendations

State of the art computer vision models by Triton on V100 process 300B pins(images) every night to extract embedded features

High throughput batch inference - 3x higher throughput with Triton dynamic batching





THE RIGHT PRODUCT AT THE RIGHT PRICE… EVERY TIME

Fashion is a fast-moving business. Customers' needs and desires change quickly to keep up with the latest trends. Zalando uses consumer data and sophisticated algorithms, powered by NVIDIA GPUs in Amazon EC2 P3 instances, to improve product recommendations.

When shoppers place items in their carts, the AI responds with personalized suggestions for additional items — giving each customer a hyper-personalized shopping experience.

NVIDIA zalando aws

PERSONALIZED FASHION AT SCALE

- Over 150M active users and growing rapidly
- Users sent in Pinterest images of products they like – and stylists matched to existing inventory
- Impossible for humans to service this volume and complexity
- Powered by GPU-accelerated style recommendation algorithms, Stitch Fix is able to deliver personal-ized fashion at scale

STITCH FIX

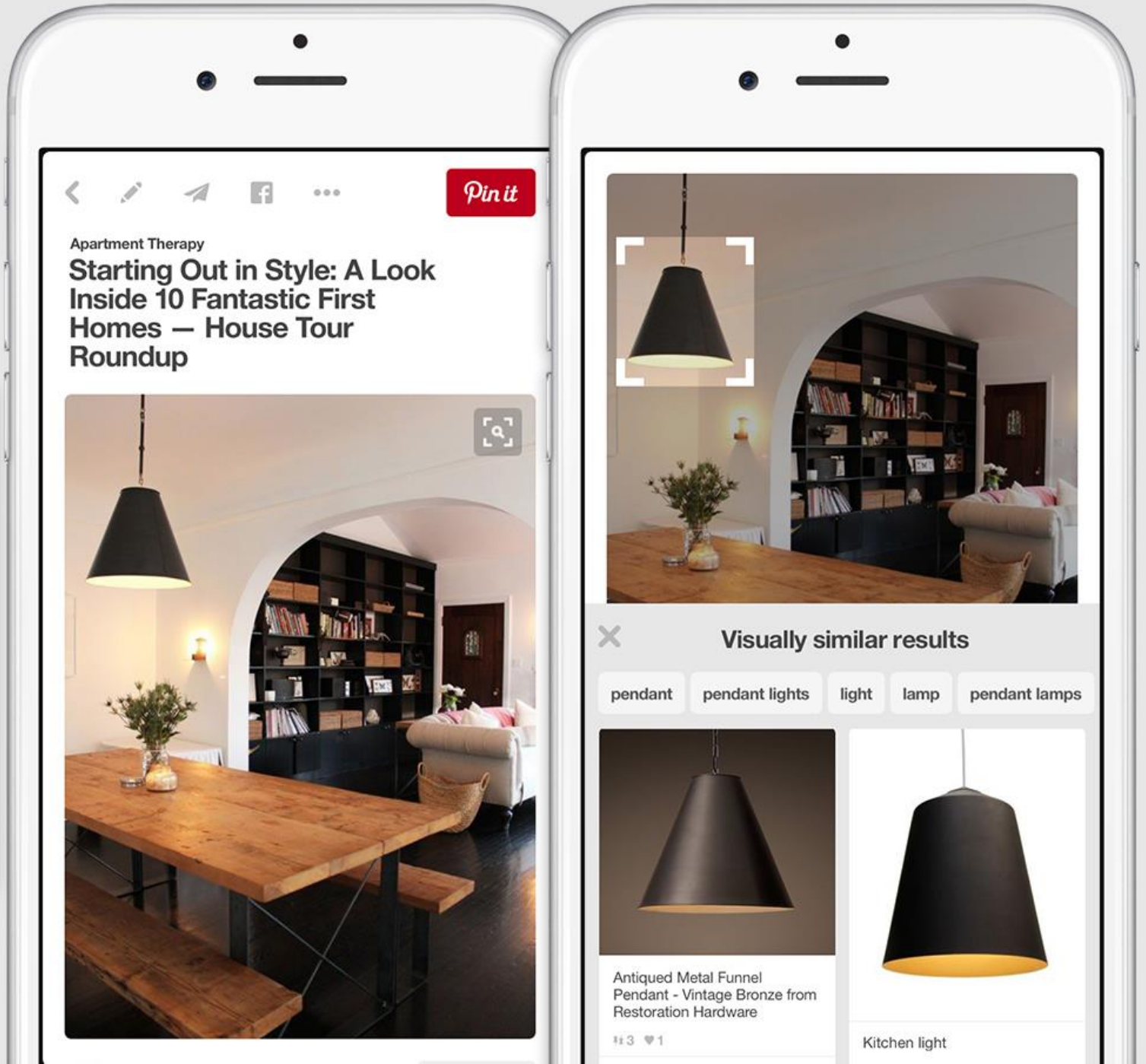
NVIDIA

Take the Stitch Fix algorithms tour

IMPROVED SEARCH FOR A BETTER SHOPPING EXPERIENCE

"Pinterest uses state of the art computer vision technology to build a sophisticated understanding of over 175B pins. We rely on GPUs for training and evaluating our recognition models and for performing real-time inference."
–Andrew Zhai, Visual Search Tech Lead, Pinterest

NVIDIA

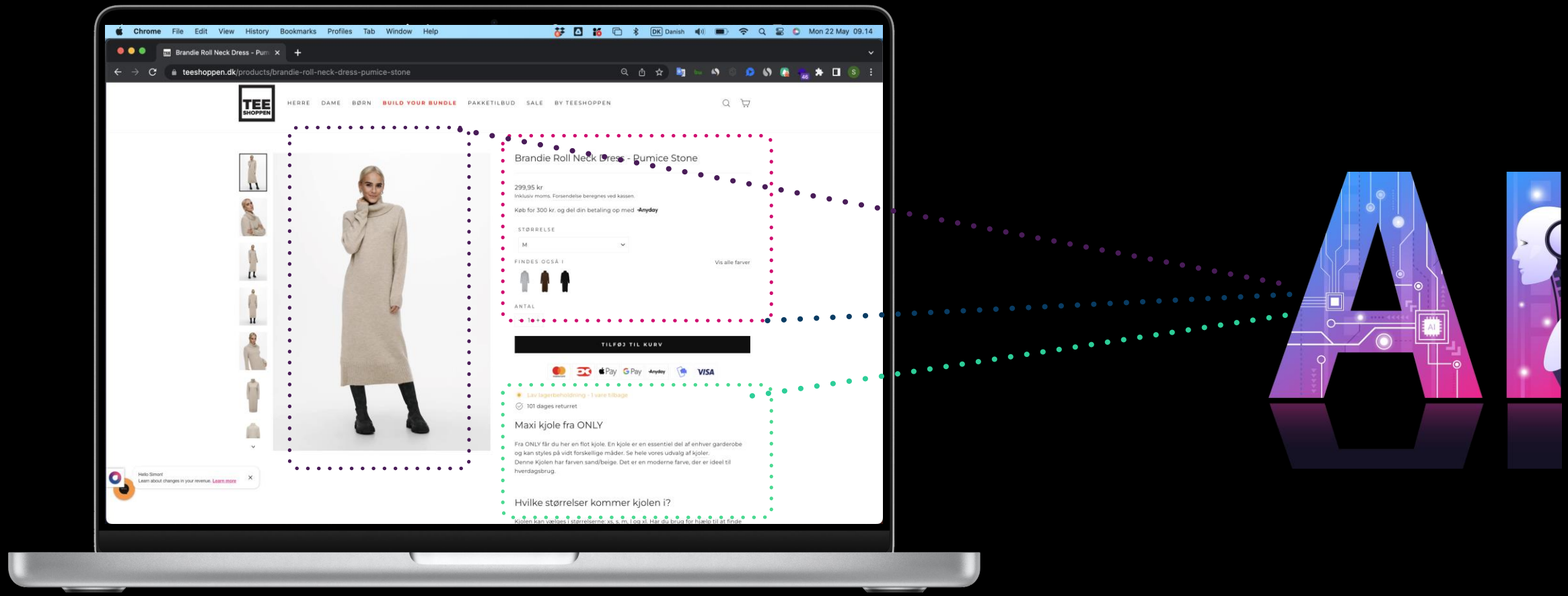


Product Intelligence is the process of storing anonymous behaviour on products instead of users

Protecting user privacy is crucial, given the significance of the Cookie Law, GDPR, and user sentiment.

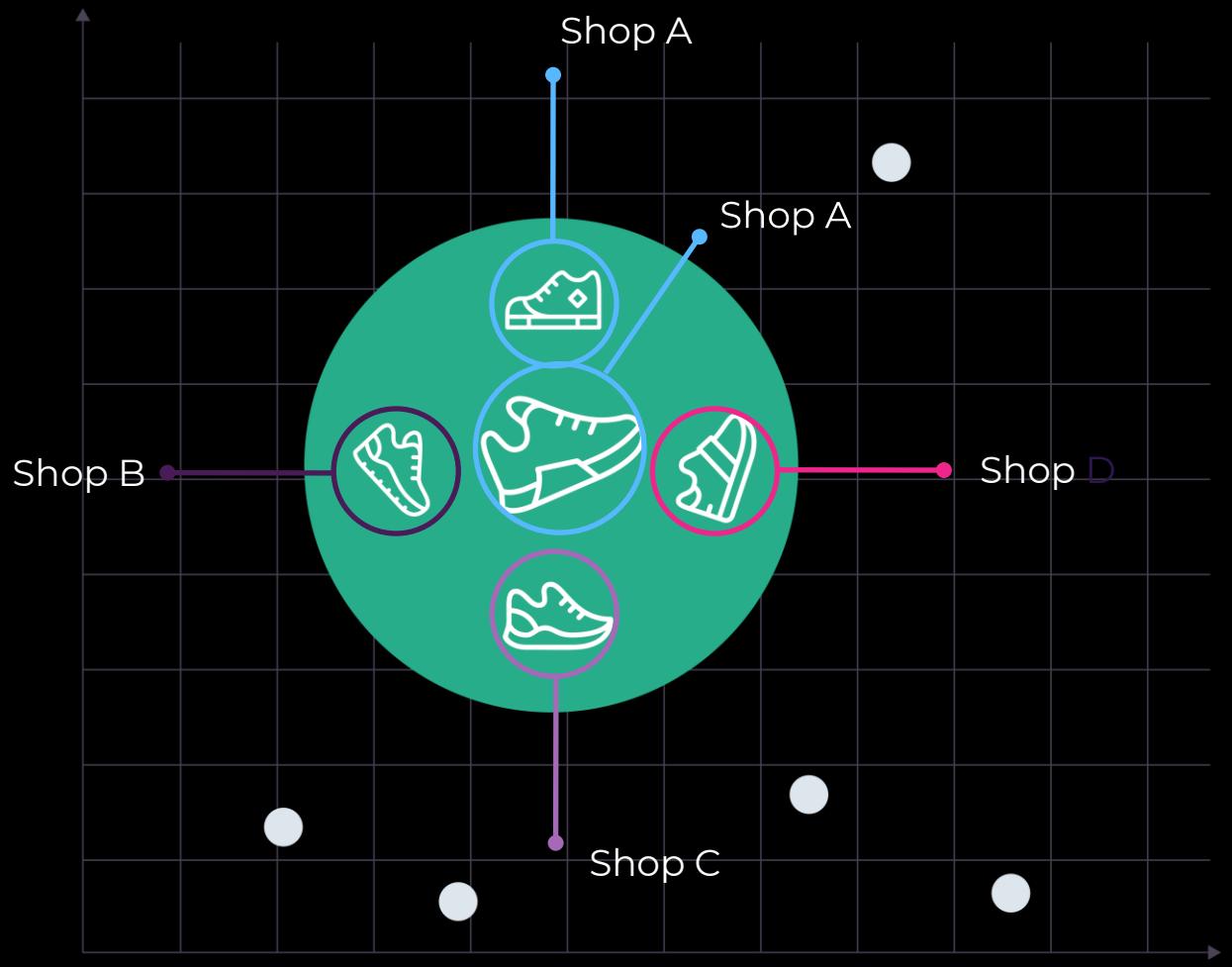
Browsers make it harder to track users across sessions

Product Intelligence uses machine learning to analyze product data like text, images, price, etc.,
and plot it on a multi-dimensional graph.



hello·retail

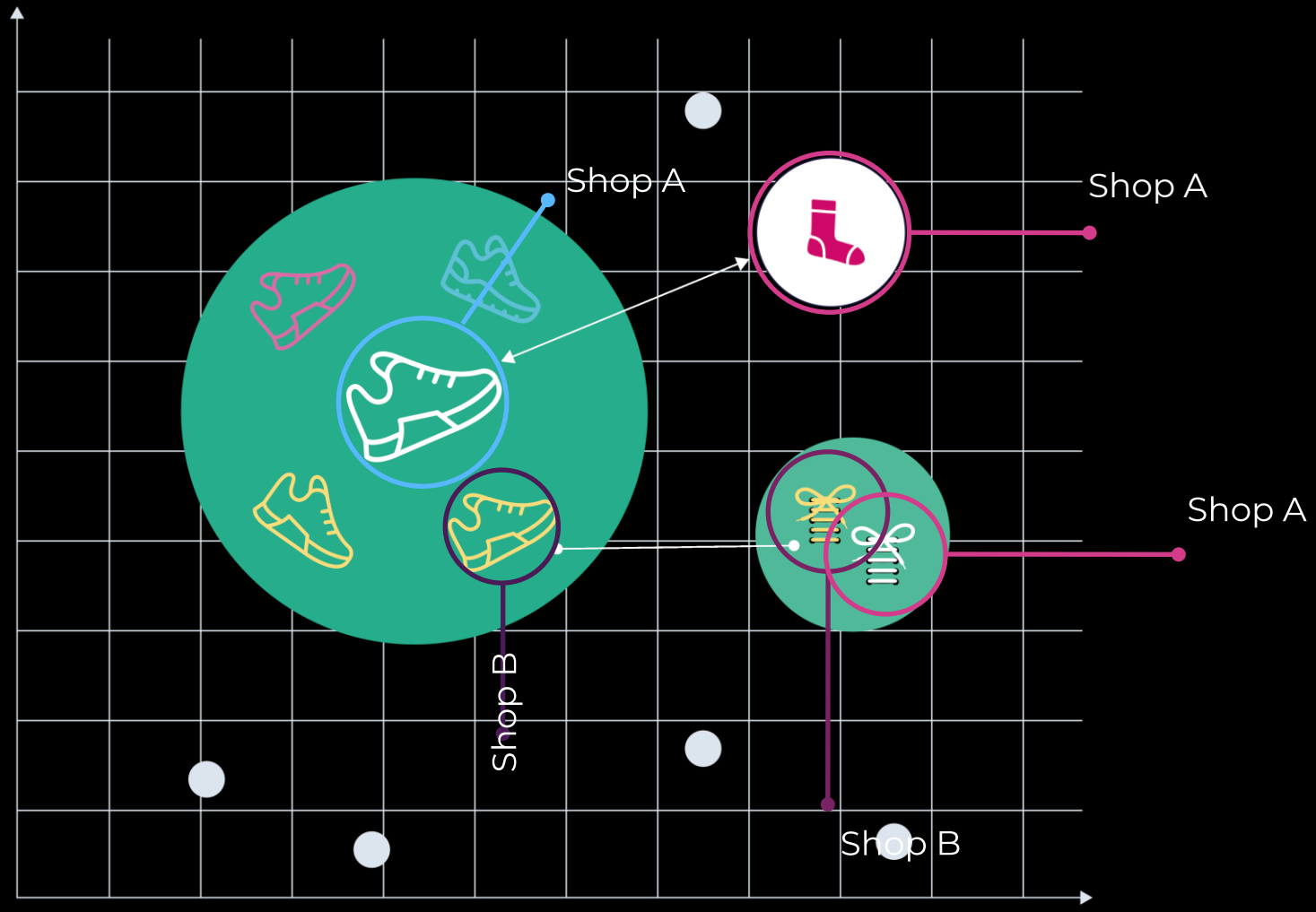
Understanding Similar Products
- Globally



hello·retail

Anonymous Purchase

Relations across Webshops



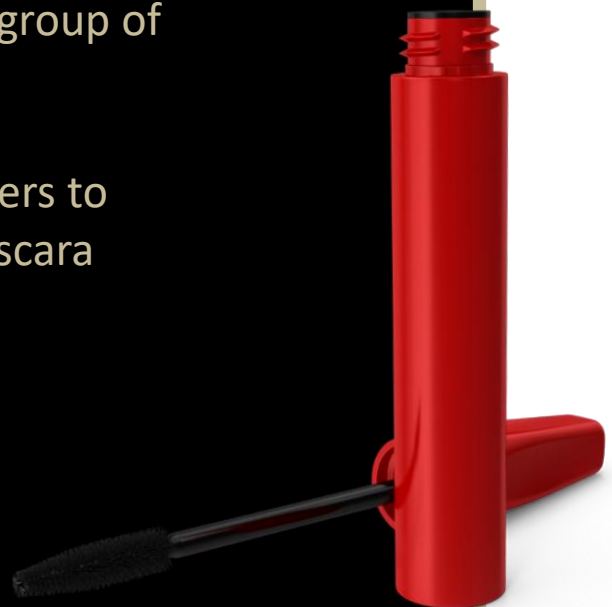
hello·retail

The process of grouping similar groups of products into one product, to reuse product behaviour

With clustering you can begin seeing
trends across large product catalogs

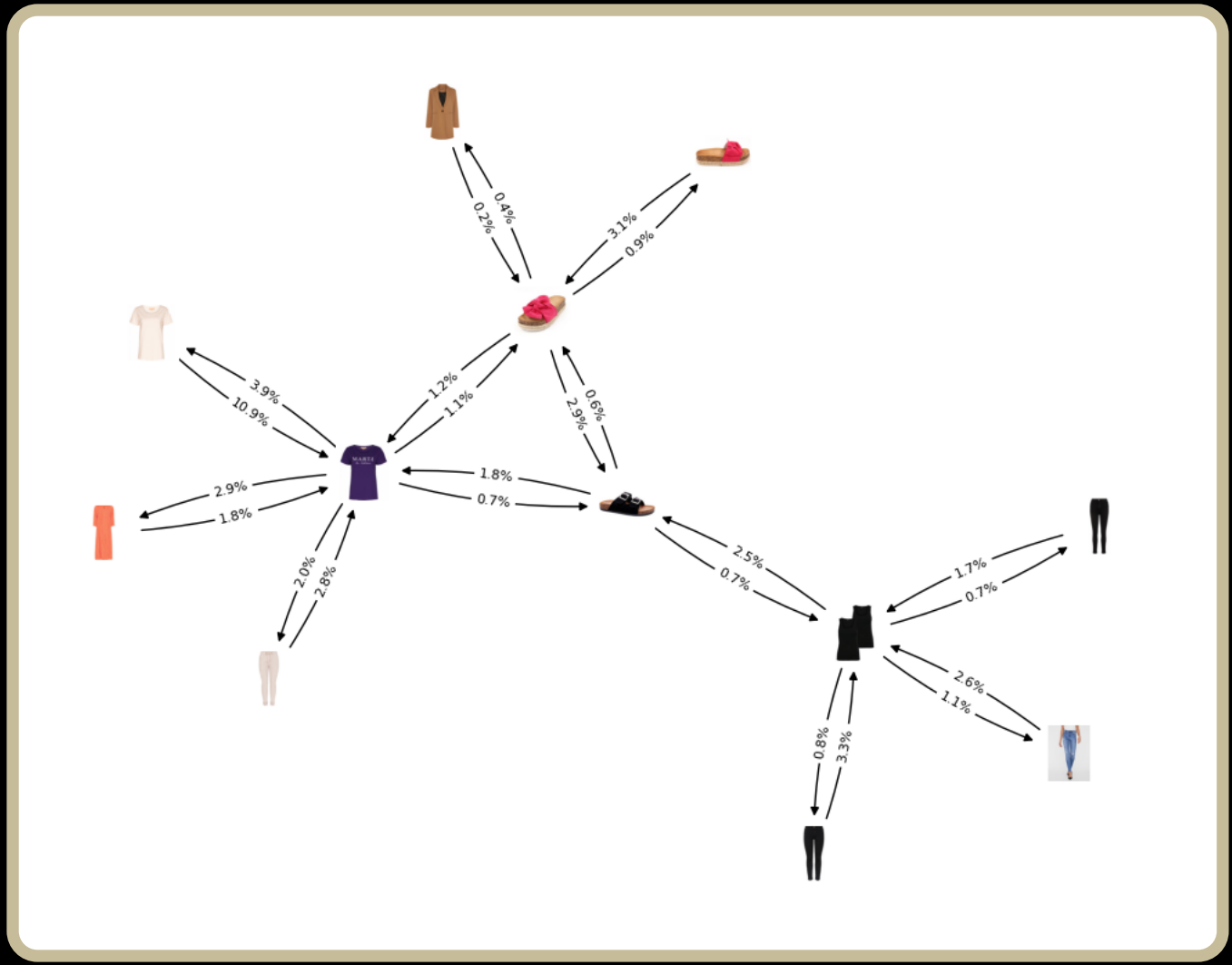
Understand when to push recurring products by looking at a similar group of products (cluster)

Don’t push customers to buy this type of mascara again after week 25



Understand the optimal
product buying journey

Understand:
What order products are purchased, with what likelihood and when



hello·retail