Role of AI in marketplaces

November 9th 2023, ICMA, Amsterdam

Davor Anicic, CEO Velebit.AI

Today's **agenda**

01 Introduction to Velebit AI

We are VELEBIT AI

CO-FOUNDERS

Davor Aničić

CEO

Mladen Fernežir

DATA SCIENTIST

Ivan Borko

ML ENGINEER

Tomislav Levanić

ML ENGINEER

What we do for **you**

- Solution provider for **AI-enhanced buyer-seller matchmaking**
- **9 years of experience** with AI & horizontal marketplaces
- Fast prototyping

AI CONSULTANCY

We support our customers on AI implementation from ground zero to deployed at large scale

OFF-THE-SHELF PRODUCTS

Developed text & images based models for product tagging, discovery & recognition

CUSTOM AI SOLUTIONS

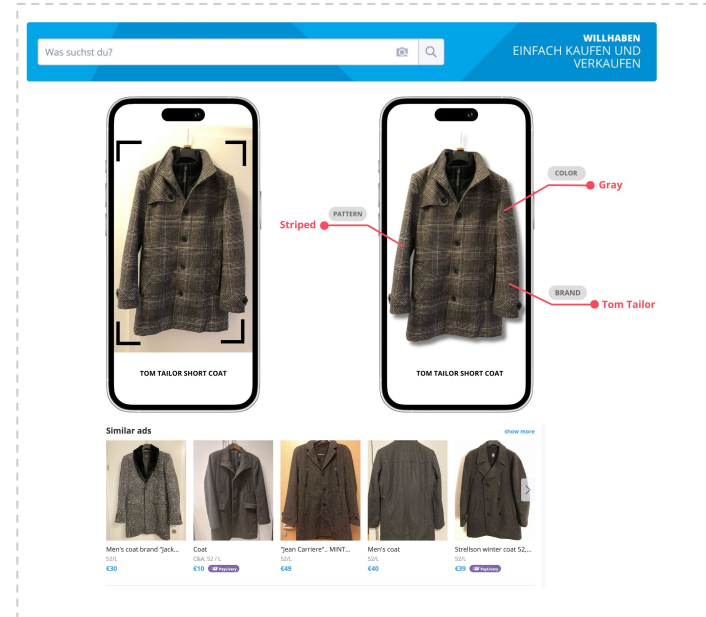We provide end-to-end AI solutions tailored to meet your specific business needs

Companies that trust us

Visual search on

- Visual search for marketplaces
  - **1st in the World**
  - Since 2016
- Color, pattern & brand detection
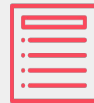
02 AI for Marketplaces

AI for Marketplaces in 2023

**IN-HOUSE
USE OF LLMs**

**IN-HOUSE
USE OF AI TOOLS**

**PRODUCT DESCRIPTION
GENERATION**

**IMAGE
TAGGING**

SMART SEARCH

**DUPLICATE
DETECTION**

In-house use of LLMs for knowledge

- Large Language Models (Generative AI) are compressed knowledge filters
- Conversational interface
- Easier to turn **data → searchable knowledge**

- **LLMs + graf database**
  - Fine Tuning Mistral or LLama2
  - Prompting ChatGPT or Claude
  - Custom GPT (OpenAI ecosystem)

In-house use of **LLM based tools**

PRODUCTIVITY TOOLS

AI DATA ANALYTICS

Microsoft 365 Copilot   Power BI

SOFTWARE DEVELOPMENT

GitHub Copilot   aws CodeWhisperer   Bard

MARKETING & SALES

ChatGPT   Bing   DALL·E   Midjourney   Adobe

Product Description **Generation**

- AI generated product description
- **LLMs** formulate compelling yet informative product descriptions
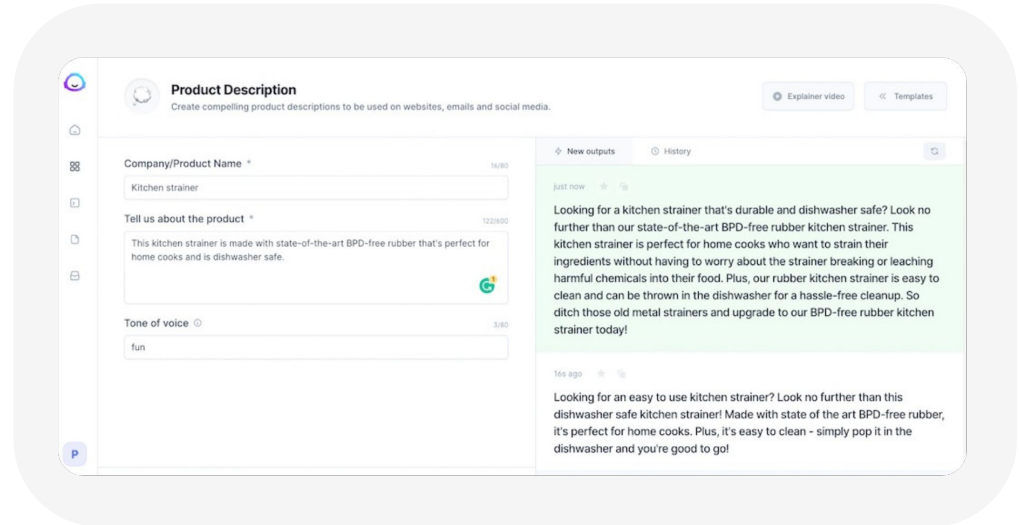- Improves overall quality of descriptions

Image **Tagging**

Men's suit jacket

Dark blue

Long sleeves

Men's suit pants

- AI automatically **labels and categorizes** product photos
- Improves product discovery and searchability
- Accelerates ads posting
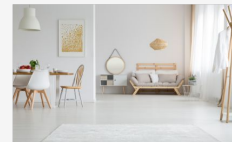- Ensures consistent and precise labeling

Smart **Search**

- Any **AI-enhanced search** technology helps (recommenders based on similarity, collaborative filtering or soft filtering)
- Using **LLMs** for search
- **Searching in a more natural language**
- **Personalized** - match the right buyers with the right products at the right time

Find similar

Duplicate **detection**
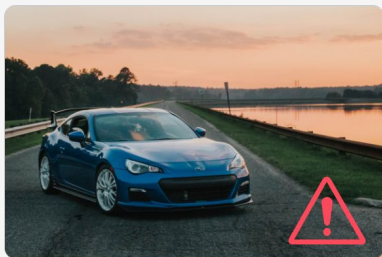
- AI identifies duplicate products if imported from multiple sources
- Manually taking down duplicate products is costly and time-consuming
- The goal is to provide **better user experience**
- Enhanced overall **content quality**

03 Generative AI

Generative AI in 2023

- Generates images
- Writes software code
- Generates & summarizes texts

- Less hallucinations
- Aligned (positive & polite)
- Allows longer context conversations
- Multimodal - reads text & documents, sees images
- Connected to internet

Visualize your ideas with Generative AI

AI **challenges**

INTEGRATION & USER ADOPTION
Needs to be seamlessly integrated into existing workflows and tools to drive user adoption

TRAINING & MAINTENANCE
Requires significant resources to train and maintain, which can be a challenge for smaller organizations

HALLUCINATIONS
Sometimes generates text that is factually incorrect or misleading

LEGAL UNCERTAINTY
Legal uncertainty around the use of ChatGPT, such as liability for its output

OpenAI promises to defend business customers against copyright claims (November 6th 2023)

BIAS
Responses can be biased, reflecting the biases of the data it was trained on

SENSITIVE DATA CONSIDERATIONS
All valuable or sensitive company data should stay within company or trusted AI providers

OpenAI DevDays (November 6ᵗʰ)

GPT-4 Turbo (improved & multimodal)
- **128k context** length (~300 pages of text)
- Knowledge up to **April 2023**
- 3x **cheaper** for input, 2x cheaper output tokens that GPT 4
- **Assistant API** that includes a code interpreter, threading, improved function calling, retrieve outside knowledge, Dall-E 3, and text-to-speech.

App store → **GPT store**
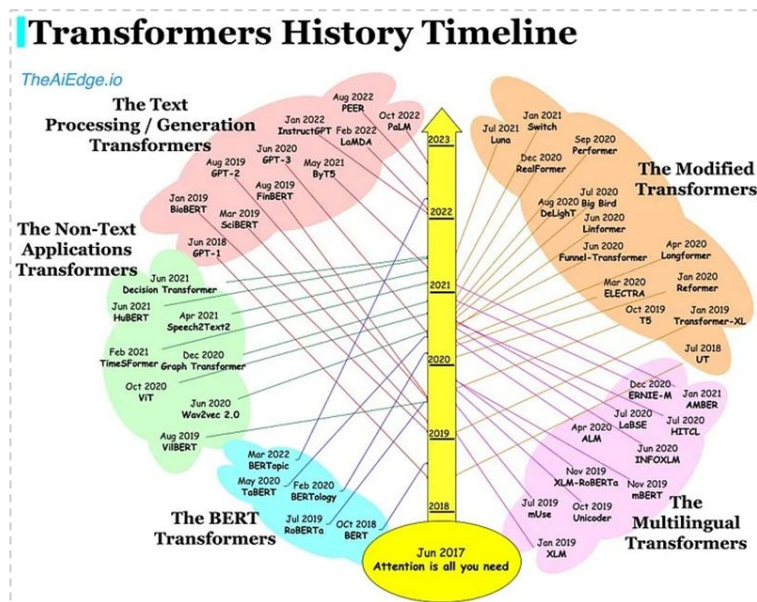- GPTs are customized versions of ChatGPT. Anyone can create their own GPT by using natural language. No coding is needed.
- Users can share their customized GPTs - revenue share with the best ones.

04 AI concepts worth remembering

Large Language Models (LLMs) **History**

Transformers History Timeline
TheAiEdge.io

- **ChatGPT is a LLM**

- Transformer neural network

- GPT family: predicting the next probable word

LLMs **ecosystem**

- **Dozens** of foundation models
- **Hundreds** of derivative models

- Closed proprietary systems
- Open Source Apache 2.0 license

- Stanford's Foundation Model Transparency Index

Rank	Size	Model	Total	MT-Bench	CoT	Python Code	LM-Eval
1		GPT-4-0613	77.78	8.89	0.71	0.66	
2		GPT-3.5-Turbo-0613	65.17	8.22	0.52	0.53	
3		GPT-3.5-Turbo-0301	63.67	8.11	0.48	0.55	
4	70B	WizardLM 70B V1.0	56.38	7.68	0.47	0.27	72.23
5	70B	Stable Beluga 2	55.60	7.42	0.47	0.29	
6	70B	OpenAssistant Llama-2 70B V10	55.01	7.13	0.47	0.31	74.89
7	70B	WizardMath 70B V1.0	51.78	6.86	0.49	0.16	
8	70B	LLaMA-2 70B Chat	48.18	7.10	0.36	0.20	71.54
9	34B	WizardCoder Python 34B V1.0	47.15	6.30	0.29	0.49	
10	65B	Guanaco-65B	44.32	6.63	0.33	0.15	73.92
11	13B	Open-Assistant Code-Llama 13B V10	43.98	6.00	0.29	0.37	62.03
12	33B	Vicuna 33B V1.3	43.04	6.97	0.28	0.11	70.35
13	7B	Mistral-7B-Instruct-v0.1	42.66	6.69	0.28	0.15	
14	13B	Nous-Hermes Llama-2 13B	42.51	6.54	0.29	0.16	70.03
15	13B	OpenOrca x OpenChat - Preview2 - 13B	42.45	6.38	0.31	0.15	72.42
16	13B	OpenOrca-Platypus2-13B	41.41	6.49	0.30	0.09	72.22
17	30B	MPT-30B-Chat	41.24	6.45	0.26	0.16	69.45
18	13B	StableBeluga-13B	41.05	6.19	0.30	0.14	71.31
19	13B	LLaMA-2 13B Chat	40.94	6.66	0.26	0.11	68.48
20	40B	Open-Assistant Falcon-40B SFT-Mix	40.30	6.16	0.28	0.16	74.41
21	13B	Dolphin Llama 13B	39.76	6.04	0.31	0.07	67.12
22	13B	Nous-Hermes-13B	38.49	6.23	0.24	0.12	68.84
23	40B	Open-Assistant Falcon-40B SFT-Top1	38.25	6.30	0.22	0.12	73.95
24	13B	WizardMath 13B V1.0	38.03	5.52	0.34	0.05	
25	13B	WizardCoder Python 13B V1.0	37.09	5.01	0.21	0.40	
26	7B	LLaMA-2 7B Chat	35.36	6.07	0.21	0.05	65.23
27	7B	Vicuna 7B V1.3	34.50	6.11	0.18	0.06	65.83
28	40B	Falcon-40B-Instruct	31.75	5.21	0.19	0.09	72.67
29	7B	MPT-7B-Chat	30.97	5.34	0.15	0.10	64.76
30	12B	Open-Assistant Pythia-12B SFT-8	28.41	5.12	0.13	0.07	60.30
31	15B	WizardCoder-15B	27.48	3.56	0.14	0.38	48.44
32	12B	Open-Assistant Pythia-12B SFT-1	23.30	4.28	0.11	0.03	58.83
33	7B	Falcon-7B-Instruct	21.78	3.88	0.12	0.01	62.50

How LLMs (ChatGPT) **work**

- **Predicting next word** is better when understanding broader context is better
- Capital of Netherlands is…
  - … Amsterdam.
  - … a beautiful city.
  - … where ICMA 2023 conference is held.
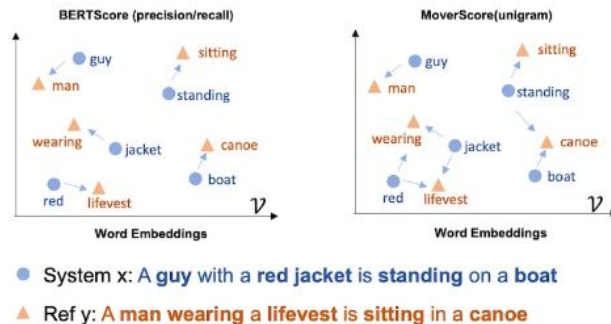
Figure 1: An illustration of MoverScore and BERTScore.

BERTScore (left) vs. MoverScore (right; source)

Source: https://eugeneyan.com/writing/llm-patterns/

Neural networks are simple

- Transforms are 2K lines of code
- Combinatorial creativity

- Stable diffusion image model:
  - 100.000 GB → 2 GB single file

- **To summarize/compress is to learn**
  - Understanding principles
  - LLMs are **knowledge filters**

Old woman driving a cab in Amsterdam in 1935

Reinforcement **learning**

- Trial and error with human feedback
- **Like teaching a toddler**
- Efficient learning of complex concepts
- Teaching Alignment

- Teaching Ethics

A small child looking at a cat and a dog, and the cat and the dog looking back at the child

Natural Language/Chat interface

- **Chat**GPT (Generative Pre-trained Transformers)
- **English instead of Python**
- Like a texting with a person
- Prompt engineering WILL NOT be job.

- September 25th 2023:
  - ChatGPT can now see, hear and speak
    - ➜ multimodality and voice interface
    - ➜ like a conversation with a person

Inevitabilities

- Mathematics, physics, electronics, hardware & software development, Internet, digitalization of our data AI
- **AI is here to stay and change how we work**
- AI's will be smarter than us
- There will be millions of different AI's in use
- AI agents & AI systems

"image of genie that came out of the bottle and smiles"
Made by Bing Image Creator   Powered by DALL·E

05 Where we are at AI timeline

AI today - comparison to cars

1885: <u>Carl Benz first car</u>

1908: <u>Henry Ford model T</u>

2023.

1934: <u>Ford Deluxe Roadster</u>

2024.

2022: <u>Rimac Nevera</u>

The 5 Biggest AI Trends in 2024

- Democratization of AI
- Generative AI hype
- **Working alongside AI - Augmented Working**
- Open source models vs. Models as closed ecosystems
- Ensemble of models & AI agents

06 How to get started with AI

Things to consider

Ways of interaction
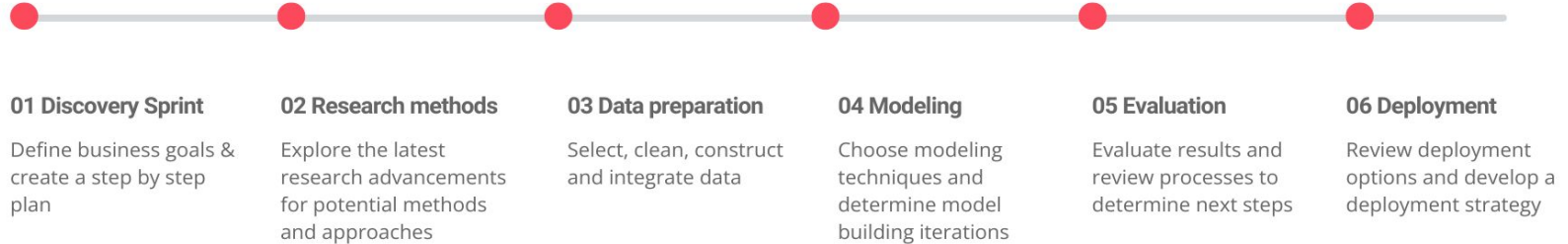
Cost and commercial model

Data protection & privacy

Integration capabilities

AI **implementation journey**

01 Discovery Sprint

Define business goals & create a step by step plan

02 Research methods

Explore the latest research advancements for potential methods and approaches

03 Data preparation

Select, clean, construct and integrate data

04 Modeling

Choose modeling techniques and determine model building iterations

05 Evaluation

Evaluate results and review processes to determine next steps

06 Deployment

Review deployment options and develop a deployment strategy

AI strategy for any business

> **Customer Value Optimization**
>   - Identify the pain points and areas where you can add value to your customers or improve your processes

> **AI implementation strategy**
>   - Put an AI enthusiast in charge of AI transformation
>   - Build internal AI skills
>   - Partner with a reputable AI development company

> **Adaptive AI approach**
>   - Iterate ideas fast (trial & error)
>   - Evaluate the performance of AI solutions

> **AI skills development**
>   - Train all your employees to embrace productivity tools with AI
>   - Train your developers in AI methods

Focus of successful AI strategy

Have top line in mind - not bottom line.

Ask yourself - "What is **the most valuable question to be answered using data** in our company with this new technology?"

Have in mind data inside but also outside your organization**.**

Thank you

Davor Anicic, CEO
Velebit Artificial Intelligence LLC
Zagreb, Croatia
e-mail: davor.anicic@velebit.ai
mob. +385 91 469 2222
web: https://www.velebit.ai/

Backup slides

AI capabilities in 2025 - **AGI?**

No one knows!

AGI - Artificial General Intelligence

Words of top experts in the field:
- Geoffrey Hinton
- Yoshua Bengio
- Yann LeCun
- Andrew Ng

AI challenges - alignment

- We are not sure how in detail big LLM models work
- **Open sourced models** vs. **closed proprietary models**
- Not your model - not your brain
- Transparency in training data
- AI know-how is not hard to acquire and learn,
- Hardware constraints are lowering rapidly
- Model weights are easily copied and can't be contained form spreading

AI challenges - disruption

- Fast changes to societies and labour markets:
  - Creative & White collars jobs
  - Analytical jobs, Translators
  - Speed & Scale… all jobs… too disruptive to predict
  - Disruption of education system

AI challenges - abundance & mixed reality

- For the first time we (will) have large scale problems with:
    - an abundance of content
    - an abundance of creativity
    - believing our eyes & ears (fake or not)
- All this started 25 years ago when we entered digital living

AI **promises**

KNOWLEDGE FOR ALL:
- GPT4 knows a lot about almost everything
- Specialized models for calculus, literature, genomics will know everything about their respective field (GPT4 + Wolfram Alpha)
- 1:1 tutor that works with a classroom but has capacity to adopt to every single individual
- Teacher that is always emphatic, positive and focused

HEALTHCARE:
- diagnostics at scale

Velebit AI Product portfolio

PRODUCT TAGGING

Improve your retail product data and speed up the time it takes to bring your items to the market using our AI product tagging solutions

Automatic categorization

Color detection

PRODUCT DISCOVERY

Elevate your customer experience and assist them in quickly finding what they're searching for

Visual search

Similarity Recommendation

PRODUCT RECOGNITION

Simplify the process of product scanning and automatic categorization with our advanced real-time object detection

Armchair
55 eur

Real-time object detection

CONTENT MODERATION

Optimize your product catalog by removing any duplicates in user-generated content

Duplicate detection

Color detection API

| Returns up to three main colors | Fast response time (under 100 ms) | Integrated background removal | Trained on human-labeled dataset |

- **Automatic detection of product colors**
- Fixed set of **22 colors + colorful property**
- Use it as a simple **JSON REST API**
- Color detection is not based on the pixel values, but on the actual look of the item
- **Customizable** - the set of colors can be customized
- Suitable for *Marketplaces with user-generated images, Fashion & Real-estate E-commerce, Content management systems, Design inspiration websites, Fabric manufacturers*

ORANGE

COLOR

WHITE	MAROON	GREEN	DARK BLUE
BEIGE	RED	OLIVE	BLUE
YELLOW	PURPLE	MUSTARD	BLACK
GOLD	MAGENTA	TEAL	GRAY
ORANGE	PINK	CYAN	SILVER
BROWN	COLORFUL	LIGHT BLUE	

Automatic categorization

- **Predicting the exact product category** from a user-generated image and/or text
- Those categories have a hierarchical structure, from general to more specific in most cases
- We use our proprietary algorithm to categorize products in hierarchical structures based on text, image or both
- Significantly **reduces the time to market**
- Encourages users to upload more content
- Improved user experience

- **Clicks saved** - 2.7 (Average categories depth = 3.1)
- **Performance** - 90.2% Accuracy
- **Ad placing time** shortened up to 70% - from 98s to 29s (iOS)
- **Average user rating** - 4.77 (1-5=highest)

WOMEN'S TOPS

WOMEN'S BAGS

WOMEN'S JEANS

Visual Search & Similarity Recommendation

- Visual search & Similarity recommendation **allows people to find what they are looking for** without needing the words to describe it
- Adds **Natural Language Processing** (NLP) component
- Combines **text + image**
  - More weight on images in "visual" categories (fashion, furniture, antiques)
  - More weight on text in other categories (services, jobs)
- Visual Search is used to boost profits and personalize customer experiences
- Provide better product recommendations with visual similarity search
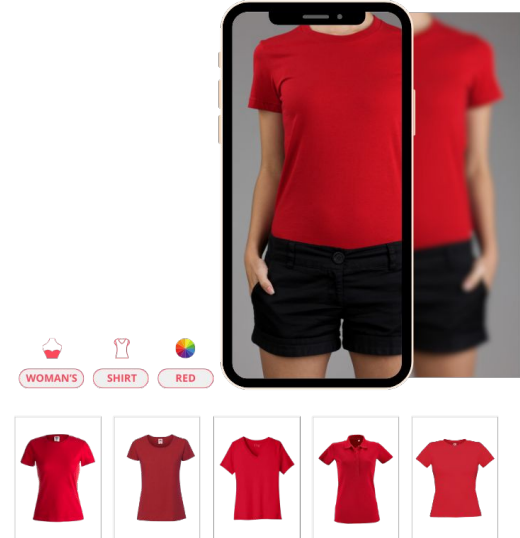
Real time object detection

- Instantly locate and detect objects in image or video
- Detected object can then be used for:
  - **Searching** - to find the most similar objects, assisting users in discovering more products
  - **Categorization** - using the object's class for an action, such as translation to more languages ([Babbly app use case](#)) or for price prediction at Online Marketplaces ([Njuškalo use case](#))
- Production ready
- Implement in weeks
- Cloud-based inference
- Ad posting made faster & easier
- Harmonized product inventory
- Matching quickly sellers & buyers

ONLINE MARKETPLACE

CATEGORY: ARMCHAIRS

PRICE RANGE: 60 - 350 EUR

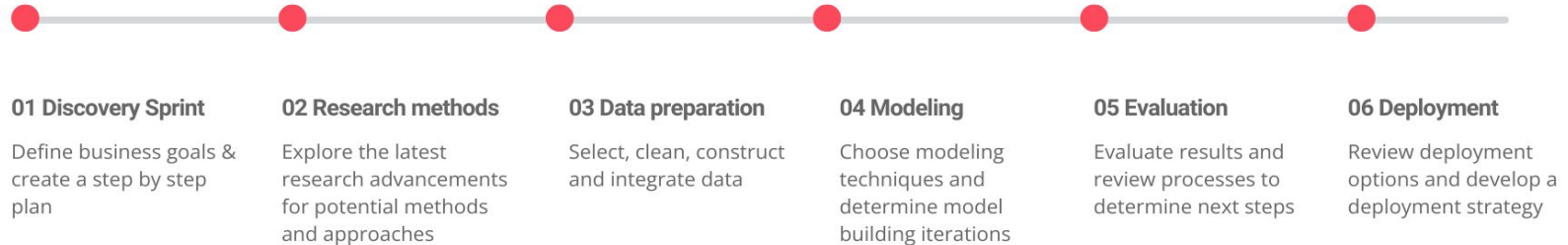**LAST SOLD:** 5H AGO
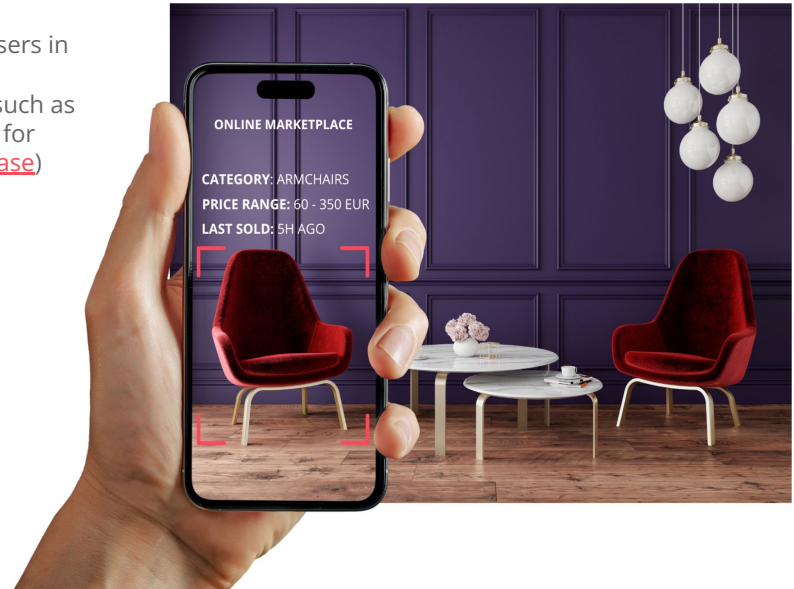
Duplicate detection

- **Identify and eliminate** duplicate content uploaded by users
- Maintain the quality of content on your platform
- Improve your product inventory and management
- Real-time monitoring of user-generated content
- Seamlessly integrates with content management systems
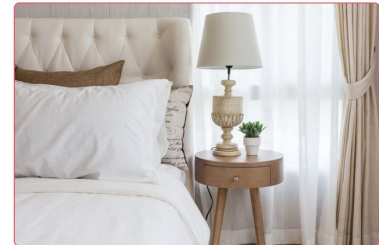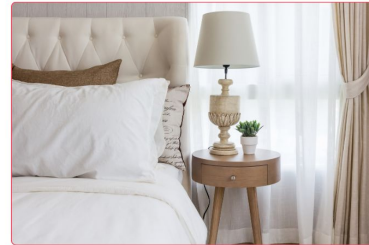
Custom AI solutions

- Machine learning solutions:
  - ○ Natural Language Processing (NLP)
  - ○ Computer Vision
  - ○ Generative AI
  - ○ Predictive Analytics
- We support you in achieving your business goals from initial idea to model implementation & deployment
- **AI solution development phases:**

01 Discovery Sprint

Define business goals &
create a step by step
plan

02 Research methods

Explore the latest
research advancements
for potential methods
and approaches

03 Data preparation

Select, clean, construct
and integrate data

04 Modeling

Choose modeling
techniques and
determine model
building iterations

05 Evaluation

Evaluate results and
review processes to
determine next steps

06 Deployment

Review deployment
options and develop a
deployment strategy