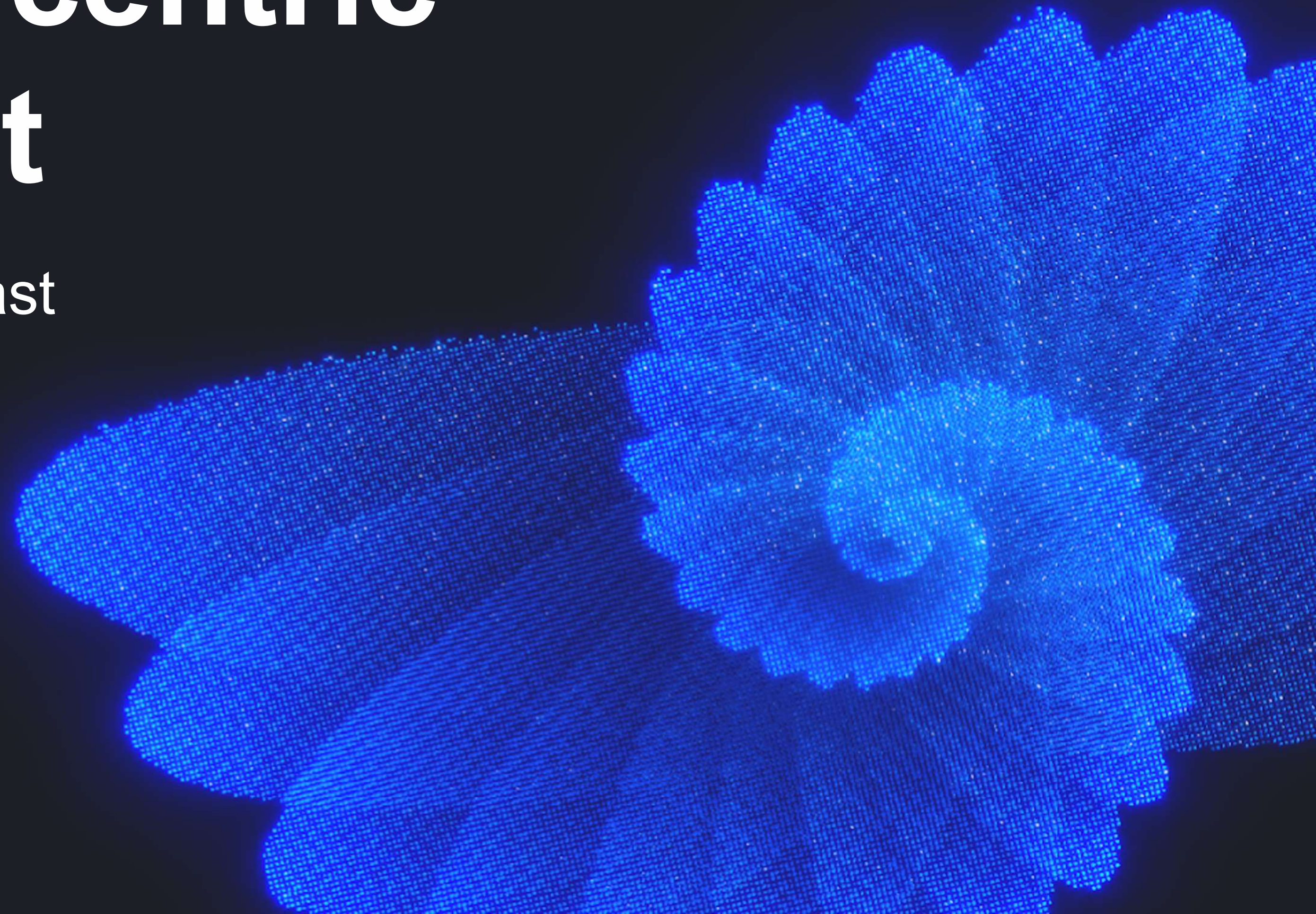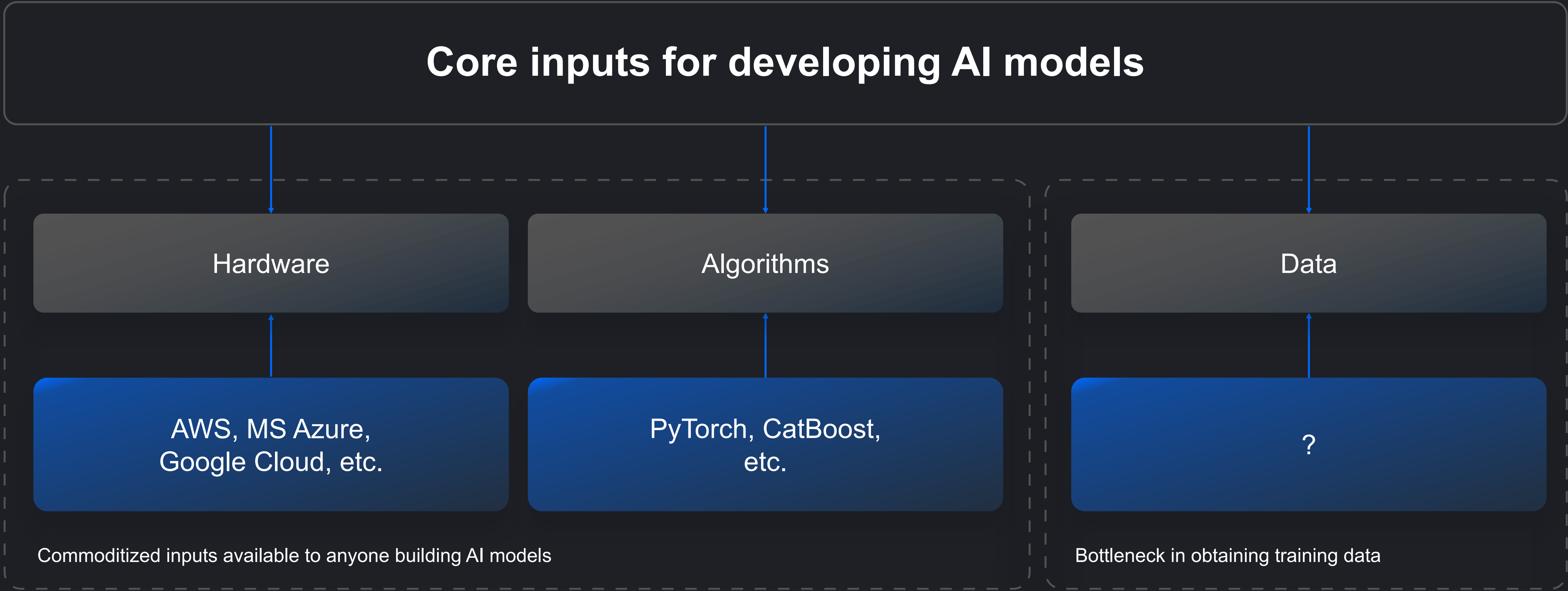# Toloka

Powering data-centric AI development

A unified environment to support fast
and scalable AI/ML development

Context: Data-centric AI is trending in the ML world

Everyone has access to hardware and algorithms.
Data is the key to gaining a competitive advantage.

Core inputs for developing AI models

Hardware	Algorithms	Data
AWS, MS Azure, Google Cloud, etc.	PyTorch, CatBoost, etc.	?

Commoditized inputs available to anyone building AI models

Bottleneck in obtaining training data

✳ Toloka

Toloka supports data-related processes across the entire ML lifecycle

From data collection and annotation to model training, deployment and monitoring

Data collection

Data processing
Store, process and clean data

Data annotation

Data analysis

Model training, deployment, and evaluation

Model monitoring

Founded in 2014 after years of research and experimentation, Toloka is a global tech company that develops a platform and environment to support data-related processes across the entire ML lifecycle.

Designed by engineers for engineers, Toloka environment allows data scientists and ML teams to get AI solutions to production faster by:

— Testing hypotheses
— Boosting the success rate of prototypes
— Building optimal data production pipelines that can be integrated into the ML production cycle

✳ Toloka

Our products & open-source libraries for seamless ML workflows

Data collection & labeling platform

Our [crowdsourcing platform](#) is purpose-built for scaling and acceleration to meet any data labeling demands.

— Diverse global crowd or your own in-house workforce
— Support for any type of data and task
— Design your own data flow or use preset projects

Adaptive ML models `New`

Our [collection of pre-trained models](#) are ready to use out of the box and adapt to fit your data.

— Pre-trained on large datasets with human verification for high accuracy
— Available via API with low latency for model predictions

Toloka ML Platform `In development`

Our [ML platform](#) is engineered for comparing datasets, tracking experiments, calculating metrics, and tuning models.

— Versioning for models and datasets
— Visualizations, reports and diffs
— Python API for access from any environment

Powerful open API

[Toloka-Kit](#) is a Python library for working with Toloka via API.

[Toloka-Java-SDK](#) allows working with API functionality using JVM-based languages.
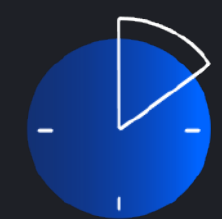
Tools for labeled data

[Crowd-Kit](#) is an open source Python library that simplifies working with crowdsourced data.
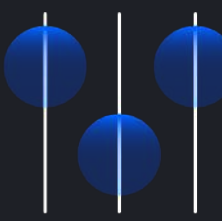
Pipeline management

Integrate your data labeling processes with popular workflow management platforms using our open-source Python libraries: [Apache Airflow](#), [Prefect integration](#)

Toloka

Loved by leading ML teams

Accelerate experiments
with fast data iterations

Tune your solutions
with whatever new data you need, when you need it

Scale your projects
with high-performance tools and unlimited crowd power

Integrate
with end-to-end ML pipelines

Dbrain
"With Toloka we were able to resolve even the most difficult cases of recognizing handwritten text in documents."

Founder and CTO of Dbrain, Y Combinator alum

ID R&D
"We choose Toloka because of high throughput for large data volumes. We collected the world's largest database of 200,000 unique photos and videos."

Science Director and Co-founder

AliExpress
"Thanks to Toloka, we're able to run numerous data projects on a regular basis. What we gain is a dependable approach to data labeling."

Crowd Solutions Architect

Yandex Search
"We were really impressed with how fast we got our project done in Toloka — 10,000 ads were reviewed in just 12 hours."

Special Projects Team
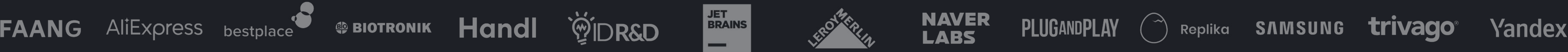
Yandex Self Drive
"Toloka is the first place we go to prepare data for AI. We get a full set of quality control tools and it's 10 times cheaper than our previous solution."

Head of Technologies

TINKOFF
"It allows us to quickly test different markup methods and hypotheses and then choose the optimal approach"

Data scientist

Read more reviews and case studies

FAANG · AliExpress · bestplace · BIOTRONIK · Handl · ID R&D · JET BRAINS · LEROY MERLIN · NAVER LABS · PLUGANDPLAY · Replika · SAMSUNG · trivago · Yandex

Toloka

Why Toloka is the platform of choice

Toloka's capabilities are rooted in Intelligent technologies, a diverse global crowd, and secure infrastructure to support the most demanding projects

State-of-the-art technologies

Advanced tools and unique approaches backed by 10 years of industry experience and research

- Multiple quality control methods and adaptive crowd selection
- Smart matching mechanisms
- Autolabeling (on demand)
- Flexible foundation models pre-trained on large datasets
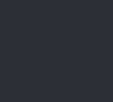- API and open-source libraries for seamless integration

Global crowd

Millions of Tolokers across every time zone for on-demand labeling, instant scaling, and multilingual projects

- **100+** Countries
- **40+** Languages
- **200k+** Monthly active Tolokers
- **800+** Daily active projects
- **24/7** Continuous data labeling

Robust secure infrastructure

Fault-tolerant high-load system for rapid knowledge enrichment that prioritizes data security and privacy

- **GPDR** Privacy-first, GDPR-compliant focus on data protection
- **ISO** Information security management system (ISMS) is ISO 27001 — certified
- Multiple data storage options, including Microsoft© Azure cloud for world-class protection
- Automatic scaling to handle any volumes

❄ Toloka

Fueling data-driven solutions

Innovative applications integrate Toloka for on-demand data streams

Handl
Document recognition

Stream any documents via API for fast extraction of structured data with AI plus human-in-the-loop

ASK THE CROWD
Usability testing

Test your designs and ideas on the spot — ask people to choose the best option and get feedback in 5–10 minutes
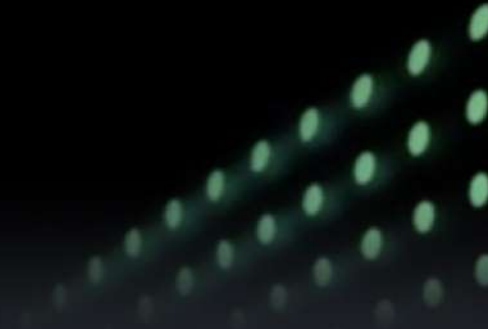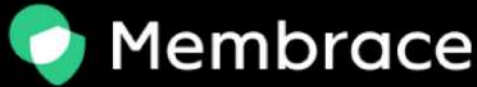
Meteum
Hyperlocal weather forecast

Get real-time weather data for your business with accurate forecasts and analytics at the neighborhood level, anywhere in the world

Smelter
Competitor and media analysis

Gain ML-based insights into what people are saying about your brand and your competitors online to fine-tune your communication strategy

[Learn more](#)

Membrace
Content moderation

Flexible content moderation platform that combines ML and Human-in-the-loop into effective ready-to-go solutions

NeuroDub
Automated video translation

Automatically generate voice-over translations and subtitles of videos for 6 languages in under 60 seconds

SupportAI
Customer support

Automate customer support and hiring with an integrated AI chatbot that covers 80% of incoming requests

Toloka

Recognized as an industry advocate

We are committed to making an [Impact on the AI industry](#) and helping companies unlock AI opportunities

Gartner®

Featured in Gartner's Hype Cycle Report for Data Science and ML as a notable data labeling solution
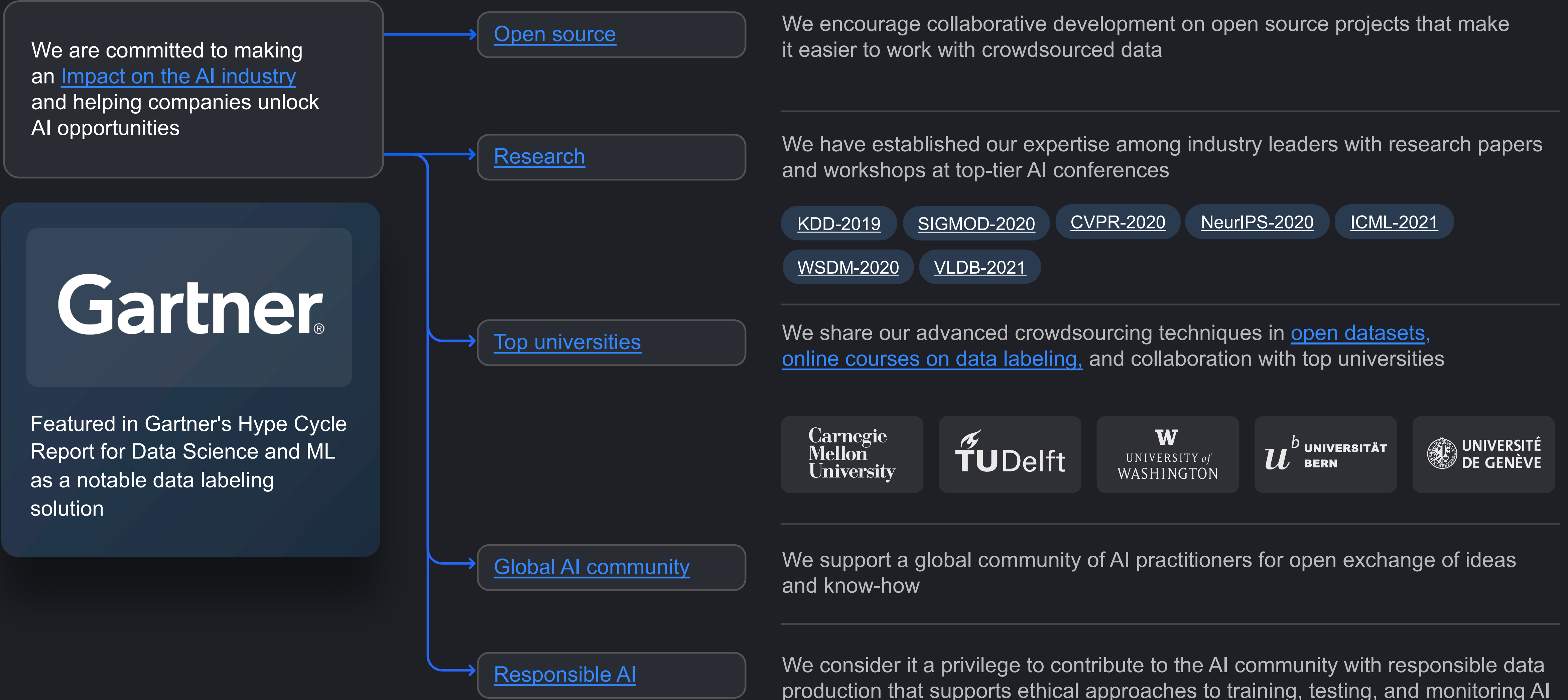
Open source

We encourage collaborative development on open source projects that make it easier to work with crowdsourced data

Research

We have established our expertise among industry leaders with research papers and workshops at top-tier AI conferences

KDD-2019   SIGMOD-2020   CVPR-2020   NeurIPS-2020   ICML-2021

WSDM-2020   VLDB-2021

Top universities

We share our advanced crowdsourcing techniques in [open datasets,](#) [online courses on data labeling,](#) and collaboration with top universities

Carnegie Mellon University   **TU**Delft   UNIVERSITY of WASHINGTON u^b UNIVERSITÄT BERN   UNIVERSITÉ DE GENÈVE

Global AI community

We support a global community of AI practitioners for open exchange of ideas and know-how

Responsible AI

We consider it a privilege to contribute to the AI community with responsible data production that supports ethical approaches to training, testing, and monitoring AI

✳ Toloka

Featured in top media outlets

Creating responsible AI products using human oversight

Case studies   Jan 28, 2021

Training Data: The Overlooked Problem Of Modern AI

Data Science   Jan 28, 2021

Why businesses take a month or more to deploy ML models and what you can do

Crowdsourcing   Jan 28, 2021

UNESCO-Backed IRCAI Names Northeastern's and UNAM's Civic AI Research —in Collaboration with

Company news   Jan 28, 2021

See all

Toloka

Get in touch

www.toloka.ai

LinkedIn

Twitter

GitHub

Join our Slack community

Toloka