InstaDeep
AIchor Platform

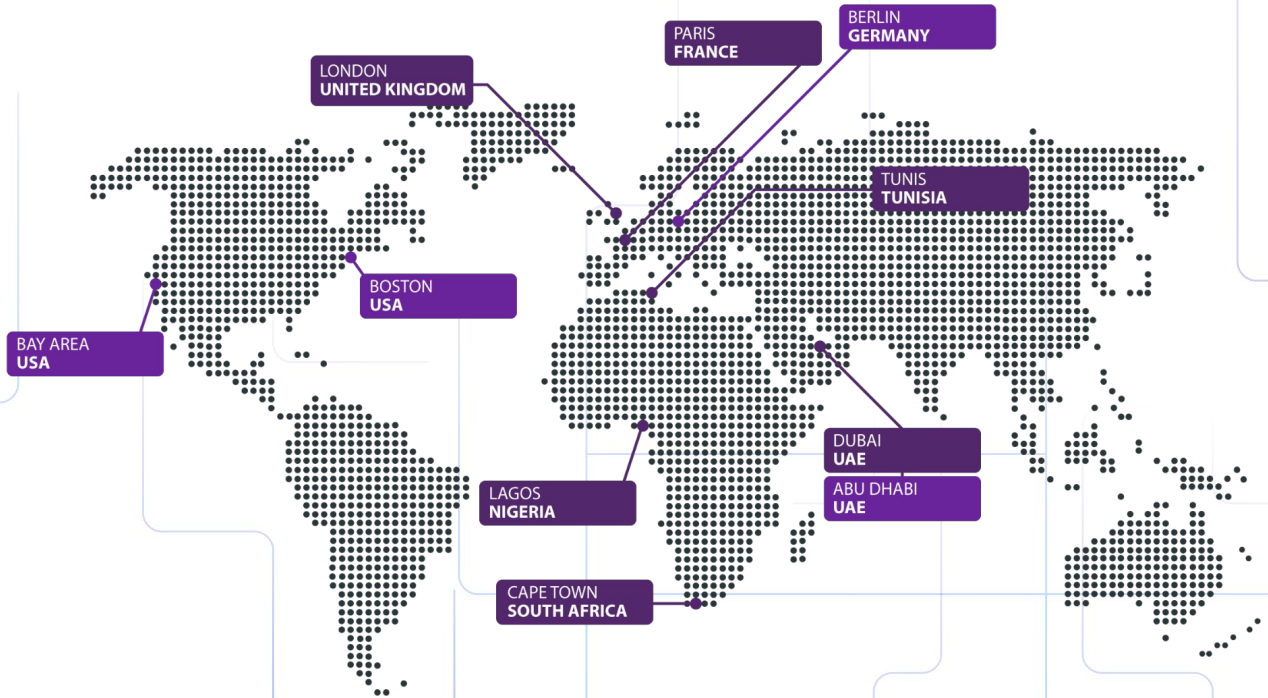
Azure

InstaDeep™

Agenda

- Instadeep - Our mission
- Alchor Platform - High level overview
- Use cases
- Pricing model
- Usual next steps / Q&A

InstaDeep: Leader in Decision-Making AI



InstaDeep™

InstaDeep: Leader in Decision-Making AI

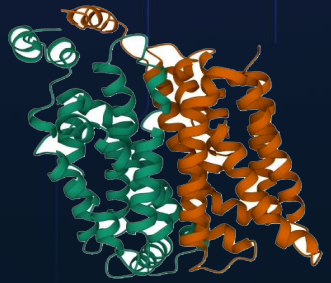
450+ **AI Research, Engineering & BioAI**
ML Engineers, Research Engineers, Research Scientist, Computational Biologists

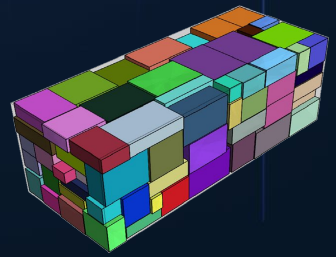
35+ **MLOps**
HPC Engineers, DevOps Specialists, SREs, Data Engineers, Info Sec

35+ **Insights & Visualisation**
UX/UI Designers, Front & back-end software engineers & QA Engineers

★ **Historical expertise with**

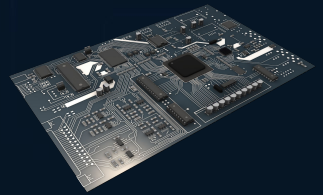
- Bio (Design new candidate cures and vaccines)
- HW (Route complex printed circuit boards rapidly)
- Logistic (Pack items more efficiently)
- Fleet Mgt (Optimize train scheduling and mobility)





★ **Annual Internal usage**

- 3M hours of CPU
- 130k hours of GPU (A100)
- 80 000 experiments



InstaDeep™

The AI training platform providing everything but AI !

InstaDeep™

AI teams: A very specific world to manage

Business pains

- **Uncontrolled cost** management
- **Inefficient** compute **resource utilization**
- **Inconsistent** workflows and processes
- **Heavy reliance** on **ML/DevOps** teams
- **Frequent missed** project **deadlines**

Technical constraints

- Tooling **Complexity** and **Overload**
- **Compute** Resources **Availability**
- **Exposure** to **Security Vulnerabilities**
- **Scalability challenges** in handling growing workloads

InstaDeep™

AIchor™ - Benefits

Cost Control

- **Optimized** resources **allocation preventing over-provisioning**
- **Visibility** into **resource consumption** for better **budgeting** and **forecasting.**

Simple & Effective

- **Remove** infrastructure **constraints** allowing **users** to focus only on **AI**
- **Streamline** team's **process** while ensuring **shared** working methods

Security & Reliability

- **Role-based** solution ensuring **strict security control**
- **Containerized** environment for **secure, isolated** projects

AIchor™ - Distributed AI Workloads Manager

AIchor supercharges AI/ML workloads with efficient resource allocation and scalable infrastructure, optimizing training performance. AIchor empowers teams to fine-tune models faster and deploy AI solutions with ease.

Git Integration and CLI Access

Researchers and Engineers can trigger experiments and trainings with a **single command-line**.

Experiment Management

Manage, track, and compare multiple experiments within a unified interface

Integration with ML Frameworks

Integrate with popular ML frameworks leveraging the use of dedicated operators.

Prototyping to Training

Easily scale your projects from initial Jupyter notebooks to advanced distributed training environments.

Infrastructure Management

AIchor runs on top of a **public cloud infrastructure** and **on-premise environments.**

Security and Compliance

Safeguard sensitive data and ensure compliance with industry standards with dataplane segregation.

InstaDeep™

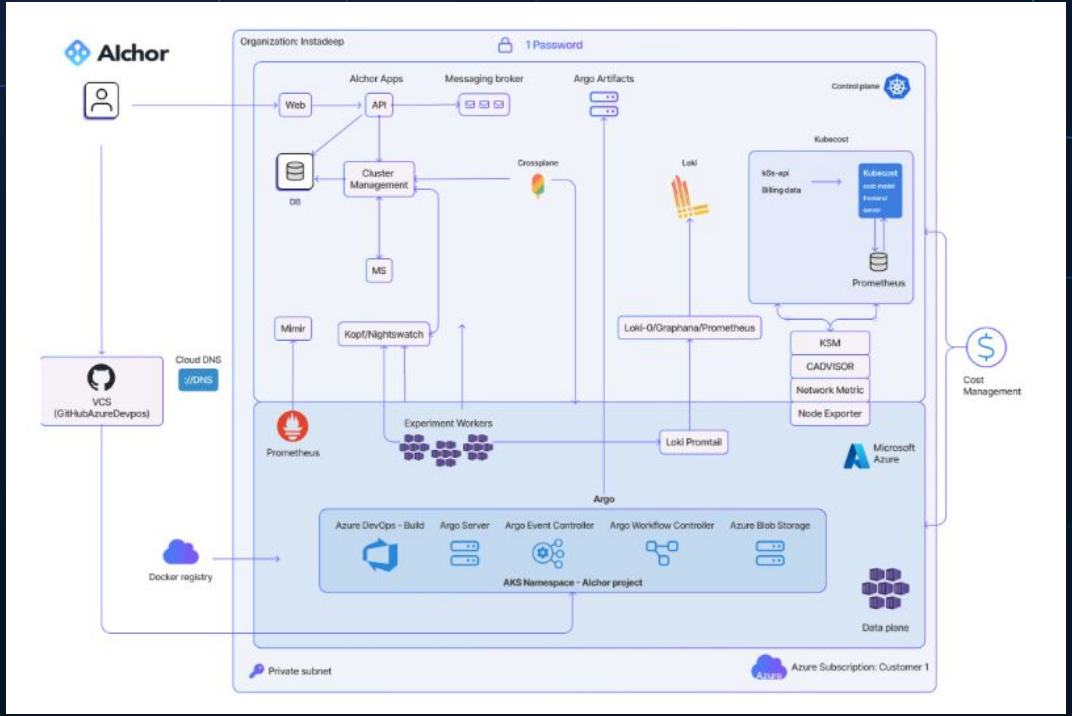
AIchor Platform: Architecture

On demand AI training:

Leveraging our in-house built tools and know-how, AIchor is an event driven automation ecosystem for running AI Workloads.

By simply providing a code repo and a manifest. A VCS Push event will trigger large scale experiments on a cluster.

The AIchor platform provides all resources needed to run, log and manage the runs in a self contained environment removing overhead of setting up and managing an infrastructure.



InstaDeep™

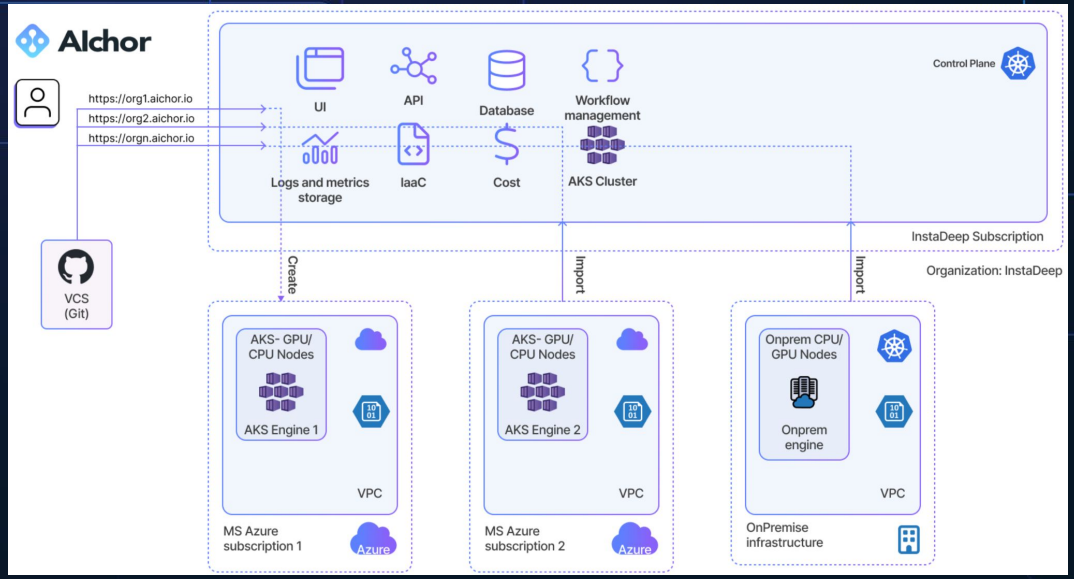
AIchor Platform management

Engines:

Workloads can run on EKS engines on customers subscriptions: AKS clusters can be **created** or **imported** by AIchor; AND On Premise kubernetes clusters which can be **imported** to AIchor.

Access to AIchor is restricted to users defined on Git repositories and administrators defined at the organization level, **SSO is supported**.

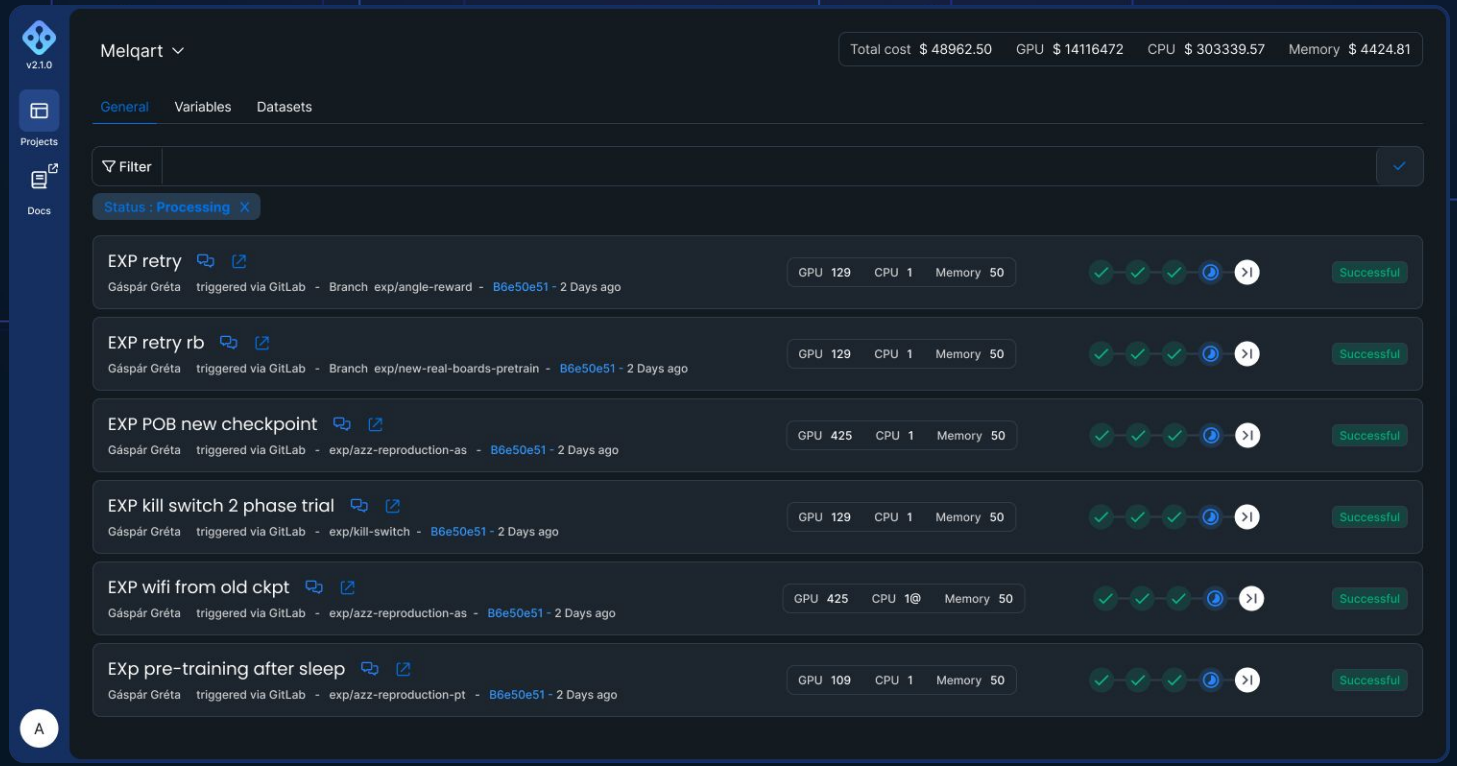
Data is stored in Azure Blob Storage and Persistent volumes can be created for a faster and less expensive interaction during experiments.



MS Azure managed services that AIchor leverages are: **AKS, Blob storage, Azure devOps, Azure Postgres SQL, ACR**.

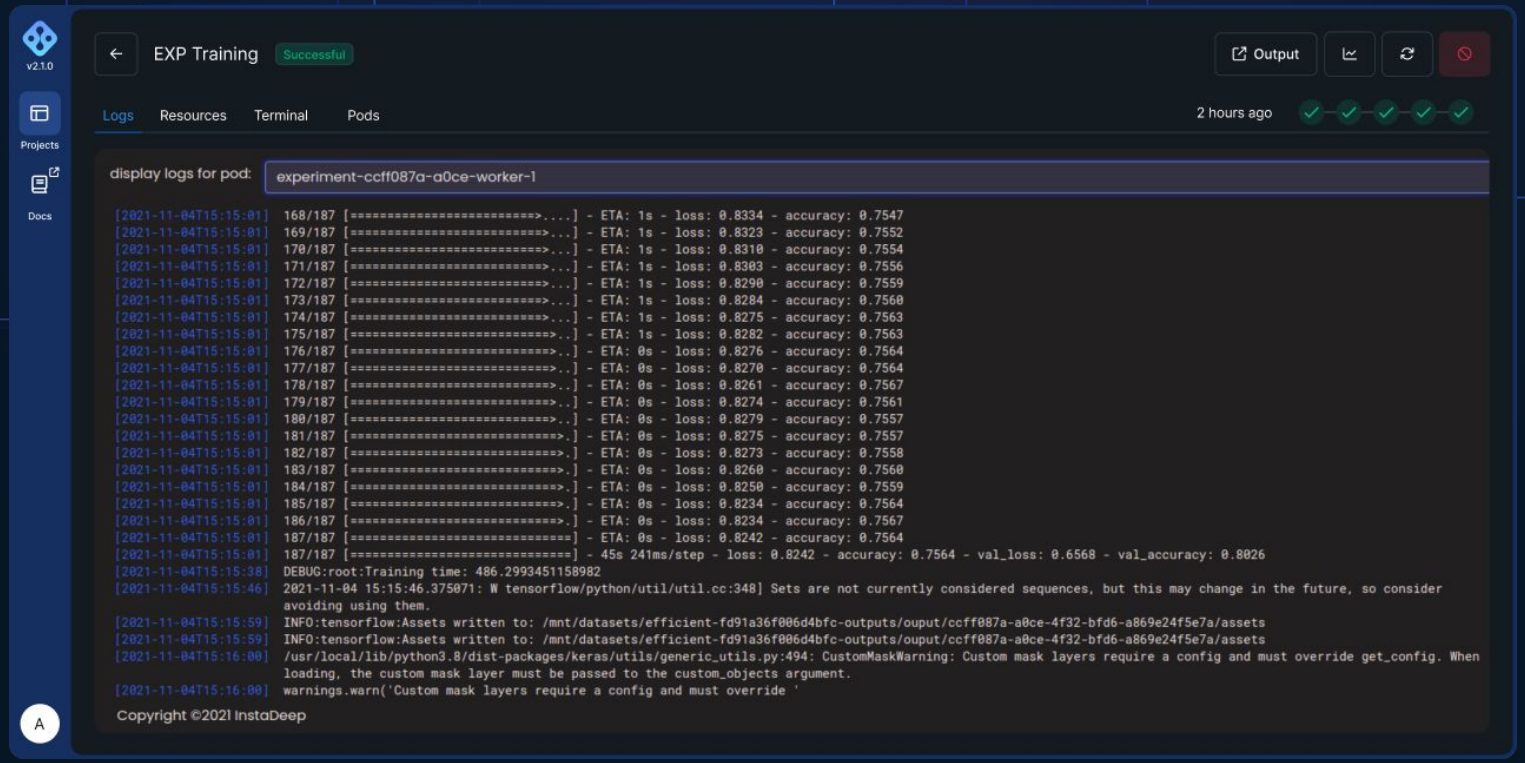
InstaDeep™

AIchor Platform: Workloads Management (1/3)



InstaDeep™

AIchor Platform: Workloads Management (1/3)

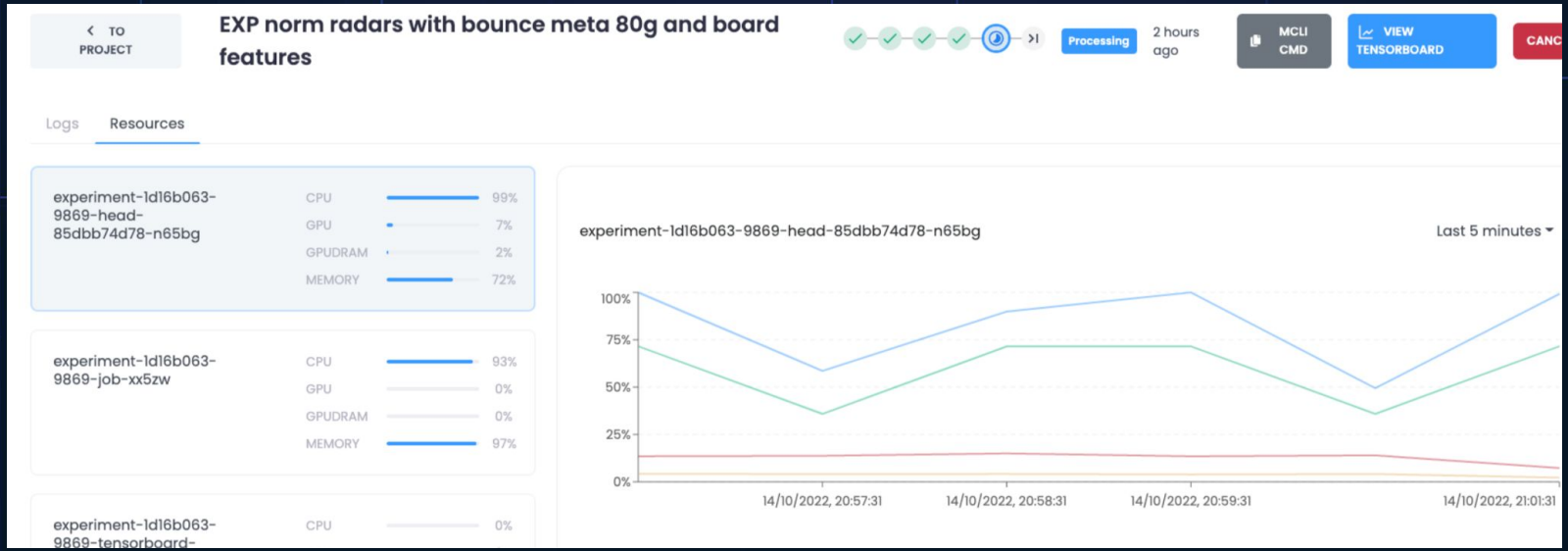


InstaDeep™

AIchor Platform: Workloads Management (3/3)

Platform capabilities

Ability to Monitor real-time resource utilization from the user interface



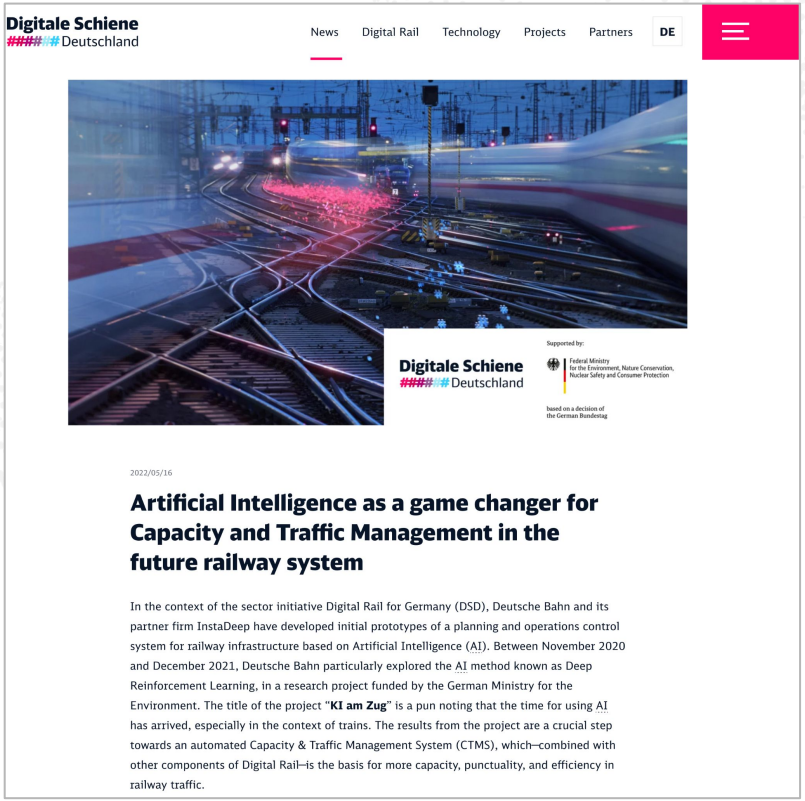
InstaDeep's AI-Powered Solution to Boost Deutsche Bahn's Efficiency

Problem and Motivation: Deutsche Bahn grapples with delays and inefficiencies due to manual rail management in dense traffic, causing customer dissatisfaction and increased costs.

Goal: Revolutionize rail operations by implementing Automatic Train Operation (ATO), and employing AI-driven Capacity and Traffic Management Systems (CTMS) for increased capacity and reduced delays.

Technology: Multi-Agent Reinforcement Learning (RL) using AIchor for automated schedule construction, live re-dispatching, and decision-making in a simulated railway environment.

Value Proposition: Target of 25% reduction in high-speed train delays to 2021 baseline. Anticipated annual savings of more than EUR 65 million.



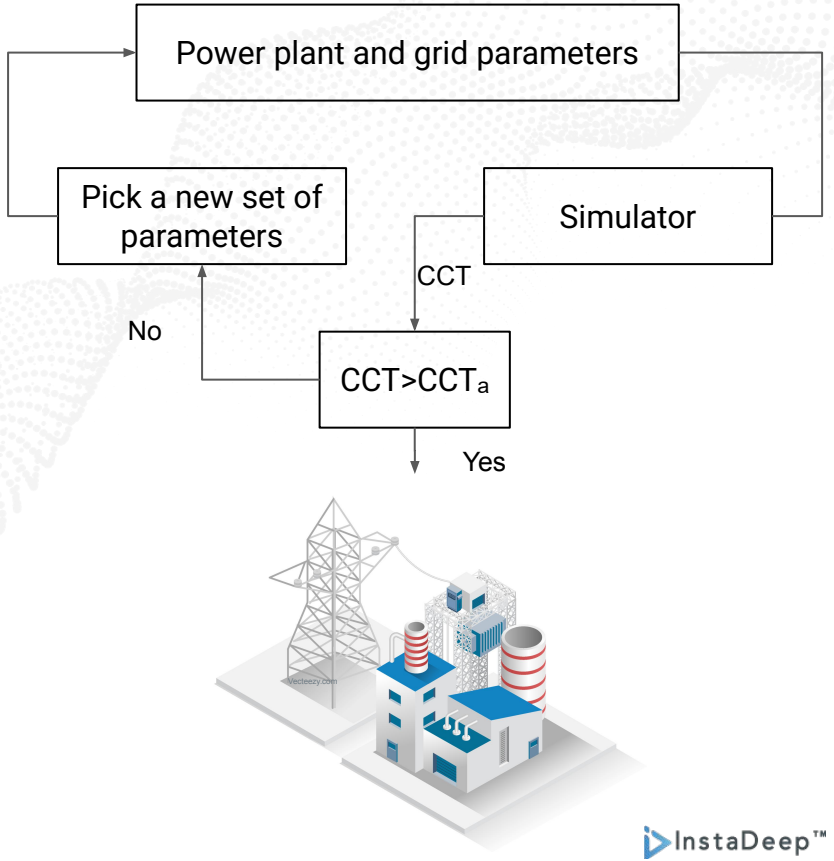
Fast-Track Power Plant Integration with ML

Problem Statement: Commissioning new power plants comes with **challenges in optimizing control design parameters** for safe grid connection, with **traditional methods being slow and resource-intense**.

Objective / Opportunity: Increase the number of search iterations for control design parameter optimization. **Achieve cost savings** in design changes and **ensure seamless integration of new power plants** into the grid.

Technology: Advanced **ML models trained through AIchor leveraging data** to generate **insights** to streamline the commissioning process and ensure optimal grid connection parameters are identified more efficiently.

Value Proposition: Cost savings in design modifications necessary for **safe integration** of power plants into the grid. **Shorten the commissioning timeline** and re**duce the risk of integration issue.**,



AIchor™

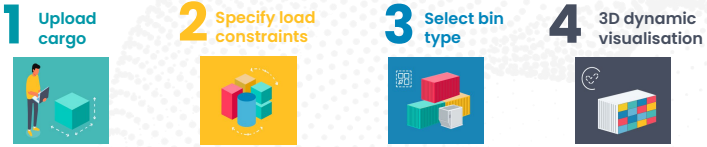
AI-Powered 3D Load Planning Tool for Containers/Trailers/ULDs

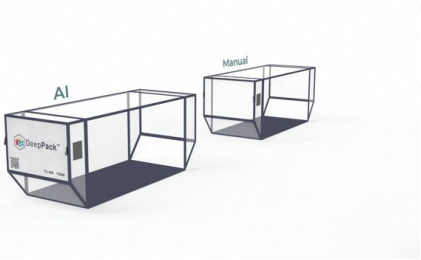
Problem and Motivation: Load planning is a slow process and not optimised for complex and varying loads. Human planners with required experience are increasingly scarce commodity.

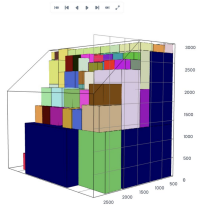
Goal: Maximize cargo space utilization & generate optimal load plans fast with minimal planning time and effort.

Technology: Reinforcement Learning algorithm using AIchor for optimal load planning respecting any operational and regulatory constraints, flexibly, despite varying shapes, sizes & constraints.

Value Proposition: Measurable volume yield optimisation translating to enhanced revenue management and transportation cost savings for logistics companies.







Smart, self-learning tool using AI

Handle **complex shapes**

API enabled

Collaborative workspace

InstaDeep™

Infrastructure models

CSPs

Come with your AKS cluster

- Quick & Compatible
- 0 invest
- Marketplace ready

SaaS

Just launch your model training

- 0 management
- 0 invest
- Scalability++

OnPrem

Use your on Premise cluster

- Simple approach
- Infra compatibility
- Security++

Pay as you go approach

Pricing

AIchor pricing model *

Consumption	Gold	Platinum	Diamond
CPU (1 hour)	$0.020	$0.015	$0.010
GPU (1 hour)	$1.000	$0.900	$0.800
RAM (1 Go)	$0.001	$0.001	$0.001
CPU / Month (hour)	< 10k	10k to 500k	> 500k

(*) Unit costs above are subject to change depending on infrastructure model

Usual process

- Cost review / estimation
- Free Trial (Small / Medium customer)
- PoC (1 month)
  - Success Metrics
  - Deadline
  - Decision criteria
- Legal / paper discussion

InstaDeep™

Thank you!

InstaDeep™

InstaDeep™