

InstaDeep

Alchor Platform

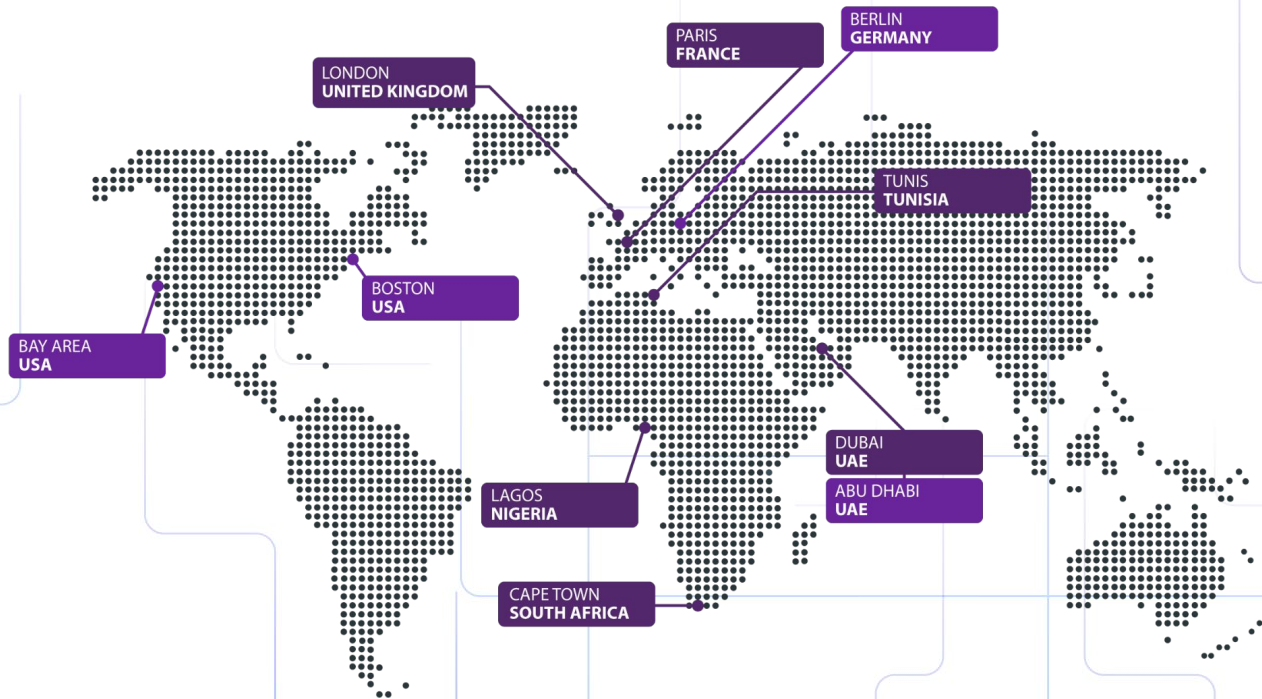


Alchor

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: 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+

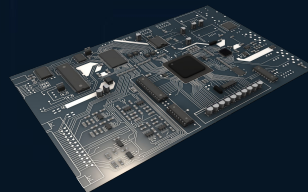
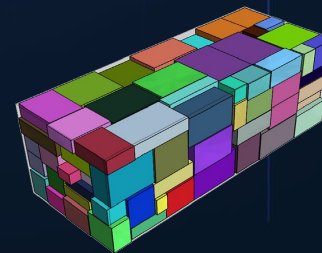
Insights & Visualisation
UX/UI Designers, Front & back-end software engineers & QA Engineers

★ 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





Alchor

The AI training platform providing everything but AI !

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



Alchor™ - 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



Alchor™ - Distributed AI Workloads Manager

Alchor supercharges AI/ML workloads with efficient resource allocation and scalable infrastructure, optimizing training performance. Alchor 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**.



Prototyping to Training

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



Experiment Management

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



Infrastructure Management

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



Integration with ML Frameworks

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



Security and Compliance

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

Alchor Platform: Architecture

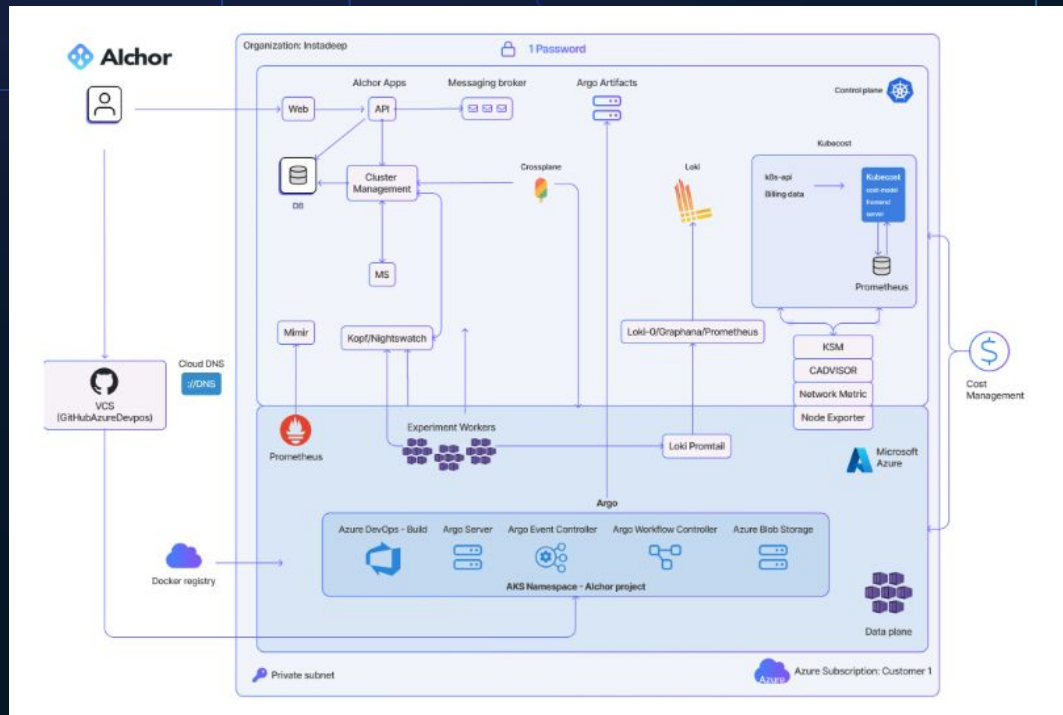
On demand AI training:

Leveraging our in-house built tools and know-how, Alchor 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 Alchor 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.



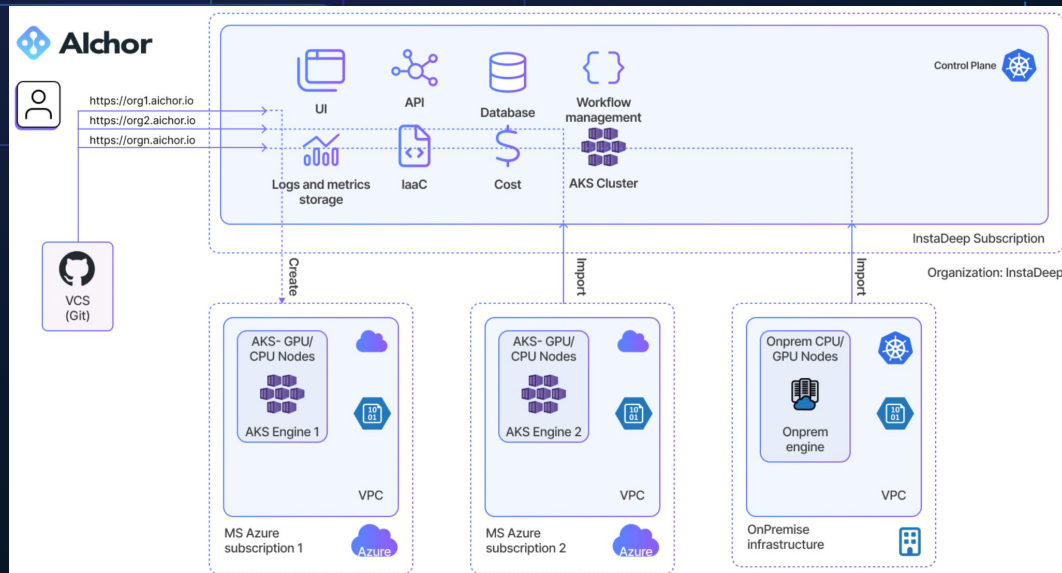
Alchor Platform management

Engines:

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

Access to Alchor 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 Alchor leverages are: **AKS, Blob storage, Azure devOps, Azure Postgres SQL, ACR.**

Alchor Platform: Workloads Management (1/3)

The screenshot displays the Alchor Platform interface for managing workloads. The top navigation bar shows the project name 'Melqart' and a dropdown menu. To the right, a summary of total costs is provided: Total cost \$ 48962.50, GPU \$ 14116472, CPU \$ 303339.57, and Memory \$ 4424.81. Below this, there are tabs for 'General', 'Variables', and 'Datasets'. A search filter is present, and a status filter is set to 'Processing'. The main area lists several jobs, each with a title, trigger information, resource usage (GPU, CPU, Memory), a progress indicator, and a 'Successful' status label.

Job Name	Trigger	Branch	GPU	CPU	Memory	Status
EXP retry	triggered via GitLab	exp/angle-reward	129	1	50	Successful
EXP retry rb	triggered via GitLab	exp/new-real-boards-pretrain	129	1	50	Successful
EXP POB new checkpoint	triggered via GitLab	exp/azz-reproduction-as	425	1	50	Successful
EXP kill switch 2 phase trial	triggered via GitLab	exp/kill-switch	129	1	50	Successful
EXP wifi from old ckpt	triggered via GitLab	exp/azz-reproduction-as	425	1@	50	Successful
Exp pre-training after sleep	triggered via GitLab	exp/azz-reproduction-pt	109	1	50	Successful

Alchor Platform: Workloads Management (1/3)

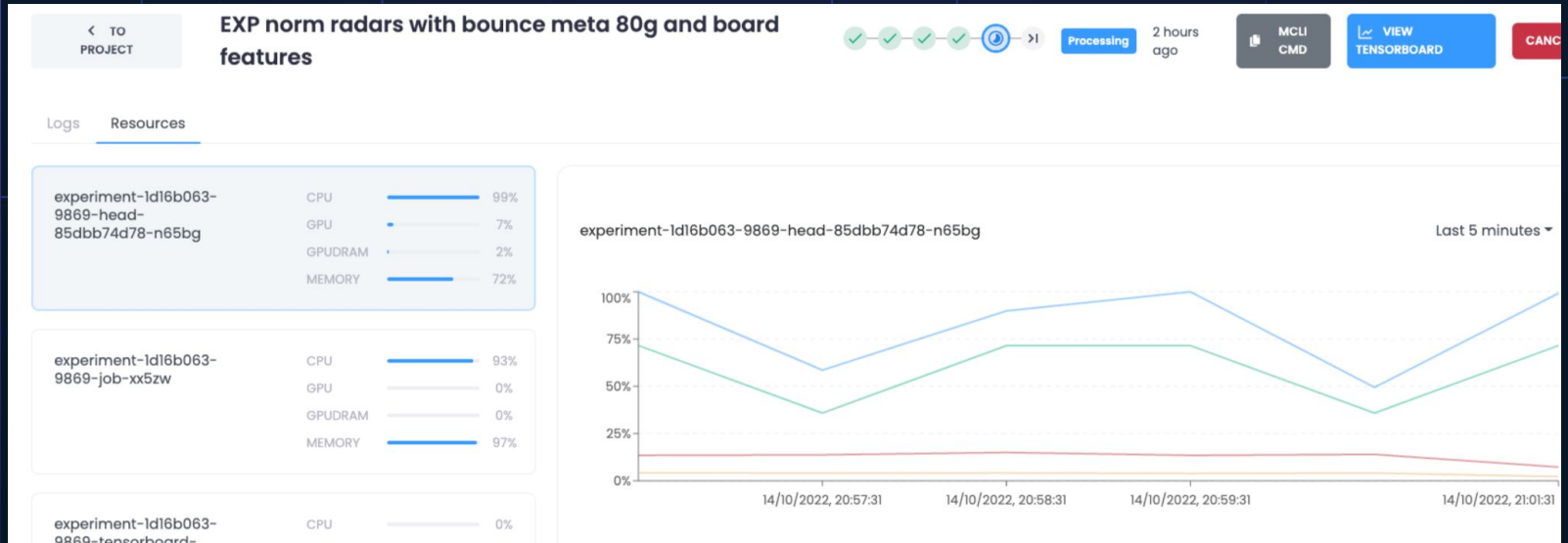
The screenshot displays the Alchor Platform interface. At the top, a navigation bar shows 'EXP Training' with a 'Successful' status indicator. On the right, there are buttons for 'Output', a search icon, a refresh icon, and a close icon. Below the navigation bar, a secondary bar contains 'Logs Resources Terminal Pods' and a timestamp '2 hours ago' followed by five green checkmarks. The main area is a terminal window titled 'display logs for pod: experiment-ccff087a-a0ce-worker-1'. The terminal output shows a series of training progress lines, each with a timestamp, a step number (e.g., 168/187), and performance metrics: ETA, loss, and accuracy. The final line shows 'DEBUC:root:Training time: 486.2993451158982' and a warning message: 'WARNING:tensorflow:Assets written to: /mnt/datasets/efficient-fd91a36f086d4bfc-outputs/ouput/ccff087a-a0ce-4f32-bfd6-a869e24f5e7a/assets /usr/local/lib/python3.8/dist-packages/keras/utils/generic_utils.py:494: CustomMaskWarning: Custom mask layers require a config and must override get_config. When loading, the custom mask layer must be passed to the custom_objects argument. warnings.warn('Custom mask layers require a config and must override '.

Copyright ©2021 InstaDeep

Achor 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 Alchor 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.

Digitale Schiene
Deutschland

News Digital Rail Technology Projects Partners DE

Supported by:
Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection

Based on a decision of the German Bundestag

2022/05/16

Artificial Intelligence as a game changer for Capacity and Traffic Management in the future railway system

In the context of the sector initiative Digital Rail for Germany (DSD), Deutsche Bahn and its partner firm InstaDeep have developed initial prototypes of a planning and operations control system for railway infrastructure based on Artificial Intelligence (AI). Between November 2020 and December 2021, Deutsche Bahn particularly explored the AI method known as Deep Reinforcement Learning, in a research project funded by the German Ministry for the Environment. The title of the project "KI am Zug" is a pun noting that the time for using AI has arrived, especially in the context of trains. The results from the project are a crucial step towards an automated Capacity & Traffic Management System (CTMS), which—combined with other components of Digital Rail—is the basis for more capacity, punctuality, and efficiency in railway traffic.

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-intensive.



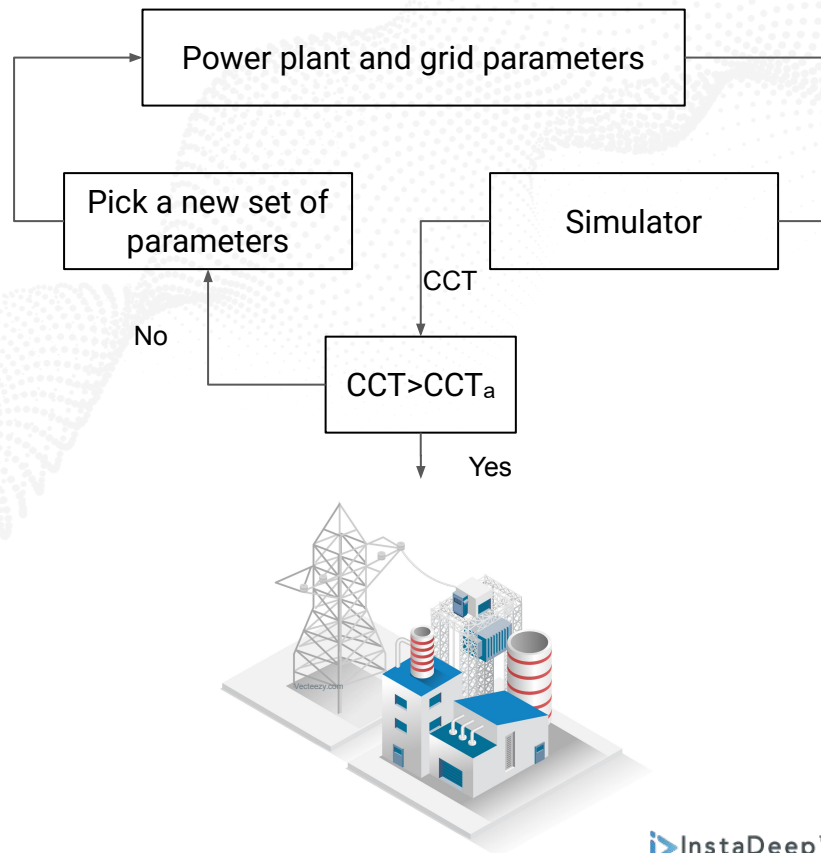
Objective / Opportunity: Increase the number of search iterations for control design parameter optimization. Achieve cost savings in design changes and ensure seamless integration of new power plants into the grid.



Technology: Advanced ML models trained through Alchor 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 reduce the risk of integration issue.,



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.



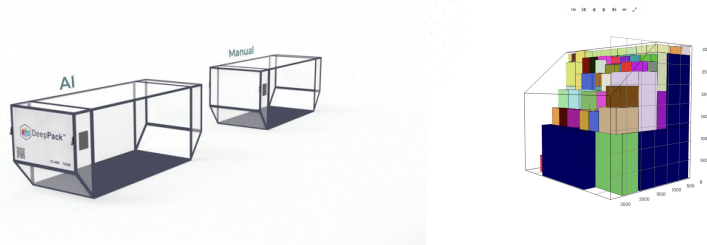
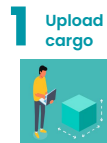
Goal: Maximize cargo space utilization & generate optimal load plans fast with minimal planning time and effort.



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

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

Anchor 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

Thank you!

