1st Digital Edition
2026

HammerHAI

HammerHAI

Germany's AI Factory for Manufacturing, Engineering, and Research

First Edition of User and Success Stories



HammerHAI Success Stories
1st Digital Edition

Publisher
HLRS on behalf of the HammerHAI Consortium
Nobelstraße 19
70569 Stuttgart, Germany
info@hammerhai.eu

Published
September 2026

Copyright
© 2026 Members of the HammerHAI Consortium. All rights reserved.

Editor:
HLRS on behalf of the HammerHAI Consortium

Content contributions
Success stories, images, and logos provided by members of the HammerHAI Consortium and participating users.

The acquisition and operation of the EuroHPC Al-optimised supercomputer is funded jointly by the EuroHPC Joint Undertaking, through the European Union's Digital Europe Programme, as well as by the German Federal Ministry of Research, Technology and Space (BMFTR) and the Baden-Württemberg Ministry of Science, Research and the Arts.
Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the EuroHPC JU. This project has received funding from the European High Performance Computing Joint Undertaking under grant agreement No. 101234027. It is co-funded by the European Commission, the German Federal Ministry of Research, Technology and Space (BMFTR), the Baden-Württemberg Ministry of Science, Research and the Arts, the Bavarian State Ministry of Science and the Arts and the Lower Saxony Ministry of Science and Culture.

HammerHAI

Germany's AI Factory for
Manufacturing, Engineering, and Research



Index

About us 6

The Consortium 7

Success Stories

Accelerating Real-Time Shopfloor Intelligence with HammerHAI. 8

Accelerating AI Robot Planning with Large Language Models 10

GenCAD – Foundation Model for Engineering Design Generation 12

Securing AI across the Machine Learning Operations chain: Piloting cybersecurity and data sovereignty of AI in supercomputing. 14

Testing a RAG Prototype with Real Engineering Questions 16

User Stories

Bridging Simulation and Reality for Safer Autonomous Driving. 18

JAWS: Anticipating Industrial Operating Conditions from Point-of-View Video. 20

Training Europe's Construction Embedding AI with HammerHAI. 22

List of Abbreviations 24

Further Information 25

About us

Established by the EuroHPC Joint Undertaking under its AI Factories initiative, HammerHAI is helping to accelerate the artificial intelligence revolution in Europe. Supporting start-ups, SMEs, industrial companies, and academic research — particularly in manufacturing and engineering fields — HammerHAI provides a comprehensive portfolio of AI computing infrastructure, services, and tools. These resources are empowering organizations to more quickly seize the opportunities that emerging AI technologies now offer.

Going into service in early 2027, the HammerHAI supercomputer is a cost-efficient option for European organizations whose AI computing needs exceed their in-house capabilities. With up to 15 exaflops of AI inference performance and cloud-like usability, it can support medium- to large-scale workloads for model training, inference, and data analytics. Located at the High-Performance Computing Center Stuttgart, it is operated securely and in compliance with European data protection regulations.

HammerHAI also offers professional training courses, consulting and strategic planning services, software tools, and data resources, as well as personalized support from our team of AI experts. Regardless of the maturity of an individual application, HammerHAI can provide scalable solutions, addressing challenges at all stages of the AI lifecycle.

The following User and Success Stories highlight innovative applications of artificial intelligence, machine learning, and data analytics being developed with HammerHAI's support. The stories give insights into our users' needs and experiences, including their goals, needs, and progress achieved. Together, they illustrate the value and impact of HammerHAI's resources and services.



The Consortium

HLRS — High-Performance Computing Center Stuttgart

Leads HammerHAI and will host the AI-optimized supercomputer, including computing and storage resources, software, and secure access.

KIT — Karlsruher Institut für Technologie

One of Europe's leading research institutions, contributing expertise in trustworthy AI, data-driven methods, and innovation for SMEs.

SICOS — Wir vermitteln Höchstleistung

Supports companies, especially SMEs, in adopting simulation, HPC, data analytics, and AI through consulting, training, and technology transfer.

GWDG — Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Provides HPC services and AI infrastructure, with a strong focus on secure AI solutions for research, SMEs, and startups.

lrz — Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

A leading German supercomputing centre providing supercomputing resources, AI expertise, user support, and training across disciplines for research and science.

Accelerating Real-Time Shopfloor Intelligence with HammerHAI

Highlights

- **Reduced model training time from 6–8 hours to less than one hour.**
- **Supported AI applications in manufacturing, robotics, and digital twins.**
- **Enabled industrial testing of AI-based shopfloor intelligence.**

Challenges

ARENA2036 stands for „Active Research Environment for the Next generation of Automobiles“ and is one of nine research campuses of the „Research Campus - Public-Private Partnership for Innovation“ funding initiative in Germany. ARENA2036 is supported by the Federal Ministry of Research, Technology and Space and is run as a registered association with members from science and industry. These members are active in various disciplines – from the automotive industry to aerospace technology, textile and materials research to ergonomics.

Modern manufacturing environments generate vast amounts of spatial data that must be processed in real time to support robotics, autonomous systems, and digital twins. At ARENA2036, we developed a deep-learning-based Shopfloor Model capable of mapping, segmenting, and classifying industrial environments. However, training these large neural networks on local GPU workstations required overnight computation (6–8 hours), significantly slowing research iterations and limiting rapid model optimization.

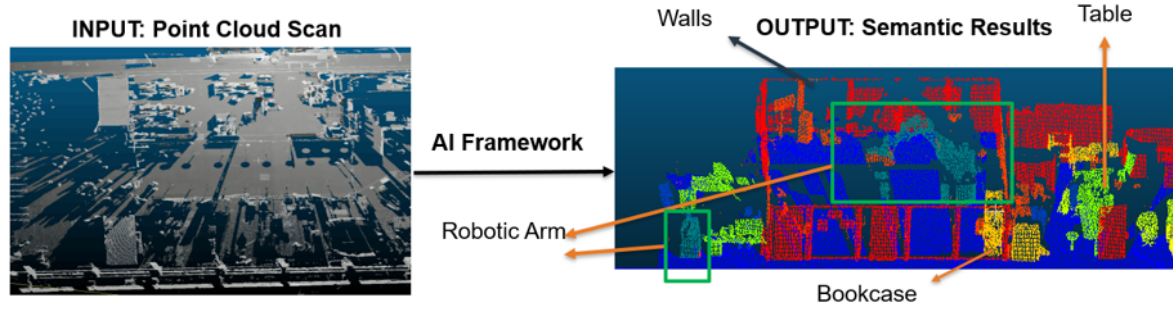
Solution

Through the HammerHAI project, ARENA2036 gained **access to the HLRS Hunter supercomputer, enabling the migration of our training pipeline** from sequential local execution to a highly optimized parallel multi-node supercomputing environment. Hunter's scalable computing resources accelerated deep learning workflows while maintaining reproducible training across large datasets. HammerHAI provided the computational foundation that made this acceleration possible.

Benefits and Impact

The accelerated Shopfloor Model now serves as a key building block for intelligent manufacturing, robotic navigation, and digital twin applications. Through AI-MATTERS, these technologies are being transferred into industrial testing and experimentation services, allowing companies to evaluate AI-based shopfloor intelligence in realistic production environments.

Using HLRS Hunter, model training time was reduced from 6–8 hours to less than one hour, enabling significantly faster experimentation and validation. Through HammerHAI, companies can access the required HPC capabilities without investing upfront in dedicated computing infrastructure. Combined with AI-MATTERS, the solution can be tested in realistic manufacturing environments, reducing the technical and financial risks of industrial AI adoption.



Organizations Involved

Contact

Muhammad Saeed, ARENA2036, muhammad.saeed@arena2036.de

Accelerating AI Robot Planning with Large Language Models

Highlights

- **Reduced average inference latency from 9.03 seconds to 3.98 seconds.**
- **Achieved more than 2× higher throughput.**
- **Enabled benchmarking and scalable testing of open-source LLMs for industrial robotics.**

Challenges

ARENA2036 stands for „Active Research Environment for the Next generation of Automobiles“ and is one of nine research campuses of the „Research Campus - Public-Private Partnership for Innovation“ funding initiative in Germany. ARENA2036 is supported by the Federal Ministry of Research, Technology and Space and is run as a registered association with members from science and industry. These members are active in various disciplines – from the automotive industry to aerospace technology, textile and materials research to ergonomics.

Programming industrial robots traditionally relies on manually designed rule-based systems that are difficult to scale. Large Language Models offer a new paradigm by translating natural-language instructions into executable robot actions, but evaluating hundreds of planning tasks requires substantial computational resources.

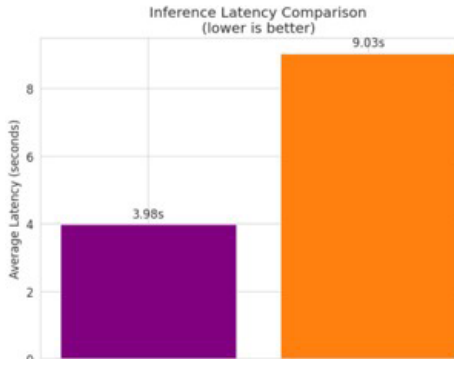
Solution

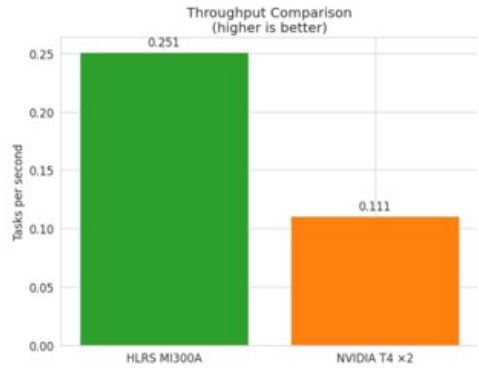
Through HammerHAI, ARENA2036 **leveraged the HLRS Hunter supercomputer to benchmark an LLM-based robot task planning framework** using DeepSeek-Coder. Hunter enabled large-scale evaluation across 90 manipulation tasks comprising 450 robot actions, providing the computational performance required for systematic benchmarking and optimization. HammerHAI provided the computational infrastructure that transformed experimental AI concepts into scalable industrial solutions.



Benefits and Impact

Using HLRS Hunter, average inference latency was reduced from 9.03 s to 3.98 s, with more than 2× higher throughput across a benchmark of 90 tasks and 450 robot actions. HammerHAI enables scalable evaluation of LLM-based robotic intelligence without requiring companies to build dedicated AI computing infrastructure. Through AI-MATTERS, these approaches can be tested in realistic industrial robotics environments before deployment.





Organizations Involved

Contact

Muhammad Saeed, ARENA2036, muhammad.saeed@arena2036.de

GenCAD – Foundation Model for Engineering Design Generation

Highlights

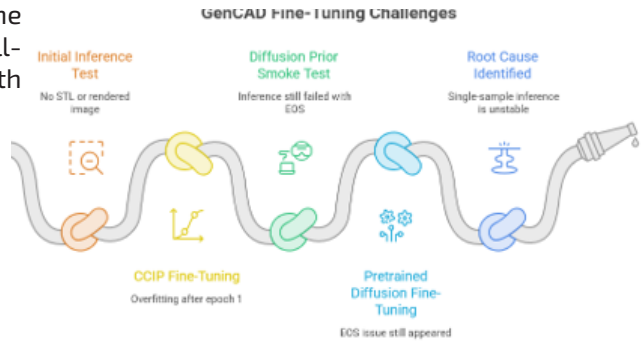
- **Accelerated development and training of a generative AI framework for Computer-Aided Design (CAD).**
- **Enabled training on large engineering datasets.**
- **Supported model optimization and robustness evaluation.**

Challenges

ARENA2036 stands for „Active Research Environment for the Next generation of Automobiles“ and is one of nine research campuses of the „Research Campus - Public-Private Partnership for Innovation“ funding initiative in Germany. ARENA2036 is supported by the Federal Ministry of Research, Technology and Space and is run as a registered association with members from science and industry. These members are active in various disciplines – from the automotive industry to aerospace technology, textile and materials research to ergonomics. Generative AI is rapidly transforming engineering design by enabling AI models to automatically generate, understand, and optimize CAD models from limited inputs such as sketches, point clouds, or natural-language descriptions. Training these large foundation models requires processing millions of geometric primitives and complex 3D representations, resulting in computational demands that exceed conventional workstation hardware.

Solution

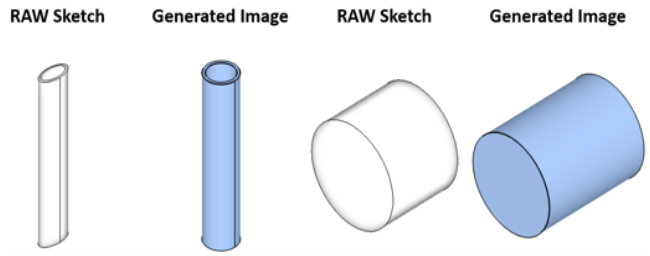
Through the HammerHAI project, ARENA2036 utilized the High-Performance Computing Center Stuttgart (HLRS) **Hunter supercomputer to accelerate the development and training of GenCAD,** a generative AI framework for computer-aided design. Hunter's scalable computing resources enabled efficient training of deep neural networks on large engineering datasets while significantly reducing model development time. The supercomputing environment also allowed extensive experimentation with model architectures, hyperparameter optimization, and large-scale benchmarking that would have been impractical using local GPU resources. HammerHAI transformed what had previously been resource-constrained experimentation into a scalable AI development workflow and provided the computational backbone that enabled this next generation of engineering AI research.

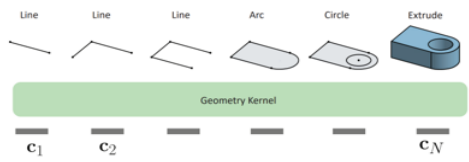


Benefits and Impact

GenCAD represents an important step toward AI-assisted engineering design and digital product development. The technology has strong potential for automated CAD generation, design optimization, digital engineering, and future industrial AI assistants. Through AI-MATTERS, these capabilities can be demonstrated and validated with industrial partners, accelerating the adoption of generative AI in manufacturing and product development. HammerHAI provided the computational backbone that enabled this next generation of engineering AI research.

Using HLRS Hunter, GenCAD training became more scalable, supporting larger engineering datasets and systematic evaluation of different model and design-generation strategies. HammerHAI enables companies to explore compute-intensive Generative AI for engineering without making major upfront investments in dedicated HPC infrastructure. Through AI-MATTERS, resulting Proofs of Concept can be tested with industrial partners before larger deployment decisions.





Organizations Involved

University of Stuttgart · HLRS · Swinburne · HammerHAI · AI matters

Contact

Muhammad Saeed, ARENA2036, muhammad.saeed@arena2036.de

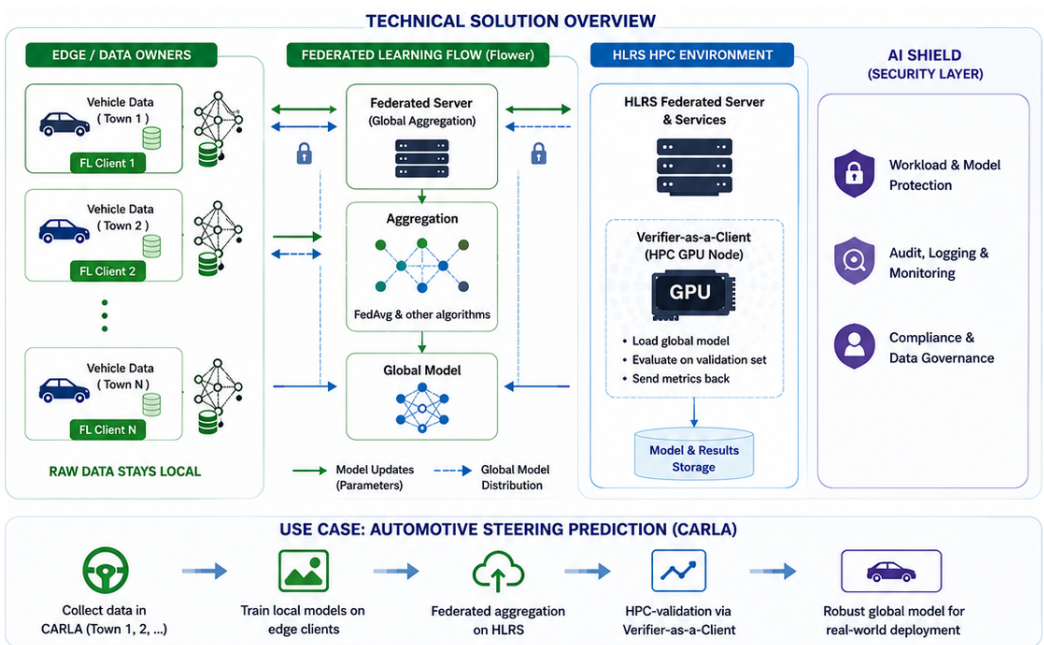
Securing AI across the Machine Learning Operations chain: Piloting cybersecurity and data sovereignty of AI in supercomputing

Highlights

- **Federated learning applied to steering angle prediction reduced the need for data transfers.**
- **The federated steering-angle model reduced estimated data network transfer approximately by 60%.**
- **Continuous model evaluation integrated as a "verifier-as-a-client" architecture.**

Challenges

During industrial partnerships, cross-organizational AI development often depends on large, distributed and sensitive datasets that cannot be freely transferred to a central supercomputing environment because of privacy, security, regulatory and bandwidth constraints. Centralized training typically requires moving substantial volumes of raw data, increasing network load, storage demand, computing time and operational cost, while also exposing AI workloads and models to additional cybersecurity risks. This can limit collaboration between organizations and slow the validation of AI models on supercomputing infrastructure.



Solution

Bosch Global Software Technologies GmbH (BGSG), located in Stuttgart, Germany, develops software, digital systems, and IT solutions for enterprises. Working with HammerHAI, BGSG tested an approach that combines federated learning and secure model validation to train automotive AI models without the need to exchange raw vehicle datasets.

Distributed clients trained local steering-angle models using the Flower open-source federated learning framework.
No data were exchanged among clients except for model parameters, which were aggregated into a global model. **HLRS GPU infrastructure provided through HammerHAI** was integrated into a customized validation workflow, which included a "verifier-as-a-client" concept for evaluating global models during training. The AI security platform "AI Shield" was deployed in the HPC environment to strengthen protection of AI workloads. The approach enabled collaborative model development among data owners and sources and created a reusable edge-to-HPC workflow for secure industrial AI development and validation.

Benefits and Impact

The project demonstrated that federated learning can combine knowledge from distributed automotive datasets without requiring the transfer of the underlying raw data.

Compared to a centralized approach, the federated steering-angle model prediction demonstrated significantly better generalization across operating environments, without a huge drop in accuracy. Estimated network traffic was reduced by approximately 60%, lowering bandwidth, storage and data-management requirements. Integration with **the HLRS supercomputing environment** established a reusable workflow for GPU-based model validation by internal teams, while AI Shield extended the approach toward securing AI workloads on HPC infrastructure.

This solution offers a scalable foundation that will make it easier for industrial partners to collaborate on the development and validation of secure AI models without exposing proprietary or sensitive data. It holds the additional benefit that reduced data movement can also decrease infrastructure load and associated energy consumption. Direct energy or monetary savings were, however, not measured in this project.

Contact

Rane Abhijeet, Bosch Global Software Technologies GmbH,
Abhijeet.Rane@in.bosch.com

Testing a RAG Prototype with Real Engineering Questions

Highlights

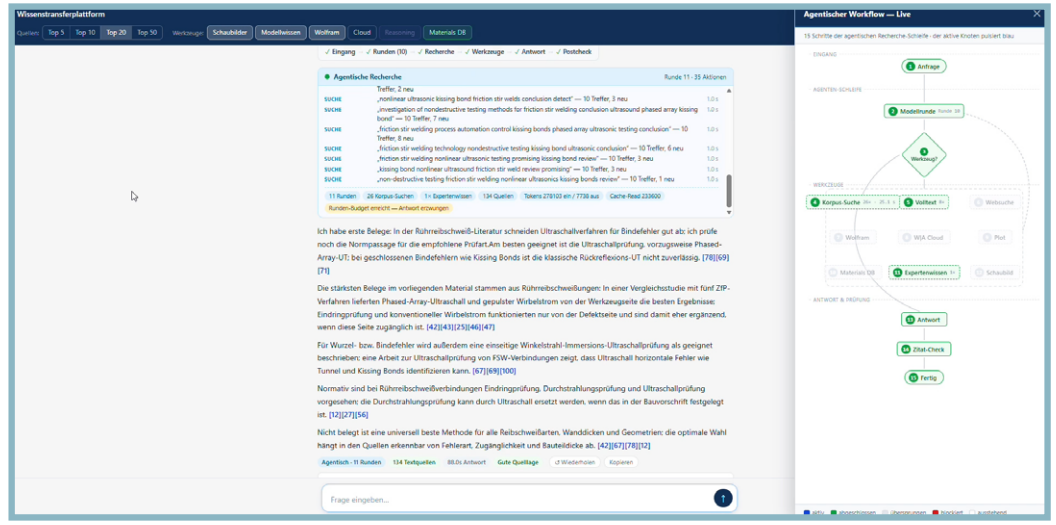
- Developed a retrieval-augmented generation (RAG) prototype
- RAG prototype uses 7,282 MPA publications on welding and joining
- MPA engineers write benchmark questions and evaluate answers and sources

Challenges

The Material Testing Institute (MPA) at the University of Stuttgart combines materials testing, joining research, and industrial transfer. Its knowledge is spread across publications, material tables, formulas, standards, and expert experience. A retrieval-augmented generation (RAG) system — a large language model (LLM) that retrieves and uses information from external knowledge sources — could improve access to this information. However, fluent answers alone do not guarantee engineering correctness. Relevant sources must be retrieved and cited traceably, calculations must be reliable, and uncertainties must remain transparent. Earlier work by the High-Performance Computing Center Stuttgart (HLRS) and the University of Stuttgart's Institute for Metal Forming Technology (IFU) showed that language models can partially predict temperature-dependent yield curves.

Solution

The collaboration with HammerHAI began with a **roadmapping workshop** that identified a benchmark test for a RAG system and initiated the development of a RAG prototype. **MPA and HLRS ran benchmark tests** to see where the RAG prototype helps, where it fails, and how much expert review is required before industrial use could be considered. A living roadmap and weekly meetings now guide prototype reviews and planned tests. The RAG testbed works with 7,282 MPA Stuttgart publications on welding and joining. Its interface allows experiments with source counts, multi-step retrieval, source, page-level references, PDF links, calculations, plots, and a second verification pass. These are prototype options, not yet integrated or validated production features. MPA engineers develop representative questions and evaluate the answers, cited sources, and calculations using the benchmark tests developed by HammerHAI and MPA.



Benefits and Impact

The main result is an evaluation approach and early RAG prototype. The benchmark records assessments conducted by MPA engineers on technical correctness, source traceability, and reproducibility. An exploratory comparison of 11 openly available models illustrates the need for systemic evaluation. This solution has the potential to support future AI-assisted engineering workflows by improving access to specialized engineering knowledge, reducing information retrieval efforts, and supporting the selection of welding parameters to accelerate the calibration of welding machines.

Organizations Involved

Contact

Gerrit Koch, Material Testing Institute (MPA), University of Stuttgart, Gerrit.Koch@mpa.uni-stuttgart.de

Bridging Simulation and Reality for Safer Autonomous Driving

User Goals

- **Convert driving simulations into photorealistic video.**
- **Scale synthetic data generation to support verification and validation of autonomous vehicles.**

Current State and Key Challenges

cogniBIT is a Munich-based startup that develops cognitive traffic agents for the validation of automated driving and driver-assistance systems. Our software, cogniBOT, creates controllable and repeatable scenarios involving vehicles and other road users. The simulated agents behave based on insights from neuroscience, which enables them to model human perception, decision-making, and behavior more realistically than conventional traffic agents while also ensuring that cogniBOT remains computationally tractable for large-scale simulations. The scenarios are independent of any visualization platform.

In the project for which we use HammerHAI resources, we use the open-source autonomous driving simulation software CARLA to visualize our scenarios. We export synchronized RGB, depth, edge, and semantic-segmentation data. We then evaluate generative video models that transform CGI outputs into photorealistic dashcam-style footage while preserving the logic of the scenario being investigated. This allows us to obtain photorealistic footage of any scenario we wish, including the ones including safety-critical events. This addresses a key industry challenge, as safety-critical events are rare, costly, and difficult or impossible to reproduce in real-world testing. Photorealistic outputs could help autonomous vehicle (AV) developers test their perception and decision-making software against scenarios that might not occur during normal operational testing.

We require HammerHAI's computing infrastructure and AI expertise to overcome limitations of other available computing systems, including their limited computational power and insufficient GPU-memory capacity. Being able to use HammerHIA resources is essential for cogniBOT's ability to fulfil the objectives of this project, which is to obtain photorealistic videos from CGI inputs that maintain temporal consistency and visual fidelity.

Support from HammerHAI

Through HammerHAI, we seek **access to high-performance NVIDIA GPU infrastructure and expert support** in selecting, configuring, and evaluating generative video models. This work requires intensive computing resources that would be prohibitively expensive for us at standard commercial rates. We currently use NVIDIA Omniverse and Cosmos, making NVIDIA-based systems particularly suitable.

Intended Value and Impact

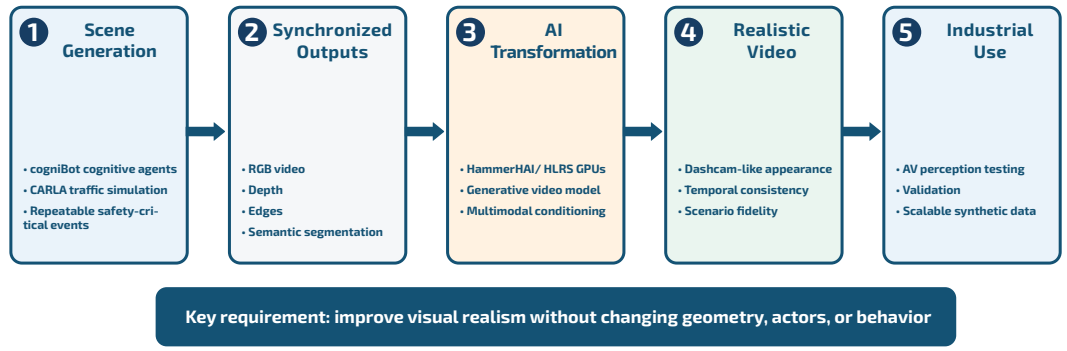
Our goal is to design a reliable pipeline that transforms controllable simulated scenarios into realistic sensor videos while preserving the underlying ground truth. This would let AV developers test perception an downstream driving functions on diverse, repeatable, safety-critical situations without depending solely on costly road recordings. Faster video generation and systematic quality metrics could shorten autonomous vehicle development cycles and improve confidence in simulation-based validation. This also addresses a very serious shortcoming of the EU-based AV companies.

Organizations Involved

Contact

Dr. Vahid S. Bokharaie, Chief Scientist at cogniBIT, vb@cognibit.ai

From Controllable Simulation to Photorealistic AV Test Data



JAWS: Anticipating Industrial Operating Conditions from Point-of-View Video

User Goals

- **Anticipate future scene graphs and process states from egocentric industrial video**
- **Fine-tune compact multimodal models that output validated, machine-readable results**
- **Enable real-time planning and decision-making for robotics and manufacturing**

Current State and Key Challenges

Ramblr GmbH is a German AI startup that builds models that understand the physical world to support robotics and industrial production.

Our technology turns egocentric video (i.e., video shot from an individual subject's perspective) into structured representations of objects, attributes, spatial relations, activities, and process states. A new technology currently under development at Ramblr—called JAWS—goes further, anticipating plausible short-term futures as scene graphs in validated, machine-readable form. These graphs can then be used for downstream planning.

We train with a production-tested stack based on PyTorch FSDP, Ray, and containerized workflows for distributed GPUs. However, we currently face the challenge of scaling. Fine-tuning multimodal models on industrial video data requires substantial computing power, fast connections between computing resources, reliable storage, and repeatable execution across multiple nodes. In addition, industrial data are not consistent across sites, processes, camera positions, and conditions. Models must learn robust representations while producing outputs in a format upon which machines can act.

For a startup, limited GPU access makes systematic experimentation and validation difficult. HammerHAI provides the infrastructure to move from smaller experiments to production-relevant training.

Support from HammerHAI

Through HammerHAI, we have been **using scalable NVIDIA GPU infrastructure, high-performance networking, and storage for distributed multimodal training at KISSKI (AI service center for critical and sensitive infrastructures)** in Göttingen. Technical onboarding and support for efficient multi-node execution have helped us to validate our FSDP, Ray, and container-based pipeline at production-relevant scale.

Intended Value and Impact

With JAWS we aim to deliver a compact multimodal world model that predicts scene graphs and process states from industrial video in near real time, targeting inference within 200 milliseconds per frame. Machine-readable outputs will help robots and assistance systems recognize production situations, anticipate what happens next, and react more reliably to change. Our goals are to shorten development cycles, improve robustness across industrial environments, and build a scalable foundation for AI-enabled manufacturing.

Contact

Philipp Schubert, Ramblr GmbH, philipp@ramblr.ai

Training Europe's Construction Embedding AI with HammerHAI

User Goals

- **Scale our embedding model into a full cross-modal foundation model for construction data.**
- **Prove that domain-trained, EU-models can outperform large general-purpose ones.**
- **Give the European Architecture, Engineering, and Construction sector a sovereign semantic infrastructure layer.**

Current State and Key Challenges

flinq is a Stuttgart-based deep tech startup and is currently a technology spin-off (TGU) of TTI GmbH at the University of Stuttgart.

We are training a domain-specific, cross-modal embedding model for the European construction industry — an industry that generates enormous amounts of data every year, including tenders and an increasing number of Building Information Modeling (BIM) models and product catalogs.

As a cross-modal model, our embedding model processes many data types simultaneously, converting BIM models (IFC), tenders (GAEB), manufacturer data, construction site images, and other construction-specific formats into a common vector space. We make this available via an OpenAI-compatible API.

Our finely tuned 600-million-parameter Proof of Concept (PoC) model shows promising benchmark results compared to OpenAI's "text-embedding-3-large" and the "Qwen3" model, which is 13 times larger.

Initial tests also show that 3D geometry and text can be embedded in the same space. This approach could address the challenge of integrating diverse construction data.

To date, flinq has run all training on a single rented GPU. For scalability and robustness, we require more computing power than we currently have, highlighting the need for access to HammerHAI's computing infrastructure.

Support from HammerHAI

We are looking for continuous access to GPUs that can scale with us as we transition from today's fine-tuning experiments to the significantly larger training runs for new models that lie ahead. In addition to **raw computing power, HammerHAI's expertise in efficient large-scale training and its connections to the broader European AI ecosystem** are also of great value.

Intended Value and Impact

Using larger, fully cross-modal models, we aim to make flinq the semantic interoperability layer that the construction industry lacks, connecting BIM, tenders, products, and cost data. While the field moves toward horizontal, general-purpose models, we want to establish flinq as a dedicated embedding model for the construction industry and a trusted, vertical, EU-hosted solution. This will provide a common foundation that will reduce the need for time-consuming manual data reconciliation.

Contact

Matthias Hornung, flinq (TTI GmbH, Universität Stuttgart), info@flinq.ai

List of Abbreviations

AEC	Architecture, Engineering, and Construction
AI	Artificial Intelligence
AV	Autonomous Vehicle
BIM	Building Information Modeling
CAD	Computer-Aided Design
CGI	Computer-Generated Imagery
EU	European Union
GPU	Graphics Processing Unit
GWDG	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
HLRS	High-Performance Computing Center Stuttgart
HPC	High-Performance Computing
KISSKI	KI-Servicezentrum für sensible und kritische Infrastrukturen
KIT	Karlsruhe Institute of Technology
LLM	Large Language Model
LRZ	Leibniz Supercomputing Centre
PoC	Proof of Concept
RAG	Retrieval-Augmented Generation
RGB	Red, Green, Blue
SME	Small and Medium-sized Enterprises

Further Information

Learn more about HammerHAI, its activities, and latest updates:

Project website: www.hammerhai.eu
Contact: info@hammerhai.eu

Follow us:

Bluesky LinkedIn YouTube

Funded by:

The acquisition and operation of the EuroHPC AI-optimised supercomputer is funded jointly by the EuroHPC Joint Undertaking, through the European Union's Digital Europe Programme, as well as by the German Federal Ministry of Research, Technology and Space (BMFTR) and the Baden-Württemberg Ministry of Science, Research and the Arts.

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the EuroHPC JU. This project has received funding from the European High Performance Computing Joint Undertaking under grant agreement No. 101234027. It is co-funded by the European Commission, the German Federal Ministry of Research, Technology and Space (BMFTR), the Baden-Württemberg Ministry of Science, Research and the Arts, the Bavarian State Ministry of Science and the Arts and the Lower Saxony Ministry of Science and Culture.

