CAPABILITIES OF THE NATIONAL AI PLATFORM

Capacities of the existing supercomputer and plans for further expansion

eУПРАВА
КАНЦЕЛАРИЈА
ЗА ИТ И еУПРАВУ

Existing Supercomputer

VENDOR: **NVIDIA**
SUPPLIER: **ЕВИДЕН (ATOC)**
DECEMBER **2021.**
VALUE: **2M** EURO

- **4 x nVIDIA DGX A100 systems**
- **1024 CPU threads (AMD)**
- **32 GPU A100, 40GB RAM per GPU**
- **DDN Storage 150TB**
- **Data Science software**
- **HDR (InfiniBand), 100Gb/s**
- **5 PetaFLOPS AI performance**
- **Consumption 34,5kW**

Data Science Platform

- FMLe (Fast Machine Learning engine)

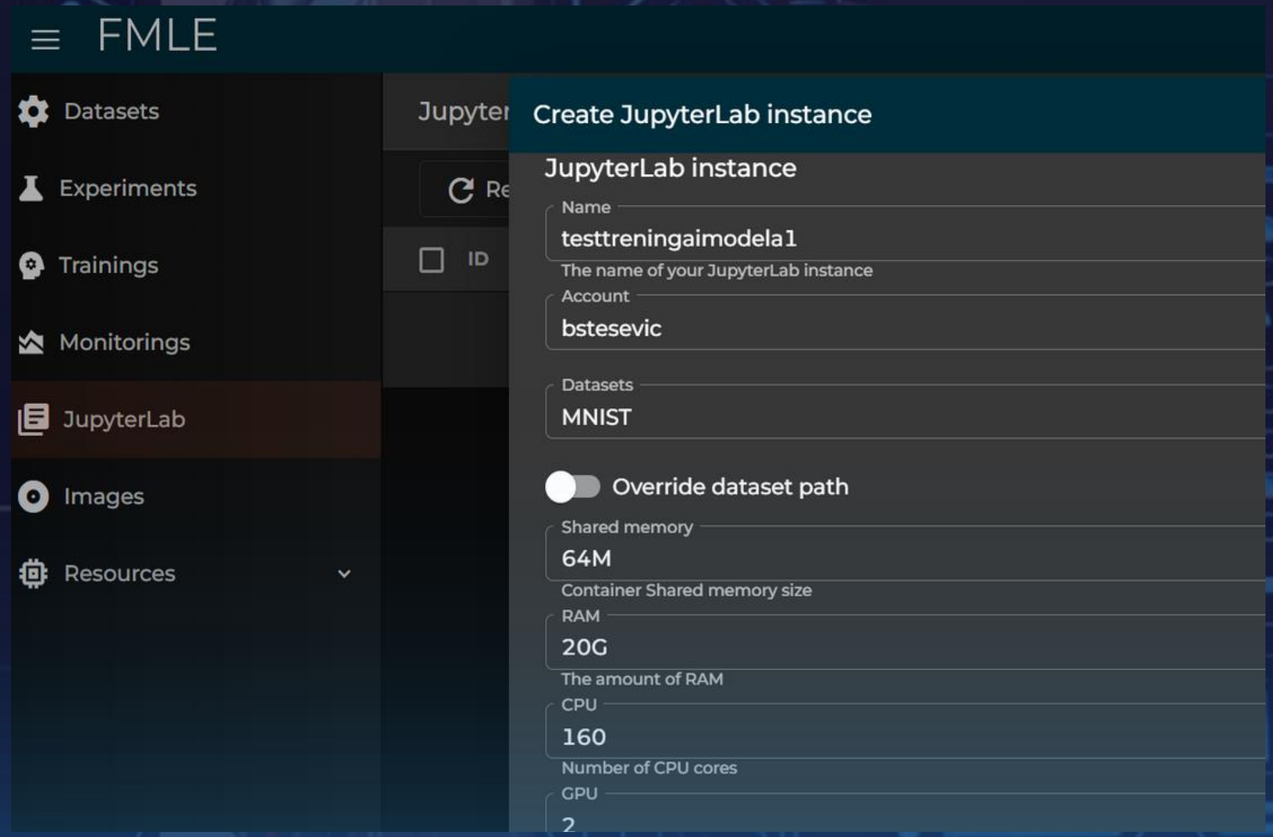
- User-friendly GUI

- Easily create AI training, experiments, and monitoring

- Easy manipulation of the data sets

- Monitoring resource usage, training duration, occupancy

- Jupyter Lab as a service

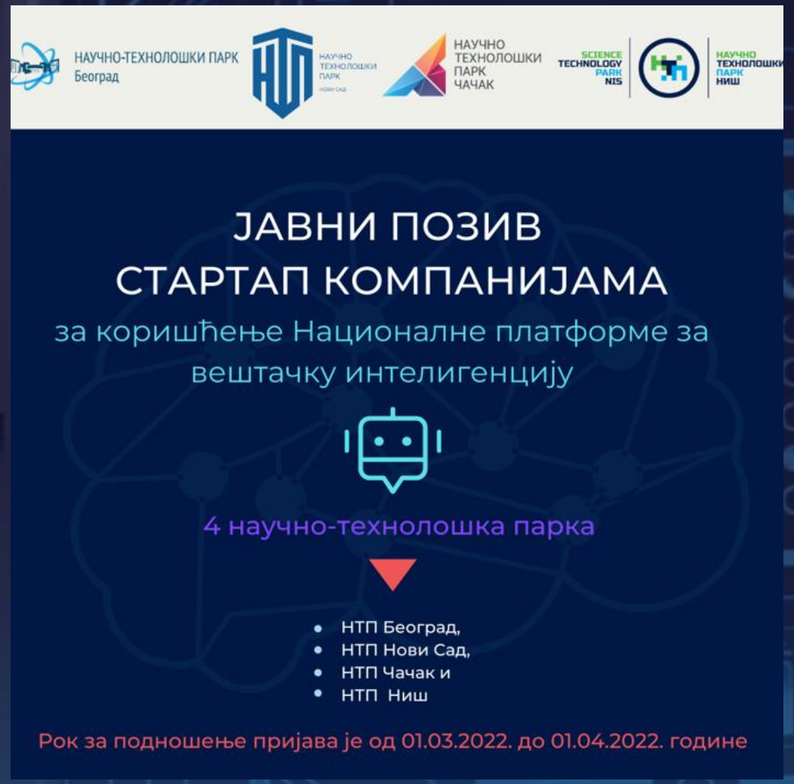


Users of National AI Platform

- **38 INSTITUTIONS:**
  - **UNIVERSITIES**
  - **FACULTIES**
  - **INSTITUTES**
  - **SCIENCE-TECHNOLOGY PARKS**
  - **INNOVATION FUND**

- **38 STARTUP COMPANIES**

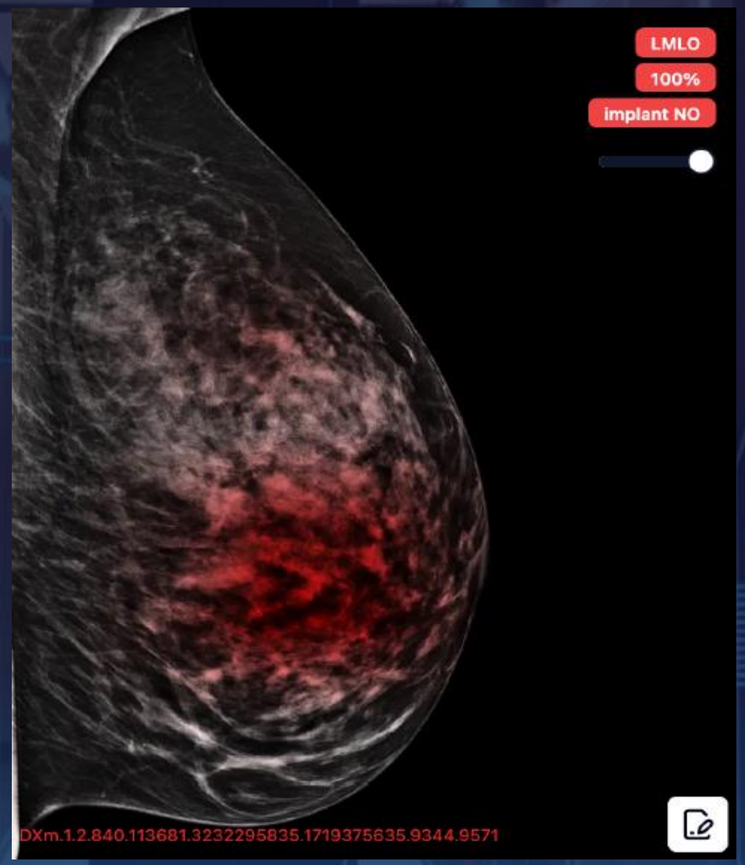
- **400 USER ACCOUNTS**

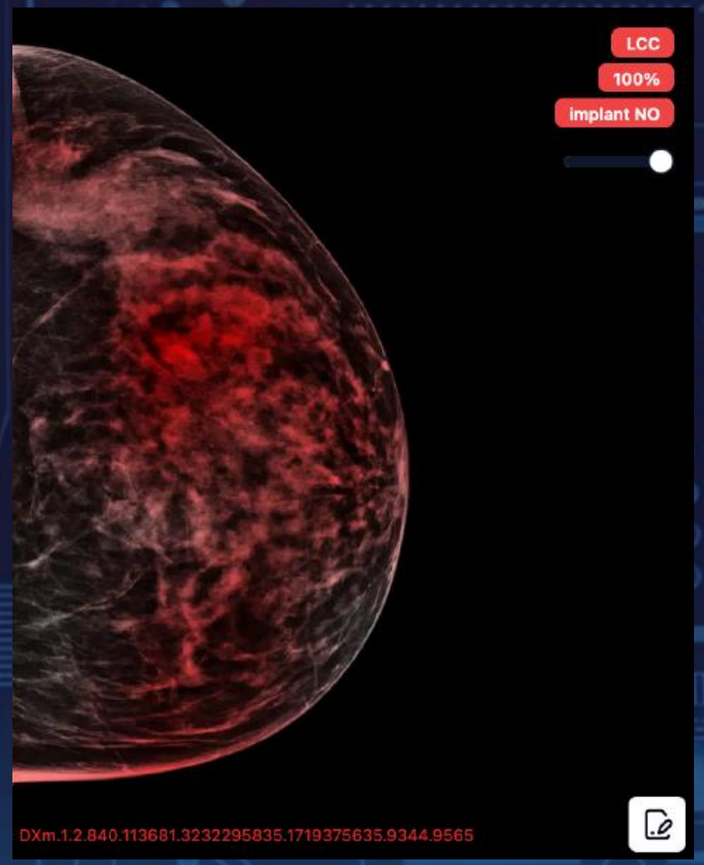


Automatic Ranking of Mammographic Images.

AI Mamography Project

- DEVELOPED BY AI INSTITUTE OF SERBIA
- DEVELOPED ON NATIONAL AI PLATFORM
- APPLICATION IS HOSTED IN THE STATE DATA CENTER IN KRAGUJEVEC
- AUTOMATIC RANKING OF MAMMOGRAPHIC IMAGES
- EXPLAINABLE AI – MARKED REGIONS WHERE THE SOFTWARE DETECTED A POTENTIAL ISSUE
- IORS FIRST USER





Expansion of the National AI Platform Phase 1

VENDOR: **NVIDIA**
SUPPLIER: **EVIDEN (ATOS)**
FIRST HALF OF **2024.**
VALUE: **5M** EURO

- **6 x nVIDIA DGX H200 systems**
- **1344 CPU threads (Intel)**
- **48 GPU H200, 141GB RAM per GPU**
- **DDN Storage 120TB**
- **NDR (InfiniBand) network technology, 200Gb/s**
- **32 PetaFLOPS AI performance**
- **Consumption 72kW**
- **Planned Launch - Second Half of 2025.**

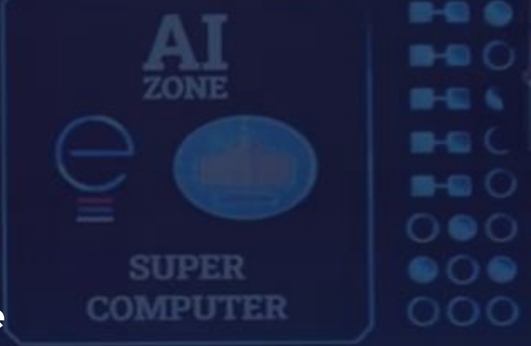
*** 7 X faster performance than the existing GPUs**

Expansion of the National AI Platform Phase 2

VENDOR: **NVIDIA, EVIDEN**
SUPPLIER: **EVIDEN (ATOS)**
FIRST HALF OF **2025.**
VALUE: **36M** EURO

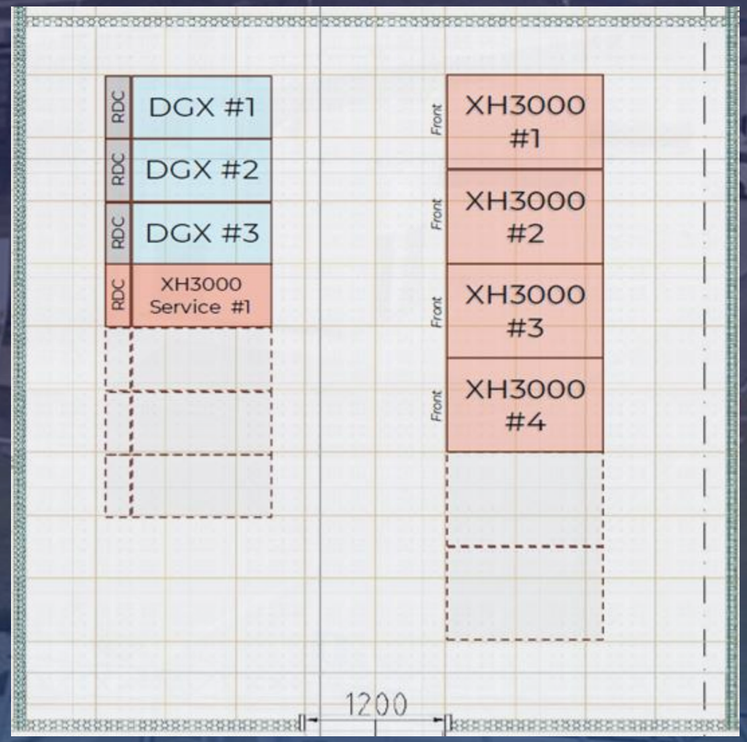
- **Loan from the French government**
- **HPC + Software stack + Mistral AI**
- **4 x BullSequana XH3000 racks**
- **640 x nVIDIA GraceHopper Superchip**
- **96GB RAM per GPU**
- **Jarvice AI**

- **Inference**
- **DDN Storage 2.5PB**
- **NDR (InfiniBand), 200Gb/s**
- **22.5 PetaFLOPS AI performance**
- **Consumption 550kW**
- **Planned Launch - 2026.**



Integration of 3 Supercomputers into one cluster

- **Planed integration of three supercomputers**
- **Designed mini module with a water-cooling system (two technologies)**
- **Double floor designed to support a weight of 2.5 tons per reck**
- **Integration of all storages so that users perceive it as a one name space**
- **Total consumption 656,5kW**



Thank you for your attention!

In 2023, the OECD ranked the National Artificial Intelligence Platform as one of the 9 best innovative projects in the public sector, from a competition of 1,048 projects across 94 countries worldwide.

eУПРАВА | КАНЦЕЛАРИЈА ЗА ИТ И еУПРАВУ

Contact: Богдан Стешевић
bogdan.stesevic@ite.gov.rs