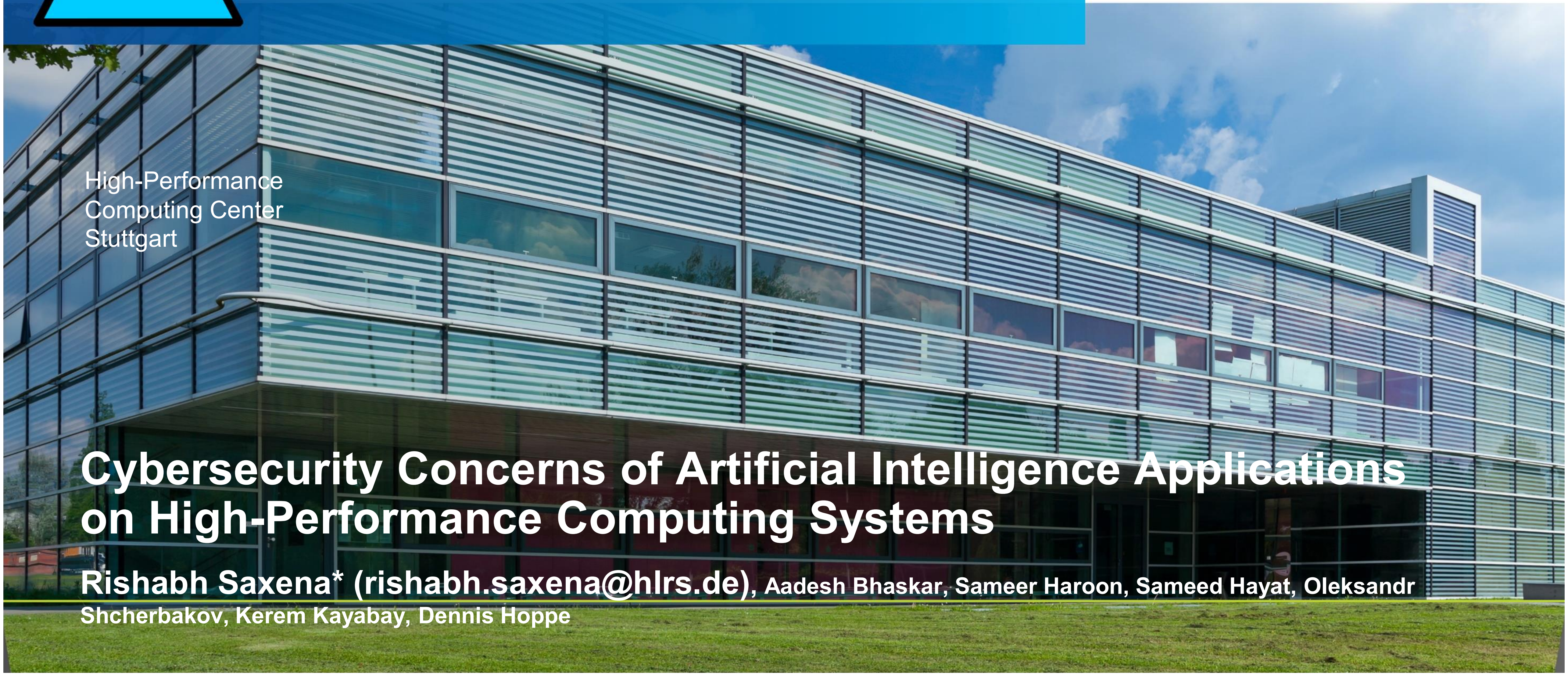


High-Performance
Computing Center
Stuttgart

Cybersecurity Concerns of Artificial Intelligence Applications on High-Performance Computing Systems

Rishabh Saxena* (rishabh.saxena@hlrs.de), **Aadesh Bhaskar, Sameer Haroon, Sameed Hayat, Oleksandr Shcherbakov, Kerem Kayabay, Dennis Hoppe**

About the Presenter

- **Current:** Working as an ML Engineer/Research Scientist with HLRS under Converged Computing department.

- **Education:** Master of Science, Data Science from RWTH Aachen, Germany; Bachelor of Engineering in Computer Science from VIT University, India.

- **Research Interests:** HPC AI Convergence, with specific focus on MLOps and ML Engineering pipelines.

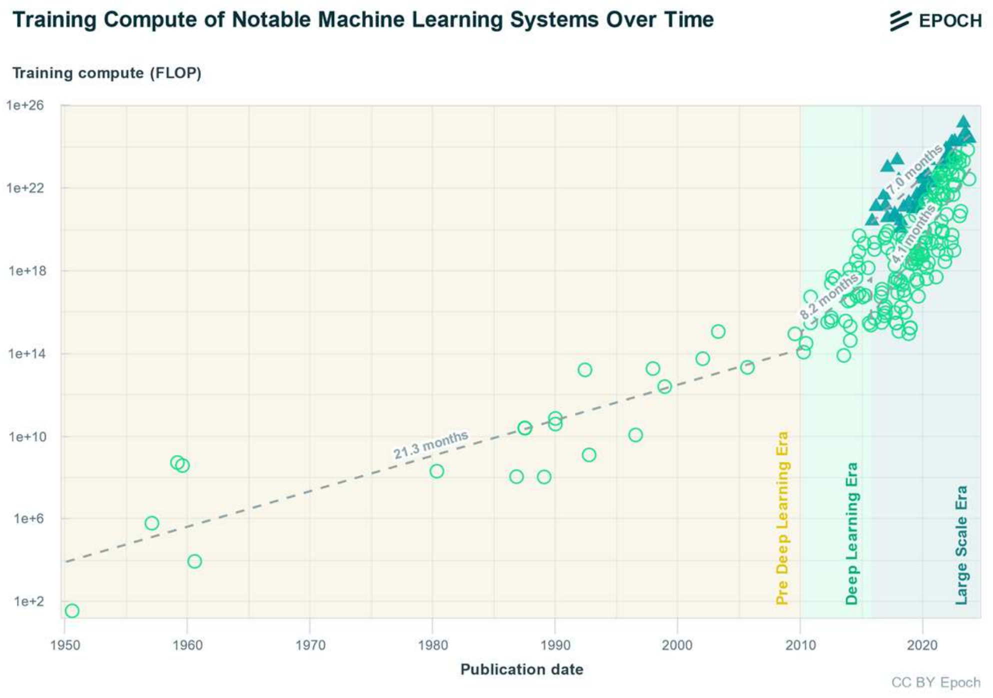
Introduction

Introduction

AI and HPC Intersection: AI demands high processing power, which HPC can provide. Likewise, simulation workflows on HPC can be enhanced using AI.

Emerging Security Challenges:

- AI frameworks (e.g., TensorFlow, PyTorch) introduce instability, requiring HPC experts to adapt to evolving AI needs.
- Resistance from traditional HPC administrators due to security concerns and unfamiliarity with AI software.

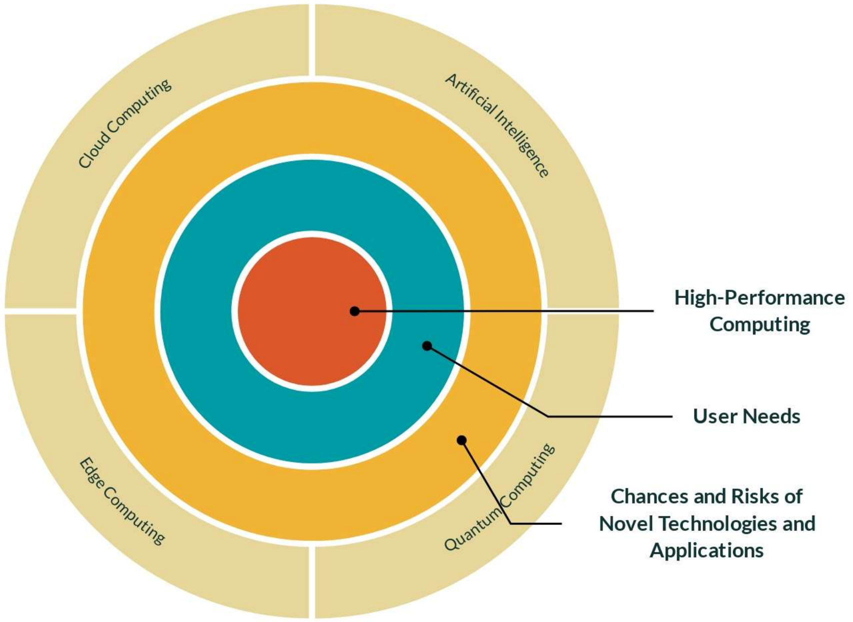


CC BY Epoch

Motivation

- **Why AI on HPC Matters:** SMEs and start-ups are increasingly reliant on foundational AI models. With access to HPC resources, smaller enterprises can compete better.

- **Technical Challenges:** Seamless hybrid HPC/AI workflows can be achieved by tackling technical obstacles like containerization, usage patterns, and cybersecurity aspects.

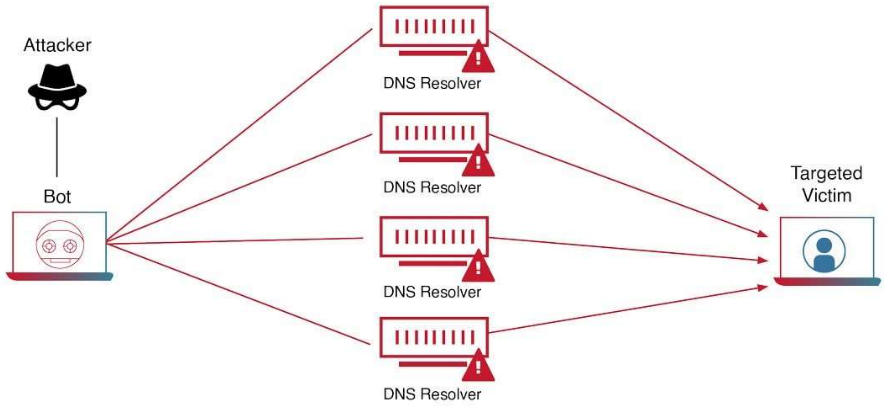

Cloud Computing
Artificial Intelligence
Edge Computing
Quantum Computing
High-Performance Computing
User Needs
Chances and Risks of Novel Technologies and Applications

Cybersecurity on HPC

Cybersecurity on HPC (1)

- **HPC Systems' Vulnerabilities:** Similar to general IT systems, HPC systems face threats like:
    - Unauthorized access
    - Data breaches
    - Denial of Service (DoS) attacks
    - Compute cycle theft and misuse

- **Key Differences from General IT Systems:**
    - Batch Scheduled Jobs instead of virtualized occupancy.
    - Ingress/egress traffic points and user access are tightly controlled (e.g., via login and data transfer nodes).



Cybersecurity on HPC (2)

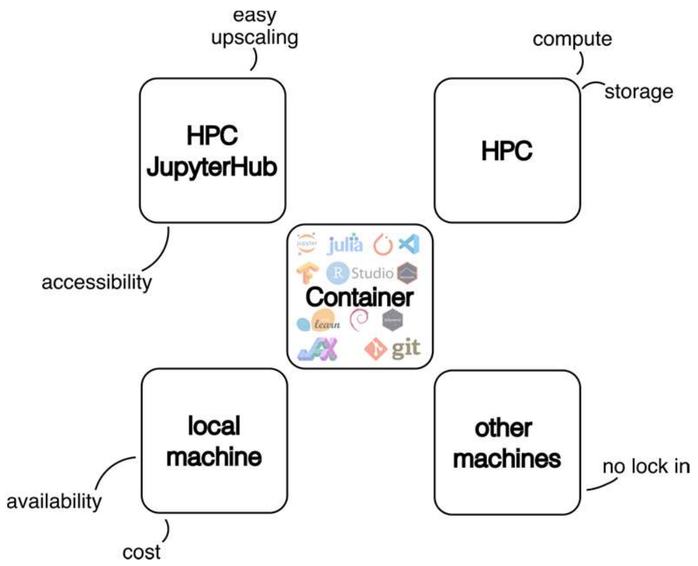
- **Security Challenges:**
  - Security tools introduce performance overheads.
  - Multi-tenancy increases risks of cross-job attacks.
  - New technologies like containers improve portability but add extra attack surfaces.

- **Emerging Solutions:**
  - Secure tools and frameworks with restricted privileges can mitigate multi-tenant risks.
  - Enhanced file-system access control mechanisms (e.g., Discretionary and Mandatory Access Control) are vital.

Cybersecurity for AI/ML on HPC (1)

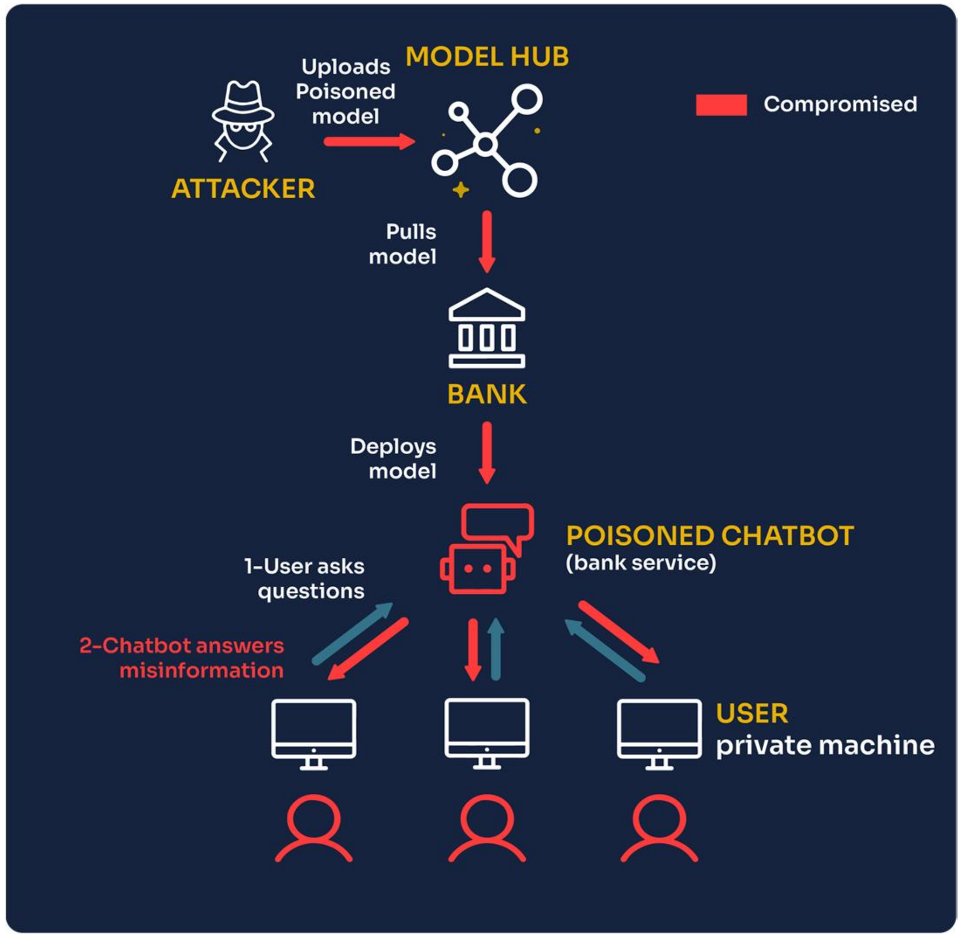
- **Problem Definition**:
  - Malicious AI models (e.g., deepfakes) can be developed on HPC.
  - Requires stringent oversight of project purposes and periodic reviews to prevent misuse.

- **Data Exploration**:
  - Vulnerabilities occur when opening ports for Exploratory Data Analysis (EDA).
  - HPC, when connected to external cloud/edge systems, has increased exposure to threats.
  - Containerization for EDA, like Docker, presents security risks in shared HPC environments.



Cybersecurity for AI/ML on HPC (2)

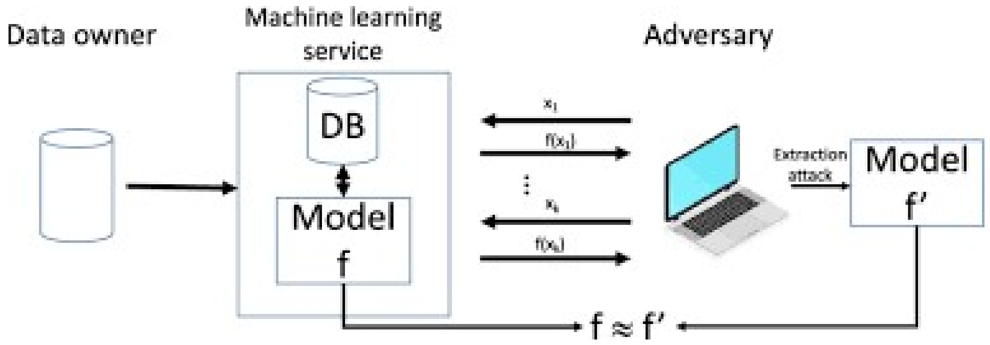
- **Data Ingestion**:
  - Challenges in securely transporting large datasets on HPC; risk of man-in-the-middle attacks.
  - Encryption and data transmission steps differ between cloud and traditional HPC.
- **Data Engineering**:
  - Risks like data poisoning, unauthorized access, and tampering of data pipelines extend to HPC environments.
- **Model Training**:
  - Security risks include open backdoors, malicious code execution, and compromised pre-trained models.
  - Cloud-HPC pipelines can expose privileged steps to attackers.



Cybersecurity for AI/ML on HPC (3)

- **Model Deployment**:
  - Risks like Distributed Denial of Service (DoS), model extraction, and membership inversion attacks when deploying models in production.
  - Attackers may exploit GPU sessions or inject malicious code during inferencing.

- **Monitoring & Maintenance**:
  - CI/CD pipelines introduce vulnerabilities such as privilege escalation in containers.
  - Malicious code in CI/CD can infect the HPC cluster, creating large-scale security issues.

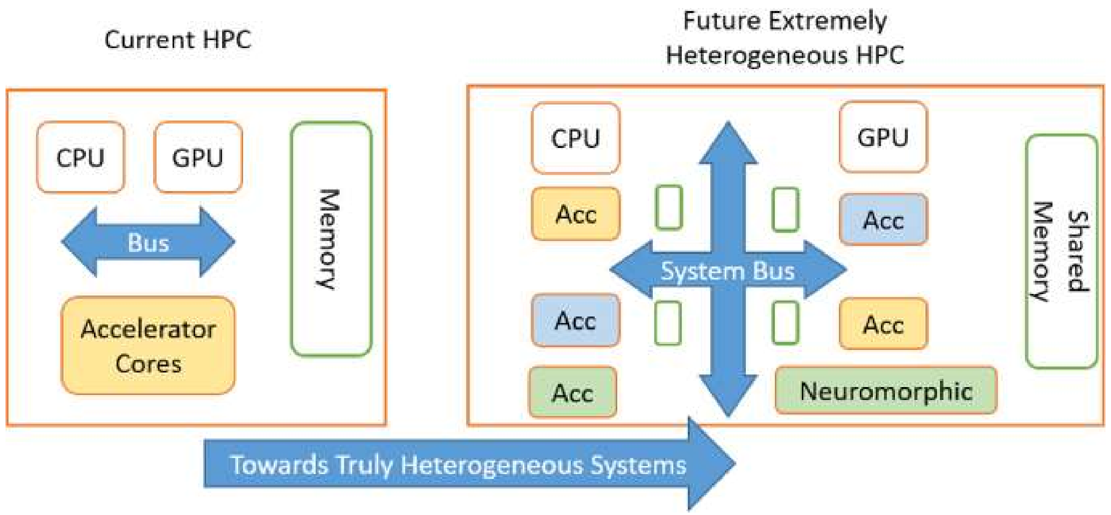


Challenges and Mitigation Strategies

Challenges (1)

- **Diverse Hardware Components**
  - Integration of CPUs, GPUs, TPUs, quantum processors
  - Unique vulnerabilities (e.g., side-channel attacks, memory exploits)
  - Requires specialized knowledge for secure configurations

- **Performance vs Security**
  - HPC prioritizes performance with minimal security overhead
  - Shared environments increase risk compared to virtualized cloud setups



Challenges (2)

- **Diverse Software Ecosystems**
  - AI, Big Data, and HPC software have different ecosystems
  - Managing dependencies and compatibility is challenging without introducing security risks (Supply Chain Attacks)

- **ML Frameworks & HPC Security**
  - ML frameworks (e.g., TensorFlow, PyTorch) are constantly changing, bringing new vulnerabilities
  - More resources are dedicated to cloud security, while HPC security is usually an afterthought

Challenges (3)

- **Managing Diverse Systems**
    - Coordinating security across different platforms (HPC, AI accelerators, Big Data systems)
    - Different access control and update needs for each system type

- **Delegated Security Risks**
    - Users expected to manage their own network security
    - Varying expertise leads to weak security practices during data transfer and software use

Mitigation Strategies

Mitigation Strategy	Cloud-Based GPU Systems based on [26]	HPC Systems with GPU Acceleration

Conclusion

Conclusion

- **AI Security in HPC**:
  - HPC systems present unique AI security challenges due to their focus on performance and scale, lacking the virtualization benefits seen in cloud environments. Distributed AI workloads increase risks such as data poisoning, adversarial attacks, and model inversion.

- **Complex Hardware Integration**:
  - AI security in HPC is complicated by the diverse hardware, including CPUs, GPUs, and TPUs, which require tailored security strategies across the system.

- **Gaps in Literature**:
  - Current research does not adequately address AI-specific security vulnerabilities in HPC environments, particularly within the lifecycle of machine learning applications.

Outlook

- **Future Threat Mitigation**:
  - Incorporating tools like NVIDIA Morpheus and DPUs can enhance AI security in HPC by isolating cybersecurity processes and improving traffic analysis across clusters.
- **Research Directions**:
  - Future studies should focus on systematically ranking vulnerabilities, validating threat mitigation strategies, and developing secure AI adoption roadmaps for HPC systems.
- **HPC Accessibility**:
  - To lower barriers for SMEs, start-ups, and researchers, future work should focus on integrating secure AI processes into HPC environments, promoting innovation with robust security practices.

IARIA

Thank you!

E-Mail: rishabh.saxena@hlrs.de