**Alexey Shvechkov**  
Greater Boston, MA | [alexey@shvechkov.com](mailto:alexey@shvechkov.com) | +1.774.278.1743 | [[LinkedIn profile]](https://www.linkedin.com/in/alexey-shvechkov/) | [[github.com](https://github.com/shvechkov)]  
  
Summary

## **Technical Director | Hands-On Systems, Cloud & AI Architect | From Kernel Drivers to Agentic AI Platforms**

With over two decades of expertise, I architect, prototype, and deliver scalable enterprise solutions, blending hands-on systems programming, cloud infrastructure, and AI-driven platforms. As a Technical Director, I lead global teams of up to 10 engineers while actively coding and prototyping mission-critical systems, from low-level kernel drivers (Linux/Windows) to advanced AI platforms leveraging LLMs, agentic AI, and cloud-native technologies (AWS, Azure, Kubernetes). My hands-on contributions include designing high-throughput systems, fine-tuning transformer-based models, and rapidly prototyping solutions like cyber-resilient appliances, reducing delivery cycles by 40%. I drive innovation through agile leadership, R&D excellence, and direct coding in C++, Golang, Rust, Python, and modern AI frameworks.

## Education

* **M.Sc. in Applied Mathematics** – Institute of Telecommunications and Computer Science, Russia (1996)
* **Certified in Machine Learning (**[**ML**](https://github.com/shvechkov/coursera_ml_ex/blob/main/completion_cert.png)**)**
* **Holder of multiple** [**U.S. patents**](https://patents.google.com/?q=(Alexey+Shvechkov)) in storage and infrastructure technologies

## Skills

**Systems Programming**

* Languages: C++, Golang, Rust, Python
* Kernel-Mode Development: Linux/Windows (eBPF, WDM, KMDF, minifilters, file systems), FUSE
* Hands-On Contributions: Architect and code kernel-level components, file systems, and high-throughput backup/replication systems

**AI/ML & LLMs**

* Frameworks: TensorFlow, PyTorch, scikit-learn, LangChain, ONNX Runtime
* Expertise: BERT, agentic RAG, supervised/unsupervised learning, transformer fine-tuning
* Hands-On Contributions: Prototype AI-driven malware detection, build agentic AI assistants, and optimize LLMs for edge inference

**Cloud & DevOps**

* Technologies: AWS SDK (Python, Go), Azure SDK (Python), Kubernetes, Docker, Terraform, REST/SOAP APIs, CI/CD
* Expertise: Scalable internet-based applications, resilient architectures
* Hands-On Contributions: Architect and prototype cloud-native solutions, including agentless protection for AWS/Azure workloads and immutable object-store servers

**Databases**

* RDBMS: Oracle, PostgreSQL, MSSQL, MySQL
* NoSQL: Redis, Cassandra

**Leadership**

* Technical Direction, Agile Delivery, R&D Management, M&A Due Diligence
* Hands-On Leadership: Mentor teams while coding and prototyping solutions

**Additional Skills**

* React, Next.js, Bash scripting

## Work Experience

**Technical Director, Software Engineering**

*Arcserve – Greater Boston, MA | Aug 2014 – Present*

*Progressed from Staff SWE → Sr Architect → Product Owner → Technical Director*

**Leadership & Strategy**

* Provide hands-on technical leadership for AI/ML initiatives and platform modernization, reporting to CTO.
* Lead multiple high-impact teams of up to 10 engineers across storage, cloud protection, and AI-driven features, defining roadmaps and delivering on schedule.
* Mentor engineers while actively coding and prototyping, ensuring technical excellence and measurable customer impact.

**AI/ML Integration & Agentic Systems**

* Architected and prototyped malware/anomaly detection features: coded EDA pipelines, performed feature engineering, trained ML models, and fine-tuned LLMs using Python and C++.
* Developed and optimized transformer-based masked LLMs via ONNX Runtime for edge inference, writing performance-critical C++ code.
* Prototyped and coded agentic AI assistants using LangChain and OpenAI, integrating with legacy systems via REST, SOAP, and MCP protocols.
* Designed and implemented an interactive RAG-based AI assistant, writing core logic and optimizing retrieval pipelines.
* Architected and coded MCP servers to bridge agentic interfaces with legacy Arcserve UDP APIs.

**Systems & Product Engineering**

* Architected and coded core Arcserve features, including data deduplication, file system/server replication, high availability (HA), and agentless protection for AWS/Azure workloads (VMs, containers, storage), with hands-on contributions in C++ and Go.
* Prototyped and developed a Linux-based immutable object-store server (cyber-resilient appliance), writing core components and reducing delivery cycle by 40%.
* Coded system and kernel-level components (file systems, filter drivers) on Windows and Linux for high-throughput backup and replication products.

**R&D & Innovation**

* Won hackathons for AI and infrastructure prototypes, rapidly coding proof-of-concepts in Rust and Python.
* Led M&A-influencing projects, performing due diligence and coding prototypes to validate technical feasibility.
* Developed kernel-mode anti-malware solutions and email archiving engines, contributing both architecture and implementation.

**Early Career**

**Software Engineer / Principal Engineer**

*Yandex, XOsoft, Computer Associates | 1998–2014*

* Designed and coded high-traffic web applications at Yandex, including search engine, advertising/banner networks, polling systems, and discussion forums in C++ and Perl, driving 50%+ traffic growth and 20% revenue increase for ad platforms.
* Developed key modules for a content delivery network (CDN) at XOsoft, comparable to Akamai, coding network and system software in C++ to support low-latency content delivery for 10M+ daily users with 99.9% uptime.
* Implemented system and networking modules for high-availability data replication at XOsoft, writing C++ code to achieve 25% improved throughput for enterprise backup systems.
* Led a kernel-mode development team of 5 engineers at XOsoft and CA, mentoring on driver development and delivering scalable storage solutions.
* Built and automated CI/CD pipelines using PHP and Bash, reducing build times by 30% and enabling rapid iteration for development teams.