As an NVIDIA Elite Partner, Quantiphi enables customers with end-to-end AI workflows using the NVIDIA AI Enterprise software stack and NVIDIA DGX platform. Our goal is to deliver customizable solutions that optimize efficiency, reduce costs, and elevate overall customer satisfaction.

- **Enterprise Tailored RAGs**
- **Digital Avatar**
- **Intelligent Video Analytics**
- **Medical Imaging**
- **Drug Discovery**

### Industry Specific Solutions

#### Healthcare & Life Sciences
- Clinical Trial Matching
- Adverse Event Detection

#### Banking & Financial Services
- Equity Research Analysis
- Banking Relationship Management

#### Telecom
- LLM Powered Chatbot To Assist Field Technicians
- RAG Powered Network Management

#### Retail
- Interactive Recommendation Assistant
- Customer Traffic Analytics

#### Media & Entertainment
- Real Time Closed Captioning
- IVA for Safety & Surveillance

#### Manufacturing
- Generative Search for Technician Support
- Visual Quality Inspection

### Deploy & Scale

- **Platform Modernization**
- **MLOps / LLMOps**
- **Training & Inference at Scale**
- **Data Management**

Unlock the Potential of AI for your Enterprise
### Customer Success Stories

#### Field Assistant chatbot for Technicians

Developed an LLM powered Q&A chatbot application to train client’s Fiber installation technicians and minimize the frequency of supervisor calls.

**Business Impact**
- Improved Technician Efficiency
- 90.5% Q&A pipeline accuracy

#### LLM-Powered Digital Avatar for Troubleshooting Assistance

Created an LLM-powered Avatar for user product queries, FAQs, and network device troubleshooting.

**Business Impact**
- Reduced Agent Workload
- Faster Query Resolution
- Accurate Responses & Troubleshooting

#### Automated Clinical Trial Patient Matching

Developed a RAG pipeline with domain specific fine tuning for Clinical Trial Patient Matching to automate the initial screening of patients for clinical trials.

**Business Impact**
- Accelerated clinical trial timelines with optimized patient recruitment process
- Reduced selection bias with greater adherence to study design

#### Computer Vision Based Visual Inspection

Developed an end-to-end IVA solution hosted on NVIDIA Triton Inference Server that can detect missing retainer valves in locomotives & generate alerts.

**Business Impact**
- 62% Reduction in Latency using TensorRT
- >97% Precision and Recall

---

**baioniq is a powerful enterprise-ready generative AI platform that enables enterprises to solve real world business problems and boost knowledge-worker productivity. baioniq, powered by NVIDIA AI Enterprise - NeMo, TensorRT LLM & Triton** is making generative AI adoption easier and accelerating solution development while adhering to responsible AI principles.

---

**Get in touch with our experts to learn more.**

Email us at: info@quantiphi.com

---

Quantiphi is an award-winning AI-first digital engineering company driven by the desire to reimagine and realize transformational opportunities at the heart of business. We are passionate about our customers and obsessed with problem-solving to make products smarter, customer experiences frictionless, processes autonomous and businesses safer.