VISHAL SOOD

(+41) 79 942 0489 · vishal.chandra.sood@protonmail.com · Bussigny, CH-1030
Swiss Permit C · github.com/visood · linkedin.com/in/physicistvisha

SENIOR RESEARCH ENGINEER

Designing platforms that abstract computational complexity while preserving analytical power

A **Senior Research Engineer / Computational Biology Specialist** with extensive experience developing high-performance bioinformatics pipelines and clinical-grade software. Specialized in architecting scalable C++ / Python solutions for processing complex biological data, from large-scale genomics to multi-terabyte scientific simulations. Proven ability to translate complex scientific and clinical requirements into robust, maintainable, and user-friendly software for precision medicine and advanced research.

CORE EXPERTISE

Analytical Thinking | Research Design | Scientific Writing | Problem Solving | Project Management | Data Interpretation | Process Optimization | Critical Reasoning | Cross-Disciplinary Collaboration | Experimental Design | Knowledge Management | Technical Documentation | Workflow Development | Strategic Planning | Quality Assurance

TECHNICAL SKILLS

Programming `C++, Python, R, JavaScript, Scala, Haskell`
HPC `SLURM, Spark, MPI, Dask, Joblib`
Scientific `NumPy, SciPy, Pandas, HDF5, Matplotlib, Seaborn, Jupyter`
Databases `PostgreSQL/PostGIS, MySQL, SQLAlchemy, NoSQL, spatial-indexing, optimization`
Development `Git, CMake, Jenkins, Docker, CI/CD, Test-driven development, cross-platform, Emacs`
Algorithms Data structures, pattern-detection, geometry, computational parsers, optimization
Statistics Monte Carlo, Bayesian, hypothesis testing, time-series analysis, signal processing
Modeling Stochastic processes, complex-networks, multi-scale, volumetric modeling
ML / AI Classification, ensemble methods, validation, feature engineering
Biocomputing Neuroscience, genomics, bioinformatics, brain circuits, variant annotation
Physics Statistical mechanics, reaction-diffusion & kinetics, complex systems, molecular dynamics
Data systems FAIR principles, metadata, knowledge graphs, provenance tracking, controlled vocabularies
Geospatial GIS, cartographic modeling, spatial statistics, urban analytics
Languages English (native), French (intermediary), Hindi (native), Punjabi (mother-tongue)

SPECIALIZED EXPERTISE

High-Throughput Biological Data Systems

Architected and built high-performance C++ backends for processing complex genomic data. Designed and deployed REST APIs for clinical-grade, distributed data services with a focus on low-latency and high reliability.

Clinical Genomics & Variant Annotation

Developed end-to-end variant annotation and interpretation pipelines. Implemented ACMG guideline-based classification systems and integrated multiple large-scale genomic databases (e.g., ClinVar, gnomAD) for comprehensive analysis.

Scientific Software Architecture & API Design

Engineered robust, maintainable software using modern C++ and Python. Designed intuitive APIs, configuration-driven pipeline architectures, and extensible plugin-based frameworks to bridge complex backend capabilities with scientific user needs.

Mathematical Modeling & Advanced Statistical Analysis

Developed and applied sophisticated statistical and stochastic models for pattern detection in high-dimensional, noisy datasets. Expertise in Monte Carlo simulation, Bayesian methods, and robust hypothesis testing to validate algorithms and discover novel biomarkers.

High-Performance Computing & Algorithm Development

Developed and optimized scalable algorithms using C++ and Python (Joblib, Dask, Multiprocessing) for high-throughput data streams on large-scale Linux systems and HPC clusters (SLURM).

Scientific Workflow Development & Automation

Designed and automated end-to-end, configuration-driven reproducible computational pipelines for scientific and clinical applications, incorporating robust error handling, dependency management, and automated quality control.

Data Engineering & Computational Pipelines

Engineered computational pipelines managing the full data lifecycle. Mastered large-scale data storage and access using HDF5 and relational databases. Designed hierarchical data structures and metadata schemas to ensure data integrity (FAIR principles).

Collaboration & User-Centric Development

Excelled in multidisciplinary agile teams, translating scientific and clinical needs into robust technical specifications and intuitive software solutions. Authored technical documentation, co-authored scientific publications, and developed scientific web-portals.

`LaTeX`

SPECIALIZED EXPERTISE

- **High-Throughput Biological Data Systems:** Architected and built high-performance C++ backends for processing complex genomic data. Designed and deployed REST APIs for clinical-grade, distributed data services with a focus on low-latency and high reliability.
- **Clinical Genomics & Variant Annotation:** Developed end-to-end variant annotation and interpretation pipelines. Implemented ACMG guideline-based classification systems and integrated multiple large-scale genomic databases (e.g., ClinVar, gnomAD) for comprehensive analysis.
- **Mathematical Modeling & Advanced Statistical Analysis:** Developed and applied sophisticated statistical and stochastic models for pattern detection in high-dimensional, noisy datasets. Expertise in Monte Carlo simulation, Bayesian methods, and robust hypothesis testing to validate algorithms and discover novel biomarkers.
- **High-Performance Computing & Algorithm Development:** Developed and optimized scalable algorithms using C++ and Python (Joblib, Dask, Multiprocessing) for high-throughput data streams on large-scale Linux systems and HPC clusters (SLURM).
- **Advanced Software Architecture & API Design:** Designed extensible, plugin-based frameworks (Python ABCs) and configuration-driven pipelines, creating maintainable and reproducible systems for collaborative scientific development.
- **Scientific Workflow Development & Automation:** Designed and automated end-to-end, configuration-driven reproducible computational pipelines for scientific and clinical applications, incorporating robust error handling, dependency management, and automated quality control.
- **Data Engineering & Computational Pipelines:** Engineered computational pipelines managing the full data lifecycle. Mastered large-scale data storage and access using HDF5 and relational databases. Designed hierarchical data structures and metadata schemas to ensure data integrity (FAIR principles).
- **Collaboration & User-Centric Development:** Excelled in multidisciplinary agile teams, translating scientific and clinical needs into robust technical specifications and intuitive software solutions. Authored technical documentation, co-authored scientific publications, and developed scientific web-portals.

PROFESSIONAL EXPERIENCE

SENIOR SCIENTIFIC SOFTWARE DEVELOPER 2017 – 2024

Blue Brain Project, EPFL

Architected and built the core computational frameworks that enabled large-scale analysis and validation of the world's most detailed brain circuit models, directly empowering a large team of neuroscientists.

- Translated complex neuroscientific inquiry into robust, automated computational workflows; architected configuration-driven systems (Python, HPC/Slurm) that allowed scientists to execute and manage petabyte-scale analyses reproducibly.
- Led the design and development of a novel, plugin-based scientific validation framework (Python), creating an extensible ecosystem for rigorously comparing computational models against diverse experimental data.
- Engineered user-centric Python APIs that provided an intuitive, high-level interface to complex, terabyte-scale hierarchical data stores (HDF5), significantly lowering the barrier for exploratory data analysis.

SENIOR PRODUCT DEVELOPER 2015 – 2017

Saphetor SA

Applied software engineering rigour to build a high-performance clinical genomics platform from the ground up, rapidly mastering a new scientific domain to deliver production-critical systems.

- Engineered the high-performance C++ backend for a real-time genomic variant annotation service, translating complex clinical and bioinformatics requirements into a scalable, low-latency system.
- Developed the end-to-end data processing pipeline for integrating and querying multiple large-scale genomic databases, creating the core knowledge base for variant interpretation.
- Designed and implemented a robust Python REST API to serve complex clinical queries, bridging the gap between the high-performance backend and user-facing applications.

DATA SCIENTIST 2015- 2015

Citiviz Sárl, EPFL Innovation Park, Switzerland

MAÎTRE ASSISTANT 2011- 2014

Lab of Statistical Biophysics, EPFL, Switzerland

POST-DOCTORAL FELLOW 2008- 2010

Niels Bohr Institute, Copenhagen, Denmark

RESEARCH FELLOW 2006- 2008

Physics Dept., University of Calgary, AB, Canada

ACADEMIC FOUNDATION

Doctor of Philosophy in Physics, Boston University, Boston, MA, USA 2006

Thesis: Interacting Particle Systems on Graphs. Developed theoretical and computational models for stochastic processes, providing a deep, first-principles foundation for modelling complex systems and developing data-driven algorithms.

Bachelor of Technology in Engr. Physics, Indian Institute of Technology Bombay, India 2000

KEY RESEARCH ENGINEERING PROJECTS

My work has focused on building foundational platforms that operationalize research, transforming complex computational tasks into reproducible, scalable, and user centric workflows.

Automating Large-Scale Neuroscience Analysis on HPC

Architected a workflow engine (connsense-TAP) to manage petabyte-scale analyses of brain circuit models on distributed HPC systems.

- Scientific Enablement: Abstracted away the complexity of parallel execution (Slurm), allowing scientists to launch massive computational experiments via simple, declarative configuration files.
- Integrated Data System: Engineered a unified HDF5 data store with a lazy-loading Python API, transforming terabytes of results into a single, queryable scientific asset.

Architecting a Framework for Collaborative Scientific Validation

Conceived and built a novel software ecosystem (DMT) for rigorous, automated, and collaborative model-data comparison.

- Interoperable by Design: Used the Adapter Pattern and formal API contracts to create a plugin-based system where any community model could be tested against any experimental dataset.
- Automated Scientific Narrative: Engineered a system to programmatically generate rich validation reports, integrating text, data, and figures into a complete, reproducible scientific argument.

Architecting a Knowledge Portal for a Digital Brain Reconstruction

Architected the end-to-end data and analysis pipelines that power the SSCx Portal, a platform making one of the world's most detailed brain reconstructions accessible to the global neuroscience community.

- Data Processing Engine: Designed the core engine (connsense-TAP) to transform raw, petabyte-scale simulation outputs from HPC into a structured, queryable HDF5 knowledge base.
- Scientific Content Generation: Engineered the analysis framework (factology) to programmatically generate the portal's content hierarchical "factsheets" with integrated data and figures.

Building a Production-Grade Platform for Clinical Genomics

Engineered core components for a real-time genomic variant annotation and interpretation service (Saphetor).

- Research to Production: Translated complex bioinformatics requirements and clinical guidelines (ACMG) into a robust, high-performance C++ backend and Python REST API.
- Data Integration: Designed pipelines to ingest and integrate multiple large-scale genomic databases, creating a comprehensive knowledge base for clinical decision support.