

Arno Candel

Building at the Frontier of AI and Compute — research to production

- Location: Los Gatos, CA
- Site: <https://candel.org>
- Demos: <https://candel.org/zeta>
- GitHub: <https://github.com/arnocandel>

Profile

Shipping CTO / hands-on systems engineer for agents, RAG, AutoML, and large-scale ML training data. Former H2O.ai CTO (#1 on three revenue platforms, ~50k commits); Member of Technical Staff at xAI (computer-using agents); prior SLAC/ACE3P petascale simulation and Skytree (first technical hire).

Social

- LinkedIn: <https://linkedin.com/in/candel>
- X: <https://x.com/arnocandel>
- GitHub: <https://github.com/arnocandel>
- YouTube: https://www.youtube.com/results?search_query=%22Arno+Candel%22
- ORCID: <https://orcid.org/0009-0001-0262-3354>
- Kaggle: <https://www.kaggle.com/arnocandel>

Roles

Organization	Role	Dates
xAI	Member of Technical Staff	Oct 2025-Dec 2025
H2O.ai	CTO · technical leader	Dec 2013-Oct 2025
Skytree	Senior Member of Technical Staff · First technical hire	Jan 2012-Dec 2013
SLAC	Stanford Linear Accelerator Center · Staff Scientist	Sep 2005-Jan 2012
ETH Zurich	Computational Physics	-Sep 2005

Experience

xAI (Oct 2025-Dec 2025 · Member of Technical Staff)

Macrohard CUA — Colossus 2 (1,112k H100) · large-scale video (pre-train) · RFT→SFT & eval (post-train)

- Macrohard computer-using agent (CUA) at [xAI](#)
 - Pre-training: large-scale Ray/Spark/sglang video pipelines for ~90k hours of CUA labeling/captioning
 - Post-training: RL/RFT traces → supervised captioner data
 - Eval: CUA diagnostics across perception, action, and benchmarks

H2O.ai (Dec 2013-Oct 2025 · CTO · technical leader)

Sovereign Enterprise AI / OSS (Java/Python/GPU) — distributed ML/AutoML/generative/agentic AI · raised \$256M

- Chief Architect since 2015; CTO since 2017; lead developer of 3 main revenue-generating enterprise platforms for predictive and generative AI at [H2O.ai](#) (~50k+ commits)

- **Enterprise h2oGPTe — LLM-powered enterprise GenAI platform (2023-2025)** (agentic AI webinar): #1 contributor (~12k of 41k commits) — primary author of the platform stack (Python services, chat workers, UI, Helm/Docker; enterprise hardening, performance/accuracy, long-term migration): RAG (hybrid retrieval, chunking, embeddings, reranking), document AI (ingest/OCR/vision), agentic chat/tools, streaming LLM guardrails, PII redaction, GPU model-service memory management, token-cost-based model routing (cost/accuracy/speed, willingness-to-wait), distributed job queues and multi-worker services, Redis worker registry (k8s-wide heartbeats), S3/MinIO object storage + file-server worker (agent scale-out), LLM/RAG self-tests, prompts and evals/benchmarks; private/on-prem.
* <https://h2o.ai/platform/enterprise-h2ogpte/>
- **Driverless AI — world's best AutoML (2017-2023)** (CBA, AT&T, NYT, Forbes DL, Forbes AI): #1 contributor (~19k of 58k commits) — designed, implemented, and productized the patented, award-winning no-code AutoML platform for tabular data, used by large banking, telco, healthcare, and public-sector customers (incl. CBA, AT&T, Kaiser, and Wells Fargo); personally drove genetic feature engineering, model search, ensembles, time-series AutoML, NLP (BERT/TextCNN/BiGRU/CharCNN), image models, GrowNet, model metrics (AUC/ROC, etc.), MLI/SHAP, custom recipes (data/features/models/scorers), experiment setup Wizard, enterprise hardening/performance/accuracy/long-term migration, GPU-accelerated Python backend (ML, deep learning, ensembling), Java/C++ production scoring pipeline (every experiment parity-tests scorer preds+SHAP for accuracy), and NVIDIA Triton inference deployment (with GUI for C++ scorer testing).
* <https://h2o.ai/platform/ai-cloud/make/h2o-driverless-ai/>
- **H2O-3 — open-source distributed ML (2014-2017)**: #1 contributor (~4k of 33k commits) — pioneered open-source scalable ML in production; personally drove distributed deep learning (MLP/autoencoder), GBM, random forest (DRF), k-means, aggregator, model metrics (AUC/ROC, etc.), exportable Java production scoring, Frame/Java runtime, Python/R APIs, and Flow UI; 100k+ community users.
* <https://github.com/h2oai/h2o-3/graphs/contributors?all=1>
- Worked regularly with advisers Rob Tibshirani, Trevor Hastie, and Stephen Boyd — they even cited my work
- Also #2 contributor to open-source [h2oGPT](#) (~780 commits) — fine-tuning infrastructure (LoRA/PEFT, bf16, multi-GPU), training-data pipelines, TGI/transformers inference, prompt templates, and GPU performance benchmarks
- Top-three contributor to open-source [H2O4GPU \(K-means CUDA\)](#) — CUDA/C++ GPU solvers for K-means and GLM (cuBLAS, CUB, Thrust, multi-GPU)
- Inventor on [patent](#)

Fortune • Wired (Spotlight • 2014-)

Fortune Big Data All-Star • ETH Alumni • ETH Globe • Wired • SingularityHub • KDD, GTC, Strata, H2O World

- Named a [Fortune Big Data All-Star \(2014\)](#)
- Personally featured in [ETH Alumni \(2015\)](#), [ETH Globe \(2015\)](#), [Wired \(2017\)](#), and [SingularityHub \(2018\)](#)
- Frequent speaker at KDD, NVIDIA GTC, Strata, H2O World, and Deep Learning conferences

Skytree (Jan 2012-Dec 2013 • Senior Member of Technical Staff • First technical hire)

HPC for Enterprise ML • American Express • in-memory ~100× faster than SOTA (C++/MPI) • \$20M Series A (2013) • Infosys

- #1 committer (~3k commits) — pioneered HPC methods for machine learning
 - Designed, implemented, and optimized proprietary scalable C++/MPI machine learning (k-NN, k-means, RF, GBM, collaborative filtering) — highest-performing Ensemble GBM in production at large enterprise banking customers (incl. American Express)
 - Collaborative filtering ~100× faster than prior state of the art (C++/MPI) — helped raise \$20M Series A (2013; American Express investor)
 - Advisers included James Demmel, Michael Jordan, and Pat Hanrahan

- Sped up k-NN $\sim 50\times$ through algorithmic optimizations; exact/approximate nearest neighbors, metric learning, weighted NN classification
- Large-scale C++/MPI/OpenMP ML server — core table/tree runtime, HDFS/Hadoop I/O, streaming queries, and model I/O
- Batch and streaming k-means, KDE/KDA; QA and cluster benchmarks

SLAC (Sep 2005-Jan 2012 · Stanford Linear Accelerator Center · Staff Scientist)

HPC · Accelerator Simulations (C++/MPI) · first 3D conformal FEM-PIC · CERN · DOE/SciDAC · Cray XT5 (224k cores)

- #1 committer ($\sim 4.7k$ commits) on [ACE3P](#) — parallel finite-element electromagnetics suite for accelerators (C++/MPI; DOE/SciDAC; Cray XT5, 224k cores)
 - Personally drove Pic3P/TFE3P — world's first 3D curvilinear conformal FEM-PIC for self-consistent space-charge-dominated particle-field modeling: charge-conserving beamloading, particle push/deposit, emitters, PIC mesh regions, and Newmark time integration (LCLS/XL5 RF guns and cavities) ([charge-conserving PIC beamloading note](#)) ([ACE3P paper](#))
 - Pioneered strong scaling on DOE leadership machines via [causal relativistic space-time / moving-window constraints](#); [Cray XT5, #1 supercomputer on top500.org](#)
 - Drove time-domain wakefield solvers (T3P/TFE3P beamloading & wake monitors) and contributed to frequency-domain RF eigenmode (Omega3P) and S-parameter (S3P) codes; [key simulations](#) for LCLS and CERN CLIC (est. \$10B+)
 - Also contributed across multipacting, thermal/EM (TEM3P), solvers, meshing/tools; collaborations with LBNL, LLNL, BNL, JLab, FNAL, KEK, PSI, and CERN

ETH Zurich (Foundations · Computational Physics)

Masters summa cum laude (N-body) · PhD Electromagnetics (C++/MPI) · Beowulf ~ 1000 -cores

- Computational Physics
 - Masters, summa cum laude ([Behaviour of different Poisson Solvers in the Light of Beam Dynamics Simulations](#))
 - * Diploma thesis (PSI / ETH): reduced classical $O(N^2)$ N-body Coulomb force evaluation to $O(N \log N)$ via Barnes-Hut and $O(N + M \log M)$ via FFT particle-mesh (vs direct PP); ran benchmarks and simulations on ETH Asgard — Europe's largest Beowulf cluster of its time; relativistic beam-dynamics simulations ([diploma thesis](#)) ([Asgard / SUSE](#))
 - PhD (C++/MPI) ([Simulation of electron source for next-generation X-ray free-electron laser](#))
 - * Advisors: Prof. Dr. [Matthias Troyer](#) (now Technical Fellow and CVP, Microsoft Quantum) and Prof. Dr. [Ralph Eichler](#) (former President, ETH Zurich)
 - * Designed and implemented a massively parallel MPI/C++ distributed 3D finite-difference particle-in-cell code for HPC (~ 1000 -core) — enabled $\sim 1000\times$ larger problems than the leading commercial code ([PhD thesis](#))
 - * Applied to the design of [SwissFEL](#) (\$250M) at [PSI](#)
 - * Taught undergraduate and graduate courses in computational physics and C++

Selected publications

- **2023** — H2O Open Ecosystem for State-of-the-art Large Language Models — EMNLP 2023 System Demonstrations
 - PDF: [/papers/h2o-open-ecosystem-emnlp2023.pdf](#)
 - doi:10.18653/v1/2023.emnlp-demo.6
- **2023** — h2oGPT: Democratizing Large Language Models — arXiv
 - PDF: [/papers/h2ogpt-arxiv.pdf](#)
 - doi:10.48550/arXiv.2306.08161
- **2017** — Benchmarks and Process Management in Data Science: Will We Ever Get Over the Mess? — KDD 2017
 - PDF: <https://dl.acm.org/doi/10.1145/3097983.3120998>
 - doi:10.1145/3097983.3120998

- **2012** — Modeling and design of an X-band rf photoinjector — Physical Review Special Topics - Accelerators and Beams
 - PDF: /papers/xband-photoinjector-prstab.pdf
 - doi:10.1103/PhysRevSTAB.15.102001
- **2011** — 50 MW X-Band RF System for a Photoinjector Test Station at LLNL — PAC / conference proceedings
 - PDF: /papers/50mw-xband-rf-pac2011.pdf
 - https://inspirehep.net/literature/1187269
- **2011** — A Reduced Gradient Output Design for SLAC's XL4 X-Band Klystron — PAC / conference proceedings
 - PDF: /papers/xl4-klystron-pac2011.pdf
 - https://inspirehep.net/literature/1183631
- **2011** — Advances in Parallel Electromagnetic Codes for Accelerator Science and Development — LINAC 2010
 - PDF: /papers/ace3p-advances-linac2010.pdf
 - https://inspirehep.net/literature/889625
- **2011** — An Optimized X-band Photoinjector Design for the LLNL MEGa-Ray Project — PAC / conference proceedings
 - PDF: /papers/xband-photoinjector-megaray-pac2011.pdf
 - https://inspirehep.net/literature/1187483
- **2011** — LLNL's Precision Compton Scattering Light Source
 - PDF: /papers/llnl-compton-light-source.pdf
 - https://inspirehep.net/literature/1363709
- **2011** — Numerical Validation of the CLIC/SwissFEL/FERMI Multi Purpose X Band Structure
 - PDF: /papers/clic-swissfel-fermi-validation.pdf
 - https://inspirehep.net/literature/1363762
- **2011** — Numerical Verification of the Power Transfer and Wakefield Coupling in the CLIC Two-Beam Accelerator — PAC / conference proceedings
 - PDF: /papers/clic-two-beam-verification-pac2011.pdf
 - https://inspirehep.net/literature/924506
- **2011** — Precision X-band Linac Technologies for Nuclear Photonics Gamma-ray Sources — PAC / conference proceedings
 - PDF: /papers/precision-xband-linac-ipac2011.pdf
 - https://inspirehep.net/literature/1183324
- **2011** — X-Band RF Photoinjector Research and Development at LLNL — PAC / conference proceedings
 - PDF: /papers/xband-rf-photoinjector-llnl-pac2011.pdf
 - https://inspirehep.net/literature/1187329
- **2011** — X-Band Test Station at Lawrence Livermore National Laboratory — PAC / conference proceedings
 - PDF: /papers/xband-test-station-llnl.pdf
 - https://inspirehep.net/literature/1183683
- **2010** — ACE3P Computations of Wakefield Coupling in the CLIC Two-Beam Accelerator
 - PDF: /papers/ace3p-wakefield-clic-two-beam-2010.pdf
 - https://inspirehep.net/literature/883814
- **2010** — Dark Current Simulation for the CLIC T18 High Gradient Structure
 - PDF: /papers/dark-current-clic-t18.pdf
 - https://inspirehep.net/literature/1378733
- **2010** — On using moving windows in finite element time domain simulation for long accelerator structures — Journal of Computational Physics
 - PDF: /papers/moving-window-fetd.pdf
 - doi:10.1016/j.jcp.2010.08.037
- **2009** — High-Fidelity RF Gun Simulations with the Parallel 3D Finite Element Particle-In-Cell Code Pic3P — AIP Conference Proceedings
 - PDF: /papers/pic3p-rf-gun.pdf
 - doi:10.1063/1.3215604
- **2009** — Parallel 3D Finite Element Particle-in-Cell Simulations with Pic3P
 - PDF: /papers/pic3p-parallel.pdf

- <https://inspirehep.net/literature/823681>
- **2009** — Parallel Higher-order Finite Element Method for Accurate Field Computations in Wakefield and PIC Simulations
 - PDF: </papers/higher-order-fem-wakefield-pic.pdf>
 - <https://inspirehep.net/literature/823685>
- **2009** — State of the art in electromagnetic modeling for the Compact Linear Collider — Journal of Physics: Conference Series
 - PDF: </papers/clic-em-modeling-sota.pdf>
 - doi:10.1088/1742-6596/180/1/012004
- **2009** — Wakefield Computations for the CLIC PETS using the Parallel Finite Element Time-Domain Code T3P
 - PDF: </papers/wakefield-clic-pets-t3p.pdf>
 - <https://inspirehep.net/literature/823683>
- **2009** — Wakefield Simulation of CLIC PETS Structure Using Parallel 3D Finite Element Time-Domain Solver T3P
 - PDF: </papers/wakefield-clic-pets-t3p-sim.pdf>
 - <https://inspirehep.net/literature/823682>
- **2008** — Computational Science Research in Support of Petascale Electromagnetic Modeling — Journal of Physics: Conference Series
 - PDF: </papers/petascale-em-compsci.pdf>
 - doi:10.1088/1742-6596/125/1/012077
- **2008** — Design and Optimization of Large Accelerator Systems through High-Fidelity Electromagnetic Simulations — Journal of Physics: Conference Series
 - PDF: </papers/design-opt-large-accelerators.pdf>
 - doi:10.1088/1742-6596/125/1/012003
- **2008** — On Projecting Discretized Electromagnetic Fields with Unstructured Grids
 - PDF: </papers/projecting-em-fields-unstructured.pdf>
 - <https://inspirehep.net/literature/793050>
- **2008** — Parallel 3D Finite Element Numerical Modelling of DC Electron Guns
 - PDF: </papers/dc-electron-guns-fem.pdf>
 - <https://inspirehep.net/literature/779274>
- **2008** — Parallel Computation of Integrated Electromagnetic, Thermal and Structural Effects for Accelerator Cavities — PAC / conference proceedings
 - PDF: </papers/em-thermal-structural-cavities.pdf>
 - <https://inspirehep.net/literature/794909>
- **2007** — Parallel Finite Element Particle In Cell Code for Simulations of Space charge Dominated Beam Cavity Interactions — PAC / conference proceedings
 - PDF: </papers/pic-spacecharge-pac07.pdf>
 - doi:10.1109/PAC.2007.4440757
- **2007** — Towards Simulation of Electromagnetics and Beam Physics at the Petascale — PAC / conference proceedings
 - PDF: </papers/em-beam-petascale-pac07.pdf>
 - doi:10.1109/PAC.2007.4441117
- **2006** — A Massively parallel particle-in-cell code for the simulation of field-emitter based electron sources — Nuclear Instruments and Methods A
 - PDF: <https://doi.org/10.1016/j.nima.2005.11.059>
 - doi:10.1016/j.nima.2005.11.059
- **2006** — State of the Art in EM Field Computation — PAC / conference proceedings
 - PDF: </papers/state-of-art-em-field-computation.pdf>
 - <https://inspirehep.net/literature/727218>
- **2005** — Low Emittance X-FEL Development — eConf
 - PDF: </papers/low-emittance-xfel-development.pdf>
 - <https://inspirehep.net/literature/703609>
- **2005** — Simulation of Electron Source for Next-Generation X-ray Free-Electron Laser — PhD thesis, ETH Zürich
 - PDF: </papers/phd-thesis.pdf>
 - doi:10.3929/ethz-a-005082871
- **2004** — Electron Beam Dynamics Simulations for the Low Emittance Gun

- PDF: /papers/electron-beam-dynamics-leg.pdf
- <https://inspirehep.net/literature/670125>
- **2004** — Low emittance gun project based on field emission
 - PDF: /papers/low-emittance-gun-field-emission.pdf
 - <https://inspirehep.net/literature/671756>
- **2004** — Preliminary Results on a Low Emittance Gun Based on Field Emission
 - PDF: /papers/preliminary-low-emittance-gun.pdf
 - <https://inspirehep.net/literature/670994>
- **2001** — Behaviour of different Poisson Solvers in the Light of Beam Dynamics Simulations — Diploma thesis, ETH Zürich / PSI
 - PDF: /papers/diploma-thesis-poisson-solvers.pdf
- **1999** — Minimizing damage during focused-ion-beam induced desorption of hydrogen — J. Vac. Sci. Technol. B 17, 2443 (1999)
 - PDF: /papers/fib-hydrogen-desorption-jvstb1999.pdf
 - doi:10.1116/1.591108

Projects (hosted at <https://candel.org>)

- **Automatic Data Scientist** [AutoML, Tabular] (GPU): Tabular AutoML (classification/regression) with feature-engineering pipelines, ensembles, ONNX serving, and a training UI — LLM-steered search (MadPilot) plus MCP server for external agents.
- **Exact & Scalable Kelly** [Streamlit, Quant finance]: Portfolio optimization with exact Kelly criterion — ~100× faster than conventional covariance methods; rolling backtests and benchmarks on real market data.
- **Analytic Ray Tracing** [Streamlit, Telecom] (GPU): Telecom digital-twin demo — analytic ray tracing over a synthetic city geometry with Numba/CUDA acceleration.
- **Analytic Dosimetry** [Streamlit, Dosimetry] (GPU): Spectral RBE coprocessor — translate macroscopic physical dose into biological damage estimates with interactive spectral analysis.
- **Exact Structured Antibiotic Twin** [Streamlit, Biology]: Exact hidden-state twin for antibiotic strategy — compare regimens on viable burden, filamented fraction, and biomass spectrum with exact-vs-grid timing.
- **Analytic Cone-Beam CT Scatter** [Streamlit, CBCT] (GPU): CBCT scatter coprocessor demo — analytic multi-scatter resolvent vs Monte Carlo, SPR maps, and reconstruction artifact correction on a stylized thorax slice.
- **Analytic Dark-Field CT** [Streamlit, CT]: Dark-field imaging coprocessor demo — analytic vs Monte Carlo dark-field signal for lung and breast X-ray CT with interactive ray diagnostics.
- **Cinema Acoustics** [Streamlit, Acoustics] (GPU): Acoura parametric FEM for cinema acoustics — RP22 speaker layout, symmetric subwoofer optimization, and CuPy-accelerated Helmholtz solvers on theater meshes.
- **Casimir Transistor** [Streamlit, Physics]: Interactive Casimir transistor design — finite-T Lifshitz maps, exact planar gate-slab residual, and nonadditivity where pairwise intuition breaks down.
- **Standard Model of Physics** [Physics, Lean]: Arithmetic canonicity of a Lerch encoding for the finite SM+ ν_R matter architecture — affine-orbit minimality, Hecke eigenfunctions, finite Weyl algebras, and audited matter interfaces.

Machine-readable discovery

- <https://candel.org/llms.txt>
- <https://candel.org/llms-full.txt>
- <https://candel.org/resume.md>
- <https://candel.org/resume.pdf>
- <https://candel.org/sitemap.xml>
- <https://candel.org/robots.txt>

Generated from `src/config.ts` (same reverse-chrono eras + full credentials as Explore).