Nathan Douglas

(330) 998-8049 | olive.summit8245@tenesm.us | linkedin.com/in/nug-doug | github.com/ndouglas | bottomuplabs.ai

Software engineer and generalist moving into AI safety, 14 years in production systems, most recently the engineer of record for four healthcare services handling nearly two million requests a day. Interested in the collective behavior of interacting AI systems that single-agent evaluations can miss, and in bringing production discipline to research: reproducible experiments, tests as gates, and hand-verified evidence behind every claim.

Experience

Agile Six Applications, Inc (contract to U.S. Department of Veterans Affairs) Oct. 2020 – Present

Senior Platform Engineer, VA.gov Check-In *Feb. 2024 – Present*

- Founded the team's developer platform: Terraform and Terragrunt for a new AWS GovCloud account and 40+ reusable modules, each integration-tested by applying real infrastructure in two AWS accounts on every pull request.
- Led a multi-year migration of four production services to a new AWS account through a hard cutover, then decommissioned the legacy estate with evidence-gated, dry-run-by-default tooling.
- Traced a sub-0.01% error rate to a long-standing silent authorization defect and shipped the fix overnight; wrote the team's first postmortem and its 24-hour postmortem policy, every action item closed.
- Eliminated long-lived credentials from CI/CD with OIDC-assumed roles in every repository, GitHub App tokens, SHA-pinned actions, and an hourly vulnerability-count poller with its own dead-man monitor.
- Adopted AI-assisted engineering end to end under a spec-then-plan workflow, with specs, plans, and review artifacts committed beside the code and every claim backed by hand-verified evidence.
- Engineer of record for the four production services behind VA.gov patient check-in, nearly two million requests a day, owning application code, infrastructure, delivery pipeline, and the pager.
- Built the end-to-end system-test suite, 200+ tests exercising real EHR appointment flows against deployed infrastructure, with trace context propagated so a failing assertion links to the traces behind it.

Technical Lead, VA.gov CMS *Jul. 2023 – Feb. 2024*

- Re-architected the VA.gov publishing pipeline into five tested Drupal modules with zero downtime, shipping the replacement beside the legacy path and deleting the old code only after test parity.
- Ran and wrote up the team's production incidents with minute-level timelines and ticketed action items, then proposed the PagerDuty routing and on-call changes that followed.
- Led the Drupal 10 program for a multi-vendor CMS with dozens of contributors: seeded it over a year early with a module-by-module checklist and two dozen tickets other teams could pick up.

Senior Software/DevOps Engineer, VA.gov CMS *Oct. 2020 – Jul. 2023*

- Hired as a backend engineer; started and led the quality program that became the team's test strategy, growing unit tests fivefold, adding static analysis and accessibility checks, and writing most of the test codebase.
- Moved CI into GitHub Actions and unified every check into one command surface that runs identically on a laptop, in previews, and in CI, removing thousands of lines in the process.
- Built DORA-style delivery metrics from scratch (lead time, change failure rate, time to restore) by walking Git history and GitHub statuses, emitted from the deploy pipeline even when tests fail.
- Hardened GitHub Actions as a supply-chain surface: SHA-pinned every action, wrote the hardening guide and an auto-posted security checklist, and answered a CISA advisory the same day.

Convention News Company Jan. 2016 – Oct. 2020

Senior Software/DevOps Engineer

- Joined to improve developer experience and left owning the platform: migrated every service off a managed Drupal host onto AWS at substantially lower annual cost, and built the backend and its public APIs.
- Designed and built the subscription management service from scratch: Stripe payments, Cognito identity, DynamoDB, and Lambda with CloudWatch monitoring, at full test coverage; it ran in production for years afterward.
- Engineered a headless Drupal and Next.js architecture on AWS with independently scaling frontend and backend.

DEVONtechnologies Jan. 2012 – Dec. 2015

Software Engineer

- Owned the sync engine behind DEVONthink and its iOS app, DEVONthink To Go, for four years: took it over during a high-risk engineer transition, shipped the first version, and completed the second before leaving.
- Served as front-line support for the product at a small company, turning customer reports into reproductions, fixes, and release notes.

Projects

🦷 **Goldentooth**: 17-node Talos Linux Kubernetes cluster run entirely through Flux GitOps: Cilium, cert-manager with an internal ACME CA, SeaweedFS, Prometheus/Loki/Tempo observability, Chaos Mesh fault injection, Falco runtime security, and a hundred-plus-entry engineering journal of what broke and why (*Talos, Kubernetes, Flux CD, Cilium*)
🔍 **LLM Weight-Audit Blind Spots**: Solo Apart Research hackathon project on weight-based audits: when frontier LLMs read tiny nets' weights as text, training with Adam instead of SGD made MNIST nets read as Fashion-MNIST with no weight edits, and a message hidden in the weights was flagged reliably only at maximum effort; one reproducibility check re-derives every reported number (*Python, PyTorch, Claude API*)
🔁 **Adaptive Recursive Inference**: Looped the middle layers of Qwen2.5-7B until hidden states converged and tested convergence as a cheap uncertainty signal: it predicted math correctness (AUC 0.69) but inverted on multi-step reasoning, with a full write-up (*Python, PyTorch, Transformers*)
⚖️ **LLM Cartel Interventions**: Solo Apart Research sprint testing interventions against GPT-4.1 seller cartels in a double-auction benchmark: a legal notice given with the CEO's price-fixing order prevented the cartel, but the same notice an hour later barely moved it; benchmark fix merged upstream and all runs released (*Python, GPT-4.1*)
🛏️ **BottomUpLab**: Agent-based models of societies, markets, games, and minds, each reproduced from its source paper and measured against it, built AI-assisted with seeded experiments, parameter sweeps, replay, and regression tests; a reproducibility register records findings that do not reproduce (*Rust, WebAssembly, TypeScript*)
🕵️ **Rills**: Social-deduction game played by autonomous Claude agents, with probabilistic night actions, deliberately ambiguous trait-based evidence, and a tested resolution engine (*Python, Claude API*)

Education

ARENA (self-study) — Online
Alignment Research Engineer Accelerator curriculum, worked through in public from November 2026 — *Nov. 2026*

Apart Research — Online
AI Collusion Research Sprint, October 23–25; solo project submitted — *Oct. 2026*

BlueDot Impact — Online
Technical AI Safety course, October 26–31 cohort — *Oct. 2026*

BlueDot Impact — Online
AGI Strategy course; action plan on collective AI behavior, evaluation, and control — *Oct. 2026*

University of Nevada, Las Vegas — Las Vegas, NV
Bachelor of Arts in Computer Science; minors in Psychology and Sociology — *Dec. 2011*

Technical Skills

Languages: TypeScript/JavaScript (Node.js), Python, Rust, Bash, HCL, Clojure, Go, C
AWS: Lambda (aliases, provisioned concurrency, layers, extensions), API Gateway (private, OpenAPI import, authorizers, usage plans), DynamoDB, AppConfig, IAM (OIDC federation, permissions boundaries, tag-conditioned policies), KMS, EC2/Auto Scaling, ElastiCache, CloudFormation/SAM
Infrastructure as Code: Terraform (native tests, import/moved/removed blocks, preconditions), Terragrunt, Ansible, Packer
Delivery: GitHub Actions (composite and reusable workflows, OIDC, matrix sharding, self-hosted runners), Jenkins, Dependabot and Renovate, CodeQL, CodeBuild
Observability: Datadog (APM, monitors-as-code, synthetics, log pipelines, CI Visibility), Prometheus, Grafana, Loki and Tempo, CloudWatch, PagerDuty
Testing: Jest, Cypress, PHPUnit, PHPStan, axe / WCAG 2.1, MSW, terraform test, shellcheck
Platforms: Kubernetes (Talos, Flux CD, Cilium, Helm), Docker
AI and ML: PyTorch and Hugging Face Transformers, scikit-learn, NumPy, Jupyter, LLM agents and the Claude API, Agentic coding (Claude Code, spec-then-plan workflows)
Domain: HIPAA / PHI handling