

LITO ZARATE BIALA

HANDS-ON AI BUILDER & SITE RELIABILITY ENGINEERING LEAD | AGENTIC SYSTEMS · MCP · INCIDENT MANAGEMENT · AI- SAFETY TOOLING · AGENT EVALUATION · PROMPT-INJECTION DETECTION

Minneapolis-St. Paul, MN · lzbiala@gmail.com · linkedin.com/in/lzbiala · github.com/LZBiala · lzbiala.github.io

SUMMARY

Hands-on AI builder and site reliability lead at Wells Fargo. I find the bottleneck that needs a foundational fix, build the AI that clears it, and make the system prove it worked. In public I build AI-safety and agent-evaluation tooling and publish the numbers with their misses: six CI-verified repos where the build fails if a measured claim drifts. Day to day I run SRE and incident practice as the working link between Product, Platform, Development, and Infrastructure: I turn each group's constraints into decisions the others can act on, and I ship governed GenAI products end to end, backed by 12+ years of leadership across people management and technical delivery. The same approach took major incidents on the API gateway platform I own from 10 in 2024 to 1 year to date in 2026, and cut mean time to detect and react by 40%. Public, CI-verified evaluation portfolio at github.com/LZBiala - the build fails if a measured claim drifts.

AI LEADERSHIP & TECHNICAL STRENGTHS

AI Product	Use-case discovery with the teams who own the problem, requirements, acceptance criteria, roadmap sequencing, adoption, executive and non-engineer communication.
AI safety & evals	Prompt-injection and skill-poisoning detection (skill-tripwire), judge instruction-flip audits (adversarial-chambers), defect-seeding harness with best-of-k jury arithmetic (agent-mutation-lab), placebo-controlled evals on sealed held-out splits (memory-repair-lab), MCP servers, RAG, prompt and context engineering, agents, human-in-the-loop controls.
Delivery	Python, Streamlit, TypeScript, GitHub, CI/CD, automated tests, code review, iterative PR-based delivery.
Reliability	SRE, DevOps, incident management, 5-Why/RCA, MTTx, observability, fail-closed behavior, AI risk and compliance readiness.

PUBLIC AI EVALUATION & RELIABILITY PORTFOLIO · [GITHUB.COM/LZBIALA](https://github.com/LZBiala) · [LZBIALA.GITHUB.IO](https://lzbiala.github.io)

- skill-tripwire:** fail-closed scanner for poisoned agent skill and config files (SKILL.md, MCP tool descriptions, CLAUDE.md); catches hidden codepoints, download-and-execute, over-broad shell grants, and exfiltration; seeded million-input run: 0 crashes, 0 hangs, 0 false alarms on 450,360 synthetic benign files (fixed-seed reproducible); recall published as a self-authored upper bound with every miss named.
- oncall-router-mcp:** MCP server for incident routing; five tools, every one fails closed; mttx_review measures where an incident's minutes went (Detect/React/Mitigate) and routes to the tool that shrinks the dominant phase, no grade by design; zero dependencies, no keys; CI-regenerated docs.
- agent-mutation-lab:** behavioral defect-detection harness; the bundled reviewer flags 9/10 planted defects at the planted line with 0 false alarms on byte-identical clean controls, published as harness conformance, not a catch rate; best-of-k jury study with a copycat control; every figure also derivable in SQL (stdlib SQLite, test-asserted).
- adversarial-chambers:** proposer, skeptic, and judge with no role played twice; instruction-flip test on two scripted judges, true by construction: an evidence-led judge flips 0/8 verdicts, a deferential judge flips 3/8, exactly the ambiguous items. Proves the instrument detects instruction sensitivity, not that any live judge passes.
- wiki-memory-lab:** agent-memory evaluation; selective recall at 0.61× the tokens of full context, the 1.27× losing case published.
memory-repair-lab: placebo-controlled self-repair, 8/10 to 10/10 on a split sealed before generation zero.

All six are CI-verified on Windows and Linux: the build fails if a measured claim drifts from regenerated data. Each ships a walkthrough a stakeholder can click through.

PROFESSIONAL EXPERIENCE

Vice President, Lead Digital Platform Services Manager

Feb 2025 - Present

Wells Fargo, Digital Technology & Innovation · Individual-contributor leadership role · Minneapolis, MN

- AI product delivery:** lead delivery of two governed AI products, an incident-notification quality assistant and a Daily 5-Why QA platform, turning operational bottlenecks into buildable specs gated by human review and grounded in SRE and DevOps practice.
- Hands-on build:** daily engineering in Python with Claude Code, Codex, and Devin: focused branches, pull requests, tests, coverage, and type checks, so every change arrives reviewable.

- **Governance by design:** built the evidence architecture behind both products, pulling live incident and RCA data from ServiceNow and SharePoint so every record carries its provenance and passes schema validation, redacts diagnostics, fails closed on bad data, and needs explicit approval before touching production. I also drew the boundary that kept person-level scoring out of published outputs, keeping names only for routing work.
- **On-call AI workflows:** built AI workflows that tell an on-call engineer who to escalate to and when, measured from impact start rather than from ticket creation, cutting mean time to mitigate and mean time to react for me and my teammates during major incidents.
- **Incident reduction:** major incidents escalated against the API gateway platform I own fell from 10 in 2024 to 4 in 2025 to 1 year to date in 2026, on one of the noisiest and most business-critical surfaces in the enterprise.
- **Escalation design:** lead major-incident, vulnerability-remediation, and operating-review readiness; wrote escalation playbooks an on-call engineer can pick up cold and use to reach the right team in minutes, built on a working map of the organization's critical applications and their dependencies. I carry weekend on-call for those systems.
- **Executive reporting:** run 5-why root-cause analysis to the true cause, then present findings, preventive actions, and impact trends to senior leadership so remediation investment follows the evidence rather than the loudest incident.

Senior Assistant Vice President, Digital Product Service Manager

Oct 2022 - Feb 2025

Wells Fargo · Minneapolis, MN

- **Reliability:** cut mean time to detect and mean time to react by 40% on customer-critical applications through proactive engineering, 5-why root-cause discipline, and DevOps practice.
- **Observability:** enhanced log-based alerting and monitoring through platform migrations (Splunk, Grafana, AppDynamics), preserving operational continuity with executive-level service-health insight.
- **Platform delivery:** engineered foundational platforms for scale through cloud migration and data-center exit; led engineering release meetings across AppDev, Platform, and Product teams via CI/CD; guided distributed US/India engineering teams.

Senior Technology Business Systems Consultant

Jan 2022 - Oct 2022

Wells Fargo, Digital Technology & Innovation

Incident and problem management for production applications under SRE, ITIL, and ITSM practice.

Earlier roles, including API-health analytics (Software Product Analyst, 2021) and people leadership - full history on LinkedIn.

SELECTED ENTERPRISE AI INITIATIVES (2026)

Incident Quality Assurance Platform

Wells Fargo

Root-cause quality assurance for an enterprise incident practice

- **Scoring model:** defined the delivery plan and the scoring model: versioned rubrics, a standing rule that every root-cause claim cites its supporting evidence, quality checks on preventive actions, and acceptance tests. Outputs ship as advisory, pending human review.
- **Evidence-aware scoring:** designed the model so incomplete or non-evaluable material is never reported as a pass or a fail, which is what made the scores trustworthy enough for people to act on.

TECHNOLOGY

AI/ML	Generative AI, RAG, agentic automation, LLM evaluation frameworks, prompt-injection detection, prompt optimization, responsible AI, MCP servers, knowledge graphs, deep-wiki agent long-term memory
Engineering	Python, SQL (SQLite), Streamlit, TypeScript, Git, GitHub, GitHub Actions, APIs, CI/CD, automated testing
Operations	SRE, DevOps, Splunk, Grafana, AppDynamics, production triage, observability, incident and problem management, 5-Why RCA, MTTx, ITIL/ITSM
Leadership	AI product strategy, distributed-team leadership, technical writing, executive communication, risk and compliance
Agentic	Claude Code, Codex, Devin (Cognition)

EDUCATION & CREDENTIALS

Data Analytics & Data Visualization Boot Camp Certificate, University of Minnesota Twin Cities · B.A. Psychology, Metropolitan State University
Splunk Certified: Fundamentals 1-3 and Cloud Observability · Wells Fargo internal SRE-leadership and AI-engineering program pathways
Anthropic Academy, 2026: Claude Code in Action · Introduction to Model Context Protocol