Theofanis Tsakanikas

Responsible AI Data Engineer — GenAI & RAG · Guardrails & Governance as Code · Data Platforms

Athens, Greece · **Remote only** (EU · CET/CEST) · **available now**, B2B or employment
theofanis@tsakanikas.io · tsakanikas.io · linkedin.com/in/theofanis-tsakanikas · github.com/theofanis-tsakanikas

PROFILE

AI data engineer who builds **the trust layer that makes AI safe to ship** — deterministic guardrails, data quality and lineage, drift monitoring, governance as code, and EU AI Act technical documentation generated from the system itself — and the data platforms underneath them. Nine end-to-end systems on AWS, Azure, GCP, Databricks and Snowflake, every one provisioned by CI on live cloud, proven end to end, then destroyed. MEng ECE (NTUA); six AWS / Databricks / Terraform / IAPP certifications, including the **AIGP**.

> **At a glance:** 9 end-to-end platforms, all public — code and CI open to inspection · ~2,700 credential-free CI-gated tests · gates attacked on every run: **80 planted, 80 refused** · AWS · Azure · GCP · Databricks · Snowflake

TECHNICAL SKILLS

- **Responsible AI:** Guardrails (PII redaction, denied topics, prompt-attack filters, contextual grounding) · prompt-injection mitigation · anti-hallucination gates · adversarial red-teaming in CI · drift (PSI, KS, z-test) · EU AI Act (Annex IV, Art. 12) · model & dataset cards · GDPR Art. 9 / Art. 30
- **GenAI & RAG:** AWS Bedrock (Agents · AgentCore · action groups · Knowledge Bases · Guardrails) · retrieval design (chunking, metadata filtering, grounding thresholds, citation provenance) · Databricks Mosaic AI (Agent Framework, Vector Search, Genie) · LangGraph · LangChain · LangSmith · LLM eval & red-team harnesses
- **Vector search:** OpenSearch Serverless · Pinecone · Databricks Vector Search · Titan Text Embeddings v2 · GTE-Large
- **Data & platforms:** Spark Structured Streaming · Kafka (Avro + Schema Registry) · Airflow · dbt · Delta Lake · Apache Iceberg · Medallion · Trino · Redis TimeSeries · BigQuery · Glue/Athena · Databricks (Unity Catalog, DLT, Model Serving, Asset Bundles) · Snowflake · data contracts · SCD-2
- **Cloud & engineering:** AWS (Bedrock, Lambda, OpenSearch, MSK, ECS/EKS, S3, Secrets Manager, KMS) · Azure (AKS) · GCP (GKE Autopilot, BigQuery) · Terraform · Terragrunt · Docker · Kubernetes · GitHub Actions (keyless OIDC) · Python · SQL · Bash · pytest & offline eval harnesses · Grafana · Prometheus · MLflow · XGBoost

SELECTED PROJECTS — all nine public on GitHub, links live

FintelliGuard — enterprise RAG & compliance agent on AWS Bedrock

592 tests

AWS Bedrock (Agent · Knowledge Base · Guardrails) · OpenSearch Serverless · Lambda · Databricks Mosaic AI · XGBoost · Kafka/MSK · Terraform
github.com/theofanis-tsakanikas/fintelliguard

- **RAG grounded in law.** A Bedrock Agent answers over a Knowledge Base of verbatim EUR-Lex regulation (AMLD5, GDPR, PSD2, RTS-SCA) — OpenSearch Serverless, Titan v2 embeddings — citing its exact source; its only tool is a VPC-internal Lambda action group, so it never touches raw data.
- **Anti-hallucination as code, not as a prompt.** A five-check deterministic gate (schema · no-PII · grounding · faithfulness · decision): every cited article must exist in the retrieved context, and the agent may escalate but never soften a decision. Guardrails bound at an immutable version, and the gates themselves attacked on every pull request — **80 planted violations, each refused by the named check**.
- **Shipped, not sketched.** One CI dispatch provisioned the whole AWS + Databricks estate in **1h 37m** (sub-50 ms scoring, promotion gate passed at AUC 0.866), then destroyed it to zero cost. EU AI Act Annex IV documentation is generated from the code and CI-checked.

Manifest — document intelligence for cross-border trade

527 tests

AWS Textract · Bedrock Data Automation · Bedrock · SageMaker · Step Functions · Iceberg on S3 · Athena · OpenSearch · Redshift · EMR Serverless · Terraform
github.com/theofanis-tsakanikas/manifest

- **The confidence threshold is derived, never chosen.** Each field declares an **error budget**; the threshold is the lowest score whose 95% upper bound on published-and-wrong still fits it. Where none fits, the field is **always-review** — **31 of 36**, reason named per field. Publishing everything is 26.72% wrong, a hand-picked 0.85 is 4.68%, derived is **0.22%**.
- **Models propose; deterministic code decides.** A four-tier cascade routed by declared language eligibility, not tried in a line. A published field that cannot be located on the page is a *build failure*: **120/120** corrupted boxes refused. ECE **0.0592** on real photographed pages against 0.0815 on the generated corpus.
- **Shipped, then destroyed.** `gate-proof` plants **61** violations and requires the named gate to refuse each; the review queue is measured against a declared capacity the build fails over. One dispatch applied five Terraform layers in **49m 32s**, verified **34/34** against the real account, then destroyed it.

Attestor — multi-tenant regulated report factory on AWS Bedrock AgentCore

489 tests

AWS Bedrock AgentCore (Runtime · Gateway · Identity · Memory · policy engine · observability) · Knowledge Bases · Guardrails · OpenSearch Serverless · Iceberg / Athena · Cedar · Terraform
github.com/theofanis-tsakanikas/attestor

- **A contract cannot pre-authorize its own failure.** The reasons a figure goes undisclosed are a closed vocabulary: lawful omissions are printed and an auditor accepts them; internal failures *block the report*. A contract may declare the first and never the second — that is how "we could not compute it" becomes "it was not material".
- **The LLM never owns a number.** Deterministic code resolves every figure through declared SQL over a pinned Iceberg snapshot and owns the decision to abstain; the gate scans the *rendered* DOCX/XLSX/PPTX, not the data behind it — **537 numerals** checked, **4/4** artefacts clean across CSRD/ESRS and EU AI Act Annex IV.
- **Shipped, then destroyed. 12/12** cross-tenant routes closed, **16/16** poisoned documents flagged with **0/11** benign wrongly. `gate-proof` breaks each control and requires the *named* gate to refuse it: **25 planted, 25 refused**. One dispatch applied four Terraform layers in **21m 47s**, verified **32/32** live, then tore it down.

Watermark — real-time decision platform for an electricity distribution network 341 tests

AWS IoT Core · Kinesis · Managed Service for Apache Flink · Iceberg on S3 · Athena · dbt · SageMaker (Feature Store · Registry · endpoint) · Lake Formation · Step Functions · Terraform

github.com/theofanis-tsakanikas/watermark

- **Fail closed is the wrong reflex on a grid.** The transformer keeps heating while nobody decides, so the safe state is a deterministic action that confesses to being one: **485,442 W against a declared 450,000 W → 8 sessions throttled**, every one marked as a fallback from the actuator into the record.
- **No decision comes out of a window that has not closed.** One stream, three decisions, three definitions of "enough": curtailment in seconds, meter anomaly in hours, settlement restated over days. Live — **0 of 3,779 rows** published early, as SQL against the deployed table; **285 restatements, 285 naming what they replaced**.
- **The promotion gate refused the model I trained.** Per-group precision **1000/1000** in the most deprived tercile against **181/1000** in the least — because inspectors confirmed 66 of 66 there and 4 of 23 elsewhere. The model looks flawless where the dispatch log is densest. Written up, not deleted. A decision with a significant effect on a person cannot be actuated automatically: **20 pending, 0 actuated**.

Self-Healing Multi-Cloud Agents — LangGraph multi-agent system 380 tests

LangGraph · Pinecone RAG · AWS · Azure · GCP · Databricks · Terraform · GitHub Actions
github.com/theofanis-tsakanikas/self-healing-multi-cloud-agents

- **Four platforms, one code path.** A Supervisor/Architect/Infra/Medic graph designs, deploys and *repairs* data pipelines on AWS, Azure, GCP and Databricks — the Medic reads the real CI logs, patches the exact line and redeploys with no human in the loop.
- **Standards-first RAG, evidence-gated fixes.** Agents retrieve versioned engineering standards from Pinecone at generation time; a fix request is refused unless its evidence quote appears verbatim in real output. The heal loop is bounded and fail-closed.
- **Agent reliability, measured.** An offline harness scores routing and the evidence gate over 17 cases across 14 failure classes with no LLM, cloud or keys (100%), plus a live mode that catches model regressions.

Multi-Cloud Governance Platform — governance-as-code + a bounded copilot 137 tests

Databricks Unity Catalog · Snowflake · Terragrunt · AWS · Azure · GCP · OPA/Rego
github.com/theofanis-tsakanikas/multicloud-governance-platform

- **One JSON contract, two engines.** The same contract drives Unity Catalog across three clouds *and* Snowflake, with a checker that proves both enforce equivalent capabilities.
- **Governance as a merge gate, not a report.** A credential-free analyzer fails any pull request that grants read on PII, cross-checked in OPA/Rego and packaged as a reusable `govgate` CLI + GitHub Action; documented exceptions expire and turn CI red again.
- **The copilot is subordinate to the deterministic core.** A Genie space caged to four read-only governance tables shows the SQL behind every answer and declines what it may not know.

Fleet Risk Lakehouse — real-time driver-risk analytics on Databricks & AWS 173 tests

Databricks · Spark Structured Streaming · Unity Catalog · Terraform (5 layers) · Amazon Managed Grafana
github.com/theofanis-tsakanikas/fleet-risk-lakehouse

- **Explainable by design.** A Medallion lakehouse correlates vehicle telemetry with driver biometrics over a ±60-second temporal join; the risk score emits each factor's contribution, defined once in code so the generated model card cannot drift from it.
- **GDPR Art. 9 enforced, not documented.** Unity Catalog column masks NULL the biometrics for unprivileged principals — the same query returns different data depending on who runs it — and outgoing Slack / PagerDuty alerts carry no special-category fields.

Real-Time Telemetry Pipeline (GCP) & Contract-Driven Data Pipeline (AWS) 102 + 63 tests

Kafka (Avro + Schema Registry) · Spark · BigQuery · dbt · Redis TimeSeries · GKE Autopilot · Airflow · PySpark · S3 · Athena
github.com/theofanis-tsakanikas/realtime-telemetry-pipeline · contract-driven-data-pipeline

- **GCP:** a keyless, 100%-IaC streaming stack on GKE Autopilot, dual-sinking a validated stream to Redis TimeSeries and BigQuery with dbt marts; a z-test drift detector catches a miscalibrated sensor whose readings are all still in range → 3σ Slack alert.
- **AWS:** an Airflow/PySpark ETL where a declared data contract is the single source of truth for validation, rejection lineage and PII pseudonymisation; the generated data dictionary is CI-checked against it.

CERTIFICATIONS

- AWS Certified Generative AI Developer — Professional
- AWS Certified Data Engineer — Associate
- HashiCorp Certified: Terraform Associate (003)
- Databricks Certified Generative AI Engineer — Associate
- Databricks Certified Data Engineer — Associate
- IAPP AI Governance Professional (AIGP)

EARLIER CAREER — leadership · digital transformation · public speaking

Programs Director & Regional Lead — national non-profit organisation · Thessaloniki, Greece · 2018–2025

- Led the organisation's **youth programs nationwide** across all departments while directing the Northern Greece branch — setting direction, coordinating local teams, and owning the full annual programs budget.
- Built regional programs from the ground up — new initiatives, events and regional outreach that **tripled the number of participants**.
- **Drove the digital transformation** of the branch's operations and workflows.

Operations Lead — Logistics & Digital · publishing & distribution arm · Athens, Greece · 2014–2018

- Ran **digital operations** — the organisation's e-shop and website — and supervised the warehouse and international fulfilment operation, managing staff and worldwide shipping.

EDUCATION & LANGUAGES

MEng, Electrical & Computer Engineering — National Technical University of Athens (NTUA) · Greek (native) · English (C2 — Cambridge CPE) · German (C1)

Infrastructure as Code, always · Security by default · Observability from day one · Tested before shipped · Deterministic gates around probabilistic models