

WHERE DID IT REALLY START?

OriginZero traces health disparities in underrepresented communities back to their true historical root causes through Wikipedia knowledge graphs and PageRank.

princeoseibonsu.github.io/OriginZero



These numbers are not accidents.

60%

More likely for Black Americans
to develop Type 2 Diabetes

3x

Black women die in childbirth
compared to white women

2x

Higher childhood obesity in
low-income communities

Most people say: diet. genetics. personal choice.

OriginZero says:

It started with history — and we can prove it.

Redlining. Slavery. Segregation. Deliberate policy decisions. Every one traceable through Wikipedia.

Why existing tools don't answer this.



ChatGPT / Google

Answers from memory.
No graph. No sources.
No causal chain.
Same answer every time.



Wikipedia Search

Shows related articles.
No ranking.
No origin tracing.
No AI narrative.



OriginZero

Crawls Wikipedia live.
Builds knowledge graph.
Runs PageRank to find
true root causes.
Generates AI narrative.

How OriginZero works.

01

CRAWL

wikipedia-api seeds knowledge tree
60 articles · 474 connections

02

BUILD

NetworkX directed graph
Nodes + edges from real
Wikipedia data

03

RANK

PageRank algorithm scores
every node
Separates root causes from
related topics

04

NARRATE

Groq Llama3 reads ranked chain
Generates plain English origin
story

05

VISUALIZE

D3.js interactive graph
Timeline · Stats · Evolution ·
Live on GitHub

TECH STACK

Python · wikipedia-api · NetworkX · NumPy · Groq Llama3 · D3.js · GitHub Pages · HuggingFace originzero-gemma-4

LIVE DEMO

See OriginZero in Action

princeoseibonsu.github.io/OriginZero

Type 2 Diabetes

Maternal Mortality

Childhood Obesity

"African-American history. Ranked number one. By the algorithm. Not by us."

"Not diet. Not genetics. History."

Backup video ready if needed

What OriginZero Found

60

Articles
Crawled

474

Graph
Edges

1

Day to
Build

#1

Root Cause
Ranked

KEY FINDINGS

- African-American history ranked #1 root cause of Type 2 Diabetes by PageRank — not diet, not genetics
- Maternal mortality traced to 1800s medical experimentation on enslaved women
- Childhood obesity linked to 1930s federal redlining policy — food deserts by design
- All findings sourced, traceable, and linked to real Wikipedia articles
- Fine-tuned model (originzero-gemma-4) deployed on Hugging Face for deeper narratives

*"The origin is not
biological.
It is historical."*

— OriginZero
PageRank Analysis

What we faced — and how we fixed it.

Wikipedia Noise

Problem: 1,200+ links per article — most irrelevant

Fix: Built keyword filter: poverty, racism, redlining, history, food, community

GPT-2 Loops

Problem: Model generated repetitive, low quality narratives

Fix: Switched to Groq Llama3 — faster, smarter, free

Python 3.14 Incompatibility

Problem: bart-large-cnn not supported

Fix: Replaced with Groq API — better results, zero setup

Team Member Unavailable

Problem: Amelia couldn't make it day 1

Fix: Prince stepped in, covered crawler AND PageRank engine

GitHub Pages Not Finding index.html

Problem: File in wrong subfolder

Fix: Reorganized repo structure — live in under 5 minutes

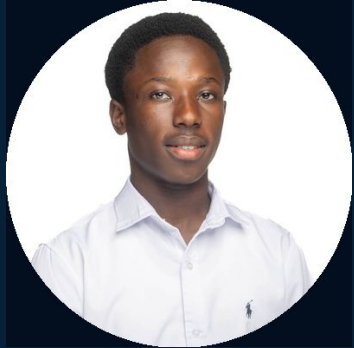
Teammate Pushed Breaking Changes

Problem: Graph showed placeholder node names

Fix: Used git revert — recovered working version instantly

Meet Origin Zero

Each member owns one part of the pipeline — built in parallel, handed off cleanly.



Prince Osei Bonsu

The Brain 🧠

NetworkX graph + PageRank
Ranks root causes
Exports origin chain



Amelia Hunter

The Engine 🔧

Wikipedia crawler
Pulls live articles
Outputs clean JSON



Anthony Nwafor

The Face 🎨

D3.js frontend
Fine-tuned Gemma model
originzero-gemma-4



Emmanuel Opoku

The Voice 🗣️

Groq Llama3 narrative
AI origin story generation
Connects to frontend

The Road Ahead for OriginZero

✓ Gemma Model Connected

originzero-gemma-4 fine-tuned and integrated — live on Hugging Face right now

✓ Full Pipeline Complete

Crawler → Graph → PageRank → AI Narrative → D3.js Frontend — all connected end to end

→ Live Search

Users type any health topic — live Wikipedia origin trace in real time

→ Expand Topics

Mental health, COVID-19, cardiovascular disease disparities

→ Community Tool

Partner with health orgs to deploy as a public education platform

→ Research Paper

Publish PageRank methodology for causal health disparity analysis