AI-Ready

or AI-Hopeful?

Eugene

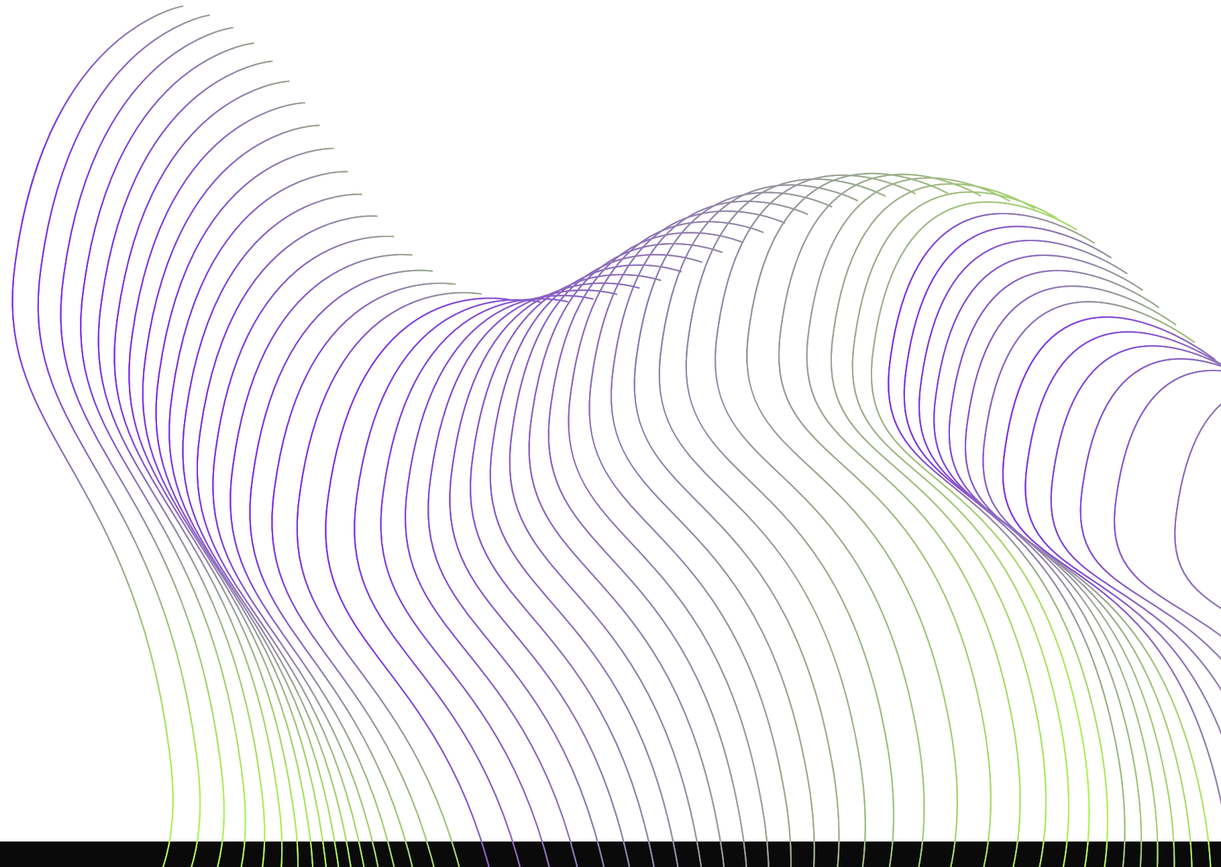
Noel

Sam

GL

Agenda

- Problem Introduction
- Personal Experience
- Case-Studies
- Live Jupyter Notebook Demo
- Live Norma Demo



GL

PROBLEM INTRODUCTION

85%

AI projects fail

to deliver on their promises

*according to Gartner

70%

AI projects fail

because of *data quality* and integration *issues*

*according to McKinsey & Company

5%

Global annual revenue lost

due to underperforming AI programs built on *low-quality data*

*according to Fivetran

GL

PERSONAL EXPERIENCE

Constant Cleaning
Around the clock cleaning on different fronts

Good Technology, Bad Input
The most advances algorithms could not compete

No Predictive Signal
“Gold” data could not produce meaningful reults

occupants	number	origin_id	display_name	document_id	estimated_repair_cost	location_code	name	parent_reference	parent
> Sandhya ...	SPA0025673		06-124	Location: BOW-06-124	$ CAD0.00	BOW-06-124	Location: BOW-06-124	Floor: Floor 06	Floor: Floor 06
	SPA0014637		10-163	Location: BOW-10-163	$ CAD0.00	BOW-10-163	Location: BOW-10-163	Floor: Floor 10	Floor: Floor 10
	SPA0019047		01-109	Location: F45-01-109	$ CAD0.00	F45-01-109	Location: F45-01-109	Floor: Floor 01	Floor: Floor 01
	SPA0020228		01-107	Location: TWS-01-107	$ CAD0.00	TWS-01-107	Location: TWS-01-107	Floor: Floor 01	Floor: Floor 01
	SPA0022161		02-002	Location: BOL-02-002	$ CAD0.00	BOL-02-002	Location: BOL-02-002	Floor: Floor 02	Floor: Floor 02
	SPA0018323		29-004	Location: BPC-29-004	$ CAD0.00	BPC-29-004	Location: BPC-29-004	Floor: Floor 29	Floor: Floor 29
	SPA0017131		02-026	Location: CL7-02-026	$ CAD0.00	CL7-02-026	Location: CL7-02-026	Floor: Floor 02	Floor: Floor 02
	SPA0018857		32-095A	Location: BPC-32-095A	$ CAD0.00	BPC-32-095A	Location: BPC-32-095A	Floor: Floor 32	Floor: Floor 32
Justin Wenger	SPA0018333		29-046B	Location: BPC-29-046B	$ CAD0.00	BPC-29-046B	Location: BPC-29-046B	Floor: Floor 29	Floor: Floor 29
	SPA0025808		02-014I	Location: SR4-02-014I	$ CAD0.00	SR4-02-014I	Location: SR4-02-014I	Floor: Floor 02	Floor: Floor 02
	SPA0026864		01-096	Location: CL1-01-096	$ CAD0.00	CL1-01-096	Location: CL1-01-096	Floor: Floor 01	Floor: Floor 01
	SPA0001013		Rainbow Lake	Campus: Rainbow Lake	$ CAD0.00	Rainbow Lake	Campus: Rainbow Lake	Region: Canada	Region: Canada
	SPA0027255		01-024	Location: C15-01-024	$ CAD0.00	C15-01-024	Location: C15-01-024	Floor: Floor 01	Floor: Floor 01
	SPA0016870		01-138B	Location: REF-01-138B	$ CAD0.00	REF-01-138B	Location: REF-01-138B	Floor: Floor 01	Floor: Floor 01

*workplace-occupancy tracking view

- No PK/FK
- Not Even 1NF
- Data Types All Strings
- Lack of Domain Constraints

GL

CASE STUDIES

It’s not just me

IBM	IBM Watson for Oncology failed because messy, synthetic data led to unsafe, unreliable AI recommendations.
GlobalCPG	A global retailer’s $20M AI project failed because messy, uncleaned data made its forecasts useless.
HCA Healthcare℠	HCA’s sepsis AI only succeeded after overcoming messy, inconsistent hospital data through deep standardization and cleanup.

GL

IBM LOSSES BILLIONS

Even **Watson** Failed Without Clean Data

AI Genius?

- $4B **Loss**
- **Incorrect** Treatment Suggestions
- Trained on Fake Patients?

Big Promise, Bigger Problems

- Marketed to **revolutionize** cancer treatment
- Couldn’t understand **messy**, real-world patient data
- Recommended **unsafe** therapies in some cases
- Abandoned after costing IBM **billions**
- **Ignored** by doctors globally

GL

CLEAN DATA SAVES LIVES

$20M Forecasting AI **Flops** from Messy Data

Big Tech, No Prep = Big Failure

- $20M AI **investment** with minimal data prep
- Built on **fragmented, error-prone** data
- Forecasts **unusable**, project scrapped after 1.5 years

Rapid Approach Gone Wrong

- Built a sophisticated algorithm on top of **poor data**
- No upfront **outlier** handling or data integration.
- Lacked data governance—AI was “learning” from **messy, inconsistent** historical records.

“Many AI projects fail due to wrong data, bad planning, and unrealistic expectations.”

— “10 Reasons” LinkedIn article

GL

CLEAN DATA SAVES LIVES

Clean Data Powers **Life-Saving** Sepsis AI

Big Data, Bigger Impact

- 31 million patient records **standardized**
- SPOT flags sepsis up to 18 hrs **earlier**
- **Saved** ~8,000 lives over 5 years
- Enabled **rapid** emergency response during hurricanes

Numbers Talk

- 708 Lab Names **Standardized**
- 31M Records **Unified**
- 8 000 Lives **Saved**

❝ SPOT monitors all available data every moment… when combinations… consistent with sepsis are detected, the system responds… alerting clinicians so they can quickly intervene. ❞

— Dr. Jonathan Perlin, HCA CMO

GL

Jupyter Notebook Demo

Cleaner Data - Better ML?

https://github.com/epaulia/untappedEnergy

IPYNB DEMO

GL

WISH UPON A STAR

I need a **Hero**...

Data Processing is Necessary, but Time Consuming

Data is the root foundation for how everything in this world runs, with messy data, you get no where.

Too Many Tools, Not Enough Flow

Teams juggle SQL editors, notebooks, and dashboards and nothing feels connected or efficient.

Insights Take Too Long

By the time data is cleaned and visualized, the moment to act has already passed.

GL

THANK YOU!

Get In **Touch**

Website

grouplabs.ca

Thank You

GL

NORMA

Introducing **NORMA**

Time Saved **..90%**

Less Iterations **.... 8x**

Operational Cost **..63%**