Innovative Data Processing: AI/ML and Algorithms Driving Efficiency in Operational and Safety Projects

November 2025

AtkinsRéalis

Why Automation Matters

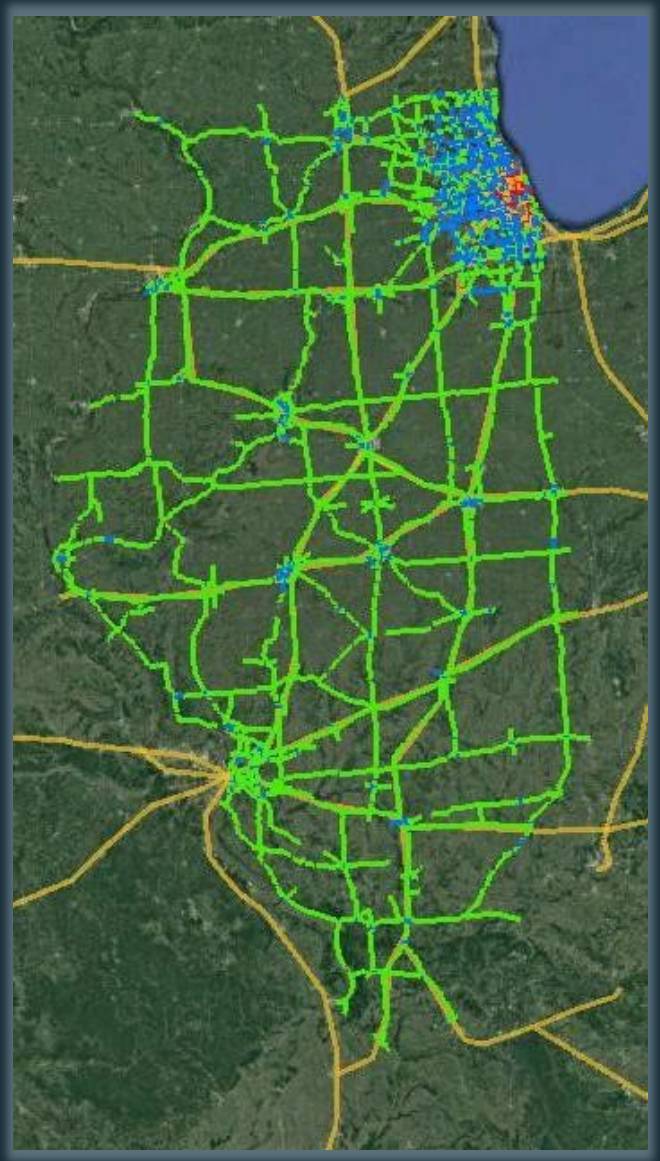
- Surge in data from sensors, geospatial tools, and open datasets is growing fast | 3V’s of big data (Volume, Velocity, Variety)
- Complexity overwhelms manual processing
- Automation unlocks efficiency and insight
- Goal: Transforming raw data into actionable insights for operational and safety projects.

Presentation Roadmap

- Tools and algorithms to process big data
- Five real-world case studies
  - Statewide TT Reliability - HERE and PostgreSQL
  - Machine learning for Crash Reports
  - Python workflows for ARCGIS and roadway data
  - Python workflow for Statewide crash data assembly
  - AI/ML-driven image data collection
- Strategies for implementation

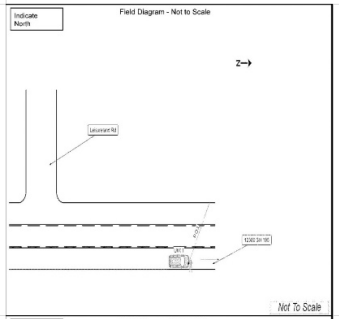
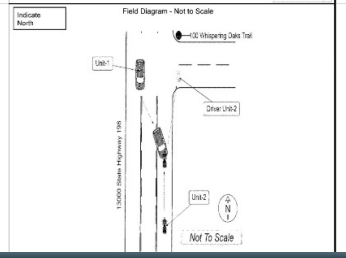
Case Study 1 – Statewide Travel Time Reliability Analysis

- Project: IDOT Statewide TSMO Plan
- Input: 2017-2019 HERE data. 16B records.
- System: Most IDOT and ISTHA roadways plus some local roads (15K+ miles)
- Tools: HERE and HPMS, Postgres GIS for spatial analysis.
- Method: Calculated TT reliability metrics, which in combination with User Delay Costs were used to identify candidate locations for TSMO Improvements
- Outcomes: Identified congestion hotspots.



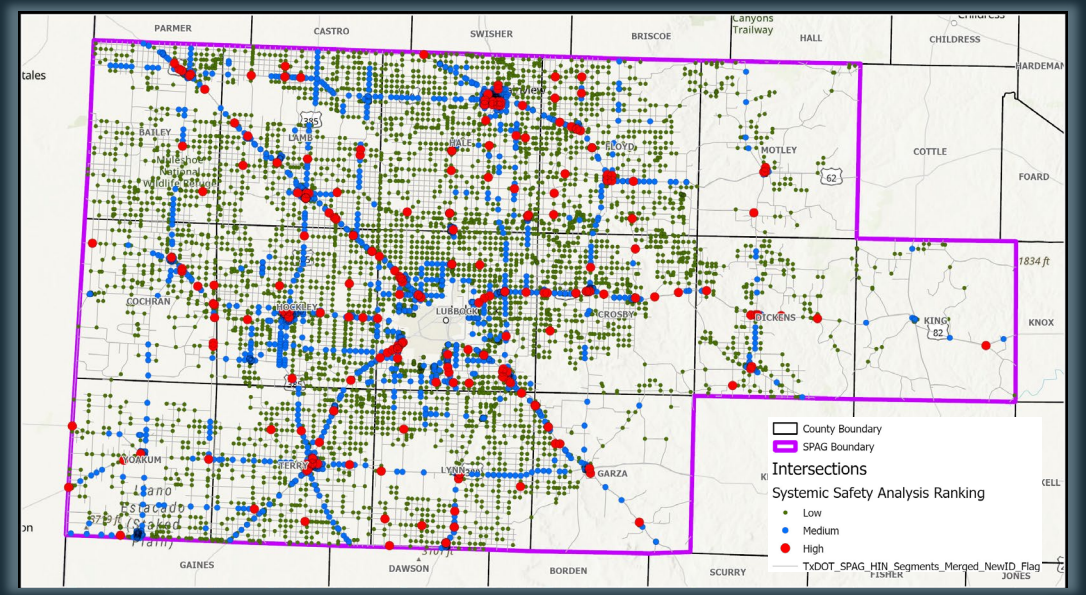
Case Study 2 – Crash Reports Interpretation

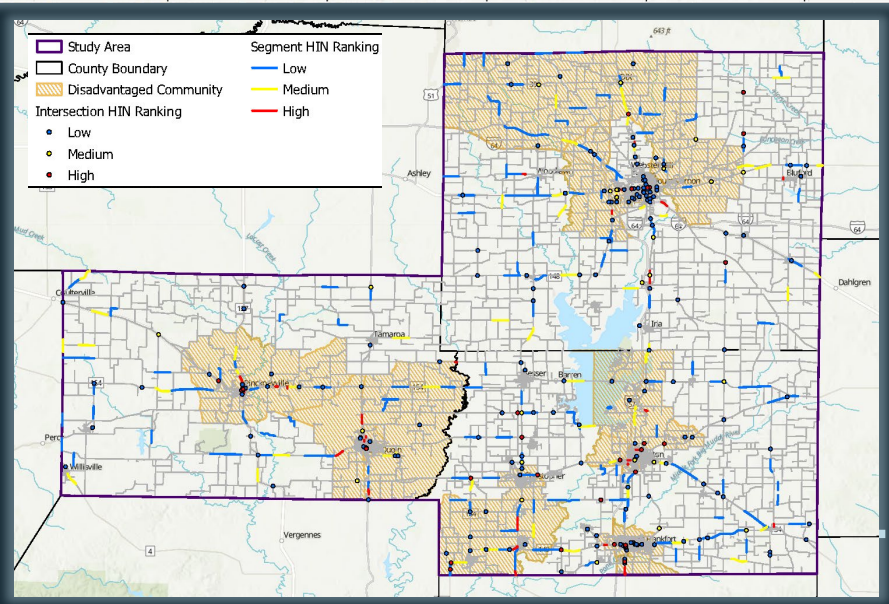
- Project: Various, TxDOT
- Input: Police crash reports inconsistencies, Direction of travel
- Tools: ML/Python workflow identifies Crash ID, interprets police narrative, assign direction of travel.
- Outcomes: Assist data preparation step by addressing inconsistencies.

Crash ID	TXDOT ID_Unit	TXDOT ID	Unit#	Narrative	Direction_4	Direction_8	Crash Diagram
2023598449	19908875.1	19908875	1	Unit 2 was standing on the side of the northbound lane wearing dark clothing. Unit 1 was traveling northbound when he stuck Unit 2 of Health ambulance #957 witnessed the incident and responded and began caregiving. Unit 2 driver and Unit 1 was flown to hospitals in Tyler . Witness Houston 903-821-5521 1 Witness Smith 903-340-7782 . Unit 1 Davis 430-302-7641	NB	NB	
2023569999	19880425.1	19880425	1	Unit-1 was traveling southbound in the 13000 block of Highway 198 Unit-2 was traveling southbound in the 13100 block of Highway 198 Unit-1 began to make a left turn into a private driveway. The driver of Unit-1 advised he did not see Unit-2 coming. Is Unit-1 entered the path Unit-2 . Unit-2 could not avoid the collision Impact resulted in the rider of Unit-2 being ejected from the motorcycle and both front airbags being deployed in Unit-1 The Investigating Officer could not locate liability insurance for either vehicle involved in the collision. The driver of Unit-2 was injured and was transported by Air Ambulance	SB	SB	

Case Study 3 – Geospatial Workflow

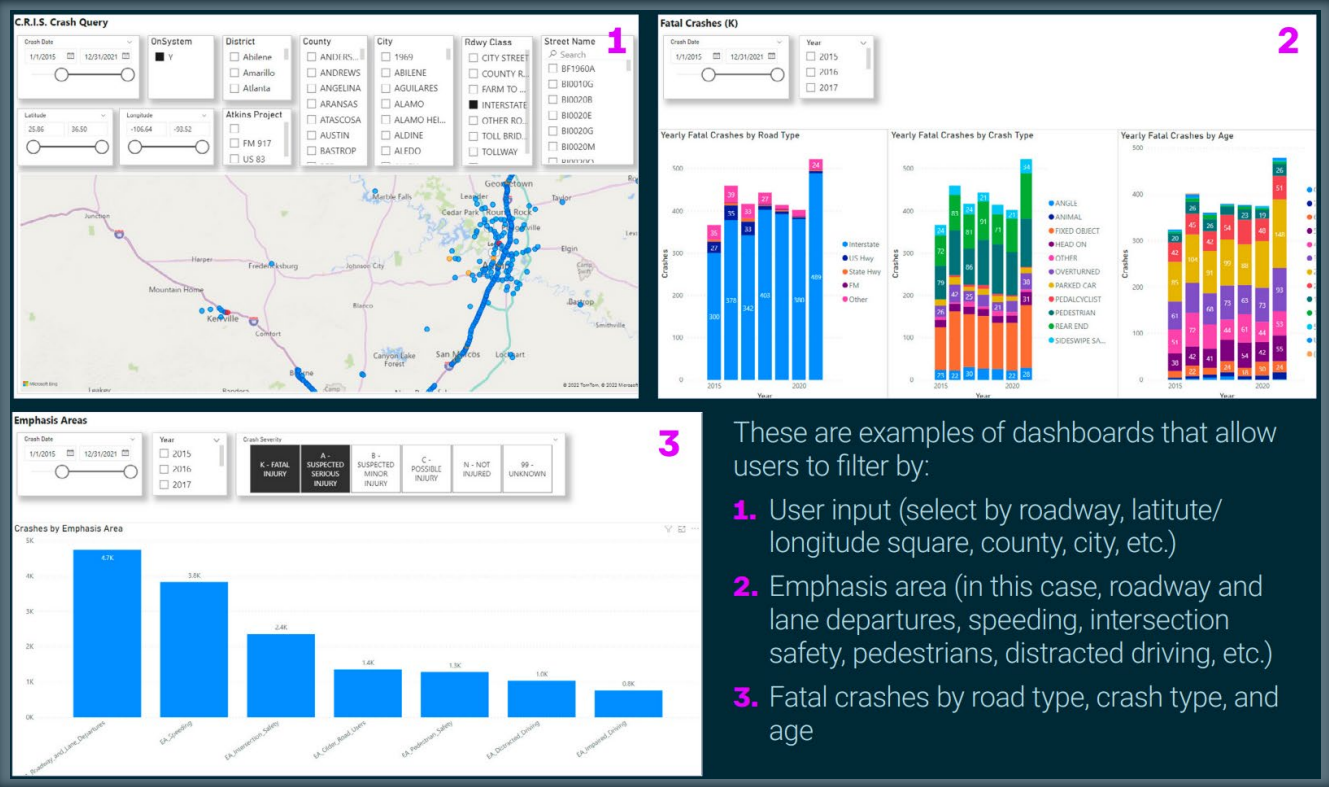
- Project: Various site-specific, corridor, and regional studies
- Input: ARCGIS + OpenStreetMap / NAVTEQ / NMPRDS /HPMS
- Tools: Python scripts streamline aggregation.
- Method:
  - Extracted roadway variables from multiple data source.
  - Created crosswalks between roadway and intersection layers
- Outcomes: Enhanced roadway layers. Developed intersection layers from scratch





Case Study 4 – Statewide Crash Data Process

- Project: Statewide Crash data dashboard (2012-2023)
- Input: Assembled 6 million+ crash, person, unit records
- Method: Automated Python workflow for data integration
- Outcomes: Enabled large-scale analysis for safety and planning



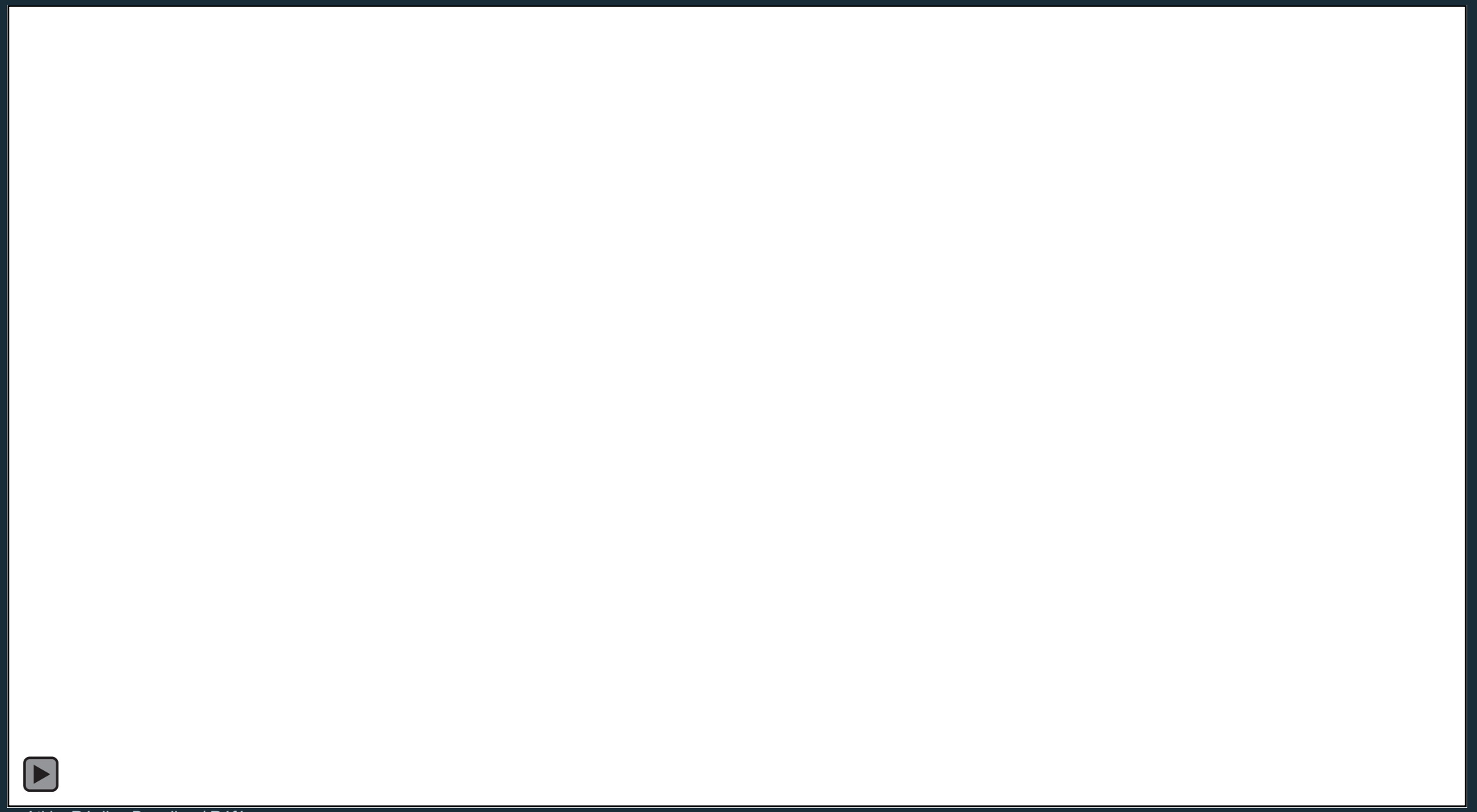
These are examples of dashboards that allow users to filter by:

1. User input (select by roadway, latitute/ longitude square, county, city, etc.)
2. Emphasis area (in this case, roadway and lane departures, speeding, intersection safety, pedestrians, distracted driving, etc.)
3. Fatal crashes by road type, crash type, and age

Case Study 5 – AI– Driven Data Collection

- Project: Automates attribute collection using machine vision and crowd-sourced data.
- Input: Google Street View images (12K intersections, 48K images).
- Tools: Python, Spyder IDE, CUDA for GPU acceleration
- Method: Real-time detection with high accuracy. Processes images in milliseconds.
- Outcomes: Enabled large-scale data collection to enhance analysis for safety and planning

AI Data-Driven Data Collection



Benefits of Automation

- Streamlined workflows through reduced manual effort and minimize human error.
- Enhanced decision-making with faster, data-driven insights.
- Shifted focus to analysis and solution development.

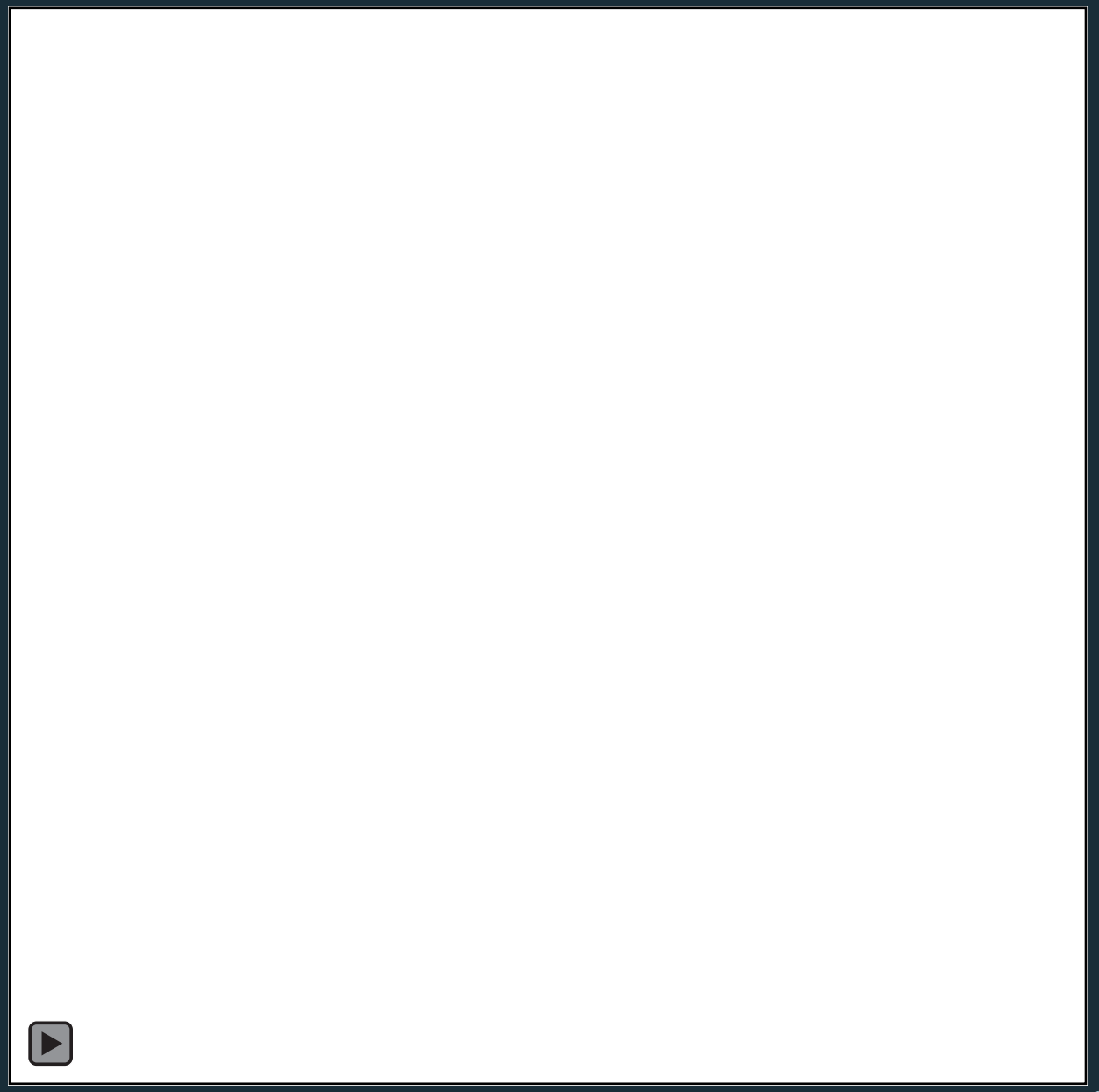
Strategies for Implementation

- Use modular tools (e.g., Python, PostgreSQL), or other open-source technologies for flexibility.
- Integrate AI/ML incrementally based on project needs.
- Prioritize data quality and validation in automated workflows.
- Iterate and improve – Automation is a journey, not a one-time fix

Conclusions

- Recap: Automation and algorithms transform complex data into actionable outcomes.

- Takeaway: Leveraging these tools is key to tackling big data challenges effectively.



Dante Perez-Bravo, PE, PTOE, RSP1

Dante.perez-bravo@atkinsrealis.com

AtkinsRéalis