PROJECT

G.R.I.D

Green, Reliable Infrastructure for Data

"He doesn't get enough credit for all the credit he gave us" - Hamilton The Musical

Meet Our Team

Tanushree Duggirala
University of Texas Austin

Destiny Osemwengie
Macalester College

King Igbozuruike
Grambling State University

Austin Nguyen
University of Houston

Agenda

1. Introduction
2. Short Term
3. Long Term
4. BNY Adoption
5. Conclusion

Our Approach

Short Term

Model and Query Optimization

Long Term

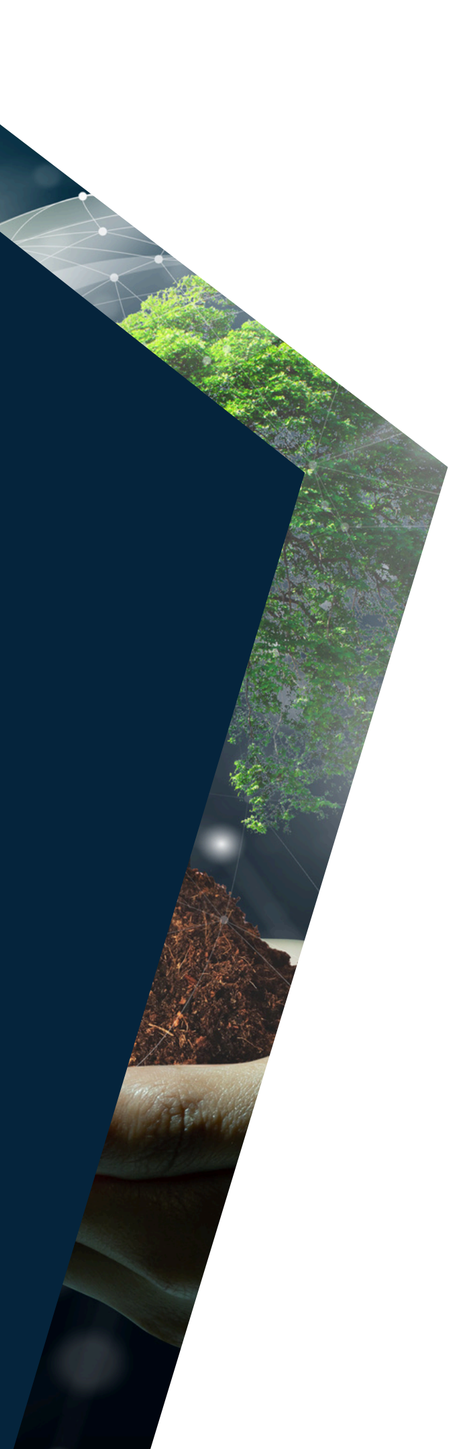
Green Data Warehousing

BNY Strategy

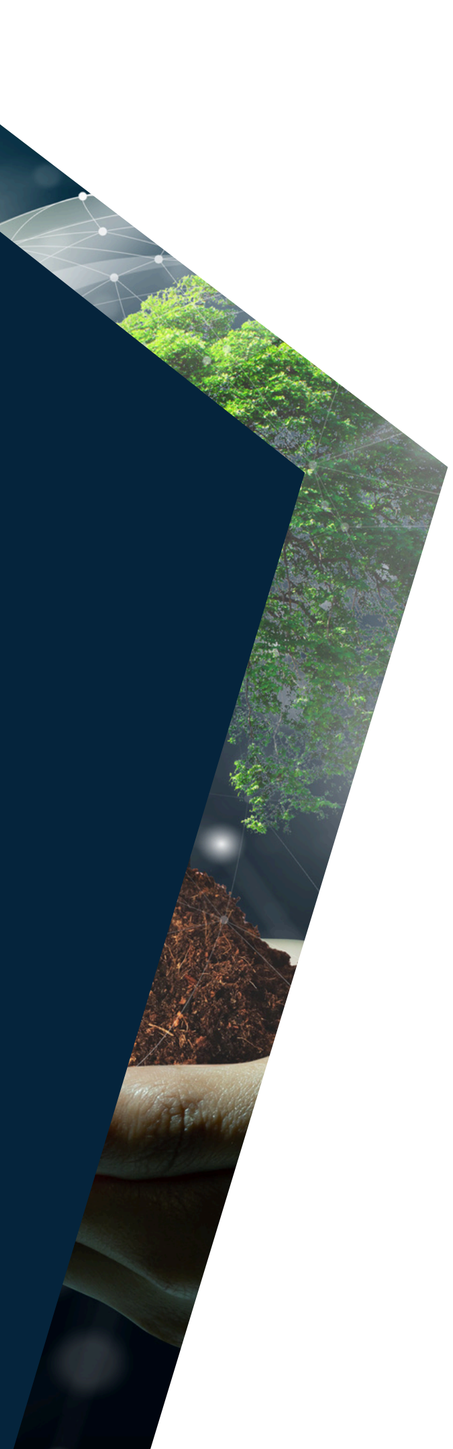
Sustainable Company Outlook

Short Term Solution

2-Pronged Approach: Model Optimization and Query Policing

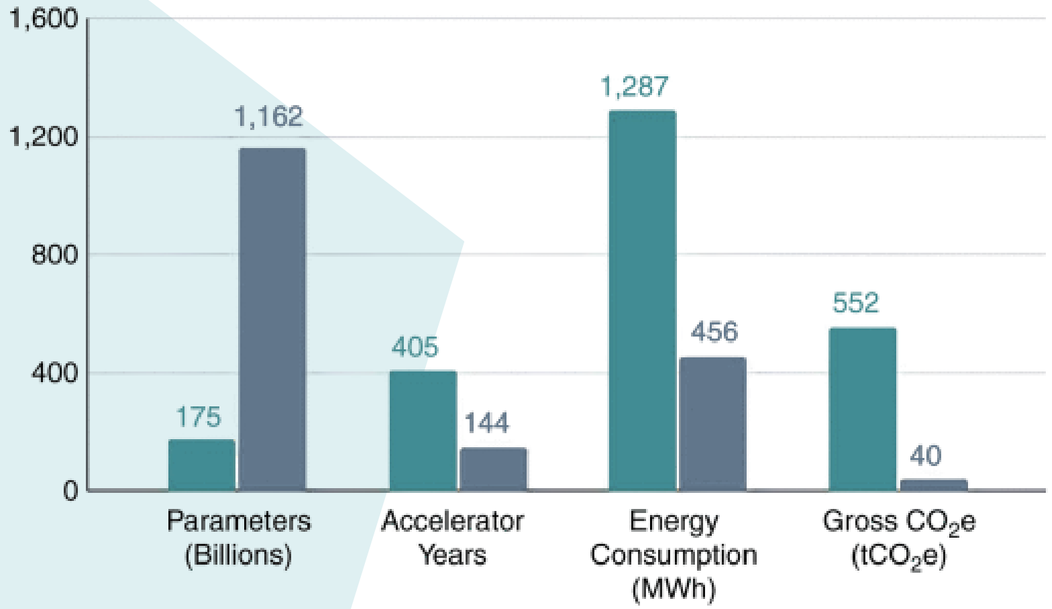


Model Optimization



Lifecycle Management

Minimizing Environmental Cost at Every Stage



The parameters, accelerator years of computation, energy consumption, and gross CO2e for GPT-3 (V100 in 2020, in green) and GLaM (TPU v4 in 2021, in gray). If instead of outperforming GPT-3 on quality scores, GLaM were trained only to match, it would halve the time, energy, and CO2e. Google's renewable energy purchases further reduce the impact to zero.

Modeling

Model: Selecting efficient ML model architectures while advancing ML quality, such as sparse models versus dense models, can reduce computation by factors of 5–10.

In one month, that's ~12,000 water bottles worth of carbon savings

SOURCE 1

SOURCE 2

Query Policing

Simple Questions, Big Energy Costs.

Employee Asks:

What's the largest transaction Client X made in Q4?

Eliza will:

Simple Questions, Big Energy Costs.

Employee Asks:

What's the largest transaction Client X made in Q4?

Multiply that by over 26,000 users, prompting multiple times, every day.

Scaling AI is Expensive for our Planet

High Scale comes with High Costs:

Heavy models consume **high energy** per query

Redundant queries are reprocessed, **not reused**

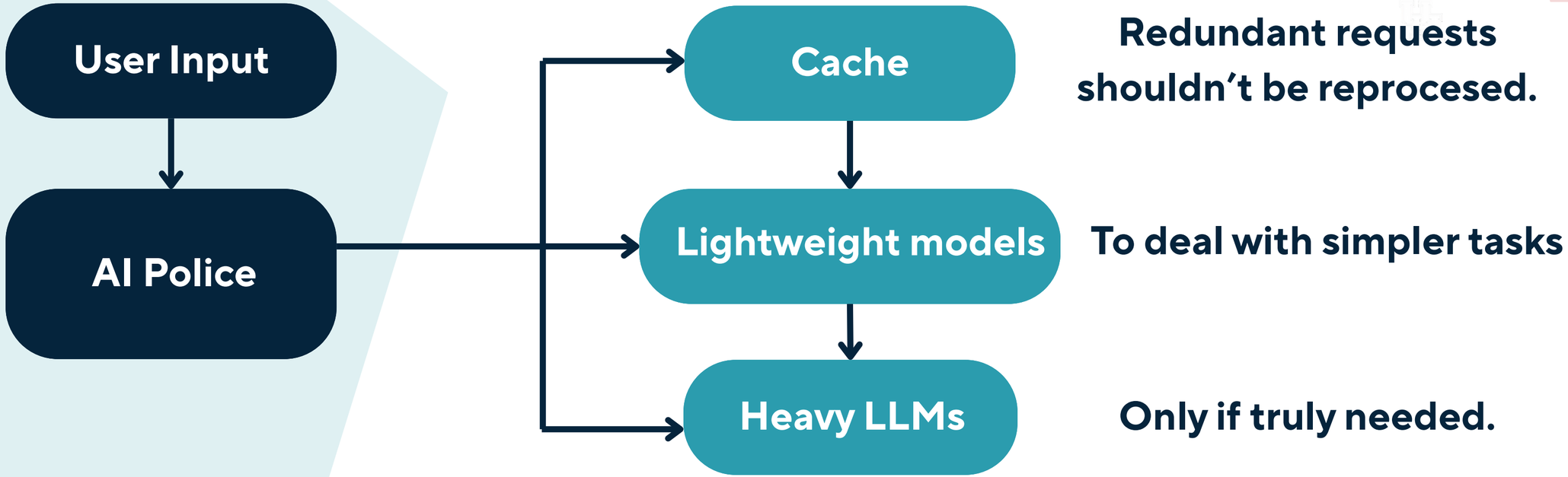
Simple tasks are often executed using **unnecessarily high-compute models.**

47% of BNY Mellon's electricity is **already consumed by data centers**

Growing carbon emissions, avoidable financial costs, and inefficiency that undercuts BNY's carbon-neutral commitments.

AI Police for Eliza: Smarter Routing, Smaller Footprint

Purpose: Reduce energy use, minimize compute waste, and boost sustainability



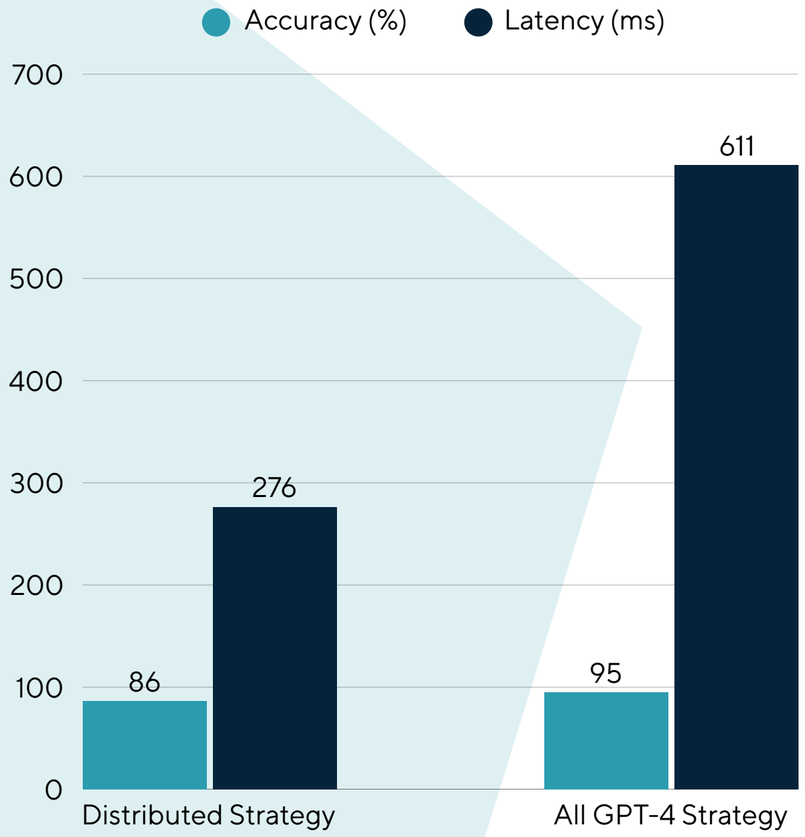
A smart, sustainable layer between Eliza and the compute behind her.

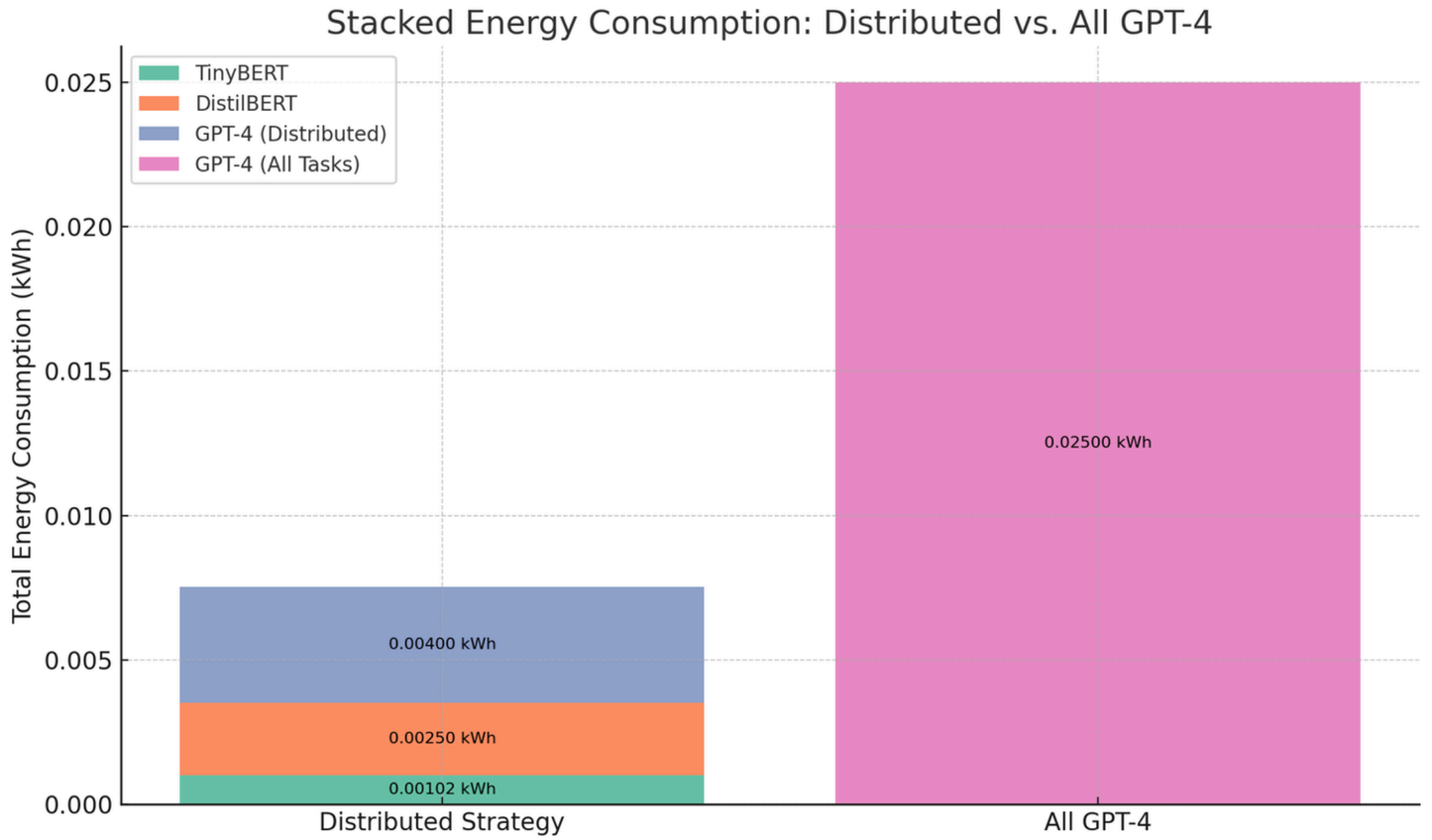
Small Models, Big Impact

TinyBERT	9.4× faster, 96.8% performance
DistilBERT	60% faster, 97% of BERT accuracy
Eliza-tuned for	Classification, Summarization, Query routing

We do not need a semi-truck to grab pizza down the street when a bicycle will do just fine.

Real Savings at Scale

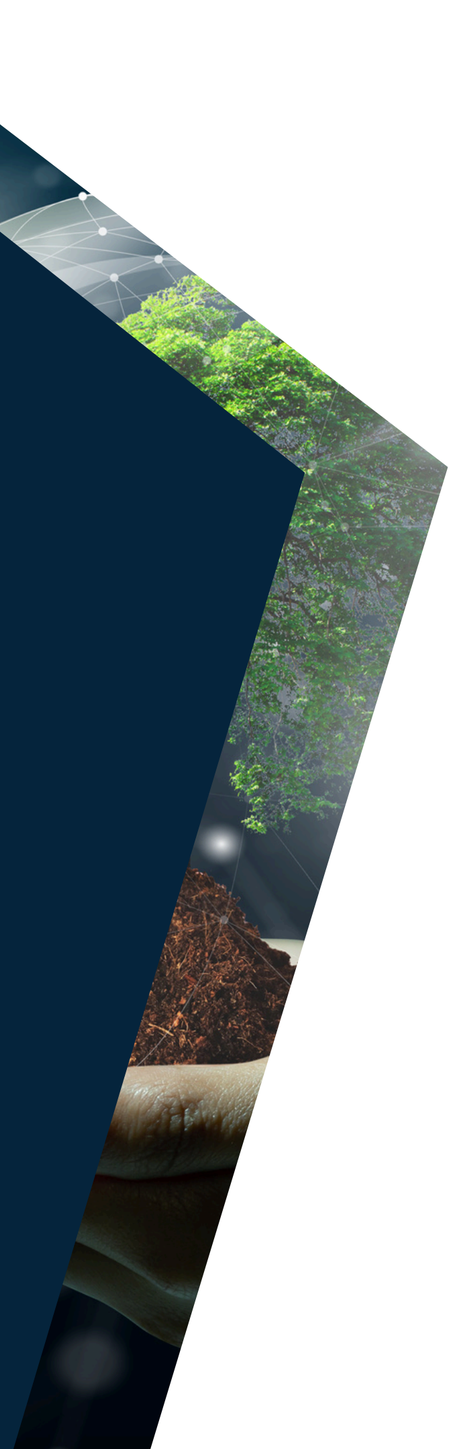




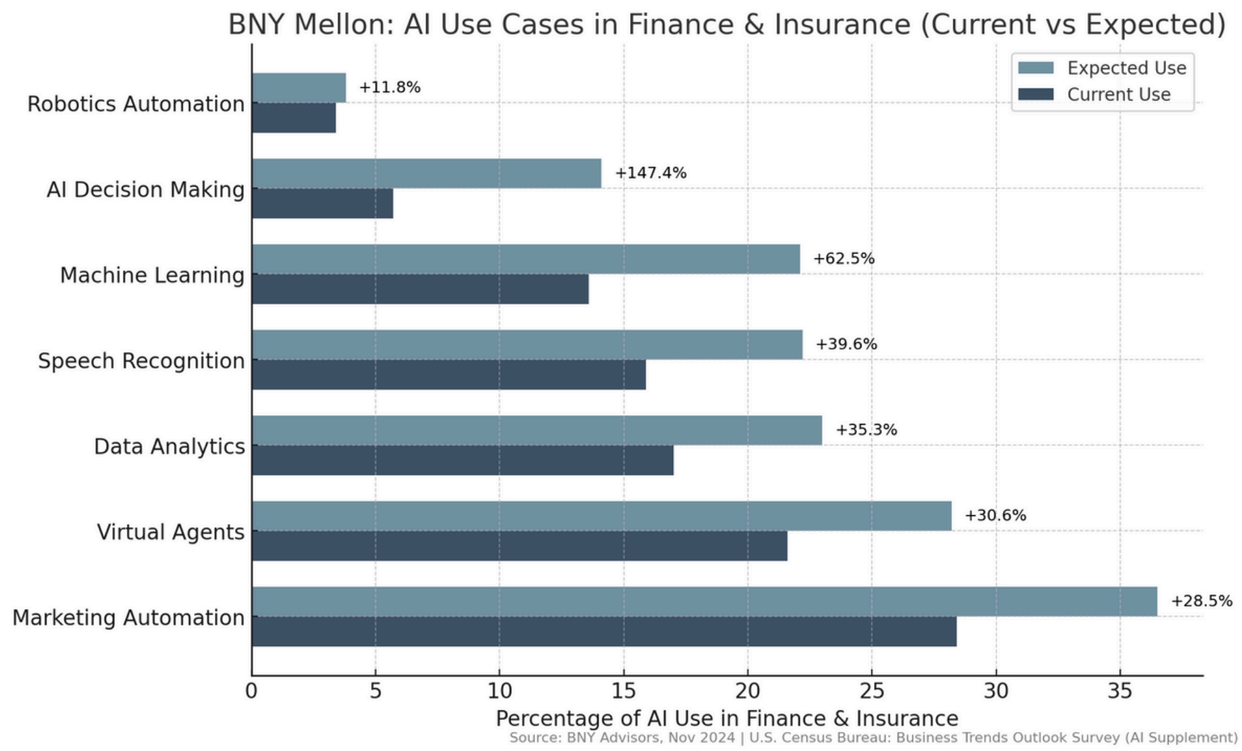
Energy Consumption Reduction: ≈ 69.92%

Long Term Solution

How can we keep up with rising, future demand for energy?



The Case



The Facts:

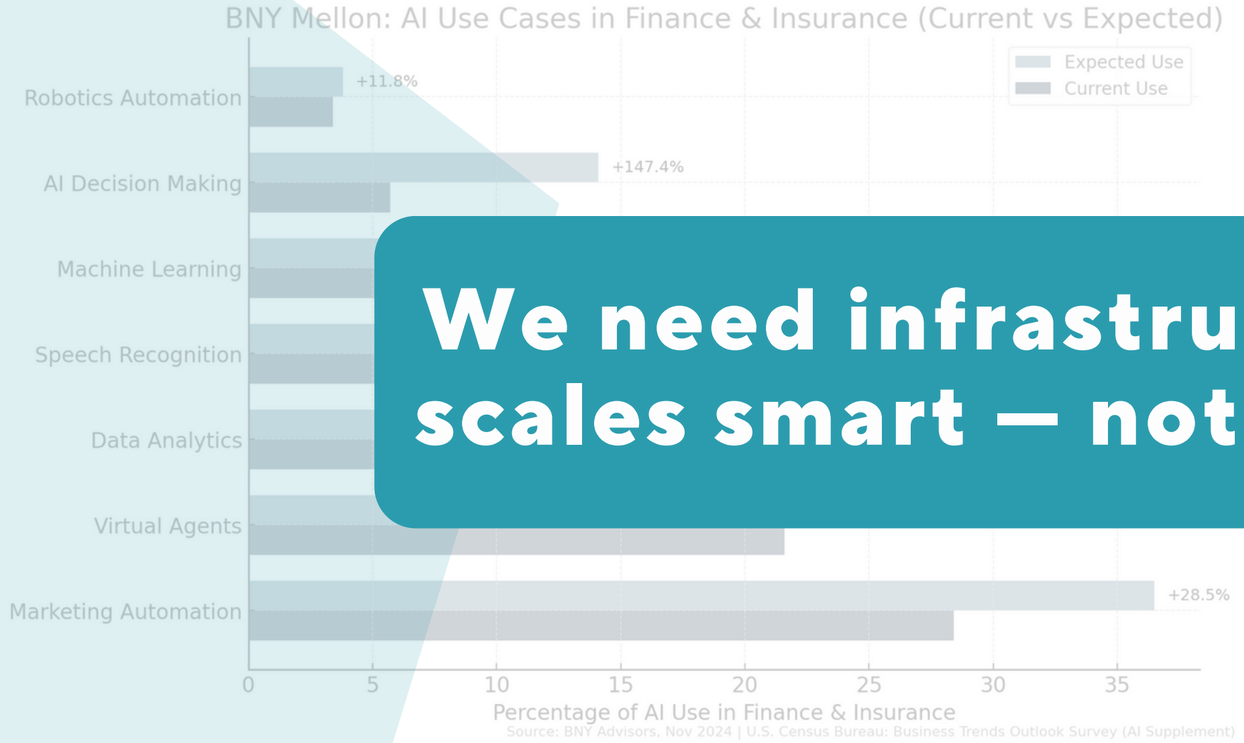
U.S. data: costly and high-emissions

AI growth = rising energy and compute needs

ESG and compliance demands are tightening

Legacy systems lack scalability and sustainability

The Case



We need infrastructure that scales smart — not just wide.

The Facts:

U.S. data: costly and high-
emissions

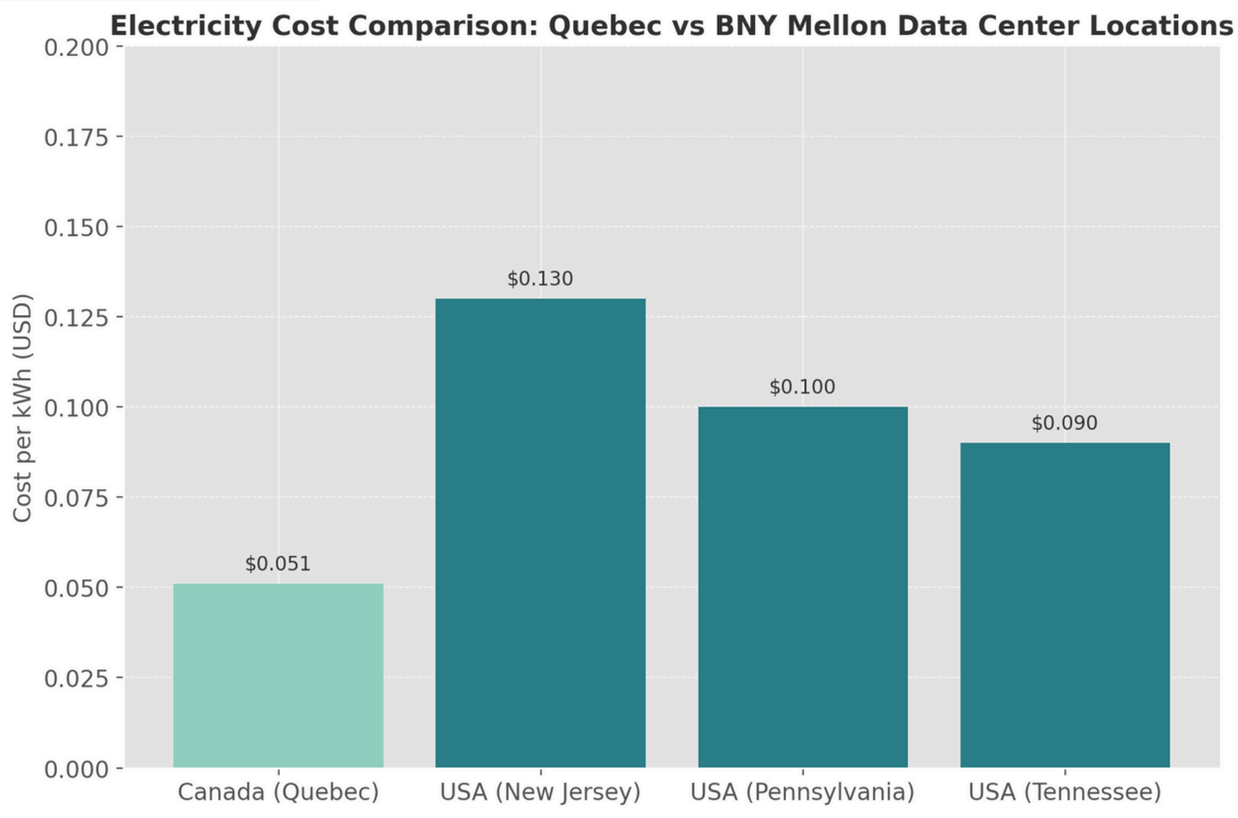
ising energy
ute needs

ompliance
e tightening

Legacy systems lack scalability and sustainability

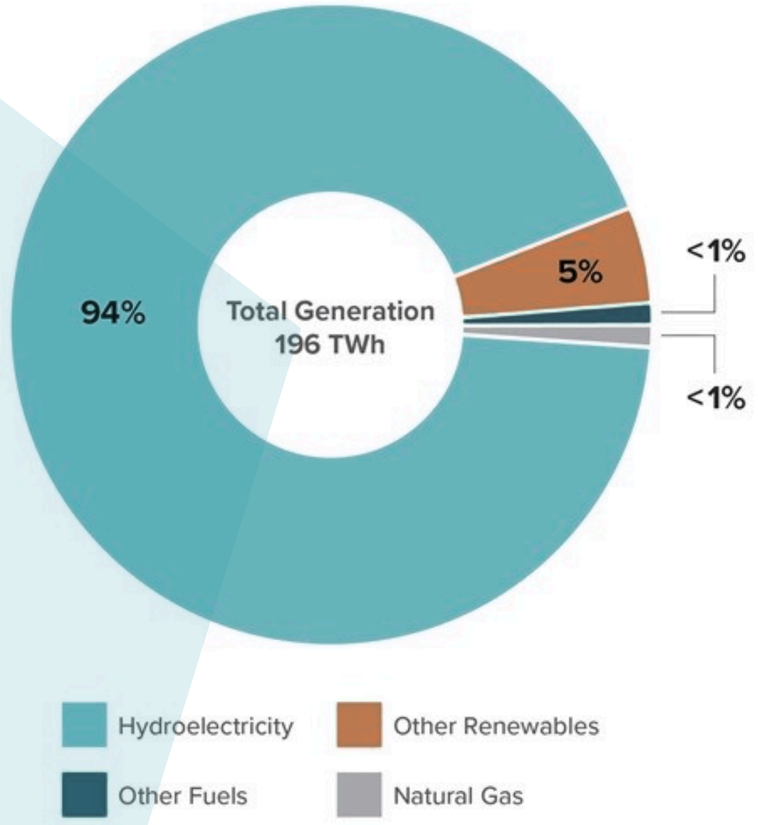
A Sustainable Future for AI at BNY Mellon

How Quebec Supports Our ESG Goals



- Global leader in clean energy infrastructure
- One of the cleanest energy profiles globally
- Highly reliable grid with low disaster risk and data privacy standards
- Up to 50% lower electricity cost supports scalable AI growth

Low Emissions. High Performance.



- 94% hydroelectricity = near-zero emissions.
- 196 TWh of clean power supports large-scale AI compute
- <1% fossil fuels — among the cleanest grids globally
- Stable, secure, and ESG-aligned
- Perfect fit for energy-intensive model training

Redesigning AI for Scale, Cost, and Climate

Ontario Cloud: R&D and model testing

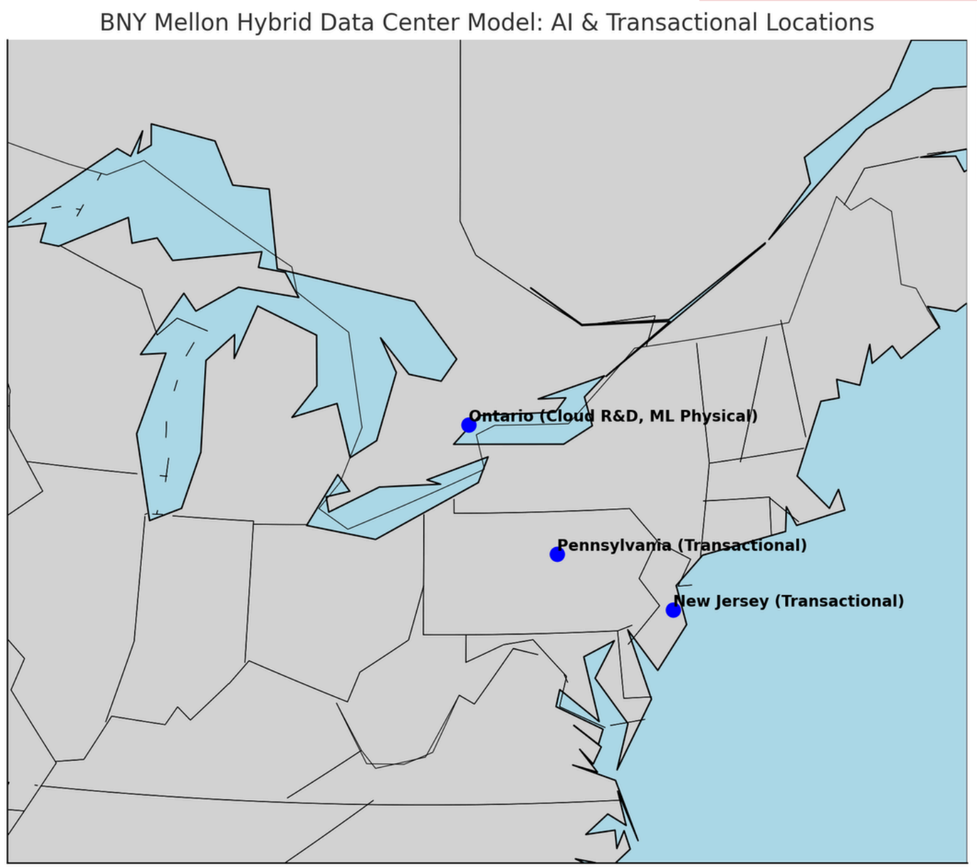
- Fast, scalable compute for AI testing and iteration
- Clean energy grid with strong compliance standards

Quebec Data Center: Large-Scale ML Training

- High-efficiency, low-emission training at scale
- Powered by 99%+ renewable hydroelectricity

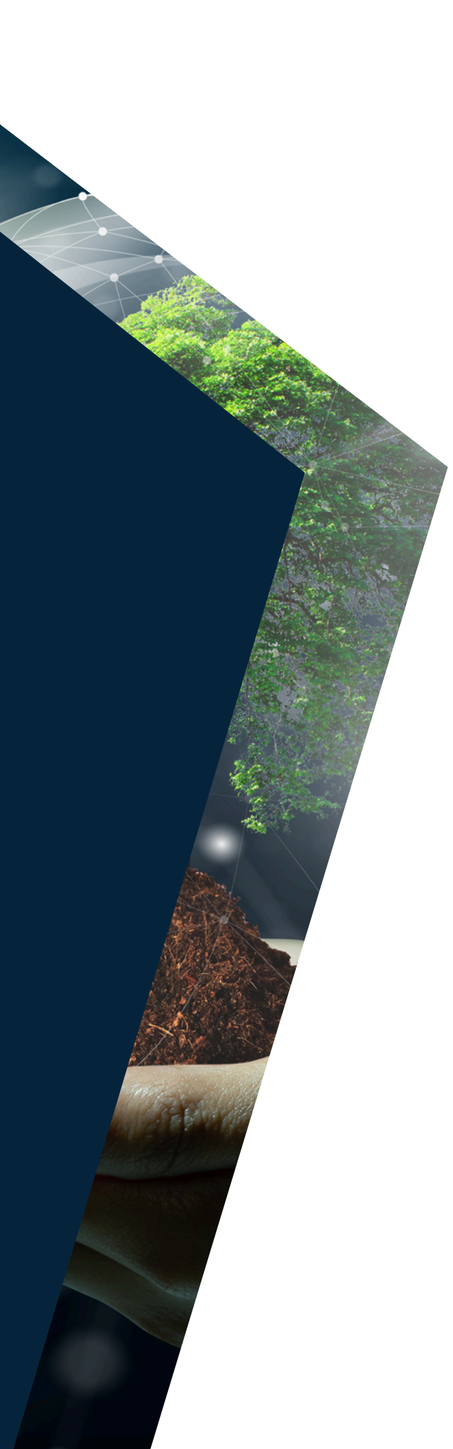
U.S. Sites (NJ, PA): Real-Time Transactions

- Low-latency AI for client transactions
- Meets U.S. regulatory and data localization needs



BNY Integration

How can BNY prove themselves a leader in the green space?



Current Landscape

BNY has a strong foundation in ethical AI use and ESG reporting.

- GHG EMISSIONS
- RE100
- ENERGY EFFICIENCY PROJECTS
- PROPER REPORTING STANDARDS

Current Landscape

BNY has a strong foundation in ethical AI use and ESG reporting.

The Gap: Environmental Impact of AI systems

EFFICIENCY PROJECTS

NG STANDARDS

Integrating Sustainability into AI Ethics

Position BNY Mellon as a leader in sustainable financial technology

Sustainable AI Development Policy

Introduce a formal policy framework that embeds environmental criteria into the AI development lifecycle:

Environmental Impact Scorecard

Develop an internal scorecard system to quantify the sustainability of AI models:

Policy Driven Procurement & Collab

Regulate collaboration with third-party or outsourced AI models (e.g., vendor LLMs)

Integrating Sustainability into AI Ethics

Position BNY Mellon as a leader in sustainable financial technology

Sustainable AI Development Policy

Framework:

Deploy **Emissions Thresholds**
Reported **Energy Budgeting**
Approve via **Greenlight Reviews**

Environmental Impact Scorecard

Develop an internal scorecard system to quantify the sustainability of AI models:

Policy Driven Procurement & Collab

Regulate collaboration with third-party or outsourced AI models (e.g., vendor LLMs)

Integrating Sustainability into AI Ethics

Position BNY Mellon as a leader in sustainable financial technology

Sustainable AI Development Policy

Framework:

Deploy **Emissions Thresholds**
Reported **Energy Budgeting**
Approve via **Greenlight Reviews**

Environmental Impact Scorecard

Metrics:

- **Model Energy Use**
- **Carbon Footprint**
- **Deployment Efficiency**
- **Data Center Carbon Intensity Score Rating**

Policy Driven Procurement & Collab

Regulate collaboration with third-party or outsourced AI models (e.g., vendor LLMs)

Integrating Sustainability into AI Ethics

Position BNY Mellon as a leader in sustainable financial technology

Sustainable AI Development Policy

Framework:

Deploy **Emissions Thresholds**
Reported **Energy Budgeting**
Approve via **Greenlight Reviews**

Environmental Impact Scorecard

Metrics:

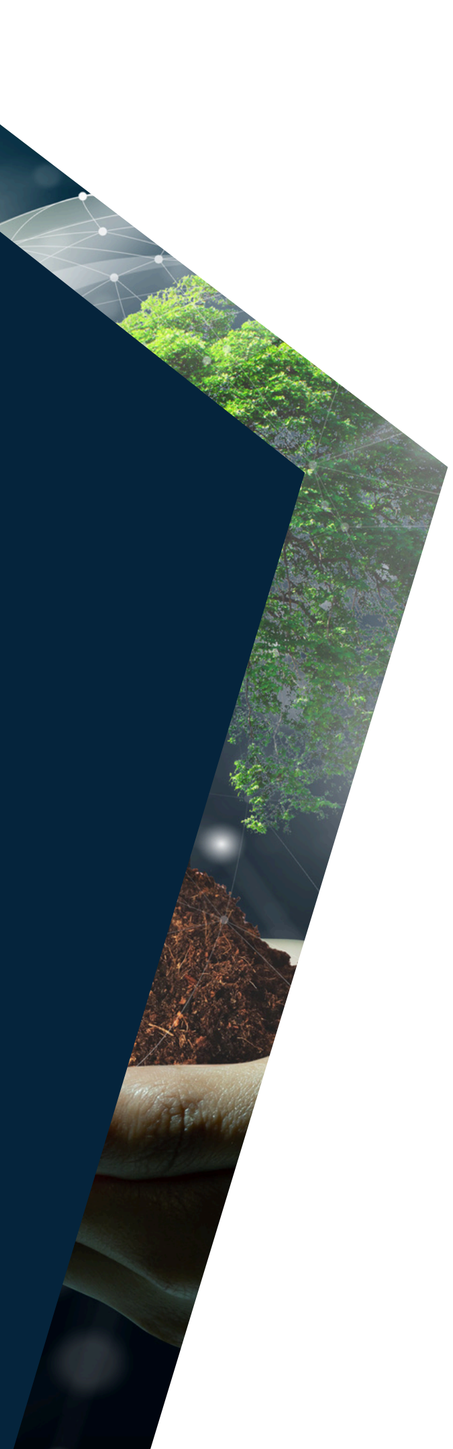
Model Energy Use
Carbon Footprint
Deployment Efficiency
Data Center Carbon Intensity
Score Rating

Policy Driven Procurement & Collab

Guidelines:

Require **environmental metric disclosures**
Partner with labs and start-ups focused on **efficient or low-power AI**

Conclusion



Overview

Case Q

"How can BNY Mellon mitigate the environmental costs of AI/ML while maintaining innovation and trust?"

Goals

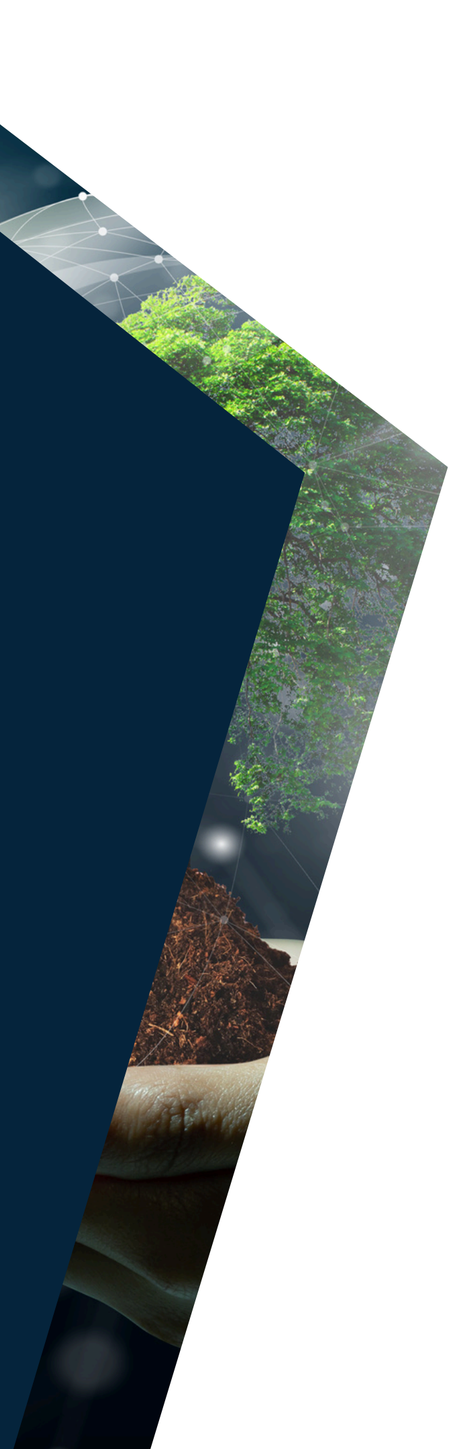
- Reduce AI-related Energy Consumption
- Minimize Carbon Emissions from AI
- Embed Environmental Accountability into AI Governance

Solution

Impact

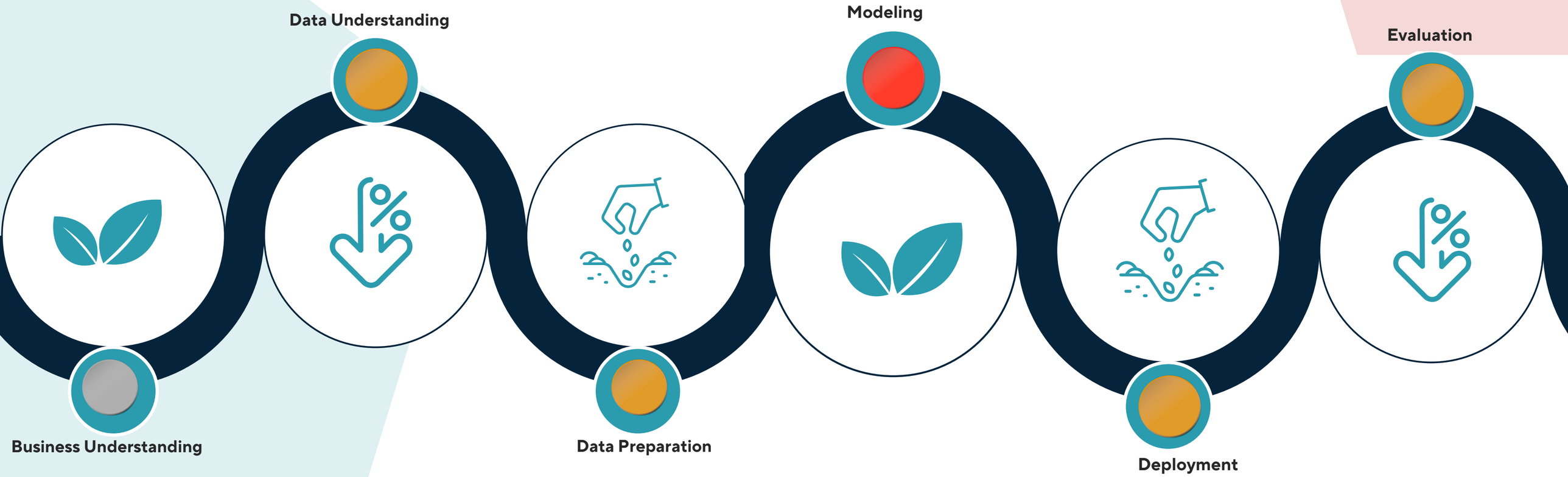
- Aligns AI initiatives with ESG targets (e.g., TCFD, SASB)
- Builds stakeholder trust & brand leadership in sustainable finance
- Enables scalable carbon tracking for AI systems

Appendix



Lifecycle Management

Minimizing Environmental Cost at Every Stage



Model Lifecycle Breakdown

Modeling

Use model complexity thresholds: favor lighter over heavier models unless necessary

Quantize or prune models to reduce inference cost.

Deployment

Set up auto-scaling rules and idle resource shutdowns. Batch predictions where latency isn't critical. Cache repeated inference results

Evaluation

Test models for energy usage per inference/training epoch using tools like CodeCarbon or MLCO2.

Business Understanding

Define the environmental goals alongside the business goals.

Data Understanding

Use data profiling to identify and remove redundant or unused features early on.

Data Preparation

Use dimensionality reduction techniques (PCA, autoencoders) to reduce input space.

Apply feature selection algorithms (e.g., Lasso, Recursive Feature Elimination).

Normalize data formats to allow for efficient loading and computation.

Use efficient storage types (Parquet over CSV for tabular data)

Seen This Before? Don’t Rerun It.

Embeds every query

Searches semantic vector DB (e.g. FAISS)

Matches similar questions like:
“Summarize Q4 trades”
“What did Client X do last quarter?”

*Up to 68% fewer model calls with semantic caching.

SOURCE

Powering AI with the Cloud

- Scalable compute (TPUs, GPUs) on demand
- No more idle servers - pay only for what you use
- Faster deployment, seamless scaling
- Access to the latest hardware (TPU v5e, H100s)
- Built for AI-native partners like OpenAI & NVIDIA