Governance, Security, and Data Ethics

Project: Real-Time Weather Analytics Pipeline
Platform: Google Cloud Platform

Data Sources
All data used in this project comes from publicly available sources.
- Historical weather data from a Kaggle dataset
- Live weather data from the Open-Meteo public API

No proprietary, private, or user-generated data is collected or stored.

Data Privacy
This project does not process personally identifiable information (PII).
Weather data includes only environmental measurements such as temperature, humidity, wind speed, and timestamps.
As a result, there are no privacy risks related to individual users or sensitive attributes.

Security Controls
The pipeline follows standard cloud security best practices:
- No API keys or secrets are hard-coded in source code
- All services authenticate using Google-managed service accounts
- IAM roles are scoped using the principle of least privilege
- Access is limited to required services such as Pub/Sub, Dataflow, BigQuery, and Cloud Functions

---

Access Management
Project access is controlled through Google Cloud IAM.
- Team members are granted Editor-level access for collaboration
- Service accounts are granted only the permissions required to run the pipeline
- No public write access is enabled on any cloud resources

Infrastructure Governance
The following controls are in place to ensure safe and reliable operation:

- Enabled APIs are limited to those required for the pipeline
- Streaming jobs are monitored using Cloud Monitoring
- Dataflow jobs can be stopped or restarted to control cost and execution
- BigQuery datasets are append-only for streaming tables

Operational Risk & Failure Handling
Potential risks include API outages, network interruptions, or streaming job failures.
Mitigation strategies include:
- Monitoring Dataflow job status and throughput
- Restarting streaming jobs when failures occur
- Validating data freshness through timestamp checks in BigQuery

---

Ethical Considerations
The project is designed for educational and analytical purposes only.
- No automated decisions affecting individuals are made
- No forecasting outputs are used for safety-critical decisions
- Model predictions are interpreted as informational, not authoritative

---

Compliance Summary
This project complies with data governance best practices by:
- Using only public, non-sensitive data
- Restricting access via IAM
- Avoiding unnecessary data retention
- Documenting assumptions, risks, and limitations

Conclusion
The system is governed using secure cloud practices, respects data privacy principles, and operates within ethical and educational boundaries. All data usage and access patterns are transparent, auditable, and appropriate for an academic analytics project.