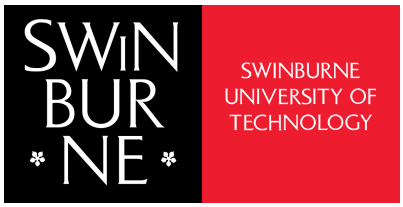

Swinburne University of Technology

Intro to AI – COS30019

Assignment 2C

Ethical Consideration When Designing AI Systems

Marco Giacoppo (104071453)

Tuesday 8:30

2025 Semester 1

Responsible AI in Transportation: Ethical Design Reflections on TBRGS and Its Future Extension TBRGS+

Marco Giacoppo
Department of Computer Science
Swinburne University of Technology
Melbourne, Australia
104071453@student.swin.edu.au

***Abstract*—The integration of Artificial Intelligence (AI) into urban transport systems through the Traffic-Based Route Guidance System (TBRGS) presents a range of ethical considerations. These include ensuring transparency in AI-driven decision-making, mitigating bias in route selection, protecting user privacy, and maintaining accountability across the system's predictive and search components. This report reflects on key technical and design decisions made in the development of TBRGS and its proposed extension, TBRGS+, which introduces real-time GPS tracking and personalized route recommendations. It highlights how each of these features may pose ethical challenges—particularly regarding surveillance, consent, and fairness.**

Through a responsible AI lens, the report analyzes the implications of using algorithms such as A*, UCS, and GBFS, and machine learning models like LSTM and GRU. Design alternatives for core components—such as tracking mechanisms and data storage methods—are compared based on their ethical trade-offs. The report concludes with a set of practical recommendations: prioritizing user consent and data minimization, ensuring transparency in route explanations, and establishing robust auditing and safety mechanisms. These measures aim to align both current and future system designs with ethical best practices for public-facing AI applications.

***Index Terms*—Artificial Intelligence, Ethical AI, Responsible AI, Traffic Prediction, Route Optimization, LSTM, GRU, A* Search, Transparency, Privacy.**

I. Introduction

The increasing integration of Artificial Intelligence (AI) into urban infrastructure has led to the development of intelligent transport systems such as the Traffic-Based Route Guidance System (TBRGS). These systems leverage historical and real-time data to predict traffic conditions and optimize route recommendations for users. In this project, TBRGS was implemented using machine learning models (GRU, LSTM, TCN), with GRU showing the closest match to real traffic trends in predictions, coupled with a selectable range of search algorithms—A*, UCS, GBFS, BFS, Bidirectional, and DFS—allowing users to compare outcomes interactively.

While the technical performance of TBRGS has shown promise, it also raises significant ethical concerns that must be addressed before real-world deployment. These concerns include the transparency of AI decision-making, fairness in route distribution, data privacy, and the accountability of predictions made by black-box models. Furthermore, a proposed extension of the system—TBRGS+—would introduce a mobile interface with GPS tracking and personalized routing features, which adds additional layers of ethical complexity.

This paper reflects on these ethical challenges by examining key design decisions made during development, analyzing their implications, and comparing alternatives through the lens of responsible AI principles. It provides recommendations for ethical design practices to ensure the system aligns with privacy, fairness, and accountability standards in public-facing AI applications.

II. Responsible AI Reflection

The development of the Traffic-Based Route Guidance System (TBRGS) incorporated several principles aligned with Responsible AI frameworks [1], [2]. At its core, TBRGS utilizes machine learning models to predict traffic volumes across SCATS sites and applies search algorithms to recommend optimal routes. While technically robust, responsible AI practices were considered in areas such as transparency, fairness, safety, and privacy.

A. Transparency and Explainability

The system was designed to maintain transparency in its decision-making pipeline. All model predictions and routing decisions are backed by interpretable outputs, such as estimated travel times, identified road segments, and clearly enumerated SCATS site transitions. By providing a step-by-step breakdown of how routes were derived, users can understand the logic behind system recommendations [3].

B. Fairness and Bias Mitigation

To ensure fair routing across all regions, we trained the predictive models (LSTM, GRU, TCN) using traffic data from a wide range of SCATS locations. The encoding and preprocessing steps were generalized to avoid bias toward any specific site or area. However, we acknowledge that fairness in AI-driven transportation systems also depends on equitable infrastructure and data coverage, which is influenced by external factors beyond the scope of the project [4].

C. Privacy and Data Protection

TBRGS operates entirely on anonymized traffic flow data provided by VicRoads. No personal information or vehicle tracking data was used or stored. The model respects data minimization principles by processing only relevant SCATS site IDs, timestamps, and aggregated volume counts. This reduces the risk of individual privacy violations while still enabling accurate travel time estimation.

D. Accountability and Safety

System outputs are derived from reproducible models with documented behavior and metrics. For instance, the route planner provides consistent results when given the same inputs, and all trained models are saved in standardized formats (.keras) alongside their scalers and encoders. To prevent misrouting or misleading estimates, the system enforces geographical validation using Haversine distance and verifies metadata integrity before making predictions.

III. Project Strengths and Limitations

The Traffic-Based Route Guidance System (TBRGS) demonstrates multiple strengths in its design and execution. However, several limitations were also encountered during the implementation and evaluation phases. This section critically reflects on both aspects.

A. Strengths

1) Modular Architecture: The system adopts a modular architecture, separating data preprocessing, model training, prediction, search algorithms, and visualization components. This design enhances maintainability, testing, and scalability. Each module can be updated or extended independently—for example, by swapping out the LSTM model for a GRU or TCN model without modifying the route planner logic.

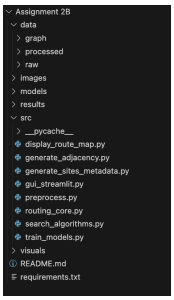

Fig. 1. Files and Folder Structured Nicely

2) Real-Time Route Estimation: TBRGS enables real-time estimation of travel time by dynamically predicting traffic volume using trained ML models. These predictions are used to adjust route costs in search algorithms, making the route recommendations context-aware and sensitive to temporal patterns.

3) Search Method Flexibility: The platform supports multiple search strategies where users can choose between A*, Bidirectional, UCS, GBFS, BFS, and DFS through the interface, enabling comparative performance testing and algorithm transparency. This allows users understand the performance and behavior of each algorithm in a transport context.

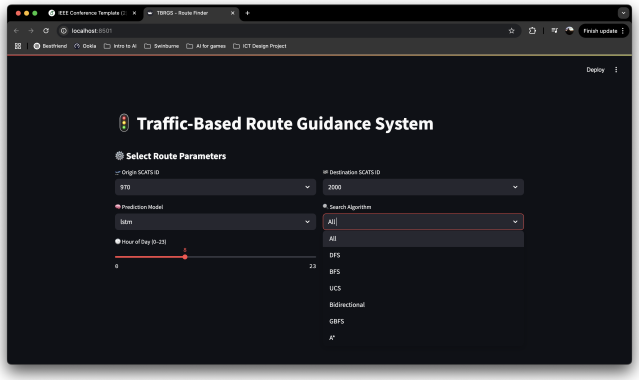

Fig. 2. Different Types of Algorithms Users Can Choose From

4) Geospatial Accuracy: To improve accuracy, the system uses the Haversine formula to compute inter-site distances, ensuring reliable spatial measurements. Combined with dynamic cost functions, this makes the estimated route distance and time more realistic.

```python
# === Haversine util ===
def haversine(lat1, lon1, lat2, lon2):
    R = 6371
    phi1, phi2 = radians(lat1), radians(lat2)
    dphi = radians(lat2 - lat1)
    dlambda = radians(lon2 - lon1)
    a = sin(dphi / 2)**2 + cos(phi1) * cos(phi2) * sin(dlambda / 2)**2
    return R * 2 * atan2(sqrt(a), sqrt(1 - a))
```

Fig. 3. Haversine Formula To Compute Distance Between 2 Points

5) Visual Route Representation: A key strength lies in the integration of a visualization tool using Folium. Folium maps were enhanced with AntPath animations and marker tooltips to visualize SCATS site transitions and routes clearly.

B. Limitations

1) Limited Data Granularity: The ML models rely on aggregate hourly traffic volume data from October 2006, which may not capture short-term fluctuations, special events, or real-time anomalies. This affects the precision of the travel time estimates.

2) Lack of Road Geometry or Lane Data: The system models the transport network as a graph of SCATS sites without including road curvature, intersection topology, or traffic light delay modeling. This can result in an oversimplified route that misses practical road constraints.

3) Inference Time Bottlenecks: While inference is fast for small queries, traversing large sections of the network using A* or UCS can cause delays due to repeated travel time predictions and distance computations. Caching was implemented to reduce this, but a more optimized graph representation (e.g., using NetworkX or precomputed edge weights) may improve performance.

4) Overgeneralized Cost Function: The cost function treats all traffic volumes and site types uniformly, assuming a parabolic relationship between flow and speed. This does not differentiate between arterial roads, minor roads, or highways—affecting the realism of estimated travel times.

5) Missing Segment Reporting: The printed route occasionally skips intermediary SCATS sites that were part of the actual computed path. While the map visualization correctly includes all segments, the console output may underreport nodes, suggesting the need for improved tracking during path reconstruction.

6) Non-Optimal Behavior in DFS: In several cases, DFS failed to traverse paths as expected, prematurely exiting branches or favoring suboptimal routes. This made it unsuitable for reliable routing despite being included for comparison.

Despite these limitations, the system offers a robust and extensible foundation for integrating AI with traffic-aware route planning.

IV. Evaluation and Lessons Learned

The development of the Traffic-Based Route Guidance System (TBRGS) provided a rich learning opportunity across data processing, machine learning, algorithmic search, and geospatial reasoning. This section discusses the system's evaluation metrics, observed outcomes, and insights gained during implementation.

A. Model Evaluation

Three machine learning models were developed to predict traffic volume per SCATS site: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN). These models were trained using hourly data from October 2006 and evaluated using common metrics:

- **Mean Absolute Error (MAE)**: Measures the average magnitude of error in predictions.
- **Root Mean Squared Error (RMSE)**: Penalizes larger errors more heavily, highlighting outliers.
- **R-squared (R^2)**: Indicates how well the model explains variance in the dataset.

model	MAE	MSE	RMSE	MAPE	R2
LSTM	12.1654	312.0532	17.665	480805.7901	0.9402
GRU	12.314	331.7622	18.2143	613251.0173	0.9364
TCN	12.7939	341.3596	18.4759	1294626.6081	0.9346

Fig. 4. Different Metrics Evaluation for each models

Among the models, while both GRU and LSTM performed well, GRU slightly outperformed LSTM in aligning predictions with actual traffic data in several key sites.

The volume predictions were then converted into estimated travel times using a parabolic formula derived from prior literature. These estimates were successfully integrated into the A* and UCS search algorithms to support realistic cost modeling.

B. Search Algorithm Performance

Different algorithms demonstrated different strengths:

- **A* Search**: Provided the most efficient paths by balancing travel time and distance using heuristics.
- **Bidirectional Search**: Reduced computation by simultaneously searching from start and goal, offering efficiency in symmetrical networks.
- **UCS**: Returned optimal paths but at higher computational cost due to full graph traversal.
- **GBFS**: Returned fast but often suboptimal paths due to reliance on heuristic only.
- **BFS and DFS**: Useful for demonstration but impractical for real-time routing due to their exhaustive nature and lack of cost-awareness.

Visual inspections and console comparisons confirmed that A* most consistently included the expected SCATS nodes in sequence, aligning well with real-world road usage. However, output inconsistencies in the printed route vs. visual map indicated minor bugs in segment tracking that were later patched.

C. Lessons Learned

1) Data Preprocessing is Critical: Early stages revealed that raw SCATS Excel data contained inconsistencies (e.g., missing site IDs, malformed road names). Thorough cleaning and normalization were essential for building an accurate metadata and adjacency structure.

2) Caching Improves Search Efficiency: Originally, predicting travel time for every edge during A* traversal created performance bottlenecks. Introducing a travel time cache significantly reduced runtime by avoiding redundant model inference.

3) Visualization is Essential for Debugging: The folium-based map visualization revealed path discrepancies not visible in console logs. This proved essential in validating both the correctness and completeness of route outputs.

4) Model Generalization Needs Real-Time Data: All models were trained on a fixed month (October 2006). As such, predictions may not generalize to contemporary or special-event traffic. Future improvements would require live SCATS feeds or simulation data.

5) Path Tracking Requires Explicit Logging: Some correct paths omitted intermediate SCATS nodes in text outputs. To resolve this, segment costs were explicitly tracked and passed along with the final path—ensuring accurate reporting.

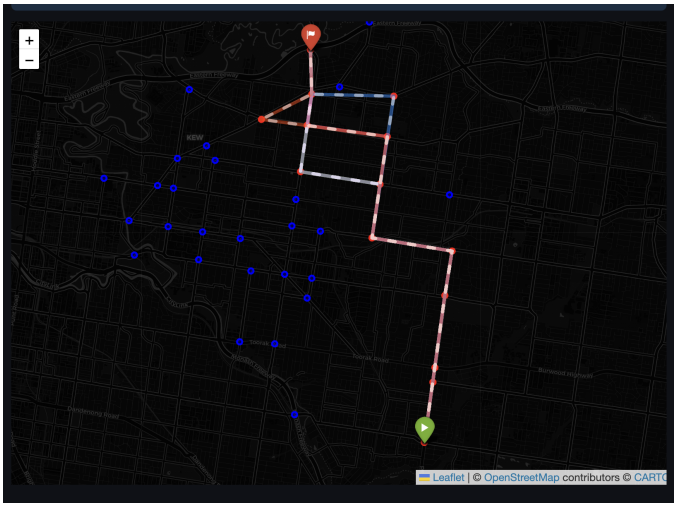

Fig. 5. Folium Map in GUI

6) Combining Multiple Datasets is Error-Prone: A major early challenge was figuring out how to integrate the different provided datasets. Initially, it seemed that all should be used simultaneously, which introduced inconsistencies. Later refinement involved selecting and cleaning only the most relevant data—especially for road names and SCATS connectivity.

Overall, the project strengthened understanding of how AI and classical algorithms can be integrated for intelligent transportation solutions.

V. Ethical Considerations in TBRGS+

While the current TBRGS system uses only anonymized SCATS traffic data, a future extension—TBRGS+—could offer a mobile app interface with real-time location tracking, personalized route suggestions, and user profile integration. However, such capabilities introduce critical ethical concerns related to surveillance, data privacy, algorithmic bias, and consent.

A. User Tracking and Consent

TBRGS+ may require continuous access to a user's GPS location to provide live route updates. Without explicit and informed consent, this constitutes a violation of personal privacy. Additionally, persistent tracking raises concerns about data security, especially if logs are stored or analyzed without proper anonymization.

B. Profiling and Personalization

To enhance user experience, TBRGS+ could integrate user preferences. However, personalization algorithms may inadvertently discriminate by reinforcing socio-economic or racial biases, especially if training data reflects historical disparities [5].

C. Data Retention and Reuse

Storing user routes or behavioral patterns introduces long-term risks if such data is repurposed without consent, sold to third parties, or exposed during a breach. Ethical design mandates implementing strict data retention policies, encryption at rest, and clear opt-in/opt-out mechanisms [6].

D. Algorithmic Transparency and Fairness

TBRGS+ should clearly explain why specific routes are chosen—especially if certain users receive slower or less direct suggestions. Users must be able to understand and challenge recommendations. This aligns with the OECD principle of transparency [6] and the CSIRO framework for Responsible AI in Australia [7].

VI. Findings: Design Alternatives and Ethical Implications

This section outlines the key technical decisions made in TBRGS and TBRGS+, along with alternative approaches and their respective ethical impacts.

A. TBRGS – Decision 1: Choice of Search Algorithm

Chosen Option: A*, UCS, GBFS. **Alternative:** Real-time traffic APIs (e.g., Google Directions). **Ethical Impact:** Using in-house algorithms ensures data sovereignty and avoids third-party dependence but requires maintaining fairness in route evaluation. External APIs offer higher fidelity but may expose user queries to surveillance.

B. TBRGS – Decision 2: Model Transparency

Chosen Option: Use of explainable metrics (travel time, route segments). **Alternative:** Black-box models with opaque outputs. **Ethical Impact:** The current design supports explainability and trust, whereas black-box solutions may violate the user's right to understanding and accountability.

C. TBRGS – Decision 3: Algorithm Selection Availability

Chosen Option: Allow user to choose between 6 different algorithms. **Alternative:** Restrict to only one or two well-tested algorithms (e.g., A* or Bidirectional). **Ethical Impact:** Giving users algorithm choice promotes transparency and education but may lead to misleading outputs (e.g., from DFS) if not explained. Restricting choices may reduce user control but improves reliability.

D. TBRGS+ – Decision 4: Real-Time GPS Tracking

Chosen Option: Continuous tracking. **Alternative:** Intermittent tracking or manual check-ins. **Ethical Impact:** Real-time tracking increases route accuracy but risks privacy loss. Intermittent tracking respects privacy but may degrade service performance.

E. TBRGS+ – Decision 5: Personal Data Storage

Chosen Option: Store user profiles for route personalization. **Alternative:** Use stateless personalization (e.g., in-memory session-based). **Ethical Impact:** Stored profiles improve UX but increase breach risk and consent complexity. Stateless alternatives reduce long-term risk and align with data minimization principles.

TABLE I
COMPARISON OF ETHICAL IMPLICATIONS ACROSS DESIGN DECISIONS

Decision	Ethical Benefit	Potential Risk
Use of A* Search	Transparent routes	Biased node weights
Use of Bidirectional	Faster computation	May miss asymmetrical traffic
GPS Tracking	Accurate routing	Privacy violation
User Profiles	Personalized UX	Data misuse

VII. CONCLUSION AND RECOMMENDATIONS

The integration of AI into transportation systems such as TBRGS and its future extension TBRGS+ offers numerous operational advantages. However, these benefits must be balanced against ethical concerns related to privacy, fairness, and accountability.

Our analysis shows that TBRGS upholds several Responsible AI principles, including transparency and safety, though it could benefit from broader data granularity and real-time model validation. The proposed TBRGS+ system introduces significant ethical risks, particularly regarding personal data handling and algorithmic bias.

To address these challenges, we recommend:

- Implementing consent-first, opt-in design patterns for any user tracking.
- Using data minimization and anonymization strategies to reduce exposure.
- Ensuring algorithmic transparency and explainability in personalized recommendations.
- Establishing regular audits of model behavior and data usage.

By proactively embedding ethical safeguards into both current and future system designs, developers can promote user trust, ensure regulatory compliance, and contribute to the responsible deployment of AI in public infrastructure.

GENERATIVE AI DECLARATION

Portions of this report, including structural outlines, grammar revision, and synthesis of public research, were assisted by OpenAI's ChatGPT (GPT-4). The author reviewed and verified all content to ensure factual accuracy and academic integrity. No fabricated content or references were used.

REFERENCES

[1] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.

[2] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "The initial design of a system for ai ethics governance: The papa framework," *Science and Engineering Ethics*, vol. 26, pp. 2141–2168, 2020.

[3] I. D. Raji, A. Smart, R. White *et al.*, "Closing the ai accountability gap: Defining an end-to-end framework for internal algorithmic auditing," *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33–44, 2020.

[4] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, pp. 1–21, 2016.

[5] M. Coeckelbergh, *AI Ethics*. MIT Press, 2020, used for theoretical grounding on ethical principles.

[6] OECD, "Oecd principles on artificial intelligence," https://www.oecd.org/going-digital/ai/principles, 2019, accessed May 2025.

[7] C. Scientific and I. R. O. (CSIRO), "Artificial intelligence: Australia's ethics framework," https://www.csiro.au/en/work-with-us/services/consultancy-strategic-advice-services/csiro-futures/futures-reports/AI-ethics, 2021, accessed May 2025.