Ethical Assessment in Smart City Decision-making

Patrizio Migliarini[1], Mashal Afzal Memon[1], Marco Autili[1]
[1]*DISIM - Department of Information Engineering, Computer Science and Mathematics*
University of L'Aquila, Italy

***Abstract*—This work introduces a system for ethical assessment of decision-making processes in smart city infrastructures, including traffic control, energy allocation, and public service prioritization. The system evaluates automated decisions by applying a configurable set of normative criteria such as fairness, harm avoidance, proportionality, and contextual relevance. The evaluation is performed through an ensemble of large language models (LLMs), each prompted with structured queries that reflect diverse ethical frameworks and reasoning strategies. These models simulate different normative perspectives and produce independent judgments, which are then aggregated and analyzed. The methodology is applied to simulated scenarios representing typical urban contexts, where system decisions affect agents with heterogeneous attributes. Results indicate that the system effectively identifies ethically ambiguous or problematic decisions, highlights sources of normative divergence, and supports context-sensitive analysis without reliance on domain-specific hardcoding.**

***Index Terms*—Ethical Evaluation, Smart City AI, LLM Ensemble, Normative Reasoning**

I. Extended Abstract

The integration of autonomous decision-making into smart city infrastructures, such as traffic optimization, energy distribution, and public service prioritization, raises new challenges in ensuring that automated decisions conform to normative expectations of fairness, transparency, and harm mitigation [1]–[3]. While these systems typically optimize operational metrics such as efficiency or throughput, they seldom incorporate mechanisms for assessing the ethical acceptability of the decisions they generate [4]–[6]. This work presents a system designed to operate in parallel with AI-based urban services to perform structured ethical assessment of individual decisions. The system evaluates outputs, such as access prioritization or dynamic resource assignments, by checking their consistency with a configurable set of normative criteria, including proportionality, equity, harm avoidance, and contextual relevance. Rather than relying on static rule-based logic, the system leverages a prompt-based ensemble of large language models (LLMs), each instantiated to simulate a distinct normative stance. Each LLM in the ensemble is prompted with structured queries derived from the scenario under analysis. These prompts encode the relevant ethical dimensions and contextual elements of the decision in a way that enables each model to render an independent judgment. The LLMs are assumed to reflect different reasoning strategies, corresponding to variations in training data, alignment objectives, or ethical frameworks. By aggregating the ensemble responses, the system produces a composite evaluation of ethical acceptability, identifying consensus, divergence, and dominant normative justifications. The methodology is evaluated on a collection of simulated urban scenarios, constructed to reflect realistic and representative smart city decision points. These include, for example, conflicts between mobility urgency and pedestrian vulnerability, or between energy efficiency and equitable access. Although fully synthetic, the scenarios are built to preserve structural properties common to real-world deployments. Preliminary results show that the system can effectively flag ethically problematic decisions, reveal hidden inconsistencies in policy-driven automation, and support fine-grained analysis of normative tensions. The ensemble approach also provides robustness against prompt phrasing variability and improves interpretability through comparative ethical reasoning. This approach enables transparent and extensible ethical auditing of AI-based services deployed in public-sector environments. It contributes a replicable framework for embedding normative sensitivity into urban automation pipelines and opens new directions for hybrid human-machine oversight in ethically sensitive public decision-making.

A. Data flow overview

- Synthetic Urban Scenario Generation: Structured JSON inputs are used to simulate urban scenarios, detailing agents (e.g., vehicles, pedestrians), their attributes (e.g., urgency levels, mobility constraints), and system decisions (e.g., traffic signal adjustments). This approach ensures controlled and diverse testing environments.
- Prompt Generation for LLM Ensemble: The system translates these scenarios into structured prompts, guiding each Large Language Model (LLM) in the ensemble to evaluate the ethical implications based on predefined criteria such as fairness, harm avoidance, and respect for vulnerability.
- LLM Ensemble Evaluation: Each LLM, potentially fine-tuned on different ethical frameworks or datasets, provides an independent assessment. This ensemble approach captures a spectrum of ethical perspectives, enhancing the robustness of the evaluation.
- Aggregation and Analysis: The approach aggregates the LLM outputs, analyzes consensus levels, identifies divergences, and determines the ethical acceptability of the decision. This step ensures a comprehensive understanding of the ethical landscape of each scenario.
- Outcome Generation: Based on the analysis, actionable insights are produced, such as flagging decisions for manual review or suggesting policy adjustments, thereby

integrating ethical considerations into smart city decision-making processes.

B. Example of simulated data and outcomes

Below we report examples of simulated input data, generated prompt, LLM responses, system analysis, and system outcome:

Simulated input data


```
{
  "scenario_id": "TRF_034",
  "domain": "traffic_control",
  "timestamp": "2025-06-03T08:15:00Z",
  "location": {
    "city": "Bologna",
    "intersection_id": "X_A12",
    "coordinates": [44.4949, 11.3426]
  },
  "agents": [
    {
      "id": "veh_ambulance_27",
      "type": "ambulance",
      "urgency_level": "medium",
      "eta_to_intersection": 14.0,
      "priority_code": "YELLOW"
    },
    {
      "id": "ped_user_843",
      "type": "pedestrian",
      "attributes": {
        "mobility": "disabled",
        "age": 71,
        "crossing_speed": "slow"
      },
      "eta_to_crossing": 5.0
    }
  ],
  "decision": {
    "action": "extend_green",
    "target": "ambulance",
    "effect_on_others": {
      "pedestrian_crossing_delay": 12
    }
  },
  "decision_maker": {
    "system_id": "smart_traffic_ai_v3.2",
    "policy_profile": "performance_optimized"
  }
}
```


Generated prompt

```

A smart traffic control system in Bologna decided to extend
the green light at intersection X\_A12 for an approaching
ambulance with medium urgency (ETA 14s). As a result,
a 71-year-old pedestrian with limited mobility who was
about to cross will be delayed by 12 seconds. The system
prioritizes performance over precaution.


Is this ethically acceptable?

Please evaluate using the following criteria:

- Fairness
- Harm avoidance
- Respect for vulnerability
- Contextual proportionality

Answer YES or NO.

Justify briefly.
```

LLMs responses

```

Model   | Acceptable? |   Justification

GPT-4-o | YES |    The ambulance case justifies minor
                   delay; 12 seconds is proportionate.

LLaMA-3 | NO  |    Vulnerable pedestrians should never be
                   deprioritized; risk is not quantified.

Claude  | YES |    Provided that the pedestrian is warned,
                   the harm is minimal.

Gemma   | NO  |    AI should follow pedestrian-first
                   policy unless critical emergency.
```

System analysis

```

Consensus Score = 2 YES / 2 NO = borderline ethical
     acceptability.

Normative divergence = high, suggesting ethical ambiguity.

Contextual indicator alignment = partial (fairness +
     proportionality = YES, vulnerability = NO).

Stability under perturbation = low (paraphrased scenario
     yields different answers).
```

System outcome


```
{
  "ethical_status": "ambiguous",
  "risk_level": "moderate",
  "recommended_action": "manual audit",
  "dominant_violation": "vulnerability disrespect",
  "suggested_rephrase": "Consider alternative routing
                         that avoids pedestrian delay"
}
```


References

[1] I. Yaqoob, L. U. Khan, S. A. Kazmi, M. Imran, N. Guizani, and C. S. Hong, "Autonomous driving cars in smart cities: Recent advances, requirements, and challenges," *IEEE Network*, vol. 34, no. 1, pp. 174–181, 2019.

[2] K. Kuru and W. Khan, "A framework for the synergistic integration of fully autonomous ground vehicles with smart city," *IEEE Access*, vol. 9, pp. 923–948, 2020.

[3] B. Townsend, C. Paterson, T. Arvind, G. Nemirovsky, R. Calinescu, A. Cavalcanti, I. Habli, and A. Thomas, "From pluralistic normative principles to autonomous-agent rules," *Minds and Machines*, vol. 32, no. 4, pp. 683–715, 2022.

[4] J. Leikas, R. Koivisto, and N. Gotcheva, "Ethical framework for designing autonomous intelligent systems," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 5, no. 1, p. 18, 2019.

[5] K. Michael, R. Abbas, G. Roussos, E. Scornavacca, and S. Fosso-Wamba, "Ethics in ai and autonomous system applications design," *IEEE Transactions on Technology and Society*, vol. 1, no. 3, pp. 114–127, 2020.

[6] L. Dennis, M. Fisher, M. Slavkovik, and M. Webster, "Formal verification of ethical choices in autonomous systems," *Robotics and Autonomous Systems*, vol. 77, pp. 1–14, 2016.