AiMTT AI Learning Initiative for Multi-modal Traffic and Transportation

Large Language Models: Shaping the Future of Transportation?

AiMTT tutorial #1
Published: July, 2025

The authors
Ting Gao, Mahsa Movaghar, Theivaprakasham Hari and Alex Roocroft are researchers at TU Delft's DAIMOND Lab. They are supervised by Dr. Ir. Marco Rinaldi and Dr. Yanan Xin, co-directors of the Lab, and Prof. Dr. Ir. Serge Hoogendoorn, advisor to the Lab.

Introduction

Are you using tools like ChatGPT in your daily life to help write an email or even draft a construction plan? Just ten years ago, these kinds of capabilities would have seemed unimaginable. Today, they're becoming part of everyday life for ordinary people. Behind these powerful tools are technologies known as Large Language Models (LLMs)—AI systems that can understand and generate human-like text and now even create images and videos. But what exactly are LLMs? Could they help transform fields like transportation and traffic management? Can they really do everything, or are there still limitations? In this article, we'll walk you through a general introduction to LLMs: what they are, how they work, and what opportunities—and challenges—they bring to the transportation sector.

What is a LLM?

Today, terms like "LLM" and "GPT" are becoming increasingly common in everyday conversations. LLM stands for Large Language Model, a broad category of artificial intelligence (AI) systems trained on massive text datasets to understand and generate human-like language. GPT, short for Generative Pretrained Transformer,

refers to a specific type of LLM developed by OpenAI, the organization behind the widely recognized tool ChatGPT. This platform enables users to perform various tasks, from summarizing text and translating languages to completing phrases and solving more complex challenges like mathematical problems.

Since 2019, the development of large language models has accelerated rapidly, with OpenAI's GPT series gaining significant attention, particularly following the release of ChatGPT in 2022. Meanwhile, other major players have advanced in the field, including Meta's LLaMA, Google's Gemini, Anthropic's Claude, and Alibaba's Qwen, each offering distinct strengths and improved accessibility. By 2025, models like DeepSeek-R1 have gained recognition for their strong reasoning capabilities, achieved with significantly lower computational demands and costs.

Just as the human body is made up of countless cells, an LLM consists of millions or even billions of parameters—numerical values arranged in specific structures. While young humans learn through sensory experiences and feedback, LLMs acquire knowledge by training on vast datasets and refining their internal parameters based on the feedback they receive during training. Unlike traditional machine learning models, LLMs are notable for their scale. They are trained on terabytes of data drawn from diverse domains such as websites, books, academic papers, code repositories, news articles, social media, and multimedia transcripts.

Behind the scenes, an LLM operates very differently from how humans use language. At its core, it handles language as numerical data. To make text understandable for a machine, a step called tokenization is applied, which breaks text into smaller units called tokens. A simple analogy would be assigning each word a unique ID so a sentence can be represented as a sequence of IDs. However, this approach can lead to issues with words not found in the predefined vocabulary—so-called out-of-vocabulary problems. To solve this, models use advanced techniques such as byte-pair encoding (BPE), WordPiece, and SentencePiece, which break words into smaller, reusable subword units. This allows the model to handle virtually any input using a limited set of components.

At the heart of modern LLMs is a concept called self-attention mechanism. Imagine you're reading a long document and trying to understand the meaning of a sentence—you don't treat every word equally. Instead, your brain naturally focuses more on the words that seem most important. Attention mechanisms work similarly: they allow AI models to "pay more attention" to the most relevant parts of the input when generating a response. This idea forms the foundation of the transformer architecture, which powers modern language models like GPT.

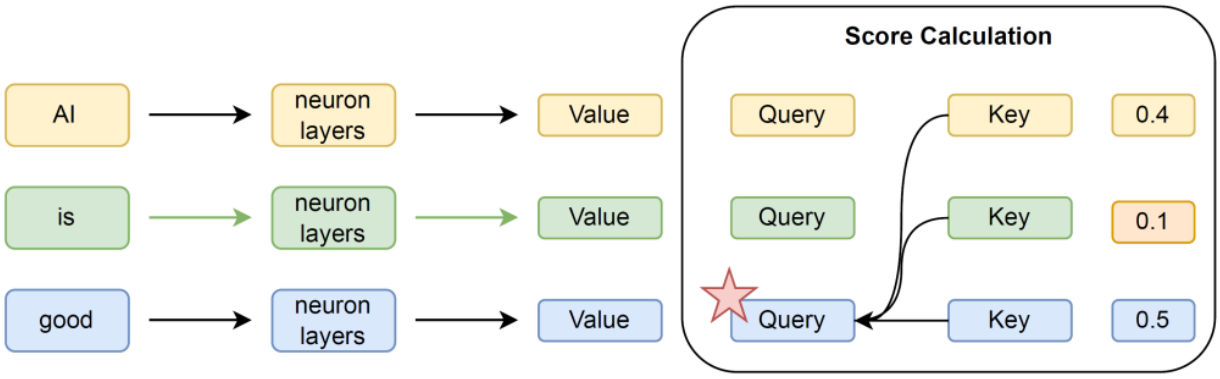

Illustration of how the word "good" is processed and updated through the self-attention mechanism when the input phrase is "AI is good".

So, how does self-attention actually work? Let's take the example phrase as in Figure 1: "AI is good." For simplicity, we treat each word as a single token. After several processing layers, each word is transformed into three vectors: a *query*, a *key*, and a *value*. To update the representation of the word "good", we compute how much attention it should pay to every other word in the sentence. The *query* vector comes from "good," while the *key* vectors come from all the words. By comparing the *query* with each *key*, we obtain attention scores that determine how much weight to give each word. In this example, the word "good" assigns a score of 0.5 to itself and 0.4 to "AI," indicating it pays the most attention to these two words. Finally, "good" is updated by summing the value vectors of all the words, weighted by these attention scores.

The self-attention mechanism is, in practice, the most effective and scalable method for enabling LLM to extract meaningful contextual information. It updates the representation of each word based on its relationship with all the other words in the sentence. When the model uses multiple sets of attention vectors (*queries, keys, and values*), it can explore a richer representation space and capture a broader range of semantic features and relationships. This approach is known as the multi-head self-attention mechanism.

At the final stage, the model outputs a probability distribution over all tokens in its vocabulary. Text is then generated by selecting tokens based on these probabilities. One simple approach is to always choose the token with the highest probability at each step, similar to always clicking the first suggestion in your phone's autocomplete—this often results in deterministic and less-than-ideal responses. To introduce variety and enhance creativity, sampling techniques are used instead. For

example, the model can sample from only those tokens whose probabilities exceed a certain threshold, enabling more diverse and contextually rich outputs.

One of the most intriguing phenomena in LLMs is the emergence of new capabilities as these models scale in size and data. Jason Wei, a research scientist at Google Brain, and his colleagues define an emergent ability as one that "is not present in smaller models but is present in larger models," emphasizing that such abilities "cannot be predicted simply by extrapolating the performance of smaller models." Examples include improved reasoning, more accurate translation, effective summarization, and more human-like text generation, marking a significant leap in AI performance.

LLMs have the potential to positively impact transportation. They are being explored to enhance autonomous vehicles by improving perception and decision-making. LLMs could also be employed to optimize routes for logistics and act as multilingual mobility assistants. Furthermore, LLMs are being explored as policy interpreters, assisting planners in navigating complex traffic laws. Finally, they could be employed in traffic management, analyzing vast datasets to forecast congestion and suggest countermeasures.

Applications in transportation

LLM in Autonomous Vehicles

Autonomous vehicles (AVs) have long promised safer roads, smoother mobility, and smarter transport systems. Now, LLMs are emerging as a powerful tool to accelerate that vision. Thanks to their unique capabilities, LLMs can help AVs not only perceive the world but also explain their actions and interact naturally with human users.

At first glance, using language models for driving may seem abstract, as driving does not feel like writing. But once perception is handled, what the vehicle 'sees' can be translated into a stream of tokens, much like a sentence. Recent models like GPT-Driver and DriveGPT-4 harness this by treating driving as a language modeling task. For instance, the model might generate a sentence such as "Changing lanes to maintain a safe distance". The result is not just a control command, but an interpretable, human-readable rationale for the action. This ability to generate and explain behavior in natural language could reshape foundational components of AV systems, including perception, strategic and tactical decision-making.

In perception, multimodal LLMs integrate diverse sensor inputs, such as LiDAR and cameras, with contextual understanding. Unlike conventional deep learning systems that often struggle in rare or unpredictable situations (the so-called long-tail

problem), LLMs can generalize more effectively and adapt with less retraining. With their reasoning capabilities, these tools can not only detect pedestrians but would also be able to reason about their actions, e.g. understanding why people are gathering at a crossing or interpret a cyclist's hand signal.

In decision-making, LLMs improve natural language interfaces that allow passengers to issue spoken instructions like "merge left" or "slow down for the cyclist". These models assess the feasibility of the command using perception data and can respond with verbal explanations, enhancing transparency and trust. The UK-based company Wayve has taken this further, integrating a text-based LLM into its driving operation so passengers can ask questions such as, "Why aren't you moving?" and receive real-time answers.

Motion planning and control are also evolving. Traditional trajectory planners often rely on rigid, rule-based systems that work well in structured environments but falter in complex or unfamiliar situations. In contrast, LLM-based models like GPT-Driver generate vehicle trajectories using language modeling techniques applied to coordinate descriptions. This draws on LLMs' general reasoning ability to produce smoother, context-aware paths across diverse scenarios. In control, precision, and comfort remain essential for safety and acceptance. LLMs can be increasingly fine-tuned with human feedback to mimic human-like driving better. Unlike conventional modular pipelines, where perception, planning, and control are separate, LLM-based systems can unify these into a single reasoning loop, enabling more fluid, adaptive responses. However, a significant challenge remains as motion planning and control are real-time tasks that require fast algorithms capable of making decisions within milliseconds, something current LLMs are not yet able to deliver reliably for operational use.

While integrating LLMs into AV systems shows great promise, it also brings new challenges. The same flexibility that allows these models to generalize and articulate reasoning can also lead to unsafe behavior, such as imitating risky human maneuvers or "hallucinating" plausible but incorrect decisions. To manage these risks, hybrid AI approaches are emerging. At U.S.-based Nuro, for example, rule-based and AI-based planners run in parallel, and their proposals are evaluated for safety. This kind of architecture may offer a safety net for future LLM integration, balancing the explainability of rules with the adaptability of language models, thereby relying on an Ensemble decision. Advancing reliability, interpretability, and robustness will be key to the safe and effective deployment of LLMs in future AV systems.

LLMs in Logistics

LLMs are driving innovations in logistics, with one key role in improved route optimization, as shown in a recent survey by Chinese researcher Dingkai Zhang and his colleagues[1]. LLMs are adept at finding patterns in complex datasets – they can crunch historical transit schedules, real-time GPS info, and rider demand to design better routes and timetables. In public transport, researchers have shown that an LLM can analyze past ridership and delays to help planners tweak bus routes or train schedules for shorter wait times and more reliable service. Logistics companies are likewise experimenting with LLMs to chart efficient delivery routes. One global shipping firm "Uber Freight" reportedly used an LLM to suggest optimal truck delivery sequences by learning from traffic reports and past deliveries, saving time and fuel on its routes. It's as if these AI tools serve as hyper-intelligent dispatchers or navigators, always one step ahead of traffic conditions.

LLMs in Mobility Assistance

LLMs are also stepping up into the role of passenger assistants. Transit agencies and travel apps are deploying chatbots powered by models like ChatGPT to help riders in real time. These assistants can provide personalized travel advice – from helping a tourist figure out the best subway connection, to alerting a daily commuter about delays on their route, all through a natural conversation. Early studies found that LLMs can handle complex commuter questions and give largely accurate, context-aware answers about transit services, improving the overall passenger experience. Crucially, these AI assistants can communicate in multiple languages, a boon in global cities. For instance, an LLM-based system can instantly respond to a rider's query in Dutch, Spanish, or Chinese as well as in English, ensuring non-native speakers get equal access to transit information. This kind of always-available, multilingual support can make public transportation more inclusive and user-friendly.

In the realm of traffic law assistance, LLMs can serve as policy interpreters. Self-driving car developers are using language models to interpret the labyrinth of traffic regulations and safety guidelines. In one recent experiment at UCLA, researchers fed an LLM a library of driving rules and taught it to reason through them during different driving scenarios. The AI system could infer/retrieve relevant regulations (like speed limits or right-of-way rules) based on the situation and then explain in plain terms

---

[1] Zhang, D., Zheng, H., Yue, W., Wang, X. (2024). Advancing ITS Applications with LLMs: A Survey on Traffic Management, Transportation Safety, and Autonomous Driving. In: Hu, M., Cornelis.

whether the car's planned action was legal and safe. Essentially, the LLM acted like a legal scholar for the vehicle, parsing through dense policy documents and translating them into practical guidance. This shows how LLMs might help planners and engineers as well, by quickly answering questions about transportation policies, interpreting new traffic laws, or summarizing lengthy regulation drafts. From predicting traffic jams to chatting with passengers or decoding policy, LLMs are becoming versatile new players in the transportation sector.

LLM in Traffic Management and Safety

Modern traffic management faces numerous challenges, including congestion, safety risks, and decision-making in real-time conditions. LLMs ability to process and integrate heterogeneous transportation data—ranging from text, sensor data, and images to user feedback—makes them powerful tools for understanding complex traffic patterns. This ability not only enhances traffic flow analysis but also supports more efficient and timely decision-making, thereby improving overall system performance.

For example, LLMs can help with safety-critical decision-making by generating actionable recommendations or providing language-based instructions. "AccidentGPT," conceptualized by groups of researchers[2] in 2024, exemplifies this application by acting as a safety advisor that converts multi-sensor data to predict accidents, generate safety warnings, and offer dialogue-based recommendations for preventative actions. In addition, LLMs can process accident reports to uncover contributing factors, identifying under-reported issues such as alcohol use, fatigue, or distracted driving.

Beyond safety, LLMs can also be beneficial in traffic data imputation and macroscopic mobility prediction. They can decode complex dependencies across diverse datasets, enabling the modeling of interactions between infrastructure, environmental conditions, and user behavior. This capability allows for more accurate traffic forecasting, optimized infrastructure management, and a better understanding of how various factors influence traffic dynamics.

[2] Wang, H., Ma, S., Dong, L., Huang, S., Zhang, D., Wei, F., 2024b. Deepnet: Scaling transformers to 1,000 layers. IEEE Transactions on Pattern Analysis and Machine Intelligence 46 (10), 6761–6774.

Challenges of LLMs

Technical Challenges

Deploying large language models in transport may look like a quick software upgrade, but under the hood, it is a four-part engineering marathon. First comes training: a general-purpose model must be fine-tuned on highly specialized data—timetables, sensor logs, policy texts—so it speaks the "dialect" of traffic without forgetting its broader knowledge, a process that demands sizeable labelled datasets, GPU hours, and staff expertise that smaller agencies rarely have. Next is integration and trust. Intelligent-transport-system stacks built along decades need new interfaces so an LLM can read live feeds and return commands; meanwhile, operators must guard against "hallucinations", which still appear in 3% of ChatGPT answers and up to 27% in some rival models. Even if the logic is solid, performance hinges on infrastructure: thousands of commuter queries or routing calls per minute can overwhelm cloud budgets or introduce lag seconds, pushing designers toward edge deployments where trimmed-down models sit on local servers—cheaper and faster, but less refined. However, a significant challenge for this edge approach is the current scarcity and high hardware cost suitable for embedded LLM deployment with local or node-level compute. What limited hardware does exist is often prohibitively expensive, especially for systems requiring enough memory to run competitive large models like DeepSeek-R1 671B, which can demand around 1,543GB of RAM [3]. Finally, there is the energy bill: training GPT-3 alone consumed about 1287 MWh and emitted 552 tons of carbon dioxide [4], while each live query keeps power-hungry GPUs humming—ironic when the same agencies are trying to cut transport emissions. Together, these hurdles explain why most projects still start as limited pilots, giving engineers time to refine data pipelines, bolt on safety checks, and explore greener hardware before letting AI take the wheel.

Ethics

If LLMs are to become embedded in transport systems, the opportunities also come with significant ethical challenges, particularly around safety, fairness, and privacy.

---

[3] DeepSeek-AI, Guo D, Yang D, Zhang H, Song J, Zhang R, et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning [Internet]. 2025. Available from: arxiv.org/pdf/2501.12948.

[4] Tomlinson, B., Black, R.W., Patterson, D.J. et al. The carbon emissions of writing and illustrating are lower for AI than for humans. Sci Rep 14, 3732 (2024).

In transport, safety is paramount. Poorly trained or deployed LLMs could misinterpret traffic scenarios or provide misleading guidance. In high-stakes contexts, such as autonomous vehicle crash responses, this could lead to serious harm and loss of public confidence.

Fairness and equity concerns are also vital. LLMs can unintentionally reinforce existing biases in training data, disadvantaging already under-served populations. For example, route optimization tools may neglect low-income or low-mobility areas, extrapolating from existing, bias-affirming mobility trends, unless explicitly designed with inclusivity in mind. Furthermore, ensuring equitable access is essential to prevent digital divides in mobility.

Privacy poses another critical concern. When LLMs process sensitive transport data, such as travel histories, there is a risk of misuse or exposure. Privacy-preserving techniques, like anonymising location data, are needed to protect individual rights while maintaining system performance.

To integrate LLMs ethically, the transport sector must use robust frameworks for accountability, transparency, and risk mitigation. This includes clear documentation of model behavior, auditability of decision processes, and stakeholder involvement in deployment strategies.

Conclusion

We are living in the age of AI, where its pace and innovation have never been faster. Advancements in Large Language Models (LLMs) are transforming not only industries but also our daily interactions with machines, reshaping personal experiences. With their unique capabilities, they are becoming powerful tools for handling and automating many of our daily tasks, including technical and engineering procedures.

Their ability to understand and analyze heterogeneous traffic data sources, generate actionable and language-based instructions and recommendations, and represent domain knowledge from unstructured data opens up new opportunities for the next generation of "Intelligent Transportation". Companies and industries can make the most of these recent advances, from automating tasks and procedures to solving more technical and challenging problems.

But do all these opportunities come with no cost or risk? Can we ensure that long-term use of LLMs contributes to a safe, equitable, and sustainable society? Should we blindly trust these machines to make decisions in critical domains, especially in safety and policy making, which affect large groups of people? Should we expect bias

and unfairness in the solutions and recommendations they provide? Are these models transparent enough to be fully trusted? Can they be a threat to the privacy of people?

Considering all these critical questions surrounding the use of LLMs, we must recognize that we are still in the early stages of their development and cannot yet provide proper and valid answers. LLMs offer transformative potential, but they also have limitations. Although they excel in reasoning and responding to natural language, LLMs lack a coherent understanding of physical truth and instead rely on proxy models formed through language relations. This means that, while LLMs can grasp the structure and function of language, they do not possess a genuine understanding of meaning in the way humans do. It is up to engineers and modelers to ensure that these models are sufficiently aligned with reality for high-stakes decision-making. Although LLMs can empower industries, human judgment, critical thinking, and domain expertise remain irreplaceable for their safe and responsible use.