

## Proposition de sujet de thèse CIFRE

### **Génération automatique de scénarios de simulation à partir de descriptifs de scénarios de conduite automobile**

**Localisation :** AMPERE/RENAULT & LAMIH UMR CNRS 8201/UPHF.

**Equipe d'encadrement :** Chouki Sentouh ([Chouki.Sentouh@uphf.fr](mailto:Chouki.Sentouh@uphf.fr))  
Jean-Christophe Popieul ([Jean-Christophe.Popieul@uphf.fr](mailto:Jean-Christophe.Popieul@uphf.fr))  
Hakim Mohellebi ([hakim.mohellebi@renault.com](mailto:hakim.mohellebi@renault.com))

**Financement :** Thèse CIFRE

#### **Contexte scientifique**

Une validation robuste des AD/ADAS atteignant les niveaux de sécurité requis est un enjeu stratégique tant pour les constructeurs automobiles que pour les pouvoirs publics qui ont la responsabilité de garantir la sécurité des conducteurs et des usagers de la route. Pour atteindre les objectifs de sécurité définis par les pouvoirs publics, selon le système à valider, les constructeurs automobiles réalisent plusieurs milliers voire plusieurs millions de kilomètres de roulage pour tester les véhicules face à un maximum de situation de conduite dans diverses conditions météorologiques et d'infrastructures. Bien que cette approche soit nécessaire, elle présente principalement deux inconvénients, le coût exorbitant des roulages ainsi que le manque de capitalisation des scénarios de conduite pertinents, en particulier les cas rares et les cas de presque accident. Ces inconvénients font de la simulation une solution clé pour réduire les coûts de validation et augmenter la couverture des plans de test en simulant les scénarios pertinents capitalisés dans un format numérique.

Ces dernières années, les autorités réglementaires ont vivement recommandé l'utilisation de la simulation et de catalogues de scénarios dans le processus de développement et de validation des systèmes avancés d'aide à la conduite (ADAS) et des véhicules automatisés. Dans ce contexte, l'initiative française appelée Adscene a réuni un grand nombre de scénarios de conduite visant à établir une base de données de scénario à l'état de l'art pour la conception, la validation et l'homologation des ADAS et des véhicules automatisés. Différents conteneurs sont dédiés pour capitaliser différents types de scénarios, notamment les accidents, les situations de presque-accidents, les scénarios réglementaires, Ncap, ainsi que les scénarios de conduite normale. Cette base de scénarios est équipée d'un modèle de données et d'une grammaire spécifique permettant de décrire de manière précise tous les types de situations de conduite. Chaque scénario est accompagné de plusieurs scènes soigneusement sélectionnées et d'un texte explicatif pour illustrer la situation de conduite de manière pertinente.

Actuellement, l'interprétation des descriptions de scénarios ADscene nécessite l'intervention d'un opérateur humain, car elles ne sont pas directement compréhensibles par les logiciels de simulation. Pour convertir ces descriptions en simulations exploitables, elles doivent être manuellement codées dans un format adapté au logiciel de simulation utilisé. Ce processus manuel est non seulement chronophage, mais aussi sujet à des erreurs, ce qui peut affecter la fiabilité et l'efficacité des simulations. Par conséquent, il est important de développer une solution automatisée qui pourrait interpréter et convertir ces descriptions de manière précise et efficiente, réduisant ainsi le temps nécessaire à la préparation des simulations et minimisant les risques d'erreurs.

Deux types de méthodes de génération de scénarios existent dans la littérature : les méthodes basées sur les modèles

et les méthodes basées sur les données (data-driven). Les approches basées sur les modèles, comme par exemple, le modèle IDM [1], peuvent générer un flux de scénario de conduite continu. Ces méthodes ont une interprétabilité claire, mais ne peuvent pas bien gérer les interactions complexes. Le processus d'exploration et de génération de scénarios particuliers, rares ou imprévus peut nécessiter un temps très important.

Pour les méthodes basées sur les données [7,16], elles apprennent généralement les modèles de conduite par apprentissage profond ou par renforcement sur de grandes quantités de scénarios naturalistes. Cependant, dans le processus de génération de scénarios, une compréhension sémantique limitée peut conduire à des scénarios qui manquent d'informations contextuelles nuancées requises pour les simulations. Les scénarios particuliers difficiles générés sont parfois irréalistes ou incontrôlables. De plus, les méthodes basées sur les modèles ou les données ne sont pas capables de suivre le raisonnement humain, ce qui réduit la fiabilité et la facilité d'utilisation des scénarios générés.

Récemment, le développement rapide des modèles de base, en particulier les Large Language Models (LLM), nous offre une nouvelle façon de résoudre la génération de scénarios [2,5]. Les LLM ont démontré de grandes capacités de compréhension sémantique et de raisonnement, ce qui leur permettent de générer des descriptions et des contextes de scénarios plus informatifs, tels que l'interprétation des comportements des véhicules en interaction [12,14].

### **Objectifs de la thèse**

Les travaux de recherches de la thèse proposée visent à relever le défi actuel associé à la nécessité d'une intervention manuelle pour convertir les descriptions de scénarios en simulations exécutables, un processus qui est long et susceptible d'erreur. L'objectif principal de ce projet de thèse est donc de développer des algorithmes avancés, notamment ceux fondés sur l'intelligence artificielle, en particulier les Large Language Models (LLM), pour l'analyse automatique et l'interprétation des descriptions de scénarios. L'ambition est que ces algorithmes soient capables de générer automatiquement des scénarios de simulation dans un format compatible avec des logiciels de simulation de conduite spécialisés. En effet, les plateformes de simulation, notamment celles dédiées à la simulation de conduite, ainsi que leur architecture et leur format informatique, sont généralement conçus et développés en fonction des objectifs de validation et ajustés en fonction des moyens de simulation cibles.

### **Missions principales**

Le travail s'organisera autour de quatre tâches principales :

- Il convient tout d'abord de réaliser un état de l'art sur les approches et algorithmes adaptés à la problématique de transcriptions de scénarios descriptifs vers des scénarios simulables. Dans cette phase, une analyse des difficultés rencontrées pour l'automatisation du codage des scénarios simulables sera réalisée.
- La 2<sup>ème</sup> étape vise à proposer et à mettre au point une ou plusieurs approches permettant d'automatiser le développement des scénarios.
- L'objet de la 3<sup>ème</sup> phase est de mettre en place un environnement expérimental permettant d'évaluer les approches proposées.
- La 4<sup>ème</sup> phase vise à étudier et valider l'approche proposée sur quelques cas d'application.

### **Profil du candidat**

Les candidats doivent détenir un diplôme de Master ou d'ingénieur en Automatique, informatique, intelligence artificielle, mathématiques appliquées ou équivalent. Plus généralement, le candidat retenu devra avoir un fort intérêt pour le domaine du véhicule autonome et le développement et réalisation de scénarios de conduite en simulateur. Le candidat doit montrer un fort intérêt à s'engager dans des recherches innovantes de haut niveau,

des compétences en développement logiciel (développement en C, C++, Python, Bibliothèques d'intelligence artificielle, Systèmes décisionnels), de l'autonomie et de l'organisation. Une bonne maîtrise de l'anglais est nécessaire pour publier et présenter les résultats lors de conférences internationales et dans des revues internationales.

### **Mots clés**

Génération automatique de scénarios, intelligence artificielle, Large Language Models (LLM), simulation de conduite, compréhension sémantique, raisonnement.

**Pour candidater :** Merci d'envoyer à [hakim.mohellebi@renault.com](mailto:hakim.mohellebi@renault.com) et [Chouki.Sentouh@uphf.fr](mailto:Chouki.Sentouh@uphf.fr):

1. CV comprenant votre expérience et vos connaissances professionnelles pertinentes,
2. Lettre de motivation et d'intérêts pour la recherche (avec les coordonnées d'au moins deux personnes de référence), en expliquant pourquoi vous souhaitez poursuivre en doctorat, quels sont vos intérêts académiques, comment ils se rapportent à vos études précédentes et à vos objectifs futurs.
3. Copie du ou des certificats et relevés de notes universitaires de vos établissements universitaires antérieurs.

### **Références bibliographiques :**

- [1] M. Treiber and A. Kesting, "The intelligent driver model with Stochasticity-new insights into traffic flow oscillations," *Transp. Res.Procedia*, vol. 23, pp. 174–187, Jan. 2017.
- [2] C. Chang, S. Wang, J. Zhang, J. Ge and L. Li, "LLMScenario: Large Language Model Driven Scenario Generation," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, doi: 10.1109/TSMC.2024.3392930
- [3] C. Chang *et al.*, "VistaScenario: Interaction Scenario Engineering for Vehicles with Intelligent Systems for Transport Automation," in *IEEE Transactions on Intelligent Vehicles*, doi: 10.1109/TIV.2024.3400323
- [4] S. Wang, Y. Zhu, Z. Li, Y. Wang, L. Li, and Z. He, "ChatGPT as your vehicle co-pilot: An initial attempt," *IEEE Trans. Intell. Veh.*, vol. 8, no. 12, pp. 4706–4721, Dec. 2023, doi: 10.1109/TIV.2023.3325300.
- [5] Z. Xu *et al.*, "DriveGPT4: Interpretable end-to-end autonomous driving via large language model," 2023, arXiv:2310.01412.
- [6] Cao, L.; Feng, X.; Liu, J.; Zhou, G. Automatic Generation System for Autonomous Driving Simulation Scenarios Based on PreScan. *Appl. Sci.* **2024**, *14*, 1354. <https://doi.org/10.3390/app14041354>
- [7] E. Esenturk, S. Khastgir, A. Wallace and P. Jennings, "Analyzing Real-world Accidents for Test Scenario Generation for Automated Vehicles," *2021 IEEE Intelligent Vehicles Symposium (IV)*, Nagoya, Japan, 2021, pp. 288-295, doi: 10.1109/IV48863.2021.9576007.
- [8] Duan, J.; Gao, F.; He, Y. Test scenario generation and optimization technology for intelligent driving systems. *IEEE Intell. Transp. Syst. Mag.* 2020, *14*, 115–127.
- [9] Guo, P.; Gao, F. Automated scenario generation and evaluation strategy for automatic driving system. In *Proceedings of the 2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, Changsha, China, 18–20 December 2020; pp. 1722–1733.
- [10] Shu, H.; Lv, H.; Liu, K.; Yuan, K.; Tang, X. Test scenarios construction based on combinatorial testing strategy for automated vehicles. *IEEE Access* 2021, *9*, 115019–115029.
- [11] D. Wu *et al.*, "Language prompt for autonomous driving," 2023, arXiv:2309.04379.
- [12] A. Elhafi *et al.*, "Semantic anomaly detection with large language models," *Autonomous Robots*, vol. 47, pp. 1035–1055, Oct. 2023.

- [13] L. Wang et al., “AccidentGPT: Accident analysis and prevention from V2X environmental perception with multi-modal large model,” 2023, arXiv:2312.13156.
- [14] H. Sha et al., “Language MPC: Large language models as decision makers for autonomous driving,” 2023, arXiv:2310.03026.
- [15] L. Wen et al., “DiLu: A knowledge-driven approach to autonomous driving with large language models,” 2023, arXiv:2309.16292.
- [16] L. Feng, Q. Li, Z. Peng, S. Tan, and B. Zhou, “TrafficGen: Learning to generate diverse and realistic traffic scenarios,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2023, pp. 3567–3575.