

# Summary: Autonomous Robot Navigation Using Large Language Models (LLM) and Semantic 3D Neural Reconstruction

---

## Internship Context

This internship focuses on advancing autonomous robotics by integrating Large Language Models (LLMs) and semantic 3D neural reconstruction for long-term action planning and navigation. Traditionally, AI used in autonomous robots focuses on short-term planning within the robot's immediate environment, constrained by predefined queries. To increase versatility, AI systems must evolve to plan actions based on natural language requests, using past observations, even outside the robot's current field of view.

Recent advancements in LLMs have shown that these models can plan sequences of actions in various domains by leveraging general world knowledge (known as chain-of-thought reasoning). However, for robots to fully exploit this capability, the LLM must be connected to the robot's abilities and the real world (grounding). This connection should not only be limited to the current perceivable environment but should include a 3D representation of the surroundings previously explored by the robot.

## Internship Objectives

The intern will explore the Vision-Language Navigation (VLN) domain, a multidisciplinary research area at the intersection of natural language processing, computer vision, and robotics. Key objectives include:

- Understanding and utilizing neural 3D scene representations for navigation tasks.
- Implementing a LLM in REACT mode to analyze user queries, plan navigation tasks, and invoke necessary tools (e.g., 3D scene analysis, trajectory calculation).
- Testing and evaluating the method in a simulation environment (e.g., Habitat).
- Deploying the demonstrator on a real robotic platform using the ROS framework for integration of localization and control algorithms.

## Skills Developed

During this internship, the student will gain knowledge in AI for robotics, combining advancements in 3D reconstruction (Neural Fields, open-vocabulary segmentation, etc.), robot control, and language models. The intern will also work with widely-used frameworks like ROS and NeRFStudio, alongside experienced researchers and PhD students.

## Desired Skills

- Proficiency in Python.
- Knowledge of robotics.
- Experience with PyTorch and/or ROS is a plus.

## General Information

- **Education Level:** Engineer, Master 2 (Bac+5).
- **Duration:** 6 months.
- **Location:** Palaiseau (91) – Centre d'intégration de Nano-INNOV.
- **Compensation:** Between €700 and €1400, depending on qualifications, with housing/transport/meal support.
- **Application:** Send CV and cover letter to [steve.bourgeois@cea.fr](mailto:steve.bourgeois@cea.fr).

## CEA Tech LIST Research Areas

CEA Tech LIST's research focuses on software-driven systems in three main areas: Embedded Systems, Interactive Systems, and Sensors & Signal Processing. They collaborate with industry leaders in nuclear, automotive, aerospace, defense, and medical sectors to develop innovative solutions.

## French Original

---

Exploitation d'un grand modèle de langage et d'une reconstruction 3D neurale sémantique pour la navigation autonome d'un robot

Contexte du stage

Si la robotique autonome connaît actuellement de grandes avancées, les intelligences artificielles qui guident ces robots se limitent généralement à une planification à court terme, sur l'environnement immédiat du robot, et pour des domaines de requêtes prédéfinies. Pour aller vers

plus de polyvalence, il est nécessaire de développer des Intelligences Artificielles (IA) capable de planifier, sur la base d'une requête en langage naturel, une série d'actions sur des horizons temporels lointain, impliquant des lieux ou des éléments qu'ils ont observés par le passé mais qui sont hors du champ de perception actuel du robot.

Or, l'essor récent des grands modèles de langage (LLM) a démontré la capacité de ces IA à exploiter

leur connaissance générale du monde pour planifier des séries d'actions (chaînes de pensées ou chain-of-thought) sur des domaines variés. Pour exploiter ces connaissances générales dans le cas d'un robot spécifique dans un environnement spécifique, il devient nécessaire de connecter le LLM aux capacités du robot et au monde réel qui l'entoure (encrage ou grounding). Pour tirer parti au maximum des capacités de planification à horizon lointain des LLM, cette connexion ne doit pas

se limiter au seule environnement perceptible à l'instant courant par le robot, mais à une représentation 3D de l'environnement qu'il a pu précédemment exploré.

Objectifs du stage

Dans le cadre de ce stage, nous proposons d'étudier le domaine du Vision-Language Navigation [1,2,3], domaine de recherche interdisciplinaire à la frontière du traitement naturel du langage, de la vision par ordinateur et de la robotique. L'étudiant aura pour objectif de mettre en place un

démonstrateur de navigation robotique autonome exploitant un LLM pour la compréhension de la

requête utilisateur et la planification d'actions, ainsi qu'une représentation neurale 3D sémantisée de la scène et d'outils d'analyse de cette dernière pour connecter le LLM au monde réel.

En raison de cet aspect interdisciplinaire, ce stage impliquera à la fois un laboratoire de vision par ordinateur (LVML) et un laboratoire de robotique (LCSR) du CEA.

L'étudiant aura pour charge de :

- Prendre en main les représentations neurales 3D développées au laboratoire LVML et mettre en place les outils d'analyse adaptés à la tâche de navigation (recherche de la destination, détection des obstacles...).
- Mettre en place un LLM en mode REACT pour analyser la requête utilisateur, réaliser la planification de tâches de navigation et invoquer les outils nécessaires (analyse de scène 3D, calcul et suivi de trajectoire...).
- Tester et évaluer la méthode dans un simulateur (eg. Habitat)
- Mettre en place le démonstrateur sur une plateforme robotique réelle (intégration des algorithmes de localisation, de contrôle-commande, etc., dans le framework robotique ROS).

[1] Liu, R., Wang, W., & Yang, Y. (2024). Volumetric Environment Representation for Vision-Language Navigation.

In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16317-16328).

[2] Wang, Z., Li, X., Yang, J., Liu, Y., Hu, J., Jiang, M., & Jiang, S. (2024). Lookahead Exploration with Neural Radiance

Representation for Continuous Vision-Language Navigation. In Proceedings of the IEEE/CVF Conference on Computer

Vision and Pattern Recognition (pp. 13753-13762).

[3] Gu, J., Stefani, E., Wu, Q., Thomason, J., & Wang, X. E. (2022). Vision-and-language navigation: A survey of tasks,

methods, and future directions. arXiv preprint arXiv:2203.12667.

Compétences développées au cours du stage

Ce stage permettra à l'étudiant de découvrir le domaine de l'intelligence artificielle pour la robotique,

domaine à l'intersection des dernières avancées en terme de reconstruction 3D (Neural Fields, segmentation en vocabulaire ouvert...), du contrôle robotique, et des modèles de langage.

L'étudiant découvrira aussi les aspects plus appliqués, avec l'usage de framework largement employés dans la communauté (ROS, NeRFStudio...). L'étudiant aura l'opportunité de travailler avec une équipe de chercheurs seniors mais aussi de doctorants.

Compétences souhaitées

Le candidat devra disposer d'une bonne maîtrise de python et des connaissances en robotique.

Une expérience sur Pytorch et/ou ROS sera appréciée.

Informations générales

Formation / Niveau d'étude Ingénieur, Master 2 / Bac+5

Possibilité poursuite Oui, en thèse ou CDD selon profil.

Durée 6 mois

Lieu Palaiseau (91) – Centre d'intégration de Nano-INNOV

Indemnités de stage

Entre 700 € et 1400 € suivant formation.

Aide au logement / transport / restauration.

Candidatures

- 
- 
- 
- 

Joindre CV + lettre de motivation à [steve.bourgeois@cea.fr](mailto:steve.bourgeois@cea.fr) avec le nom du stage auquel vous postulez

Ne pas hésiter à détailler les projets ou cours auxquels vous avez participé

Indiquer les dates de début/fin de stage envisagées.

Ce stage pourra prendre une orientation recherche ou industrie en fonction du profil du candidat

#### CEA Tech LIST

Les activités de recherche du CEA Tech LIST sont centrées sur les systèmes à logiciel prépondérant. Ces activités s'articulent autour de trois thématiques: les Systèmes Embarqués (architectures et conception de systèmes, méthodes et outils pour la sûreté des logiciels et des systèmes, systèmes de vision intelligents), les Systèmes Interactifs (ingénierie de la connaissance, robotique, réalité virtuelle et interfaces sensorielles) et les Capteurs et le traitement du signal (instrumentation et métrologie des rayonnements ionisants, capteurs à fibre optique, contrôle non destructif).

Le CEA Tech LIST a de nombreux partenariats avec les grands acteurs industriels du nucléaire, de l'automobile, de l'aéronautique, de la défense et du médical pour étudier et développer des solutions innovantes adaptées à leurs besoins. Il réalise une recherche qui va du concept de système jusqu'au démonstrateur, contribuant au transfert de technologies et à l'innovation par l'émergence de nouvelles entreprises.

#### Laboratoire Vision pour la Modélisation et la Localisation (LVML)

Laboratoire Vision pour la Modélisation et la Localisation (LVML) du CEA Tech LIST mène des recherches en vision par ordinateur et intelligence artificielle. Nous adressons en particulier les problématiques suivantes :

- Géolocalisation et cartographie d'environnement par vision et fusion de capteurs (robotique mobile, drones...)
- Systèmes et de vision pour la robotique : préhension, manipulation, assemblage d'objets...
- Contrôle de conformité, détection de défauts géométriques, colorimétriques, etc...
- Analyses hyperspectrales : détection de matériaux, tri, ....
- Correction, amélioration d'images et vidéos (superrésolution, upframing, ...)
- Compression de réseaux de neurones
- ...

#### Laboratoire Contrôle et Supervision Robotique (LCSR)

Le Laboratoire Contrôle et Supervision Robotique (LCSR) du CEA Tech LIST travaille à une interaction optimale dans un environnement où coexistent Hommes et Robots Mobiles. C'est dans

ce contexte que s'inscriront les travaux de ce stage, au regard de l'état de l'existant et du savoir-

faire du laboratoire, développé notamment lors de projets industriels de transitique et de projets

de recherche de conduite de véhicules. En particulier, le LCSR proposera l'exploitation d'une application logicielle de navigation autonome d'un robot mobile dans le cadre de ce stage.

## Proposition de stage 2025

### Exploitation d'un grand modèle de langage et d'une reconstruction 3D neurale sémantique pour la navigation autonome d'un robot

---

Contact : [steve.bourgeois@cea.fr](mailto:steve.bourgeois@cea.fr)

#### Contexte du stage

Si la robotique autonome connaît actuellement de grandes avancées, les intelligences artificielles qui guident ces robots se limitent généralement à une planification à court terme, sur l'environnement immédiat du robot, et pour des domaines de requêtes prédéfinies. Pour aller vers plus de polyvalence, il est nécessaire de développer des Intelligences Artificielles (IA) capable de planifier, sur la base d'une requête en langage naturel, une série d'actions sur des horizons temporels lointain, impliquant des lieux ou des éléments qu'ils ont observés par le passé mais qui sont hors du champ de perception actuel du robot.

Or, l'essor récent des grands modèles de langage (LLM) a démontré la capacité de ces IA à exploiter leur connaissance générale du monde pour planifier des séries d'actions (chaînes de pensées ou *chain-of-thought*) sur des domaines variés. Pour exploiter ces connaissances générales dans le cas d'un robot spécifique dans un environnement spécifique, il devient nécessaire de connecter le LLM aux capacités du robot et au monde réel qui l'entoure (encrage ou *grounding*). Pour tirer parti au maximum des capacités de planification à horizon lointain des LLM, cette connexion ne doit pas se limiter au seul environnement perceptible à l'instant courant par le robot, mais à une représentation 3D de l'environnement qu'il a pu précédemment exploré.

#### Objectifs du stage

Dans le cadre de ce stage, nous proposons d'étudier le domaine du *Vision-Language Navigation* [1,2,3], domaine de recherche interdisciplinaire à la frontière du traitement naturel du langage, de la vision par ordinateur et de la robotique. L'étudiant aura pour objectif de mettre en place un démonstrateur de navigation robotique autonome exploitant un LLM pour la compréhension de la requête utilisateur et la planification d'actions, ainsi qu'une représentation neurale 3D sémantisée de la scène et d'outils d'analyse de cette dernière pour connecter le LLM au monde réel.

En raison de cet aspect interdisciplinaire, ce stage impliquera à la fois un laboratoire de vision par ordinateur (LVML) et un laboratoire de robotique (LCSR) du CEA.

L'étudiant aura pour charge de :

- Prendre en main les représentations neurales 3D développées au laboratoire LVML et mettre en place les outils d'analyse adaptés à la tâche de navigation (recherche de la destination, détection des obstacles...).
- Mettre en place un LLM en mode REACT pour analyser la requête utilisateur, réaliser la planification de tâches de navigation et invoquer les outils nécessaires (analyse de scène 3D, calcul et suivi de trajectoire...).
- Tester et évaluer la méthode dans un simulateur (eg. Habitat)
- Mettre en place le démonstrateur sur une plateforme robotique réelle (intégration des algorithmes de localisation, de contrôle-commande, etc., dans le framework robotique ROS).

- [1] Liu, R., Wang, W., & Yang, Y. (2024). Volumetric Environment Representation for Vision-Language Navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16317-16328).
- [2] Wang, Z., Li, X., Yang, J., Liu, Y., Hu, J., Jiang, M., & Jiang, S. (2024). Lookahead Exploration with Neural Radiance Representation for Continuous Vision-Language Navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13753-13762).
- [3] Gu, J., Stefani, E., Wu, Q., Thomason, J., & Wang, X. E. (2022). Vision-and-language navigation: A survey of tasks, methods, and future directions. *arXiv preprint arXiv:2203.12667*.

## Compétences développées au cours du stage

Ce stage permettra à l'étudiant de découvrir le domaine de l'intelligence artificielle pour la robotique, domaine à l'intersection des dernières avancées en terme de reconstruction 3D (Neural Fields, segmentation en vocabulaire ouvert...), du contrôle robotique, et des modèles de langage. L'étudiant découvrira aussi les aspects plus appliqués, avec l'usage de framework largement employés dans la communauté (ROS, NeRFStudio...). L'étudiant aura l'opportunité de travailler avec une équipe de chercheurs seniors mais aussi de doctorants.

## Compétences souhaitées

Le candidat devra disposer d'une bonne maîtrise de python et des connaissances en robotique. Une expérience sur Pytorch et/ou ROS sera appréciée.

## Informations générales

<b>Formation / Niveau d'étude</b>	Ingénieur, Master 2 / Bac+5
<b>Possibilité poursuite</b>	Oui, en thèse ou CDD selon profil.
<b>Durée</b>	6 mois
<b>Lieu</b>	Palaiseau (91) – Centre d'intégration de Nano-INNOV
<b>Indemnités de stage</b>	Entre 700 € et 1400 € suivant formation. Aide au logement / transport / restauration.

## Candidatures

- Joindre CV + lettre de motivation à [steve.bourgeois@cea.fr](mailto:steve.bourgeois@cea.fr) avec le nom du stage auquel vous postulez
- Ne pas hésiter à détailler les projets ou cours auxquels vous avez participé
- Indiquer les dates de début/fin de stage envisagées.
- Ce stage pourra prendre une orientation recherche ou industrie en fonction du profil du candidat

## **CEA Tech LIST**

Les activités de recherche du CEA Tech LIST sont centrées sur les systèmes à logiciel prépondérant. Ces activités s'articulent autour de trois thématiques: les Systèmes Embarqués (architectures et conception de systèmes, méthodes et outils pour la sûreté des logiciels et des systèmes, systèmes de vision intelligents), les Systèmes Interactifs (ingénierie de la connaissance, robotique, réalité virtuelle et interfaces sensorielles) et les Capteurs et le traitement du signal (instrumentation et métrologie des rayonnements ionisants, capteurs à fibre optique, contrôle non destructif).

Le CEA Tech LIST a de nombreux partenariats avec les grands acteurs industriels du nucléaire, de l'automobile, de l'aéronautique, de la défense et du médical pour étudier et développer des solutions innovantes adaptées à leurs besoins. Il réalise une recherche qui va du concept de système jusqu'au démonstrateur, contribuant au transfert de technologies et à l'innovation par l'émergence de nouvelles entreprises.

## **Laboratoire Vision pour la Modélisation et la Localisation (LVML)**

Laboratoire Vision pour la Modélisation et la Localisation (LVML) du CEA Tech LIST mène des recherches en vision par ordinateur et intelligence artificielle. Nous adressons en particulier les problématiques suivantes :

- Géolocalisation et cartographie d'environnement par vision et fusion de capteurs (robotique mobile, drones...)
- Systèmes et de vision pour la robotique : préhension, manipulation, assemblage d'objets...
- Contrôle de conformité, détection de défauts géométriques, colorimétriques, etc...
- Analyses hyperspectrales : détection de matériaux, tri, ....
- Correction, amélioration d'images et vidéos (superrésolution, upframing, ...)
- Compression de réseaux de neurones
- ...

## **Laboratoire Contrôle et Supervision Robotique (LCSR)**

Le Laboratoire Contrôle et Supervision Robotique (LCSR) du CEA Tech LIST travaille à une interaction optimale dans un environnement où coexistent Hommes et Robots Mobiles. C'est dans ce contexte que s'inscriront les travaux de ce stage, au regard de l'état de l'existant et du savoir-faire du laboratoire, développé notamment lors de projets industriels de transique et de projets de recherche de conduite de véhicules. En particulier, le LCSR proposera l'exploitation d'une application logicielle de navigation autonome d'un robot mobile dans le cadre de ce stage.