



## Offre de Stage Génération de données continues et conditionnelles : application aux séries temporelles

**Mots clefs** Séries temporelles, génération de données, apprentissage automatique,

**Sujet du stage** La génération de données synthétiques « en continu » est un challenge important pour différentes applications telles que la simulation ou la prédiction, et dans de nombreux domaines en Industrie 4.0 comme en Santé numérique. Cette génération doit en outre être *fidèle* autant que possible aux données réelles ce qui complexifie ce challenge.

Différentes méthodes ont été développées dans ce cadre et pour ces différents types de données. Dans ce stage, même si une compréhension des grandes tendances d'approches proposées indépendamment du type de données est primordiale, un focus particulier sera apporté sur la génération de série temporelle.

Parmi les approches pour ce type de données nous pouvons citer notamment les GAN mais également les approches plus récentes basées sur les modèles LLM. Cependant, selon les contextes applicatifs, cette génération doit être conditionnée par des « événements/contraintes ». En effet, ces événements impliquent une adaptation au fur et à mesure du temps des données générées (exemple : un changement météo qui aura un impact sur la prédiction de la consommation énergétique). Différentes approches ont été développées pour prendre en compte ces conditions (instantané, fenêtre temporelle, contraintes. . .) dans la génération de données synthétiques.

Les objectifs du stages sont multiples :

1. Construire un état de l'art général des approches récentes [1] sur le domaine de la génération de données conditionnelles en faisant un focus particulier sur la génération conditionnelle de séries temporelles univariées (TSLLM, TSFM) [2, 4, 5];
2. Sur la base de cette étude, au moins deux modèles seront sélectionnés et implantés pour réaliser différentes versions d'un générateur de série temporelle conditionnel. Pour cela l'apprentissage sera réalisé sur la base de données type « données physiologiques » disponibles. Les données issues notamment du jeu de données MOVER[3] pourront être mobilisées.;
3. Enfin, suite à la définition d'un protocole expérimental et de l'identification de métriques adaptées, une comparaison objective de ces modèles sera proposée ;

**Profil des candidat-e-s** Étudiant-e en cycle master ou ingénieur, en Informatique ou mention similaires.

**Capacités attendues** : En complément du niveau de formation défini dans le profil, les éléments suivants sont souhaitables dans le profil des candidats :

- Capacité de réflexion et d'analyse d'un problème, définition d'un protocole et mise en place d'expérimentations ;
- Rigueur en programmation (Python) et capacité à générer de la documentation, avec usage des outils standard (Git, Doctest, Sphinx ou équivalents) ;
- Maîtrise des outils et techniques de l'apprentissage automatique ;
- Capacités de synthèse d'information et à rendre compte de son travail régulièrement ;
- Capacité à produire efficacement des documents de communication scientifique.

**Langues** : Maîtrise du Français (Niveau A2), conversationnel et écrit ; Maîtrise de l'anglais scientifique, écrit. **Les certifications attestant de votre niveau en langues sont vivement encouragées.**

**Conditions de réalisation du stage** Le stage sera réalisé au sein de l'**Institut de recherche en Informatique de Toulouse** (IRIT) sur son site de Castres (Tarn) où la thématique **Santé-Numérique** est une priorité stratégique du laboratoire et du territoire

**Equipe de recherche** : SIG - Systèmes d'Informations Généralisés

**Dates** : à partir du 1<sup>er</sup> mars 2026 (dates souples)

**Durée** : 5 mois - Rémunération au tarif légal en vigueur

**Lieu du réalisation** : IUT de Castres, 5 allée du Martinet 81100 CASTRES

**Modalité** : Présentiel sur le site de Castres exclusivement.

#### Important

L'IRIT est soumis au régime des **zones à régime restrictif (ZRR)** dans le cadre de la *protection du potentiel scientifique et technique national*. A ce titre, le processus de recrutement peut-être allongé le temps que votre profil soit validés par les autorités compétentes.

**Processus de recrutement** Envoi par courrier électronique, aux coordonnées ci-dessous, d'un dossier de candidature comprenant :

- un *Curriculum Vitae* ;
- une **lettre de motivation**<sup>1</sup>, présentant votre parcours, votre projet professionnel, les raisons expliquant votre intérêt pour le sujet ;
- un **portfolio** présentant vos réalisations en informatique et/ou programmation (par exemple un profil *github* ou équivalent).

A l'issue de votre candidature un ou plusieurs entretiens en visioconférence pourront vous être proposés. Ces entretiens peuvent impliquer des tests techniques. **Toutes les candidatures sérieuses seront considérées.**

**Date limite de candidature** : 15 Février 2026

#### Contact

**Dr Hugo BOISAUBERT** (hugo.boisaubert@irit.fr)

**Pr Max CHEVALIER** (max.chevalier@irit.fr)

IRIT - Equipe SIG (Systèmes d'Informations Généralisés)

IUT de Castres - 5 allée du Martinet - 81100 CASTRES

## Références

- [1] J. Ho and T. Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv :2207.12598*, 2022.
- [2] X. Liu, T. Aksu, J. Liu, Q. Wen, Y. Liang, C. Xiong, S. Savarese, D. Sahoo, J. Li, and C. Liu. Empowering time series analysis with synthetic data : A survey and outlook in the era of foundation models. *arXiv preprint arXiv :2503.11411*, 2025.
- [3] M. Samad, M. Angel, J. Rinehart, Y. Kanomata, P. Baldi, and M. Cannesson. Medical informatics operating room vitals and events repository (mover) : a public-access operating room database. *JAMIA open*, 6(4) :ooad084, 2023.
- [4] A. Shankar, L. Chen, A. van Deursen, and R. Hai. Wavestitch : Flexible and fast conditional time series generation with diffusion models. *Proceedings of the ACM on Management of Data*, 3(6) :1–25, 2025.
- [5] L. Shen and J. Kwok. Non-autoregressive conditional diffusion models for time series prediction. In *International Conference on Machine Learning*, pages 31016–31029. PMLR, 2023.

---

1. Les lettres générées par un outil d'Inférence Automatique (IA) sont à proscrire. C'est votre histoire et vos projets qui nous intéressent.