RESEARCH INTERNSHIP PROPOSAL

Title	Efficient RL algorithms for online and continual dialogue policy learning
Advisor	Prof. Fabrice Lefèvre

Topic: Artificial Intelligence, machine Learning, Reinforcement Learning, Deep Learning.
City and country: Avignon, France
Team or project in the lab: LIA – Vocal Interactions Team
Name and mail of the advisor: Prof. Fabrice Lefèvre (@univ-avignon.fr)
Name and mail of the head of the department: Prof. Yannick Estève (@univ-avignon.fr)

General presentation of the project
Most RL policy learning algorithms are exclusively oriented towards convergence to an optimal solution. Yet in many cases dynamics of the environment might not be totally stationary. For instance in Natural Language Processing, and Dialogue Systems in particular, users' preferences, abilities etc may greatly evolve over time (in relation to increased experience and familiarity with the interface). In such cases an initial one-for-all training is clearly sub-optimal and online and continual learning should be used instead. Many solutions have been proposed recently to improve stochastic policy learning but not necessarily implemented and tested in an NLP context such as Dialogue Modeling.

Objective of the internship
The objective of the project will be to study the possibility to address optimality efficiency of deep learning for RL (as instantiated in (double)-DQN [1], Soft Actor-Critic [2] and others algorithms [3]) in the context of data sparsity (sample efficiency) and online dynamic learning. Various tasks could serve as testbed during the project, yet the main overall application foreseen is the training of a vocal interaction dialogue manager, as in [4]. Showcase of the techniques studied and proposed could be done in any language (Python, R, Matlab etc).

Bibliographic references
[1] D. Silver, A. Huang, C. Maddison, *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529,** 484–489 (2016).
[2] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 5, pp. 2976–2989, 2018.
[3] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A Brief Survey of Deep Reinforcement Learning," *IEEE SIGNAL Process. Mag. Spec. ISSUE Deep Learn. IMAGE Underst.*, Aug. 2017.
[4] E. Ferreira and F. Lefèvre, "Reinforcement-learning based dialogue system for human-robot interactions with socially-inspired rewards," *Comput. Speech Lang. Spec. issue Speech Lang. Interact. Robot.*, vol. 34, no. 1, pp. 256–274, 2015.

Expected ability of the student
Interest in mathematical foundations of machine learning applied to NLP. Good skills in math (theory and programming). Good command of programming environments (OS, CLI, SDK, virtual env etc).

Duration	6 months
Remuneration	Around 529€ / month
Topics	Artificial Intelligence, Machine Learning, Reinforcement Learning, Deep Learning.