

Etablissement **Université de Lorraine**

École doctorale **IAEM - INFORMATIQUE - AUTOMATIQUE - ELECTRONIQUE - ELECTROTECHNIQUE - MATHEMATIQUES**

Spécialité **Informatique**

Unité de recherche **LORIA - Laboratoire Lorrain de Recherche en Informatique et ses Applications**

Encadrement de la thèse **Maxime AMBLARD**

Financement du 01-10-2025 au 30-09-2028 *origine* **ENACT + PEPR Santé numérique** *Employeur* **UNIVERSITE DE LORRAINE**

Début de la thèse le **1 octobre 2025**

Date limite de candidature (à 23h59) **24 avril 2025**

Mots clés - Keywords

TAL, données, santé mentale, LLM, discours, sémantique

NLP, Data, Mental Health, LLM, discourse, semantics

Description de la problématique de recherche - Project description

Le projet s'intéresse au développement de solution logicielle pour le diagnostic précoce, le diagnostic et le suivi de patient souffrant de pathologie mentale. Pour cela, la thèse consistera dans un premier temps à recueillir des données dans le milieu hospitalier, puis à proposer des modèles de langues (LLM) adaptés aux spécificités de ces données et enfin proposer un transfert des résultats vers les cliniciens pour une prise en charge améliorées des patients

Le projet s'inscrit d'une part dans l'axe sur la santé autonome du PEPR Santé Numérique et d'autre part dans la collaboration avec l'hôpital de Nice dans la suite du projet MePheSto, projet Inria-DFKI.

L'objectif global du projet est de développer un cadre/une base méthodologique pour la validation scientifique des phénotypes numériques pour les troubles psychiatriques basés sur des données multimodales, intégrant la parole, la vidéo et les signaux biologiques provenant d'interactions sociales cliniques. La collaboration avec le partenaire hospitalier à Nice reste ouverte car plusieurs perspectives sur des patients non diagnostiqués sont ouvertes.

Sur la base de biomarqueurs captées en situation réelle (voie, comportement oculaire, discours), une étape sera de définir automatiquement des élément caractéristiques pourtant être utilisés pour définir des traitements personnalisés plus efficaces, plus acceptables et plus fiables sur une longue période, tout en réduisant le risque d'effets secondaires indésirables. Ces enjeux sont particulièrement utiles dans le cas des troubles du sommeil, des addictions et des troubles mentaux, pathologies où l'évolution de la maladie peut varier significativement d'un patient à l'autre, car sur des affections de longue durée qui nécessitent souvent une prise en charge médicale continue (réévaluation des symptômes et des traitements, réévaluation paraclinique), nécessitant un traitement de longue durée avec des effets secondaires potentiellement graves ou un coût élevé et pouvant engager le pronostic vital des patients .

Dans l'ensemble, les thérapeutiques intégratives basées sur des biomarqueurs pronostiques ajustables sont un outil utile pour optimiser le traitement et améliorer les résultats des soins pour les patients souffrant de troubles. La création de ces nouvelles interventions confirmera la valeur ajoutée de la combinaison de biomarqueurs physiologiques (par exemple, vocaux) et comportementaux (par exemple la capacité de dialogue).

Le travail du doctorant sera de participer au recueil des données auprès des patients pour lesquels la contractualisation sera effective, à commencer par celui de Nice. Dans le même temps, le travail du doctorant le conduira à utiliser les outils de traitement de la langue du type LLM pour produire des représentations sémantiques des interactions. A partir de ces représentation, il aura pour mission de mettre au point une solution algorithmique pour identifier les caractéristiques de la défaillance dans le discours des patients.

Le partenaire PEPR Santé numérique sera très directement engagé dans la thèse avec un co-financement. Les problématiques du projet et de sante autonome restent distinctes et suffisamment proches pour que le travail doctoral bénéficie des deux dynamiques.

L'objectif est concrétiser un outil l'analyse de données spécifiques que sont des entretiens avec des patients souffrant de pathologie mentale pour le transférer aux experts cliniciens pour qui l'impact du projet serait très significatif. En effet, parvenir à proposer des

éléments théoriques permettant d'expliquer le dysfonctionnement cognitif, et rendre tangible l'analyse de ces éléments par des outils automatiques présente un fort potentiel tant pour le diagnostic précoce, que le diagnostic au moment des crises, que le suivi au long cours des patients. Chaque prise en charge accélérée a un impact direct sur la santé du patient, son intégration sociale, et sur la vie des aidant du patient.

The project concerns the development of a software solution for the early diagnosis, diagnosis and follow-up of patients suffering from mental pathology. To this end, the thesis will first involve collecting data in the hospital environment, then proposing language models (LLM) adapted to the specificities of this data, and finally proposing a transfer of results to clinicians for improved patient management.

The project is part of the autonomous health axis of the PEPR Santé Numérique program, and is also part of the collaboration with Nice Hospital as a follow-up to the Inria-DFKI MePheSto project.

The overall aim of the project is to develop a framework/methodological basis for the scientific validation of digital phenotypes for psychiatric disorders based on multimodal data, integrating speech, video and biological signals from clinical social interactions. Collaboration with the hospital partner in Nice remains open, as several perspectives on undiagnosed patients are open.

Based on biomarkers captured in real-life situations (pathway, eye behavior, speech), one step will be to automatically define characteristic elements that can be used to define personalized treatments that are more effective, more acceptable and more reliable over a long period, while reducing the risk of undesirable side-effects. These issues are particularly useful in the case of sleep disorders, addictions and mental disorders, pathologies where the evolution of the disease can vary significantly from one patient to the next, as they are long-term conditions that often require ongoing medical management (re-evaluation of symptoms and treatments, paraclinical re-evaluation), necessitating long-term treatment with potentially serious side-effects or a high cost that can engage the patient's vital prognosis.

Overall, integrative therapies based on adjustable prognostic biomarkers are a useful tool for optimizing treatment and improving care outcomes for patients suffering from disorders. The creation of these new interventions will confirm the added value of combining physiological (e.g. vocal) and behavioral (e.g. dialogue ability) biomarkers.

The doctoral student's job will be to help collect data from the patients for whom the contract will be signed, starting with the one in Nice. At the same time, the PhD student's work will involve using LLM-type language processing tools to produce semantic representations of interactions. Based on these representations, he will develop an algorithmic solution to identify failure characteristics in patient discourse.

The PEPR partner Digital Health will be directly involved in the thesis, with co-financing. The issues involved in the project and in autonom health remain distinct, but are sufficiently close for the doctoral work to benefit from both dynamics.

The aim is to develop a tool for the analysis of specific data, such as interviews with patients suffering from mental pathology, and to transfer it to clinical experts, for whom the impact of the project would be highly significant. Indeed, succeeding in proposing theoretical elements to explain cognitive dysfunction, and making the analysis of these elements tangible through automatic tools, has great potential for early diagnosis, diagnosis at the time of crises, and long-term follow-up of patients. Each accelerated treatment has a direct impact on the patient's health and social integration, and on the lives of the patient's carers.

Thématique / Domaine / Contexte

Traitement Automatique des Langues et Santé
NLP and Health

Informatique Computer Science

Développement d'un analyseur automatique de données de santé pour l'identification précoce de pathologie mentale.
Development of an automatic health data analyzer for the early identification of mental pathology.

Références bibliographiques

- [Amb-14] M. Amblard, S. Pogodalla – Modelling the Dynamic Effects of Discourse: Principles and Frameworks, in Manuel Rebuschi and Martine Batt and Gerhard Heinzmann and Franck Lhoreau and Michel Musiol and Alain Trognon. Dialogue, Rationality, and Formalism, 3, Springer, pp.247-282, Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics. Logic, Argumentation & Reasoning, 2014
- [Amb-15] M. Amblard, K. Fort, C. Demily, F. Nicolas, M. Michel - Analyse lexicale outillée de la parole transcrite de patients schizophrènes, Natural Language Processing and Cognition, Traitement Automatique des Langues, vol 55-3 2015
- [Are-20] A.C. Arevian, D. Bone, N. Malandrakis, V.R. Martinez, K.B. Wells, D.J. Miklowitz - Clinical state tracking in serious mental illness through computational analysis of speech. PLoS ONE 15(1): e0225695, 2020
- [Bar-18] Barnett I, Torous J, Staples P, Sandoval L. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology. 2018.
- [Bar-19] K. Bar, V. Zilberstein, I. Ziv, H. Baram, N. Dershowitz, S. Itzikowitz, E.V. Harel Semantic Characteristics of Schizophrenic Speech Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, Minneapolis, Minnesota, June 2019
- [Boc-17] C. Bocéréan, M. Musiol - Verbal interaction structures and repetition's functions: a comparison of exchanges between adults and severely disabled adolescents or young children. Discourse Processes. vol 54, N°7, 524-544, 2017
- [Bre-17] E. Breitholtz, C. Howes - Dialogical reasoning in patients with schizophrenia - Invited Talk. In (In)Coherence of Discourse 4.

Nancy, France, 2017

[Hin-15] W. Hinzen, J. Rosselló - The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. *Front. Psychol.* 6, 971, 2015

[Hol-18] Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crian, G. C., Pinte, C.-M., & Palade, V. (2018). Interactive machine learning: Experimental evidence for the human in the algorithmic loop. *Applied Intelligence*, 1–14.

[kup-10a] G. R. Kuperberg - Language in schizophrenia Part 1: An Introduction, *Lang Linguist Compass*. 4(8): 576–589, 2010

[kup-10b] G. R. Kuperberg - Language in schizophrenia Part 2: What can psycholinguistics bring to the study of schizophrenia...and vice versa? *Lang Linguist Compass*, 4(8): 590–604, 2010

[Lav-19] M. Lavelle, C. Howes - Turn-taking and other interactional issues in patients with schizophrenia. In *Workshop on Formal Approaches to (In)coherence and Dynamics in Dialogue*, Goteborg, October, 2019

[Pad-16] S. Padroni, C. Demily, N. Franck, C. Bocéréan, C. Hoffmann, M. Musiol - Ajustement comportemental et mouvements de saccades oculaires dans la schizophrénie. *L'Evolution Psychiatrique*, 81/2, 365-379, 2016

[Reb-14] M. Rebuschi, M. Amblard, M. Musiol - Using SDRT to analyze pathological conversations. Logicality, rationality and pragmatic deviances. In M. Rebuschi, M. Batt, G. Heinzmann, F. Lihoreau, M. Musiol & A. Trognon (Editors). *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics. (Dialogue, Rationality, and Formalism)*. Coll “Logic, Argumentation and Reasoning” (Interdisciplinary Perspectives from the Humanities and Social Sciences) volume 3. Springer Cham Heidelberg New York Dordrecht London, Ch 15, 345-371, 2014

[Tso-12] I. F. Tso, M. Mui, S. F. Taylor, P. J. Deldin - Eye-contact perception in schizophrenia: Relationship with symptoms and socioemotional functioning. *Journal of abnormal psychology*, 121(3), 616. 2012

Précisions sur l'encadrement - Details on the thesis supervision

Suivi hebdomadaire du travail par l'encadrant

Cours de formation complémentaires obligatoire organisés par l'Ecole Doctorale

Présentation des travaux en séminaires et conférences

Conditions scientifiques matérielles et financières du projet de recherche

Cofinancement ENACT-PEPR Santé numérique

Co-funding ENACT-PEPR Santé numérique

Objectifs de valorisation des travaux de recherche du doctorant : diffusion, publication et confidentialité, droit à la propriété intellectuelle,...

Publication dans des conférences et revues internationales du domaine. Transfert technologique

Publication in international conferences and journals of the field. Technology transfer

Profil et compétences recherchées - Profile and skills required

Master en NLP , en informatique ou dans un domaine connexe.

Maîtrise des langages de programmation (Python) et des bonnes pratiques de codage

Compétences en conception d'algorithmes

Connaissance en psychologie (un diplôme serait apprécié)

Expérience en apprentissage profond

Maîtrise de la logique, des graphes et des représentations sémantiques des énoncés en langue naturelle

Capacité à travailler de manière autonome et à travailler en équipe

Excellentes compétences en anglais, à l'oral et à l'écrit

Master's degree in NLP , Computer Science or a related master program

Proficiency in programming languages (Python) and good coding practices

Skills in algorithm design

Knowledge of psychology (a degree would be appreciated)

Experience in deep learning

Proficiency in logic, graphs and semantic representations of natural language utterances

Ability to work independently and also to work in a team

Excellent oral and written English skills