

- Faculté des sciences
- www.unine.ch/sciences

Modeling and Scaling of Generative AI Systems (3IN2079)

Filières concernées	Nombre d'heures	Validation	Crédits ECTS
Master en informatique	Cours: 2 ph Exercice: 2 ph	Voir ci-dessous	5

ph=période hebdomadaire, pg=période globale, j=jour, dj=demi-jour, h=heure, min=minute

Période d'enseignement:

- Semestre Automne

Equipe enseignante

Professeure responsable : Lydia Y. Chen

Équipe enseignante : professeure, assistant-es et/ou teaching assistants (TA) du cours.

Contenu

Ce cours présente les principaux modèles et techniques modernes de Generative AI (GAI), avec un accent particulier sur la compréhension des modèles, leur adaptation à des applications concrètes et l'optimisation de leur utilisation.

Le cours couvre notamment les modèles de diffusion, les VAE et GAN, les Large Language Models (LLM) et les Transformers, ainsi que le pré-entraînement, le fine-tuning supervisé, l'alignement et le reinforcement learning pour les LLM. Une attention particulière est accordée aux techniques permettant de construire des applications GAI efficaces, notamment le context engineering, l'évaluation et la robustesse des modèles, ainsi que l'optimisation de l'inférence à travers le test-time/inference-time compute, les KV caches et les positional embeddings.

Le cours est fortement orienté vers la pratique. Chaque séance est accompagnée, lorsque pertinent, d'un laboratoire permettant de mettre en œuvre les concepts étudiés. Le projet final consiste à développer une application GAI et à en améliorer les performances, la qualité ou l'efficacité.

Forme de l'évaluation

L'évaluation est continue et il n'y a pas d'examen final.

Quiz (36 %) : trois quiz sont organisés pendant le semestre (semaines 4, 9 et 12), chacun comptant pour 12 % de la note finale. Les quiz évaluent la compréhension des notions vues en cours et dans les laboratoires. Le format exact de chaque quiz sera annoncé lors du laboratoire correspondant. Il est nécessaire d'obtenir plus de 55 % à chaque quiz pour être éligible au projet final.

Projet de groupe (70 %) : les étudiant-es développent une application GAI et cherchent à en améliorer la qualité, l'efficacité ou les performances. Le projet comprend une proposition initiale, une réunion intermédiaire avec l'équipe enseignante, un rapport final et une présentation de 20 minutes.

Les laboratoires sont non notés, mais ils font partie intégrante de l'apprentissage et leur contenu peut être évalué dans les quiz.

Modalités de rattrapage

Il n'y a pas d'examen final à rattraper. En cas d'échec, les modalités de rattrapage sont définies conformément au règlement de la formation et de l'Université de Neuchâtel.

Documentation

Les références principales comprennent des ouvrages et ressources consacrés au deep learning, aux modèles génératifs et aux LLM, complétés par des articles scientifiques et des ressources techniques récentes.

Understanding Deep Learning — Simon J. D. Prince

Dive into Deep Learning — Alex Smola et al.

Build a Large Language Model (From Scratch) — Sebastian Raschka

Deep Learning — Ian Goodfellow, Yoshua Bengio et Aaron Courville

Hugging Face Transformers

Hugging Face Diffusers

Articles et ressources complémentaires fournis au cours.

Pré-requis

Une connaissance de base en machine learning et deep learning est fortement recommandée. Les étudiant-es devraient notamment être

URLs	1) https://github.com/lydiaYchen/26MSGAI.3IN2079
------	--

- Faculté des sciences
- www.unine.ch/sciences

Modeling and Scaling of Generative AI Systems (3IN2079)

familiarisé·es avec les notions fondamentales d'apprentissage supervisé, réseaux de neurones, optimisation et Python.

Il est possible de suivre le cours sans avoir une formation approfondie en machine learning. Toutefois, les étudiant·es dans ce cas doivent être prêt·es à fournir un effort important pour rattraper les prérequis. L'expérience montre qu'un manque de connaissances préalables peut rendre difficile le suivi du rythme du cours et des laboratoires.

Une expérience de programmation en Python est également fortement recommandée.

Forme de l'enseignement

Le cours combine cours magistraux, laboratoires pratiques et projet de groupe.

Les laboratoires proposent des exercices pratiques permettant de relier directement les concepts présentés en cours à leur implémentation et expérimentation en code. La première partie du laboratoire est consacrée au travail individuel ou en groupe sur les exercices. Dans la seconde partie, les TA répondent aux questions des étudiant·es et discutent les difficultés rencontrées.

La participation aux laboratoires est fortement encouragée. Ils constituent une composante importante de l'apprentissage et permettent de préparer les étudiant·es aux quiz ainsi qu'au projet final.