Module title	**Generative AI Models in Life Sciences**
Code	CO5
Degree Programme	Master of Science in Life Sciences (MSLS)
Group	Computation
Workload	3 ECTS (90 student working hours) Lesson contact: 32 hours, preparatory Self-study: 30 hours, post-school project Work: 28 hours
Module Coordinator	**Name**: Dr. Ahmad Aghaebrahimian **Phone**: +41 (0) 58 934 45 04 **Email**: agha@zhaw.ch **Address**: ZHAW Life Sciences und Facility Management, Schloss 4, 8820 Wädenswil
Lecturers	• Dr. Ahmad Aghaebrahimian, ZHAW • Simon Dürr, HES-SO
Entry requirements	• **Required**: Programming, Data Structures, and Algorithms module or equivalent • Basic understanding of Machine Learning
Learning outcomes and competences	After completing this module students will have a solid understanding of generative AI models (e.g. Large Language Models (LLMs) for Biomedical Natural Language Processing (NLP) and diffusion models) and their applications in Life Sciences. Precisely, they will be able to: • Explain generative AI models and their applications on an abstract level. • Work with, tune, and deploy generative models for text and other modalities • Apply, tune, and deploy generative models for one selective topic from a pool of subjects such as Literature-based Discovery (LBD) for drug discovery/repurposing, Electronic Health Record (EHR), patient review/sentiment analysis, regulatory/compliance, pharmaceutical process development, manufacturing support, Real-World Evidence (RWE) extraction and bioimage analysis.

Module contents	*Theme 1: (Ahmad Aghaebrahimian: 17 lessons)*
	• Introduction to Biomedical Natural Language Processing
	• Representation techniques for linguistic and biomedical string data.
	• Essential tasks in textual data analytics include Named Entity Recognition (NER), Named Entity Linking, and relation extraction.
	• Large Language Models (LLMs), and their applications in life sciences and biopharmaceutical research.
	  ○ Medical information accessibility with Retrieval Augmented Generation (RAG)
	  ○ Automatic patient report summarization
	
	Theme 2: (Ahmad Aghaebrahimian: 9 lessons)
	• Developing and evaluating Information Retrieval (IR) systems over research publications in the life sciences and biopharma domains
	• Developing and evaluating Question-Answering (QA) systems for large datasets of scientific literature, drug information, manufacturing information, and clinical trial records.
	• System evaluation using standard benchmarks.

	Theme 3: (Simon Dürr HES-SO: 16 lessons) • Types of generative models (autoregressive, autoencoder, GAN, diffusion) • Generative models for non-text data modalities (images, volumetric data, molecular structures) • Fine-tuning generative foundation models (e.g CLIP, DINO) for biological applications
Teaching / learning methods	This module consists of five sessions, all centrally administered on campus. Students are required to cover designated reading materials before school. Each session in the school consists of a morning lecture focused on theories, and an afternoon session focused on doing hands-on exercises supervised by the lecturers. Students propose a mini project up until the last session, discuss it in the last session, and deliver it accompanied by a short report two weeks after the school completion at the latest.
Assessment of learning outcome	– Entry exam on preparatory self-study materials (20%) – Delivering solutions for at least 60% of the exercises (30%) – Final assessment as a project work (50%)
Format	1-week winter school: preparatory self-study before school, 5 days of school sessions, two-week deadline for final project submission.
Timing of the module	Autumn semester, CW 04
Venue	Olten
Bibliography	Selected parts of the following resources will be used as the study material. **A:** Speech and Language Processing **B:** Slides **C:** Selected papers **D:** The Illustrated Stable Diffusion **E:** What are Diffusion Models?
Language	English
Links to other modules	-
Comments	-
Last Update	04.04.2026