

Lecture & Date	Topic	Presented papers
Introductory classes (Week 1-2)		
Lecture 1 01.09.2026	Introduction to the course	
Lecture 2 03.09.2026	Multilingual transformer models	
Lecture 3 08.09.2026	Introduction to Machine translation	
Lecture 4 10.09.2026	Introduction to Speech processing	
Review topics & papers		
Week 3		
Lecture 5 15.09.2026	Tokenization Lead discussion: TBD	Language Model Tokenizers Introduce Unfairness Between Languages Do All Languages Cost the Same? Tokenization in the Era of Commercial Language Models Zero-Shot Tokenizer Transfer Further reading: Julie Kallini's slides
Lecture 6 17.09.2026	Cross-lingual Transfer Lead discussion:	Extending Multilingual BERT to Low-Resource Languages Expanding Pretrained Models to Thousands More Languages via Lexicon-based Adaptation XLM-V: Overcoming the Vocabulary Bottleneck in Multilingual Masked Language Models Additional reading: Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of BERT Revisiting Machine Translation for Cross-lingual Classification
Week 4		
Lecture 7 22.09.2026	Sentence embedding Lead discussion:	Text Embeddings by Weakly-Supervised Contrastive Pre-training M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation Improving Text Embeddings with Large Language Models Additional reading: MMTEB: Massive Multilingual Text Embedding Benchmark

Lecture 8 24.09.2026	Machine Translation Lead discussion:	Scaling neural machine translation to 200 languages Multilingual Machine Translation with Open Large Language Models at Practical Scale: An Empirical Study NiuTrans.LMT: Toward Inclusive and Scalable Multilingual Machine Translation with LLMs Additional reading: X-ALMA: Plug & Play Modules and Adaptive Rejection for Quality Translation at Scale
Week 5		
Lecture 9 29.09.2026	Machine Translation Evaluation Lead discussion:	COMET: A Neural Framework for MT Evaluation MetricX-24: The Google Submission to the WMT 2024 Metrics Shared Task SSA-COMET: Do LLMs Outperform Learned Metrics in Evaluating MT for Under-Resourced African Languages? Additional reading: Large Language Models Are State-of-the-Art Evaluators of Translation Quality
Lecture 10 01.10.2026	Ethical data collection & participatory research Lead discussion:	Building Better: Avoiding Pitfalls in Developing Language Resources when Data is Scarce Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages SEACrowd: A Multilingual Multimodal Data Hub and Benchmark Suite for Southeast Asian Languages Additional reading: Global PIQA: Evaluating Commonsense Reasoning Across 100+ Languages and Cultures
Week 6		
Lecture 11 06.10.2026	Guest lecture	Prof. Skyler Wang (McGill & Handshake AI): Ethical data collection and Scaling NLP to several low-resource languages
Lecture 12 08.10.2026	LLM architecture Lead discussion:	EuroLLM-9B: Technical Report OLMo: Accelerating the Science of Language Models Apertus: Democratizing Open and Compliant LLMs for Global Language Environments Additional Reading: Olmo 3 TinyAya

Week 7		
13.10.2026	Reading week	
Lecture 13 15.10.2026	LID & Multilingual pre-training data Lead discussion:	Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets GlotLID: Language Identification for Low-Resource Languages FineWeb2: One Pipeline to Scale Them All — Adapting Pre-Training Data Processing to Every Language Additional Reading: MADLAD-400: A Multilingual And Document-Level Large Audited Dataset
Week 8		
Lecture 14 20.10.2026	LLM-as-a-Judge Lead discussion:	Lower-Resource, Higher Scores: Language Bias in LLM Evaluators Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena Are Large Language Model-based Evaluators the Solution to Scaling Up Multilingual Evaluation? Additional Reading: Challenges and Recommendations for LLMs-as-a-Judge in Multilingual Settings and Low-Resource Languages
Lecture 15 22.10.2026	LLM adaptation Lead discussion:	SEA-LION: Southeast Asian Languages in One Network AfriqueLLM: How Data Mixing and Model Architecture Impact Continued Pre-training for African Languages Instructing Large Language Models for Low-Resource Languages: A Systematic Study for Basque
Week 9		
Lecture 16 27.10.2026	LLM safety Lead discussion:	JailbreakBench: An Open Robustness Benchmark for Jailbreaking Large Language Models Multilingual Jailbreak Challenges in Large Language Models MPO: Multilingual Safety Alignment via Reward Gap Optimization Additional Reading: The State of Multilingual LLM Safety Research: From Measuring The Language Gap To Mitigating It All Languages Matter: On the Multilingual Safety of LLMs
Lecture 17 29.10.2026	Project presentations (5 minutes each)	All students presents their project (in group of one or two) - in person participation only

Week 10		
Lecture 18 03.11.2026	LLM reasoning Lead discussion:	Language Models are Multilingual Chain-of-Thought Reasoners Question Translation Training for Better Multilingual Reasoning Long Chain-of-Thought Reasoning Across Languages Additional Reading: PolyMath: Evaluating Mathematical Reasoning in Multilingual Contexts
Lecture 19 05.11.2026	Guest lecture	Dr. Julia Kreutzer (Cohere Labs): Multilingual reasoning
Advanced Topics		
Week 11		
Lecture 20 10.11.2026	Multimodal Multicultural LLMs	Multilingual Diversity Improves Vision-Language Representations BLEnD-Vis: Benchmarking Multimodal Cultural Understanding in Vision Language Models Jagle: Building a Large-Scale Japanese Multimodal Post-Training Dataset for Vision-Language Models Additional Reading: X-LLaVA: Optimizing Bilingual Large Vision-Language Alignment
Lecture 21 12.11.2026	Guest lecture	Dr. Min Ma (Google DeepMind): Speech processing
Week 12		
Lecture 23 17.11.2026	Multilingual Speech Representations	
Lecture 24 19.11.2026	Speech-to-Speech Translation	
Class Project Presentations		
Week 13		
Lecture 24 24.11.2026	AudioLLMs	
Lecture 25 26.11.2026	Final presentations (12 minutes talk + 3 minutes Q/A)	

Lecture 26 01.12.2026	Final presentations (12 minutes talk + 3 minutes Q/A)	
Lecture 27 03.12.2026	Final presentations (12 minutes talk + 3 minutes Q/A)	