# CAS Natural Language Processing Study Plan

j)ł

b UNIVERSITÄT BERN

D

15.09.2022 The CAS Natural Language Processing is a university study program leading to a "Certificate of Advanced Studies in Natural Language Processing" awarded by the University of Bern as laid out in the regulations for the Studiengänge in Extended Intelligence of 2021-12-09. 1. Course objectives About The field Natural Language Processing (NLP) has exploded in recent years. It belongs to both computational linguistics as its engineering domain and artificial intelligence as an increasingly important subdomain. In particular the applications based on deep neural networks have reached a performance level which cannot be ignored by any field that is processing natural languages. Machine translation, information extraction, question answering and nonsense natural language generation performed by NLP algorithms are already part of everyday life. However, the new semantic understanding capabilities seen in recent trained models not only expand the NLP application space enormously, but also represent possible first steps towards consciousness and mind outside human and animal nerve systems. The program targets practitioners who aim for an overview of the NLP domain with main focus on recent developments (deep learning models) and hands-on learning. Some programming experience or the willingness to learn basic scripting is a prerequisite. The CAS format is designed to align with the participants' main professional and study activities. The workload is on average about 20% of a full time position over a year. The teaching and learning approaches are team and discussion oriented, aimed at developing practical competency. All modules can also be attended online. Competence The course competencies are developed in six modules. At the end graduates objectives will: 1. have an overview of the NLP domain and common applications, 2. be able to perform relevant preprocessing tasks needed for advanced NLP, 3. be able to understand neural networks and practice them on own NLP applications, 4. be able to understand transformers and practice transfer learning with transformers for own applications,

- 5. know discussions related to philosophical and ethical aspects around NLP and artificial intelligence
- 6. be familiar with active research in the NLP domain.

### 2. Duration, outcomes and objectives of the modules

Duration and scope The CAS consists of about 18 days of courses (126 hours attendance) and 16 ECTS credit points (total effort about 480 hours) whereof 4 ECTS result from the CAS project work.

> The program consists of four modules in weekly blocks and two modules on eight to ten half days. All modules support distant learning.

Attendance at one introductory event *About the CAS Natural Language Processing* before course admission is mandatory.

## Module 1 NLP Fundamentals

In which linguistic and machine learning concepts are introduced, and an overview of the NLP field and common applications is given.

ECTS	2 ECTS credit points	Duration	4.5 days +	
	(incl. self-studies and		essay	
	project)		writing	
Assessment	Written essay	Req. attendance	80%	
Learning	Graduates will			
outcomes	• have an understanding of the field of linguistics and			
	study of languages			
	• understand different ta	sks of NLP and thei	r	
	requirements			
	• know most common NLP frameworks and tools			
	• know the history of NLP in computer science and the			
	humanities			
	• develop a first project involving NLP methods, which			
	can be used as foundation for the CAS project work			
	• be able to reflect upon the intertwining of NLP with			
	machine learning and artificial intelligence			
Learning	Basic linguistics			
objectives	Common NLP tasks			
	NLP frameworks			
	NLP history			
	NLP concept development			
Learning and	<ul> <li>Onsite classes with online</li> </ul>	line transmission		
teaching	• Lectures, self studies, team work, discussions			
methods	<ul> <li>Essay assessment</li> </ul>			
Prerequisites	Own laptop			
Language	English			

# Module 2 NLP Preprocessing and Basic Analysis

st.

In which we learn to perform basic preprocessing and analysis of natural languages.

ECTS	2 ECTS credit points	Duration	3 days +	
	(incl. self-studies and		project	
	project)	· · · · · · · · · · · · · · · · · · ·		
Assessment	Project presentation	Req. attendance	80%	
Learning	Graduates will/can			
outcomes	• perform basic preproce	ssing and segmentat	ion of text	
	for NLP purposes			
	• perform basic information	on extraction (know	v forms of	
	annotation and correspo	onding evaluation)		
	• perform basic critical te	ext analysis (epistem	nological	
	foundations and theoret	ical classification)		
Learning	Basic preprocessing techniques for text data:			
objectives	• cleaning spatial characters and numbers			
	• removing punctuation, misspells and contractions etc			
	Sequence creation:			
	• using tokenizers			
	padding sequences			
	• embedding enrichment			
	Information extraction and critical text analysis			
	Named Entity Recognition			
	Part-of -Speech Tagging			
Learning	<ul> <li>Onsite classes with online transmission</li> </ul>			
and teaching	• Lectures and inverted classroom with hands-on tutorials			
methods	Teamwork on projects			
	Project presentations with peer feedback			
Prerequisites	Previous modules			
Language	English			

## Module 3 NLP Neural Networks

In which we study and learn how neural networks work, are trained, tuned, assessed and applied for NLP tasks.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	3 days = 21 hrs attendance	
Assessment	Project presentation	Required	80%	
		attendance		
Learning	Graduates will/can			
outcomes	know different neural network architectures			
	• understand the role of the various network components			
	train neural networks			
	• evaluate the performance of neural network models			
	apply neural networks to NLP tasks			
	• understand the principles of recurrent neural networks			
Learning	• Neural network architectures and components for NLP			
objectives	Model training			

Language	English		
S			
Prerequisite	Previous modules		
	<ul> <li>Project presentations with peer feedback</li> </ul>		
methods	• Teamwork on project		
and teaching	• Lectures and inverted classroom with hands-on tutorials		
Learning	Onsite classes with online transmission		
	Long-term dependencies		
	analytics, text comprehension		
	• Model application to sequence-to-sequence tasks, text		
	• Transfer learning		
	Model performance		
	<ul> <li>Model predictions and analysis</li> </ul>		

## Module 4 NLP Transformers

In which we study transformers and learn why they have changed the NLP field.

ECTS	2 ECTS credit points	Duration	3  days = 21	
	(incl. self-studies and		hrs	
	project)		attendance	
Assessment	Project presentation	Req. attendance	80%	
Learning	Graduates will/can			
outcomes	• understand the inner we	orkings of the transf	former	
	architecture including t	he attention mechan	nism and	
	positional embeddings			
	• understand the differen	ces between autoreg	gressive and	
	autoencoding language	models		
	• know current limitations of transformer-based models			
	such as the input length			
	• apply a modern transformer-based pipeline to various			
	NLP tasks			
Learning	Attention mechanism			
objectives	• Positional embeddings	Positional embeddings		
	Autoregressive and Autoencoding Language Models			
	• BERT and GPT as examples			
	Encoders and Decoders			
	Current limitations			
Learning and	Onsite classes with online transmission			
teaching	• Lectures and inverted classroom with hands-on tutorials			
methods	Teamwork on project			
	<ul> <li>Project presentations w</li> </ul>	ith peer feedback		
Prerequisites	Previous modules			
Language	English			

# Module 5 NLP Philosophical and Ethical Aspects

In which we study and discuss ethical and philosophical aspects related to machines being capable of natural language processing

ECTS	2 ECTS credit points (incl. self-studies)	Duration	3  days = 21  hrs
Assessment	Written essay	Required attendance	80%
Learning outcomes	attendance         Graduates will/can         have an overview of the history and the philosophy of artificial intelligence         know philosophical and scientific presuppositions of artificial intelligence         can relate techniques of AI to well-known scientific methods         know main philosophical discussions on artificial intelligence         know main philosophical discussions on artificial intelligence         know main moral challenges related to artificial intelligence and can discuss solutions from the perspective of ethics		
Learning objectives	<ul> <li>Infaster best practices for ethics dealing with artificial intelligence</li> <li>Philosophical conceptions of AI, weak vs. strong AI</li> <li>Extended mind hypothesis</li> <li>Philosophical concepts of data and data analysis</li> <li>AI and scientific inference</li> <li>Ethical challenges due to AI in the light of ethical theories</li> <li>Machine ethics</li> </ul>		
Learning and teaching methods Prerequisites Language	<ul> <li>Onsite classes with online transmission</li> <li>Lectures, discussions and independent studies</li> <li>Essay writing</li> <li>Module 1</li> <li>English</li> </ul>		

# Module 6 NLP Frontier and Applications

Æ.

In which we use knowledge from the previous modules to focus on prominent NLP topics and applications with a feasibility study for our final CAS project.

ECTS	2 ECTS credit points (incl. self-studies and project)	Duration	3 days = 21 hrs attendance
Assessment	Project presentation	Req. attendance	80%
Learning outcomes	<ul> <li>Module graduates</li> <li>can design, formulate and communicate an NLP project</li> <li>know current research topics within the NLP field</li> <li>apply newer NLP approaches to own text</li> </ul>		
Learning objectives	<ul> <li>In addition to the final CAS project planning, objectives</li> <li>may be among others</li> <li>Information retrieval</li> <li>Search engines</li> <li>Multilingual NLP</li> <li>Domain-Specific NLP, e.g. legal NLP</li> </ul>		

Learning and	Onsite classes with online transmission		
teaching	• Lectures and inverted classroom with hands-on tutorials		
methods	• Teamwork on project		
	Project presentations with peer feedback		
Prerequisites	Previous modules		
Language	English		

CAS Project Work **CAS Project Work** 

In which participant teams perform and present a comprehensive Natural Language Processing project based on all six CAS modules.

ECTS	4 ECTS credit points	Duration	120 hours
Assessment	Exhibition work	Req. attendance	None
Learning	Graduates will be able to pe	rform and present a	complex
outcomes	Natural Language Processing project.		
Learning objectives	Selection of objectives from all modules. Furthermore, graduates will create an enticing presentation that reports on the practical work carried out, combined with comparison to similar state-of-the-art applications/proof-of- concepts.		
Teaching methods	Project work in teams with	remote supervision	
Prerequisite	Module 1 - 6		
S			
Language	English		

#### Assessment **3. Performance assessment**

The performance assessment includes 80% attendance and passed performance assessments from all modules and the CAS project.

- a) For Module 1 a written essay is assessed.
- b) For Module 2, 3, 4 and 6 project presentations are assessed.
- c) For Module 5 a written essay is assessed.
- d) For the CAS project a report, publication or an app with a presentation is assessed.

### 4. Final regulations

Entry into force The present plan shall enter into force on 01.11.2022.

15.09.2022 Released by the program management:

mister 1velle >

Prof. Dr. Christiane Tretter (Chair)

06.10.2022

Ż

The Dean

Prof. Dr. Marco Herwegh