CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Course Introduction

Raghava Mukkamala

Professor, Copenhagen Business School, Denmark
Email: rrm.digi@cbs.dk,

Sippo Rossi
Assistant Professor, Hanken School of Economics
Email: sippo.rossi@hanken.fi

PhD Course on Natural Language Processing and Large Language Models for Research Data Exploration and Analysis
Aalto University, Finland

EQUIS ACCREDITED AACSB ACCREDITED ASSOCIATION OF MBAS ACCREDITED CEMS PIM

About Myself Raghava

IT Development Experience

- 2000-2009: Software Programmer, software Consultant in Danish IT Industry (C++, C# and Java)
- 2003-2005: M S in Internet Technology from IT University of Copenhagen (ITU), Denmark.

Academic/Research Experience

- 2009-2012: PhD in Theoretical Computer Science & Formal Models from ITU, Denmark.

(PhD Topic: A Formal Model for Declarative Workflows: Dynamic Condition Response Graphs)

- 2012-2015: Postdoc at ITU on Software Verification and Domain-specific Modelling languages.
- 2015-2017: Assistant Professor at dept. of IT Management at Copenhagen Business School (CBS), Denmark.
- 2017: Tenured Associate Professor at the Department of Digitalization at CBS.
- 2018: Program Director for Masters in Business Administration and Data Science Programme at CBS.
- 2019: Director for Center for Business Data Analytics at the Department of Digitalization at CBS.
- 2025: Full Professorship at Copenhagen Business School, Denmark.
- Research: Natural Language Processing, Deep Learning,
    - Blockchain and Distributed Ledger Technologies
    - Cybersecurity and Cybersecurity Analytics
    - Quantum Computing and Programming Qbits

About Myself
Sippo

Corporate experience

- 2018: Technology Consulting Trainee at Deloitte
- 2019-2020: Data / Business Analyst at RELEX Solutions

Academic/Research Experience

- 2020-2024: PhD in Data Science from CBS, Denmark.
    - (PhD Topic: Bots on Social Media: The Past, Present and Future)
- 1 - 9.2024: Postdoc at University of Copenhagen on AI in healthcare
- 10.2024- : Assistant Professor of Business Analytics at Hanken School of Economics.

My Research in Large Language Models

Champions for Social Good: How Can We Discover Social Sentiment and Attitude-driven Patterns in Prosocial Communication?

Raghava Rao Mukkamala*, Robert J. Kauffman, Helle Zinner Henriksen

*Corresponding author for this work

Journal of the Association for Information Systems, 2023
AJG-4* / ABDC A*
Custom Classifiers using BERT Language Model

Using LLMs like BERT, RoBERTa, HateBERT

Framing the refugee debate on Social Media - What can we learn about frames in refugee-related events on Twitter using deep learning?

Nikodém Olšavský, Jan-Niklas Wolters, Sippo Rossi, Raghava Mukkamala, Jason Thatcher
(under review in FT-50 / AJG 4* journal)

Detecting Online Hate Speech on Refugees Using Deep Learning Methods

Frederik Jensen, Henry Stoll, Sippo Rossi, Raghava Mukkamala
(under review in FT-50 / AJG 4* journal)

Hacking the Conversations: Prompt Injection Attacks on Large Language Models

Using GPT-4/ChatGPT direct and indirect prompt injections

Are Deep Learning-Generated Social Media Profiles Indistinguishable from Real Profiles?

Sippo Rossi
Copenhagen Business School
sr.digi@cbs.dk

Youngjin Kwon
Temple University
youngjin.kwon@temple.edu

Odd Harald Auglend
Copenhagen Business School
oha.digi@cbs.dk

Raghava Rao Mukkamala
Copenhagen Business School
rrm.digi@cbs.dk

Matti Rossi
Aalto University
matti.rossi@aalto.fi

Jason Thatcher
Temple University
jason.thatcher@temple.edu

Hawaii International Conference on System Sciences, 2023
Using GPT-3 to generate fake profiles/tweets for experiments

Accepted in IJIM 2023
ABDC A level

Augmenting Research Methods with Foundation Models and Generative AI.
Rossi, S., Rossi, M., Mukkamala, R., Thatcher, J., & Dwivedi, YK.

Course Introduction

CBS

Expectations for this course

Why this course?

- Even the Management/Social Sciences researchers can not hide from technology.
- These disciplines, which were once thought to be an independent discipline, are now becoming more and more interdisciplinary.
- Huge amount of text data - which contains valuable information like people's opinions, sentiments, and others

Expectations for this course

What will this course provide?

- A robust understanding of how NLP and Generative AI work.
- We focus on the understanding of high-level architectures of NLP, Deep Learning, and Large Language Models.
- We focus on why and how these technologies work so that you will have a better understanding of how things work behind these technologies.
- We will take an impartial view/understanding of the technology.

Expectations for this course

What will this course **not** provide?

- You won't become an expert in deep learning and Gen. AI. That requires several years of hardcore training in Machine Learning and Deep Learning.
- We don’t write code to pre-train and fine-tune the LLMs; we write code to use them to perform various NLP tasks!
- We don’t discuss how organizations can use these technologies to their competitive advantage, which means we don’t discuss innovation or business Strategy regarding LLMs. Rather, we focus on understanding the technology.

Course Schedule and Exam

CBS

Course Schedule - Day 01

Session	Time	Topic & Objective	Study Material
	Day-01: Tuesday, 11-03-2025		
00	10:00 – 10:15 [RRM + SR]	Course Introduction and Practicalities	
01	10:15-12:00 [SR]	Fundamentals of Machine Learning and Natural Language Processing (NLP) • Types of Machine Learning • Performance Measure • Basic Text Processing and Tokenization, • Word normalization and Parts-of-speech tagging	Slides, articles and other reading materials
Lunch Break			
02	13:00 – 14:45 [RRM]	Supervised approaches for NLP: text classification sentiment analysis, Naïve Bayes Classifier	Slides, articles and other reading materials
03	15:00 - 17:00 [RRM]	Case Studies and Hands-on Session: • Analyzing text from discussion forums on Type-2 diabetes for domain-specific text classification • Sentiment analysis and text classification for movie reviews using NLTK	Jupyter notebooks and Python scripts

Course Schedule - Day 02

Session	Time	Topic & Objective	Study Material
	Day-02: Wednesday, 12-03-2025		
04	09:00 - 11:30 [RRM]	Unsupervised and Deep Learning approaches for NLP: • Topic modeling • Word Vectors/Word Embeddings	Slides, articles and other reading materials
Lunch Break			
05	12:30 – 14:30 [RRM]	Case Studies and Hands-on Session: • Analyzing newspaper headlines using Topic modeling • Building word embeddings for your own text corpus	Jupyter notebooks and Python Scripts
06	14:45 - 17:00 [SR]	Introduction to Generative AI and Large Language Models (LLMs): transformers architecture, attention mechanism and generating text with transformers	Slides, articles and other reading materials

Course Schedule - Day 03

Session	Time	Topic & Objective	Study Material
	Day-03: Thursday, 13-03-2025		
07	08:30 – 10:00 [RRM]	Configuring and fine-tuning LLMs for specific applications, e.g. text classification, text summarization.	Slides, articles and other reading materials
08	10:15 - 11:30 [RRM]	Case Studies and Hands-on Session: Hands-on: text summarization and text classification using LLMs using Google Cloud and Gemini LLM	Jupyter notebooks and Python Scripts
Lunch Break			
09	12:30 - 13:30 [SR]	LLMs use cases, challenges, opportunities, and ethical considerations	Slides, articles and other reading materials
10	13:30 – 14:30 [RRM + SR]	Wrap-up: Discussion about exam projects! Feedback and reflections on the course	

Course Materials

Tiny URL:
https://tinyurl.com/NLP-PHD-Course

Dropbox URL:
https://www.dropbox.com/scl/fo/tkv23m6yv3lnahv319mvh/AJCuXYHT2MaCAdO5acjQ5yQ?rlkey=fwrcxrg81xxxusoc3b238305t&dl=0

Questions?

Thank you!

Raghava Mukkamala

rrm.digi@cbs.dk

https://www.cbs.dk/staff/rrmdigi

https://raghavamukkamala.github.io/

https://cbsbda.github.io/