BACHELOR'S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Janata Pavel** Personal ID number: **465810**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Open Informatics**

Branch of study: **Computer and Information Science**

II. Bachelor's thesis details

Bachelor's thesis title in English:

Transfer Learning for Textual Topic Classificaton

Bachelor's thesis title in Czech:

Transfer learning pro klasifikaci textu

Guidelines:

Recently a significant result was achieved by using transfer learning in natural language processing (NLP) [1]. The main breakthrough was the use of a model pre-trained on Wikipedia corpus to obtain state of the art performance on a classification of textual data in a different dataset.
The student will verify that the performance of this approach is consistent on a different dataset containing textual data along with their classes. The goal of the thesis will be accomplished by successfully performing the following steps:
• Study the state-of-the-art approaches to transfer learning in the field of NLP.
• Analyze existing datasets containing textual data and their corresponding class labels (e.g., Routers Dataset) and choose the one most suitable for validating the approach.
• Use the existing pretrained model "Wikitext 103" provided by FastAI. Fine-tune the model on the dataset chosen in the previous task.
• Evaluate the performance of the model on this dataset and compare it to the results published in [1].

Bibliography / sources:

[1] Radford, Alec, et al. "Improving language understanding by generative pre-training." URL https://s3-us-west-2. amazonaws. com/openai-assets/research-covers/language-unsupervised/language_ understanding_paper. pdf (2018).
[2] Howard, Jeremy, and Sebastian Ruder. "Universal language model fine-tuning for text classification." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. 2018.

Name and workplace of bachelor's thesis supervisor:

Ing. Jiří Čermák, Ph.D., Blindspot Solutions, Prague

Name and workplace of second bachelor's thesis supervisor or consultant:

doc. Ing. Jiří Vokřínek, Ph.D., Artificial Intelligence Center, FEE

Date of bachelor's thesis assignment: **14.01.2019** Deadline for bachelor thesis submission: **24.05.2019**

Assignment valid until: **30.09.2020**

Ing. Jiří Čermák, Ph.D.	doc. Ing. Tomáš Svoboda, Ph.D.	prof. Ing. Pavel Ripka, CSc.
Supervisor's signature	Head of department's signature	Dean's signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

. _____
Date of assignment receipt

_____
Student's signature