

# Veľké jazykové modely pre generovanie kódu so zameraním na vstavované systémy



Matej Vadovič  
xvadov01@vutbr.cz  
Vedúci práce: doc. RNDr. Pavel Smrž, Ph.D.

## Abstrakt

Cieľom práce bola adaptácia predtrénovaného jazykového modelu pre účely generovania kódu v oblasti vstavovaných systémov. Práca predstavuje novú dátovú sadu pre ladenie modelov generovania kódu, ktorá obsahuje 50 tisíc dvojíc zdrojového kódu a komentárov. Na dátach nového korpusu boli ladené dva nové jazykové modely pre generovanie kódu založené na architektúre transformer. Prvý model, MicroCoder, je založený na modeli CodeLLaMA. Druhý model, MicroCoderFIM, založený na modeli StarCoderBase podporuje vyplňovanie kódu na základe okolia. Modely boli porovnávané na základe metrik BLEU, CodeBLEU, ChrF++ a ROUGE-L, pričom MicroCoderFIM ukázal viac ako 120% zlepšenie vo všetkých sledovaných metrikách.

## Nová dátová sada

Vytvorená dátová sada obsahuje približne 50 tisíc funkcií, z rôznych repozitárov z platformy GitHub v jazykoch C a C++, pre tréovanie jazykových modelov v oblasti vstavovaných systémov. Každá vzorka obsahuje zadanie (textový popis a signatúra funkcie), a referenčné riešenie (vnútorný kód funkcie).



## Ladenie modelov

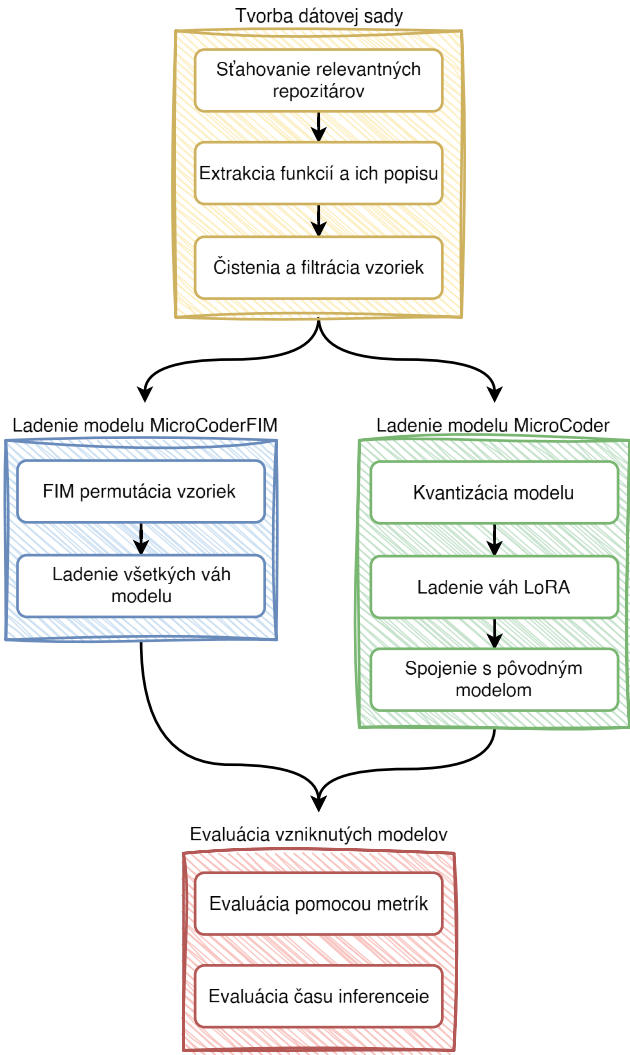
Model MicroCoder bol tréovaný na generovanie kódu funkcie z textového zadania a signatúry funkcie. Pri jeho ladení bola pre minimalizáciu výpočtových nárokov ladenia využitá technika kvantizácie váh na 4 bity a tréovanie nízkorozmerných matic LoRA. MicroCoderFIM je menší model určený pre vyplňovanie kódu na základe okolia, ladený technikou fill-in-the-middle. Tréovanie modelov prebiehalo na grafickej karte NVIDIA GeForce RTX 3090, pričom trvalo 15,5, respektíve 5,5 hodiny.

```
1 /* Unregisters an interrupt handler for the timer
2 interrupt. */
3 void TimerIntUnregister(unsigned long ulBase,
4 unsigned long ulTimer)
5 {
6     ASSERT((ulBase == TIMER0_BASE) || (ulBase ==
7 TIMER1_BASE));
8     ASSERT((ulTimer == 0) || (ulTimer == 1));
9     if (ulTimer == 0)
10     {
11         IntDisable(INT_TIMER0A);
12         IntUnregister(INT_TIMER0A);
13     }
14     else
15     {
16         IntDisable(INT_TIMER0B);
17         IntUnregister(INT_TIMER0B);
18     }
19 }
```

Výstup modelu MicroCoder.

Úloha	Model		Metriky			
			BLEU	CodeBLEU	CHRF++	ROUGE-L
NL - PL	GPT-3.5 TURBO	175B	15,21	15,00	25,92	27,84
	CODELLAMA-INSTRUCT	7B	4,67	18,40	19,50	16,10
	DEEPSEEK-CODER	6.7B	9,63	17,18	25,79	18,45
	MICROCODER	7B	10,43	16,85	25,33	20,84
	MICROCODER	7B	8,24	17,49	24,92	17,55
FIM	STARCODERBASE	1B	11,85	11,58	19,54	19,18
	MICROCODERFIM	1B	31,74	40,53	51,54	43,31

Výsledky evaluácie modelov na testovacej sade.



## Evaluácia modelov

Modely boli porovnávané na výsledkoch metrik BLEU, CodeBLEU, ChrF++, ROUGE-L a pass@k.

```
1 /* This function is registered to the SPI framework for this SPI master controller. It will
2 queue the spi_message in the queue of driver if the queue is not stopped and return. */
3 static int pl022_transfer(struct spi_device *spi, struct spi_message *msg){
4     struct pl022 *pl022 = spi_master_get_devdata(spi->master);
5     unsigned long flags;
6     spin_lock_irqsave(&pl022->queue_lock, -EBUSY);
7     if (pl022->run == QUEUE_HALTED)
8         return -ESHUTDOWN;
9     if (pl022->stopping) {
10         spin_unlock_irqrestore(&pl022->queue_lock, flags);
11         return -ESHUTDOWN;
12     }
13     if (!list_empty(&msg->queue))
14         pl022->busy = 1;
15     else
16         pl022->busy = 0;
17     pl022->msg = msg;
18     msg->actual_length = msg->status = -EINPROGRESS;
19     msg->state = STATE_START;
20     list_add_tail(&msg->queue, &pl022->queue);
21     if (pl022->run == QUEUE_RUNNING && !pl022->busy)
22         queue_work(pl022->workqueue, &pl022->pump_messages);
23     spin_unlock_irqrestore(&pl022->queue_lock, flags);
24     return 0;
25 }
```

Výstup modelu MicroCoderFIM. Zvýraznená je vygenerovaná nápoveda.

Model	pass@1	pass@10	pass@100
CODEX 12B	28,81 %	46,81 %	72,31 %
CODELLAMA-INSTRUCT 7B	34,8 %	64,3 %	88,1 %
STARCODERBASE 1B	15,17 %	—	—
MICROCODERFIM	7,83 %	12,73 %	19,07 %

Výsledky modelov na sade HumanEval.