

Sumarizační modely v kontextu hlubokého učení jsou tradičně trénovány metodou maximální věrohodnosti s použitím referenčních souhrnů. Aktivní je také výzkum v oblasti učení s vlastním dohledem (self-supervised), kde reference nejsou vyžadovány a výsledné modely jimi nejsou limitovány. Na tento výzkum navazujeme v této práci návrhem nové funkce odměn (reward function), která hodnotí kvalitu jednotlivých tokenů souhrnu. Tuto funkci pak aplikujeme ve zpětnovazebním učení. Celou trénovací logiku implementujeme modulárně, kdy nezávisle na sobě vyhodnocujeme a ladíme modul učení s učitelem, algoritmus zpětnovazebního učení a funkci odměn. Stejně tak ladíme i výsledný program po propojení těchto komponent. Výsledné modely vyhodnocujeme na 12 automatických a 3 manuálních metrikách. V téměř všech případech náš přístup zlepšil skóre na metrikách nevyžadujících referenční souhrn (reference-free).