



PŘEDNÁŠKOVÁ SÉRIE LECTURE SERIES

Vít Gvoždiak & Vladimír Havlík

Generativní AI, význam a velké jazykové modely

Generativní umělá inteligence (AI) zpracovává přirozený jazyk (NLP) takovým způsobem, že její výstupy jsou často neodlišitelné od jazykové produkce lidí (Sadasivan et al. 2023), a to včetně kontextuálního „porozumění“ významu, intence a nálady mluvčího nebo pisatele. Tato skutečnost má v zásadě tři možná vysvětlení. Velké jazykové modely (LLMs)/generativní transformery (GPT) (1) provádějí pouze statisticky podmíněnou manipulaci s výrazy a jejich výstupy postrádají obsah/význam; jsou jen „porovnávači vzorů“, „tabulkovými vyhledávači“ nebo „stochastickými papoušky“, jejichž výsledky antropomorfizujeme (např. Bender et al. 2021, Marcus 2022), a jejich užívání jazyka je jazykovým užitím pouze „jako by“ (srov. Cappelen – Dever 2021); (2) jsou nějakým způsobem schopny přirozenému jazyku skutečně porozumět, aniž by ale bylo toto porozumění podmíněno intencionálními mentálními stavy (Søgaard 2022); (3) jazyku nejen rozumí, ale porozumění je činí vnímavými (sentient) entitami vykazujícími intencionální postoje (Lemoine 2022). V přednášce se budeme zabývat legitimitou těchto postojů vzhledem ke klasickému argumentu čínského pokoje (Searle 1980, 1984), symbol grounding problému (Harnad 1990), vector grounding problému (Mollo – Millière 2023), Octopuss testu (Bender – Koller 2020) nebo Thai Library problému (Bender 2023), atd. S ohledem na současné empirické výzkumy schopností různých jazykových modelů (LLMs) a prokazatelné projevy tzv. emergentních vlastností, které se v modelech objevují, aniž by na ně byly modely předem trénovány, se pokusíme charakterizovat budoucí perspektivy generativní AI.

08/06/2023 | 15.00 CET

Meeting room of the Centre for Medieval Studies, Institute of Philosophy, Jilská 1, Prague