

Divoké výstřelky umělé inteligence a jejich vztah k nám

Umělá inteligence (AI) je poměrně nová technologie, jednak proto, protože nám dříve nestačil výpočetní výkon, a také proto, že jsme v měli v počítačové vědě historicky tendenci posunovat význam slov „umělá inteligence” – např. McCarthy a jeho průkopníci považovali za AI rozpoznávání textu na skenech, dnes by jen málokdo.

Formy a aplikace umělé inteligence jsou v dnešní době mnohem pokročilejší, a třeba dosti odlišné, než jak si to již dříve zmíněný John McCarthy představoval, když jí v 50. letech poprvé pojmenoval, ale jedna věc se s ní táhne už od počátku. A tím je možnost dostat se do stavu, kdy takříkajíc začne dělat, co by dalo nazvat jen jako „naprostá blbost” nebo tropit divoké výstřelky, mnohdy i velmi směšné.

Dnes se však AI zpravidla učí od nás a neurální sítě jsou trénovány vstupem pocházejícím od každodenních lidí, a proto, když se chatovací bot na Twitteru stane na základě komunikace s lidmi stane neofašist(k)ou a z Cleverbota je připrostlý vtipálek, můžeme se právoplatně ptát: Co tyto jejich projevy říkají o nás?

A když dva boti, nastavení ke komunikaci jen mezi sebou, „zdivočí” a vymyslí si vlastní jazyk, přizpůsobený jejich podmínkám a schopnostem, získáváme poznatky o vývoji primitivní, umělé mysli, a můžeme se ptát, co to znamená pro nás. Umělá inteligence zpravidla začíná jako naprostá, perfektní tabula rasa, a její „osobnost” vzniká pouze z nezaujatého přístupu k extrémnímu množství lidmi vygenerovaných, mnohdy ničím nefiltrovaných dat. Vystaven těmto podmínkám, vyvíjel by se předem ničím neposkvrněný člověk stejně? Na kolik se vlastně lišíme od našich primitivních modelů AI?