

Grafické využití AI

V červnovém čísle VTEI jsme se formou „rozhovoru“ seznámili s AI nástrojem ChatGPT. Na téma umělé inteligence přinášíme pokračování a představujeme zkušenosti s nástrojem více „obrázkovým“. Naším záměrem bylo pomocí textového zadání, tzv. „promptu“, nebo z předlohy fotografie vytvořit různé vizualizace, např. revitalizace vodního toku či návrhů podoby vodárenské věže v krajině. Než se však dostaneme k samotným ukázkám, řekněme si na toto téma pár slov.

Existuje několik AI nástrojů, jež umožňují uživatelům vygenerovat požadované obrázky na základě textových zadání, nazývaných prompty. Tyto nástroje využívají pokročilých technologií strojového učení a generativních modelů a dokážou vytvářet realistické obrázky podle popisu, který uživatel poskytne. Mezi takové nástroje patří např. DALL-E od společnosti OpenAI nebo MidJourney od stejnojmenné americké společnosti Davida Holze. Tyto nástroje mají potenciál být využity pro různé aplikace, včetně tvorby vizuálního obsahu, vizuálního designu, či dokonce návrhu nových produktů.

Pro naše účely jsme si vybrali nástroj MidJourney, což je služba pro generování grafiky pomocí umělé inteligence. Nástroj byl spuštěn na začátku roku 2022 a uživatelé vytvářejí grafiku pomocí příkazů zadaných chatbotu v aplikaci Discord.

Funkce MidJourney spočívá v rozpoznávání vztahu mezi obrazem a textem, kdy se strojově učící algoritmus trénuje na velkém množství obrázků s textovými popisy. Pokud uživatel zadá svůj požadavek/výzvu neboli tzv. prompt do chatovacího okna, umělá inteligence mu umožní vytvořit obrázek, který odpovídá jeho popisu.

Fungování/využití AI nástroje MidJourney jsme si ověřili na čtyřech příkladech.

Návrh revitalizace vodního toku Jezerka

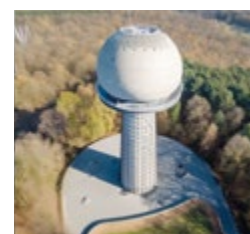
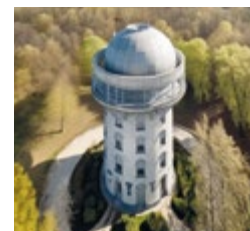
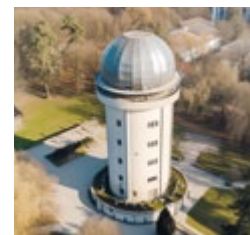
Základem byl v tomto případě obrázek vizualizace revitalizace přítoku do nádrže s mostem a mokřadní vegetací uveřejněný v letošním dubnovém vydání VTEI [1]. Celý proces vygenerování výsledku probíhal v následujícím pořadí – nahrání reálné fotografie parku před revitalizací (obr. 1a), vygenerování mostu přes vodní tok (v pořadí až cca 20. prompt s uspokojivým výsledkem, obr. 1b), propojení obou výstupů (obr. 1c) a doladění výsledného obrázku (obr. 1d). Časová náročnost tohoto procesu byla cca tři hodiny.

Vodárenská věž

V dalším případě bylo naším záměrem vykreslit stavbu vodárenské věže. Také zde byl zdrojem snímek z článku o vodárenských věžích publikovaný ve VTEI



Obr. 1a, b, c, d. Vodní tok Jezerka, situace vyústění drenáže podzemní vody jakožto občasného přítoku do vodní nádrže (Foto: T. Hrdinka, následná úprava nástrojem MidJourney)



Obr. 2a, b. Věžový vodojem v Kolíně navržený architektem Františkem Jandou ve funkcionalistickém stylu (Foto: O. Civiň, následná úprava nástrojem MidJourney)

6/2022 (obr. 2a) [2]. Použitý prompt sepsaný ve tvaru „a tall concrete tower with a metal dome of the tower, featured on cg society, danube school, arial shot, water-tank, germany, low pressure system, awe – inspiring award – winning, watedrops, manufactured in the 1920s, aquiline features, parks and monuments, brenizer method --v 5“ pak vykreslil náhledový obrázek čtyř variant (obr. 2b). Jednotlivé varianty lze poté vytvořit samostatně ve vyšším rozlišení. Časová náročnost procesu byla cca 10 minut.



Obr. 3a. Výsledek zadání „rak“ – první pokusy MidJourney



Obr. 3b, c. Finální výsledek zadání „rak“ nástrojem MidJourney



Vodní živočich

Kreativitu a schopnosti AI nástroje MidJourney jsme vyzkoušeli i na tvorbě vyobrazení živých organismů. Pomocí textového zadání jsme nástroj nechali vykreslit raka (obr. 3a). Ukázalo se, že nástroj MidJourney raky generuje s obtížemi – přidává jim nesprávnou anatomii. Sestavení promptu vyžadovalo cca 10 pokusů. Ukázka neúspěšného promptu:

„A captivating, hyper-realistic underwater photograph of a crayfish with two antennae, gracefully navigating the crystal-clear waters of a mountain creek, showing the intricate details and beauty of this fascinating aquatic creature. This stunning image is skillfully captured using a Nikon D850 DSLR camera, equipped with a NIKKOR AF-S 105mm f/2.8G IF-ED VR Micro lens, renowned for its exceptional sharpness and ability to render vivid, lifelike colors, even in challenging underwater environments. The camera settings are meticulously chosen to highlight the delicate features of the crayfish and the serene ambiance of its habitat, with an aperture of f/11, ISO 800, and a shutter speed of 1/125 sec. The composition is taken from a close perspective, immersing the viewer in the aquatic world of the crayfish as it scuttles among the rocks and submerged plants that line the creek bed. The scene is softly illuminated by natural sunlight filtering through the water's surface, casting shimmering patterns that dance across the crayfish's intricate exoskeleton and the surrounding environment. This awe-inspiring, high-resolution photograph transports viewers beneath the surface of the mountain creek, offering a rare and privileged glimpse into the secret underwater realm of the crayfish. --ar 4:3 --q 2 --v 5.“

Po tomto „neúspěchu“ byl nakonec použit jednoduchý prompt: „A crayfish, captivating, hyper-realistic photograph --ar 4:3 --q 2 --v 5“. Srovnání prvního, rozsáhlého zadání a finální podoby dobře demonstruje rčení, že někdy méně znamená více :- (obr. 3b, c). Časová náročnost procesu byla cca jedna hodina.

Budova VÚV TGM

Posledním příkladem, na němž jsme zjišťovali schopnosti AI, bylo zadání vizualizace budovy pražské centrály VÚV TGM, a to nejen v reálné podobě (obr. 4a, b), ale např. i v „lego“ podobě (obr. 4c, d). Předlohou byla fotografie budovy VÚV TGM, která byla nahrána do AI nástroje s příkazem „image to text“. Některé z prvků z popisu „image to text“ byly využity a doplněny o popis osvětlení, fotografických a výtvarných stylů a barev. Výsledný prompt „a large red and white brick building, in the style of agfa vista, dark bronze and blue, vray, school of london, computer-aided manufacturing, dark brown and navy, lively and energetic --ar 31:22 --v 5“ pak vytvořil následující výsledek.



Obr. 4a. Budova VÚV TGM (foto: archiv VÚV TGM)



Obr. 4b. Budova VUV TGM (vizualizace pomocí nástroje MidJourney)



Obr. 4c, d. Budova VUV TGM v lego stylu – náhledový obrázek vizualizace a jejich variant pomocí nástroje MidJourney



Vytvořit budovu VUV TGM z lego kostek si vyžadovalo úpravu promptu do tvaru „a large red and white brick building, in the style of agfa vista, dark bronze and blue, vray, school of london, computer-aided manufacturing, dark brown and navy, lively and energetic, as lego. --ar 31:22 --v 5“. Výsledkem bylo vygenerování náhledového obrázku (obr. 4c). Jednotlivé varianty lze pak opět vytvořit samostatně ve vyšším rozlišení (obr. 4d). Časová náročnost procesu byla cca 15 minut.

Závěr

Nástroj MidJourney dokáže úspěšně generovat opravdu hezké obrázky, v některých případech až kýchovité. Ukázalo se však, že umělá inteligence neví, co na fotce přesně je. Objekty sice pozná (řeknete jí to), ale nedokáže posoudit, zda vytvořený obraz je v souladu s námi vnímanou realitou. Příkladem může být vizualizace písma (v našem případě název naší instituce na fasádě nástrojem vygenerované budovy), kdy AI nástroj zatím neumí z dotazů vzít písmo/nápisy jako parametry. Ovšem např. Stable Diffusion si již s texty poradí.

Vzhledem k poměrně dynamickému vývoji v oblasti umělé inteligence se i funkcionality a kvalita výstupu v AI aplikacích neustále posouvají. Např. současná verze MidJourney generuje již vysoce realistické obrázky ve velkém rozlišení s mnoha detaily oproti předchozím verzím. Na druhou stranu není k dispozici podrobná dokumentace modelu, na němž MidJourney běží, proto se výsledné grafické výstupy liší v závislosti na podobě promptu, který uživatelé „vybrušují“ na základě zkušenosti práce s tímto nástrojem, a pomocí „reverzního inženýrství“ tak objevují možnosti a skrytá nastavení modelu. K tvorbě takového promptu se hojně používají další AI ve formě webových aplikací, jež dovolují sestavit výzvu „na míru“ požadované představy výstupu. Velmi dobře k tomuto účelu poslouží např. ChatGPT.

Konkrétně MidJourney nyní disponuje nejen funkcí vytváření obrázků, ale umí i jiné vygenerované obrázky po vložení do nástroje textově popsat a nabídnout vlastní verzi. Lze v něm rovněž míchat několik obrázkových vstupů, výsledek je pak jejich složeninou. Též má četné volby stylů, v nichž grafiku generuje (od napodobenin stylů různých umělců přes animované a anime výstupy až po fotorealistické grafiky, např. ve fantasy prostředí). Umožňuje také téměř libovolně variovat nabízené výstupy.

Nutno podotknout, že v současné době je využití tohoto nástroje zpoplatněno a vyžaduje registraci a přihlášení přes službu Discord.

Literatura

[1] HRDINKA, T. Plánované revitalizace vodních ekosystémů v Praze 4. *Vodohospodářské technicko-ekonomické informace*. 2023, 65(2), s. 50. Dostupné z: <https://www.vtei.cz/2023/04/planovane-revitalizace-vodnich-ekosystemu-v-praze-4/>

[2] KORÍNEK, R., HORÁČEK, M. Představení některých výsledků výzkumného projektu mapující vývoj staveb věžových vodojemů na území České republiky. *Vodohospodářské technicko-ekonomické informace*. 2022, 64(6), s. 8. Dostupné z: <https://www.vtei.cz/2022/12/predstaveni-nekterych-vysledku-vyzkumneho-projektu-mapujici-vyvoj-staveb-vezovych-vodojemu-na-uzemi-ceske-republiky-2/>

Autoři

Ing. Josef Nistler
✉ josef.nistler@vuv.cz

Tomáš Pojeta
✉ tomas.pojeta@vuv.cz

Výzkumný ústav vodohospodářský T. G. Masaryka, Praha