

KELEMEN ATTILA

HOGYAN VÁLTOZTATJA MEG A MESTERSÉGES INTELLIGENCIA A KREATÍV ÖKOSZISZTÉMÁT

A Midjourney tanulságai



A téma forró, a vélemények szélsőségesek, ez pedig egy olyan szöveg, ami már akár egy év múlva sem lesz érvényes.

■ 2012-ben, egy júniusi reggelen a Yahoo kutatói igazgatója arról számolt be a nemzetközi kreatív szakma elitjének, hogy ébredés után az amerikai tinik többségének első dolga már nem a metabolizmusuk diktálta agenda szerinti. Közbeékelődött valami új, idegen elem. Mert, amint felébrednek, még az ágyból megnézik a közösségi médiás feedjüket a párnájuk környékén tartott mobileszközükön.

Az egészen friss kutatás a Cannes Lions fesztiválon volt bemutatva. Ez a rendezvény a kommunikációs és kreatív szakma legfontosabb fesztiválja és egyféle gondolatörzsdéje, az új ötletkollekciók kifutója. Mai perspektívából túlzásnak tűnhet: a teremben a közönség el volt képedve. Elég élénken emlékszem erre a Lionskiadásra, mert ez volt az az év, amikor a reklámszakma hirtelen azzal szembesült, hogy a közösségi média megváltoztatja és vélhetően pillanatokon belül leuralja a teljes marketingkommunikációs ökoszisztémát. Mi lehet ebből? Senki nem tudta. Abban az évben egyféle kollektív pszichózis alakult ki, mert látszott a gátszakadás, de nem volt sejthető, hogy az árhullám mit sodor el, és mit termékenyít meg. Mindenkinek kellett gondolnia valamit, a stratégiai osztályoknak meg kellett győzniük ügyfeleiket, hogy minden kontroll alatt van. Csakhogy a reklámosok gondolati szabadságát némileg korlátozza, hogy ők nem lehetnek nem optimisták. Eszerint az előadók és kerekasztal-résztevők vagy azon a véleményen voltak, hogy „It’s so

amazing” (amerikaiak) vagy azon, hogy „It’s inspiring” (európaiak). Az év buzzwordje ennek megfelelően az engagement lett, és abban az értelemben használták, hogy a brandek cselekvésre kell fogják a közösségi média fogyasztóit, és ez egy remek lehetőség arra, hogy a brandek cselekvés által teljesüljenek ki, mert ettől organikusabb kötődést remélhetnek. Az amerikaiak mámorosan lelkesedtek, az európaiak visszafogottabbak voltak, de a szakma egyöntetű álláspontja az volt, hogy a közösségi média végre megnyitja az embereket egymás irányába, a kommunikáció decentralizálódik, a világ demokratikusabb lesz, és az új játékból a brandek úgy veszik ki korrektül a részüket, hogy megtanulnak zenélni ebben az új Woodstockban. A helyzetet szellemes, oda nem illő analógiákon és precedenseken keresztül interpretálták: mit hozott a nyomda, a gőzgép, a rádió, a tévé, az internet, a push-up melltartó stb.

Miért írom le ezt, egy generatív AI-tematikájú szövegben? Mert jelezhet valamit, hogy a világ legtöbb kutatási eredményével rendelkező, legjobban tájékozott, legkreatívabb, legsokszínűbb szakmájának, a reklám- és kommunikációs szakmának az elitje a legutóbbi ökoszisztéma-váltás kapcsán nem is tévedhetett volna nagyobbat. A social media transzformatív erejét senki nem sejtette, vagy ha mégis, a transformáció irányát csak nagyon kevesen. A közösségi média nem tett bennünket nyitottabbá egymás iránt, viszont az algoritmus, a hirdetési bevételek maximalizálása végett aktívan és sikeresen fundamentálja a frusztráció és gyűlölet társadalmát, a narratívaipart, a brandek kiüresedését, az újfajta patológiákat. A brandek kiszolgáltatottabbak lettek, a narratívaalkotási technikák pedig leváltották a ténymegfelelés ódon receptje szerinti igazságalkotási hagyományt. És ennek megfelelően a brandek egy része elkezdett lemaradni önmagáról, és ezzel értékvilágunk is kicsit felszínesebb lett.

Vagyis sekély az esélye annak, hogy régi intuíciónkkal sejtjük, mi lesz abból, ha valami felforgatja az eddigi berendezkedést. Nem tudhatjuk előre, mit hoz a kreatív szakmákban és a popkultúrában egy akkora dopamin-serotonin meteorit becsapódása, mint a generatív AI-ok térnyerése. Kicsi az esélye, hogy azok a hipotézisek, amelyekkel ezt a szöveget zárom, vizionárius felismerések legyenek. Gondolatkísérletem lényege nem is ez. Inkább csak azoknak a kliséknek a tudatosítása, amelyek a generatív képalkotással kapcsolatban jelenleg forgalomban vannak.

Egy friss *New York Times*ban a vezető véleménycikk azzal foglalkozik, hogy nagyobb-e a füstje, mint a lángja az AI-nak.¹ A szerző AI-szkeptikus. Arra jut, hogy középszerű munka elvégzésére remek az AI, minőségire viszont nem alkalmas. Az AI a sulimenza, a hagyományos a belvárosi étterem. Ezzel radikálisan szembe megy azzal a mainstream nézettel, amely egészen apokaliptikus. Néhány hónapja hatalmas médiavisszhangja volt a vezető fejlesztők és techbefektetők nyílt levelének, amely egyes AI-fejlesztések hibernálását követelte.² Ugyanígy szembemegy Harari³ előadás-sorozatával, aki a legbefolyásosabb történészként az AI elszabadulása miatt a civilizációnk feletti sapiensi kontroll elvesztését vizionálja.

A *New York Times*-cikk AI-szkepszisével szembe megy saját tapasztalatom is. Egy éve használok különféle AI kép- és szöveggenerátorokat, és csak ez alatt a rövid időszak alatt a technológia többször is szintet ugrott. A technológia máris elkezdte újraírni a kreatív szakma munkaköri leírásait. A téma forró, a vélemények szélsőségesek, ez pedig egy olyan szöveg, ami már akár egy év múlva sem lesz érvényes. Ki tudna ellenállni a kísértésnek, hogy tanulmányt publikáljon generatív AI-témában?

Mit tudnak már most az AI-képközpontok?

■ Az AI a kreatív szakma minden aspektusában jelen van, úgy értem, már a napi rutin részeként. Zenét szerez, szöveget ír, megadott témára közösségi médiás tartalmat készít, le is teszteli ezeket, automatizálja a kampányokat, influenszer-tartalmat gyárt, weboldalakat állít össze, stb. Jelen elemzésnek a fókuszában a Midjourney áll, amelyet napi 1-2 millió user használ képközpontásra.⁴ A Midjourney nem az egyedüli eszköz a piacon, legnagyobb konkurensa a DALL-E 3, ami a ChatGPT fizetős verziójával és más Microsoft-termékekkel integrálva működik, de az élbolyban ott van még a Leonardo AI vagy a Stable Diffusion is, hogy csak a legismertebbeket említsem. Ez utóbbi ingyenes és nyílt forráskódú, vagyis minden nehézség nélkül ki is próbálható.

A Midjourney „text to image” statikus képeket készít. Vannak mozgóképkészítők is, például a Haiper.ai,⁵ amelynek a mostani béta verziója 4 másodperces mozgóképet készít szöveges, képes vagy ezek kombinált parancsaiból kiindulva. A Haiper.ai ingyenes, könnyen használható, pont ezért is említtem – a téma iránt érdeklődők könnyen, szoftvertelepítés nélkül kipróbálhatják az élményt. A video AI szcena is élénk, kitermelte a krónikáit, szakértőit, influenszereit.⁶ Jelen szöveg írásának pillanatában a nagy fejlemény, hogy a Google bejelentkezett a piacra egy videogenerálóval, a VEO-val. A szakértők ezekben a napokban lázasan tesztelik, jobb-e, mint a SORA, amelyet az iparág jelenleg vezető megoldásának tekintenek.

A generatív videó AI területe még erősen kísérleti, a statikus képközpontok viszont professzionálisak, és teljes értékű eszközei azoknak, akik vizuális kommunikációval foglalkoznak.

Hogyan működik a Midjourney?

■ A Midjourney egy generatív mesterséges intelligencia. Legtöbb versenytársával ellentétben független és saját finanszírozású projekt, amelynek fejlesztése zárt forráskódú. Ennek következtében a fejlesztés pontos részletei homályosak. Ami egészen biztos, hogy a program nagyban támaszkodik két gépi tanulási technológiára: egy nagy nyelvi modellre (LLM) és egy szofisztikált képdiffúziós modellre. A nagy nyelvi modell lehetővé teszi a program számára, hogy megértse a felhasználók utasításainak jelentését. A diffúziós modell lényege, hogy a képeket zajjára, majd a zajt képpé alakítja vissza, több lépésben, több hierarchiában, vagyis újrakomponál. A két modell összekapcsolva véletlenszerű, azon belül paraméterezzhető eredményt hoz.⁸ Ez a gyakorlatban úgy néz ki, hogy egy szöveges leírás alapján az algoritmus értelmezi az igényt, képi megoldásokat gyárt, olyanokat, amelyek generatívok, vagyis soha nem lesznek ugyanolyanok. Szöveg leírása, majd kép generálása, majd kép és/vagy szöveg finomhangolása, és ha nem jön össze a várt eredmény, ennek adaptált, újragondolt ismétlése. Tulajdonképpen ez az alkotási ciklus. A szöveges parancsokhoz már képek is társíthatók, és ez főleg a stílus finomhangolásában tesz hozzá sokat, mint később látni fogjuk.

A Midjourney esetében több modell és alkotói mód közül lehet választani. A RAW módban generált képek a legtermészetesebbek. Ha fotókat generálok, leginkább ezt a módot használom. A Stylize szintek által (4 ilyen van) az esztétikai faktor intenzitását állítjuk be – minél magasabb, annál kevésbé természetes, ám annál látványosabb, intenzívebb, „művészibb” az eredmény. A Chaos és

Weird parancsok azok számára hasznosak, akik rábízzák magukat a véletlenre, inspirációt szeretnének, radikálisan új közelítéseket.

Stílus-e az AI?

■ Kreatív körökben gyakran hallani az „ez olyan éjajós” kifejezést. A verdikt azokra a képekre áll, amelyek túl szépek, túl tökéletesek, valószínűtlenül látványosak, túlfilterezettek, idilliek, tónusuk intenzív, a hatás erősen emocionális, a részletek pazarul kidolgozottak, álomszerűek, egészen sajátosan giccsbe hajlók, jellegzetes AI-hibákat vagy anomáliákat produkálnak, olyan bódítóan ambivalens hatással vannak az idegvégződéseinkre, mint egy augusztusi triatlon után egy cukros, túlhűtött kóla. Az AI ember által trénel (tanított), így hajlamos arra, hogy olyan képeket gyártson, amelyeket a felhasználók milliói validálnak, és az elcsúszás a „túl szép, hogy igaz legyen” kétes territóriumába tulajdonképpen elkerülhetetlen.

A különféle trendfigyelők azzal számolnak, hogy a 2020-as években nagyságrendileg 15-20 markánsan különböző mainstream vizuális stílus azonosítható. Évi 1-2 kiszorul vagy bejön a trendbe. A trendjelentésekben az AI mint stílus is megjelent, méghozzá két nagyon különböző értelemben. Az egyik a fent körülírt, kicsit kétes vonatkozásban. A másik már erősen progresszív. Ez az AI organikus dizájnban elért szuperképességeire utal.⁹ Vagyis, ha némi fogalomtisztázásra is szorul, az AI ez idő tájt valóban tekinthető vizuális stílusnak is. Ezzel egy időben viszont a modellek komplexitása és minősége inkább egy sokkal általánosabb generatív eszközzé teszi az AI-t, mint egyes stílusok célszerszámává. Egy új morfizmus kezdetén vagyunk: a kép- és formaalkotásban az ecset, a véső, a kamera, a Photoshop termékcsalád stb. mellett megjelent a modell és az algoritmus is, és ez arra kényszerít bennünket, hogy új fogalmakat alkossunk – legyen szó esztétikáról, marketingről vagy akár politikáról és etikáról. Vagyis az AI stílus is, eszköz is, az alkotás szociológiájára és gazdaságára ható tényező is.

Különbözik a különféle AI-platformok stílusa? Igen, alapbeállításban, egyszerű parancsok esetén bizonyos különbség van kromatikában, fénytörésben, karakterexpresszivitásban stb. Viszont a Midjourney, mint ahogy más platformok is, saját magához képest is különbözik stílárisan, mert minden új modell „esztétikai evolúciót” is hoz. Ember vagy kiborg legyen a talpán vagy kerekén, aki meg tudja mondani, hogy egy komplex kép Midjourney 6-ban vagy DALL-E 3-ban készült.

A képgenerálás folyamata

■ A Midjourney univerzumában a varázspálca a prompt, vagyis a szöveges parancs. Ezeket az utasításokat a Midjourney Discord¹⁰ szerveren kell beírni a megfelelő Midjourney szerver képképző chatjébe (igaz, bétában tesztelhető egy felhasználóbarátabb webes interfész is).

A prompt mérnök számára az angol nyelvű szöveges leírás, képreferenciák megadása és többféle, egyre komplexebb paraméterezés áll a rendelkezésére ahhoz, hogy megkonstruálja a képet. Az algoritmus egy ilyen parancs, prompt megadását követően négy előnézeti verziót gyárt, vagyis négy mintaképet készít. Ha valamelyik kép iránya megfelel a prompt mérnök elvárásainak, azt a képet kiválasztva kinagyítja – ez alatt az algoritmus finomít is a részleteken –, és eset-

leges további paraméterek megadását követően tovább hangolja a látványt. Tapasztalatom szerint egy professzionális szempontból elfogadható, nem klisées kép húsz perc és két óra közötti munkát igényel, 5–20 nekifutást feltételez. Folyamatosan nyitottnak kell lenni arra, hogy új és új módon közelítsük meg a témát, vagyis a folyamat intellektuális prezenciát is igényel. A jó stratégia az, ha kezdetben kevesebb paramétert tesztlünk, a működőket megtartjuk, és a koncentrikus körök logikája szerint építjük a látványt képről képre. Minthogy generatív képalkotásról van szó, némi tudást igényel, hogy a karakterekben, hangulatokban, stílusokban nagy értékű konzisztenciát érzünk el. Ez jellemzően akkor fontos, ha megépítettünk egy karaktert, szcénát, hangulatot, és azt koherensen és konzisztensen tovább akarjuk vinni egy egész sorozaton át. Általában elmondható, hogy a szokásos mesterséges intelligenciás klisék elkerülése és a képalkotási kontroll elnyerése pár hónapos intenzív gyakorlást és sok referenciataralom megtekintését feltételezi. Viszont ez a játék nem olyan, mint a hangszer-tanulás, mert a jutalmazás azonnali, első pillanattól sikerélménnyel jár.

A Midjourney professzionális használata professzionális kompetenciákat igényel, viszont egy vicces diáknapos logót, egy látványos szuperhősös mozijelenetet vagy egy részletgazdag álomszerű tájképet pillanatok alatt el lehet készíteni tulajdonképpen mindenféle előzetes beavatás nélkül. A műkedvelő és a profi prompt mérnök között hatalmas a különbség. A profi nem csak ismeri az eszköz potenciálját, de ráérez arra, hogyan hajlítsa rá a parancssort az algoritmusra. Ez egyféle jedi-tulajdonság.

A professzionális promptoknak kialakult egy piaca is, néhány dollárért szofisztikált képgenerálási parancsokat vásárolhatunk. A dolog iróniája az, hogy a Midjourney esetében generatív, folyamatosan fejlődő platformról van szó, így a megvásárolt prompt nem hozhatja garantáltan azt az esztétikai értéket, amely a referenciaképen megtetszett nekünk. Ha reverzív folyamatot szeretnénk, a Midjourney meg tudja saccolni, milyen promptokkal lehetne reprodukálni a megadott képet.

A „state-of-the-art”

■ A generatív képalkotás a Photoshop piacvezető képfeldolgozó szoftver beépített funkciója is. Az Adobe Firefly több mint alkalmas arra, ha egy tárgyat el akarunk tüntetni, ki akarunk cserélni más tárggyal, vagy a tárgy egyes paramétereit, mondjuk a színét, meg kívánjuk változtatni, mindezt egy rövid szöveges parancssal. Egy kollégám egy prezentáció számára kerti virágtartót készített az Agrosel új logójából, és magam is zavarban voltam, mikor el kellett döntenem, hogy a kép generált-e. Erős AI-funkciókkal bír a Canva is, ami azért fontos, mert a Canva az új nemzedék közösségi média tartalmat gyártó legnépszerűbb eszköze. A Shutterstock, amely a stock képek egyik legnépszerűbb lelőhelye, elindította saját AI szolgáltatását. Az ő modelljük kizárólag az adatbázisukban fellelhető, nem kis számú képből merít, vagyis a forrás jogtiszta és etikus. Nem így a Midjourney, amely civilizációnk legnagyobb műalkotásrablója – és ennél a tényállásnál érdemes kicsit elidőzni.

Arról, hogy mekkora eltulajdonítást végez a Midjourney, némi támpontot adhat az, hogy például a Nidlibrary.io¹¹ galéria-adatbázisa százasával jegyez olyan művészeket, akiknek a stílusát zavarba ejtően jól modellezi a képalkotó. Kérhetek tőle például regényhős-tekintetű portrét Kees van Dongen stílusában,

szofisztikált textúrát William Morris stílusában vagy didergős női aktot Egon Schiele stílusában, de ezek kombinációját is, ha éppen nincs jobb ötletem arra, mivel foglalkozzam, ha már mindent láttam a Netflixen. További részletes katalógusok állnak rendelkezésünkre fotóstílusokról, grafikusok stílusairól, beállításokról, anyagokról, textúrákról, technikákról.

Januári hír, hogy a Midjourney 16 ezer művész munkáit tartalmazó adatbázist használt arra, hogy trenírozza a modelljét,¹² ami azt jelenti, hogy nagyon sok olyan művész stílusát is megtanulta, akit a felhasználók még nem is azonosítottak. Tulajdonképpen elismert tény, hogy a kereskedelmi képadatbázisokat és művészek online fellelhető alkotásait a Midjourney a jogtulajdonosok hozzájárulása nélkül használja fel modelljének trenírozásában. Folyamatosak a hírek arról, hogy perek indulnak a magát kutatóintézetnek nevező, idén 200 millió dolláros bevételre számító San Franciscó-i vállalat ellen. Egy a kreatív ipari esszéjében Weiler Péter képzőművész úgy fogalmaz, hogy „az AI kreativitása is egy kaleidoszkópé, nem értelmezhető a forgató kéz vagy a befogadó, arról döntést hozó emberi szem és agy nélkül. Nem hasonlítható össze az emberi kognitív kreativitással”. Na ja, az emberi kreativitással nem, bár stílusfúziós képességeit tekintve vitára érdemes a kérdés. A Midjourney-hez ember kell, a Midjourney csak egy eszköz. Mégis, kicsit olyan ez, mint az a képzeletbeli szolgáltatás, amely előre bereszelve, testhőmérsékletűvé melegítve, ergonomikus székkal együtt készíti be a mackónak a kreatív tolvajkulcsot.

A művészvilág szorongása érthető. Arról, hogy miként élnek meg a művészek azt, hogy a Midjourney és más hasonló AI-platformok esztétikai és hosszú távon nem csak esztétikai értelemben is kismizmizik őket, számtalan nagyon kemény állásfoglalás született az utóbbi hónapokban.¹³ A perek zajlanak, a törvényhozók bottal piszkálják a kérdést, mert a stílus vagy művészeti esztétikai innováció nehezen definiálható, izolálható, mérhető egység.

A megoldások egy része jöhet technológiai oldalról is. A Chicagói Egyetem kutatócsapata fejlesztett egy még nem tökéletes, de ígéretes eszközt arra, hogy a képek értelmezhetetlenség legyenek az AI-tanuló algoritmusai számára. A Glaze¹⁴ úgy működik, hogy egy 20-tól 120 percig tartó újrakódolási folyamat által a digitális kép új forrása megtéveszti az AI-t méghozzá úgy, hogy az emberi szem számára tulajdonképpen azonos marad az eredetivel. A szoftver ingyen letölthető, tesztelhető és használható, ám használata egyelőre meglehetősen nehézkes.

Egyféle megoldást hozhat a törvényalkotó is. A legkézenfekvőbb, igaz, nem a legegánsabb megoldás az lehetne, hogy a művészek „kiiratkozhassanak” az AI-modellekből. Csakhogy ezek a modellek nem lineárisan trenírozottak, vagyis nem úgy törölsz ki egy stílust, mint egy oda nem illő mondatot egy bekezdésből.

Az igazi megoldás egy új üzleti modellből jöhet. Hasonló válságcikluson túl van már a zene- és a filmipar is. A technológiai innovációknak mindig több-kevesebb etikai ára van, a jogi keretrendszerek lemaradásra vannak ítélve. Az mp3 megjelenésére több próbálkozási kör után a zenei streamingszolgáltatások hoztak némi megoldást. Az, hogy miként legyen visszaosztható egy művész részére az, hogy az AI tőle is tanult, és az ő tudását is kapitalizálta, jelenleg teljesen megfoghatatlan.

A totális stílusmágnes, a –sref prompt

■ Ha nyugtalanítónak tűnik az, hogy a Midjourney reprodukálja ikonikus képzőművészek ezreinek a stílusait, és ehhez csak a művész nevét kell megadni a parancssorban, ennél még nyugtalanítóbb az, hogy a Midjourney mestere a spontán másolásnak is. Az új feature az év elejétől elérhető. A Midjourney már nem egyszerűen csak alkalmaz korábban megtanult stílusokat, hanem stílusosan értelmezi a referenciaképeket, és a stílus matematikáját alkalmazza is egy új kép generálásánál. A parancs neve Style reference, szintagmája –sref. A szintagma után kell belinkelni egy vagy több referenciaképet, és nehéz eldönteni, hogy az első eredmények után becsületesen borzong-e az ember, vagy inkább valamiféle elohimi mohóság önti el. Az alábbiakban ezt a vizuális élményt reprodukálom, nagyon egyszerűen, primitíven, minden speciális technikai háttértudás vagy trükk tudás nélkül.

Első lépés: kiválasztottam egy tetszőleges Gy. Szabó Béla-metszetet.¹⁵ Azért pont metszetet és pont ettől az alkotótól, mert az eredmény monokróm, vagyis jól látszik ezen a felületen. Azért ettől az alkotótól, mert kulturális hálózatunkban jól ismert, ám nem annyira meghatározó, hogy a nagy kiállítótermek múzeumi shopjai hűtőmágneseken árulják a művész reпрóit. Tehát adott a forráskép.



Második lépés: egy nagyon egyszerű prompittal kértem a Midjourney-től egy hasonló, téli tájképet. Kreált is 4 előnézetet:

Kiválasztottam a második képet, és kértem, tolja fel az arányokat 3:2-re
winter in the garden –sref <https://s.mj.run/jHmGv5CmeEQ> —ar 3:2



Ugyanezt a folyamatot megismételtem pár, egyre abszurdabb prompittal: the pyramids of Giza –sref <https://s.mj.run/jHmGv5CmeEQ> —ar 3:2



Marilyn monroe and a cadillac –sref <https://s.mj.run/jHmGv5CmeEQ> —ar 3:2



Fontos hozzátenni, hogy a közzétett képek egyszerű parancsok eredményei, első javaslatok, a promptok nem részletesen kidolgozottak, és mindössze egy referenciaképpel írtam körül a stílust, vagyis az eredmény és a stílus pontosságát meg jelentősen fejleszthető.

A Midjourney álláspontja

■ A Midjourney nem kommunikál, nincs PR-ja, nem hirdet. Vélhetően jó az együttműködése a terület influenszereivel, akik első kézből értesülnek az új funkciókról, de marketingértelemben jegyezhető erőfeszítéseket nem tesz. Két, nem túl hosszú, nem túl precíz interjú áll a rendelkezésünkre, ha valamilyen módon ki akarjuk hámozni a Midjourney általános vízióját vagy álláspontját olyan nagyon konkrét kérdésekben, mint a stíluslopás. Ezeket az interjúkat 2022 nyarán adta David Holz, a platform alapítója, vagyis még azelőtt, hogy a művész-társadalom egy része fellépett volna a Midjourney ellen, és az előtt, hogy a Midjourney üzleti sikere nyilvánvaló lett volna.

Holz nem jogász, nem esztéta, hanem – romantikus szóhasználattal élve – egy nyáron is pamutsapkát hordó matekzseni. A NASA-nak és a Max Plancknak dolgozott, többször utasította vissza az Apple-t, több startupot indított, az egyik az emberi kéz mozgását interpretálta és alakította át interfészé. Ebből exitelt, és indította a generatív AI-ra épülő új vállalkozását. Az egyik interjút a trendekről szóló The Verge¹⁶-nek, a másikat a gazdasági glossy magazinnak, a Forbes¹⁷-nak adta. Két éve, a Midjourney 4-es verziójának megjelenésekor, vagyis mind teljesítményben, mind monetizálási szempontból a platform korábbi korszakában, ami nagyon fontos. Ekkor a Midjourney identitása szerint kutatóműhely, ami nem ismert, de nagyon ígéretes, és alig tízen tagjai. Az alapító ideológiája kicsit homályos: „nem gondoljuk, hogy ez igazán a művészetről vagy a deepfake-ek készítéséről szól, hanem inkább arról, hogyan bővítsük az emberi faj képzelőerőjét. És mit jelent ez? Mit jelent, amikor a számítógépek jobbak a vizuális képzelőerőben, mint az emberek 99 százaléka? Nem jelenti azt, hogy felhagynánk a képzelődéssel. Az autók gyorsabbak, mint az emberek, de ez nem jelenti azt, hogy abbahagytuk a sétálást. Amikor hatalmas mennyiségű dolgot kell nagy távolságokra szállítanunk, szükségünk van motorokra, legyen az repülőgép, hajó vagy autó. És ezt a technológiát a képzelet motorjának tekintjük. Tehát ez egy nagyon pozitív és humanisztikus dolog.”

David Holz nem mestere a mély analógiáknak.

Kérheti-e egy művész, hogy a Midjourney vegye ki a modelljeiből? Erre nincs lehetőség, mondja David Holz, de töpregenek a kérdésen. Aztán arról beszél, hogy a képek metaadatai nem megbízhatók, tehát nem is tudni biztosan, hogy melyik kép, melyik stílus tulajdonosa kihez tartozik valójában. Persze némileg ellentmond ennek, hogy a Midjourney a képeket azok szöveges adataival (kép-aláírások, fájlnevek, szövegkontextus stb.) együtt szántja fel, vagyis ha nem is precízen, de hozzávetőlegesen nem kevés „metaadat” rekonstruálható lenne.

Hipotézisek, amelyek csak spekulációk

■ Hogyan hat a mesterséges képalkotás a kreatív ökoszisztémára? Van-e még egyáltalán értelme fotósnak, grafikusnak, formatervezőnek készülni? Mit foglal el, és mit szabadít fel a generatív képalkotás? Olyan mérvű átalakulás várható,

hogy korábbi intuícióinkra nem alapozhatunk. A művészeti oktatás rendje, a kreatív piac szerepkörei, a műalkotáspiac logikája is jelentősen megbillenni látszik. Az egyéni stílus nem izolálható a kontextustól, mégis felismerhető és modellezhető. De mi a mértékegysége, hogyan kompenzálható a lemásolása? Lehet művész az, akinek nincsenek konkrét műalkotásai, de kidolgozott egy piacilag releváns stílusmátrixot, és abból tarja fenn magát, hogy ezt értékesíti? Mi zajlik most a kreatív munkaköri leírásokban?

Vegyük sorra.

Stockfotósok. Az első hullámban a leginkább kitett üzletág a stockfotózás volt. A stockfotósok azok, akik szorgosan készítik azokat a fotókat, mozgóképeket, amelyeket többnyire a kis és közepes költségvetésű reklámkampányok használnak fel. A fotó- és mozgóképtételek árai néhány eurótól a néhány száz eurós nagyságrendbe kerülnek. A stocküzletág irányítói a nagy értékesítő platformok, például a Shutterstock, iStock, Getty Images stb. Katalógusuk a kezdeti időszakban nyomtatott volt, majd hatalmas online képadatbázissá vált. A generatív AI-megoldások máris igényspecifikusabb és olcsóbb megoldást nyújtanak akár kevesebb idő alatt, mint amennyi időbe telik egy stockkép megtalálása egy online katalógusban. Sok esetben az AI-megoldás olcsóbb, könnyebb, gyorsabb és pontosabb.

Van még esélye a stocküzletnek? Ha az etikus képalkotás piaci normává válik, és egyes brandek önlogikájuk szerint elutasítják a mesterséges képalkotást, akkor igen. Lehet, hogy a jogalkotó erősen megdrágítja a mesterséges képek gyártását, például stílusadókat vezet be, ezzel védve a vizuális innovátorokat, ami némileg megdrágíthatja az AI-vizuálok előállítását, de nehéz elképzelni, hogy a stock piaci relevanciája megmarad.

Grafikusok. Paradox módon minél popkulturálisabb egy grafikus stílusa, annál nagyobb eséllyel helyettesíthető már az AI-jal. A nyugati képregény és az anime alkotói 1-2 éven belül tökéletesen kiválthatók lesznek egy profi prompt-mérnök és scriptíró által. Leteszteltem a MIJI-vel, ami a Midjourney speciális képregénymodellje, és pár óra alatt meg tudtam kreálni néhány olyan képregényoldalt, ami egy profi képregényrajzolónak akár több napi munkájába is kerülhet. A Lucer's Archive az egyik nagy tekintélyű reklámgrafikai szaklap, amely rangsorolja a piac vezető alkotóit. A Midjourney 2022-ben nem szerepelt a grafikusok között, 2023-ban viszont már a negyedik helyen volt.¹⁸ Komoly szakmai kérdés, hogy a mesterséges intelligenciát szabad-e, értelmes-e bent tartani a képalkotói versenyben. Mielőtt még rávágnánk, hogy nem, mert az helikoptert sem indítjuk a maratonton ember ellen, érdemes figyelembe venni a következő szempontot: egy díjnyertes reklámgrafikának a creditlistáján ott lehet egy stratégia, egy fotós, egy art director, egy copywriter és aztán legalább egy grafikus, egy utómunkás, aki például fényezi a képet, valaki, aki a betűket megtervezte, stb. Egy nyertes Midjourney-képre alapozott reklám sok ember előzetes és utólagos munkáját feltételezi. Vagyis nem biztos, hogy ebben az értelemben egy Midjourney által kreált képre alapozott díjnyertes reklám, aminek a promptját egy vagy több ember írta, mesterségesebb, mint a digitális morfológiai szakaszok sokaságán átvezetett egyéb reklámkép.

Bár az áttörésre még várni kell, ám elég nagy a verseny a generatív mozgóképek területén is. Folyamatosan jelennek meg az új eszközök, a Curiousrefuge szakértői szintű YouTube-csatorna¹⁹ napról napra követi a terület innovációit. A mozgóképpalkotás kiindulópontja többnyire egy mesterségesen előállított kép, amiből néhány másodperces mozgóképet generál az algoritmus. A kísérletezők

ezeket a többnyire lassú mozgásokat fűzik egymásba, a megoldás kicsit melan-
kolikus, de jól áll például a slow burn retrofuturista sci-fi kísérleteknek vagy új
spirituális YouTube-csatornák vizuáljainak. Vagyis ezen a területen is javában
kialakulóban van egy eszközszerű esztétikai kliséisztika. Katsukokoiso²⁰
például egy milánói művész, és nagyon jól szemlélteti, mi a jelenlegi video AI-
határa; egy olyan radikális szürrealizmust hoz, ami az új technika, a generatív
AI nélkül elképzelhetetlen esztétikát vetít előre.

Hol tart most a mozgóképes AI? Referenciának javaslom a mostanában meg-
jelent zenevideót, amit Paul Trillo 100%-ban a SORA nevű mesterséges intelli-
genciával készített, címe *Washed Out: The Hardest Part*.²¹

Kreatív ügynökségi emberek, vagyis reklámügynökségek creative, digital és art
directorai, szövegírói, stratégiái. Az AI gyökeresen felforgatta és leegyszerűsítette a
munkájukat. Összemérhetetlenül gyorsabban készülnek a moodboardok, olcsób-
ban és gyorsabban A-B tesztelhetők az ötletek, ugyanazon időegység alatt több kre-
atív készíthető, és az AI inspirál is, ötleteket is prototipizál. Ennek ellenére a nagy
ügynökségek mindent megtesznek majd annak érdekében, hogy tartsák az AI meg-
jelenése előtti áraikat. Ez középtávon azt jelenti, hogy több hely lesz a piacon a
kisebb ügynökségeknek, akik több szolgáltatástípusban ugyanazt tudják, mint
a nagyok, de jelentősen olcsóbban. Eszerint az AI az ügynökségi bizniszre is
transzformatív hatással lesz. Helyzetbe kerülnek azok a profik, akik 1-2 munka-
szakaszra szakosodott kreatív szabadúszók voltak, de az AI segítségével végig tud-
nak vinni olyan projekteket is, amelyeket korábban nem vállaltak el.

A formatervezők, építészek, fashion designerek számára az AI már jelenlegi
fejlettségi szintjén is kimeríthetetlen inspirációs forrás. A Nike megkreatálta első
AI tervezte cipőit, és ez beállította az új időszámítást. Ha egy laikus számára is
hatalmas élmény lekövetni a generatív AI erejét, amikor stílusok fúzióját hajtja
végre, egy formatervező számára semmivel sem összemérhető az új eszköz ere-
je. Ez nemcsak kombinatorika, hanem egy új morfizmus. A nagy áttörés akkor
várható, amikor megjelennek a 3D-re szakosodott AI-ok és azok a perifériák, pél-
dául nyomtatók, amelyek igazi prototípusokat is képesek legyártani.

A működvelő dilettánsok univerzuma számára a generatív képalkotás sem-
mivel sem összehasonlítható élményt nyújt. Gyufát kaptak a gyerekek! Ezt pon-
tosan tudom, mert ehhez a demográfiához tartozom magam is. Ha valaki tudja,
mit, de nem tudja, hogyan, ott a Midjourney. Aki tudja, hogyan, de nem tudja,
mit is, ott a Midjourney. Ha valaki azt sem tudja, mit, azt sem tudja, hogyan, ott
a Midjourney. A Midjourney kicsit olyan, mint a felelősség nélküli, ám esztéti-
kai ambícióktól sem mentes viktoriánus hobbikertészkedés.

Elveszi-e a kreatív szakma munkáját a generatív AI? Erre a kérdésre nem le-
het „bölcs”, két bites választ adni. Vannak szakemberek, akik szerint az AI adott
esetben több állást teremt, mint amennyit elvesz, mert a modellek táplálása és
az eredmények validálása nem működhet humán asszisztencia nélkül. A gene-
ratív képalkotás második-harmadik évében nehéz lenne nem észrevenni, hogy
az új technológia a teljes kreatív ökoszisztéma minden szakaszára kihat. Olyan
szakmák, kompetenciák jöhetnek létre, amelyek eddig nem léteztek, a kompe-
tenciamixek megváltoznak, és nagy erőfeszítést követelő kreatív munkaszaka-
szok picicéllal átértékelődnek.

A Midjourney alapítójának egyik állítása mélységesen elgondolkodtató. Azt
állítja, eljön a pillanat, amikor a modelleket már nem úgy trénelik, hogy újabb
kép-szöveg kombókat vezetnek be, hanem elegendő lesz az a tudás, amire a fel-

használók munkájuk során tanítják meg az algoritmust. És akkor bezárul valami hatalmas nagy, táguló kör.

■ JEGYZETEK

1. Julia Angwin: *Will AI ever live up to its hype?* The New York Times 2024.05.16.
2. <https://futureoflife.org/open-letter/pause-giant-ai-experiments>
3. <https://www.youtube.com/watch?v=LWiM-LuRe6w>
4. <https://www.demandsage.com/midjourney-statistics>
5. <https://haiper.ai/explore>
6. Arról, hogy miként néz ki ez a szubkultúra, remek hangulatképet ad az, ahogy a szakértők 2023-as karácsonyi AI-videókat elemeznék: <https://www.youtube.com/watch?v=p4iONZkE3qg>
7. <https://www.digitaltrends.com/computing/google-veo-ai-reveal/>
8. Egy bevezetés a diffúziós modellek működésébe: <https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction>
9. <https://www.youtube.com/watch?v=FHFoKr4YLTl&t=62s>
10. A Discord a gamer szubkultúra egyik vívmánya, egy alapvetően szöveges szerver chatszolgáltatásból nőtt ki magát gamerhellyé, aztán nerd-szigetté. Mára egészen mainstreammé vált. A Midjourney a legnagyobb Discord server, 9 millióan használják. A Discord egy chat platform, mint amilyen például a WhatsApp, de a felhasználói élmény jelentősen különbözik. A Discord elsősorban a gameerek számára készült, és olyan funkciókat kínál, mint az szöveges és audiochat vagy a képernyőmegosztás, így ideális a gamerközösségek számára. Szerverekre és csatornákra van osztva, hierarchikus struktúrára biztosítva a közösségeken belüli kommunikációhoz. Fontos még, hogy a Discord különböző játéklatformokkal és alkalmazásokkal való integrációkat kínál, lehetővé téve a felhasználók számára, hogy összekapcsolják játékaikat, játékmenetet streameljenek, és egyáltalán: élőbbé és közösségibbé tegyék az interakciójukat. Ebből a zónából jön kulturálisan a Midjourney, amelynek kezdeteit nehéz megérteni a gamer szubkultúra nélkül. De van egy másik vonatkozás is, ami miatt a Midjourney a Discordon indult: az egész fejlesztő csapat remote dolgozik.
11. [Nidlibrary.io](https://www.nidlibrary.io)
12. <https://www.artnews.com/art-news/news/midjourney-ai-artists-database-1234691955>
13. Pár YouTube-állásfoglalást érdemes jegyzetbe venni. Steven Zapata egy grafikai projektjét közvetíti, közben mondja fel esszéjét: <https://www.youtube.com/watch?v=tjSxFAGP9Ss>. Martin Perhiniak digitális grafikus többrétegű elemzését nyújtja a problémának: <https://www.youtube.com/watch?v=ZJ59g4PV1AE>, az 1,2 millió követővel rendelkező SamDoesArts pedig, aki anime stílusú grafikai printjeinek köszönhetően népszerű, közvetlen érintettként mesél a AI várható hatásáról: <https://www.youtube.com/watch?v=5Vy3Cu3DLk>.
14. <https://glaze.cs.uchicago.edu/downloads.html>
15. <https://www.mutualart.com/Artwork/Summer/8D6A8F2F8746F31E1C784890EA2CB1AD?login=1>
16. <https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz>
17. <https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=2630d4f72d2b>
18. https://www.luerzersarchive.com/rankings/?rankings_year=2023&rankings_taxonomy=c_illustrator&rankings_country=all
19. <https://www.youtube.com/@curiousrefuge>
20. Katsukokoiso.ai
21. <https://vimeo.com/941713443>