

DRAGON ZOLTÁN

Szegedi Tudományegyetem Angol-Amerikai Intézet

Fotografikus percepció és kogníció a generatív média korában

Absztrakt

A fotográfia kompozíciós szabályrendszerét a festészettől örökölte, és a digitalizáció térnyerésével egyre prominensebbé váló számítástechnikai fotográfia percepció szempontjából nem hozott újdonságot, hiszen a klasszikus fotográfia esztétikai modelljét reprodukálja. Azonban a kortárs generatív mesterséges intelligencia már elmozdulást mutat a korábbi szabályszerűségektől. Ha a fotografikus kép maga kognitív struktúra (Sandström 2007), a kép jelentésének létrehozásában az interakció során az emberi megértés stratégiái folyamatosan visszacsatolásként, adattáplálékként értelmezhetőek a generatív rendszerek algoritmusai számára. Ennek következtében ezek a rendszerek elvileg egyre pontosabban képezik le nem pusztán a fotográfiai kép alapvető (vagy akár azon jóval túlmutató) jellegzetességeit, de azt is, hogy a néző pontosan mit lát meg, mit néz, mit keres egy képen.

Mindemellett azonban a generatív kép kísérteties marad a befogadó számára, ami abból a furcsa jelenségből ered, hogy a percepció során észlelt figyelemfelkeltő aspektus a kognitív, motoros válaszreakciók gyakorta felülírnak (ld. Moshel et al. 2022). Tanulmányomban arra keresem a választ, hogy miként ragadható meg ez az elcsúszás vagy kisiklás a befogadó szempontjából, illetve milyen elméleti és kritikai keretben vizsgálható az, ahogyan a kognitív folyamatok felülírják a primér érzékletet.

Kulcsszavak: fotográfia, analógia, perceptív szaliencia, generatív kép, szintetikus kép

A mesterséges intelligencia a neurális hálózatok egyre hatékonyabb működtetésével érkezett el abba az állapotba, amikor már átlag felhasználói felületeken is képes olykor megtévesztően emberi intelligenciához hasonló módon megnyilvánulni, akár nyelvi, akár képi, vagy mozgóképi reprezentáció tekintetében. Sok aggodalom és frusztráció övezi a generatív média termékeinek terjedését, legfőképpen azért, mert aláássa a képbe vetett hitet, a képmás autentikusságát, a valósághoz való megkér-

dőjelezhetetlen viszonyt, vagyis mert hétköznapi szinten talán most érkezett el a pillanat, amikor ténylegesen megéljük a kép kiüresedését, a baudrillardai szimulákrum jelenét. A generatív média képe ugyanis technikai szükségszerűsége ürügyén mindig már eleve szimulákrum, a másolat másolata, másodlagossága ontologikus tényező, nem tud más lenni (vagyis mindig már eleve “más”), részben ebből ered furcsa kísértetiessége, ismerős idegensége.

E kísérteties jelenségnek azonban az elméleti tételezésen túlmenően egy furcsa perceptív-kognitív disszonancia adja a konkrét vizsgálati szükségszerűségét, nevezetesen az, ahogyan az érzékelés folyamán azonosított paraméterek alapján MI-generált képnek helyesen felfogott kép az ezt követő kognitív folyamat során felülíródik és autentikus fényképként regisztrálódik. Különös helyzet áll elő optikai alapon nyugvó és generatív kép megkülönböztetésére irányuló tesztek esetén: nem humán tárgyat, például ételt megjelenítő kép esetén egy kutatás azt mutatja, a résztvevők nagy arányban helyesen állapították meg a kép eredetét, ontológiai státuszát (Califano és Spence, 2024). Azonban Moshel és kollégái egy kísérlet során arra lettek figyelmesek, hogy az emberi arc vonatkozásában jóval nagyobb szórással és kiemelkedő hibaszázalékkal válaszoltak a kutatás alanyai (Moshel et al. 2022).

Ez annak vonatkozásában meglepő eredmény, hogy Shlomo Bentin és kutatócsoportjának eredményei óta tudjuk, hogy az emberi arc látványa az úgynevezett N170-es reakciót generálja az agyban (Kropotov 2016, 137–169), mely alapján sokkal nagyobb valószínűséggel várhatnánk, hogy könnyedén azonosítható a valódi arc a generálttal szemben (Gauthier és Tarr 2002, 431). A kutatók arra a konklúzióra jutottak az EEG vizsgálat során, hogy bár a szem, a percepció folyamatában azonnal észleli a két kép közötti különbséget, vagyis nagy arányban képes azonosítani a fotografikusan előállított, valós képet a generatív, vagyis hamis képpel szemben, a kognitív feldolgozási folyamat, illetve a képpel való további foglalkozás során ezt a primér észlelést valami felülírja, ami ilyen értelemben egyfajta disszociációt, törést vagy eltolódást okoz a képpel történő interakció során. A generatív kép tehát kognitív aspektusát tekintve kísértetiesé válik. Úgy is fogalmazhatunk e tekintetben, hogy az exogén stimulusként érkező perceptív szalencia a kognitív folyamatok során elvész az MI-generált kép esetén, míg a fotografikus alapú fényképek képesek ezt a szalenciát megtartani.

Tanulmányomban nem arra teszek kísérletet, hogy megpróbáljam kitalálni, vajon mi történhet a percepció és a kogníció közötti pár századmásodpercnyi intervallumban, ami a kísérteties jelleget létrehozza: céлом e helyett az, hogy ezt a kísérteties jelleget elemezzem, rámutatva egyrészt a generatív média logikájára, másrészt arra, hogy miért ragadja meg figyelmünket a fotografikus generatív kép a fotográfiához képest eltérő módon. Az eltérés, az alteritás problematikájából kívánom kibontani azt a kognitív-perceptív folyamatot, ami alapvető különbséget jelent nem csupán a kép befogadásának tekintetében, de – a generatív technológiák

logikájának értelmében – a kép létrejöttét is alapvetően befolyásolja. Annak érdekében, hogy ezt a folyamatot behatóan tudjam értelmezni, több esetben fogok a fotográfia esztétikai-elméleti keretein túlra nyúlni, legfőképpen némely filmelméletben gyökerező terminológiához, hiszen ezek többsége nyilvánvalóan vagy kevésbé nyilvánvalóan, már eleve táptalaja volt több fotóelméleti traktátumnak is.

Fontos leszögezni, hogy amikor a mesterséges intelligencia képgeneráló eszközeiről teszek említést, valójában mindig eleve kérdőjelesen teszem, hiszen nap mint nap újabb és újabb lehetőségek és minőségek jelennek meg a különböző interfészekben. Olyan korszakát éljük a generatív médiának, mint amilyen talán a fotográfia megjelenésének első évtizedeit jellemezte: megannyi technológia, megoldás és minőség versengett egymással, mielőtt a huszadik századra valamelyest konszolidálódott a képkalkotó megoldás, és megszilárdult nem csupán a fotográfia definíciója, hanem a fotografikusság, a fotorealisztikus reprezentáció magától értetődősége. Az MI sem egyféle, a generatív megoldások is különböznek egymástól, korántsem konszolidált, megállapodott módszer és technológia. Ennek ellenére reprezentáció tekintetében több olyan visszatérő jellegzetességet produkál, ami percepció tekintetében hasonlatossá teszi őket, és szembe állítja az eredményt a fotográfiai képpel.

1. Generatív techno-logika

A mesterséges intelligencia projektjét én magam az automatizáció kontextusában értelmezem: ez annak az évszázados elképzelésnek a magas szintű megvalósítása, amely a humán munkavégzést – legyen az fizikai vagy szellemi – részben vagy egészben gépi megoldással váltja ki. Az alap tételezés ebben a kérdésben az, hogy a repetitív vagy redundáns munkavégzési fázisokat, illetve folyamatokat delegálni lehet egy, az adott problémára létrehozott gép számára, ami ráadásul gyorsabban és pontosabban is képes az adott feladat elvégzésére. Talán nem véletlen, hogy bár keveset emlegetjük, de a számítástechnika kezdetei nagyrészt egybeesnek az ipari forradalom kirobbanásával: a szövőszék például valójában már egy bináris rendszerben működő kalkulátor, ami ráadásul matematikai kódot fordít át képi reprezentációra.

Charles Babbage megvalósíthatatlannak bizonyult differenciálgépét, majd az analitikai gép tabulátoros megoldását követően az ipari forradalom gyorsan megteremtette azt a háttérrel, amiből a 20. század fejlesztései profitáltak (Manovich 2001, 22). A második világháborút követően az automatizációs lehetőségek új fázisba léptek a digitalizáció alapjainak megteremtésével, hiszen elkezdődhetett a mesterséges intelligencia projekt, mely számos vargabetűt, csalódást és reményt keltő sikert követően a GPT-rendszerek megjelenésével csúcsonyosodott ki.

A kortárs generatív MI-rendszerek működése meglehetősen összetett, de alapvetően a neurális hálózati modell, illetve a mélytanulás alapvető logikája az, hogy

egy nemlineáris, rétegekből álló rendszerben a visszacsatolások következtében lehetőség van a rétegzettség, vagy a gráfalapú elágazások gazdagítására, mely segíti a rendszer absztrakciós potenciálját, ezzel egyre bonyolultabb és összetettebb feladatok elvégzésére nyílik lehetősége (LeCun et al. 2015).

A generatív képalkotó rendszerek alapvetően úgy működnek, hogy nagy mennyiségű képi adatbázison programozzák be azokat az aspektusokat, jellegzetességeket, amire a rendszernek szüksége van a kimeneti, képalkotói fázisban, majd ezeket a rendszer elkezd alkalmazni, és a visszacsatolások alapján finomhangolja az eredményeket. Valójában tehát szöveges irányítás alapján tanulja meg a képek különböző jellegzetességeit, majd ezeket képi alakzatokban jeleníti meg, amit szintén szöveges parancsbeviteli formákkal irányíthatunk. A generatív MI rendszerek tehát nem arra használják a rendelkezésükre álló adatbázist, hogy a fellelhető képrészleteket montázsszerűen átváriálják és így hozzanak létre újabb és újabb képeket, hanem a betáplált képek alapján megtanulják, hogy például milyennek kell lennie egy fényképnek vagy egy fotogramnak. Szimulálják és modellezik tehát a képalkotást.

Ezen túlmenően érdekes, hogy a valóságra történő reflexióval kapcsolatos, kiértékelő mechanizmus is része a generatív logikának. A generatív kép létrehozása két neurális hálózat – egy generátor és egy diszkriminátor – versengésére épül, ahol a generátor célja, hogy reálisnak tűnő képeket állítson elő, míg a diszkriminátor feladata ezeknek a képeknek az értékelése: ez a rendszer hivatott megkülönböztetni a valódi adatokat a generáltaktól. Ezzel a folyamatban mindkét hálózat folyamatosan fejlődik, egyre kifinomultabbá válik, a generált tartalmak pedig minőségileg egyre közelebb kerülnek a valódiakhoz (Goodfellow et al., 2014).

Ezek a rendszerek nem öncélúan és pláne nem önmagukban fejlődnek, hanem a felhasználói visszajelzések alapján, ami azzal is jár, hogy mivel a programozók ideológiai diszpozíciója mellett a felhasználó elvárásai horizontja ugyanannak a kulturális közegnek, ideológiai narratívának az alárendeltje, minden interakció tovább erősíti, ha úgy tetszik, optimalizálja a generatív szisztéma reprezentációs logikáját. Ilyen értelemben állítja azt Eryk Salvaggio “How to Read an AI Image” című tanulmányában, hogy “minden MI-generált kép a mögöttes adatállomány infografikája” – ha úgy tetszik, mondja, “kollektív mítoszok és kimondatlan feltételezések” példázata (Salvaggio 2023, saját ford.).

Manapság ráadásul a legtöbb generatív interfész által kibocsátott reprezentációk figuratív megjelenítése Jens Schröter szerint többnyire valamiféle szürreális minőséget hordoz, mely a freudi álommunka bizonyos aspektusain túl kísérteties hasonlóságot mutat Francis Galton, az utópista eugenika kidolgozójának (Charles Darwin félunokatestvérének) korai kompozit-kép kísérleteivel, melynek során a fényképeket különböző tematikák mentén illesztette össze, vagyis szintetizált portrékat hozott létre (Schröter, 2023). Bizonyos értelemben ez a szürreális reprezentációs aspektus is erősíti ezeknek a képeknek a kísérteties jellegét.

A generatív fotografikus képet Jay David Bolter meglátása szerint tekinthetjük remediációnak, mely a William J. T. Mitchell által definiált metaképet hozza létre a szöveg-kép kombinatorikával. Bolter felhívja a figyelmet arra, hogy ez nem a sima remix örököse, hiszen a generatív rendszerek nem korábban elkészült képek összeollózáásával készítik el az új képet (Bolter 2023). Hozzáteszem, Manovich mély-remix elképzelésétől (Manovich 2013, 267–276) is eltér, hiszen nem különböző mediális logikák kombinálódnak a kép megszületéskor, ilyen értelemben talán ki lehet jelenteni, hogy a hipermédián túli generatív logikát kell emlegetnünk.

A generatív média mindezek alapján tehát minden esetben már eleve meghatározott elemek és alakzatok mentén hozza létre képeit, melyek előre programozott figyelemirányítási technikákban gyökereznek. A generatív kép ugyanis csak úgy képes reprezentációt létrehozni, hogy tudja, milyen elemekre figyel majd a felhasználó.

2. A képet érteni kell, és punctum!

A klasszikus fotografiai elméletek alapvetően befogadásorientáltak, esztétikai igényűek, és közös jellemzőjük, hogy a megörökített pillanatot valamiképpen a halálhoz kapcsolják. Ennek a tanatográf megközelítésnek alaptétele, hogy a fotó mindig *memento mori*: azt üzeni, ami egyszer ott volt, már elmúlt. A kép ontogenezise tehát egyben biztosíték-bizonyíték is arra nézvést, hogy a láthatóvá tett pillanat *valóban ott és valóban úgy* volt. Amint azt ezzel kapcsolatban Jacques Derrida megállapítja: a fénykép evidens „ontológia” (Derrida 1995, 92).

Ennek a szakirodalomnak talán a legtöbbet idézett képviselője Roland Barthes, akinek kései műve, a *Világoskamra*, melyben a nyilvánvalóan személyes és interpretatív jellege mellett Barthes egészen konkrétnek tűnő dichotómiát ajánl a fénykép értelmezéséhez. A *studium* azon értelmezői gyakorlat, mely előzetes tudásra épít: lett légyen az konkrét művészeti vagy mediális tudás, vagy tágabb, kulturális tájékozottság, netán a valóságra vonatkozó tapasztalat (Barthes 1985, 33). Leegyszerűsítve azt mondhatnánk, hogy minden, ami egy képpel kapcsolatosan leírható, szavakba önthető, értelmezhető, megérthető és felfogható, a *studium* kategóriájába tartozik.

Ezzel szembeállítva definiálja a *punctum* terminusát, ami megzavarja a *studium*ot, valami „szúrás, kis lyuk, kis folt,” másképp szólva „az a véletlen valami, ami rögtön belémszúr” (uo. 34). Míg tehát a *studium* valamiféle ideális értelmezői gyakorlatot feltételez, vagyis a vizuális információ szekvenciális logika mentén történő feldolgozását, megértését jelenti, így egy adott kognitív útvonal bejárását posztulálja, addig a *punctum* az ezt megakasztó, kizökkentő, kibillentő perceptív illeszkedési problémát előhozó aspektus. Egyfajta dekonstruktív hozzáállás ez a fénykép értelmezésében, hiszen a gondosan felépített (azaz előzetesen keretezett és

komponált) befogadási helyzet középpontját marginalizálja annak érdekében, hogy alternatív olvasatot hozzon létre.

Barthes dichotómiájának szubverzív aspektusa tulajdonképpen a szaliencia egy-fajta esztétikai leírását adja, így kognitív értelemben sokkal inkább szól a percepció működéséről, mintsem valamiféle vágymechanizmus vizuális modelljéről. Erre enged következtetni az oppozíció fenntarthatóságának problematikája is. Amint arra Jacques Ranciére rámutat, amikor Barthes a *Világokamrában* előhozakodik a konkrét képek elemzésével, azt vesszük észre, hogy a studium és a punctum egyre inkább átfedésbe kerül: mintha a punctum éppúgy függene a befogadó pszichés állapotától, előzetes műveltségétől, mint a studium: „hiszen amit, punctum címén, látni állít, ugyanannak a logikának engedelmessé válik, mint a studium, amit, mint mondja, nem lát” (Ranciére 2011, 76). Ranciére maga a „tűnődő kép” fogalmát ajánlja a probléma áthidalására, ami egy körforgásszerű összekapcsolást jelent a kép tárgya, a fotografiai kép és a befogadó között, és nem limitálja a figyelem működését egyoldalúan, hiszen ez a modell szerinte alapvetően egy többféle reprezentációs mód közötti feszültséget jelent (uo. 78), ami számomra azt sejteti, hogy az ontológiai aspektus, a technológia maga, valamint a befogadó aktuális helyzete oda-visszahatása jellemzi a kép működését.

Számomra ezek a megközelítések összhangban vannak a vizuális kultúra tanulmányok azon igyekezetével, hogy – amint azt Irit Rogoff megfogalmazza – szakítsunk a művészettörténeti, hagyományos, tekintélyelvű értelmezési rendszerekkel (ami a *good eye*, vagyis a „jó szem” típusú interpretációt alapozza meg, és nagyban hajaz a barthesi studiumra), és egy sokkal inkább befogadásorientált, a figyelem aktuális megnyilvánulásaira apelláló megközelítés felé mozduljunk el, aminek működtetője az ún. *kíváncsi szem* (*curious eye*), ami a képi befogadás és az értelmezési stratégia tekintetében a saját útját járja: mindig az adott képi jelenség és a befogadás helyzetének megfelelően (Rogoff 2002, 28). Ebben a felvetésben a ranciére-i feszültség maga a figyelemfelkeltő aspektus a Rogoff-féle kíváncsi szem számára.

A figyelemfelkeltés, a percepció inger tudatosulása, illetve hogy egy bejövő stimulus hogyan aktivizál korábbi emlékeket, és mindez milyen reakciókat képes kiváltani az emberben, már Sigmund Freud topografikus modelljében vizsgálódás tárgyává vált. Az általa lelki szerkezetnek elnevezett „összetett műszer” meglehetősen erős vizuális konnotációval bír, lényege pedig az, hogy az ún. Ψ -rendszerekből összeálló szisztéma miként tudatosít bizonyos ingereket és irányítja rá a tudatos figyelmet, és mi alapján hagy másokat elsiklani (Freud 1996, 374–377). Amint azt a bevezetőben említettem, a generatív kép perceptív szalienciája elvész a képi információk kognitív feldolgozása során, miközben az optikai alapú fotografikus kép képes ezt megtartani. A freudi modell azt tételezi, hogy a bemeneti, vagyis az érzékelési póluson áthaladva az inger emléknymokhoz kapcsolódva válhat szaliens

elemmé: olyan érzékletté, amely aktiválhat tudatos válaszreakciót.

Mark Solms és Jaak Panskepp neuropszichoanalitikai kutatása, mely az affektív és kognitív idegtudomány lehetséges kapcsolódási pontjait vizsgálja, megerősíti az alapvető freudi elképzeléseket a percepció működésével kapcsolatban (Solms és Panskepp 2012, 150–151). Meglátásuk szerint Freud jogosan osztotta két külön rendszerré az érzékletek feldolgozását, valamint helyesen mutatta ki ezek szaliens jellegzetességeit, melyeket az emlényomok aktivizálására vezetett vissza (uo. 152). A percepció, ami eredetileg a Φ -rendszer volt Freud modelljében, majd az előzetes feldolgozást végző Ψ -rendszer határozza meg, hogy a bejövő érzéklet képes-e kapcsolódni korábbi emléktöredékekhez, amelyek el tudnak jutni a tudat szintjére, vagyis beérkezhetnek az ω -rendszerbe (ibid). Solms és Panskepp szerint Freud 1917-es revíziója, mely integrálta az eredetileg elkülönülő rendszereket a percepció-tudat szisztémába, pontosan vázolja az agyban végbemenő folyamatokat, melyek a külső, exteroceptív stimulustól a motoros kéreg felé vezet (uo. 158).

Freud később ezt az elképzelést, illetve szisztémát már kifejezetten fotografikus alapon modellezte, hiszen a tudattalan folyamatokat, amelyek a bejövő stimulus alapján beindulnak, negatív képként gondolja el, amelyeknek esélye van majd „előhívásra”, tehát pozitív fordítássá, pozitív képpé válnia (Freud 1986, 242). Nem minden negatív válik pozitív képpé persze, és talán ebben – visszautalva a korábbi rendszer értelmezésére – ismét az emlényomok aktivizálódása vagy épp érintetlensége játszik szerepet, aminek következtében felkelti a tudatos figyelmet vagy sem. A freudi modell alapján azt mondhatjuk tehát, hogy a figyelem felkeltése és irányítása, a punctum megléte összetett pszichés folyamat, amely percepciós folyamattal indul, de nem feltétlenül válik kognitív észleléssé: csakúgy, mint a bevezető kutatásban, a fotográfia kontra generatív kép megkülönböztetésének szituációjában.

Kognitív oldalról az egyik legalapvetőbb figyelemre vonatkozó elképzelés, az alak-háttér észlelés fontos alapfeltétele a vizuális percepció működésének, azonban nyilvánvalóan nem fed le minden olyan aspektust, ami számomra fontos lépés ahhoz, hogy rámutassak a fotográfiai kép és a generatív média által létrehozott jelenségek alapvető különbségeire – még ha azok nem is olyan szembetűnőek néha. A fotográfia kapcsán a kompozíciós elemek elhelyezkedésén, valamint a keretezésen túl fontos kérdés az élesség, ami alapvetően befolyásolja például az alak felismerhetőségét és lokalizálhatóságát a háttérhez képest, valamint a kontraszthatás teljesen primér, perceptív, vizuális vonatkozása (ami természetesen szorosan kapcsolódik a kognitív pszichológiai aspektushoz, vagyis egy inger értékelésének a negatív, pozitív, illetve a szimultán vagy szukcesszív fajtáihoz, ám itt kifejezetten a konkrét vizuális hatásra értem). Ezen túlmenően nem elhanyagolható az érzékelés kiértékelésének referencialitása, vagyis a fotorealizmus kifejezésben rejlő valósággal való vagy hasonlósági (vagyis ikonikus) vagy oksági (vagyis indexikális) viszonya, ami ismételten elsődlegesebb, mint a szöveg alapvetően szimbolikus és

absztrakt modellje. Ehhez kapcsolódik a kép terének érzékelése és értelmezése, vagyis a mélység-élesség figyelemirányító szerepe, ami tehát szintén nem pusztán szimbolikus, hanem nagyon is vizuálisan konstruált viszonyrendszer.

Az alakzat-háttér azonosítás egyfajta kognitív térkép, ha úgy tetszik, e térkép bejárása egyfajta okularizációs séma-rendszer, reprezentációs ontopológia. A képen megjelenő információk, a technikai-mediális közvetítettség, valamint a figyelem irányítása fontos aspektusa annak, ahogyan az előttünk feltáruló világ bizonyos aspektusai keretezetté válnak. Az okularizáció kérdése a fotográfia kapcsán elsőként ott merül fel, amikor a *camera obscurá*ba érkező fény által megjelenített, befelé vetített fényjátékból egy szerzői intenció jól körülhatárolható keretbe foglalt állóképet hoz létre (Silverman 2015, 70). Ha úgy tetszik, az okularizáció aktusa a mozgó, feltáruló, megmutató képfollyamból hoz létre megfigyelésre, elmélyült kutatásra alkalmas állóképet – vagyis fényképet. Természetesen ez az aktus nincs messze a filmelméletben használt okularizáció terminusától, vagyis attól az aspektustól, ami a film előtti lehetőséget keretezi és filmivé varázsolja. A Francois Jost által bevezetett okularizáció fogalma a fokalizáció kognitív aspektusát támogatja és egészíti ki, kifejezetten a látványra, a látásra koncentrálva. Az okularizáció a filmben a látás pozicionálását, a képi nézőpont kijelölését és narratív kontextualizációját írja le. A fotográfia tekintetében az okularizáció tehát feltételezi a kép keretezettségét, amit viszont – Ranciére-hez visszatérve, és a kép tárgyára és a technológiára vonatkozó megjegyzéseket figyelembe véve – megelőz egy elsődlegesebb stádium: az elemek által indukált narratívát megalapozó mimetikus, vagyis nagyon is ikonikus réteg.

Amint arra Silverman Lady Eastlake és Oliver Wendell Holmes 19. századi fotográfusok kapcsán rámutat, a fotográfiát azonban nem biztos, hogy az indexikalitás és az ikonicitás peirce-i jelrendszeri meghatározottságában érdemes értelmezni, sőt, a Walter Benjamin által sokat hangoztatott, a valósághoz inherensen kötődő bizonyíték jellemző sem feltétlenül a leggyümölcsözőbb út annak feltérképezésére, miként tud ez az általam az imént mimetikusnak bélyegzett aspektus utat törni magának az okularizációban (uo. 8–10). Mindezek helyett az analógia kifejezést ajánlja a fénykép működésének definiálásához, vagyis azt az emberi képességet, amit Leonardo da Vinci is elsődlegesnek gondolt a humán percepció kapcsán: a világról alkotott kép egy vizuális analógia, mely az egyén saját emlékezetéhez kapcsolódva hasonlóság alapján értelmeződik (uo. 17). Ez a percepció leírás hasonlatos Ranciére tünődő képének logikájához, vagyis ahhoz az elképzeléshez, hogy a kép valójában a létező világ befogadó előtti feltárulkozását valósítja meg, melyben tehát az analógia a lényegi aspektus.

Mindezek alapján amellet érvelek, hogy a világ vagy valóság okularizáción keresztüli feltárulása egyfajta percepció eltolódással vezeti a szemlélőt a szaliens elem felé. Ennek az a folyamata, hogy míg az okularizáció fenntartja, szervezi és

működteti az alakzat-háttér struktúrát, a feltároló látvány destabilizálja ezt, a keretből kimutató, mimetikus tüskéként, egyfajta felfüggesztett, köztes tér/elemeként. Percepció tekintetében mielőtt tehát az okularizáció működésbe lépne, van egy előzetes lépés, ami ösztönszerű készítés: a látvány felé fordulás, majd a látvány befogadása – azé a látványé, ami még nem érett képpé. Ez maga a monstatív pillanat, a szó szerint értendő megmutatkozás, a még-fel-nem-ismert látványa, a látvány mint olyan. A kép, aminek vizuális tere bejárhatóvá válik, keretezettsége révén határokat jelöl ki, útvonalat teremt a kogníciónak, tehát már a látvány okularizált mivolta ebben az értelemben.

Ha már technikai oldalról belegondolunk abba, ahogyan az MI-generálta kép keletkezik, nyilvánvalóvá válik, hogy az sohasem volt látvány, hiszen szöveges parancsokból, címkékből, adatbázisból nyert modellek alapján jön létre, vagyis eleve okularizált, sőt, nyelvi struktúrája miatt akár fokalizáltként is hivatkozhatnánk erre az aktusra. Azonban az eredendően nem okularizált képi alakzatok, részek (azon reprezentációs foltok, melyeket a rendszer „kitölt”) a pusztá analógia jelen esetben szubverzív jelentéstartományát hozzák mozgásba, de a fotográfiától alapvetően eltérő módon.

Amint azt a fotográfia és a fotografikus kép alteritása apropóján Ranciére megfogalmazza: a Kép és a Látvány között az az alapvető különbség, hogy a Kép mindig másvalamire hivatkozik, míg a Látvány csupán önmagára (Ranciére 2007, 2). A látvány megmutatkozik, feltárol, míg a kép megmutat, keretezi a látványt. Másképp megfogalmazva, a fotográfia, nem pusztán ikonikus-indexikális, de sokkal inkább analogikus kapcsolata alapján kénytelen másra utalni, míg a látvány, kizárólagos szimbolikusságában csak önmagát képes másolni, miközben önmaga is csupán másolat. Parergon, ontologikusan másodlagos, szimulákrum, amelynek szubverzív szalienciája önnön essenciája: önmaga válik kísértetiessé, hiszen kénytelen keretezni a kísértetiést.

3. Az össze-tettségek problematikája

Fontos kiemelni, hogy a generatív kép nem valamiféle képi kompozit – sem létrejöttében, sem működésében. A hagyományosan pixelekből felépülő, tisztán számítógép generálta látvány David Rodowick szerint alapvetően szintetikus (Rodowick 2007, 15). Amellett érvelek, hogy a kifejezés etimológiáját figyelembe véve szintetikusnak minősítsük a generatív média vizuálit is, de nem feltétlenül és nem elsősorban a számítástechnikai háttér ürügyén, hanem a görög eredeti jelentés mentén: a *szüntitánái*, később *szüntetikosz* kifejezés annyit tesz, „összeillesztés”, „össze-tétel”.

Az MI-generált kép szintetikus, azaz össze-tett: összerakott, összefércelt (a német eredeti *verziehen* jelentéseire utalva, vagyis arra, hogy nagy vonalakban összeállított képi produktumról, össze-öltött lehetőségről van szó); a jelenkori generatív

interfészek először próbákat, előképeket mutatnak meg, majd ezekből készítik el a végső, finomított verziót.

Egy kép kísérteties aspektusát, vagyis a figyelem felkeltésének kiszámíthatatlan és tervezhetetlen alakzatát a valóság referencialitása és indexikalitása, azaz az analógia adja, hiszen ez a kapocs értékeli és érzékelteti az esetleges eltéréseket. A generatív kép esetében azonban, a valóság referencialitásának hiánya nem teszi lehetővé, hogy a képen belüli aspektus váljon kísértetiessé: maga a kép az ugyanis. Tehát a figyelem nem a befogadó összevetéséből irányul egy adott aspektusra, mint a fénykép esetén, hanem maga a kép össze-tettsége hívja fel magára a figyelmet, ahogyan a studium és a punctum megkülönböztethetetlenül egybemosódik. Éppen ezért figyelmünket ez a kép nem kép-es irányítani, hiszen eleve okularizált, össze-tett mivolta nélkülözi azanalógiát.

Az MI-generálta kép az emberi percepció (a fotó nézése) valamint kogníció (a látvány képként való felfogása) leképezése, egy kognitív térkép, mely szimulálja a befogadási folyamatot. Ha a tiszta fotográfiai kép maga kognitív struktúra, ahogyan azt Sven Sandström (2007, 7) állítja, akkor a generatív kép egy kognitív visszafordítás eredményeképpen jön létre: nem a képet termeli újra (adatbázisára támaszkodva, milliárdnyi kép részleteit újrahasznosítva), hanem azt a struktúrát (vagyis kompozíciót, térhatást stb.), ahogyan a kép perceptív-kognitív folyamata működésbe lép.

A generatív média visszafelé dolgozik: a képből indul ki, hogy képet hozzon létre, de nincs látvány, amit keretezne. A szürreális hibák, a valóság referencialitásának kiiktatása azzal jár, hogy a feltárulkozó világ analogikusan nem kínál látványt, amit az okularizáció keretezhetne. Az MI-kép nem analogizál, hanem szintetizál, következésképp a generatív kép mindig már eleve csakis szintetikus, azaz „össze-tett” lehet. Ezen techno-logikai össze-tettsége mindig már eleve, *a priori* tételezi a látvány kiiktatását és a kép kísérteties figyelemfelkeltő aspektusát.

Photographic perception and cognition in the age of generative media

Photography inherited its compositional rules from painting, and computational photography, which has become increasingly prominent with the rise of digitalisation, has not brought anything new in terms of perception, since it reproduces the aesthetic model of classical photography. However, contemporary generative artificial intelligence is already showing a shift away from the previous regularities. If the photographic image is itself a cognitive structure (Sandström 2007), the strategies of human understanding in the creation of meaning in the interaction with the image can be interpreted as continuous feedback and data loop for the algorithms of generative systems. As a consequence, these systems are in principle increasingly accurate in depicting not only the basic (or even far beyond) charac-

teristics of a photographic image, but also what exactly the viewer sees, looks at, or looks for in an image.

Nevertheless, the generative image remains haunting for the viewer, which stems from the curious phenomenon that the salient aspect during perception is often overridden by cognitive, motor response reactions (see Moshel et al. 2022). In my study, I seek to answer how this slippage or derailment can be tackled from the perspective of the perceiver, and within what theoretical and critical framework can the ways in which cognitive processes override primary perception be examined.

Keywords: photography, analogy, perceptive salience, generative image, synthetic image

Hivatkozások

Barthes, Roland. *Világoskamra*. Ford. Ferch Magda. Budapest: Európa Kiadó, 1985.

Bolter, Jay David. „AI generative art as algorithmic remediation”. *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft* 19.1 (2023), 195–207. doi: 10.25969/mediarep/22321.

Califano, Giovanbattista & Spence, Charles. „Assessing the visual appeal of real/AI-generated food images”. *Food Quality and Preference* 116 (2024), 105149. doi: 10.1016/j.foodqual.2024.105149.

Derrida, Jacques. *Marx kísértetei*. Ford. Boros János, Csordás Gábor & Orbán Jolán. Pécs: Jelenkor, 1995.

Freud, Sigmund. *Álomfejtés*. Ford. Hollós István. Budapest: Helikon, 1996.

Freud, Sigmund. *Bevezetés a pszichoanalízisbe*. Ford. Hermann Imre. Budapest: Gondolat Kiadó, 1986.

Gauthier, Isabel & Tarr, Michael. „Unraveling mechanisms for expert object recognition. Bridging brain activity and behavior”. *Journal of Experimental Psychology. Human perception and performance* 28 (2002. ápr.), 431–46. doi: 10.1037/0096-1523.28.2.431.

Goodfellow, Ian és tsai. „Generative adversarial networks”. *Advances in Neural Information Processing Systems* 3 (2014. jún.), 1–9. doi: 10.1145/3422622.

Kropotov, Juri .D. *Functional neuromarkers for psychiatry. Applications for diagnosis and treatment*. Academic Press, 2016.

LeCun, Yann, Bengio, Yoshua & Hinton, Geoffrey. „Deep learning”. *Nature* 521 (2015), 436–44. doi: 10.1038/nature14539.

Manovich, Lev. *Software takes command*. New York: Bloomsbury, 2013.

Manovich, Lev. *The language of new media*. Cambridge, MA: The MIT Press, 2001.

Moshel, Michael és Tsai. „Are you for real? Decoding realistic AI-generated faces from neural activity”. *Vision Research* 199 (2022. okt.), 108079. doi: 10.1016/j.visres.2022.108079.

Rancière, Jacques. *A felszabadult néző*. Ford. Erhardt Miklós. Budapest: Műcsarnok Kiadó, 2011.

Rancière, Jacques. *The future of the image*. Ford. Gregory Elliott. London: Verso, 2007.

Rodowick, David N. *The virtual life of film*. Cambridge MA: Harvard University Press, 2007.

Rogoff, Irit. „Studying visual culture”. *The visual culture reader*. Szerk. Nicholas Mirzoeff. London: Routledge, 1998, 24–36.

Salvaggio, Eryk. „How to read an AI image. Toward a media studies methodology for the analysis of synthetic images”. *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft* 19.1 (2023), 83–99. doi: 10.25969/mediarep/22328.

Sandström, Sven. *Explaining the obvious. A theory of visual images as cognitive structures*. 2007/2008:1 köt. Scripta Minora Regiae Societatis Humaniorum Litterarum Lundensis. Kungliga Humanistiska vetenskapssamfundet i Lund, 2007.

Schröter, Jens. „The AI image, the dream, and the statistical unconscious”. *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft* 19.1 (2023), 112–120. doi: 10.25969/mediarep/22315.

Silverman, Kaja. *The miracle of analogy, or the history of photography, Part 1*. Stanford: Stanford University Press, 2015.

Solms, Mark & Panksepp, Jaak. „The “Id” knows more than the “Ego” admits. Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience”. *Brain sciences* 2 (2012), 147–75. doi: 10.3390/brainsci2020147.