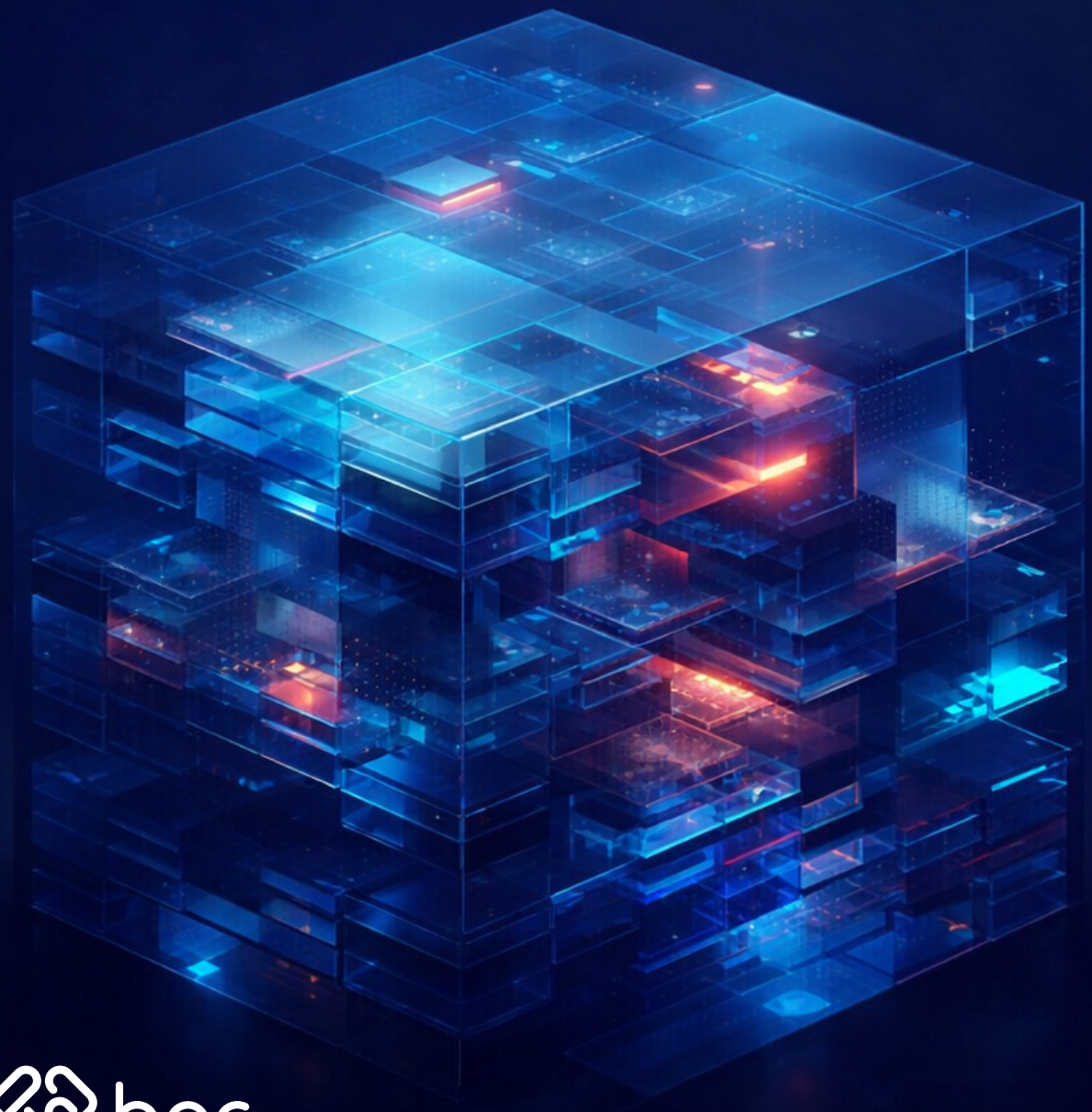
HPC ECHO

Tudománytól Innovációig



hpc

2026

Tartalom

A KOMONDORRÓL	3
Egy HPC és ami mögötte van...	4
Mágneses láncok szupravezetőkön, a nagy hibatűrésű kvantumbitek nyomában	5
Ultragyors magyar nyelvi modellt tanítottunk a Komondor szuperszámítógépen!	8
Vegyipari kismolekulák titkos élete	10
HIV és SARS-CoV-2 vírus fehérjék gátolhatóságának nyomában	14
Belepillantás a bennünk élő mikrobák univerzumába: a bélrendszer ökoszisztémájának nagyszabású számítógépes szimulációi	17
Komondor a kutyák genetikai vizsgálatának szolgálatában	20
Hogyan tanulhatunk a rendszermérnöki gondolkodásról a szuperszámítógéptől?	23
A számítógépes szimuláció és a virtuális valóság	27
Amikor a mesterséges intelligencia visszaadja a mozgás örömét: így forradalmasítja egy magyar startup a térdműtéteket	30
Hazai infrastruktúrára építve kínál biztonságos AI-alternatívát a TENGR.AI	32

Csúcson a HPC szolgáltatás

Köszöntöm az olvasót éves kiadványunk lapjain, amelyben a magyar HPC (High Performance Computing), azaz szuperszámítástechnika infrastruktúrán elért legfrissebb eredményeket összegezzük immár ötödik alkalommal. Az elmúlt időszak ismételten bebizonyította, hogy a szuperszámítástechnika már nem csupán egy szűk tudományos elit eszköze, hanem a gazdasági versenyképesség és az innováció egyik legfontosabb hajtómotorja. Legyen szó domén specifikus kutatásokról vagy az ipari folyamatok optimalizálásáról, a HPC infrastruktúra stratégiai jelentősége vitathatatlanná vált.

Ahogy az is látszik, hogy a mesterséges intelligencia (MI) és a HPC szimbiózisban működik! Az MI számításokhoz nagyszerű alapot ad a minden elemében eleve csúcsteljesítményre hangolt HPC. Az új technológia soha nem látott felfedezéseket tesz lehetővé, ugyanakkor új követelményeket is támaszt. A hardverek energiahatékonysága és a fenntartható üzemeltetés kérdései ma már éppolyan fajsúlyosak, mint a számítási kapacitás növelése. Büszkén állíthatom, hogy ebben élen járunk, de a várva-várt kapacitásnövelésre sem kell már sokat várnunk.

Hangsúlyozandó azonban, hogy a gépek teljesítménye mögött mindig az emberi szaktudás és a kutatói közösség elhivatottsága áll. A működésének ötödik évébe lépő Nemzeti HPC Kompetencia Központ segítségével felépült egy olyan közösség, amely rekordidő alatt teljesen megtöltötte hasznos kutatási feladatokkal a Komondort, amely jelenleg Magyarország legnagyobb számítási teljesítményű szuperszámítógépe, így az immáron folyamatosan csúcskapacitáson üzemel. A HPC-t működtető intézmény 2025-ben ugyan megváltozott, de a szakmaiság és a célok nem. Ez a kiadvány tisztelgés a szakemberek és kutatók munkája előtt, akik nap mint nap feszegetik a saját és az infrastruktúra határait, nemzetközileg is figyelemre méltó eredményeket produkálva. Bízom benne, hogy az itt összegyűjtött írások nemcsak informálják, hanem inspirálják is Önöket a következő év szakmai kihívásainak megoldásában.

Kiss Zoltán
HPC Technológiai és Szolgáltatási Irodavezető

A KOMONDORRÓL

A szuperszámítástechnika, vagyis a High Performance Computing (HPC) a modern kutatás és innováció egyik alapköve. A HPC rendszerek olyan számítási kapacitást biztosítanak, amely nagyságrendekkel meghaladja a hagyományos számítógépek teljesítményét. Ennek köszönhetően komplex szimulációk, nagyfelbontású modellezések, hatalmas adathalmazok elemzése vagy mesterséges intelligencia modellek tanítása valósítható meg lényegesen rövidebb idő alatt – akár hónapok helyett órák vagy percek alatt. A HPC ma már nem csupán a tudományos kutatás eszköze, hanem az ipari innováció, a gyógyszerfejlesztés, az energetika, a pénzügyi modellezés és a kormányzati adatfeldolgozás meghatározó technológiai háttere is.

Magyarország legnagyobb kapacitású és teljesítménysűrűségű szuperszámítógépe, a Komondor szuperszámítógép, a hazai HPC infrastruktúra zászlóshajója. A rendszer több mint 6 petaflop számítási teljesítményre képes, 200 Gbps sebességű belső hálózattal és 400 TB ultragyors tárolóegységgel rendelkezik, miközben kiemelkedő, 28,166 GFlops/watt energiahatékonysági mutatót ér el. A Komondor célja, hogy különböző tudományterületeken biztosítsa nagy mennyiségű számítási művelet gyors és megbízható lefuttatását.

A Komondor nemzetközi mezőnyben is megmérettetni magát: a TOP500 lista – amely a világ legerősebb általános célú szuperszámítógépeit rangsorolja a LINPACK (HPL) benchmark alapján – 2025. júniusi kiadásában a 381. helyen szerepelt. Emellett bekerült a Green500 rangsorba is, ahol a 76. helyet szerezte meg, bizonyítva, hogy nemcsak számítási teljesítményben, hanem energiahatékonyságban is versenyképes. Ez különösen fontos a fenntartható, modern HPC rendszerek esetében, ahol a teljesítmény és az energiafelhasználás egyensúlya kulcskérdés.

A Komondor működtetését és a hozzáférés biztosítását a Digitális Kormányzati Fejlesztés és Projektmenedzsment Kft. (DKF) végzi, amely Magyarország egyedüli HPC szolgáltatója és fejlesztési felelőse. A rendszer kapacitásai elérhetők a felsőoktatás, a közoktatás, a kutatóközpontok és kormányzati intézmények mellett a piaci szereplők számára is.

Az akadémiai szektorban jelenleg több száz futó projekt kapcsán használják a rendszert, elsősorban természettudományos kutatások – például molekuláris biológia, kémia, gyógyszerkutatás vagy kvantumfizika – területén. A vállalati felhasználók körében egyre meghatározóbbak a mesterséges intelligencia modelltréningek, míg a kormányzati projektek főként nagymennyiségű adat feldolgozására épülnek. A HPC Kompetencia Központ szakmai támogatásának köszönhetően a Komondor kihasználtsága 2025 második félévében elérte a 100%-ot, a számítási feladatok folyamatos sorban állása mellett.

A Komondor tehát nem csupán egy nagy teljesítményű számítógép, hanem a hazai kutatás, innováció és digitális fejlődés stratégiai infrastruktúrája, amely egyszerre képviseli a nemzetközi versenyképességet és a fenntartható technológiai jövőt.

Egy HPC és ami mögötte van…

A magyarországi HPC szolgáltatás és az elérhető infrastruktúrák idén már 25 éves jubileumi múltra tekintenek vissza. Ennek a hosszú útnak az utóbbi 5 évét már a Nemzeti HPC Kompetencia Központ (NCC Hungary) ökoszisztéma építő tevékenysége formálta.

A kezdetektől kiemelt cél volt a legmodernebb technológiák és rendszerek biztosítása. A felhasználók számára ugyanakkor legalább ilyen fontos fejlesztések és új lehetőségek bővülése a Nemzeti HPC Kompetencia Központ megalakulásával gyorsult fel látványosan.

A Nemzeti HPC Kompetencia Központ támogatásával elérhető szolgáltatások tárháza igazán széles, kezdve az egyszerű HPC hozzáférés és autentikáció biztosításától, egészen az adott Ügyfél számára testre szabott megoldások kidolgozásáig és adatelemzésig. A Központ által nyújtott szolgáltatási mátrix a lehetséges magyarországi ügyfélkör teljes spektrumát lefedi: az induló start-up vállalkozásoktól a több száz éves kutatói múlttal rendelkező akadémiai felhasználókig. A kapcsolati háló szintén rendkívül széles, a hazai érdeklődő egyéni vállalkozásoktól és nemzetközi, több száz fős kutatóintézetektől egészen a legmagasabb szintű európai HPC közösségekig terjed.

A Nemzeti HPC Kompetencia Központ segítségével lehetőség nyílik bárki számára az egyszerűbb, gyorsabb és rugalmasabb HPC hozzáférésre, a szükséges elméleti tudás és gyakorlati ismeretek elsajátítására, nemzetközi kapcsolatok kialakítására vagy éppen a világpiacra történő kilépésre.

A szuperszámítógép használatával szinte bárki nagyon komoly versenyelőnyre tehet szert önállóan is, ugyanakkor azok, akik a Kompetencia Központon keresztül találkoznak ezzel a lehetőséggel, ennél sokkal többet kapnak.

Ebben a kiadványban olyan sikeres kutatásokat és jó megoldásokat ismerhet meg a kedves olvasó, amelyek a Kompetencia Központ által biztosított szolgáltatások igénybevételével lettek igazán eredményesek és bemutatásra alkalmasak.

Terjék Mihály
HPC Kompetencia Központvezető



Mágneses láncok szupravezetőkön, a nagy hibatűrésű kvantumbitek nyomában

Szerző: *Nyári Bendegúz*

A BME és a Wigner FK kutatóinak együttműködésével létrejött **Számítógépes mágnesség és szupravezetés** kutatócsoport tagjaként közel 10 éve foglalkozom mágnesek, majd később szupravezető mágnes hibrid rendszerek elektronszerkezetének számításával. Jelenleg olyan topológikus szupravezető rendszereket vizsgálunk, melyek forradalmasíthatják a számítástechnikát.

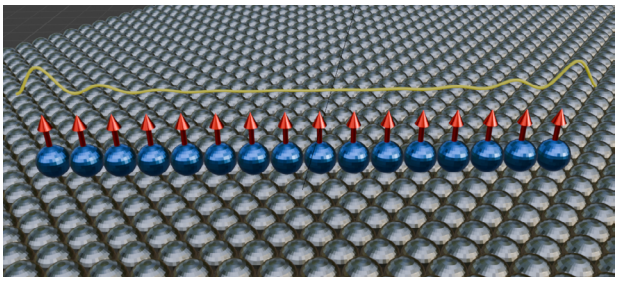
A cél tehát nem a mai szuperszámítógépek gyorsítása, hanem egy teljesen új típusú számítógép létrehozása, annál is inkább, mivel a hagyományos számítógépek bitekkel dolgoznak: egy bit vagy 0, vagy 1. Ez rendkívül sok feladatra jó, de bizonyos problémák (pl. gyógyszerkutatás, molekulák szimulációja, nagyon bonyolult optimalizálás) rendkívül lassúak még a legerősebb szuperszámítógépeken is. Erre lehet megoldás a kvantumszámítógép. A kvantumszámítógépek kvantumbitekkel (qubitekkel) dolgoznak, amelyek egyszerre lehetnek 0 és 1 is. Emiatt bizonyos számításokat nagyságrendekkel gyorsabban tudnának elvégezni. Ugyanakkor komoly kihívást jelent, hogy a kvantumbitek nagyon instabilak. A legkisebb zavar (hő, rezgés, elektromágneses zaj) is tönkreteszi a számítást.

A kvantumszámítógépek korának hajnalát éljük, és sokakkal együtt azon dolgozunk, hogy olyan fizikai rendszereket találjunk, melyek kvantumbitként a jövő számítógépeinek alapjaként szolgálhatnak.

A bemutatott eredmények nem jöhettek volna létre kutatócsoportunk többi tagjának munkája nélkül, köszönet illeti Lászlóffy Andrást (Wigner FK), Rózsa Leventét (Wigner FK), Újfalussy Balázst (Wigner FK) és Szunyogh Lászlót (BME), akik szakmai tudásukkal, ötleteikkel és elkötelezettségükkel nagyban hozzájárultak a közös eredményekhez.

A szupravezetőkre épített atomi mágneses láncok az egyik legígéretesebb modellt jelentik az új kvantumállapotok kutatásában. Ezek a hibrid nanostruktúrák – amelyek atomról atomra építhetők fel a szupravezető felületeken – lehetővé teszik a Majorana zéró módusokként ismert egzotikus kvantumállapotok kialakulását, amelyek várhatóan kulcsfontosságú építőelemei lesznek a jövőbeli topológikus kvantumszámítógépeknek. Az ilyen rendszerek tervezésének, modellezésének és megértésének képessége közelebb visz minket a környezeti zajtól eredendően védett kvantumbitek létrehozásához.

Amikor egyetlen mágneses atom kerül a szupravezetőre a környező Cooper-párok felbomlanak és lokalizált kvantumállapotok jönnek létre a szupravezető tiltott sávon belül, melyeket Yu–Shiba–Rusinov-állapotoknak (YSR) nevezünk. Ha több atom láncba rendeződik, akkor a lokalizált állapotok hibridizációja következtében egy-dimenziós YSR-sávok képződnek. Megfelelő feltételek mellett, mint mágnesetettség, szimmetria és spin-pálya csatolás, a YSR-sávok topológikus karakterre tehetnek szert, és nulla energiás lokalizált végállapotokat hordozhatnak. Ezek az 1. ábrán illusztrált Majorana-állapotok.



1. ábra. *Szupravezetőre növesztett mágneses lánc illusztrációja, mely mutatja a lánc végeire lokalizált Majorana zéró módusokat.*

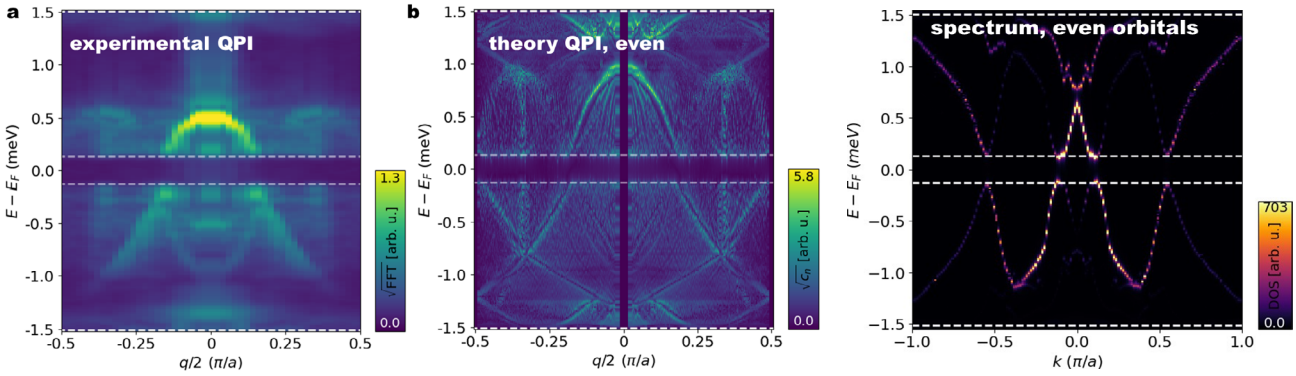
A jelenség részletesebb megismerése érdekében ötvöztük az atomi felbontású kísérleteket és a szuperszámítógépekre épülő első elvű számításokat. Az elméleti keretrendszert a Budapesti Műszaki és Gazdaságtudományi Egyetem, valamint a HUN-REN Wigner Fizikai Kutatóközpont kutatói fejlesztették. A módszer alapja az árny kolt Korringa–Kohn-Rostoker-módszer (SKKR) *[1]*, melynek keretében a Dirac–Bogoliubov–de Gennes-egyenletet megoldva *[2]* lehetőség nyílik a szupravezetés, a mágnesség és a relativisztikus hatások, mint a spin-pálya kölcsönhatás, egységes kezelésére a fizikai rendszerek egy széles választékára, mint felületek, atomfürtök *[3]* és egydimenziós láncok *[4, 5]*. A számítások elvégzéséhez szükség volt a Komondor szuperszámítógép számítási kapacitásának kiha sználására, mely lehetővé tette számunkra, hogy a szupravezető nanoszerkezetek elektronszerkezetetét anyagspecifikus módon modellezzük.

Legutóbbi munkánkban *[6]* két szupravezető anyagon, niobiumon [Nb(110)] és tantálon [Ta(110)] előállított mangán (Mn) atomláncokat tanulmányoztunk. Ezek az elemek kémiailag közel állnak egymáshoz, de bennük a spin-pálya csatolás erőssége különböző, ugyanis a tantálban, mint nehezebb elemben, erősebbek a relativisztikus effektusok. Eredményeink kimutatták, hogy ezekben a rendszerekben több, topológiailag különböző tulajdonságú sáv létezhet egymás mellett. Az elméleti elemzés és a kísérletek egyaránt azt mutatják, hogy ezek a láncot tartalmazó tükörsíkra nézve párosak és páratlanok. A nióbium hordozón a páros szimmetriájú állapotok a spin-pálya kölcsönhatás által kinyitott, széles másodlagos tiltott sávot mutatnak, de topológiailag triviálisak maradnak, ami azt jelenti, hogy a másodlagos tiltott sávon belül nem jelennek meg lokalizált végállapotok. A kísérleti és az elméleti sávszerkezet összehasonlítását mutatja a 2. ábra. Ezzel szemben a páratlan állapotok alacsony energiás jellegzetességeket mutatnak, melyek a spin-pálya kölcsönhatás erősödésével nulla energiához közelítenek és a lánchossz növelésével részlegesen lokalizálttá válnak, rájuk a Majorana-állapotok előfutáraként tekintünk.

A nióbiumot tantálra cserélve erősödik a spin-pálya kölcsönhatás és megváltozik a rendszer topológikus karaktere. Miközben a felerősített csatolás növeli a hibridizációt a páros és páratlan állapotok között, a Majorana-szerű végállapotok megjelenése nagy mértékben érzékennyé válik a mágnesezettség irányára. Számolásaink azt mutatják, hogy a mágneses momentumokat a lánccal párhuzamos irányból, vagyis az alapállapotból, a síkra merőleges irányba forgatva szignifikánsan megnő a másodlagos tiltott sáv szélessége és megjelennek az egyértelműen lokalizált végállapotok, ami olyan elméleti feltételeket teremt, amelyek mellett stabil Majorana-állapotok alakulhatnak ki.

Ezek az eredmények bemutatják, hogy a kristályszimmetria és a relativisztikus hatások hogyan határozzák meg együttesen a topológikus szupravezető fázisok megjelenését a mágneses nanoszerkezetekben. A munkánk kiemeli a HPC fontosságát a komplex kvantumanyagok modellezésében, ahol az elektronok és a mágneses tulajdonságok önkonzisztens tárgyalása szükséges.

A legmodernebb első elvű számítások és az atomi manipulációval végzett kísérletek kombinálásával azonosítottuk azokat az alapvető szimmetria mechanizmusokat, amelyek a szupravezetőre helyezett mágneses láncokban a Yu—Shiba—Rusinov-sávokat szabályozzák. E mechanizmusok megértése elengedhetetlen ahhoz, hogy olyan anyagokat tudjunk tervezni, amelyek egy napon kontrollálható Majorana-kvazirészecskéket hordozhatnak, és így közelebb kerüljünk a skálázható topológikus kvantumeszközök megvalósításához.



2. ábra. *A kísérleti és az elméleti adatok összehasonlítása Mn láncokra Nb(110)-on, a Az STM kí- sérletekben mért kvázirészecske interferencia (QPI) spektrum b az elméletből meghatározott QPI spektrum, c a spektrál függvény.*

[1] L. Szunyogh, B. Újfalussy, P. Weinberger, and J. Kollár, „Self-consistent localized KKR scheme for surfaces and interfaces," Physical Review B, vol. 49, no. 4, p. 2721, 1994.

[2] G. Csire, A. Deák, B. Nyári, H. Ebert, J. F. Annett, and B. Újfalussy, „Relativistic spin-polarized KKR theory for superconducting heterostructures: Oscillating order parameter in the Au layer of Nb/Au/Fe trilayers," Physical Review B, vol. 97, no. 2, p. 024514, 2018.

[3] B. Nyári, A. Lászlóffy, L. Szunyogh, G. Csire, K. Park, and B. Ujfalussy, „Relativistic first- principles theory of Yu-Shiba-Rusinov states applied to Mn adatoms and Mn dimers on Nb (110)," Physical Review B, vol. 104, no. 23, p. 235426, 2021.

[4] B. Nyári, A. Lászlóffy, G. Csire, L. Szunyogh, and B. Újfalussy, „Topological superconductivity from first principles. I. Shiba band structure and topological edge states of artificial spin chains," Physical Review B, vol. 108, no. 13, p. 134512, 2023.

[5] A. Lászlóffy, B. Nyári, G. Csire, L. Szunyogh, and B. Újfalussy, „Topological superconductivity from first principles. II. Effects from manipulation of spin spirals: Topological fragmentation, braiding, and quasi-Majorana bound states," Physical Review B, vol. 108, no. 13, p. 134513, 2023.

[6] B. Nyári, P. Beck, A. Lászlóffy, L. Schneider, K. Palotás, L. Szunyogh, J. Wiebe, B. Újfalussy, and L. Rózsa, „Symmetry aspects of Yu-Shiba-Rusinov bands in magnetic atomic chains on a superconductor," Physical Review B, vol. 112, no. 11, p. 115414, 2

Ultragyors magyar nyelvi modellt tanítottunk a Komondor szuperszámítógépen!

Szerzők/A kutatócsoport tagjai:
Csibi Zsolt – PhD-hallgató, MI Tanszék,
Gyöngyössy Natabara – PhD-hallgató, MI Tanszék,
Nagy Kornél – kutatócsoport menedzser, DH-Lab,
Nemeskey Dávid – tudományos munkatárs, DH-Lab,
Palkó Gábor – tanszékvezető, DH-Lab,
Sallai Martin – AI mérnök, DH-Lab,
Simonyi András – NLP szakértő, MI Tanszék,
Szekeres András – tudományos munkatárs, DH-Lab

A Racka-kutatócsoport az ELTE két tanszéke – a Bölcsészettudományi Kar Digitális Bölcsészet Tanszéke (DH-Lab) és az Informatikai Kar Mesterséges Intelligencia Tanszéke (MI Tanszék) – közös együttműködésében dolgozik. Kutatásaink célja, hogy az AI és a digitális örökség egymást gazdagítsák a tudományos munka és az oktatás területén.

> **A projekt ötlete egy esti NLP-olvasókör beszélgetése során született, amikor a nagy nyelvi modellek méretezési törvényeit és a magyar nyelv korlátozott adatkészletét vitattuk meg. Ekkor merült fel a gondolat: Magyarországnak szüksége van egy saját, kis méretű, de optimálisan skálázott nyelvi modellre, amely segíti nyelvünk digitális életének megőrzését és fejlődését.**

Célunk az AI-hoz való hozzáférés demokratizálása, a helyi nyelvi képességek erősítése, és Magyarország digitális függetlenségének előmozdítása. A modell tanítása során számos kihívással szembesültünk – a HPC infrastruktúrához való alkalmazkodástól a tokenizáló finomhangolásáig –, de ezek a nehézségek egyben mérföldkövekké is váltak.

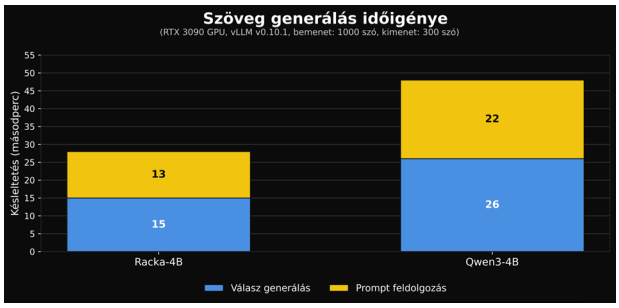
Bízunk benne, hogy munkánk megalapozza a jövőbeli egyetemi nyelvi modell-fejlesztéseket, inspirál hasonló kezdeményezéseket, és hogy a jövőben egyre több magyar LLM, azaz nyelvi modell csatlakozik a Racka „nyájához”.

A Racka-4B egy alacsony memória igényű, gyors, folyamatosan tovább tanított nagy nyelvi modell (LLM), amelyet az Eötvös Loránd Tudományegyetemen (ELTE) fejlesztettünk. Célunk az volt, hogy erősítsük a magyar nyelv jelenlétét a modellben, miközben megőrizzük annak versenyképességét a nagy nyelvi erőforrásokkal rendelkező nyelvek területén is, mint például az angol és a német. A modell a Qwen-3 4B architektúrára épül, amelyet paraméterhatékony folyamatos tovább tanítással LoRA-kat (Low-Rank Adaptation) használva adaptáltunk. Az alapmodell képességeit megtartva a Racka-4B képes szövegek generálására, csevegő (chat) és érvelő (reasoning) üzemmódban való működésre.

A nemzetközi kutatási trendeket követve modellünket a rackáról, egy magyar hosszúszőrű juhfajtáról neveztük el. Szerény mérete ellenére a Racka-4B az egyik legkisebb, magyarul is beszélgetésre képes nyelvi modell, ami már most is figyelemre méltó tudással rendelkezik. További optimalizálás és kvantálás után akár offline is futtatható lehet mobil eszközökön, megteremtve ezzel egy biztonságos, könnyen elérhető, drága hardverigény nélküli magyar nyelvű mesterséges intelligencia alkalmazásának lehetőségét.

A modell alapját egy 160 milliárd tokenből álló, többnyelvű adatbázis képezi, amelyet internetes forrásokból és gondosan válogatott anyagokból állítottunk össze. Az adathalmaz az alábbi arányban tevődik össze: 50% magyar, 20% angol, 20% német, 10% programkód. Ez az arány lehetővé teszi, hogy a magyar nyelv kiemelt szerepet kapjon, miközben csökkentjük a „katasztrofális felejtést", azaz a más nyelvek képességeinek elvesztését. Emellett a programkódok jelenléte révén megőrizzük a logikai következtetési képességeket is. Összesen 130 millió magyar, angol és német dokumentumot, valamint programkódot használtunk fel a modell tanításához. Ez a korpusz képviseli jelenleg a magyar nyelvre fókuszáló nyelvi modellezés egyik legkorszerűbb alapját.

Kutatásunk egyik legfontosabb gyakorlati eredménye a tokenizáló (szófelbontó rendszer) átdolgozása. A legtöbb többnyelvű tokenizáló ugyanis angolra van optimalizálva, emiatt a magyar szavakat túl sok kisebb egységre bontja, ami hosszabb szövegsorokat és gyengébb hatékonyságot eredményez. Ennek megoldására új tokenizálót tanítottunk kifejezetten magyar adatokon, majd ezt összefésültük a Qwen-3 tokenizálójával. Az így létrejött hibrid szótár 32 000 új, magyar-specifikus tokent tartalmaz, miközben megőrizte a latin ábécét, valamint az angol és német szavak tokenjeit is. A szótár-átviteli módszerrel az új tokenek beágyazását jelentéstartalmuk alapján inicializáltuk, ami már a kezdetektől javította a modell teljesítményét. Ennek eredményeképpen a Racka-4B gyorsabban és hatékonyabban dolgozza fel a hosszabb magyar szövegeket, valamint kevesebb nyelvtani hibát is vét.



A modellt Magyarország legnagyobb szuperszámítógépén, a Komondor HPC-n tanítottuk, 16 nodeon, 64 darab NVIDIA A100 (40 GB) GPU felhasználásával. A Fully Sharded Data Parallel (FSDP) helyett a Distributed Data Parallel (DDP) technológiát választottuk, mivel az jobban skálázódott a Komondor rendszerén. A tanítás 12 napig tartott, ami körülbelül 2 év és 1 havi GPU-időnek felel meg, ezalatt a modell egyszer ment végig az egész adathalmazon.

Az eredményeink megerősítették, hogy a tokenizáló-adaptáció és a forrásérzékeny tanítás egyaránt hatékony megközelítés. Az egy szóból képzett altokenek száma jelentősen csökkent, ami gyorsabb feldolgozást, alacsonyabb számítási igényt és jobb reprezentációs képességet eredményezett. A magyar nyelvű teljesítmény számottevően javult az alapmodellhez képest, miközben az angol és német nyelvi tudás megmaradt. A modell megőrizte érvelési és csevegési képességeit is, így egy sokoldalú, ígéretes magyar-központú LLM vált belőle. Ugyanakkor a javulás még mérsékelt a magyar specifikus alkalmazásokban, ami rávilágít a közepes erőforrású nyelvekhez való alkalmazkodás nehézségeire.

A jövőben célunk a hiperparaméterek szisztematikus optimalizálása, a magyar nyelvű NLP-benchmarkok szélesebb körű kiértékelése, valamint a felügyelt és preferenciaalapú finomhangolás a gyakorlati alkalmazásokhoz. Ha a magyar AI-szabályozás is lehetővé teszi, tervünk a modell és a tokenizáló nyílt forrásúvá tétele, ezzel erősítve a magyar nyelv digitális szuverenitását, mintát adva más, kisebb nyelvek számára is.

Köszönetnyilvánítás

A kutatást a Digitális Örökség Nemzeti Laboratórium támogatta. A 2022-2.1.1-NL-2022-00009 számú projekt a Kulturális és Innovációs Minisztérium támogatásával, a Nemzeti Kutatási, Fejlesztési és Innovációs Alap finanszírozásával valósult meg. A kutatást az EKOP-24 Egyetemi Kiválósági Ösztöndíjprogram támogatta, a Kulturális és Innovációs Minisztérium és az NKFI Alap finanszírozásában.

Köszönet illeti a Digitális Kormányzati Fejlesztési és Projektmenedzsment Zrt.-t, amiért hozzáférést biztosított a Komondor HPC szuperszámítógéphez.

Vegyipari kismolekulák titkos élete

Szerző: *Szőri Milán*

A Miskolci Egyetem Kémiai Intézetének Szuperszámítógépes Modellezés Kutatócsoport eredményeinek összefoglalása

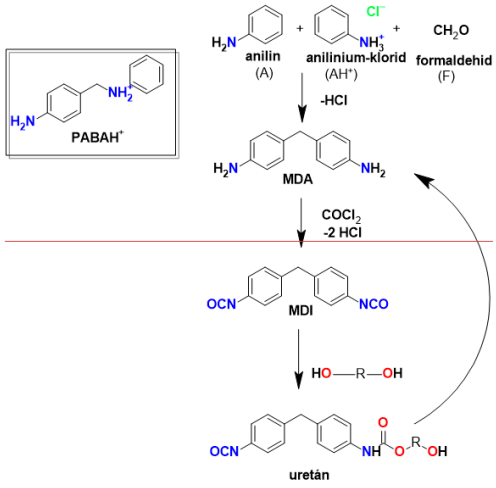
Az Észak-Magyarországi régió a magyar vegyipar egyik meghatározó központja: a hazai értékesítés mintegy ötödét adja, és itt működnek a poliuretán-gyártás kulcsfontosságú alapanyagainak nagyüzemi gyártói. A Miskolci Egyetem Kémiai Intézetének Szuperszámítógépes Modellezés Kutatócsoportja a térség ipari szereplőivel együttműködve vizsgálja a poliuretánok előállításának kémiai folyamatait, nagy pontosságú kvantumkémiai számítások segítségével. A poliuretánok olyan sokoldalú műanyagok, amelyek számos mindennapi termékben megtalálhatók, például bútorhabokban, hőszigetelő anyagokban és műbőr bevonatokban. A Komondor szuperszámítógépen végzett virtuális mérések molekuláris szinten tárják fel a reakciók mechanizmusát és energetikáját, ami segíti a hatékonyabb és fenntarthatóbb technológiák fejlesztését.

Poliuretán-gyártás reakciómechanizmusának vizsgálata

Magyarország vegyipari klaszterei közül az Észak-Magyarországi régió a nemzeti vegyipari értékesítés mintegy 20 százalékát adja. A térségben találhatók a poliuretán-gyártás két kulcsfontosságú komponensének nagyüzemi előállítói: a BorsodChem-Wanhua kazincbarcikai gyárában évi 350 000 tonna kapacitással termelik a 4,4′-metilén-difenil-diizocianátot (MDI, 1. ábra), míg a MOL tiszaújvárosi poliolüzeme évi 200 000 tonna poliéter-poliol kapacitással rendelkezik.

A Miskolci Egyetem Kémiai Intézetének Szuperszámítógépes Modellezés Kutatócsoportja számára a régió vegyiparához kapcsolódó kérdések kiemelt kutatási területet jelentenek, különös hangsúllyal az uretánok ipari előállításának reakciómechanizmusára, amelynek vizsgálata szoros együttműködést teremtett az egyetem és a vegyipar között.

A vegyipari technológiák fejlesztésének alapja a gyártás során végbemenő kémiai reakciók és energetikai viszonyaik pontos ismerete, amely kísérleti úton gyakran csak korlátozottan vizsgálható. A Komondor szuperszámítógépen végzett számításokon alapuló virtuális mérések molekuláris szinten értelmezik a makroszkopikus jelenségeket, elősegítve a folyamatok megértését, a gyökérokok feltárását és a hatékony tervezést, valamint hozzájárulva a körforgásos gazdaságot támogató technológiák kidolgozásához. Mivel a reakciók tényleges irányát alapvetően az dönti el, hogy a lehetséges reakcióutak közül energetikailag melyik a legkedvezőbb, a legvalószínűbb útvonalak azonosításához nagy pontosságú energiameghatározásra van szükség. Ennek érdekében G3MP2B3 és G4MP2 kvantumkémiai protokollokat alkalmaztunk, SMD oldószermodellel kiegészítve.



***1. ábra**. Lineáris gazdaságtól a körkörös gazdaságig: a poliuretángyártás főbb reakciólépései és az uretánlebontás.*

Kutatócsoportunk G3MP2B3 kvantumkémiai módszerrel feltérképezte az anilin–formaldehid rendszer nyolclépéses, protonkatalizált diaminképződési reakciómechanizmusát, amely az MDA kialakulásához vezet (1. ábra). Azonosítottuk a legvalószínűbb mellékreakciókat, és meghatároztuk kulcsfontosságú köztitermékek saverősségi állandóit. Kimutattuk, hogy a PABA protonált formája (PABAH+, 1. ábra) a legalacsonyabb energiájú intermedier, amely felhalmozódásával rontja a kitermelést, és szerkezeti instabilitása hozzájárul melléktermékek képződéséhez. A reakciót vizes és anilines közegben is vizsgáltuk; az energiaprofil alapján a vizes közeg mechanizmusa kedvezőbbnek bizonyult.

Az MDA kétféle reakcióúton alakítható MDI-vá (1. ábra): mindkét amincsoport foszgénezésével dikarbaminsav-diklorid intermedieren keresztül („foszgénezés először”), vagy egymásutáni izocianát kialakulásával járó mechanizmussal. Számításaink szerint az előbbi energetikailag kedvezőbb, az első foszgén-addíció pedig pszeudónegatív aktiválási energiával bír. A keletkező HCl amin-hidroklorid intermedier formában gátolhatja a további foszgénezést. A reakcióban szereplő köztitermékek, reaktánsok és termékek termokémiai adatait meghatároztuk gázfázisban, és javaslatot tettünk a fenilgyűrűhöz kapcsolódó NCO- és NHCOCl-csoportok entalpia hozzájárulására, amely a MDI-gyártás entalpiamérlegének becsléséhez használható.

Az izocianátok és poliolok reakcióját modellezve a fenil-izocianát–1-propanol rendszert vizsgáltuk (1. ábra). Az uretánképződés mechanizmusát alkohol- és izocianát-feleslegben egyaránt feltártuk G4MP2/SMD kvantumkémiai számításokkal, és izocianát-felesleg esetében új, allofanáton keresztüli kétlépéses mechanizmust javasoltunk. A reakciók aktiválási energiáit kísérletileg is meghatároztuk HPLC-vel, amelyek jól korreláltak a számítási eredményekkel.

Az uretánkötés kémiai stabilitása a poliuretánok újrahasznosításának fő akadálya, ezért egy modellvegyület (metil-N-fenil-karbamát, MPCate) bontási reakcióit vizsgáltuk G3MP2B3 kvantumkémiai protokollal. Hidrogénezés, hidrolízis, metanolízis, peroxidáció, glikolízis, ammónolízis, metilaminos redukció és dimetil-foszfit keresztüli uretánbontási reakciók energiaprofilját tártuk fel. Ígéretesnek bizonyult a H_2-, H_2O_2- és CH_3NH_2-vel való bontás, különösen vizes közegben, mivel viszonylag alacsony aktiválási energiával és exoterm folyamatokkal jártak. Az így képződő termékek (pl. anilin, dimetil-karbonát) iparilag értékesek, így illeszthetők a körforgásos gazdaság szemléletéhez. Vizsgáltuk az uretán modellvegyület ionos bontási útvonalait is, ahol az alkalikus hidrolízis mutatkozott a legkedvezőbbnek.

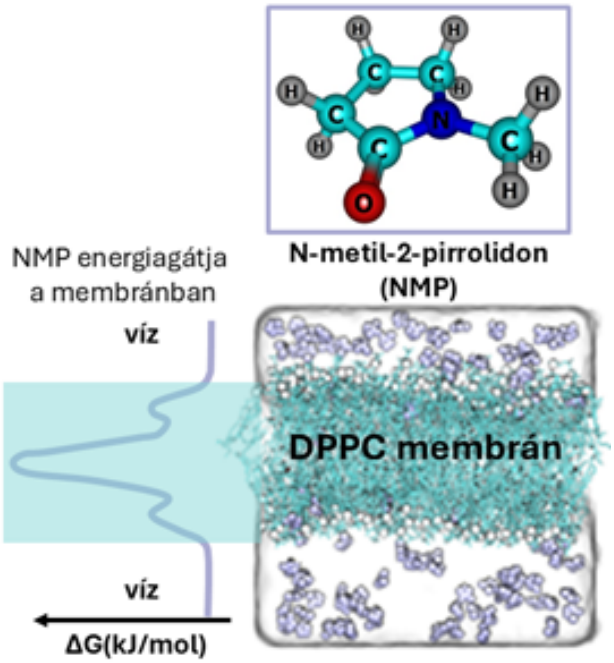
Biológiai membránok áteresztőképessége vegyipari kismolekulákkal szemben

A kutatócsoport nemcsak a vegyipari technológiák mögött álló kémiai reakciókat vizsgálja, hanem a vegyipar által előállított molekulák biológiai rendszerekkel való kölcsönhatásait is, amelyek potenciális környezeti kockázata felmérhető és becsülhető. Az ipari méretekben előállított vegyi anyagok szervezetbe jutása ugyanis jórészt feltáratlan – néhány riasztó adatot a 2. ábra is szemléltet –, pedig felhalmozódásuk súlyos egészségügyi problémákat okozhat. A vizsgálatok elmaradását sokszor az indokolja, hogy a specifikus analitikai eljárások kidolgozása nehézségekbe ütközik.



2. ábra. *A vegyipar globális termelése és a kémiai anyagok veszélyességének ismertségi szintje.*

A molekuladinamikai (MD) szimulációk hatékony módszerként szolgálnak a környezetbe kikerülő szennyező anyagok biomembránokon való átjutásának vizsgálatára. Négy jellegzetes vegyipari kismolekulát tanulmányoztunk: a 1,4-dioxánt (széles körben alkalmazott oldószer), a fenolt (amely nagy mennyiségben fordul elő az izocianát-gyártás szennyvizeiben), az oxánt (ígéretes következő generációs üzemanyag-komponens), valamint az N-metil-2-pirrolidont (NMP, lítiumakkumulátor-gyártásban alkalmazott oldószer). Ahogy az NMP példáján keresztül a 3. ábra is mutatja, a molekulák passzív transzportját részletesen elemeztük, különös tekintettel a membránszerkezeti jellemzőkre és az átjutás energetikai viszonyaira.



3. ábra. *Az N-metil-2-pirrolidon (NMP) szerkezete, DPPC membránban történő beoldódásának szabadenergia-profilja és a hozzá tartozó jellemző konfiguráció.*

Szemben a NMP-vel, az oxán felhalmozódása a membrán kettősrétegének közepén termodinamikailag a legkedvezőbb, míg a dioxán elősegíti a vízmolekulák bejutását a membrán belsejébe. Világossá vált, hogy a vizsgált DPPC, DPPE, DPPA és DPPG membránok lipidfejcsoportjainak kémiai jellege – polaritása, mérete és töltése – alapvetően meghatározza, hogy a vizsgált vegyületek milyen mélyre épülnek be, illetve milyen szerkezeti változásokat idéznek elő. Bár egyik molekula esetében sem találtunk magas szabadentalpia-gátat az átjutás folyamatához, a fenol és az oxán különösen problémásnak bizonyult, mivel jelentős szerkezeti zavarokat okoztak a membránban. Ez arra utalhat, hogy az emberi sejtek membránintegritását is károsíthatják.

https://cefic.org/landscape-of-the-industry/hungary/?utm (elérés: 2025.09.26.)

https://borsodchem.com/uploads/documentitem/30/bor-sustainability-report-2021-2022-eng-1707925684.pdf (elérés: 2025.09.26.)

https://molgroupchemicals.com/en/news/press-releases/mols-1-3-billion-polyol-complex-in-tiszaujvaros-inaugurated (elérés: 2025.09.26.)

https://fiek.uni-miskolc.hu/ (elérés: 2025.09.26.)

Boros, R. Z. Z.; Farkas, L.; Nehéz, K.; Viskolcz, B.; Szori, M. An Ab Initio Investigation of the 4,4′-Methlylene Diphenyl Diamine (4,4′-MDA) Formation from the Reaction of Aniline with Formaldehyde. Polymers (Basel). **2019**, 11, 398. https://doi.org/10.3390/polym11030398.

Boros, R. Z.; Koós, T.; Wafaa, C.; Nehéz, K.; Farkas, L.; Viskolcz, B.; Szőri, M. A Theoretical Study on the Phosgenation of Methylene Diphenyl Diamine (MDA). Chem. Phys. Lett. **2018**, 706, 568–576. https://doi.org/10.1016/j.cplett.2018.06.024.

Cheikh, W.; Rózsa, Z. B.; López, C. O. C.; Mizsey, P.; Viskolcz, B.; Szori, M.; Fejes, Z. Urethane Formation with an Excess of Isocyanate or Alcohol: Experimental and Ab Initio Study. Polymers (Basel). **2019**, 11, 1543. https://doi.org/10.3390/polym11101543.

Horváth, T.; Kecskés, K.; Jordán Csábrádiné, A.; Szőri-Dorogházi, E.; Viskolcz, B.; Szőri, M. Searching for the Achilles' Heel of Urethane Linkage—An Energetic Perspective. Polymers (Basel). **2024**, 16, 1126. https://doi.org/10.3390/polym16081126.

Rózsa, Z. B.; Horváth, T.; Viskolcz, B.; Szőri, M.; Influence of Membrane Composition on the Passive Membrane Penetration of Industrially Relevant NSO-Heterocycles Int. J. Mol. Sci. **2025**, 26, 7427; https://doi.org/10.3390/ijms26157427.

Rózsa, Z. B.; Horváth, Thangaraj, R.; Viskolcz, B.; Szőri, M. Foreseeing the future of green Technology. Molecular dynamic investigation on passive membrane penetration by the products of the CO2 and 1,3-butadiene reaction J. Mol. Liquids, **2022**, 361, 119581 https://doi.org/10.1016/j.molliq.2022.119581.

Rózsa, Z. B.; Szőri-Dorogházi, E.; Viskolcz, V.; Szőri, M. Transmembrane penetration mechanism of cyclic pollutants inspected by molecular dynamics and metadynamics: the case of morpholine, phenol, 1,4-dioxane and oxane. Phys. Chem. Chem. Phys., **2021**, 23, 15338-15351.

HIV és SARS-CoV-2 vírus fehérjék gátolhatóságának nyomában

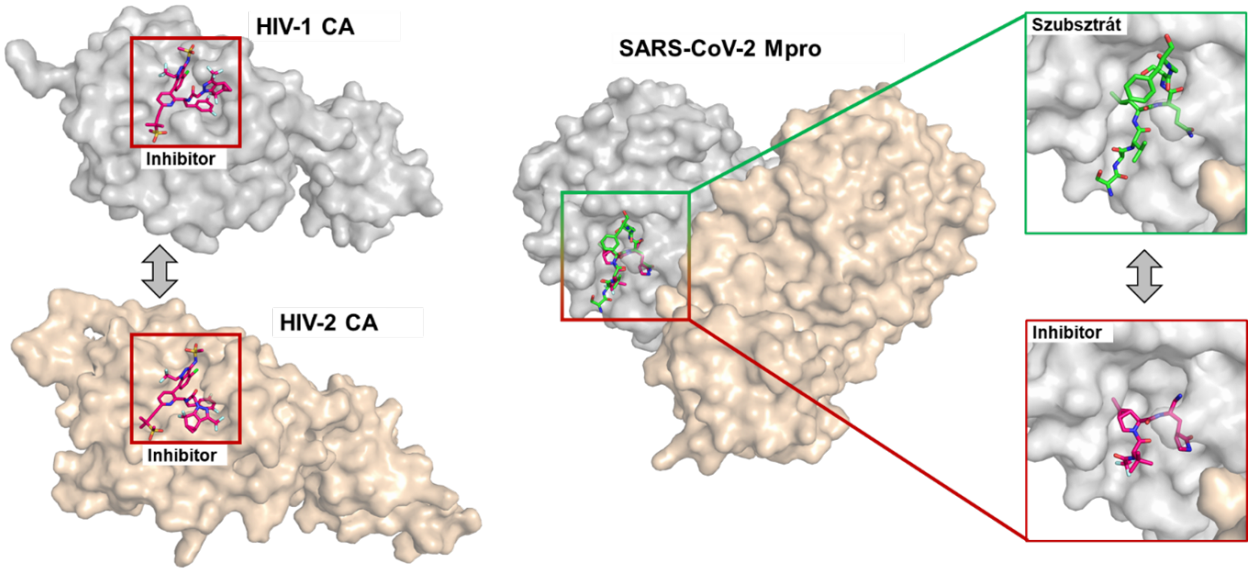
Szerző: *Dr. Mótyán János András*

A Retrovirális Biokémiai Kutató Laboratóriumban végzett kutatásaink célja többek között vírusfehérjék biokémiai és enzimológiai tanulmányozása kísérletes és számítógépes eljárások alkalmazásával. A Komondor erőforrásait több különböző fehérje számításos kémiai módszerekkel történő vizsgálatához is felhasználtuk. Ezek közül szeretnénk röviden bemutatni két, egészségügyi szempontból is jelentős vírus, a humán immundeficiencia vírus (HIV) és a súlyos akut légzőszervi szindróma-koronavírus 2 (SARS-CoV-2) fehérjéinek vizsgálatát. Mindkét vírus fehérjéi kiemelt gyógyszercélpontnak számítanak, ezért vizsgálatuk a gátlószerek működésének megértése, a hatékony(abb) gátlószerek fejlesztése és a gyógyszerekkel szemben kialakuló rezisztencia szempontjából is jelentőséggel bír.



A HIV a legismertebb vírusok egyike, mely a szerzett immunhiányos tünetegyüttes (AIDS) kórokozója. Két fő típusa ismert, a HIV-1 és a HIV-2. Világszerte mintegy 40 millió HIV-fertőzöttet tartanak számon. A fertőzéses esetek túlnyomó részéért a HIV-1 vírus felelős, ehhez képest a HIV-2 által okozott fertőzéses esetek száma jelentősen kisebb, a több mint egy évtizeddel ezelőtti becslések alapján a HIV-2 fertőzöttek száma körülbelül 2-3 millió lehet. Ennek hátterében jelentős részben a HIV-1 és HIV-2 közti különbségek állnak, ugyanis a HIV-2 vírus esetében a szervezetet érő vírusterhelés kisebb mértékű, a fertőzést követő látenciaidő hosszabb és az AIDS kialakulásának üteme is lassabb. Mivel a fertőzések túlnyomó részéért, így a világméretű járványért elsősorban a HIV-1 tehető felelőssé, ez a típus sokkal részletesebben tanulmányozott, míg a HIV-2-ről jelentősen kevesebb információval rendelkezünk. A vírusfehérjék a fertőzöttek kezelésében alkalmazott gyógyszerek kiemelt célpontjai, a HIV-vírusok fehérjéi is a legalaposabban tanulmányozott molekulák közé tartoznak. Bár ezidáig számos gyógyszert fejlesztettek ki a HIV-vírusfertőzés kezelésére, a gyógyszermolekulák vizsgálatára és fejlesztésére az igény folyamatos, hiszen a vírusok könnyen rezisztenssé válhatnak ezekkel a szerekkel szemben. A HIV-vírus-fertőzés kezelésére a terápiás gyakorlatban alkalmazott gyógyszereket jellemzően a HIV-1 vírus ellen fejlesztették ki. A HIV-2 kevésbé tanulmányozott, ezért sok esetben ezeknek a szereknek a hatékonysága a HIV-2 esetében nem, vagy nem kellő részletességgel ismert.

Kutatásaink célja annak összehasonlítása volt, hogy a HIV-1 gátlására kifejlesztett szerek milyen hatékonysággal képesek a HIV-1 és HIV-2 vírusfehérjék működését gátolni. Ehhez kísérletes és számításos módszereket egyaránt felhasználtunk, a fehérjék és a gyógyszerek közti kölcsönhatások atomi szinten történő vizsgálatához pedig a Komondor erőforrásait vettük igénybe. A szuperszámítógép által nyújtott nagy számításos kapacitás lehetővé tette számunkra, hogy a vírusfehérjéket hosszú-idejű szimulációkkal, magas számítási igényű szoftverekkel vizsgáljuk, részletesebb képet kapva a fehérjék viselkedéséről. Kutatócsoportunk a HIV-vírusok két különböző fehérjéjét, az úgynevezett integrázt és kapszidot célzó gátlószereket is tanulmányozta. Kísérletes vizsgálatokkal kimutattuk, hogy a vizsgált, kereskedelmi forgalomban kapható gátlószerek képesek mindkét vírus fehérjéinek működését gátolni, akár összehasonlítható mértékű hatékonysággal. A kísérletes megfigyeléseket a szerkezeti vizsgálatok is alátámasztották, amelyek alapján a vizsgált, HIV-1 ellen kifejlesztett gyógyszermolekulák a HIV-1 és a HIV-2 integráz és kapszid fehérjéihez hasonló módon képesek kötődni, ami magyarázhatja hatékonyságukat a HIV-2 fehérjék gátlásában [**Kiarie és mtsai, 2024**].



Enzim-inhibitor komplex szerkezetek vizsgálata.
Bal oldal: a lenacapavir nevű kapszid inhibitor kölcsönhatása a HIV-1 és HIV-2 kapszid (CA) fehérjékhez. Jobb oldal: a SARS-CoV-2 fő proteáz (Mpro) szerkezete, valamint az enzim aktív helyére kötődő szubsztrát és inhibitor kötődési módjának összehasonlítása.

A SARS-CoV-2 által okozott első fertőzéses eseteket 2019-ben azonosították, az azóta eltelt idő alatt közel 800 millió fertőzéses esetet regisztráltak világszerte. A világméretű járvány 2020-ban történő kirobbanását követően kutatócsoportunk úgy döntött, hogy a rendelkezésre álló kísérletes és számítógépes eljárásokat felhasználja a SARS-CoV-2 fehérjék vizsgálatára is, ezért új kutatásokba kezdtünk, melyeket azóta is folyamatosan végezzük. Vizsgálataink legfontosabb célpontjai a SARS-CoV-2 vírus gyógyszercélpontként szolgáló vírusfehérjéi, a tüske fehérje, valamint a fő proteáz, különös tekintettel a gyógyszermolekulák ellen kialakuló rezisztencia vizsgálatára [**Mahdi és mtsai, 2025**]. Ehhez felhasználtuk a kutató laboratórium vírusfehérjék, köztük a HIV-1 fehérjék vizsgálata során szerzett évtizedes tapasztalatait. A HIV-1 és SARS-CoV-2 vírusok összehasonlításához a fehérjék és gyógyszermolekulák közti kölcsönhatások feltérképezését számítógépes eljárásokkal végeztük el, ami a rezisztencia kialakulása mögött álló mechanizmusok alapvető hasonlóságait tárta fel a két vírus esetében. Érdekesség, hogy a terápiás felhasználásra engedélyezett, a SARS-CoV-2 proteázt célzó gyógyszer egyik hatóanyagát korábban a HIV-1 proteáz gátlására fejlesztették ki. A vizsgálatok eredményeként azonosítottuk a fehérje azon részeit, melyek a gyógyszermolekula (nirmatrelvir) kötéséért felelősek, és amelyeknek megváltozása a vírus evolúciója során, azaz az újonnan megjelenő vírus-variánsokban potenciálisan hozzájárulhat a rezisztencia kialakulásához [**Mótyán és mtsai, 2022**]. Ezen eredményeink további, jelenleg is folyamatban lévő kutatásainkat alapozták meg, melynek keretében részletesebben kívánjuk tanulmányozni a rezisztencia kialakulásának molekuláris mechanizmusát. A szuperszámítógépen futtatott számításokkal kapott adatok pontosságát az egyes variánsok esetében végzett kísérletes vizsgálatok eredményével kívánjuk összehasonlítani. A cél egy olyan eljárás kidolgozása, mely megbízható módon használható új, potenciálisan rezisztens variánsok azonosítására és jellemzésére, valamint segítheti hatékony gátlószerek tervezését is.

A Komondor számítógépes kapacitásait sikerrel alkalmaztuk más fehérjék tanulmányozása során is, így a HIV és SARS-CoV-2 vírusok fehérjéi mellett számításos kémiai és szerkezeti bioinformatikai módszerekkel vizsgáltuk más, egészségügyi és biotechnológiai szempontból jelentős vírusok fehérjéit [**Hoffka és mtsai, 2023**] [**Golda és mtsai, 2024**], de a szuperszámítógépen futtatott programokat használtuk emberi emésztőenzim vizsgálatában [**Hoffka és mtsai, 2024**] és mesterségesen előállított enzimfehérje tervezésében is [**Listov és mtsai, 2025**].

Golda M, Hoffka G, Cherry S, Tropea JE, Lountos GT, Waugh DS, Wlodawer A, Tőzsér J, Mótyán JA. P1′ specificity of the S219V/R203G mutant tobacco etch virus protease. Proteins. 2024; 92(9): 1085-1096.

Hoffka G, Lountos GT, Needle D, Wlodawer A, Waugh DS, Tőzsér J, Mótyán JA. Self-inhibited State of Venezuelan Equine Encephalitis Virus (VEEV) nsP2 Cysteine Protease: A Crystallographic and Molecular Dynamics Analysis. J Mol Biol. 2023; 435(6):168012.

Hoffka G, Mhana S, Vas M, Toldi V, Mótyán JA, Tőzsér J, Szabó A. Modeling protease-sensitive human pancreatic lipase mutations in the mouse ortholog. J Biol Chem. 2024; 300(10): 107763.

Kiarie IW, Hoffka G, Laporte M, Leyssen P, Neyts J, Tőzsér J, Mahdi M. Efficacy of Integrase Strand Transfer Inhibitors and the Capsid Inhibitor Lenacapavir against HIV-2, and Exploring the Effect of Raltegravir on the Activity of SARS-CoV-2. Viruses. 2024; 16(10): 1607.

Listov D, Vos E, Hoffka G, Hoch SY, Berg A, Hamer-Rogotner S, Dym O, Kamerlin SCL, Fleishman SJ. Complete computational design of high-efficiency Kemp elimination enzymes. Nature. 2025; 643(8074): 1421-1427.

Mahdi M, Kiarie IW, Mótyán JA, Hoffka G, Al-Muffti AS, Tóth A, Tőzsér J. Receptor Binding for the Entry Mechanisms of SARS-CoV-2: Insights from the Original Strain and Emerging Variants. Viruses. 2025; 17(5):691.

Mótyán JA, Mahdi M, Hoffka G, Tőzsér J. Potential Resistance of SARS-CoV-2 Main Protease (Mpro) against Protease Inhibitors: Lessons Learned from HIV-1 Protease. Int J Mol Sci. 2022; 23(7): 3507.

Belepillantás a bennünk élő mikrobák univerzumába: a bélrendszer ökoszisztémájának nagyszabású számítógépes szimulációi

Szerzők: :
Jakab Máté, Boza Gergely, Scheuring István

Az emberi bélrendszerben élő mikroorganizmusok közössége, a bélmikrobiom, rendkívül összetett ökoszisztéma, amely alapvetően befolyásolja a gazdaszervezet egészségét és anyagcseréjét. A mikrobák kulcsszerepet játszanak az olyan táplálék-összetevők lebontásában, amelyeket az emberi szervezet önmagában nem képes feldolgozni, miközben vitaminokat és más fontos molekulákat is termelnek. Kutatócsoportunk egy új számítógépes modellt fejlesztett ki a mikrobiális közösségek működésének és fejlődésének vizsgálatára, amely figyelembe veszi a fajok közötti anyagcsere-kapcsolatokat és az eltérő lebontási stratégiákat. A szimulációk szerint a változatosabb tápanyag-összetétel gazdagabb és stabilabb mikrobiális közösségek kialakulását segíti elő. A modell hozzájárulhat ahhoz, hogy jobban megértsük a mikrobiom működését, és a jövőben támogassa a személyre szabott orvoslás és a mikrobiom-alapú terápiák fejlesztését.

Az emberi és általában a gerincesek bélrendszerében élő mikrobák közössége, a bél mikrobiomja, egy rendkívül összetett ökoszisztéma, amely több milliárd mikroorganizmust és több száz fajt foglal magába. Ezek az apró élőlények alapvetően befolyásolják gazdaszervezet egészségét, viselkedését és általános állapotát. A bélrendszer térben tagolt, sokrétű környezetében számtalan mikroba lép kölcsönhatásba egymással – gyakran olyan fajok is, amelyekről még alig tudunk valamit. A gazdaszervezet emésztése során a táplálék olyan összetett anyagait, polimereit, amelyeket a gazdaszervezet nem képes lebontani, a mikrobák közössége dolgozza fel – különösen a vastagbélben –, több lépésből álló lebontási láncolatok révén. Az így keletkező kisebb molekulákat (például oligo-, di- és monomereket) részben a mikrobák maguk hasznosítják, de a gazdaszervezet is profitál ezekből az anyagokból. Emellett számos mikrobafaj vitaminokat és egyéb, a gazda egészsége szempontjából létfontosságú vegyületeket termel. Ennek a rejtett ökoszisztémának a szerkezetét, a stabilitását és működését fenntartó mechanizmusok megértése hatalmas tudományos kihívás. Bár egyre több kutatás tárja fel a mikrobiom és az egészség kapcsolatát, a rendszer dinamikájának mélyebb megértése még gyerekcipőben jár. Kutatócsoportunk ezért új modellezési módszereket fejleszt, hogy megragadjuk és leírjuk e folyamatok összetett dinamikáját.

Egy új modell születése

A tudományos szakirodalom átfogó elemzése és a meglévő modellek hiányosságainak feltárása után olyan új modellt kezdtünk kidolgozni, amely képes megragadni a biológiai valóság fontos, de gyakran figyelmen kívül hagyott elemeit. Célunk az volt, hogy a bélmikrobiom sokirányú összetettségét valósághűen, mégis egyszerűen kezelhető módon jelenítsük meg, s így bepillantást nyerhessünk a valós folyamatokba.

Az ilyen típusú modellek segítségével pontosabban megérthetjük, hogyan befolyásolja az étrend a mikrobiom összetételét, illetve hogyan épül fel újra egy ilyen közösség antibiotikum-kezelés után – mindezt számítógépes szimulációk révén. A legnagyobb kihívást az jelentette a modell tervezése során, hogy miképpen építsük be az anyagcsere-folyamatok, lebontási láncolatok bonyolult hálózatát úgy, hogy az egyszerre legyen általánosan érvényes és biológiailag hiteles. Hosszú egyeztetések és nemzetközi szakmai konzultációk után sikerült olyan megoldást találnunk, amely mind tudományosan, mind gyakorlati szempontból kielégítő. A modell fejlesztését is nagymértékben inspirálta a H2020 FindingPheno projekt keretében zajló, évek óta tartó interdiszciplináris együttműködésünk, a sok beszélgetés elméleti és kísérleti biológusokkal, adatkutatókkal és alkalmazott matematikusokkal.

Hogyan működik a modell?

Modellünk a bélmikrobiom fejlődését szimulálja, mégpedig úgy, hogy a különböző mikrobafajokat egymás után, „inváziós" eseményként vezeti be a rendszerbe. A fajok közös erőforrásokat fogyasztanak, miközben melléktermékeket bocsátanak ki a környezetbe – ezek a melléktermékek pedig más mikrobák számára jelenthetnek új tápanyagforrást. Így alakul ki a metabolitokon keresztüli kölcsönhatások bonyolult hálózata.

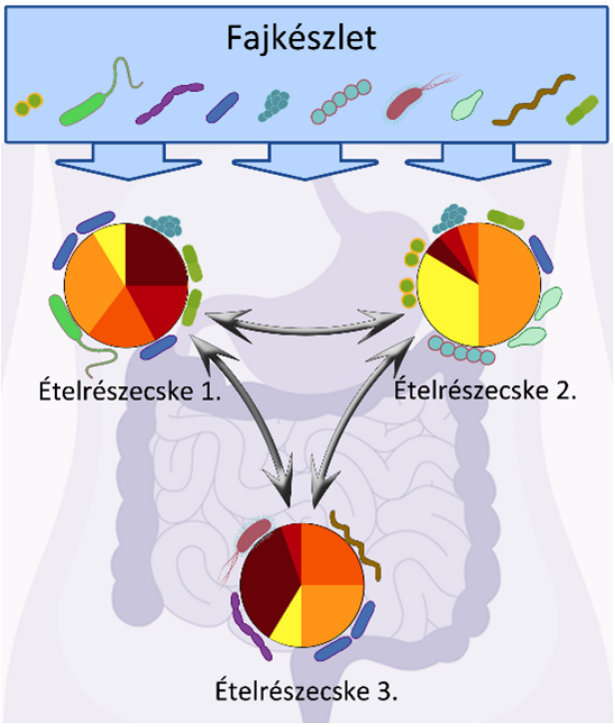
Egyik fő újításunkként a metabolitok hierarchikus rendszerét tételezzük fel. Minden erőforráshoz adott energiatartalmat rendelünk, melyet a mikrobák felhasználhatnak. További újításunk, hogy a biológiai realitáshoz illeszkedve a fajokat generalista vagy specialista lebontók közé soroljuk, ahol a generalisták nagyobb génkészletének magasabb költségét is figyelembe vesszük. A szimulációban a fajok csak lebontó (katabolikus) reakciókat végezhetnek – vagyis összetett anyagokat bontanak egyszerűbbé, miközben az így nyert energiát anyagcseréjükhöz használják fel. Azok a fajok, amelyek elegendő energiát képesek nyerni a fennmaradáshoz elszaporodhatnak, míg a többiek fokozatosan eltűnnek a rendszerből.

A modell további fontos eleme, hogy a teljes bélrendszert több, részben elkülönülő helyi közösségre bontjuk. Ezek a közösségek az emésztőrendszerben található ételrészecskéket szimbolizálják, amelyeken a mikrobák megtelepedhetnek és amelyeket bonthatnak. Minden helyi közösség saját kezdeti tápanyagkészlettel rendelkezik, míg az egyéb anyagok, melléktermékek csak a mikrobák lebontó tevékenysége során jelennek meg. A fajok vándorolhatnak a közösségek között, az erőforrások azonban csak a helyi közösségben érhetőek el.

Ezzel a modellrendszerrel jelenleg azt vizsgáljuk, hogy miként befolyásolja az eltérő tápanyag-összetétel a mikrobák fajgazdagságát. Ennek a számszerűsítéséhez kiszámítottuk a helyi közösségek tápanyagprofiljainak kompozíciós diverzitását. Ugyanezzel az mérőszámmal, illetve a klasszikus Shannon-diverzitási mutatóval értékeljük a mikrobiális közösségek sokféleségét is, így több szinten is követni tudjuk a faji diverzitás alakulását. A kompozíciós diverzitás a helyi közösségek közötti eltérésekre, míg a Shannon diverzitás a teljes rendszer egyesített fajkészletének alakulására érzékeny.

Eredményeink szerint minél változatosabb a táplálékösszetétel, annál sokszínűbb a mikrobiális közösség, úgy, hogy eközben a specialista és generalista fajok stabilan együtt tudnak élni. Ez az egyszerű, de alapvető megfigyelés alátámasztja modellünk hitelességét, és segít a további fejlesztések irányának kijelölésében.

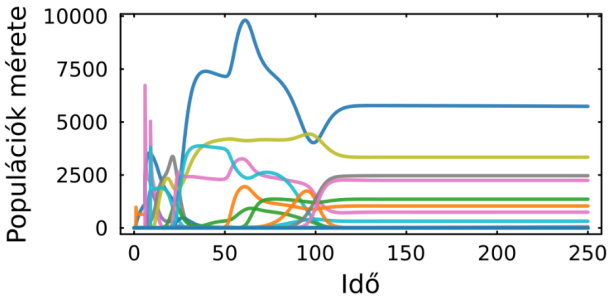
Célunk, hogy kísérleti úton is teszteljük a modell előrejelzéseit a jövőben, nemzetközi partnereinkkel együttműködve, akik laboratóriumi körülmények között tudják vizsgálni a bél mikrobiomjának működését. Az ilyen közös kutatások lehetővé teszik, hogy a szimulációk eredményeit valódi biológiai adatokkal vessük össze, és tovább finomítsuk a modellt.



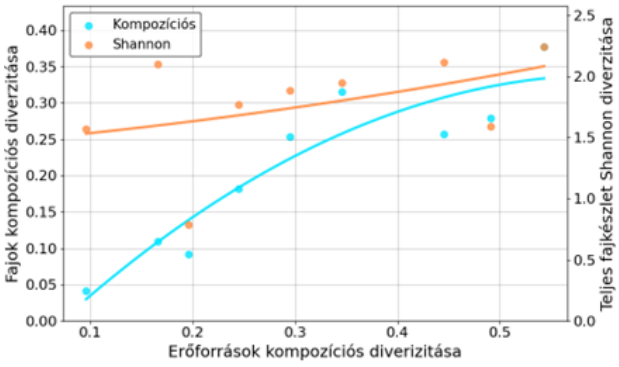
1 ábra. *A modell felépítésének áttekintése. A helyi közösségek, vagyis az ételrészecskék a fajok vándorlása révén kapcsolódnak egymáshoz (szürke nyilak). Minden helyi közösség eltérő erőforrás-összetétellel rendelkezik, amit a különböző színű kördiagramok szemléltetnek. Az inváziós fajok egy közös fajkészletből véletlenszerűen kiválasztva kerülnek rá a táplálékforrásokra.*

A jövő orvoslása és a mikrobiom

A jövő orvostudományában egyre nagyobb figyelem fog irányulni a mikrobiom irányába, mely vélhetően kulcsszerepet fog játszani a személyre szabott és precíziós gyógyászatban, és a bélbetegségek gyógyításában. A mikrobaközösségek megértésével és irányításával az orvosok a jövőben nemcsak a tüneteket kezelhetik, hanem helyreállíthatják a szervezet mikrobiális egyensúlyát is. Az új terápiás megközelítések között szerepelhetnek személyre szabott probiotikumok, mikrobiom-transzplantációk, valamint olyan étrendi beavatkozások, amelyek célzottan alakítják a bélrendszer ökoszisztémáját a testi és lelki egészség megőrzése érdekében.



2 ábra. *A fajok egyedszámainak alakulása az idő függvényében egy szimulált bélrendszerben. Minden vonal egy-egy faj időbeli változását mutatja. Az inváziós események a 100-as időpontnál leállnak. Ezt egy relaxációs időszak követi.*



3 ábra. *A fajdiverzitás alakulása a tápanyag összetétel diverzitásának függvényében. A szimulációkat addig futtattuk, amíg a rendszer elérte az egyensúlyi állapotot, miután 100 fajt adtunk hozzá. Az így kapott egyensúlyi egyedszámok alapján számítottuk ki a kompozíciós diverzitás és a Shannon-diverzitási mutatókat. A grafikonon látható pontok tíz ismételt szimuláció átlagértékei azonos paraméterek mellett.*

Komondor a kutyák genetikai vizsgálatának szolgálatában

Szerzők:
Jónás Dávid, Bucska Tamara

Tagok:
Jónás Dávid, Rékasi Zsófia, Szobonya Hanga, Tfirst Dóra, Bucska Tamara, Morvay Bence, Tarcsa Virág

Jónás Dávid bioinformatikus az ELTE Etológia Tanszékén dolgozik és több, mint 15 éves tapasztalattal rendelkezik a genetikai adatok elemzésének területén. Foglalkozott mesterséges szelekcióval, pedigré- és teljes genom adatok elemzésével, RNS- és ChIP-szekvenálási kísérletekkel, géntérképezéssel és CNV-detektálással egyaránt.

Jónás Dávid jelenleg az ELTE egyetemi adjunktusaként az öregedés genetikai hátterét vizsgálja kutyákban kutatócsoportjával, melyben szakdolgozók, TDK- és doktorandusz hallgatók egyaránt vannak. Minden hallgatói téma valamilyen kutatáshoz is kapcsolódik, melyeket a Komondor szuperszámítógépen végzik a hallgatók.

Kutatócsoportunk a kutyák (Canis familiaris) genetikájának vizsgálatára fókuszál a domesztikációtól a különböző tulajdonságok és betegségek genetikai hátterén át, egészen az öregedés genetikájáig. Célunk, hogy mélyebb rálátást szerezzünk a kutyák biológiai öregedésének folyamatára, illetve, hogy a kutyákat az öregedés kutatásának tágabb perspektívájába helyezzük.

A kutya, mint modell

Az öregedést megannyi aspektusban lehet vizsgálni klasszikus laboratóriumi modellállatokon, de az emberre jellemző komplex öregedési folyamatot nem. Az öregedést egyidejűleg több faktor is befolyásolja: genetikai összetevők – gének, szabályozó elemek stb. –, környezeti hatások és ezek többszintű interakciói, amelyek együttes hatását nem lehet modellezni loboratóriumi körülmények között.

Ezért szükséges egy megfelelő modellállat, mielőtt a klasszikus laboratóriumi állatmodelleken végzett kísérletek eredményeit és abból megszerzett tudást felhasználhatnánk és következtetéseket vonhatnánk le az emberre (például orvosbiológiai kutatások során), és a kutyák mint társállatok több okból is ideális választások (Kaeberlein et al., 2016).

Komondorok főszerepben

Összesen 1,5 TB nyers genetikai adatot állítottunk elő 745 mintából, és további 4 TB-nyi adatot töltöttünk le nyilvános szekvencia-adatbázisokból, amelyek több ezer kutyától származtak. Ezen adatok tárolása, feldolgozása, és elemzése nem lenne lehetséges szuperszámítógépek bevonása nélkül, így a Komondort, Magyarország legnagyobb teljesítményű szuperszámítógépét használtuk a kutatásainkhoz.

Számos kísérlet – egy cél

Az elmúlt években nagy számú genetikai/bioinformatikai kutatást végeztünk a Komondoron. Genetikai módszerek széles skáláját alkalmaztuk, mint például genomszintű asszociációs vizsgálatokat, összehasonlító genetikai vizsgálatokat, teljes genom- és RNS szekvenálást, vagy alacsony és közepes sűrűségű SNP array adatok imputációját teljes-genom szintre, csak hogy egy párat említsünk.

Ezen módszerek és eszközök felhasználásával vizsgáltuk extrém idős kutyák DNS-ét (melyeket úgy definiáltunk, mint azon kutyák, amelyek legalább 1,5-ször annyit éltek, mint fajtájuk átlaga), és több tízezer új egynukleinsavas polimorfizmust (SNP), rövid indel mutációt, valamint kópiaszám-változásokat azonosítottunk, amelyek potenciálisan hatással lehettek ezen kutyák extrém hosszú és egészséges élettartamához. Előzetes kutatásunk során publikáltunk ~80000 korábban ismeretlen SNP mutációt extrém idős kutyákban, melyek között 670 misszensz mutáció volt összesen 462 gének (Jónás et al., 2020); jelenleg a teljes adathalmazon végzett kísérlet eredményeiből készül egy tudományos közlemény.

RNS szinten az egészséges öregedéshez köthető genetikai markereket kerestünk agy- és vérszövetmintákból. Agyszövetben azonosítottunk 3436 különbözően expresszálódó gént a prefrontális kéreg területén (Sándor et al., 2024), de nem tudtunk azonosítani hasonló markereket egy könnyebben vizsgálható szövettípusból, a vérből (Jónás et al., 2025). Emellett leírtunk 204 olyan gént, melyben a kronológiai korral változó alternatív splicing hatására eltolódást mutatott a keletkező transzkriptek egymáshoz viszonyított aránya az agy prefrontális kéregállományában (Jónás et al., In prep.).

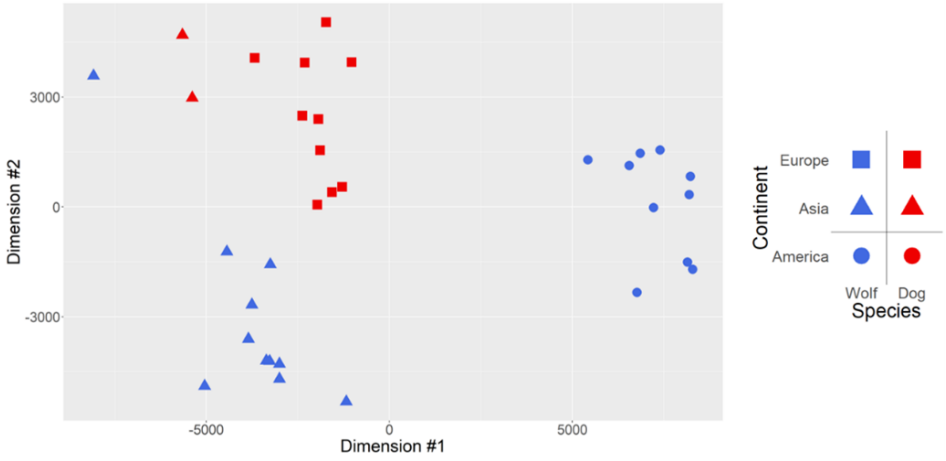
A DNS és RNS szintű eredményeink mind azt mutatják, hogy a génexpresszió szabályozása fontos szerepet játszik az egészséges öregedésben. Több misszensz mutációval rendelkező gén kapcsolatban állt a génexpresszió regulációjával, míg RNS szinten megfigyeltük, hogy a fiatal és idős kutyák között eltérő génexpressziós szintet mutató gének döntő többségénél a kifejeződés nagysága változott és nem lettek a gének teljesen ki/bekapcsolva. Az RNS szekvenáláshoz kapcsolódó génontológiai elemzés is azt mutatja, hogy a 3436 szignifikáns gén közül több génonkológiai terminusa kapcsolatban állt genetikai regulációs mechanizmusokkal. Jelenleg prefrontális kéregből származó RNS szekvenálási adatokat hasonlítunk össze klasszikus laboratóriumi modellállatokban, kutyában, rókában és emberben. A hipotézisünk szerint az általános génexpressziós profil kutyákban nagyobb hasononlóságot mutat az emberrel, mint a laboratóriumi egerek vagy patkányok, mely legalább részben a hasonló környezeti hatásokkal indokolható. Emellett DNS-szekvencia adatok segítségével szintén kutatjuk a régóta fennálló hipotézist, miszerint a retrotranszpozon aktivitás növekszik az életkorral.

A farkas és kutya genomszintű génkifejeződési profiljainak összehasonlításával vizsgáltuk a kutya evolúcióját is, és 1567 szignifikánsan eltérő szinten kifejeződő gént azonosítottunk a vérmintákból a két vizsgált faj között (1. ábra). Továbbá több tulajdonság genetikai hátterét is vizsgáltuk genomszintű asszociációs tanulmányok segítségével: kutyák szótanulásának képességét[1] (2. ábra), a dermedési viselkedést laboratóriumi állatokban – mely egy ismert védekező

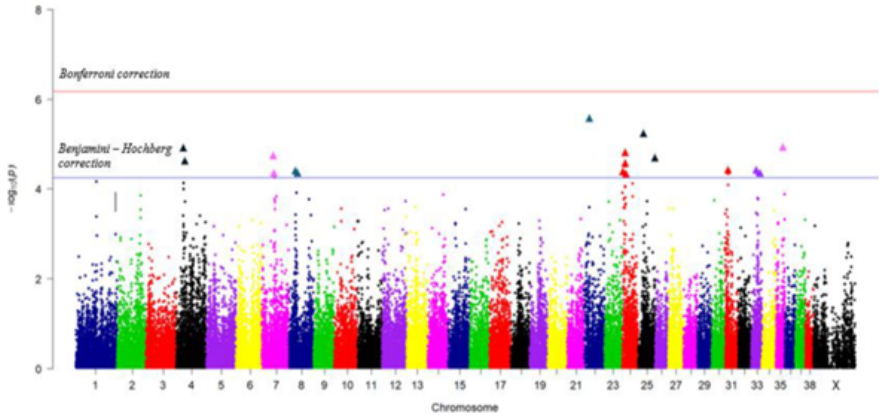
[1] *Ebben a kontextusban a szótanulás tárgynevek megtanulására vonatkozik és nem cselekvés-orientált parancsszavak megtanulására.*

mechanizmus az állatvilágban –, és a kutyák szaglási teljesítményét is, ami az egyik legfontosabb érzékszerve a kutyáknak, melyet használnak például vadászat vagy kommunikáció során is. Egy új kutatás keretein belül tervezzük azonosítani a dermedési viselkedés okozó mutációját, melyhez 50 további kutya teljes genom szekvenálása zajlik.

Végezetül, a Komondort csoportunk használja a felsőoktatásban is. Összesen tíz BSc és MSc szakdolgozathoz (5 végzett hallgató, 5 folyamatban van) és 3 TDK dolgozathoz (1 végzett: ELTE Természettudományi Kar (Bioinformatika szekció): I. helyezett és különdíj; 2 folyamatban van) is a Komondoron készültek a kutatások.



1. ábra *A 33 elemzett kutyaféle génexpressziós mátrixának multidimenziós skálázása. A 33 kutyafélében volt 10 európai, 13 ázsiai és 10 amerikai minta. Forrás: Jónás et al. (2024).*



2. ábra *A genom-szintű asszociációs vizsgálat eredményei Manhattan ábrán. Az x-tengelyen a kromoszómák láthatóak (1-38, X), különböző színekkel jelölve. A háromszögek a szignifikáns SNP mutációkat mutatják, ugyanazzal a színnel, mellyel a kromoszóma is színezve van. Forrás: Morvay Bence mesterszakos hallgató disszertációja (2025).*

Kaeberlein et al. (2016). The Dog Aging Project: Translational Geroscience in Companion Animals. doi: 10.1007/s00335-016-9638-7.

Jónás et al. (2020). A Preliminary Study to Investigate the Genetic Background of Longevity Based on Whole-Genome Sequence Data of Two Methuselah Dogs. doi: 10.3389/fgene.2020.00315.

Jónás et al. (2024). Dog Domestication Strongly Relied on Translation Regulation According to Differential Gene Expression Analysis. doi: 10.3390/ani14182655.

Jónás et al. (2025). Differential Gene Expression Analysis of Whole Blood Transcriptome Between Young and Old Border Collie Dogs. doi: 10.3390/vetsci12020086.

Jónás et al. (In prep.). Differential transcript usage of canine RNA-seq samples from brain cortical region.

Sándor et al. (2022). Poly(A) RNA sequencing reveals age-related differences in the prefrontal cortex of dogs. doi: 10.1007/s11357-022-00533-3.

Morvay B. 2025. Investigation of the word learning ability of dogs using GWAS methods. M.Sc. thesis.

Hogyan tanulhatunk a rendszermérnöki gondolkodásról a szuperszámítógéptől?

A Komondor, a nyelvi modellek, és az ipar találkozása a hibaelhárítás új korszakáért

Szerző:
Dénes Márk

A modern IT-üzemeltetésben naponta sok hibajegyet kell rövid határidővel kezelni, miközben a dokumentáció gyakran csak a végső megoldást tartalmazza, a mögötte álló mérnöki gondolkodást nem. Kutatásunk célja az volt, hogy a hibajegyek szövegéből mesterséges intelligencia segítségével visszakövethetővé tegyük a megoldási stratégiát. Szuperszámítógépes környezetben generált tanuló-készlettel több nagy nyelvi modellt finomhangoltunk, amelyek közül a nagyobb modellek már az emberi teljesítményhez közeli pontossággal készítettek javasolt megoldási lépéseket.

A modern IT üzemeltetésben a gyors és megbízható hibaelhárítás jelenti az egyik legfőbb kihívást. Egy infotechnológiai szolgáltató cégnél (ITSM) naponta több tucat hibajegy fut be, amelyeket előre rögzített rövid határidővel (SLA) kell kezelni. Nálunk, az DAM Invisible Technology Zrt-nél ez jellemzően 150-300 hibajegyet jelent. A tapasztalt rendszertámogató mérnöknél marad a megoldás módja, a munkalapokon azonban többnyire csak maga a megoldás szerepel, nem pedig az, hogy lépések és döntési logika vezetett el hozzá. Mérnökként dolgozom, és amikor felismertem, mennyi érték rejlik a hibajegyek mögötti gondolatokban, elhatároztam, hogy megpróbáljuk magát a mérnöki gondolkodást is megragadni.

A DAM Invisible Technology Zrt.-nél hatalmas mennyiségű korábban megoldott hibajegy áll rendelkezésre, de a mögöttük álló döntések, zsákutcák, illetve kockázatelemzések gyakran csak a szóban átadott információk, a mérnöki tudás részei. Elég egyszerű a kérdés, de mégis merész: **képesek vagyunk-e a mérnök gondolkodási és megoldási stratégiáját pusztán a hibajegy és a munkalap szövegéből visszakövetni?** Először a kollégákkal teszteltük, hogy szerezzünk némi érdemi tapasztalatot. Három Windows rendszermérnököt kértünk fel arra, hogy tíz-tíz olyan hibajegyhez készítsenek megoldási stratégiát, amelyeket nem ők oldottak meg, és amelyek munkalapja csak a végső eredményt tartalmazta. A stratégiákat ezt követően egy 0–3-as skálán értékeltük (0: zsákutca; 1: hibás irány, de logikus következtetés; 2: helyes irány, de a valódi megoldás hiányzik; 3: helyes stratégia, benne a tényleges megoldással). A kis felmérés átlaga 2,46 lett, ami azt jelzi, hogy a megoldási stratégia valóban „kihámozható" a szövegből: a legtöbb esetben sikerült eltalálni a helyes irányt, és sok esetben magát a megoldást is.

Innen adódott a következő kérdés: **ha egy mérnök képes erre, vajon egy nyelvi modell is képes lehet ugyanígy visszakövetni a megoldási stratégiát?** A modell finomhangoláshoz azonban nem csak nagy mennyiségű, hanem jó minőségű tanító készletre is szükség lenne, ami esetünkben nem állt rendelkezésre. Ezért egy másik megközelítést választottunk: egy többlépéses rendszerutasítással láttunk el egy erős következtetési képességekkel rendelkező nagy nyelvi modellt (LLM), a DeepSeek R1-et , hogy a bejelentés és munkalap alapján készítsen egy megoldási stratégiát. Az utasítást több körben finomítottuk mérnöki visszajelzések alapján, ezúttal nem csak Windows, hanem hálózati, Linux, Service Desk szakembereket és rendszergazdákat is bevonva, valamint a területeiknek megfelelő példákkal kiegészítve a tanulókészletet. A végleges rendszerutasításban öt mintapélda mutatta be a modell számára, hogyan kell a bejelentésekből megoldási stratégiát készíteni. Ez az ötpéldás (5-shot) rendszerutasítás már elég sokféle technikai területet és hibajegy-típust lefedett ahhoz, hogy stabilan működjön, és végül összesen 2057 új példát tudtunk vele generálni. A DeepSeek R1 a korábbi 0-3 skálájú teszten 2,4-es átlagot ért el, ami meglepően jól közelít az emberi teljesítményhez. Az eredményben fontos szerepe volt annak, hogy a modellel először egy rövid összefoglalót, problémaértelmezést és kockázatelemzést készíttettünk, és csak ezt követően kezdhetett hozzá a részletes megoldási stratégia kidolgozásához.

Hogy mindezt miért kellett HPC-n futtatni? A rövid válasz: a gondolkozó modellek rendkívül hosszú szöveget generálnak, és ehhez óriási memóriaigény társul. A DeepSeek R1 egy 671 milliárd paraméteres, Mixture of Experts (MoE) neurális hálózati architektúrájú modell. Egyetlen példánál 8000-12000 tokennyi kimenet is keletkezett (főként a belső Chain-of-Thought monológ miatt), amelyből az összefoglalók és stratégia körülbelül 3000 tokent tettek ki. A Komondor szuperszámítógép AI node-ján 8 darab NVIDIA A100-40GB GPU, összesen 320 GB memória állt rendelkezésünkre; ez tette lehetővé, hogy a kvantizált Q2_K_XL változatot a llama.cpp backenddel betöltsük. A generálási sebesség a kontextus méretétől függően 10-15 token/s volt, így egyetlen példa előállítása 8-10 percet vett igénybe. Ezzel 1000 példa legenerálása nagyjából egy teljes hetet, körülbelül 1300 GPU órát jelentett volna. Más backendet is kipróbáltunk (vLLM, exl2/exl3), de akkor még nem állt rendelkezésre a megfelelő méretű kvantizált modell, a llama.cpp viszont nem kezeli hatékonyan az egynél nagyobb batch-méretet, így a párhuzamos generálás nem működik jól vele, ezért a Komondort választottuk.

A tanulókészlet új lehetőséget nyitott meg azzal, hogy minden hibajegyhez rendelkezésre állt egy releváns megoldási stratégia. Felmerült tehát a kérdés: **betanítható-e a modell arra, hogy pusztán a hibajegy szövegéből állítson elő érvényes megoldási stratégiát, a konklúziót (munkalapot) egyáltalán nem használva?** Több nagy nyelvi modellen is végeztünk LoRA-alapú felügyelt finomhangolást

az általunk fejlesztett, több GPU-ra és HPC-környezetre optimizált Etalon tréning rendszerünkben. A tréningek a modellmérettől függően 8-24 GPU óra között mozogtak; a 6000 lépéses futásoknál jellemzően a 4000-5000-es checkpointok adták a legjobb eredményt. A kis modellek (Gemma 3 1B/4B és a DeepSeek R1-alapú 8B) nem tudtak használható stratégiát előállítani; a Gemma 3 12B már mutatott némi elvárt „gondolkodási" mintát, de a tartalom minősége még nem érte el a szükséges szintet. Végül a nagyobb modellek váltak be: Gemma 3 27B és a Qwen 3 30B A3B egyaránt 2,2-es átlagot értek el kizárólag a ticket szövege alapján, sok esetben a helyes megoldást is tartalmazó stratégiával. Üzemeltetési szempontból fontos különbség, hogy a Qwen modell a MoE architektúrájának köszönhetően 5-6-szor gyorsabban generálta a stratégiákat, mint a Gemma, ami a nagy hibajegy forgalom esetében kulcsfontosságú előnyt jelent.

Eredmények (0–3 skála, csak a ticket szövegéből, munkalap tudta nélkül):

Modell	Paraméterek	Eredmény
Gemma 3	1B	0,0
Gemma 3v	4B	0,0
Gemma 3	12B	1,4
Gemma 3	**27B**	**2,2**
Qwen 3	**30B A3B**	**2,2**
Qwen 3	32B	2,1
DeepSeek R1 0528 Qwen 3	8B	0,0

Íme egy példa arra, hogyan nézne ki egy képzeletbeli partner bejelentésére generált megoldási stratégia:

Jegy szövege:

Partner: Boltzmann Kft. | Name: Dénes Márk | Operation Service Level: Szerver üzemeltetés | Type: Incidens | Priority: 3 (normal) | Support Grade: Rendszergazda | Subject: DC (192.168.1.30) WARNING SERVICE: Volume C: | Message: DC (192.168.1.30) WARNING SERVICE: Volume C: WARNING: Disk space usage below 10%

Összegzés:

Normál prioritású monitoring incidens, amelyet egy rendszergazda szintű szakembernek kell kezelnie Szerver üzemeltetési szolgáltatási szinten. A bejelentő Dénes Márk a Boltzmann Kft. alkalmazottja. A probléma a DC (Domain Controller) szerver C: meghajtóján történő lemezterület-használatról szól, ahol a riasztás 10% alatti értéket jelez.

Probléma értelmezés:

A Domain Controller-on a C: meghajtó tipikusan a Windows operációs rendszerhez, alkalmazásokhoz és ideiglenes fájlokhoz kapcsolódik. A 10%-os alacsony szabad terület komoly kockázatot jelenthet a szerver működésére nézve:

- Az operációs rendszer frissítéseinek telepítése vagy futása akadályozva lehet, ami biztonsági réseket hagyhat nyitva.
- A Windows gyorsindítási (Hybrid Boot) funkció megszakadhat, ami a rendszerindulás lassulását okozhatja.
- Az Exchange vagy más kritikus szolgáltatások működése akár ideiglenesen is befolyásolható lehet, ha nincs elegendő hely a naplófájlok tárolására vagy átmeneti adatkezelésre.

A lemezterület csökkenésének oka lehet: gyorsan növekvő naplók, ideiglenes fájlok felhalmozódása, nem törölt frissítési csomagok vagy egyszerűen a fizikai lemezkapacitás kimerülése. Fontos megjegyezni, hogy a DC-k esetében a rendszerfelismerő (SYSVOL) könyvtár és a GPO-alkönyvtárak is jelentős területet foglalhatnak.

Megoldási stratégia:

Az elsődleges cél azonnali lemezterület-felszabadítás, hogy elkerüljük a 10%-os küszöb alatti csökkenést, míg a mögöttes ok kivizsgálása másodlagos. A megoldási folyamatnak biztosítania kell a szerver stabil működését és a kritikus szolgáltatások zavartalan működését.

1. A lemezterület-használat részletes elemzése eszközökkel (pl. WinDirStat, TreeSize) a legnagyobb fájltípusok és könyvtárak azonosítására. Ez segíthet kiszűrni a felesleges vagy elavult adatokat.

2. Az ideiglenes internetfájlok (Temp Internet Files) törlése, valamint a „Törlés" mappában található elemek eltávolítása, amelyek gyakran felhalmozódnak.

3. A Windows Update frissítések manuális törlése (Disk Cleanup segítségével vagy a „Component Store" kényszerített tisztogatásával), mivel ezek jelentős területet foglalhatnak.

4. Az Exchange szolgáltatás esetén az üzenetsorok (Message Queues) és a legfrissebb naplók áttekintése, különös figyelemmel az „ActiveSync" vagy „Mailbox Transport" szolgáltatásokra, amelyek időnként nagy mennyiségű adatot gyűjthetnek.

5. Ha a fenti lépések nem elegendőek, javasolt a felhasználóval egyeztetni a legnagyobb fájlok (pl. videók, nagyméretű mentési adatok) meglétével kapcsolatban, amelyeket áthelyezhetnek más tárolóeszközökre.

6. Végül a szerver teljesítményének és hőmérsékletének monitorozása az üzemeltetési állapot értékelése során, mivel a lemezterület

A modell vállalati rendszerbe épülve az újonnan érkező hibajegyekhez automatikusan, néhány másodperc alatt egy háromrészes kimenetet készít: rövid összegzést, probléma értelmezést (kockázat és hatás), és megoldási stratégiát. Ez jelentős segítséget nyújt a Service Desk munkatársainak a hibajegy delegálásakor és a mérnök számára is számottevő időmegtakarítást jelent a probléma pontos körének meghatározásában.

A HPC-n futó projektünkben saját fejlesztésű és menedzselt ügynököket (LLM Agenteket) tanítunk automatikus problémamegoldásra, ahol a stratégiagenerálás jelentette az első lépést és egyben a további fejlesztések alapját.A fent bemutatott tananyag-előállítási módszer számos más területen is jól működött, így nem csupán ipari vagy üzleti környezetben lehet hasznos.

Legyen szó üzemeltetési problémák mérnöki igényességű megoldásának kutatásáról, vagy egyéb számunkra hétköznapi jelenség vizsgálatáról, a Komondor szuperszámítógép megbízható és ideális környezetet biztosít mindehhez, teret adva a valódi technológiai áttörések megvalósulásának.

Köszönet: *a mérnökcsapatnak az értékeléseket, a hazai HPC közösségnek és a Komondor üzemeltetőinek a megbízható kapacitást. Megtiszteltetésnek érzem, hogy „vendégkutatóként" én is részesülök ebből – nagy élmény, hogy a két világ találkozhat, és izgatottan várom, hogy mit hoz a jövő.*
Dénes Márk László, MS-HW Rendszermérnök, AIOPS csapatvezető • DAM Invisible Technology Zrt. • denes.mark@damit.hu

A számítógépes szimuláció és a virtuális valóság

Szerző: *Boda Dezső*

A tudomány egyik alapvető kérdése, hogy miként alakul ki rendezett viselkedés az atomok és molekulák hatalmas számú kölcsönhatásából. Kutatásunkban nanoméretű csatornákat, úgynevezett nanopórusokat vizsgáltunk, amelyekben ionok vándorolnak át elektromos tér hatására. Kísérletek szerint bizonyos esetekben a pórus nem a vártnak megfelelően viselkedik: például a $CaCl_2$ oldatban nem mutat ion-szelektivitást. A jelenség okát nagyméretű molekuláris szimulációkkal tártuk fel, amelyek atomról atomra modellezik a rendszer viselkedését szuperszámítógépen.

Az ember kíváncsi természetű lény, és az emberek halmazán belül a tudósok kiváltképp kíváncsi természetűek. Szeretnénk megérteni, hogy miért működik úgy az Univerzum, ahogy azt tapasztaljuk. És nem csak a szemmel vagy a távcsővel látható dolgok működése érdekel bennünket, hanem a kicsiké is, az atomoké és a molekuláké, amikből minden körülöttünk, még a saját testünk is, építkezik. Az alapkérdés az, hogy hogyan vezet az elképzelhetetlen számú atom spontán viselkedése egy olyan rendhez, amit a saját szemünkkel vagy a műszereinkkel észlelni tudunk?

Mit értünk itt rend alatt? Alapvetően bármit, amiben mintázatokat fedezhetünk fel, már egy rendezett rendszernek tekinthető. A folyadék rendezettebb, mint a gáz. A kristály rendezettebb, mint a folyadék. A galaxisok a véletlenszerűtől szemlátomást különböző struktúrákba rendeződnek. Érdekes, hogy ezek a struktúrák (hálózatok) félelmetesen hasonlítanak azokhoz, amilyenekbe az agysejtek rendeződnek. Ha a technológia felől közelítjük meg a kérdést, akkor képzeljük el, hogy van egy rendszerünk, amit meglövünk valami jellel (elektromos impulzus, fény, nyomás, bármi lehet). Hogyan reagál erre a rendszer, azaz mi lesz a kimeneti jel? Ha a bemenet és a kimenet között valami szabályszerűséget fedezünk fel, akkor ebben a rendszerben ez valamiféle rendezettségről tanúskodik. Azt szeretnénk tehát megérteni, hogy miért reagál a rendszer egy behatásra éppen úgy, ahogy reagál.

Igen ám, de lehet, hogy a kimenet és a bemenet közötti összefüggés lent az atomok szintjén dől el. Hogyan tudjuk azt a szintet tanulmányozni? Hogy rátérjünk konkrétan a minket érdeklő rendszerekre, a nanopórusok olyan csatornák, amiknek az átmérője a nanométer tartományba tartozik. Ez az atomok mérettartománya. Egy vízmolekula átmérője kb. 0,3 nm (0,3 x 10-9 m). Mi történik egy olyan szűk pórusban? Milyen módszerrel tudunk egy ilyen nanopórust tanulmányozni?

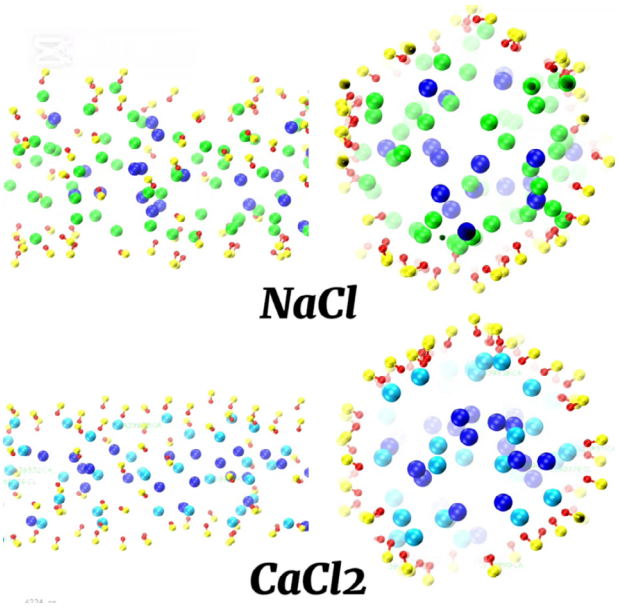
Először is beszéljünk arról, hogy mi is ez a nanopórus. Képzeljünk el egy vékony fóliát, ami két elektrolit oldatot választ el egymástól és ezen a fólián egy apró lyukat, azaz a nanopórust. Az elektrolit oldatban az oldószer általában a víz; ebben ,,úszkálnak" az ionok. A legismertebb elektrolit oldat a vízben oldott konyhasó (NaCl), azaz a tengervíz. Ha külső elektromos teret kapcsolunk erre a rendszerre, akkor a nanopóruson keresztül ionok fognak vándorolni az egyik oldalról a másikra. Az egyik fő kérdés, hogy melyik ion fog nagyobb számban áthaladni, azaz mi a nanopórus szelektivitása. Ez sok mindentől függ, de ha például a pórus fala negatívan töltött, akkor a pórus a pozitívan töltött ionokra (kationokra, Na+) lesz szelektív. Legalábbis ezt gondolnánk. A valóságban ez a NaCl-ra így is van, de $CaCl_2$-ra nem. Vajon miért nem? Ezt a kérdést tettük fel magunknak.

Kísérleti eszközökkel meg lehet mutatni, hogy $CaCl_2$ esetében a pórus nem szelektív, azaz hasonló arányban engedi át a Ca2+ ionokat, mint a Cl- ionokat [1]. Arra a kérdésre azonban, hogy miért van ez így, mi ennek a viselkedésnek a molekuláris mechanizmusa, a kísérletek nem adnak választ. Ehhez bele kéne néznünk egy olyan mikroszkópba, ami nem csak megmutatja nekünk az atomokat, de számot ad a mozgásukról is. Ilyen mikroszkóp nincs. Azaz mégis van, úgy hívjuk, hogy számítógépes szimuláció. A lényege az, hogy a rendszert alkotó részecskéket (az ionokat, a vízmolekulákat, a membránt alkotó atomokat és minden mást) a közöttük ható kölcsönhatásokkal modellezzük, majd ezeknek a részecskéknek a viselkedését a számítógépben szimuláljuk [2,3]. Hasonlatos ez ahhoz, ahogy a Gyűrűk Urában a nagy csatajeleneteket szimulálták: modellezték az orkok és a tündék viselkedését, majd a számítógép legenerálta a csatajelenetet. A számítógép tudta, hogy ha egy ork és egy tünde találkozik, mit fognak kezdeni egymással. A molekuláris szimuláció is ilyen, csak az orkok és tündék szerepét atomok és molekulák játsszák. Ilyen módon remek mozikat lehet készíteni (https://youtu.be/SBHBw7req98?si=36Fj-RDBWG510xQEB), de minden pillanatban tudatában kell lennünk, hogy nem a valóságot nézzük, hanem egy virtuális valóságot. Van azonban egy eszköz a kezünkben, hogy megkülönböztethessük a kettőt: ha a szimuláció által adott eredmény ugyanazt az eredményt adja, mint a kísérlet, akkor elfogadjuk a modellt. Nincs ezen mit csodálkozni, így működik a fizika. A fizika csak egy modellt tud tanulmányozni, a valóságot nem. Papp Sándor Tanár Úr szokta mondani: „a modell egy amputált valóság".

Ezek a szimulációk azonban rendkívül számításigényesek. Részecskék tízezreit kell modelleznünk, de néha millióit. Ehhez bizony erős számítógépek kellenek. A Komondor segítségével képesek voltunk ezeket a szimulációkat végrehajtani, az eredményeket kielemezni, és az elemzésből következtetéseket levonni.

Osszuk a nanopórust képzeletben két részre: a fal melletti felületi rétegre és a középső tartományra. A lenti ábrán és a videóból látható, hogy a fal mellett felhalmozódnak a kationok (Na+ zöld golyók, Ca2+ világoskék golyók), hogy a fal negatív töltését (kis piros golyók) semlegesítsék. Ebbe a tartományba az anionok (Cl- sötétkék golyók) nem igazán jutnak el. A NaCl és a $CaCl_2$ viselkedése merően különböző ebben a tartományban. A Na+ ionok tudnak mozogni a felületi tartományban, de mivel többen vannak, mint a Cl- ionok, a pórus Na+-szelektív lesz. A Ca2+ ionok (kettős töltésük miatt) erősen hozzáragadnak

a fal negatív töltéseihez, és ezáltal megbénulnak: a mozgékonyságuk nagyon kicsi lesz. Emiatt a felületi tartományban a Ca2+ ionok nem járulnak hozzá az áramhoz. Ebben a tartományban azonban a Cl- ionok sem járulnak hozzá, mivel nincsenek jelen. $CaCl_2$ esetében tehát csak a középső, nem szelektív tartomány fog vezetni, így a pórus is nem szelektív lesz.



A felső ábra NaCl-ot, míg az alsó ábra a $CaCl_2$-ot tartalmazó nanopórus szimulációjának egy pillanatfelvételét mutatja oldalról és a hossztengely tengely irányából. A zöld, a világoskék és a sötétkék gömbök a Na+, Ca2+ és Cl- ionokat jelölik. Az ábra az áttekinthetőség érdekében csak a szilanol csoportokat mutatja (piros: O atom, sárga: Si atom), de ezek egyúttal a pórus falát is kijelölik, mivel a felszínen vannak. Az ábrához tartozó video linkje: https://youtube.com/shorts/sNgINE4XPgM?si=opkDPiZovIkL8YZw

Ezeket az állításokat természetesen különböző, a szimulációkból kinyerhető adatokkal, függvényekkel is alá tudjuk támasztani, de a fenti egyszerű érvelés szemléletesen írja le azt, amit tapasztaltunk. A számítógépes szimulációk természetesen sokféle célra használhatók, ma már oda jutottunk, hogy könnyen használható szoftvercsomagok formájában ugyanúgy megvásárolhatók, mint ahogy az ember megvásárol egy mérőműszert. A szimulációs technikák tehát kinőtték a gyermekkorukat és felsorakoztak a kísérleti technikák mellé, azokat kiegészítő hatékony vizsgálati módszerré váltak.

[1] Y. He, D. Gillespie, D. Boda, I. Vlassiouk, R. S. Eisenberg, and Z. S. Siwy. Tuning transport properties of nanofluidic devices with local charge inversion. JACS, 131(14):5194-5202, 2009.

[2] M. Valiskó, B. Matejczyk, Z. Ható, T. Kristóf, E. Mádai, D. Fertig, D. Gillespie, D. Boda. Multiscale analysis of the effect of surface charge pattern on a nanopore's rectification and selectivity properties: from all-atom model to Poisson-Nernst-Planck. J. Chem. Phys. 150(14):144703, 2019.

[3] S. Shabbir, D. Boda, Z. Ható: Transport properties of CaCl2 are determined by the balance of ion-ion and ion-water interactions as revealed by molecular dynamics simulations. J. Mol Liq. 426:127308, 2025

Amikor a mesterséges intelligencia visszaadja a mozgás örömét: így forradalmasítja egy magyar startup a térdműtéteket

Szerzők:
Szoldán Péter, Prof. Dr. Hangody László, Szabó Endre, Dr. Egyed Zsófia

A csapat: ahol az orvostudomány találkozik a mesterséges intelligenciával

A GenuForm AI-t olyan szakemberek alkotják, akik egyedülálló módon ötvözik az orvosi és a technológiai tudást:

Szoldán Péter (CEO) - AI-mérnök, a Harvard Egyetemen szerezte diplomáját, korábban a McKinsey & Company tanácsadójaként dolgozott

Prof. Dr. Hangody László (CMO) - 30+ év klinikai gyakorlattal rendelkező ortopéd sebész, a mozaikplasztika világhírű feltalálója

Szabó Endre (CTO) - matematikus és AI-szakértő, a kombinatorikus geometria kutatója

Dr. Egyed Zsófia - 30+ év klinikai gyakorlattal rendelkező radiológus



A probléma, ami milliókat érint

Képzeljük el, hogy egy napon fájdalom nélkül tudunk leguggolni, futni, vagy akár csak sétálni, olyasmi ez, amit sokan természetesnek veszünk. Mégis, az 55 év feletti lakosság mintegy 30%-a küzd valamilyen fokú térdporc-károsodással, és ennek fele fájdalmas tünetekkel jár. A térdízület az emberi test egyik legjobban terhelt pontja: nemcsak a testsúlyunkat hordozza, de minden lépésnél, ugrásánál, guggolásánál óriási erőknek van kitéve.

A jelenlegi megoldás, a teljes térdprotézis, ugyan széles körben elterjedt, de komoly korlátai vannak. A műtét után a páciensek gyakran csak korlátozott mozgástartományt kapnak vissza, és 15-20 év múlva az implantátum kilazul, újabb műtétre van szükség. De mi lenne, ha létezne egy jobb megoldás?

A donorporc

A válasz valójában régóta ismert: a donoroktól származó porcszövet átültetése. Ez a módszer, szemben a mesterséges protézissel, képes visszaadni a teljes mozgástartományt, és egy életen át szolgálhatja a pácienst. Ráadásul immunológiai szempontból is ideális: a porcszövet átültetésekor nincs szervkilökődés.

Van azonban egy kritikus akadály: a geometriai illeszkedés. A porc merev szövet – ha a donor és a befogadó térdének alakja nem illeszkedik tökéletesen, a műtét sikeressége kérdéses. Eddig ezt az illesztést manuálisan végezték, ami rendkívül időigényes és pontatlan volt. Ezért világszerte mindössze évi 10 000 ilyen műtétet végeznek, miközben a potenciális igény ennek huszonötszöröse.

A mesterséges intelligencia

A GenuForm AI pontosan ezt a szűk keresztmetszetet oldja meg. A magyar fejlesztésű rendszer négy lépésben, mesterséges intelligencia segítségével párosítja össze a donort és a befogadó pácienst:

1. **A páciens térképezése:** MR-felvételek alapján a mesterséges intelligencia meghatározza a páciens meglévő porcának pontos geometriáját.

2. **Digitális rekonstrukció:** A rendszer „visszaállítja” az egészséges állapotot: meghatározza, hogyan nézne ki a sérült terület ép állapotban.

3. **A donor feltérképezése:** Amikor donorszövet válik elérhetővé, a mesterséges intelligencia azonnal elemzi annak geometriáját.

4. **Az illesztés:** Egy kombinatorikus geometriai algoritmus megkeresi a tökéletes egyezést a donor és a befogadó páciens között. Egyetlen donorporcból gyakran több páciens is ellátható, az algoritmus pedig optimálisan osztja fel a rendelkezésre álló szövetet.

A mesterséges intelligencia modelljeinek betanítása és a komplex geometriai számítások hatalmas számítási kapacitást igényelnek. Fejlesztésünk során a Komondor szuperszámítógép nyújtotta erőforrásokra támaszkodtunk, amely nélkül a kutatás nem valósulhatott volna meg ilyen ütemben. A nagy teljesítményű számítástechnika (HPC) lehetővé tette, hogy a 3D képfeldolgozási algoritmusainkat nagy mennyiségű orvosi képanyagon tanítsuk be, és a kombinatorikus illesztési feladatokat valós időben oldjuk meg.

Az első műtét sikere

2024-ben a GenuForm AI technológiáját először alkalmazták páciensen, kutatási engedély keretében. Az eredmények kiválóak voltak. A gyors, 24-48 órán belüli műtéti idő biztosítja, hogy a szövet életképes maradjon, ami drámaian növeli a műtét sikerességét.

Miért csináljuk?

A GenuForm AI csapatát az a meggyőződés hajtja, hogy a mesterséges intelligencia és az orvostudomány találkozásában óriási, még kiaknázatlan lehetőségek rejlenek. A 3D számítógépes látás technológiája az elmúlt években érte el azt az érettségi szintet, amely lehetővé teszi az ilyen komplex geometriai illesztési feladatok megoldását – és mi ezt a lehetőséget szeretnénk az emberek szolgálatába állítani.

Célunk, hogy évtizedes orvosi és technológiai tapasztalatainkat az egészségmegőrzés javára fordítsuk. Hiszünk abban, hogy a térdporc-átültetés 5-10 éven belül ugyanolyan rutineljárássá válhat, mint ma a protézisbeültetés, csak éppen jobb eredménnyel, teljes mozgástartománnyal és élethosszig tartó megoldással. Azt szeretnénk elérni, hogy az emberek ne kelljen lemondjanak a mozgás örömérőll: a kertészkedésről, a sétákról, a sportról, mindarról, ami az életminőséget jelenti.

Hazai infrastruktúrára építve kínál biztonságos AI-alternatívát a TENGR.AI

Szerző: *Horváth Tibor, a Tengr.ai vezérigazgatója*

A TENGR.AI egy képgenerálásra és szerkesztésre specializált cenzura nélküli platform, amely szöveges és vizuális bemenetek alapján állít elő tartalmakat. A TENGR.AI megközelítése nem elsősorban tiltásokra vagy explicit elutasításokra épít, hanem arra törekszik, hogy a felhasználói kéréseket minden esetben feldolgozható formára alakítsa, és biztonságos, a rendszer keretein belül értelmezhető kimenetet biztosítson.

A rendszer kialakítása mögött az a felismerés áll, hogy a generatív modellek gyakorlati értéke sok esetben nem az egyszeri képelőállításban, hanem az ismételhető, kontrollált és skálázható folyamatokban rejlik. Ennek megfelelően a platform nem kizárólag kreatív eszközként, hanem egyfajta vizuális feldolgozó pipeline-ként értelmezhető, ahol a generálás különböző lépései strukturált módon szervezhetők.

Ennek egyik kulcseleme az úgynevezett „receptek" használata, amelyek több lépésből álló generálási folyamatokat rögzítenek. Ezek a receptek különböző modellek, paraméterek és feldolgozási lépések kombinációját jelentik, és lehetővé teszik, hogy egy adott vizuális eredmény reprodukálható módon újra előállítható legyen. Ez a működés különösen olyan alkalmazási területeken válhat relevánssá, ahol a konzisztencia és az iteráció meghatározó, például marketinganyagok, kreatív kampányok vagy termékvizualizációk esetében. A megközelítés egyben közelebb áll a klasszikus szoftverfejlesztési pipeline-ok logikájához, ahol a determinisztikusabb működés és az újrafelhasználhatóság kiemelt szerepet kap.

A generatív mesterséges intelligencia fejlődését az elmúlt években nagymértékben a számítási kapacitás elérhetősége határozta meg. A vezető szereplők jellemzően nagylétékű, felhőalapú GPU-infrastruktúrákra támaszkodnak, amelyek nemcsak technológiai, hanem gazdasági értelemben is jelentős belépési korlátot jelentenek. Ez a dinamika különösen az európai ökoszisztémában válik hangsúlyossá, ahol a fejlesztőcsapatok gyakran korlátozottabb erőforrások mellett működnek, miközben globális szereplőkkel kell versenyezniük.

Ebben a környezetben egyre inkább felértékelődnek azok a megközelítések, amelyek nem pusztán a meglévő infrastruktúra-modellek skálázását célozzák, hanem eltérő erőforrás-stratégiák mentén szervezik újra a generatív rendszerek működését. Az ilyen alternatívák tipikusan a költség-teljesítmény kompromisszum újrapozicionálására, valamint a rendszer feletti kontroll növelésére törekszenek. A TENGR.AI egy ilyen irányt képvisel, amely hibrid módon kombinálja a saját üzemeltetésű környezetet és a hazai nagy teljesítményű számítási infrastruktúrát (HPC), és ezzel egy, a felhőközpontú modellektől részben eltérő működési logikát alakít ki.

Adatbiztonság és európai szuverenitás

A platform pozicionálásában hangsúlyos szerepet kap az adatkezelés és a felhasználói kontroll kérdése is. Az architekturális kialakítás arra törekszik, hogy a felhasználói adatok kezelése minimalizált legyen, és a generált tartalmak feletti rendelkezés elsősorban a felhasználónál maradjon. Ez a megközelítés összhangban állhat azokkal az európai szabályozási és iparági törekvésekkel, amelyek a digitális szuverenitást és a „privacy-by-design" elveket helyezik előtérbe. Ugyanakkor fontos hangsúlyozni, hogy az ilyen előnyök mértéke erősen függ az implementáció konkrét részleteitől, így ezek nem tekinthetők minden esetben általánosan érvényesnek.

Ezt a szemléletet erősíti a „Vault" funkció, amely a TENGR.AI egyik meghatározó eleme. A Vault megoldás kliensoldali titkosításra épül: a tartalmak titkosítása és visszafejtése a felhasználó eszközén történik, így a szervereken tárolt adatok titkosított formában jelennek meg. Ennek következtében a szolgáltató technikailag ugyan kezeli és tárolja az adatokat, azonban azokhoz tartalmi szinten nem fér hozzá, mivel a visszafejtéshez szükséges kulcsok nem állnak rendelkezésére.

Ez a megközelítés jelentősen növeli a felhasználói kontrollt, és csökkenti annak kockázatát, hogy az érzékeny vagy üzleti értéket képviselő tartalmak illetéktelenül hozzáférhetővé váljanak. Ugyanakkor fontos hangsúlyozni, hogy az ilyen rendszerek biztonsága nagymértékben függ a kulcskezelés módjától és a kliensoldali implementáció minőségétől. Ebben az értelemben a Vault nem abszolút, hanem architekturális értelemben vett erős adatvédelmi modellt képvisel, amely a gyakorlatban magas szintű, de nem korlátlan izolációt biztosít.

A hasonló megoldások jelenléte a piacon nem példa nélküli, ugyanakkor a generatív AI platformok esetében viszonylag ritkábban jelenik meg ilyen hangsúlyosan a kliensoldali titkosításra épülő adattárolás. Ezért a „Vault" megközelítése inkább egy tudatosan választott irányként értelmezhető, amely a funkcionalitás mellett a felhasználói adatkezelés kérdését is központi szerepbe emeli. A TENGR.AI ezzel nem csupán egy újabb generatív platformként jelenik meg, hanem olyan megoldásként, amely az európai adatvédelmi elvárásokkal és a digitális önrendelkezés iránti igénnyel is valódi összhangban van.

A hazai szuperszámítógép szerepe

A TENGR.AI működésének egyik meghatározó eleme a hibrid infrastruktúra-modell, amely a saját üzemeltetésű környezetet és a Komondor szuperszámítógép erőforrásait kombinálja. Ez a felépítés lehetővé teszi a különböző workloadok elkülönítését és optimalizált kezelését, ami a generatív rendszerek esetében különösen releváns. A napi fejlesztési és inference feladatok jellemzően a saját infrastruktúrán futnak, ahol alacsonyabb késleltetés és közvetlenebb operatív kontroll érhető el. Ezzel szemben a nagy számítási igényű folyamatok, például a modellek tanítása vagy kiterjedt kísérletek futtatása a HPC környezetbe kerülnek, ahol a nagy léptékű párhuzamosítás hatékonyabban kihasználható.

Ez a workload-szintű szétválasztás a költség-teljesítmény egyensúly finomhangolását is lehetővé teheti. A HPC erőforrások használata révén kezelhetők a nagyobb modellek és komplexebb tanítási folyamatok, miközben a saját infrastruktúra fenntartása biztosítja az operatív rugalmasságot és a gyors iteráció lehetőségét. Az így kialakuló architektúra egyfajta resource orchestration modellként értelmezhető, amely során a rendszer dinamikusan kiosztja, összehangolja és optimalizálja a rendelkezésre álló erőforrásokat, annak érdekében, hogy az egyes feladatok a számukra legmegfelelőbb környezetben fussanak.

Doménspecifikus finomhangolás az E.ON-nál

A generatív AI rendszerek működésében alapvető különbség figyelhető meg a tanítási és a kiszolgálási fázis között. A tanítási folyamatok tipikusan nagy volumenű, párhuzamosítható számításokat igényelnek, míg az inference oldal elsősorban alacsony késleltetésre és stabil működésre optimalizált. A TENGR.AI architektúrájában ez a különbségtétel explicit módon jelenik meg: a HPC környezet a fejlesztési és kísérleti tér szerepét tölti be, míg a production környezet a felhasználói igényekhez igazított kiszolgálást biztosítja. Ez a szétválasztás egyben csökkentheti a fejlesztési és éles rendszerek közötti interferenciát, és kiszámíthatóbb működést eredményezhet.

A fenti architektúra gyakorlati alkalmazását jól illusztrálja egy, az E.ON-nal közösen megvalósított, vízilabdára fókuszáló projekt. A kezdeményezés célja az volt, hogy a generatív modellek pontosabban jelenítsék meg egy specifikus sportág vizuális sajátosságait, mivel az általános modellek gyakran nem képesek megfelelően kezelni az ilyen kontextusokat. A fejlesztés során első lépésként egy domain-specifikus adatkészletet állítottunk össze, amely képi és szöveges forrásokat egyaránt tartalmazott. Ezt követően a modellek finomhangolása zajlott, amelynek célja a sportágra jellemző mozgásminták és vizuális struktúrák pontosabb reprodukciója volt.

A folyamatot iteratív validáció egészítette ki, amely során az eredményeket folyamatosan összevetettük valós sportesemények vizuális anyagaival. Az ilyen típusú fejlesztések rámutatnak arra, hogy a generatív AI gyakorlati alkalmazása gyakran túlmutat az általános modellek használatán, és jelentős erőforrásokat igénylő, doménspecifikus finomhangolást tesz szükségessé. Ebben a kontextusban a HPC infrastruktúra nem pusztán kiegészítő elemként jelenik meg, hanem a fejlesztési folyamat egyik kulcsfontosságú komponenseként.

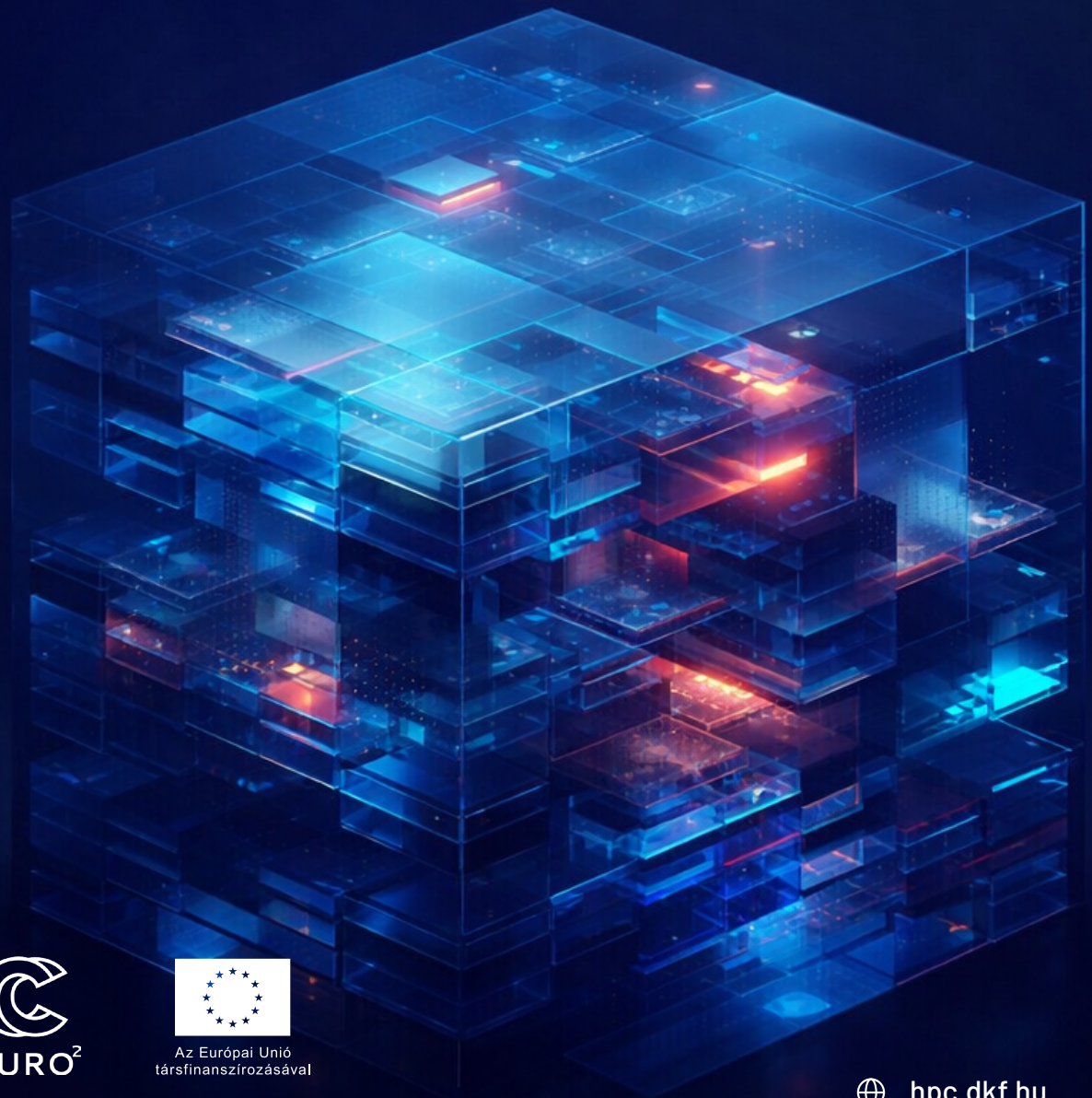
A hibrid modell együttműködési szempontból is releváns lehet. A HPC környezetek jellemzően megosztott infrastruktúrák, amelyek több szereplő számára biztosítanak hozzáférést, és ezáltal lehetőséget teremtenek kutatási és ipari együttműködésekre. Ilyen környezetben nemcsak az erőforrások, hanem a fejlesztési tapasztalatok és a doménspecifikus tudás is koncentrálódhat, ami hosszabb távon gyorsíthatja az innovációt. Ez a modell különösen olyan projektek esetében lehet előnyös, ahol a számítási kapacitás és a szakértői tudás egyaránt kritikus tényező.

Jövőkép

A jövőbeli fejlesztési irányok között megjelenik a klasszikus HPC környezetek és a modern, konténeralapú szoftverarchitektúrák közötti különbségek csökkentése. A hagyományos szuperszámítógépes rendszerek gyakran batch-orientált működésre épülnek, ami nem minden esetben illeszkedik jól az AI-fejlesztés iteratív és interaktív jellegéhez. Az olyan megoldások, mint a Slurm-alapú ütemezés kiterjesztései vagy a konténerizációs rétegek integrációja, lehetővé tehetik ezen rendszerek rugalmasabb használatát AI workloadok esetén. Ez a közelítés nemcsak technikai, hanem metodológiai szempontból is jelentős, mivel csökkentheti a fejlesztési és éles környezetek közötti különbségeket, és gyorsíthatja a time-to-market mutatókat.

A TENGR.AI példája azt mutatja, hogy a generatív AI fejlesztése nem kizárólag felhőalapú infrastruktúrákra épülhet. A megfelelően integrált nemzeti vagy regionális HPC erőforrások alternatívát kínálhatnak, különösen azokban az esetekben, ahol a költséghatékonyság, a kontroll és a doménspecifikus fejlesztés egyaránt fontos szempont. Ez a megközelítés várhatóan egyre nagyobb figyelmet kap Európában, ahol a technológiai szuverenitás kérdése stratégiai jelentőséggel bír. A generatív AI ökoszisztéma így valószínűleg nem egyetlen domináns infrastruktúra-modell köré szerveződik, hanem különböző architektúrák és erőforrás-stratégiák együttélésére épül, amelyek közül a HPC-alapú megközelítések egyre inkább életképes, bár kontextusfüggő alternatívaként jelennek meg.

DKF | hpc



EURO²

Az Európai Unió
társfinanszírozásával

Az EuroCC2 projektet az Európai Unió Horizon 2020 kutatási és innovációs keretprogramja a 101101903. számú támogatási szerződésben foglaltaknak megfelelően támogatja

hpc.dkf.hu

@HPC CC Hungary

@HPC_hu