



Kontextusablak, memóriamodulok és jelentőségük az oktatásban

Szerző: Lengyel Zsolt, iTStudy Hungary Számítástechnikai Oktató- és Kutatóközpont

Lektor: Hartyányi Mária, iTStudy Hungary Számítástechnikai Oktató- és Kutatóközpont

Tanulási cél

A mindennapi MI-használat számos meglepetést okozhat – a modell „elfelejt” egy utasítást, összekever két témát, vagy váratlanul tud rólunk valamit. Mindez két mechanizmussal magyarázható: a kontextus ablakkal és a memóriefunkciókkal. A tananyag célja, hogy bemutassa, mit képesek és mit nem képesek megőrizni az MI-rendszerek egy beszélgetésen belül és a beszélgetések között. A tananyag végére a tanárok meg tudják különböztetni az ideiglenes beszélgetési kontextust a hosszabb távú, memóriaszerű funkcióktól, el tudják magyarázni, hogyan hatnak ezek a folytonosságra, a személyre szabásra és az adatvédelemre, és gyakorlati rutinokkal tudják megbízhatóvá tenni a hosszú munkameneteket, illetve biztonságban tartani a tanulók adatait.

Bevezetés: a modell nem úgy emlékszik, mint egy ember

Amikor egy nyelvi modellel beszélgetünk, úgy tűnik, mintha emlékezne az egész beszélgetésre. Valójában maga a modell a beszélgetés alatt nem változik. A modell minden egyes válasz előtt összeállít egy bemenetet – az eddigi beszélgetést, a csatolt fájlokat, a háttérben futó utasításokat –, és kizárólag ebből számolja ki a következő választ.

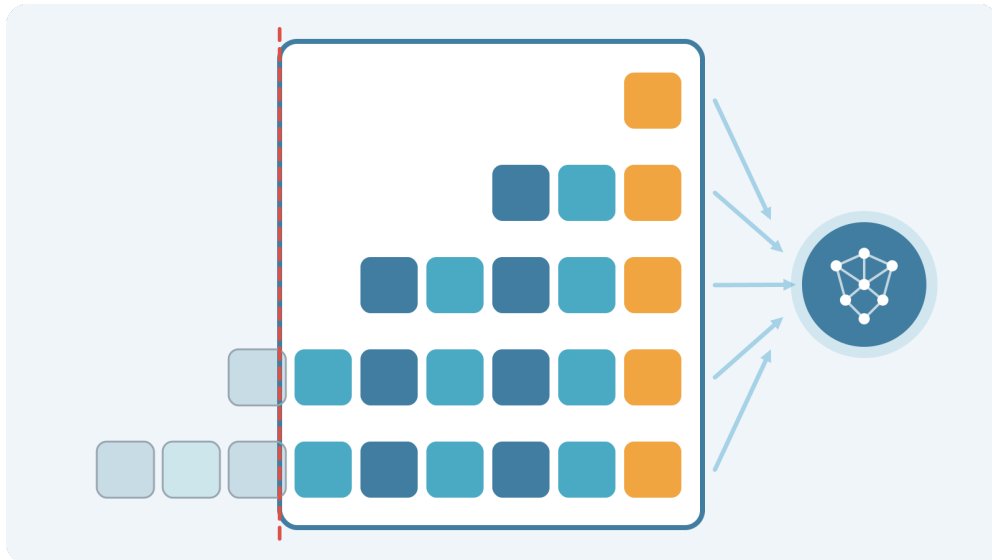
Jó hasonlat a Memento vagy az 50 első randi című film egyik szereplője: aki nem tudja megőrizni az új emlékeit, de az alaptudása megmarad. A nyelvi modell esetében az alaptudás az, amit a betanítás során megtanult; minden mást újra és újra elé kell tenni.



Az 50 első randi (Forrás: <https://nlc.hu/>)

Hogyan épül fel egy beszélgetés: a kontextus

A csevegőfelületen a bemenet minden újabb kérdéssel nő. Az első körben a modell csak az első üzenetünket (Ü1) kapja meg, és válaszol (V1). A második körben a rendszer az Ü1 + V1 + Ü2 szöveget küldi el, a harmadikban az Ü1 + V1 + Ü2 + V2 + Ü3-at, és így tovább. Mindazt, amit a modell egy adott pillanatban „lát”, kontextusnak nevezzük.



A bemenet minden körrel nő; ami már nem fér bele a rögzített méretű kontextus ablakba, kiesik, vagy a rendszer tömöríti

A kontextust tokenekben mérjük – ezek kis szövegegységek, gyakran szótöredékek. A modell által egyszerre kezelhető tokenek maximális számát kontextusablaknak (angolul: context window) nevezzük. A kontextusablak elég nagy, de véges: a tananyag írásakor a vezető modellek néhány százezer tokentől (pl. a GPT-5 osztályú modellek esetében kb. 400 000) nagyjából egymillióig (pl. Gemini 3.1 Pro) terjedő ablakkal dolgoznak, ami több vastos könyvnek felel meg. Egy csevegőalkalmazásban elérhető korlát kisebb lehet a modell technikai maximumánál, és gyakran az előfizetéstől függ. Magyar szövegnél ráadásul egy szó átlagosan több tokenből áll, mint angolul.

Miért fontos ez?

Korlát: ha a beszélgetés túl hosszú lesz, a rendszer vagy levágja a legrégebbi részeket, vagy összefoglalva lerövidíti.

Költség és sebesség: minél hosszabb a bemenet, annál több számítást kell végezni; a válaszok lassulhatnak, és gyorsabban fogy a használati keret.

Minőség: a több kontextus általában segít, de „zajt” is okoz. Kutatások szerint a modellek a hosszú kontextus elején és végén lévő információt megbízhatóbban használják, ami pedig „középen van, az elveszhet”. Ha túl sok az irreleváns régi anyag, növekszik a félreértés esélye.

Tömörítés (compaction): amikor a beszélgetés összesűrűsödik

Ha egy beszélgetés eléri a kontextuskorlátot, sok alkalmazás automatikusan tömöríti a régebbi üzeneteket: összefoglalóval helyettesíti őket. A lényeg megmarad, a részletek halványulnak – hogy az új információnak mindig maradjon hely. Azt, hogy ez mikor történik, nem tudjuk befolyásolni, hátránya pedig, hogy finom részletek vagy korábbi pontos megfogalmazások elveszhetnek. A nagyon hosszú beszélgetés a böngészőt is belassíthatja.



Tömörítés: a hosszú beszélgetésből rövid összefoglaló lesz – és néhány részlet kiesik

Gyakorlati útmutató hosszú munkamenetekhez

- Témaváltáskor nyissunk új beszélgetést. Ugyanezt kell tenni, ha egy téma nagyon hosszúra nyúlik.
- Váltás előtt kérjük meg: „Foglald össze a beszélgetés következtetéseit, követelményeit és nyitott kérdéseit annyira, hogy be tudjam másolni egy új beszélgetésbe.” Ellenőrizzük az összefoglalót – ugyanis előfordul, hogy nem tartalmaz mindent.
- Ne az üzenet szövegébe írjunk be hosszú szöveget, inkább dokumentumként csatoljuk; egyes alkalmazások a nagyon hosszú beillesztést automatikusan csatolmánnyá alakítják.

Három réteg: általános tudás, memória és aktuális kontextus

Hasznos, ha egy MI-asszisztens „tudását” három rétegben képzeljük el:



A betanításból származó általános tudás, egy kis, szerkeszthető memória és az aktuális kontextus együtt alakítja a választ

Az általános tudás a betanításból származik. Széles körű, de nem naprakész: csak a modell tudásának lezárási dátumáig (a tudáskorlátig) terjed. Friss témáknál a modellnek internetes keresést kell használnia.

A memória a felhasználóról szóló, viszonylag kis mennyiségű információ, amit az alkalmazás az új beszélgetések elején a háttérben átad a modellnek. Minden új beszélgetés üres kontextussal indul – kivéve, ha a memóriamodulban van információ.

Az aktuális kontextus az, amit éppen most adunk át a modellnek: a beszélgetés, egy fájl, egy példa, egy hibaüzenet, az elvárt formátum. A modell abból dolgozik a legjobban, amit konkrétan megmutatunk neki: egy jó minta gyakran többet ér egy hosszú magyarázatnál.

Memóriefunkciók a gyakorlatban (példaként a ChatGPT)

A legtöbb nagy asszisztens (ChatGPT, Claude, Gemini) ma már kínál memóriaszerű funkciókat. A ChatGPT-ben például jelenleg ezek érhetők el:

Mentett emlékek (saved memories) – olyan részletek, amelyek megjegyzésére kifejezetten megkértük, vagy amelyeket automatikusan elmentett. A beszélgetésektől külön tárolódnak: egy beszélgetés törlése nem törli a mentett emléket.

Hivatkozás a korábbi beszélgetésekre (reference chat history) – bekapcsolt állapotban az asszisztens a korábbi beszélgetésekből is meríthet a személyre szabáshoz.

Memória-összefoglaló és memóriaforrások – áttekintés arról, mit használhat fel az asszisztens a személyre szabáshoz, illetve egyes válaszok alatt az, hogy mely korábbi beszélgetések vagy fájlok befolyásolták a választ. Mindkettő javítható vagy törölhető.

Ideiglenes csevegés (temporary chat) – olyan beszélgetés, amely nem hoz létre és nem frissít emlékeket.

Csak projekten belüli memória (project-only memory) – egy projektben a kontextus a projekt beszélgetéseire és fájljaira korlátozódik (lásd az összetett feladatok kezeléséről szóló tananyagot).

Fontos

A memória nem adatbázis. Az asszisztens maga választja ki, mit tart fontosnak; nem őriz meg minden részletet, a tárolt információ elavulhat vagy hibás lehet, és olyan helyzetben is felhasználhatja, ahol nem számítunk rá. Rendszeresen nézzük át a memória-összefoglalónkat.

Mit jelent mindez az oktatásban?

Folytonosság

Mivel a rövid távú kontextus a beszélgetések között elvész, a folytonosságot meg kell tervezni.

A tananyagfejlesztéshez érdemes rövid „kontextuscsomagot” fenntartani: a témakör céljait, a csoport szintjét (személyes adatok nélkül), a tananyag stílusát és formátumát. Ezt új beszélgetés elején bemásolhatjuk, vagy utasításként és fájlként eltárolhatjuk egy projektben. Ez biztosabb megoldás, mint abban bízni, hogy a memória megjegyzi a részleteket.

Személyre szabás

A memória időt takaríthat meg: az asszisztens megjegyezheti például azt, hogy járműmechatronikát tanítunk 16 éveseknek, és megoldókulcsos táblázatokat szeretünk. Ugyanakkor a személyre szabás beszűkítheti a válaszokat vagy téves feltevéseket vihet tovább. Ha semleges válasza van szükségünk, vagy azt szeretnénk látni, hogyan viselkedik az eszköz egy tanuló esetében, használjunk ideiglenes csevegést.

Adatvédelem



A memória átnézhető, kikapcsolható és törölhető – de a legbiztonságosabb adat az, amelyet meg sem adunk

Soha ne adjuk meg a tanulók személyes adatait (nevét, jegyeit, egészségügyi vagy családi információit, fényképét) általános célú MI-eszközöknek. Használjunk anonimizált vagy kitalált példákat.

Tartsuk be a GDPR-t és az iskola adatvédelmi szabályait, és ahol lehetséges, használjuk az iskola intézményi MI-fiókjait.

Ellenőrizzük a memória beállításait: a memória és a korábbi beszélgetésekre való hivatkozást ki lehet kapcsolni, az egyes emlékeket ki lehet törölni. A szolgáltatók a törölt emlékek naplóját korlátozott ideig megőrizhetik (a ChatGPT esetében legfeljebb 30 napig).

Legyünk óvatosak, amikor órán kivetítjük a képernyőt: a memórián alapuló személyre szabás vagy az oldalsávban látható beszélgetéscímek olyan információt is felfedhetnek, amelyet nem szerettünk volna megmutatni.

Tanítsuk meg a tanulóknak is ugyanezt a digitális higiénéit: mit jegyez meg róluk az MI, hogyan ellenőrizhetik és hogyan törölhetik.

Tanulságok tanároknak

A modell minden választ az aktuális bemenetből számol ki; nem úgy emlékszik, mint egy ember.

A kontextusablak véges: a hosszú beszélgetések lassabbak, zajosabbak, és végül a rendszer tömöríti őket.

Új téma – új beszélgetés. A fontos döntéseket ellenőrzött összefoglalóval vagy kontextus-csomaggal vigyük tovább.

A memória kicsi, szelektív és szerkeszthető; nézzük át, és kapcsoljuk ki, ha nem helyénvaló.

Mutassunk konkrét példákat és fájlokat: a modell abból dolgozik a legjobban, ami előtte van.

Védjük a tanulók adatait: anonimizáljunk, használjunk intézményi fiókot, és érzékeny információt soha ne bízunk a memóriára.

Hivatkozások

Liu, N. F. és mtsai. (2024): Lost in the Middle: How Language Models Use Long Contexts. Transactions of the Association for Computational Linguistics, 12, 157–173.

OpenAI Help Center (2026): Memory in ChatGPT; Projects in ChatGPT; ChatGPT release notes.

Lengyel Zs. (2026): Bevezetés a mesterséges intelligencia fogalomrendszerébe. VITA képzési prezentáció.

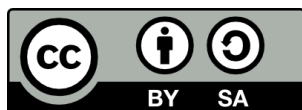
Linkek

Memória a ChatGPT-ben: help.openai.com/en/articles/8590148-memory-faq

Projektek a ChatGPT-ben: help.openai.com/en/articles/10169521-projects-in-chatgpt

Tokenizáló bemutató: tiktokenizer.vercel.app

Ez a mű a Creative Commons Nevezd meg! – Így add tovább! 4.0 Nemzetközi Licenc (CC BY-SA 4.0) feltételei szerint újra felhasználható és átdolgozható.



**Az Európai Unió
támogatásával**

Az Európai Unió finanszírozásával. A kifejtett nézetek és vélemények kizárólag a szerző(k)éi, és nem feltétlenül tükrözik az Európai Unió vagy a Nemzeti Ügynökség álláspontját.