

Beszéd felismerés, szöveg felolvasás, feliratozás és olvasástámogatás

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

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

Tanulási cél

A beszéd szöveggé és a szöveg beszéddé alakítása az MI legkiforrottabb és leghasznosabb oktatási alkalmazásai közé tartozik. A tananyag célja bemutatni, hogy hogyan lehet a beszéd felismerő, szöveg felolvasó, feliratozó és olvasástámogató eszközökkel használhatóbb tananyagokat készíteni a mindennapi tanításhoz. A tananyag végére a tanárok képesek lesznek kiválasztani azokat az eszközöket, amelyek megkönnyíthetik a speciális hallási látási, nyelvi vagy olvasás-írási igényű diákok tanuláshoz való hozzáférését, videót feliratozni, szövegekhez és magyarázatokhoz hangos változatot készíteni

Bevezetés: egy tartalom, több csatorna

A legtöbb tananyag egyetlen csatornán jut el a tanulókhoz: nyomtatott szöveggént, szóbeli magyarázatként vagy videóként. Minden egyes csatorna tanulási nehézséget okozhat valakinek – annak a tanulónak, aki nem hallja tisztán a magyarázatot, aki nem tudja elég gyorsan elolvasni a sűrű szöveget, vagy akinek félúton elfogy a figyelve.

Az MI-alapú beszéd- és olvasóeszközökkel lehetővé válik, hogy ugyanaz a tartalom egyszerre több csatornán is megjelenjen: a szöveget meghallgatjuk, a hangos beszéd tartalmát elolvashatjuk, a videót feliratozhatjuk. Ez a legkönnyebb módszer ahhoz, hogy **„a jelentés többféle megjelenítési módjának elvét”** a gyakorlatban megvalósítsuk

A „jelentés többféle megjelenítési módja” az egyetemes tanulástervezés (UDL) egyik alapelve (angolul: *multiple means of representation*). Az UDL az angol *Universal Design for Learning* rövidítése, magyarul egyetemes tanulástervezés. Az UDL olyan pedagógiai keretrendszer, amely abból indul ki, hogy nincs „átlagos tanuló”, ezért a tanítást eleve úgy kell megtervezni, hogy a lehető legtöbb ember számára hozzáférhető legyen, ahelyett, hogy utólag igazítanánk egyes tanulókhoz.

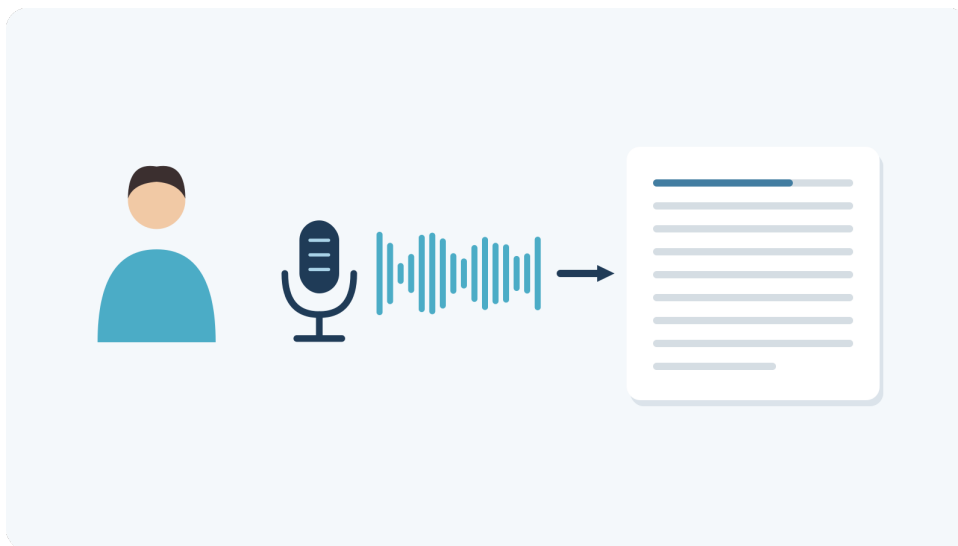
Beszédből szöveg (speech-to-text)

A beszéd felismerés mára – magyarul is – elég pontossá vált ahhoz, hogy mindennap használjuk:

- A **diktálás** be van építve a Windowsba (Win+H), a macOS-be, az Android és iOS billentyűzetekbe, a Google Dokumentumokba és a Wordbe, valamint az olyan MI-asszisztensekbe, mint a ChatGPT. Az írási nehézséggel vagy mozgáskorlátozottsággal élő tanulók beszéddel fejezhetik ki a tudásukat.
- Felvételek vagy élő beszéd **átírása**: az Androidon futó Élő átírás (Live Transcribe), a Word Átírás (Transcribe) funkciója, valamint a Teams- és

Meet-értekezletek átiratai. Online szolgáltatások percek alatt, alacsony költséggel alakítanak szöveggé egy órányi felvett beszédet.

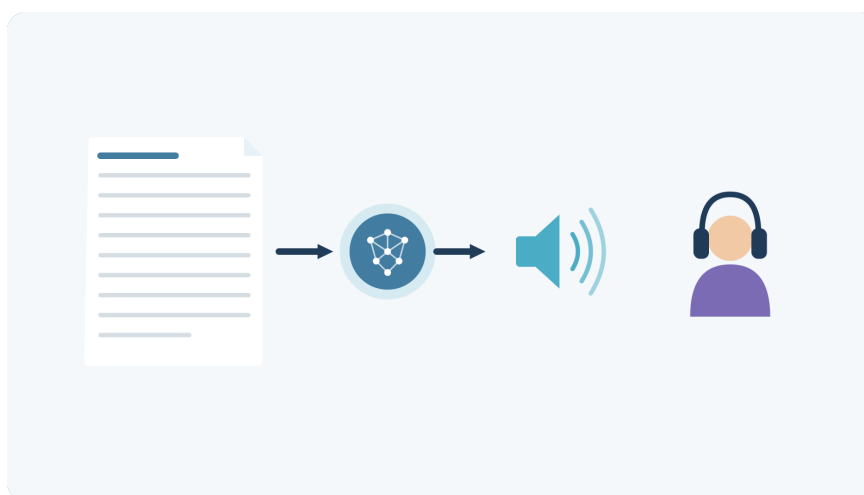
- A **magyarázatok átirata** olyan szöveget nyújt a tanulóknak, amelyet újraolvashatnak, kereshetnek benne és lefordíthatják. Ez különösen hasznos a nagyothalló tanulóknak, azoknak, akik hiányoztak, és azoknak, akik nem az anyanyelvükön tanulnak.



A beszédfelismerés a szóbeli magyarázatot szerkeszthető szöveggé alakítja

A szöveg pontosságát rontja háttérzaj, ha egyszerre többen vagy tájszólással beszélnek és ha túl sok a beszédben a speciális szakkifejezés. Beszéljünk érthetően, ha lehet, használjunk külső mikrofont, és megosztás előtt javítsuk ki a kulcsfontosságú szakkifejezéseket. Fontos: a hangfelvétel készítéséhez az érintettek hozzájárulása szükséges!

Szövegből beszéd (text-to-speech)



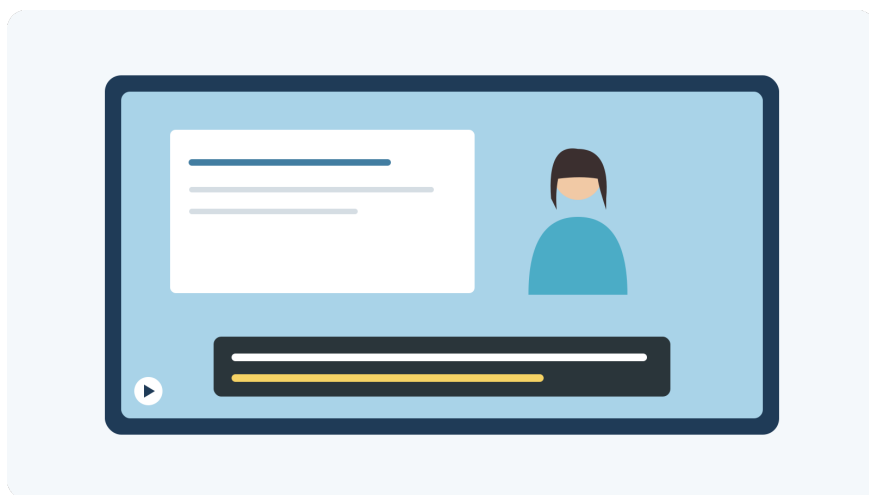
A szövegfelolvasás bármilyen szövegből természetes hangzású hangos változatot készít

A modern neurális hálók egyre természetesebb hangnemben képesek beszélni, és nagyon sok nyelven elérhetők.

- A Microsoft Edge **Hangos felolvasás** funkciója, és a Word és a OneNote **Modern olvasója** felolvassa a dokumentumokat és weboldalakokat, és közben kiemeli a felolvasott szavakat.
- A **telefonok akadálymentesítő funkciói** Androidon a **Felolvasás** (Select to Speak), iOS-en a **Képernyő felolvasása** (Speak Screen) mindent felolvasnak, ami a képernyőn látható.
- Az olyan **hanggeneráló szolgáltatások**, mint az ElevenLabs, jó minőségű narrációt készítenek a tananyagok hangos változatához.

A hangos változat segíti a diszlexiás vagy látássérült tanulókat, a nem anyanyelvű tanulókat és mindenkit, aki szívesebben hallgatja meg a tananyagot például a gyakorlati képzőhelyre utazva. Az egyidejű hallgatás és olvasás („bimodális” olvasás) javíthatja a megértést és a kitartást is.

Feliratok

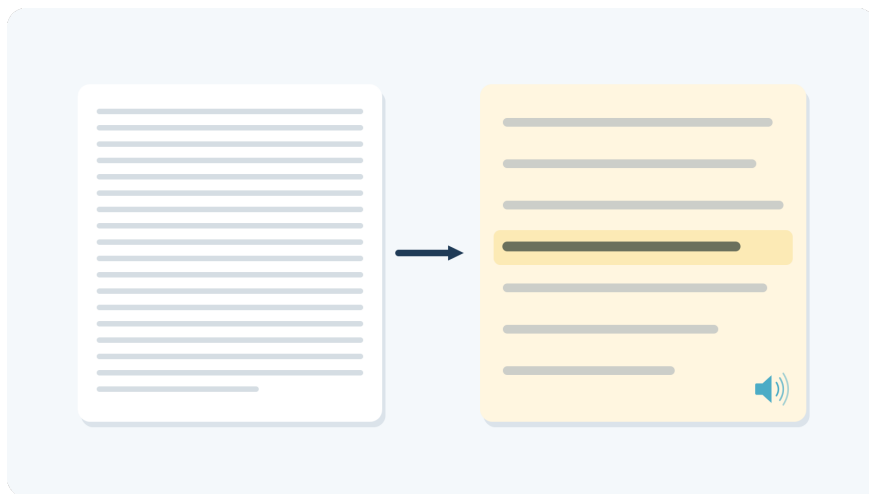


A felirat hozzáférhetővé teszi a videós magyarázatot – és zajos tanműhelyben is hasznos

A felirat nélkülözhetetlen a siket és nagyothalló tanulók számára, és a **Webes tartalmak akadálymentesítési irányelvei** (WCAG) elő is írják a feliratozást az előre rögzített videókhoz. A felirat azoknak is segít, akik zajos környezetben néznek videót, akik nem az anyanyelvükön tanulnak, vagy látni szeretnék, hogyan kell leírni az adott szakkifejezést.

- A YouTube, a Microsoft Stream és a legtöbb videóeszköz **automatikus feliratot** készít; az operációs rendszerek és a böngészők **élő feliratozást** kínálnak bármilyen hanghoz.
- Az automatikus feliratot **ellenőrizni és javítani** kell, különösen a szakkifejezéseket, a számokat és a munkavédelmi figyelmeztetéseket.
- A feliratokat automatikusan **lehet fordítani**, ami segít azoknak a tanulóknak, akik nem a saját anyanyelven tanulnak.
- A képbe égetett felirat helyett részesítsük előnyben a ki- és bekapcsolható, átméretezhető feliratot.

Olvasástámogatás



Az olvasómódok csökkentik a vizuális zsúfoltságot, kiemelik az aktuális sort, és fel is tudják olvasni a szöveget

Az olvasástámogató eszközök csökkentik a dekódolási feladatot, és segítenek abban, hogy a tanuló a jelentésre összpontosítson:

- Az **olvasómódok** (Modern olvasó, böngésző olvasási nézete) eltávolítják a zavaró elemeket, és lehetővé teszik a betűméret, a térköz, a háttérszín és a sorfókusz beállítását; egyesek szótagokra is bontják a szavakat, vagy képet mutatnak a gyakori szavakhoz.
- **MI-alapú egyszerűsítés és összefoglalás:** egy nyelvi modell alacsonyabb megértési szintre írhatja át a szöveget, rövid szakaszokra bonthatja, címsorokat és összefoglalót adhat hozzá – a szakkifejezések megtartásával és elmagyarázásával.
- Egy témakör kulcsfogalmainak **fogalomtára** egyszerű magyarázatokkal és képekkel minden tanulónak segít, különösen a második nyelven olvasóknak.

Recept: egy feladatlap, több csatorna (kb. 20 perc)

1. Ellenőrizzük a szöveg szerkezetét: világos címsorok, rövid bekezdések, hosszú mondatok helyett használjunk felsorolásokat.
2. Készítsünk hangos változatot a **Hangos felolvasás** funkcióval vagy egy szövegfelolvasó eszközzel.
3. Kérjünk az MI-től egyszerűsített változatot és 8–10 kulcsfogalomból álló fogalomtárat – majd ellenőrizzük mindkettőt.
4. Lássuk el felirattal a bemutató videót, és javítsuk ki a szakkifejezéseket.
5. Minden változatot kínáljunk fel az egész csoportnak, hogy a tanulók választhassanak.

Korlátok és jó gyakorlat

- Az MI által készített átirat, felirat és egyszerűsítés hibákat tartalmazhat – a megosztott tartalom helyességéért a tanár felel.
- Ha a hangfelvételeket és dokumentumokat a felhőben is feldolgozzuk: tartsuk be az adatvédelmi szabályokat, és ne töltsünk fel személyes

adatot. Érzékeny anyagnál a helyben futó eszközök vagy az intézményi fiókok biztonságosabbak.

- A támogató eszközök nem helyettesítik az olvasás és az írás tanítását; hozzáférést adnak a tartalomhoz, amíg ezek a készségek fejlődnek.
- Előírás helyett kínálunk választási lehetőséget: a tanulók tudják a legjobban, melyik formátum segít nekik.

Tanulságok tanároknak

- Ugyanazt a tartalmat kínáljuk több csatornán: szöveggént, hangként és feliratos videóként.
- A diktálás és az átírás segít azoknak a tanulóknak, akiknek nehézséget okoz az írás vagy a hallás utáni megértés.
- A természetes magyar hangokkal bármilyen szöveg hangos változata gyorsan elkészíthető.
- Az automatikus feliratot és átíratot mindig ellenőrizzük, különösen a szakkifejezéseket és a munkavédelmi információkat.
- Az olvasómódok, az egyszerűsített változatok és a fogalomtárak csökkentik az olvasás terhét, és segítik a megértést.

Hivatkozások

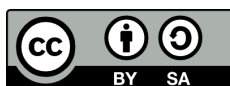
- W3C (2023): Web Content Accessibility Guidelines (WCAG) 2.2 – 1.2.2 Feliratok (előre rögzített) sikerkritérium.
- CAST (2024): Universal Design for Learning Guidelines version 3.0.
- Microsoft Support: Modern olvasó és Hangos felolvasás – nyelvek és hangok.
- Lengyel Zs. (2026): Konkrét alkalmazások tartalomgenerálásra. VITA képzési prezentáció.

Linkek

- UDL útmutató: udlguidelines.cast.org
- Webes tartalmak akadálymentesítési irányelvei: w3.org/WAI
- A Modern olvasó nyelvei és hangjai: support.microsoft.com
- Élő átírás és hangértékesítések (Live Transcribe & Sound Notifications): play.google.com
- Feliratok – W3C: w3.org/WAI/media/av/captions

Szerzői jog © 2026 VITA Konzorcium

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.



Azokat a képeket, amelyeknél nincs forrásmegjelölés, a Claude for Mac generálta



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.

