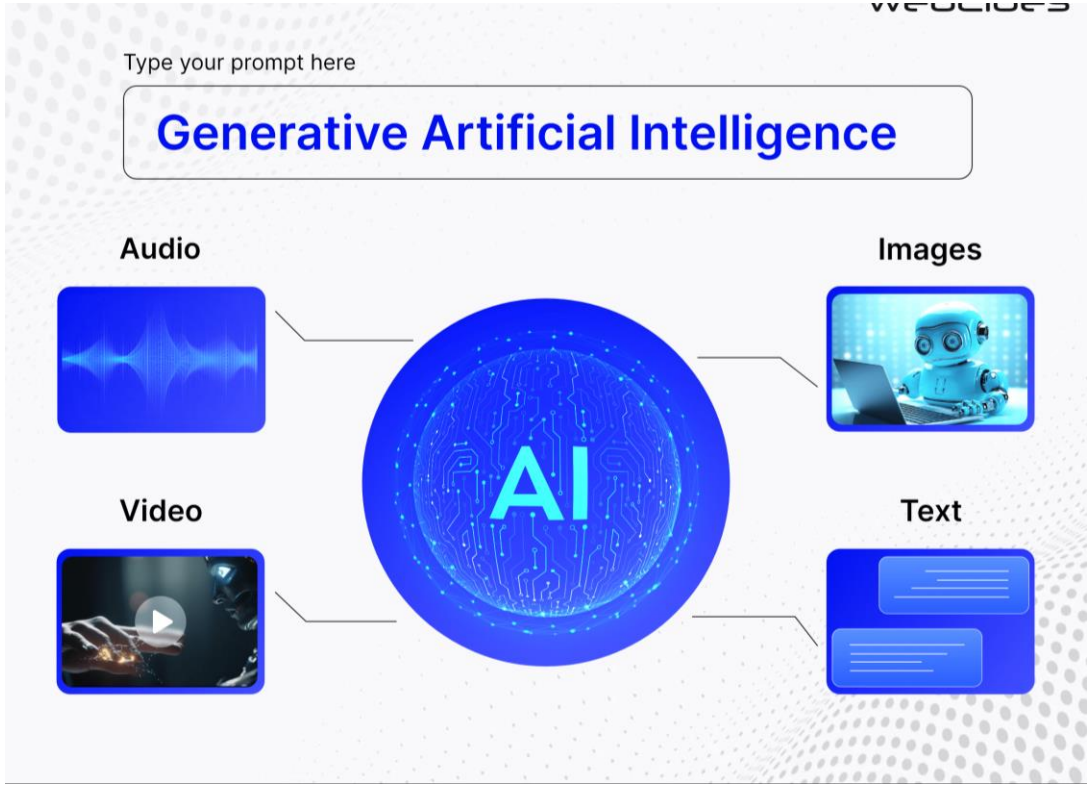
# Az európai Mesterséges Intelligencia Rendelet generatív MI-re vonatkozó szabályai

Megvéd-e bennünket a jog az MI-től?

# Generatív MI

- A generatív mesterséges intelligencia (más néven: generatív AI, GenAI, vagy GAI) a mesterséges intelligencia egy olyan alcsoportja, amely generatív modelleket használ szövegek, képek, videók vagy más adatformák előállítására.
- Ezek a modellek megtanulják a tanítóadatok mintázatait és struktúráit, és ezeket használják fel új adatok előállítására a bemenet alapján, amely gyakran természetes nyelvi kérdések (prompt) formájában érkezik.



# A generatív MI és az MI rendelet kalandos interakciója

- Az MI rendeletet 2021 áprilisában nyújtották be
- Termékmegfelelőségi szabályozás a New Legislative Framework alapján, a 27. szabályozott termékcsoport
- Logikája az, hogy nem a technológia veszélyes önmagában (és emiatt szabályozott), hanem annak *használata* bizonyos helyzetekben, területeken (pl. munkaügyi, banki döntések, határvédelem, közigazgatás egyes döntései, kritikus infrastruktúra üzemeltetése, egyes termékek biztonsági rendszerei)
- Ezeken a területeken az MI-rendszereket eleve a Rendeletben előírtaknak megfelelően kell fejleszteni, majd folyamatosan működtetni és ezt dokumentálni
- 2022 november – berobban a ChatGPT
  - NEM területspecifikus – „bármire” használható
  - MÁR készen van, nem lehet kontrollálni a fejlesztését
  - A „belseje” üzleti titok
  - Gyaníthatóan jogellenesen szerzett adatokon tanították (szerzői jogvédett és személyes adatok)

ChatGPT

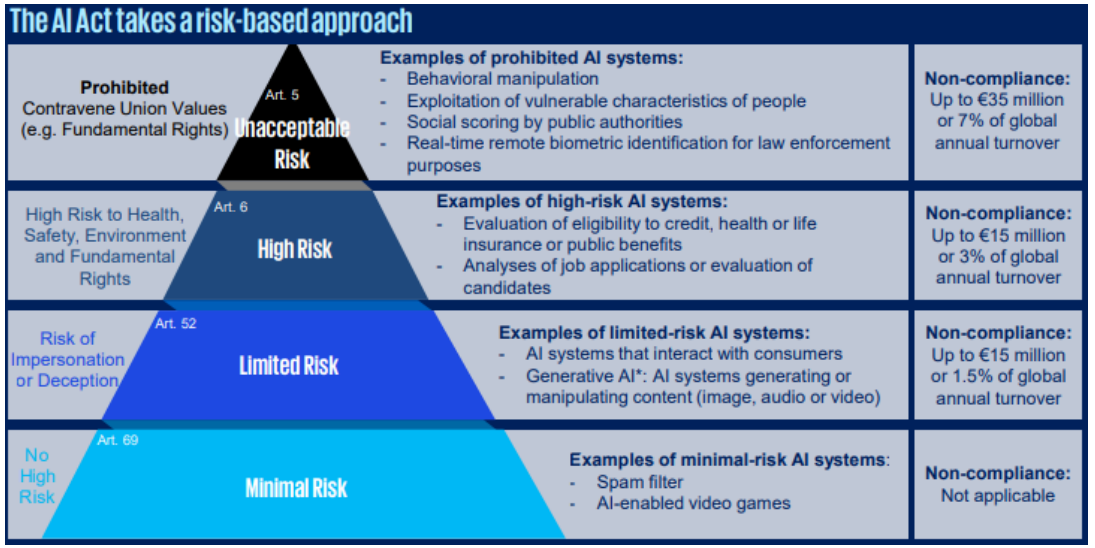
# Az MI rendelet és a generatív MI

- Az MI rendelet **nem szabályozza** a generatív MI-t kifejezetten ekként említve – két helyen szerepel egyáltalán ez a kifejezés
- az első a preambulum 99 bekezdése, amely szerint „nagy generatív MI-modellek az általános célú MI-modellek tipikus példái”
- a második a preambulum 105 bekezdésében, amelyben a szerzői jogokkal kapcsolatos aggályokat és az EU szerzői jogi irányelv adatbányászati szabályát emlegetik.

  „A szöveg- és adatbányászati technikák ebben az összefüggésben széles körben alkalmazhatók az olyan tartalmak kinyerésére és elemzésére, amelyek a szerzői és szomszédos jogok védelme alatt állhatnak. A szerzői jogi védelem alatt álló tartalom felhasználásához az érintett jogosult engedélye szükséges, amennyiben nincsenek érvényben szerzői jogi kivételek és korlátozások.”

# Akkor hogyan szabályozza?

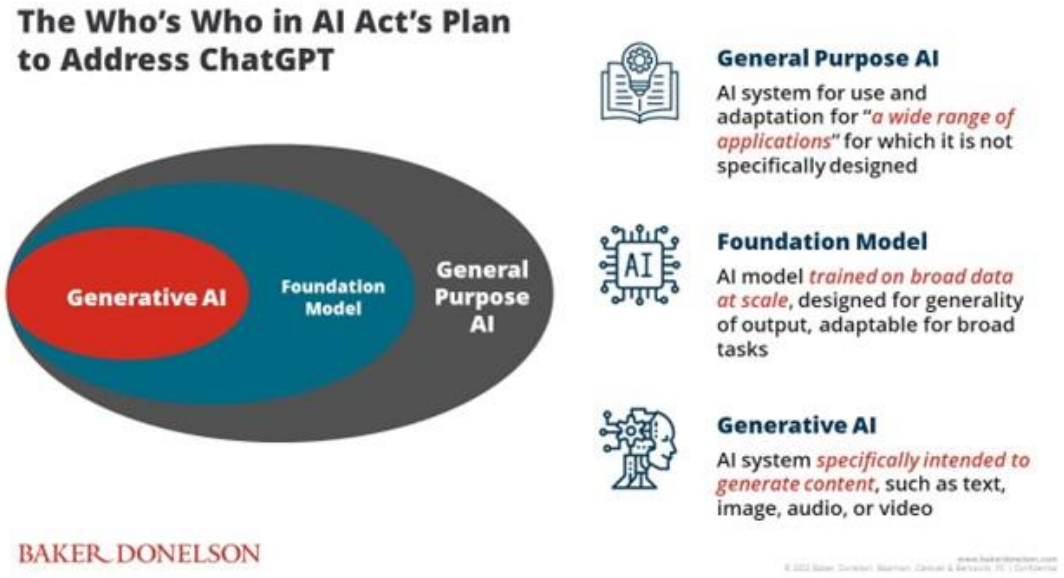
- Eredetileg az MI rendelet-tervezet négy kockázati kategóriáról beszélt
- A kockázati kategóriákat nem a technológia jellege határozza meg, hanem **a használati területek és a használati módok.**
- A generatív MI-k többféle módon (célra) és többféle területen lehet használni. Ez dönti el, hogy melyik kategóriába fog esni.
- A generatív MI a rendelet elfogadása alatt jelent meg – az eredeti logikát megtörve **külön fejezetet szúrtak be az általános célú MI modellek és a rendszerszintű kockázatot jelentő GPAIM-ek szabályozására.**



# Fogalom (3. cikk)

63. „általános célú MI-modell”: olyan MI-modell – ideértve azt is, amikor az ilyen MI-modell tanítása nagy adatmennyiséggel, nagy léptékű önfelügyelet mellett történik –, amely **jelentős általánosságot mutat**, és forgalomba hozatalának módjától függetlenül, különféle feladatok széles körének elvégzésére képes, valamint többféle downstream rendszerbe vagy alkalmazásba integrálható, azon MI-modellek kivételével, amelyeket a forgalomba hozatalukat megelőzően kutatási, fejlesztési vagy prototípus-alkotási tevékenységekre használnak;

66. „általános célú MI-rendszer”: általános célú **MI-modellen alapuló MI-rendszer**, amely – mind közvetlen felhasználás, mind más MI-rendszerekbe való integráció céljából – többféle célt képes szolgálni;



# Preambulum – célok az általános célú MI-kel kapcsolatban

(85) **Az általános célú MI-rendszerek használhatók önmagukban, nagy kockázatú MI-rendszerként vagy más nagy kockázatú MI-rendszerek alkotóelemeiként.** Ezért (...) az ilyen rendszerek szolgáltatóinak (...) **szorosan együtt kell működniük** egyrészt a releváns nagy kockázatú MI-rendszerek szolgáltatóival annak érdekében, hogy azok meg tudjanak felelni az e rendelet szerinti releváns kötelezettségeknek, másrészt az e rendelet alapján létrehozott illetékes hatóságokkal.

# Fő kötelezettség - átláthatóság

(101) Az általános célú MI-modellek szolgáltatói különleges szerepet játszanak és különleges felelősséget viselnek a mesterségesintelligencia-értéklánc mentén, mivel az általuk biztosított modellek **számos downstream rendszer alapját képezhetik, amelyeket gyakran downstream szolgáltatók biztosítanak, ez pedig a modellek és azok képességeinek alapos ismeretét teszi szükségessé, egyrészt hogy integrálni tudják a modelleket a termékeikbe, másrészt hogy eleget tegyenek az ezen, illetve más rendeletek szerinti kötelezettségeiknek. Ezért arányos átláthatósági intézkedéseket kell megállapítani**, beleértve **a dokumentáció elkészítését és naprakészen tartását, valamint az általános célú MI-modellre vonatkozó információk nyújtását a downstream szolgáltatók általi használatra**.

# Korlátozott kockázatú rendszerekre vonatkozó kötelezettségek (50. cikk)

(2) A szintetikus hang-, kép-, video- vagy szöveges tartalmat létrehozó MI-rendszerek – köztük az általános célú MI-rendszerek – szolgáltatóinak biztosítaniuk kell, hogy az MI-rendszer kimeneteit **géppel olvasható formátumban jelöljék meg**, és azok mesterségesen létrehozottként vagy manipuláltként észlelhetők legyenek.

(4) Az olyan MI-rendszerek alkalmazóinak, amelyek eredetinek vagy valóságosnak tűnő („deepfake”) kép-, hang- vagy videotartalmat hoznak létre vagy manipulálnak, **közölniük kell, hogy a tartalmat mesterségesen hozták létre vagy manipulálták**. (...) A nyilvánosság közérdekű ügyekről való tájékoztatása céljából közzétett szöveget generáló vagy manipuláló MI-rendszer alkalmazóinak közölniük kell, hogy a szöveget mesterségesen hozták létre vagy manipulálták.

# Modellszolgáltatók kötelezettségei (53. cikk)

- el kell készíteniük és naprakészen kell tartaniuk a modell **műszaki dokumentációját**, (XI. melléklet)
- információkat és dokumentációt kell kidolgozniuk, naprakészen tartaniuk és rendelkezésre bocsátaniuk az MI-rendszerek azon szolgáltatói részére, amelyek az általános célú MI-modellt **be kívánják építeni** MI-rendszereikbe (XII. melléklet)
- a **szerzői** és kapcsolódó **jogokra** vonatkozó uniós jognak való megfelelésre irányuló **politikát** kell bevezetniük
- kellően részletes **összefoglalót** kell készíteniük – az MI-hivatal által rendelkezésre bocsátott sablonnak megfelelően – és közzétenniük az általános célú MI-modell tanításához használt tartalomról.

kivéve a nyílt forráskódú rendszereket

# A XI. melléklet tartalma – általános műszaki dokumentáció

1. Az általános célú MI-modell általános leírása, beleértve a következőket:

a) a modell rendeltetése szerint **ellátandó feladatok**, valamint azon MI-rendszerek típusa és jellege, **amelyekbe integrálható**;

b) az elfogadható felhasználásra vonatkozóan **alkalmazandó irányelvek**;

c) a **forgalomba hozatal időpontja és forgalmazási módszerek**;

d) az **architektúra és a paraméterek száma**;

e) a **bemenetek és a kimenetek modalitása** (pl. szöveg, kép) és formátuma;

f) **a licenc**.

2.A modellnek az 1. pontban említett elemeinek **részletes leírása** és a fejlesztési folyamatra vonatkozó releváns információk, ideértve a következő elemeket:

a) az általános célú MI-modell MI-rendszerekbe **történő integrálásához szükséges műszaki megoldások** (pl. használati utasítás, infrastruktúra, eszközök);

b) **a modell és a tanítási folyamat – beleértve a tanítómódszereket és -technikákat is – tervezési előírásai**, a legfontosabb tervezési döntések, beleértve az indokokat és a feltételezéseket; adott esetben tervezésének megfelelően mit optimalizál a modell és a különböző paraméterek relevanciája;

c) adott esetben a **tanításhoz, a teszteléshez és a validáláshoz használt adatokra vonatkozó információk**, beleértve az adatok típusát és eredetét, valamint az adatgondozási módszereket (pl. tisztítás, szűrés stb.), az adatpontok száma, azok köre és fő jellemzői; az adatok megszerzésének és kiválasztásának módja, valamint adott esetben minden egyéb, az adatforrások alkalmatlanságának észlelésére irányuló intézkedés és az azonosítható torzítások feltárására irányuló módszerek;

d) a modell tanításához használt **számítási erőforrások** (pl. lebegőpontos műveletek száma), tanítási idő és a tanítással kapcsolatos egyéb releváns részletek;

e) a modell ismert vagy becsült **energiafogyasztása**.

# A XII. melléklet tartalma – információk a downstream szolgáltatók számára

1. Az általános célú MI-modell **általános leírása**, ideértve a következőket:

   a) **azon feladatok**, amelyeket a modell ellátni hivatott, valamint azon MI-rendszerek típusa és jellege, amelyekbe integrálható;

   b) **az elfogadható felhasználásra vonatkozóan** alkalmazandó irányelvek;

   c) **a forgalomba hozatal időpontja** és forgalmazási módszerek;

   d) adott esetben az, hogy a modell **miként működik együtt olyan hardverrel vagy szoftverrel, amely nem része magának a modellnek**, vagy miként használható fel az együttműködésre;

   e) adott esetben az általános célú MI-modell használatához kapcsolódó **releváns szoftver verziói**;

   f) **az architektúra és a paraméterek száma**;

   g) **a bemenetek és a kimenetek modalitása (pl. szöveg, kép) és formátuma**;

   h) **a modell licence**.

2. **A modell elemeinek és fejlesztési folyamatának leírása**, ideértve a következőket:

   a) az általános célú MI-modell MI-rendszerekbe történő integrálásához **szükséges műszaki megoldások** (pl. használati utasítás, infrastruktúra, eszközök);

   b) **a bemenetek és a kimenetek modalitásai** (pl. szöveg, kép stb.) és formátuma, valamint ezek maximális mérete (pl. a kontextusablak hossza stb.);

   c) adott esetben a **tanításhoz, a teszteléshez és a validáláshoz használt adatokra vonatkozó információk**, beleértve az adatok típusát és eredetét, valamint az adatgondozási módszereket.

# A rendszerszintű kockázatot jelentő általános célú MI-modellek szolgáltatóinak kötelezettségei

- **modellértékelést** kell végezniük – külsős bevonásával – ez most épp kimaradt az új verzióból
- **értékelniük és enyhíteniük** kell az esetlegesen az MI-modellek fejlesztéséből, forgalomba hozatalából vagy használatából eredő lehetséges, uniós szintű **rendszerszintű kockázatokat**
- **nyomon kell követniük,** dokumentálniuk kell és indokolatlan késedelem nélkül jelenteniük kell az MI-hivatal és adott esetben az illetékes nemzeti hatóságok részére a **súlyos váratlan eseményekre** és az azok kezelésére szolgáló lehetséges korrekciós intézkedésekre vonatkozó releváns információkat
- megfelelő szintű **kiberbiztonsági védelmet** kell biztosítaniuk a rendszerszintű kockázatot jelentő általános célú MI-modell és a modell fizikai infrastruktúrája számára
- rendszerszintű kockázatot jelentő általános célú MI-modellek szolgáltatói egy harmonizált szabvány közzétételéig támaszkodhatnak az 56. cikk szerinti **gyakorlati kódexekre**

# Merre tovább?

- Látható, hogy a GPAIM szolgáltatók számára csak nagyon általánosan vannak meghatározva a kötelezettségek, és ezek is nagyrészt csak a **transzparenciára** vonatkoznak
- A GPAIM-ekre vonatkozó gyakorlati kódexek kidolgozása a harmadik fordulójánál tart – a kötelezettségeken folyamatosan lazítottak
- Tegnapi hír, hogy Axel Voss EU parlamenti képviselő dörgedelmes levelet írt a Bizottságnak, hogy ne engedjen az amerikai GPAIM yzolgáltatóknak, és ne próbálja a kötelezettségeket az európai downstream szolgáltatókra terhelni.

*„The latest draft reveals a troubling capitulation to the wishes of a handful of powerful upstream companies. And it risks undermining many of the key concepts of Chapter V and IX in the #AIAct that have been carefully integrated by the European Parliament and Council of the European Union. (...) Backed by their governments, most foreign GPAI providers will likely ignore EU enforcement anyway.”*

# Konklúziók

- Az AI Act a generatív MI-vel kapcsolatban szinte kizárólag **transzparenciakövetelményeket** határoz meg – a tartalmi szabályozást a hagyományos jogágakra hagyja
- A deepfake és szintetikus tartalmaknál csak **címkézési** kötelezettség van.
- Az általános célú modelleknél szintén csak tájékoztatási kötelezettség van, főleg  a downstream szolgáltatók felé – de ennek konkrétumaiban (a GPAIM magatartási kódexben) is úgy tűnik **az enyhébb kötelezettségek** érvényesülnek
- **Ez nem biztos hogy akkora baj**, mert az európai szolgáltatóknak is ad egy esélyt.
- Ez a szabályozás nem fog bennünket megvédeni az MI kockázataitól, legfeljebb esélyt teremt arra, hogyha rosszfelé fordulnak a dolgok, akkor lehessen tudni mi volt a baj.