

GERVIGREIND

TÆKNIBYLTING, DEUS EX MACHINA
EÐA UPPHAF AÐ NÆSTA MARKAÐSHRUNI?



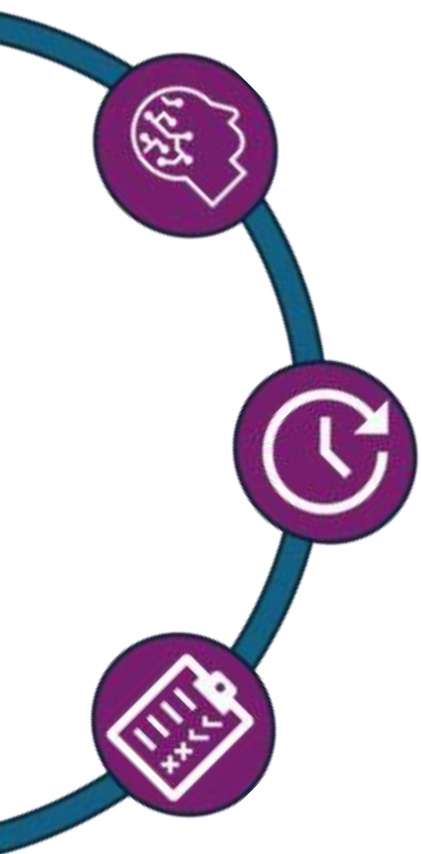
GERVIGREIND

- Hvað er gervigreind?
- Til hvers er hún nýtileg?
- Hvaða hættur hefur hún í för með sér?
- Gervigreind er GPT
- Hvað gerir gervigreind frábrugðna öðrum GPT sem hafa umbylt heiminum?

Gervigreind mun hafa í för með sér djúpstæðar breytingar á samfélaginu

... Eða hvað?

- Fjárfestingar og fjármálageirinn
- Arðbærni í rekstri
- Mannauður



Reglusetning - hvernig mun hún þróast

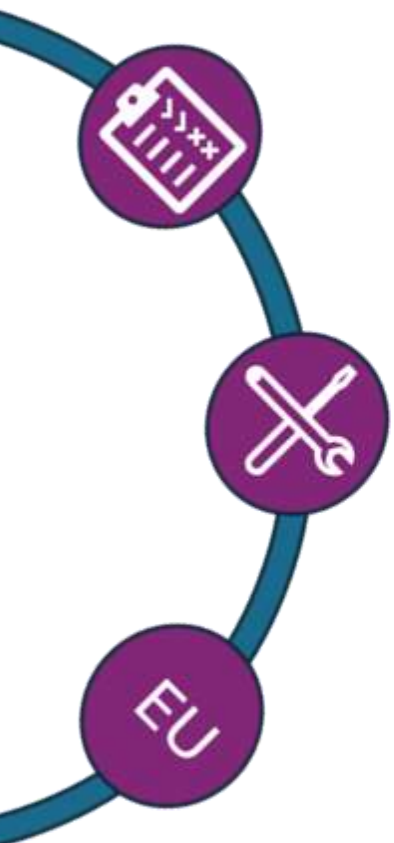
Markmiðið: fit-for-purpose og future-proof

- Við hverju á regluverk um gervigreind að bregðast?
- Tæknin þróast of hratt til þess að hefðbundin löggjöf geti brugðist við í tæka tíð.
- Mismunandi markmið og mismunandi nálganir í ólíkum heimshlutum.
- Þroskað regluverk mun vera samspil af mismunandi reglum með mismunandi gildissvið.
 - Siðareglur og sjálfseftirlit.
 - Samningar og kvaðir í nytjaleyfum.
 - Staðlar (tæknilegir og rekstrar)
 - Alþjóðasáttmálar
 - Svæðisbundnar reglur
 - Landslög



Meðferð stjórnsýslumála og stjórnvaldsákvarðanir – Helstu skyldurnar sem hvíla á stjórnvöldum

1. Hlutlægni,
2. Leiðbeiningarskylda,
3. Málshraðareglan,
4. Rannsóknarreglan,
5. Jafnræðisreglan,
6. Meðalhófsreglan,
7. Upplýsingaréttur aðila máls,
8. Gagnsæi og rekjanleiki



Hvernig getur gervigreind hjálpað við að framfylgja þessum skyldum?

1. Hlutlægni,
 - sjálfvirk könnun hagsmunaárekstra
2. Leiðbeiningarskylda,
 - spjallmenni, pre-populated meta data
3. Málshraðareglan,
 - aukin skilvirkni
4. Rannsóknarreglan,
 - margfalt öflugri leitarvél
5. Jafnræðisreglan,
 - samanburður niðurstaðna sambærilegra mála, fordæmaleit til leiðbeiningar,
6. Meðalhófsreglan,
 - möguleikar flokkaðir og metnir eftir hversu íþyngjandi
7. Upplýsingaréttur aðila máls,
 - auðveldar samantekt skjala, eyðingu persónugreinanlegra upplýsinga
8. Gagnsæi og rekjanleiki
 - Kallar á skráningu, varðveislu, loggun, tímastimplanir o.s.frv.

Er stjórnvöldum e.t.v. skylt að nýta gervigreind?



Gervigreindarlöggjöf Evrópusambandsins

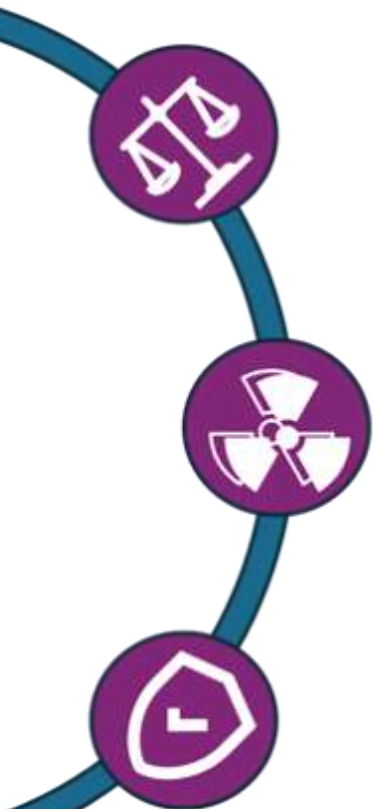
- Birt OJEU 12. júlí 2024. Tók gildi 1. ágúst 2024. Tilteknar skyldur samkvæmt reglugerðinni verða virkar í nokkrum skrefum (aðlögunartími). Að mestu bindandi frá 2. ágúst 2026, en eldri kerfi hafa aukið svigrúm til ársloka 2027 (og jafnvel lengur).
- Staðan hjá EES ríkjunum og Íslandi.
 - AI Act er EES-tæk reglugerð.
 - Sameiginlega EES nefndin hefur ekki afgreitt reglugerðina – ekki enn hluti af EES samningnum.
 - Má vænta þess að það gerist seint á árinu 2025 eða fyrri hluta 2026.
 - Eftir það þarf að innleiða reglugerðina í lög á Íslandi.
 - Aðgerðaáætlun stjórnvalda gerir ráð fyrir að frumvarp liggi fyrir vorið 2026.
 - Líklegt að það verði pressa á að innleiða í kringum ágúst 2026.
 - Ekki ólíklegt að frestir á gildistöku tiltekinna skyldna verði lengdir.



Hvað felst í EU AI Act?

- Jafnvægisæfing – reynt að hvetja til nýsköpunar og vaxtar en um leið að draga úr hættum.
- Á að stuðla að öruggri og traustverðugri tækni og að verja grundvallarréttindi borgaranna.
- Aðferðin er áhættumiðað regluverk. Skyldurnar eru strangari eftir því sem hættan er meiri.

Tegund áhættu	Dæmi	Afleiðingar
Óásættanleg áhætta: Skýr og veruleg hættu fyrir öryggi, grundvallarréttindi og samfélagsleg gildi.	Tækni sem hefur áhrif á undirmeðvitund. Tækni sem nýtir veikleika jaðarsettra hópa. Félagsleg stigagjöf stjórnvalda.	Bannað
Mikil hættu. Umtalsverð hættu á öryggi, lífsviðurværi og grundvallarréttindi.	Gervigreind í læknisfræðilegri greiningu og meðferð; einkunnagjöf; í löggæslu, réttarvörslu- og refsivörslu	Strangar kröfur
Lítil áhætta.	Þjónustuviðmót. Sýndar-aðstoð.	Aðallega gagnsæiskröfur
Kerfisbundin áhætta.	Hápróuð almenn mállíkön – ChatGPT. Gervigreind í mikilvægustu innviðum.	Umfangsmiklar og strangar kröfur sem taka til fjölda sviða



Hvað telst til mikillar áhættu?

- Annars vegar gervigreind í vörum sem þegar falla undir vöruöryggisreglur.
- Hins vegar 8 flokkar.

Svið	Dæmi
Lífkennaauðkenning. greining tilfinninga.	Greining andlita á opinberum svæðum. Lífkennaauðkenning byggð á kynþætti, trúarbrögðum, kyni o.sfrv. Kerfi sem meta tilfinningaástand.
Mikilvægir innviðir.	Kerfi sem stórna öryggi stafrænna inniðva, umferðaröryggi, rafmagnskerfum o.s.frv.
Kennsla og þjálfun.	Kerfi sem ákveða inngöngu í nám, meta námsárangur, meta menntunarstig, greina svindl í prófum
Vinnutengd.	Við ráðningu. Við ákvörðun um stöðuhækkun og úthlutun verkefna. Við frammistöðumat.
Aðgengi að mikilvægri þjónustu.	Kerfi sem meta rétt til þjónustu og aðstoðar. Meta lánshæfi. Verð líftrygginga. Forgangsröðun sjúklinga.
Löggæsla.	Kerfi sem nota gervigreind sem lygamæli. Meta gildi sönnunargagna. Meta hættu á að einstaklingur fremji afbrot byggt á persónueinkennum og fyrri hegðun.
Fólksflutningar og hælisleitendur.	Kerfi sem aðstoða við afgreiðslu umsókna. Meta einstaklinga.
Dóms- og kosningakerfi.	Rannsóknir, túlkun laga og beiting laga á tiltekin mál. Hafa áhrif á útkomu kosninga.

Hvaða auknu skyldur fylgja mikilli hættu?

Grein	Efni	Lykilatriði
10	Gagnastjórnun- og gagnagæði	• Setja gagnastjórnunarkerfi sem hentar tilgangi gervigreindartólsins. • Koma á ferlum sem tryggja gæði gagna og þjálfunargagna.
12	Skráning og skráarhald	• Kerfi verða að tryggja sjálfvirka skráningu. • Tryggja rekjanleika þegar hættu er á neikvæðum áhrifum á grundvallarréttindi. • Halda skrár yfir breytingar á kerfinu.
13	Gagnsæi	• Notendur verða að skilja notkun og túlkun niðurstaðna. • Veita skýrar leiðbeiningar um tilgang og tæknilega getu. • Dæmi: læknisfræðileg gervigreind útskýrir hvers vegna það mælir með tiltekinni meðferð og játar ef það er óvissa fyrir hendi.
14	Mannlegt eftirlit	• Skilvirkt mannlegt eftirlit þarf að vera (auð)framkvæmanlegt. • Viðmót fyrir samskipti manneskju og vélar. • Manneskja verður að geta skilið takmarkanir kerfisins og gripið inn í eða stöðvað ef þörf er.
15	Nákvæmni, áreiðanleiki og netöryggi	• Kerfi þarf að viðhalda nægri nákvæmni, áreiðanleika og öryggi. • Nota skýrar viðmiðanir (benchmarks) við mat á nákvæmni. Áreiðanleiki og mótstaða gegn bilunum og villum. Öryggisviðbragð til að koma í veg fyrir kerfishrun.

