

Spáum fyrir um framtíð gervigreindar

Spáum fyrir um framtíð gervigreindar

út frá akademísku sjónarmiði

Dr. Helga Ingimundardóttir (helga@hi.is)

Stjórnskipti 22. febrúar 2025

BRUNNARFORSKUNARDEILD

Grunnáherslur:

- Áhrif á fræðasamfélagið.
- Hagnýtar tengingar við atvinnulíf og nýsköpun.
- Mikilvægi ábyrgðar og sjálfbærni.

Spáum fyrir um framtíð gervigreindar

└ Áhrif á háskólakennslu

ChatGPT breytir kennslu. Kennarar geta lagt meiri áherslu á gagnrýni og flóknari verkefni. Nemendur nýta stoðtækni til að læra dýpra.

Nýjar áskoranir

- ChatGPT og CoPilot breyta hvernig nemendur læra.
- Þörf fyrir gagnrýni á niðurstöður frá stórum módelum.

Tækifæri

- Meiri áhersla á umfangsmeiri og flóknari verkefni.
- Stoðtæki gerir flóknar áskoranir leysanlegar fyrir nemendur.



Spáum fyrir um framtíð gervigreindar

└ Gagnasiðfræði og ábyrg notkun

Siðferðileg notkun gervigreindar verður mikilvægari. EU AI Act setur skýra ramma um ábyrgð og gagnsæi.

EU AI Act er regluverk Evrópusambandsins sem tryggir örugga, gagnsæja og siðferðilega notkun gervigreindar með áherslu á áhættumat og ábyrgð.

Við kennum gagnasiðfræði sem hluta af valáfanga í iðnaðarverkfræði, kennt af Henry Alexander Henrysson við Siðfræðistofnun HÍ, en í ljósi hversu mikilvægt þetta er fyrir framtíðina ætlum við að færa þessa lotu inn í skyldunámskeið, og hvetja aðrar verkfræðideildir til að láta sína nemendur taka gagnasiðfræði líka.

Mikilvægi siðfræði

- ▶ Siðferðileg notkun gervigreindar verður lykilatriði.
- ▶ Kennsla í gagnasiðfræði verður skylda.

EU AI Act

- ▶ Áhenda á gagnsæi, ábyrgð og áhættumat.
- ▶ Alþjóðlegur staðall sem mun hafa áhrif á þróun gervigreindar.



Spáum fyrir um framtíð gervigreindar

└ Sjálfbærni og hagkvæmni

Sjálfbærni er lykilatriði. Með nýrri tækni má draga úr kostnaði og umhverfisáhrifum. DeepSeek er gott dæmi um hagkvæma nálgun.

DeepSeek, kínverskt gervigreindarfyrirtæki, kynnti nýlega DeepSeek V3, stórt mállíkan fyrir chatbot með 600 milljarða vigra, þjálfað á 14,8 trilljónum tákna. Þjálfunarkostnaður þess var undir 6 milljónum Bandaríkjadala, sem er mun lægra en þjálfunarkostnaður GPT-4 frá OpenAI, sem var um 78 milljónir dala.



Spáum fyrir um framtíð gervigreindar

└ Réttlæti í gervigreindarmódelum

Gögn þurfa að vera fjölbreytt og sanngjörn. Sanngirni í þróun móðela skiptir lykilmáli til að byggja traust.



Ahrif gagna

- Gögnin móta niðurstöður móðela – þau þurfa að vera fjölbreytt og sanngjörn.

Markmið

- Sanngirni í gagnasöfnun og markföllum.
- Sanngirni sem lykilatriði frá upphafi líkanagerðar.

Spáum fyrir um framtíð gervigreindar

└─ArtInRare: Gervigreind og listir

Artistic Intelligence - Responsiveness, accessibility, responsibility, equity (ARTinRARE) er samevrópskt verkefni sem miðar að því að styðja við listamenn og rannsakendur. Verkefnið leggur áherslu á að þróa gervigreind sem stuðlar að nýsköpun án þess að skerða réttindi eða sjálfstæði listafólks. Það sýnir hvernig gervigreind getur orðið hluti af skapandi ferli.

Um verkefnið

- Samevrópskt rannsóknarverkefni sem hófst haustið 2024.
- Markmið: Samtíðinn gervigreindar og listrannsóknar á siðferðilegan hátt.
- Stendur yfir til haustið 2028.

Viðfangsefni

- Siðferðileg nálgun á notkun gervigreindar í listum.
- Hvernig má efla sköpunarferlið með nýsköpun í gervigreind.
- Áhersla á samstarf milli fræðigreina og listafólks.