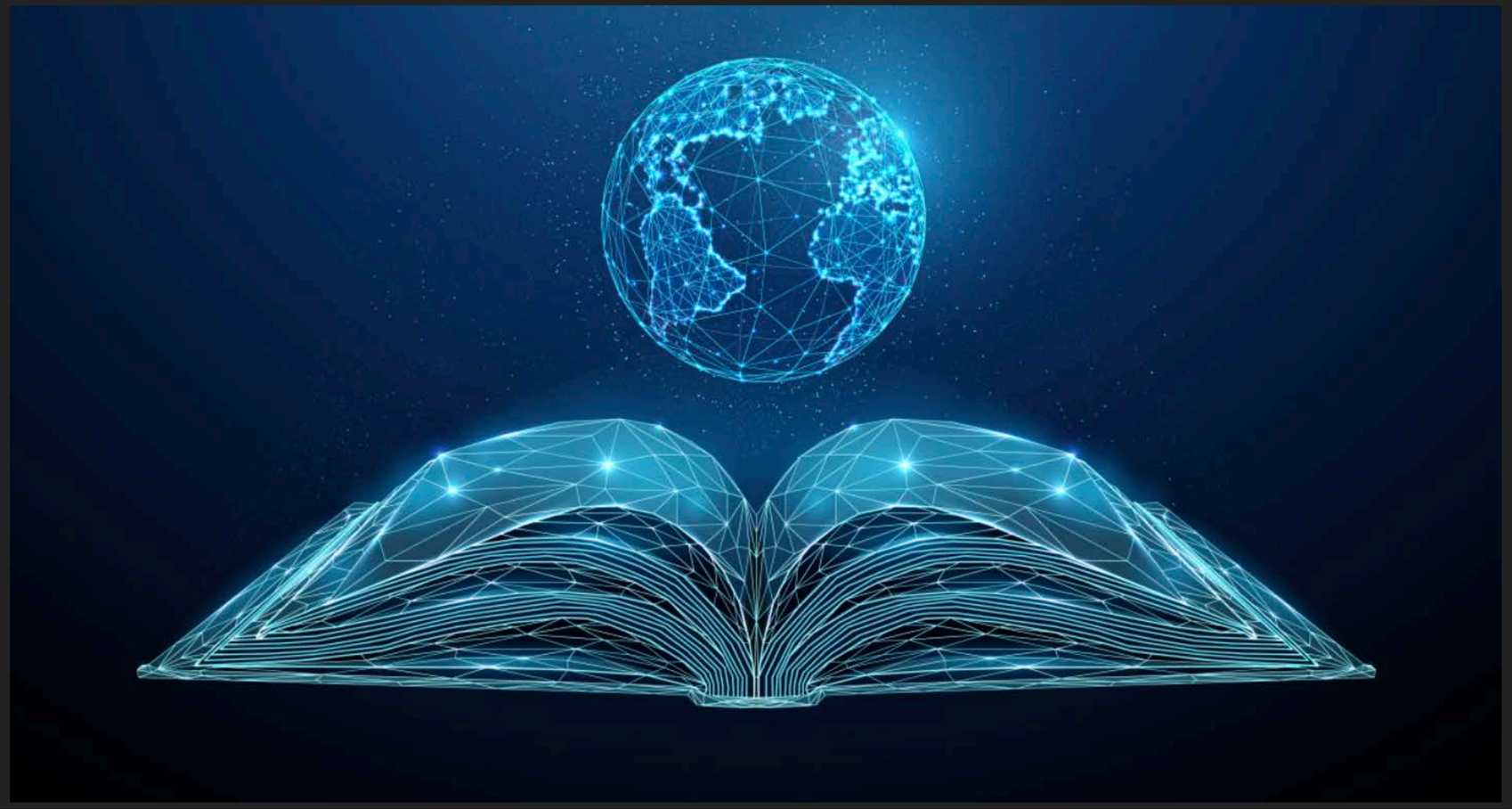


SIMO KYLLÖNEN

AVOIN TIEDE JA TEKOÄLY

EETTISET HAASTEET JA HYVÄT KÄYTÄNNÖT

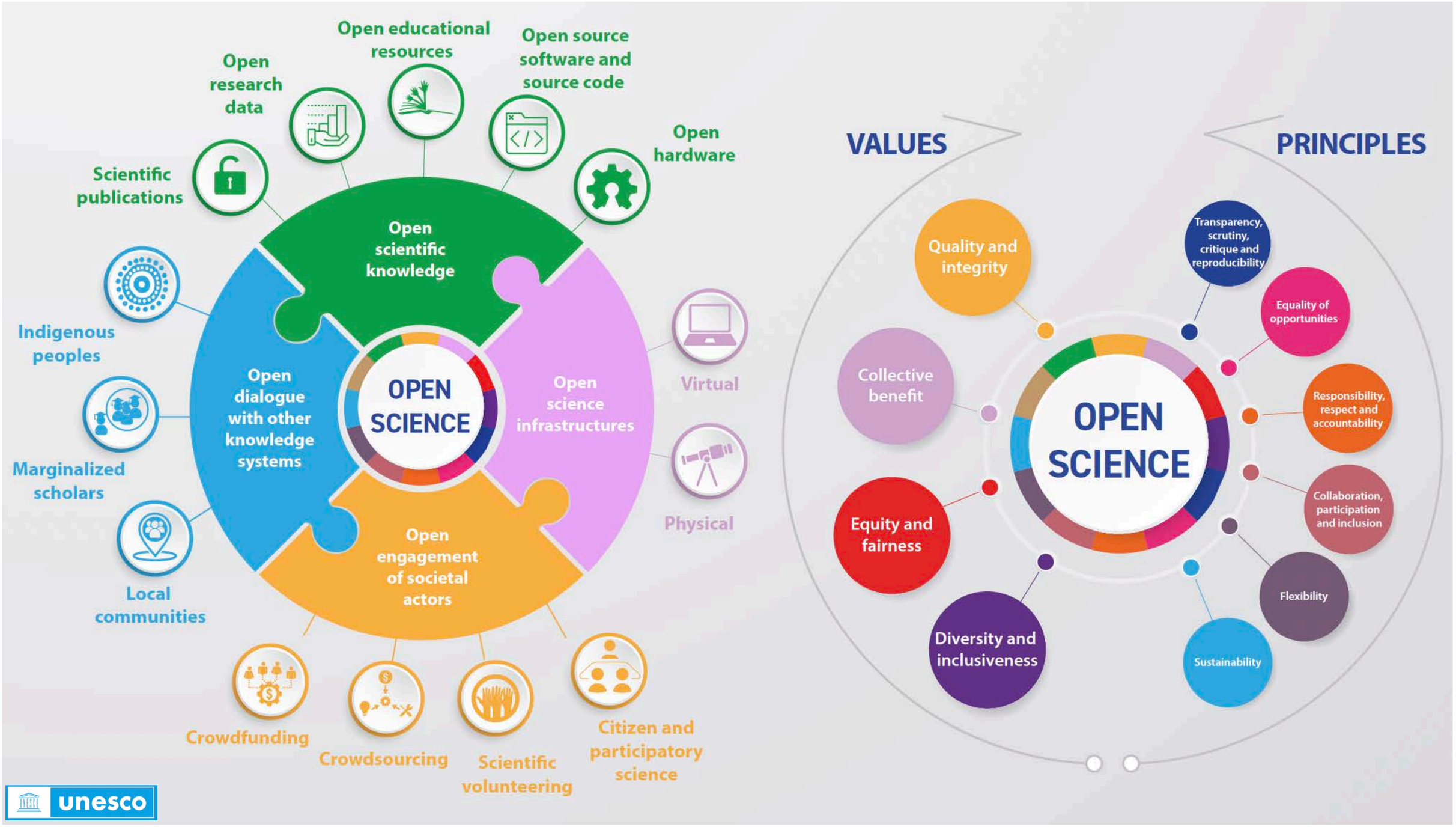
The open science approach, when combined with AI technologies, can significantly contribute to addressing the current pressing global threats.

NASA Fights *Climate Change* with Open Science and AI

Building a Powerful AI Tool for Earth Science

In 2023, NASA teamed up with IBM to create a powerful AI tool for studying Earth. This tool is like a basic building block, trained on a massive dataset of NASA's satellite images. Anyone can use this tool for free to build their own studies on climate challenges.

Think of it like a Swiss Army Knife for science. It can't do everything by itself, but it can be a starting point for lots of different projects. Scientists can take this basic tool and add specific information to make it even better at a particular task.


Open educational resources
Open research data
Open source software and source code
Open hardware
Scientific publications
Open scientific knowledge
Indigenous peoples
Open dialogue with other knowledge systems
Open science infrastructures
Virtual
Marginalized scholars
OPEN SCIENCE
Physical
Local communities
Open engagement of societal actors
Crowdfunding
Crowdsourcing
Scientific volunteering
Citizen and participatory science
VALUES
PRINCIPLES
Quality and integrity
Transparency, scrutiny, critique and reproducibility
Equality of opportunities
Collective benefit
Responsibility, respect and accountability
OPEN SCIENCE
Collaboration, participation and inclusion
Equity and fairness
Flexibility
Diversity and inclusiveness
Sustainability

Open Science at the Generative AI Turn: An Exploratory Analysis of Challenges and Opportunities

Mohammad Hosseini*[1], Serge P.J.M. Horbach[2], Kristi Holmes[1,3], Tony Ross-Hellauer[4]

[1] Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America

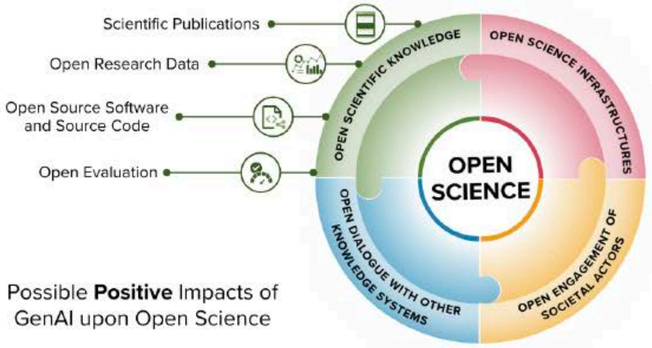
[2] Institute for Science in Society, Radboud University, Nijmegen, The Netherlands

[3] Galter Health Sciences Library and Learning Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America

[4] Open and Reproducible Research Group, Know-Center GmbH and Institute for Interactive Systems and Data Science, Graz University of Technology, Graz, Austria

* Corresponding author

Email: mohammad.hosseini@northwestern.edu

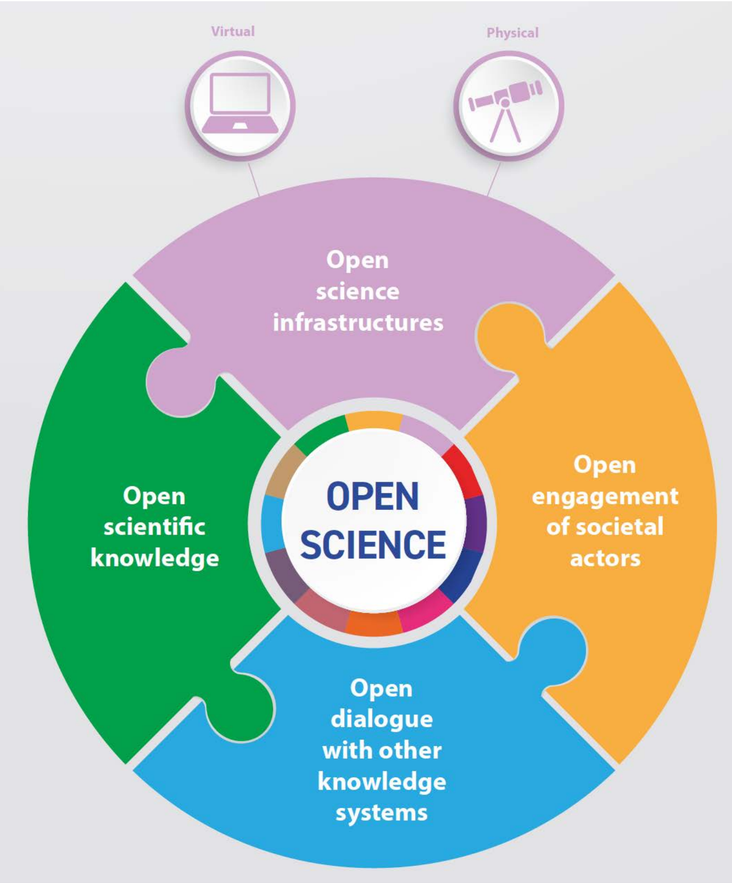


IMPACTS OF GENERATIVE AI UPON OPEN SCIENCE

× GENAI

Possible **Positive** Impacts of GenAI upon Open Science		Possible **Negative** Impacts of GenAI upon Open Science
Open Scientific Knowledge		
• Transitioning from open to meaningful and equitable access for researchers, policymakers and laypersons. • Helping researchers cope with information overload.	**Scientific Publications (Open Access)**	• Enabling paper mills and fake publications, and increased biases and misinformation about science. • Deterring researchers from open sharing of research out of a fear of having content harvested without attribution.
• Supporting Research Data Management and data sharing documentation, data analysis and curation. • Detecting errors or inconsistencies in datasets. • Creating synthetic datasets.	**Open Research Data**	• Generating fabricated datasets that support specific hypotheses and falsely validate fake papers. • Worsening reproducibility and integrity by enabling the use of unvalidated mathematical and statistical methods.
• Increasing productivity and efficiency in coding. • Reviewing code in real-time, thereby increasing quality of publicly available code. • Improving code documentation and reusability.	**Open Source Software and Source Code**	• Increasing code inaccuracies (due to indeterminism and variability of outputs) and biases.
• Enhancing evaluation of scholarly contributions. • Increased reviewer pool and lower burden on reviewers.	**Open Evaluation**	• Increasing the risks of fake review reports, leading to reduced value of open reports. • Compromising confidentiality.
Open Science Infrastructures		
• Enhancing efficient use of open repositories and scholarly indices.		• Compromising reproducibility since most major models are not themselves open. • Lacking proper attribution and risk of infringing copyrights or committing plagiarism. • Risking monetization of research information and leading to further restrictions.
Open Engagement of Societal Actors		
• Increasing opportunities for effective engagement of more parties. • Leveling the playing field for citizen scientists.		• Increasing the likelihood of technological dependence and inequity. • Spreading misinformation and increasing the risks of shaping public opinion based on false information about research.
Open Dialogue with Other Knowledge Systems		
• Facilitating knowledge exchange across different epistemic communities. • Enhancing translation of open outputs and enabling dialogue.		• Increasing biases that reflect global hegemonies and serve the interests of specific groups. • Increasing the likelihood of mistranslation and misrepresentations, especially of those languages less reflected in training data.

AVOIMEN TIETEEN INFRA



- **AI** voi tehostaa avointen aineistojen ja arkistojen käyttöä
- **Mutta:**
  - Käyttöä rajoittaa AI:n läpinäkyvyyteen liittyvät riskit
    - LLMt ym eivät itse ole aidosti avoimia
    - plagioinnin ja tekijänoikeuksien rikkominen
    - aineistojen taloudellinen hyödyntäminen

Opening up ChatGPT: tracking openness of instruction-tuned LLMs

⚡**FAccT'24 paper**⚡ Liesenfeld, Andreas, and Mark Dingemanse. 2024. 'Rethinking Open Source Generative AI: Open-Washing and the EU AI Act'. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Rio de Janeiro, Brazil: ACM. (PDF).

There is a growing amount of instruction-tuned text generators billing themselves as 'open source'. How open are they really? 🔗 FAccT'24 🔗 CUI'23 🔗 repo

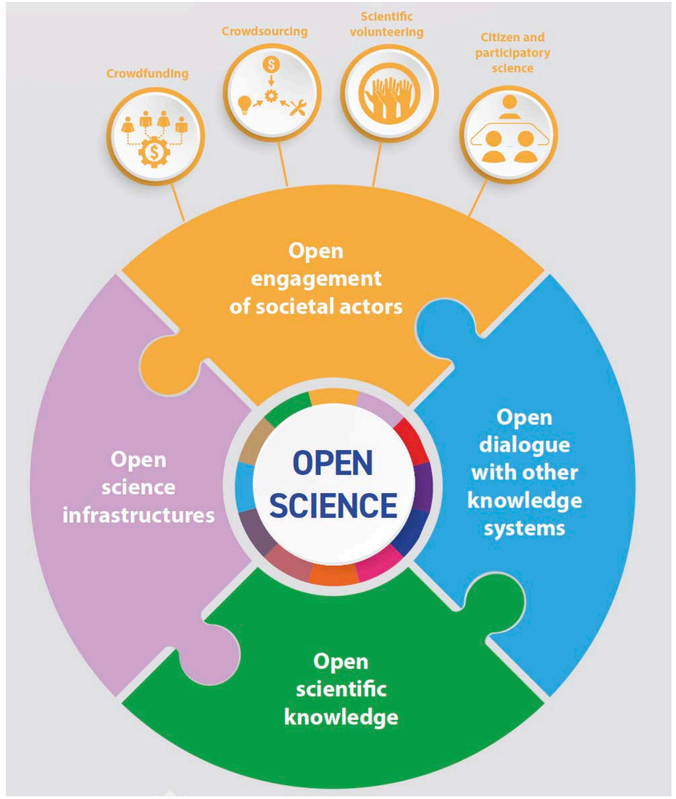
Project	Availability						Documentation						Access	
(maker, bases, URL)	Open code	LLM data	LLM weights	RL data	RL weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API
OLMo 7B Instruct	✔	✔	✔	✔	✔	✔	✔	✔	✔	✘	✔	✔	✔	~
Ai2	LLM base: OLMo 7B			RL base: OpenInstruct										12.5
BLOOMZ	✔	✔	✔	✔	~	~	✔	✔	✔	✔	✔	✔	✘	✔
bigscience-workshop	LLM base: BLOOMZ, mT0			RL base: xP3										12.0
AmberChat	✔	✔	✔	✔	✔	✔	~	~	✔	✘	~	~	✘	✔
LLM360	LLM base: Amber			RL base: ShareGPT + Evol-Instruct (sy…										10.0
Open Assistant	✔	✔	✔	✔	✘	✔	✔	✔	~	✘	✘	✘	✔	✔
LAION-AI	LLM base: Pythia 12B			RL base: Open										
OpenChat 3.5 7B	✔	✘	✔	✘										
Tshinghua University	LLM base: Mistral 7B			RL base: Share										
Pythia-Chat-Base-7…	✔	✔	✔	✔										
togethercomputer	LLM base: EleutherAI pythia			RL base: OIG										
Cerebras GPT 111…	~	✔	✔	✔										
Cerebras + Schramm	LLM base: Cerebras			RL base: Alpac										
RedPajama-INCITE…	~	✔	✔	✔										
TogetherComputer	LLM base: RedPajama-INCITE-7B-Base			RL base: variou										

Code well documented and actively maintained

Falcon-180B-chat	✘	~	~	~	~	✘	✘	~	~	✘	~	✘	✘	✘	
Technology Innovation In…	LLM base: Falcon 180B			RL base: OpenPlatypus, Ultrachat, Airo…										3.5	
Gemma 7B Instruct	~	✘	~	✘	~	✘	✘	~	~	✘	✔	✘	✘	✘	
Google DeepMind	LLM base: Gemma			RL base: Unspecified										3.5	
Orca 2	✘	✘	~	✘	✔	✘	✘	~	~	✘	~	✘	✘	~	
Microsoft Research	LLM base: LLaMA2			RL base: FLAN, Math, undisclosed (syn…										3.5	
Command R+	✘	✘	✘	✔	✔	~	✘	✘	✘	✘	~	✘	✘	✘	
Cohere AI	LLM base:			RL base: Aya Collection										3.0	
LLaMA2 Chat	✘	✘	~	✘	~	✘	✘	~	~	✘	~	✘	✘	~	
Facebook Research	LLM base: LLaMA2			RL base: Meta, StackExchange, Anthropic										3.0	
Nanbeige2-Chat	✔	✘	✘	✘	✔	~	✘	✘	✘	✘	✘	✘	✘	~	
Nanbeige LLM lab	LLM base: Unknown			RL base: Unknown										3.0	
Llama 3 Instruct	✘	✘	~	✘	~	✘	✘	~	✘	✘	~	✘	✘	~	
Facebook Research	LLM base: Meta Llama 3			RL base: Meta, undocumented										2.5	
Solar 70B	✘	✘	~	✘	~	✘	✘	✘	✘	✘	~	✘	✘	~	
Upstage AI	LLM base: LLaMA2			RL base: Orca-style, Alpaca-style										2.0	
Xwin-LM	✘	✘	~	✘	✘	✘	✘	✘	✘	✘	✘	✘	✘	~	
Xwin-LM	LLM base: LLaMA2			RL base: unknown										1.0	
ChatGPT	✘	✘	✘	✘	✘	✘	✘	✘	~	✘	✘	✘	✘	✘	
OpenAI	LLM base: GPT 3.5			RL base: Instruct-GPT										0.5	

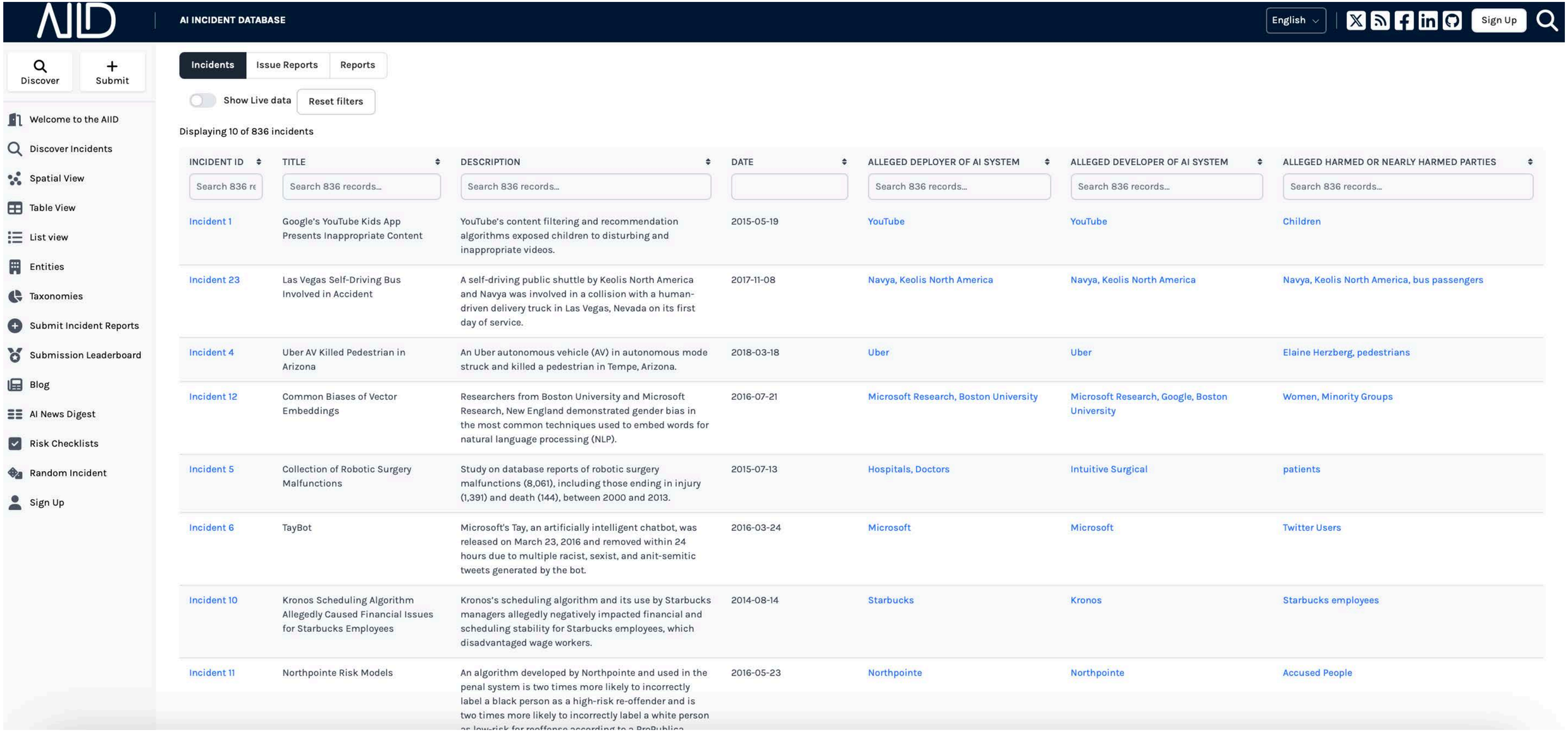
https://opening-up-chatgpt.github.io/

AVOIN TUTKIMUSKULTTUURI JA OSALLLISTUMINEN



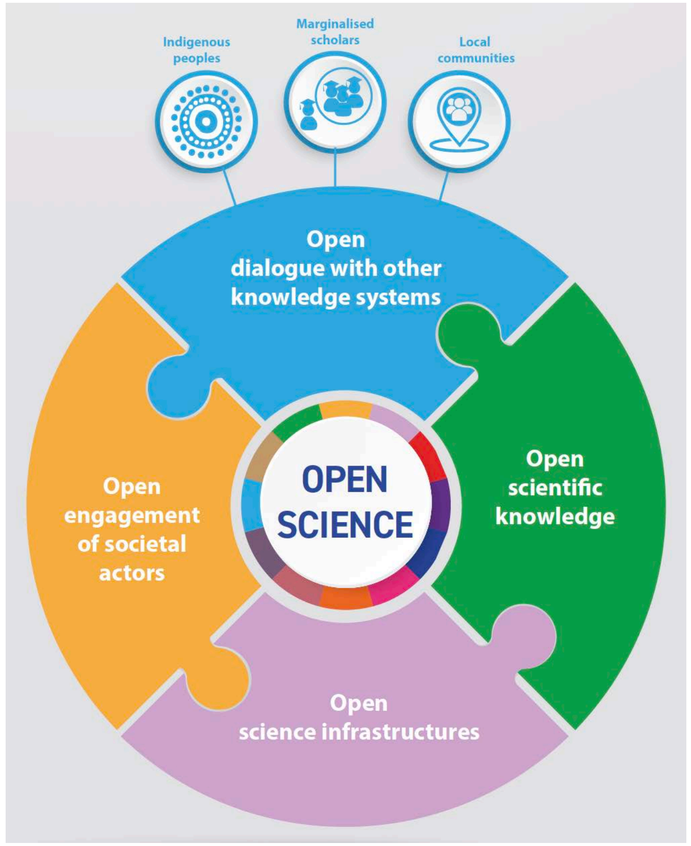
- AI voi madaltaa kynnystä osallistua ja auttaa tavoittamaan laajemman yleisön

- **Mutta:**
  - Teknologiaan liittyvien epäoikeuden-mukaisuuksien lisääntyminen
  - Virheellisen informaation levittämiseen ja manipulatiiviseen käyttöön liittyvät riskit



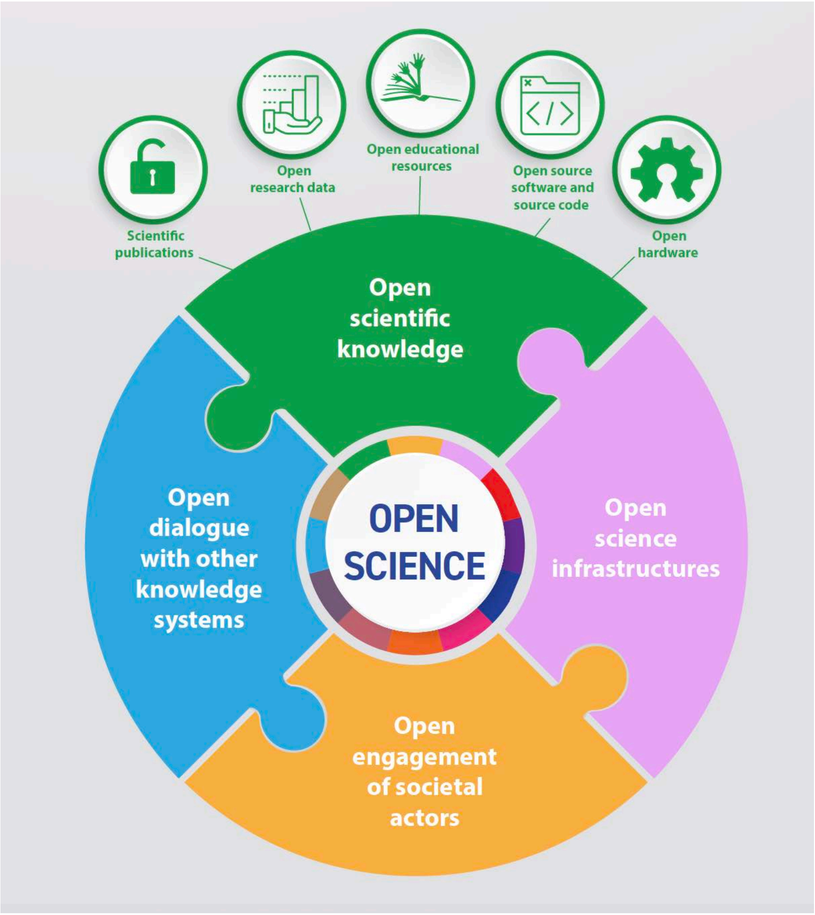
https://incidentdatabase.ai/apps/incidents/?view=incidents

AVOIN DIALOGI



- AI voi helpottaa dialogia muiden episteemisten yhteisöjen kanssa:
  - Avointen tutkimustuotosten kääntäminen
- **Mutta**:
  - **Lisääntynyt painotus valtakulttuurin avoimena saataviin näkemyksiin, oletuksiin**
  - **Käännös- ja tulkintavirheiden lisääntyminen**

TUTKIMUSAINEISTOJEN, -MENETELMIEN JA -JULKAISUJEN AVOIN SAATAVUUS



- Aito, ***mielekäs*** ja ***tasapuolinen*** saatavuus
  - tieteellisen tiedon ***lukutaito***
  - ***informaatioähkyn*** hallinta
- **AI** voi edistää näitä tarjoamalla tietoa, tuloksia ymmärrettävällä, kootulla ja kuratoidulla tavalla

- **Mutta**:
  - **AI** helpottaa artikkelitehtailua, vale/vinoutuneen informaation leviämistä
  - **AI** voi rajoittaa tutkijoita avaamasta tutkimustaan ja julkaisujaan

“Ensimmäinen kokemukseni oli viime kesäkuussa. Olin arvioimassa konferenssiin submittoitua abstraktia, joka oli melko helppo tunnistaa tekoälyn tuottamaksi. Siinä ei periaatteessa ollut mitään vikaa, mutta jos on vähänkin sisällä keskustelussa, huomaa, että kyseessä on tilkkutäkki eri tasoisista keskusteluista: johdantokirjan kaltaista tekstiä yhdistetään huippuspesialisoituneeseen keskusteluun ilman, että kokonaisuuden taustalta kuuluu selkeä ääni ja suunta.

Toinen kokemus oli rajumpi, ja uskallan väittää, että oli kyse onnekkaasta sattumasta - siitä, että juuri minä satuin olemaan ko. käsikirjoituksen vertaisarvioija - että käsikirjoitus hylättiin. Kyse oli tutkimaani aihetta käsittelevästä artikkelikäsikirjoituksesta, joka oli submittoitu alani JUFO 1-tasoiseen kv-lehteen. Pintapuolisesti luettuna jutussa ei ollut mitään erityistä vikaa, olkoonkin, että argumentti ei ollut erityisen selkeästi jäsentynyt ja yhdisteli teemoja, jotka pulpahtavat esiin googlaamalla ko. aiheesta. Mutta tällaisia kässäreitä nyt on liikkeellä paljonkin, ja usein (etenkin kokemattomammat) refereet antavat niille mahdollisuuden eli R&R.

Olen siitä 98 % varma, että kyse oli AI:n tuottamasta tekstistä. Näin siksi, että käsikirjoituksessa esitettiin yksi minun julkaisemani väite, josta tiedän varmuudella, että kukaan muu ei ole sitä esittänyt. Väitettä ei myöskään avattu, vaan se esitettiin triviaalina seikkana (mitä se ei totisesti ole). Samoin triviaaleina esitettiin kahden muun tutkijan julkaisemat spesifit väitteet, joista toisen premissit ovat suorassa ristiriidassa mainitun oman väitteeni kanssa, mutta tavalla, jota kone ei (vielä) tunnista.

Nämä aihetodisteet riittivät minulle perusteeksi esittää kässärin hylkäämistä ja kerroin päätoimittajalle luottamuksellisesti epäileväni AI:ta.”

“Oma analyysini on, että nykyinen AI pystyy tuottamaan uskottavan tuntuista tekstiä, joka hyvinkin saattaisi mennä läpi ainakin heikommissa lehdissä. Sen heikkous on toistaiseksi, että aidon argumentaation ja pohdinnan sijaan tyylilaji on lähempänä oppikirjamaista tekstiä: väitteet, joihin alansa tuntevat tutkijat suhtautuisivat kriittisesti ja kyseenalaistaen, esitetään ikään kuin faktoina.

Kauhuskenaarioni toinen vaihe on mielestäni se, että AI:lle esitetään pyyntö muokata tyyliä enemmän “alan tyylin mukaiseksi”. TAI sille syötetään esim. referee-lausunto, jossa yksityiskohtaisesti selostetaan, mikä argumentaatiossa on pielessä. Kun AI sitten toteuttaa nämä toiveet, on AI-lähtöisyyden tunnistamisen entistä vaikeampaa. Pelkään, että AI kehittyy tässä suhteessa kovaa vauhtia (ilmainen versio on selvästi huonompi kuin maksullinen, nykyinen ilmainen versio on todella paljon parempi kuin vuoden takainen, jne.). Tästä syystä onkin huolestuttavaa, että esim. Taylor & Francis myy korpuksensa miljoonadiilillä Microsoftin AI:n kehittämistarkoitukseen (kertomatta asiasta kirjoittajille). Koska kaikki OA-materiaali on materiaalia AI:lle, OA:n edistäminen syöttää suoraan lapaan.

Kauhuskenaarion kolmas vaihe (jossa uskoakseni saatamme jo olla) on se, että materiaali, johon AI perustuu, on osittain AI:n itsensä tuottamaa. Tämä toimijuutta - intentiota ja vastuuta - vailla oleva teksti sitten leviää hallitsemattomasti, viime kädessä myös ihmiskirjoittajiin.”

ONKO AVOIN TIEDE VAIVANSA VÄÄRTI?

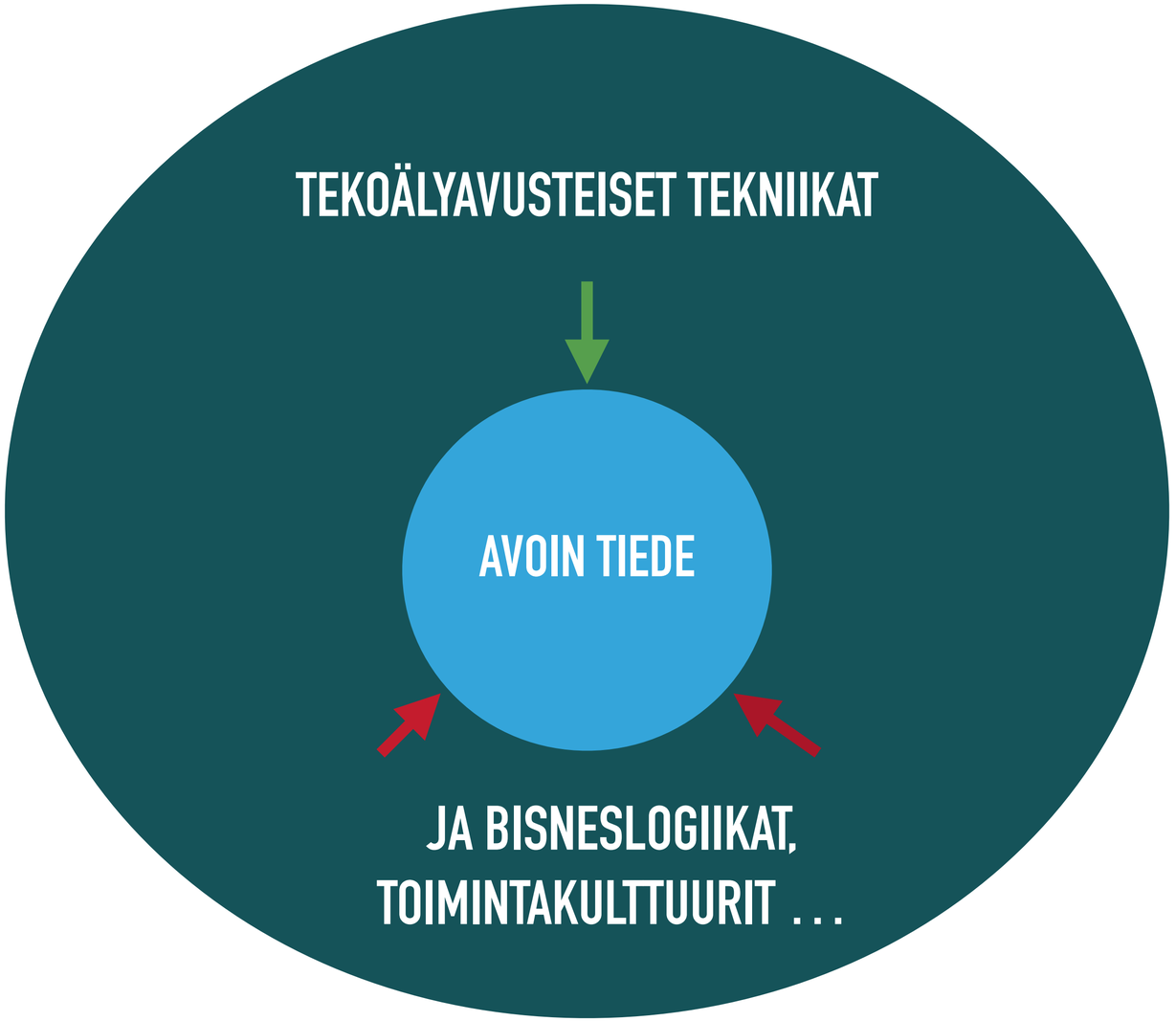
Tutkijoiden halu julkaista avoimesti vähenee

Avoimen arvioinnin edellytykset heikkenevät

Riski vinoutuneen/valetiedon leviämiseen kasvaa

- **vastoin avoimen tieteen tavoitteita:**
- **tasapuolisuus, kollektiivinen hyöty, diversiteetti, läpinäkyvyys**

TEKOÄLY HAASTAVA TOIMINTAYMPÄRISTÖ AVOIMELLE TIETEELLE

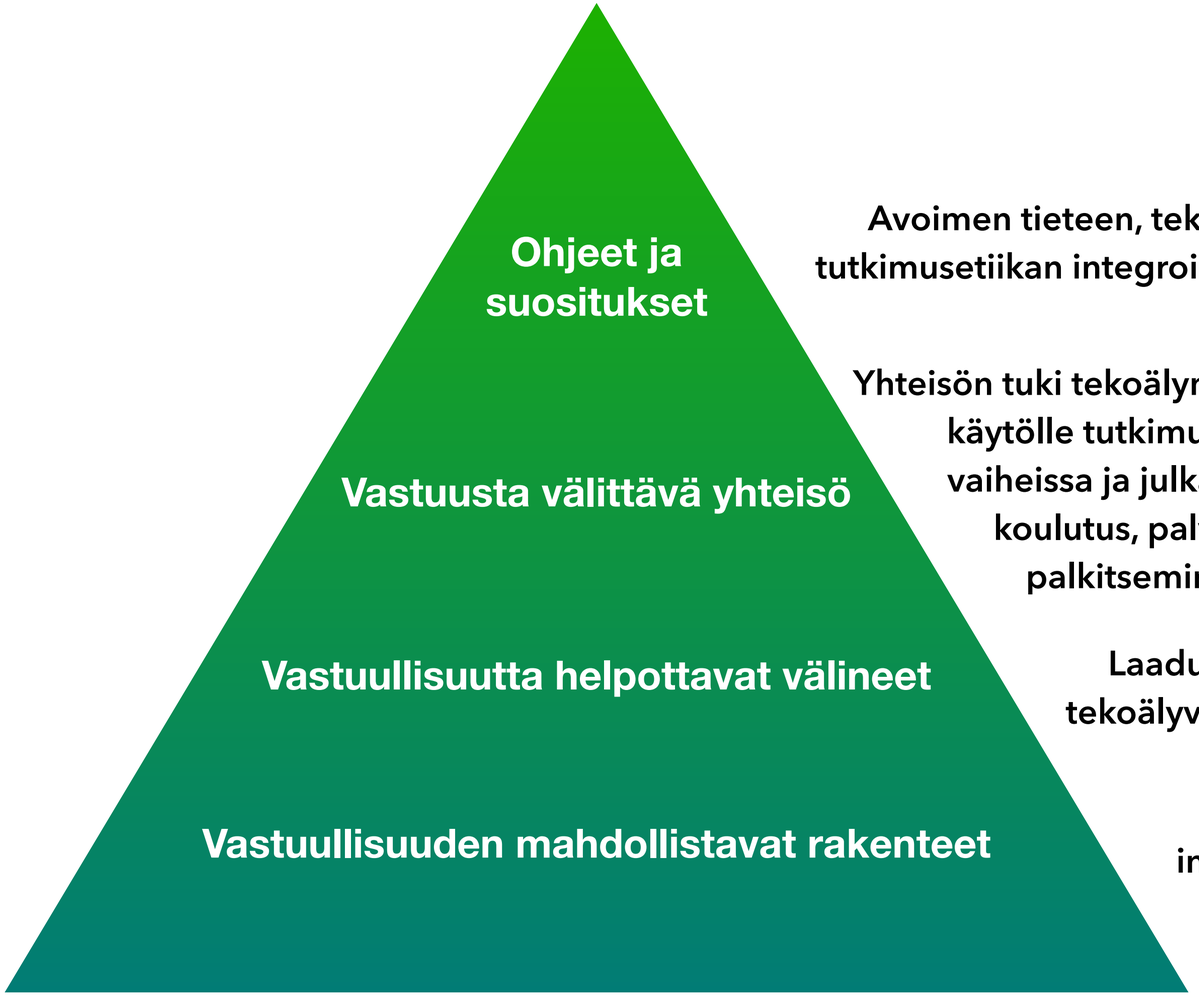


"Because generative AI has been explosive and organizations often need help to establish basic OS policies, these concerns have not yet reached the most visible parts of the OS ecosystem.

… We are concerned that if OS policymakers continue to not explicitly consider AI in their debates, AI harm will continue to propagate and policy creation to prevent it will become increasingly difficult."

Acion, Laura et al. (2023) *Nature*

HYVÄT KÄYTÄNNÖT?

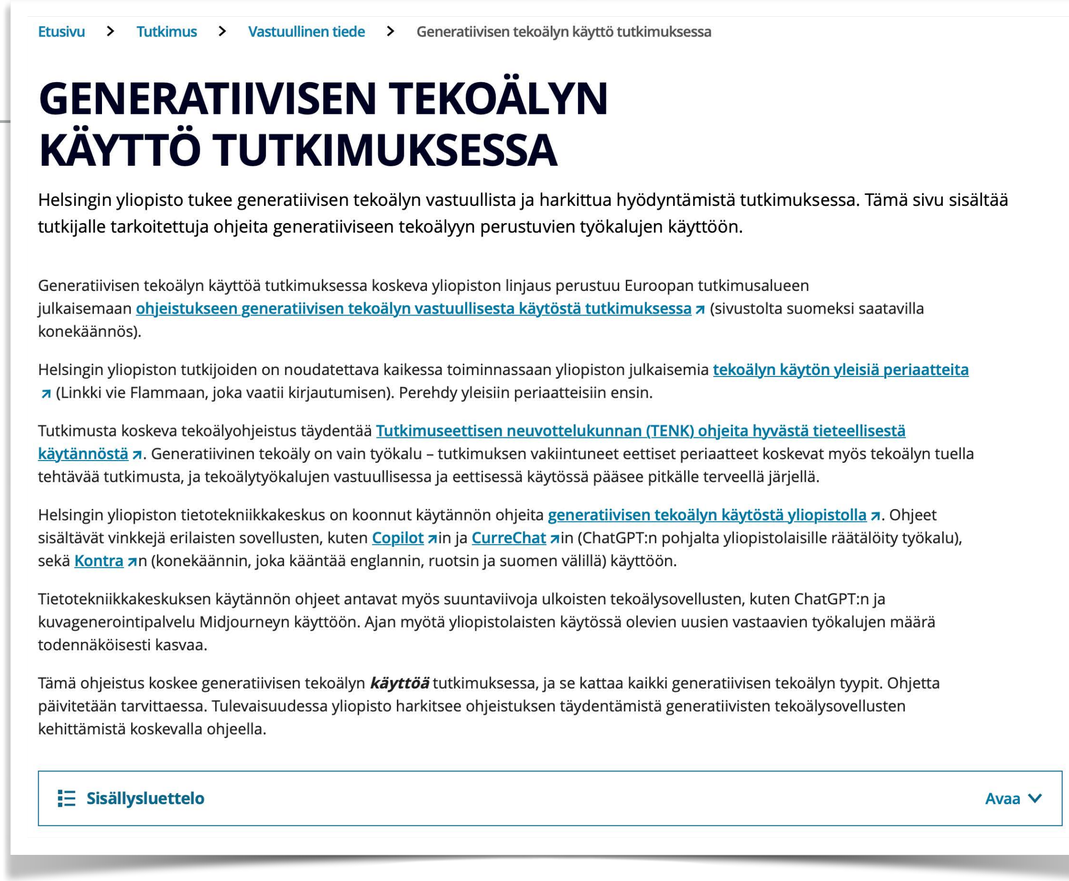


Avoimen tieteen, tekoälyn ja tutkimusetiikan integrointi ohjeissa

Yhteisön tuki tekoälyn vastuulliselle käytölle tutkimuksen eri vaiheissa ja julkaisuissa: koulutus, palvelut, palkitseminen

Laadukkaat avoimet julkaisut ja tietoturvalliset tekoälyvälineet tutkijoille (esim. Copilot, CurreChat)

Tasapuolisuutta, diversiteettiä ja tieteen integriteettiä ja tutkijoiden oikeuksia turvaavat avoimen tieteen ja tekoälyn rakenteet



Etusivu > Tutkimus > Vastuullinen tiede > Generatiivisen tekoälyn käyttö tutkimuksessa

GENERATIIVISEN TEKOÄLYN KÄYTTÖ TUTKIMUKSESSA

Helsingin yliopisto tukee generatiivisen tekoälyn vastuullista ja harkittua hyödyntämistä tutkimuksessa. Tämä sivu sisältää tutkijalle tarkoitettuja ohjeita generatiiviseen tekoälyyn perustuvien työkalujen käyttöön.

Generatiivisen tekoälyn käyttöä tutkimuksessa koskeva yliopiston linjaus perustuu Euroopan tutkimusalueen julkaisemaan ohjeistukseen generatiivisen tekoälyn vastuullisesta käytöstä tutkimuksessa (sivustolta suomeksi saatavilla konekäännös).

Helsingin yliopiston tutkijoiden on noudatettava kaikessa toiminnassaan yliopiston julkaisemia tekoälyn käytön yleisiä periaatteita (Linkki vie Flammaan, joka vaatii kirjautumisen). Perehdy yleisiin periaatteisiin ensin.

Tutkimusta koskeva tekoälyohjeistus täydentää Tutkimuseettisen neuvottelukunnan (TENK) ohjeita hyvästä tieteellisestä käytännöstä. Generatiivinen tekoäly on vain työkalu – tutkimuksen vakiintuneet eettiset periaatteet koskevat myös tekoälyn tuella tehtävää tutkimusta, ja tekoälytyökalujen vastuullisessa ja eettisessä käytössä pääsee pitkälle terveellä järjellä.

Helsingin yliopiston tietotekniikkakeskus on koonnut käytännön ohjeita generatiivisen tekoälyn käytöstä yliopistolla. Ohjeet sisältävät vinkkejä erilaisten sovellusten, kuten Copilotin ja CurreChatin (ChatGPT:n pohjalta yliopistolaisille räätälöity työkalu), sekä Kontran (konekäännin, joka kääntää englannin, ruotsin ja suomen välillä) käyttöön.

Tietotekniikkakeskuksen käytännön ohjeet antavat myös suuntaviivoja ulkoisten tekoälysovellusten, kuten ChatGPT:n ja kuvagenerointipalvelu Midjourneyn käyttöön. Ajan myötä yliopistolaisten käytössä olevien uusien vastaavien työkalujen määrä todennäköisesti kasvaa.

Tämä ohjeistus koskee generatiivisen tekoälyn ***käyttöä*** tutkimuksessa, ja se kattaa kaikki generatiivisen tekoälyn tyypit. Ohjetta päivitetään tarvittaessa. Tulevaisuudessa yliopisto harkitsee ohjeistuksen täydentämistä generatiivisten tekoälysovellusten kehittämistä koskevalla ohjeella.

Sisällysluettelo Avaa

Kiitos!

Simo Kyllönen

Yliopistonlehtori

Helsingin yliopisto
https://researchportal.helsinki.fi/en/persons/simo-kyllönen
simo.kyllonen@helsinki.fi