Google

Asia: VN/7020/2024

# Lausuntopyyntö tekoäly- ja tekijänoikeuspienryhmien loppuraporteista

## 1. Täydentäviä huomioita pienryhmien laatimiin loppuraportteihin

**Onko pienryhmä I: Tekoäly luovassa työssä loppuraportin sisältöön täydennettävää oman toimialasi osalta?**

At Google, we are excited about the promise and applications of artificial intelligence (AI). In recent years, huge breakthroughs have been seen in the use and application of AI, and this technology holds major promises for people around the world. AI has the potential to unlock significant benefits, from a better understanding of diseases to mitigating climate change and driving prosperity through greater economic opportunity. AI also has exciting and promising applications in the creative and media industries that are more commonly associated with copyright law. It has the potential to open up new opportunities for artists, journalists, and creators of all kinds. We are already seeing creators exploring new approaches to the creative writing process, to music and visual art productions, to textile design development, and more. We believe in the transformational power of AI and are excited about where it will take us next.

AI describes a wide-ranging and diverse set of technologies that are much broader than certain user-facing generative AI applications that have captured the public's imagination in recent months. When considering issues around the development of AI, it is important to grasp the sheer breadth of possible AI applications. Sectors as diverse as healthcare, media, and entertainment, retail, e-commerce, logistics, banking, finance, and IT are already using and integrating AI-based solutions in myriad different ways into their products and services.

However, innovation in AI, regardless of industry or purpose, fundamentally depends on the ability of AI models to learn - in the computational sense - from very large amounts of data. And it is the quantity of data, as a whole, that allows the models to identify features, relationships, and patterns between data points and assign corresponding weights and functions to them. These technical processes enable new knowledge and the identification of trends that can lead to breakthrough novel applications. Many legislators in the world have adopted fair use and text and/or data mining (TDM) exceptions under copyright law to support such innovation, including the ones based on AI

training. In the case of AI, these copyright laws have ensured that developers have the access to training data - ultimately the building blocks - needed for the development of AI.

All industries stand to benefit from advances in AI.  The AI copyright frameworks we adopt will affect and shape not only the creative industries - typically associated with copyright considerations - but virtually every sector of economic and scientific activity that relies on innovations in AI, including both generative and non-generative capabilities

For example in medicine, Med-Gemini, a family of Gemini models fine-tuned for multimodal medical domain applications, presents substantial potential. This family of models builds upon the Gemini large language model by fine-tuning on de-identified medical data, and achieves a 91.1% accuracy on benchmarking against the popular MedQA benchmark. These models can interpret complex 3D scans, answer clinical questions, and generate state-of-the-art radiology reports - and even calculate risk predictions. Med-Gemini demonstrates that powerful multimodal capabilities, driven by generative AI, have the potential to assist clinician, researcher and patient workflows. These are early findings but point to exciting capabilities on the horizon for healthcare applications from generative AI.

We recently introduced AI co-scientist, a multi-agent AI system built on Gemini 2.0 as a virtual scientific collaborator to help scientists generate novel hypotheses and research proposals, and to accelerate scientific and biomedical discoveries. Initial findings include multiple wet-lab validations of new insights in areas like drug repurposing for acute myeloid leukaemia, liver fibrosis and antimicrobial resistance.

In contrast to the perspectives presented above, the final report of Working Group I paints an especially concerned picture of the state of the creative industry. AI's potential is recognized only as a supportive tool, while potential negative effects are described widely. Furthermore, the report seems to illustrate somewhat problematic opposition between humans and machines, which research has shown to be present also in politics. Associate professor (social and legal implications of AI), Director (University of Helsinki Legal Tech Lab) Riikka Koulu has stated that "it seems that the early human/machine dichotomy is still reproduced in policy-making without including later critical appraisals. Hence, policy documents portray human control in opposition to unstoppable technological change, rather than as hybridization of complex sociotechnical systems, i.e. seamless collaboration between humans and artificial systems." [KOULU, Riikka, Human control over automation : EU policy and AI ethics, European journal of legal studies, 2020, Vol. 12, No. 1, pp. 9-46 - https://hdl.handle.net/1814/66992, p. 22]

Beyond societal challenges and economic growth, AI has exciting and promising applications in the creative and media industries. AI has the potential to open up new opportunities for artists, journalists, and creators of all kinds - and we are already seeing creators exploring new approaches to the creative writing process, to music and visual art productions, to textile design development,

and more. Newsrooms and journalists are also integrating AI into their work processes, with nearly three quarters of news organisations believing generative AI presents new opportunities for journalism. Our research tool, Pinpoint, helps journalists and academics analyse and explore large collections of documents, and is already being put to use in award-winning investigative reporting.

We are committed to building tools that increase access to information and create new and expanded economic and creative opportunities for artists, small businesses, and creators of all kinds. To do this, we are working closely with the creative community to put these tools in the hands of creators and to tackle new challenges as they emerge. We see AI as a complement to, and not a substitute for, human creativity; YouTube creators have embraced AI to streamline and boost their creative processes, with more than 1.7 billion views of videos related to AI tools on YouTube in 2023 alone.

Thus, we suggest adding a section presenting information on the diverse ways AI is already used in the creative industries. Related projects are currently underway in Finland, for example, the partly EU-funded https://luovain.ai/ project. The project aims to help creators find new ways to use AI tools in their work and adapt to the changes AI brings.

**Onko pienryhmä 2: Teknologia loppuraportin sisältöön täydennettävää oman toimialasi osalta?**

We appreciate the fact that there was also a separate technology group. Absent the technical knowledge it is hard to decide on the right public policy course, especially with respect to copyright. However, we would like to share the following insight in order to complement, and partly correct, the technical description presented by Working Group II.

Fundamentally, the training of models, including those underlying both generative and non-generative AI systems, captures the statistical relationships among training data, such as, in the case of an LLM, the relationships between words as they are used in writing. Like the act of reading a book and learning the facts and ideas within it, this would not only be non-infringing, it would further the very purpose of copyright law. Further, and unlike reading a book, innovation in AI fundamentally depends on the ability of LLMs to learn in the computational sense from the broadest possible variety of publicly available material. By analyzing the words and sentences in existing works, the algorithm develops a capacity to infer how new ones should be put together. This deconstructive, computational use of data in creative works in model training is fundamentally different from the communicative, aesthetic purpose for which those works were created.

Generative AI models can use what they have learned to create new content, such as text, images, music, and computer code. A "large language model" (LLM) is a generative AI model that finds patterns in human language, making it suitable for a range of writing tasks, including predicting the next words to complete a sentence or suggesting grammatical edits that preserve what you mean to say.

During training, a model evaluates the proximity, order, frequency, and other attributes of portions of words, called tokens, in its training data. In fact, the model itself selects which attributes to use.

In this way, training is the discovery of probabilities of relationships between the tokens — ultimately not in any individual text, but in all of the text on which the model is trained. The trained model then comprises a large network of weights that represent these learned relationships. The model can then respond to a prompt and generate new text with a probability of addressing the prompt as determined by its training.

Generative AI models are not databases or information retrieval systems. Critically, there are no copies of content/data in the AI model itself. The model learns from a wide variety of content during the training phase. Once trained, the model is able to produce new works/content based on 1) internal mathematical function, 2) corresponding weights, and 3) information about the data (not about the content of the data but information around patterns, constructs and syntax).

When, for instance, an LLM is prompted for facts, it can generate articulate responses that may give the impression that it is retrieving information. But, fundamentally, the model generates responses based on a statistical estimation of a satisfactory response. Put simply, it produces an average group of words, pixels, or sounds related to a prompt. Some have referred to this as not an answer but merely "answer-shaped." To understand how generative AI systems are built, it is easiest to take as an example the LLMs — like LaMDA, PaLM, and MusicLM — that underlie many of Google's latest AI advances.

The technical process of "learning" for an LLM begins with training the model to identify relationships and patterns among words in a large dataset. Through this process, a generative AI model will adjust its parameters to reflect the mathematical relationships in the data. Once the model has adjusted its parameters to reflect these relationships accurately, it can then use them to generate new outputs based on those parameters. The number of parameters needed to capture the complexities and nuances of human language and facts about the world is vast.

Consequently, AI works by learning from diverse information and data. For responsible AI development to succeed, we, therefore, need copyright systems that take a balanced, pro-innovation approach – and that continue to protect access to the works needed to train these systems while still ensuring that rightsholders can protect their creative works and the goal of copyright systems to foster creativity is honored. Existing industry standards governing web crawling are an important way to accomplish this. These standards are simple and scalable, and build on long-established machine-readable robot.txt protocols widely used across the web to control how their content is accessed by web crawlers. And now thousands of web publishers are also using the Google-Extended protocol and similar AI-specific protocols offered by other companies.

**Onko pienryhmä 3: Oikeudelliset kysymykset loppuraportin sisältöön täydennettävää oman toimialasi osalta?**

This Working Group put together, among other things, an information package on legal sources concerning AI and copyrights. However, the final report does not clearly distinguish between the

group's reasoning and the broader legal scholarly discourse. Furthermore, it remains unclear how the Working Group (or certain representatives of it) would align its suggestions with the existing Finnish and EU legislation. Thus, commenting on the suggestions, especially in relation to collective licensing, is somewhat problematic.

## 2.1 TOIMENPITEET

**1. Kannatatko esitettyjä toimenpiteitä?**

In relation to the proposed actions, Google would like to share the following comments and concerns:

a) Teksti- ja kuva-aineistojen käyttöä esimerkiksi tekoälyn kouluttamisessa ja kielimallien rakentamisessa koskeva erityinen sopimuslisenssisäännös tekijänoikeuslakiin.

b) Skaalautuvat ja reilut tavat hallita oikeuksia; huomioiden valtasuhteiden epäsuhta ja datavarantojen hyödyntämisen mahdollistaminen.

We believe that AI training is covered by fair use and other copyright exceptions and limitations, such as for TDM (in national legislation Copyright Law 13 b § "Teosten kappaleiden valmistaminen tekstin- ja tiedonlouhintaa varten"), given the transformational and profoundly different purpose of using the underlying works.

However, having a balanced copyright framework, such as TDM exceptions, does not prevent AI model providers and right holders from finding new commercial solutions for access to content: negotiating agreements and partnership deals for a variety of situations, including programmatic access to custom APIs, access to data, digitisation, etc.

At this early stage of commercial discussions, new measures to facilitate collective licensing would be premature and hinder those discussions:

Collective licensing is not a substitute for a balanced copyright regime, particularly for exceptions such as Article 4 of the Directive. Nor does a reservation to Article 4 necessarily entail that licensing is required in order to carry out acts of data-mining as other exceptions and limitations remain applicable, individually or in combination.

A collective licensing requirement would be essentially impossible, given the large amount of training data needed and the lack of comprehensive data about copyright ownership. As a result, it would effectively block the development and use of large language models and other types of cutting-edge AI. This would hinder "Europe's possibility to be at the forefront of innovation in tech

sectors that will matter in tomorrow's economy – such as Artificial Intelligence (AI)" (European Commission; A Competitiveness Compass for the EU)

If innovators are unable to leverage these building blocks needed for the development of AI, the many opportunities that come with this technology will be at risk. We will not be able to use AI to help unlock scientific discoveries and to tackle humanity's greatest challenges and opportunities - from improving cancer screening to developing solutions to tackle climate change.

Any limitation on the ability to train on publicly available material increases the risk that models will be trained on non-representative data -- including potentially excluding marginalised or alternative voices from the training data. For example, restrictions or impediments to the training of models might lead some model developers to favour older data sets (such as out-of-copyright books from more than 100 years ago that are in the public domain), which could result in model outputs being skewed based on biased or inaccurate assumptions about, e.g., race, nationality, gender roles/identity, and language. Furthermore, as there are already quite limited data sources in Finnish and Swedish, we do not advise creating any further obstacles such as national legislation that would limit the use of language and/or country specific data.

Such collective licensing requirements would be particularly detrimental to smaller AI service providers who are unlikely to have the resources to identify the relevant content and their owners; and to pay for access to data at the scale they require.

Collective licensing structures are not appropriate.

There simply aren't any copyright collectives governing the vast array of copyrightable works that are currently being used (or could potentially be) in large data sets, or that fully cover the vast array of works within a class (e.g., non-European works, etc). A collective licensing regime would also create barriers to entry for smaller, homegrown EU players, increasing transaction costs, time to market, and licensing fees. Further, in instances where there are collectives governing certain specific commercial uses of certain specific classes of works, a requirement to license a certain class would create an incentive not to develop AI models using those classes of works in favor of others that do not require a license (e.g., public domain, open source, non-copyrightable, etc.).

The TDM exception is built on the idea of freedom of choice, and on the fact that the underlying right is an exclusive one (i.e. the right grants its owner the right to authorize or prohibit copyright-relevant acts). Rightholders are free to decide whether to allow the use of their work in AI model development or not. Given the large amount of training data needed the practical implementation of these rights is based on technological solutions. We have announced Google-Extended, a standalone product token that web publishers can use to manage whether their sites help improve Gemini Apps and Vertex AI generative APIs, including future generations of models that power those

products. Google-Extended does not impact a site's inclusion or ranking in Google Search. The Google-Extended control follows the ordinary rules for the widely known Robots Exclusion Standard, through which web publishers can give instructions to web crawlers through a single plain text file. This provides web publishers with flexibility and control. Web publishers can disallow their entire website, a single or several directories, or a single or several webpage(s), or any combination thereof. We believe that making simple and scalable controls, like Google-Extended, available through robots.txt is an important step in providing transparency and control that we believe should apply with respect to content from the open web. However, as AI applications expand, web publishers will face the increasing complexity of managing different uses at scale. That's why we're committed to continued engagement with the web and AI communities to explore additional machine-readable approaches to choice and control for web publishers and we will continue to facilitate a public discussion where all voices can be heard. We believe that these new approaches need to be developed with broad input, to ensure that they are simple, scalable and broadly workable.

The services we offer have been a central part of the information society for almost 30 years. We help to find, organize, and manage information. The search engines have a long history of respecting technical rights reservations made by web page owners. AI tools and systems are creating tremendous benefits for the information society and fundamentally changing how people handle and access data.

The Finnish legal literature has noted previously the tension between constitutional right to protection of property and the right to information and communication in the copyright field (see for example Tuomas Pöysti, Tekijänoikeuksien teknisten suojakeinojen perusoikeusjännitteistä, IPRinfo 2002 and Mikael Koillinen ja Juha Lavapuro, Tekijänoikeudet tietoyhteiskunnassa perusoikeusnäkökulmasta, in Heikki Kulla et al., Viestintäoikeus 2002, s. 335-355). Again, these remarks are timely, as a balanced copyright framework does not unnecessarily hinder the emergence of advanced data processing and related services.

In addition to the above mentioned, it's important to keep in mind that AI related policy-making can have an impact on decisions related to AI infrastructure, such as datacenters. Unclear or divergent regulation from the EU level may lead data center business and investments to shift elsewhere.

As a result, policymakers should reject proposals that would require statutory licensing to train AI models. Furthermore, any proposals that are at odds with EU law and would lead to siloed national approaches to AI regulation and copyrights, should be rejected.

c) Ihmisen luovan työn tulosten erottaminen tekoälyllä tuotetusta materiaalista: läpinäkyvä sertifiointijärjestelmä.

As we invest in more capable models, we are also deeply investing in AI responsibility. That includes having the tools to identify synthetically generated content whenever you encounter it.

Provenance information, like watermarks or metadata, can reduce the risk of deception about the creator of particular material. For example, Google pioneered the industry-leading SynthID tool, and has joined the steering committee of the Coalition for Content Provenance and Authenticity (C2PA). These kinds of efforts can help consumers make informed assessments about the content they see.

Scholars and policymakers alike have recognized that AI systems do not need an incentive to create, and so there is no sound public policy reason to extend copyright protection to AI-generated works where there is no sufficient human creative intervention (as copyright exists to incentivise and reward human creativity). That said, the presence or absence of sufficient human intervention in the creative process is a nuance that will need to be addressed on a case-by-case basis. In particular, it is likely that most commercial uses of AI will entail at least some amount of human creativity. There may also be many cases where creators use these tools integratively as part of their creative process. In that circumstance, the final work product may well be protected by copyright.

d) OKM:n tulisi toimia aktiivisesti Euroopan komission tekoälytoimiston suuntaan, jotta tekoälysäädöksen läpinäkyvyysvaatimukseen ja tekijänoikeuspolitiikkaan liittyvä komission ohjaus olisi mahdollisimman selkeää ja johtaisi tekijänoikeuksien kunnioittamiseen myös tekoäly-käyttötilanteissa.

e) Eurooppa-tasoiset ja kansainväliset standardit sekä käytänne- ja soveltamisohjeet.

The AI Office is leading multiple workstreams carrying significant impact on how the AI Act will be implemented.  We believe the General-Purpose AI Code of Practice has the potential to significantly streamline AI Act compliance, reduce legal uncertainty, and lower compliance costs for both GPAI model providers and downstream developers. It represents an innovative way to balance a robust legal framework with the rapid pace of AI development. However, to fully realize this potential, The Code should specify measures already outlined in the AI Act, not introduce new obligations, and ensure technical feasibility and effectiveness. We are concerned that adding requirements beyond the AI Act creates legal uncertainty, increases compliance costs, potentially hinders Code adoption, and risks re-introducing concepts rejected during the legislative process. Google strives to ensure the Code effectively supports the AI Act's objectives by providing clarity and predictability, without inadvertently creating new hurdles or contradicting established legal frameworks.

The public summary on training content (Art. 53(1)(d) AI Act) should strike an appropriate balance of interests. Disclosures should be technically feasible, and not jeopardize confidential business information, trade secrets, and model security.

Finland should not develop work in parallel to these EU developments, as such work may overlap with EU rules and end up leading to legal uncertainty. This approach would be in line with the Finnish Cultural Policy Report (Kultturipoliittinen selonteko), which states that copyright issues related to artificial intelligence and the protection of creators of creative work must be resolved at the European Union level and through international agreements.

**2. Onko jokin toimenpide mielestäsi toista tärkeämpi?**

The most important measures are ones that are based on the AI Act, which itself forms extensive rules on AI development. Further legislative measures should not be initiated before the AI Act and the various instruments applying it have been fully implemented, and their impact has been properly assessed.

Prematurely altering copyright laws without understanding the full effects of these measures may inadvertently stifle innovation and hinder the potential benefits AI offers across various industries. We would therefore recommend against a modification of the existing copyright legal framework at this stage.

**3. Millä tavalla katsot, että toimenpiteet tulisi toteuttaa?**

As stated in the Cultural Policy Report, copyright issues related to artificial intelligence and the protection of creators of creative work must be resolved at the European Union level and through international agreements.

The European Commission has recently underlined the aim to "make the most of the opportunities offered by artificial intelligence" (Commission work programme 2025, p. 5). We believe that this forms a clear guiding principle in all areas of AI policymaking.

## 2.2 KOULUTUS- JA TIEDOTUS

**1. Kannatatko esitettyjä toimenpiteitä koulutukseen ja tiedotukseen liittyen?**

While the proposed measures in section 2.2 are mostly worth supporting, we would like to comment on the point concerning the "copyright risks" of the AI tools. Firstly, it is important to keep in mind that the possibility that AI models can occasionally, despite the best efforts of their developers, output content that replicates existing expression is a bug, not a feature, and developers are taking a range of measures and undertaking research to limit that occurrence even further. Tools like output filters can also help restrict substantially similar outputs even as models themselves learn to make more nuanced assessments of these factors.

**2. Onko jokin ehdotettu teema mielestäsi toista tärkeämpi?**

Generally, we find it important that all stakeholders continue to share insights and information as well as maintain a positive attitude and curiosity about new ways of working. Having said that, we would like to highlight that the need to learn new working methods is not limited to the creative industry but covers all parts of society worldwide.

For example, the Ministry of Finance (Finland) published recently (26.2.2025) "Guidelines for the Use of Generative AI as a Support Tool in Public Administration". Similarly, the Finnish National Agency for Education and the Ministry of Education and Culture produce AI guidelines for teaching this year. Both of these examples handle, among many other important things, also copyright questions. However, we would not weight education in AI and copyright themes more important than other fields of responsible use of AI.

Many stakeholders in the private sector are also putting lots of effort in educational projects.  We offer plenty of online courses and educational materials free of charge in relation to AI (please see https://grow.google/intl/fi/courses-and-tools/)

**3. Millä tavalla katsot, että koulutus ja tiedotus tulisi toteuttaa?**

-

## 2.3 SELVITYKSET

**1. Kannatatko esitettyjä aiheita selvitettäväksi?**

In relation to the proposed studies, Google would like to share the following comments:

a) Laadukkaiden datasettien kriteerit ja lisensiointi tekoälyratkaisujen kehittämiseen (määrittely, parhaiden käytäntöjen kuvaaminen ja jakaminen, tunnisteiden ja metadatastandardien hyödyntäminen ja laajan käyttöönoton edistäminen, miten seurata ja valvoa tekoälyn kehittämiselle asetettuja standardeja ja lisensiointiehtoja?).

As stated in section 2.1 question 1 we believe that AI training should be covered by fair use or other copyright exceptions and limitations. On top of that, direct data licensing deals are already being actively explored and agreed upon by parties. Thus, there is no need to monitor licensing conditions.

b) Mitä tekijänoikeudellisesti relevanttia tapahtuu tekoälymalleja kehitettäessä ja hyödynnettäessä? Pienryhmätyössä tunnistettuihin kontaktipisteisiin ja avoimiin kysymyksiin liittyvä selvitystyö.

Please see our answer to section 1, WG 2.

c) Mahdollistettava uudenlaisten ammattitaitojen kehittäminen ja selvitettävä luovien alojen uudet ansaintamallit

Please see our answer to section 1, WG 1 regarding the positive possibilities AI tools offer for the creative industry.

The need for additional financing should be looked into through a comprehensive impact assessment. This needs to assess whether there is a financing need, where such a financing need exits, what are the options to address it, and what is the expected impact of all options.

**2. Onko jokin selvityksen aihe mielestäsi toista tärkeämpi?**

What is important in this realm is the development of professional skills - we already see an uptake of AI tools by creators, but we believe more creators would benefit from these tools if they were given a chance to learn how to use them in their work.

**3. Millä tavalla katsot, että selvitykset tulisi toteuttaa?**

-

## 3. Lopuksi

**Muita huomioita pienryhmien laatimiin loppuraportteihin ja niissä esitettyihin suosituksiin?**

In summary, we believe that:

1. AI has the potential to unlock significant benefits, from a better understanding of diseases to mitigating climate change and driving prosperity through greater economic opportunity.

2. AI also has exciting and promising applications in the creative and media industries that are more commonly associated with copyright law.

3. Collective licensing requirements  would make AI development more expensive.

4. There are already quite limited data sources in Finnish and Swedish, no further obstacles such as national legislation that would limit the use of language and/or country specific data. Otherwise AI models may not meet the needs of people and businesses in Finland in future.

5. Unclear or divergent regulation from the EU level may also lead data center business and investments to shift elsewhere.

6. The opt-out model is a practical and balanced solution, it has a solid base in EU legislation and is implemented in Finnish Copyright Law.

While we acknowledge the ongoing discussions around AI and copyright, we believe it's essential to first assess the impact of existing regulations, particularly the AI Act. This Regulation complements

the existing EU copyright framework in order to ensure the clear articulation between them both. Prematurely altering copyright laws without understanding the full effects of these measures may inadvertently stifle innovation and hinder the potential benefits AI offers across various industries. We would therefore recommend against a modification of the existing copyright legal framework at this stage.

Furthermore, the European Commission has taken a deregulatory approach on AI liability regulation to leave room for European AI development. We recommend the same cautious and well balanced approach on AI copyright policy. A predictable and stable regulatory environment both at EU and national level forms the basis for AI development in Europe.

Jern Heidi
Google