Third Draft of the General-Purpose AI Code of Practice
COMMITMENTS BY PROVIDERS OF GENERAL-PURPOSE AI MODELS
<u>COPYRIGHT SECTION</u>

<u>Introductory note by the Chair and Vice-Chair of the Copyright Section</u>

The Copyright section of the Code of Practice describes a set of Measures that Signatories commit to taking in order to comply with their obligation under Article 53(1)c) AI Act. The third draft retains core elements of the first two drafts in a simplified and clearer form. Measure I.2.1 (former Measures 2.1 and 2.2, Draft 2) specifies what it means to "put in place" a policy to comply with Union copyright law and clarifies the set of Measures covered. While Measures I.2.2 and I.2.3 (former Measures 2.4-2.8, Draft 2) regulate the mining of web-crawled content, Measure I.2.4 (former Measure 2.3) concerns the mining of protected content not web-crawled by the Signatory. Measure I.2.5 (former Measures 2.9 and 2.10, Draft 2) sets out commitments to mitigate the risk that a downstream AI system repeatedly generates infringing output. Finally, Measure I.2.6 (former Measure 2.11, Draft 2) provides for commitments to designate a point of contact and to allow for the submission of complaints concerning the non-compliance of Signatories with their commitments under the copyright Section. As indicated in the General Introduction, this Section no longer contains separate KPIs. However, the features of some KPIs have been incorporated into the Measures (e.g. description of the policy in a single document, easily accessible information on the contact point and the possibility to lodge complaints). The recitals clarify that this Section is without prejudice to the application and enforcement of Union law on copyright and related rights and to commercial agreements between the Signatories and rightsholders authorising the use of protected content.

Alexander Peukert
Working Group 1 Co-Chair

Céline Castets-Renard
Working Group 1 Vice-Chair

RECITALS FOR COPYRIGHT SECTION

Whereas:

a) This Section aims to contribute to the proper application of the obligation of providers of general-purpose AI models placed on the Union market to put in place a policy to comply with Union law on copyright and related rights pursuant to Article 53(1), point (c) AI Act.

b) The compliance with the commitments under this Section should be commensurate and proportionate to the size and capacities of providers, taking due account of the interests of SMEs, including startups.

c) This Section is without prejudice to and in no way affects the application and enforcement of Union law on copyright and related rights.

d) This Section is without prejudice to commercial agreements between the Signatories and rightsholders authorising the use of works and other protected subject matter.

Commitment I.2. Copyright policy

LEGAL TEXT: Article 53(1)(c) AI Act

In order to fulfil the obligation to put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790 pursuant to Article 53(1), point (c) AI Act, Signatories commit to drawing up, keeping up-to-date, and implementing a copyright policy in accordance with Measure I.2.1, as well as adopting Measures I.2.2—I.2.6 for their general-purpose AI models placed on the EU market.

Measure I.2.1. Draw up, keep up-to-date and implement a copyright policy

(1) Signatories will draw up, keep up-to-date and implement a policy to comply with Union law on copyright and related rights. This policy will address all commitments pursuant to this Section and will be described in a single document approved by the Signatory. Signatories will assign responsibilities within their organisation for the implementation and overseeing of this policy.

(2) Signatories are encouraged to make publicly available and keep up-to-date a summary of their copyright policy.

Measure I.2.2. Reproduce and extract only lawfully accessible copyright-protected content when crawling the World Wide Web

In order to ensure that Signatories will only reproduce and extract lawfully accessible works and other protected subject matter if they use web-crawlers or have such web-crawlers be used on their behalf to crawl, scrape and/or otherwise compile data for the purpose of text and data mining according to Article 2(2) of Directive (EU) 2019/790 and the training of their general-purpose AI models, Signatories will:

(a) not circumvent effective technological measures as defined in Article 6(3) of Directive (EC) 2001/29 designed to prevent or restrict access to works and other protected subject matter, such as paywalls, and

(b) make reasonable efforts to exclude from their web-crawling Internet domains that make available to the public copyright-infringing content on a commercial scale and have no substantial legitimate uses ("piracy domains") and that are recognised as such by courts or public authorities in the European Union and the European Economic Area. A list of hyperlinks to relevant piracy domain lists issued by the relevant bodies in the European Union and the European Economic Area is to be made publicly available on an EU website.

Measure I.2.3. Identify and comply with rights reservations when crawling the World Wide Web

(1) In order to ensure that Signatories will identify and comply with, including through state-of-the-art technologies, machine-readable reservations of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790 if they use web-crawlers or have such web-crawlers used on their behalf to crawl, scrape and/or otherwise compile data for the purpose of text and data mining according to Article 2(2) of Directive (EU) 2019/790 and the training of their general-purpose AI models, Signatories will

    (a) employ web-crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt), as specified in the Internet Engineering Task Force (IETF) Request for Comments No. 9309, and any subsequent version of this IETF standard, and

    (b) make best efforts to identify and comply with other appropriate machine-readable protocols to express rights reservations pursuant to Article 4(3) of Directive (EU) 2019/790, for example through asset-based or location-based metadata, that have either resulted from a cross-industry standard-setting process as referred to in paragraph 3 of this Measure or are state-of-the-art and widely adopted by rightsholders, considering different cultural sectors, and generally agreed through an inclusive process based on bona fide discussions to be facilitated at EU level with the involvement of rightsholders, AI providers and other relevant stakeholders as a more immediate solution, while anticipating the development of cross-industry standards referred in paragraph 3.

(2) This commitment is without prejudice to the right of rightsholders to expressly reserve the use of lawfully accessible works and other protected subject matter for the purposes of text and data mining pursuant to Article 4(3) of Directive (EU) 2019/790 in any appropriate manner, such as machine-readable means in the case of content made publicly available online.

(3) In due consideration of relevant international and European standard-setting processes, Signatories are encouraged to support relevant standardisation efforts and engage on a voluntary basis in bona fide discussions with other relevant stakeholders, including rightsholders, with the aim to develop appropriate machine-readable standards to express a rights reservation pursuant to Article 4(3) of Directive (EU) 2019/790.

(4) Signatories will take reasonable measures to enable affected rightsholders to obtain information about the web crawlers employed and their robot.txt features and the measures that a Signatory adopts to identify and comply with rights reservations expressed pursuant to Article 4(3) of Directive (EU) 2019/790 at the time of crawling, for example by making public such information and syndicating a web feed that covers every update of the website informing about the rights reservation compliance.

(5) Signatories that also provide an online search engine as defined in Article 3(j) Regulation (EU) 2022/2065 or control such a provider are encouraged to take appropriate measures to ensure that

3

its compliance with a rights reservation expressed in accordance with paragraph 1 of this Measure does not negatively affect the findability of the content, for which a rights reservation has been expressed, through their search engine.

Measure I.2.4. Obtain adequate information about protected content not web-crawled by the Signatory

(1) If Signatories, for the training of their general-purpose AI models, mine works and other protected subject matter according to Article 2(2) of Directive (EU) 2019/790 that they have obtained by other means than by web-crawling or by web-crawling on their behalf and without an authorisation by the respective rightsholders or their authorised representatives, they will make reasonable efforts to obtain adequate information (e.g., by checking the information available on the website of the third parties or requesting information) as to whether works and other protected subject matter that have been scraped or crawled from the internet were collected by employing web-crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt), as specified in the Internet Engineering Task Force (IETF) Request for Comments No. 9309, and any subsequent version of this IETF standard.

(2) This Measure does not imply a commitment to verify or proceed to a work-by-work assessment of the data mentioned in paragraph 1 of this Measure in terms of copyright compliance.

Measure I.2.5. Mitigate the risk of production of copyright-infringing output

(1) In order to mitigate the risk that a downstream AI system, into which a generative general-purpose AI model is integrated, repeatedly generates output that infringes copyrights or related rights as protected according to Union law on copyright and related rights, Signatories will
   (a) make reasonable efforts to mitigate the risk that a model memorizes copyrighted training content to the extent that it repeatedly produces copyright-infringing outputs and
   (b) prohibit copyright-infringing uses of a model in their acceptable use policy, terms and conditions, or other equivalent documents.

(2) This Measure applies irrespective of whether a Signatory vertically integrates the model into its own AI system(s) or whether the model is provided to another entity based on contractual relations. The commitment to prohibit copyright-infringing uses pursuant to paragraph 1, point b of this Measure does not apply to general-purpose AI models that are released under a free and open-source licence.

Measure I.2.6. Designate a point of contact and enable the lodging of complaints

(1) Signatories will designate a point of contact for communication with affected rightsholders and provide easily accessible information about it.

(2) Signatories will put a mechanism in place to allow affected rightsholders and their authorised representatives, including collective management organisations, to submit, by electronic means, sufficiently precise and adequately substantiated complaints concerning the non-compliance of Signatories with their commitments pursuant to this Section and provide easily accessible information about it. Where complaints by rightsholders are manifestly unfounded or excessive, in particular because of their repetitive character, Signatories may refuse to act on the complaint.