Principles-Code for Protection of intellectual property and transparency for the appropriate use of generative AI (provisional title) (draft) Summary of disclosure items Specific examples

*These specific examples are provided by the Secretariat as examples of disclosures made by businesses under the Code for reference purposes, so that businesses can refer to them when submitting their opinions in the public comment period.

*As an example of what would happen if you selected "Explain" for a matter to be disclosed, the secretariat has provided "Examples of when to explain" for some matters for reference.

classification	item	Specific examples
Usage model related	Name (identifier, version, etc.)	○○ Ver2.1
	History including publication date (past versions, revision history, etc.)	Released on ○/○/○, fixed △△ function on △/△/△.
	Architecture and design specifications (status of licenses contracted with third parties for model development, hardware, software and licenses required for use, etc.)	**[Architecture/design specifications]** Transformer architecture **[Status of license contract with third party for model development]** Contract signed with XX company <Example of when to explain> In developing the model, we have concluded a license agreement for the model used with a third party, but due to confidentiality obligations under the agreement, and expressing that it is at the core of our businesses, we cannot disclose the name of the third party. **[Hardware and software required for use]** GPU with a capacity of ○○ GB or more, △△ Ver. 3.1 or higher

		[License policy for users] After publishing and sharing the model , we ask that you agree to a free open source license that allows the provider to freely access, use, modify, and redistribute the model or its modified versions under certain conditions. *License policy (URL of reference page)
	Terms of use (clarification of intended uses, restricted and prohibited uses, etc.)	Prohibited uses: Violent, sexual, or CBRN-related uses, or any use that violates any law or regulation * URL of the reference page
	Details of the model training process (training method, parameter settings including inference process and decision basis, etc.)	The pre-training stage is optimized based on gradient descent. The model employs an autoregressive language model and is trained to predict the next word using a large corpus of trillions of tokens. Post-training is then performed using techniques such as reinforcement learning (RLHF) using a dataset that reflects human preferences, adjusting the response to match human values and intentions.

classification	item	Specific examples
Learning data	Matters related to the data used for training and validation (type of data, matters related to private datasets obtained by web crawling or third parties, matters related to public datasets, matters related to data collected by other means, whether synthetic data is used and for what purpose, etc.)	**[Data type]** Text, images, audio, video **[Web crawl]** Conducted **[Matters related to private datasets]** Obtained through license agreement **[Matters related to public datasets]** Uses ○○ structured dataset **[Matters related to data collected by other means]** Obtained from data provider △△ **[Whether synthetic data is used/purpose]** Uses synthetic data/enhanced security
	Crawler (purpose, data collection period, name/identifier, whether or not a third-party crawler is used and its name/identifier, etc.)	**[Purpose]** Collect data to improve the model **[Name/Identifier: Data collection period]** ○○bot: Collected continuously from ○/○/○ △△bot: Collected continuously from △/△/△ **[Third-party crawler name/Identifier: Data collection period]** ●●bot: Collected continuously from ●/●/● ▲▲bot: Collected continuously from ▲/▲/▲

classification	item	Specific examples
Accountability Relationships	Ensuring that decisions made during the development, provision, and use of generative AI systems or services can be tracked and traced to the extent technically possible and reasonable (improving traceability, clarifying who is responsible, allocating responsibilities among parties involved, responding specifically to stakeholders, and documenting such information)	**[Traceability]** Model evaluation logs will be continuously recorded and retained for the period specified by the rules of ○○ Company. **[Clearance of responsible person]** The CAIO of ○○ company will be the person in charge and will be responsible for the work of the ○○ team at the company. **[Distribution of responsibility among parties involved]** In accordance with the above contract, we will clarify the allocation of responsibilities between the rights holder and users, etc., and confirm the agreement with the other party. **[Specific responses to stakeholders]** According to ○○ company's ○○ regulations **[Documentation]** Documented according to ○○ company's ○○ regulations

classification	item	Specific examples
Measures to protect intellectual property rights	In order to ensure proper compliance with rights, principles for protecting intellectual property rights will be established and the system of responsibility will be clarified. These will be reviewed at least once a year, and a summary will be made public.	·○○ principles for intellectual property protection have been established. *Outline of ○○ principles (URL of reference page) ·Establish a ○○ department to comply with the ○○ principles, review the results once a year, and publish the results on the company's website (URL of the reference page)
	When utilizing data, including for the development and training of generative AI, these businesses shall not infringe on the intellectual property rights of others.	·Establish a ○○ process to ensure that development and learning do not infringe on the intellectual property of other companies *Outline of the ○○ process (URL of reference page) ·Regarding use, include clauses regarding the prevention of infringement of rights in the terms of use * URL of reference page

classification	item	Specific examples
Measures to protect intellectual property rights	to respect access restrictions such as paywalls and employ crawlers that follow machine-readable instructions such as robots.txt . To ensure that rights holders take appropriate measures, the above measures must be published with user agents and any changes must be notified.	Comply with paywalls and use crawlers that comply with robots.txt. Publish the policy, including the above, on your company website (URL of the reference page).
	The learned logs are kept for a certain period of time.	Keep logs for the duration of use
	Work to avoid crawling so-called pirated sites.	・In order to exclude pirated sites from the collection of learning data, when information about pirated sites is provided by public institutions, etc., we will review the content and take measures such as excluding them. ・Exchange information with rights holder organizations and make efforts to identify pirated sites

classification	item	Specific examples
Measures to protect intellectual property rights	To the extent possible, technological measures shall be taken to prevent the creation of products that infringe intellectual property rights.	Equipped with a filtering function to prevent output of products that infringe intellectual property rights
	Wherever possible, use digital watermarks, C2PA and other technical measures to verify the origin and provenance of content.	C2PA compliant , digital watermark technology implemented
	Inform users that if they believe that a product infringes the intellectual property rights of others, they should not use it.	Include clauses in the terms of use to prevent infringement of intellectual property rights
	In order to ensure timely and appropriate relief for rights holders, appropriate contact points will be established, including by utilizing existing systems, application requirements will be made as clear as possible, and records of responses will be kept.	Establishment of copyright contact point (URL of reference page), keeping correspondence records