TOKYO COLLEGE
WORKING PAPER SERIES

Preliminary Insights and Recommendations for HAIP Reporting Framework

– Based on HAIP Participant Organizations Interviews –

Arisa Ema, The University of Tokyo
Fumiko Kudo, The University of Osaka
Toshiya Jitsuzumi, Chuo University

July 11, 2025

TC-E-25-1

Tokyo College, The University of Tokyo

Preliminary Insights and Recommendations for HAIP Reporting Framework
– Based on HAIP Participant Organizations Interviews –

Arisa Ema, Dr., The University of Tokyo,
Fumiko Kudo, J.D., The University of Osaka,
Toshiya Jitsuzumi, Dr., Chuo University

1. Introduction

As the social implementation of artificial intelligence (AI) accelerates, the need for collaborative efforts among industry, government, academia, and civil society to build trustworthy AI governance is becoming increasingly urgent. In response to this context, the "Hiroshima AI Process" (hereinafter referred to by its acronym "HAIP") was launched at the 2023 G7 Hiroshima Summit[1]. Under HAIP, a set of International Guiding Principles and an International Code of Conduct aimed at promoting the safe and trustworthy development of advanced AI systems were formulated and endorsed by the G7 leaders[2].

At the 2024 G7 under the Italian presidency, HAIP was further advanced with the development—supported by the Organisation for Economic Co-operation and Development (OECD)—of a reporting framework that enables AI developers and other relevant actors to self-assess and report on their adherence to the International Code of Conduct[3].

This paper presents findings based on interviews conducted with 11 of the 19 companies that had joined the reporting framework as of the end of April, as well as insights gained from a stakeholder consultation meeting held at the University of Tokyo in June 2025. Based on these findings, the paper identifies key challenges surrounding the current framework and proposes directions for its future improvement. It should be noted that this paper is a preliminary and simplified version, and a more detailed set of recommendations is scheduled for release in the summer of 2025.

These preliminary findings were already shared at the OECD Workshop on the HAIP Reporting Framework (11 June 2025), and the presentation slides are publicly available[4].

2. About the HAIP Reporting Framework

The HAIP reporting framework is implemented through a questionnaire developed and administered by the Organisation for Economic Co-operation and Development (OECD), to which participating organizations voluntarily respond[5]. Participation is not mandatory; rather, the framework is intended to encourage AI developers and other relevant actors to align with the International Code of Conduct on a voluntary basis. In this sense, HAIP functions as a framework for voluntary AI governance.

* Correspondence to: aema@g.ecc.u-tokyo.ac.jp kudo.fumiko.elsi@osaka-u.ac.jp / This paper is dated on June 23, 2025.

[1] Cabinet Office, Government of Japan “The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI” (2024) https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html

[2] Ministry of Foreign Affairs of Japan “G7 Leaders' Statement on the Hiroshima AI Process” (2023) https://www.mofa.go.jp/ecm/ec/page5e_000076.html

[3] Ministry of Internal Affairs and Communications of Japan “Launch of the ‘Reporting Framework’ for the International Code of Conduct (‘Hiroshima AI Process’)” (2025) https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2025/2/7_3.html

[4] https://drive.google.com/file/d/1Re8fApWZTVzU1xMXBBuS3V-C3pASU7Kz/view

[5] OECD “G7 reporting framework – Hiroshima AI Process (HAIP) international code of conduct for organizations developing advanced AI systems” (2025) https://transparency.oecd.ai/

Based on the HAIP International Code of Conduct's 11 actions, the reporting framework includes structured questions to help organizations align their practices with the Code. It was developed through collaboration among diverse stakeholder communities—including government bodies, private companies, academia, civil society organizations, and research institutions. The organizations that contributed to shaping this framework include Anthropic, AWS, Databricks, DFKI, Google, Google DeepMind, iGenius, LNE, Leonardo, Meta, Microsoft, Mistral AI, NEC, NTT, OpenAI, OpenText, and SaferAI[6].

The questionnaire is broadly divided into the following seven thematic areas. Each area contains multiple questions, with a total of 39 questions across the entire form[7].

1. Risk identification and evaluation
2. Risk management and information security
3. Transparency reporting on advanced AI systems
4. Organizational governance, incident management and transparency
5. Content authentication & provenance mechanisms
6. Research & investment to advance AI safety & mitigate social risks
7. Advancing human and global interests

Participating organizations report their initiatives based on the questionnaire provided by the OECD, and the submitted information is made publicly available on the OECD's website. This initiative aims to enhance transparency in AI development and foster international trust. At the same time, it aspires to serve as a global standard for information disclosure in the field of AI.

As of the end of April 2025, when the first set of responses was published, reports from 19 participating organizations had been made available[8]. They came from Japan, the United States, Germany, Canada, South Korea, Romania, and Israel. The participants included not only major technology companies but also startups and research institutions. At the time of writing this paper, one more organization, for a total of 20 organizations, has been released.

Additionally, the OECD has explicitly stated that it does not assess the content of the responses[9]. However, a blog post published on the OECD website summarizes share preliminary insights from the first wave of reports across seven thematic areas and also presents ideas for enhancing the value of the framework for future reporting[10].

3. Research Methodology: Interviews and Stakeholder Consultation

3.1 Interviews with 11 Companies

Among the 19 organizations that had participated in the HAIP reporting framework as of the end of April 2025, the following 11 companies agreed to take part in online interviews conducted by the authors. (The list below follows the order in which each organization submitted its response to the OECD.)

[6] OECD "About the reporting framework" (2025) https://transparency.oecd.ai/about

[7] OECD "Report sample questionnaire" (2025) https://transparency.oecd.ai/questionnaire-sample

[8] OECD "Submitted reports" (2025) https://transparency.oecd.ai/reports

[9] OECD "FAQ - How will submissions be processed?" (2025) https://transparency.oecd.ai/faq#question-4

[10] Karine Perset, James Gealy, Sara Fialho Esposito "Shaping trustworthy AI: Early insights from the Hiroshima AI Process Reporting Framework" (2025) https://oecd.ai/en/wonk/haip-reporting-insights

1. KDDI Corporation (JP)
2. SoftBank Corp. (JP)
3. Preferred Networks (JP)
4. NEC Corporation (JP)
5. Nippon Telegraph and Telephone Corporation (JP)
6. Microsoft (US)
7. Salesforce (US)
8. OpenAI (US)
9. Google (US)
10. Fujitsu (JP)
11. Rakuten Group, Inc. (JP)

For the following eight organizations, interview requests were submitted via their contact forms, but interviews could not be conducted, as of the time of writing on June 23, 2025, due to reasons such as no response. (The list follows the order in which each organization submitted its response to the OECD.)

1. West Lake research & education service (US)
2. Data Privacy and AI (DE)
3. KYP.ai GmbH (DE)
4. Anthropic (US)
5. TELUS (CA)
6. Fayston Preparatory School (KR)
7. ai21 (IL)
8. MGOIT (RO)

The interviews were conducted online using videoconferencing tools between April and May 2025. The interviews were conducted by the authors (Ema, Kudo, and Jitsuzumi). In general, each interview involved one or two representatives from each company, who were, in principle, the individuals responsible for drafting the responses of the HAIP reporting framework. However, in a few cases, the interviewee was solely the person in charge at the organization's Japanese branch.

Examples of the interview questions are provided below. In some cases, when interviewees were affiliated with organizations based in Japan, they were also asked for their views on the transparency reporting requirements under Japan's AI bill[11].

Examples of the interview questions

- Feedback on the HAIP Reporting Framework
    - What aspects of the HAIP framework do you find beneficial in terms of practical implementation?
    - On the other hand, what aspects of general public disclosure and reporting, in line with HAIP, do you feel need improvement?
    - How do you view the verifiability of the information disclosed under the HAIP reporting framework?
    - In your opinion, what kinds of incentive structures are necessary to ensure continued participation or to encourage new participants to join?

[11] The AI bill was enacted in Japan on May 28, 2025. See also "Japan enacts bill to promote AI development and address its risks" Japan Times (2025) https://www.japantimes.co.jp/news/2025/05/28/japan/japan-ai-law/

- What types of resources or support do you consider necessary to effectively engage with the HAIP reporting framework?

- Transparency and Accountability of AI Systems
  - Who are the intended audiences for the public disclosure of information about your company's AI systems? (e.g., consumers, industry stakeholders, regulatory authorities)
  - To what extent and in what form do you believe information regarding the risks and potential harms associated with your AI systems should be made publicly available? Why? (This includes but is not limited to HAIP-related disclosures.)
- Regulatory Interoperability (for interviewees affiliated with organizations based in Japan)
  - In relation to the AI bill currently under deliberation in the Japanese Diet, what kind of regulatory interoperability do you consider necessary between HAIP and the domestic legal framework in Japan?

The interviewees were informed that they were being asked to respond not in their official capacity as representatives of their companies, but rather in their individual capacity as the persons who had authored the responses to the HAIP reporting framework. They were also informed that their responses would be published as part of the research findings with anonymity preserved, and they agreed to participate on that basis.

While the original plan was to hold individual sessions with each company, in some cases, multiple companies participated in a single session due to scheduling constraints. Each interview lasted approximately 30 to 45 minutes per company. In some cases, not all questions could be covered, as the interviews concluded once the scheduled time had elapsed.

3.2 In-Person Stakeholder Consultation

On June 3, 2025, following the completion of the interview process, a stakeholder consultation meeting was held (Fig. 1). The event took place in person at the University of Tokyo in Japan. While not all interviewees were able to attend, the majority were present. In addition to the interviewees, participants included government officials from the Ministry of Internal Affairs and Communications and the Cabinet Office, members of the AI Safety Institute[12] and GPAI Tokyo Expert Support Center[13], and academic experts.

Following opening remarks by Mr. Yoichi Iida of the Ministry of Internal Affairs and Communications, who has been a driving force behind HAIP[14], the authors presented a summary of the interview findings. This was followed by a plenary discussion. The exchange of views took place under the Chatham House Rule, allowing for open and candid dialogue.

[12] AISI Japan https://aisi.go.jp/

[13] GPAI Tokyo Expert Support Center https://www2.nict.go.jp/gpai-tokyo-esc/en/

[14] OECD "Yoichi Iida : AI Expert" https://oecd.ai/en/community/yoichi-iida

Fig.1 Stakeholder consultation meeting

4. Key Findings

4.1 Diversity of Motivations

The interview findings revealed that the purposes and incentives for participating in the HAIP reporting framework vary widely. Five primary target audiences can be identified: international bodies, policy stakeholders, business and technical partners, the general public, and internal teams (Table 1). It became clear that the emphasis placed on each objective differed across companies, highlighting the diversity of motivations—a key finding of this study.

Some interviewees see the HAIP reporting framework primarily as a means of engagement with international institutions. For these organizations, participation in HAIP serves to signal alignment with emerging global norms. It becomes a tool to demonstrate their leadership in responsible AI governance, as well as to gain recognition and legitimacy in transnational policy arenas. Also, some interviewees expressed the expectation that establishing a certain degree of interoperability between the HAIP reporting framework and various national laws could help streamline reporting obligations and facilitate the development of coherent governance structures for companies.

Other interviewees direct their HAIP report toward policy stakeholders at the national or regional level. These include government ministries, regulatory authorities, and parliamentary actors involved in drafting or enforcing AI-related legislation. For such companies, reporting is not merely a compliance exercise but a strategic means to clarify their governance structures, articulate their risk management processes, and present themselves as responsible and policy-aware actors. By doing so, they hope to influence or contribute to regulatory development and demonstrate readiness for future oversight mechanisms.

In contrast, interviewees operating in business-to-business settings often tailor their HAIP reporting framework disclosures toward business clients and technical collaborators. For these companies, the value of the report lies in its ability to provide credible, detailed information that supports procurement decisions, vendor assessments, and ongoing technical cooperation. The emphasis is placed on transparency in development practices, model testing procedures, and risk mitigation frameworks. These reports are typically more technical in nature and are prepared with an audience of engineers, compliance officers, or integration partners in mind, rather than the general public.

At the same time, a number of interviewees focus their HAIP reporting on public communication and reputational trust. These reports are crafted for a broad audience that includes consumers, students, shareholders, and civil society organizations. In such cases, the report functions as a tool for demonstrating the company’s commitment to ethical AI development in a manner that is understandable and relatable to non-experts. Clarity, readability, and contextualization are prioritized, with efforts made to explain abstract concepts through familiar examples.

Finally, many interviewees highlight the internal benefits of the HAIP reporting process. Even when the report is outward facing, the process of preparing it often leads to valuable internal coordination across departments. The requirement to compile and structure responses to HAIP questions serves as a catalyst for internal reflection and helps to build shared understanding and accountability across teams. In this sense, HAIP reporting becomes a mechanism not just for external transparency, but also for enhancing internal governance and organizational learning.

Audience Type	**Description**	**Typical Motivation**
International Bodies	G7 / OECD Partners	-Visibility in AI governance -International alignment
Policy Stakeholders	Government bodies, regulators	-Gain trust -Influence on regulatory frameworks
Business & Technical Partners	B2B clients, external developers, corporate partners	-Contractual clarity -Risk accountability
General Public	Shareholders, citizens, job-seeking students	-Trust-building -Brand strategy
Internal Teams	Employees	-Create internal alignment and awareness on AI governance

Table 1 Motivations for participation in the HAIP reporting framework

It is also worth noting that the effort required to complete the HAIP questionnaire varied depending on each company’s existing internal practices. Some interviewees reported that they were able to respond primarily by reorganizing publicly available information, while others needed to collect new data internally and prepare additional documentation. In particular, several Japanese companies indicated that coordinating and explaining the process to relevant internal departments posed significant challenges.

Furthermore, multiple interviewees pointed out that the significance of the HAIP reporting framework was not yet well understood, either within their organizations or by external stakeholders. The interviewees emphasized the need for future efforts to raise awareness and recognition of the initiative. Furthermore, one respondent remarked that if shareholders and institutional investors come to recognize and value the HAIP reporting framework, it could serve as a strong driving force for companies.

4.2 The Trade-off Between Difficulty of Response and Flexibility

Regarding the questionnaire, many interviewees commented that "the intent of some questions was unclear and difficult to interpret," and that they were "unsure about the appropriate level of detail to include in their responses."

Specifically, there was confusion over the scope of disclosure—for example, whether companies were expected to provide information about specific AI models or about company-wide policies. Some interviewees, which act both as developers and users of AI technologies, also expressed uncertainty about which roles they should prioritize in their responses. In addition, B2B companies face unique challenges: while they may have documentation prepared for their business clients, they are often less accustomed to disclosing information in a way that is accessible to general consumers.

On the other hand, some respondents noted that the very ambiguity of the framework allows for broader participation. In particular, the broad and abstract definition of "advanced AI systems" has enabled the framework to accommodate emerging trends such as improved small language models (SLMs), the use of open-weight models, retrieval-augmented generation (RAG), and the rise of AI agents.

Related to this, some interviewees, after reviewing other companies' published responses, commented that "the level of detail varies widely." In practice, the volume and depth of responses differ significantly between companies. For example, one company submitted a report totaling around 10 pages, while another provided about 60 pages. As a result, some participants suggested that a word or page limit might be helpful, noting that "if the reports are too long, they become difficult to read."

However, part of this variation may be attributable to the previously noted diversity of motivations. That is, differences in the intended audience or communication targets—such as whether the response was aimed at government officials, customers, or the general public—may have influenced the length and content of the responses.

This raises a difficult question: Should the level of detail be standardized, or should flexibility be preserved so that each organization can tailor its responses independently? One participant remarked that while encouraging broad participation is important, a certain degree of consistency might be necessary to maintain the overall credibility of the framework.

Regarding future updates to the responses, the OECD expects update frequency is once per year. However, some interviewees expressed uncertainty about whether this frequency is appropriate. There is a recognized need for timely and appropriate reporting rules to prevent information from becoming outdated, as well as for the development of a flexible update mechanism.

4.3 Concerns About Overgeneralized Rankings

The HAIP reporting framework, by presenting organization responses in a standardized questionnaire format, facilitates comparative analysis. However, the balance between comparability and voluntariness is highly delicate. If simplistic scoring or ranking were to emerge or gain prominence, it could undermine organizations’ incentives to participate in the HAIP reporting framework.

In fact, during the stakeholder consultation, when the authors presented a set of evaluation criteria along with a sample ranking, some participants expressed strong concerns and objections.

The HAIP reporting framework should be positioned as a form of “credit for transparency efforts” that encourages voluntary disclosure. Of course, it goes without saying that verifying the integrity and appropriateness of disclosed information is important. However, such evaluation should not take the form of one-size-fits-all comparisons; rather, it should account for individual contexts, such as industry structure and cultural background. Initiatives like quality assurance and external audits might be better pursued through mechanisms separate from HAIP.

Within the HAIP reporting framework as a voluntary initiative, it is important to leverage the process toward establishing a de facto standard for information disclosure, facilitating the sharing of best practices, and enabling systematic monitoring.

In this context, it was observed that frequently used keywords in the responses—such as "safety," "quality," and "transparency"—often carry different nuances across organizations. To the extent possible, efforts should be made to promote consistency by referring to existing glossaries and terminology guidelines.

5. Proposals for Improvement

Based on insights gained through interviews and the stakeholder consultation, the authors will make the following set of proposals.

The first proposal is primarily directed at organizations participating in the HAIP reporting framework. It became evident that companies vary in both their reasons for participation and their intended audiences for the responses. However, modifying the structure of the HAIP questionnaire itself according to these varying purposes or audience types would risk reducing its flexibility and coherence, and is therefore not advisable. Instead, the authors propose that each organization clearly indicate, at the beginning of its response, which of audience types it is primarily targeting and, optionally, the reporting policy (Fig. 2).

Furthermore, the weighting of key perspectives—such as clarity, risk assessment, and technical accuracy—differs depending on the target audience type. Therefore, the authors intend to provide future guidance on good practices and recommended approaches to writing, tailored to each audience category.

In addition, it was observed that the same terms are often used with different meanings across responses, which reduces comparability. To address this issue, the authors also plan to introduce a glossary of commonly used technical terms in AI governance (e.g., those provided by the OECD and other relevant sources).

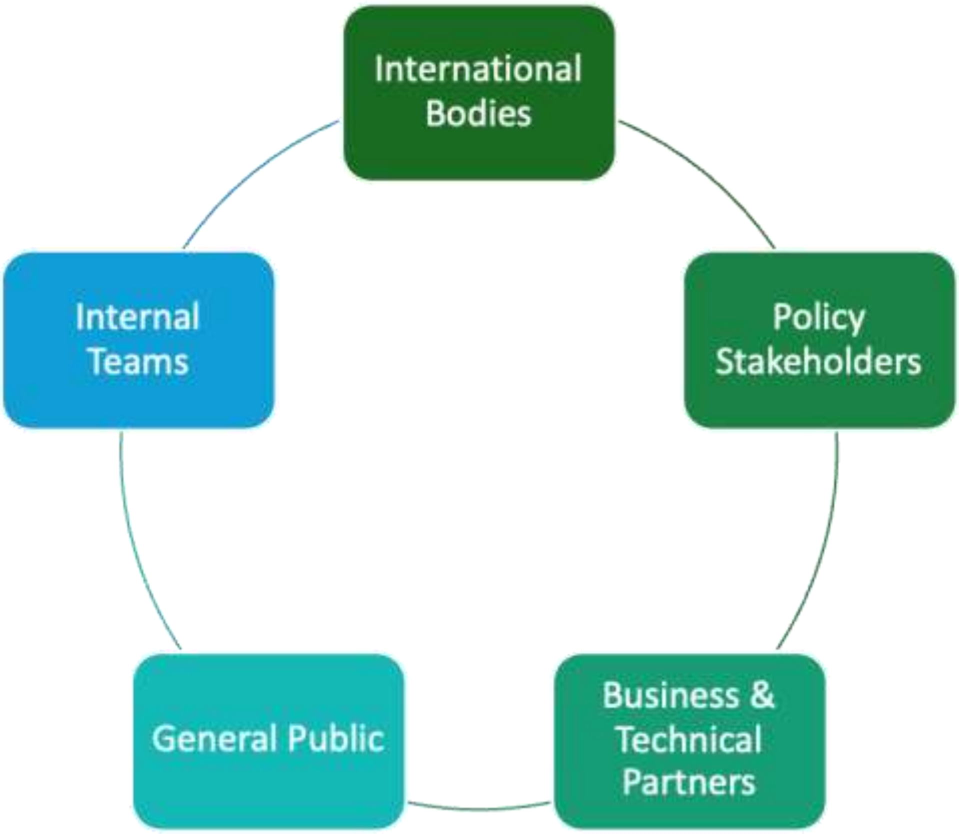

Fig. 2 Intended audience categories

The second proposal is directed at the OECD Secretariat and the governments of G7 member states, who are responsible for administering the HAIP reporting framework. Several interviewees have noted that the current questionnaire contains overlapping content, is overly lengthy, and is difficult to complete. In response, the authors plan to propose a more structured version of the questionnaire that retains flexibility while eliminating redundancy. It would also be an option for the secretariat to provide explanatory guidance and accumulate good practices.

Moreover, due to the currently low level of public awareness about HAIP, some participants expressed concerns that the initiative lacks sufficient understanding both within companies and among external stakeholders, including the general public. To address this, we propose launching a public awareness campaign to improve HAIP's visibility and credibility. In particular, considering that the rise of ESG investment has served as a major impetus for advancing corporate CSR and CSV initiatives, appealing to shareholders and institutional investors is likely to become increasingly important.

As concrete measures for the campaign, for example, participating organizations could be permitted to use the HAIP logo under certain conditions. Another option would be to foster momentum by organizing and hosting explanatory sessions at events attended by shareholders and institutional investors. Furthermore, in order to provide a platform for participating organizations to share updates on their latest initiatives and good practices, we propose holding an annual "HAIP Summit" in Hiroshima, the birthplace of HAIP.

The third proposal is addressed to those who read and assess the responses submitted under the HAIP framework. As previously discussed, the HAIP reporting framework is intended to function as a form of recognition for voluntary transparency efforts. Stakeholders—particularly audit firms, rating agencies, and the press—should be aware that (disclosure of) superficial scoring or overgeneralized rankings may lead to the sole focus on improving these indicators and that as this may ultimately undermine transparency in AI development as stated in the so-called Goodhart's Law.

As previously emphasized, verifying the accuracy and appropriateness of publicly disclosed information is a fundamental prerequisite. However, such verification should not rely on simplistic, superficial, and overly generalized comparisons. Instead, it should be based on evaluations that take into account the specific circumstances of each case, such as industrial structures and cultural contexts. Matters such as quality assurance and external audits could be addressed through frameworks separate from HAIP reporting framework.

The authors plan to formally present these proposals by July 2025, and welcome feedback and continued dialogue from all interested parties.

6. Conclusion

As discussed above, while the HAIP reporting framework faces practical challenges, it holds significant value as a mechanism for promoting voluntary information disclosure. In particular, interviews repeatedly highlighted two key points: first, that corporate transparency serves as a foundation for trustworthy AI governance, and second, that the disclosure process itself can serve as a catalyst for organizations to reevaluate and strengthen their internal AI governance practices.

Accordingly, the responses should not be viewed primarily as demonstrations of corporate superiority, but rather as public goods that advance transparency and accountability in the AI domain. Indeed, some stakeholders emphasized the importance of fostering a culture in which organizations are not penalized by participating, and where the very act of submitting a report is seen as commendable. This perspective should be reflected in future HAIP outreach and awareness efforts.

The HAIP reporting framework represents a pioneering initiative in the landscape of global AI governance. Its development and implementation have the potential to shape transparency practices in other jurisdictions. Through the improvements proposed in this paper, we hope the framework will evolve into a more effective and inclusive mechanism for promoting responsible AI.

Acknowledgements

We would like to express our sincere gratitude to the following individuals and organizations for their valuable cooperation.

Individuals (in alphabetical order):
Hiroyuki Abe, Junichi Arahori, Alexandra Belias, Claudia Fischer, Aliki Foinikopoulou, Jun Fukuoka, Kyohei Fushida, James Gealy, Akihiko Hitomi, Eunice Huang, Momoko Hosono, Yoichi Iida, Takaharu Ishikawa, Ryoki Kishi, Haruki Kojima, Takeshi Komoto, Koichi Takagi, Masakazu Takahashi, Atsushi Tsuchida, Kenji Urano, Miho Naganuma, Muneki Nemoto, Kenta Oono, Kazuya Okubo, Yasutaka Ota Osamu Saito, Motohiko Sato, Yuichi Saho, Akiko Murakami, Kenji Zaitsu, Rie Yomogida, Akira Yoshida, Kazuhiro Yoshinaga

Organizations (in order of the HAIP-OECD submission):
KDDI Corporation, SoftBank Corp., Preferred Networks, NEC Corporation, NTT, Microsoft, Salesforce, OpenAI, Google, Fujitsu, Rakuten Group.

Additional organizations were invited, and we look forward to including their input in future versions.