API-NAVER **JOINT SEMINAR**

Trustworthy Hyper-Scale AI: Cooperation Between Platforms and Start-ups

Summary

Trustworthy Hyper-Scale AI: Cooperation Between Platforms and Start-ups

Summary[1]

Contents

---

[1] Prepared by Alexandra Son (SAPI), Jooyeon Kim (NAVER), Yong Lim (SAPI).

Opening
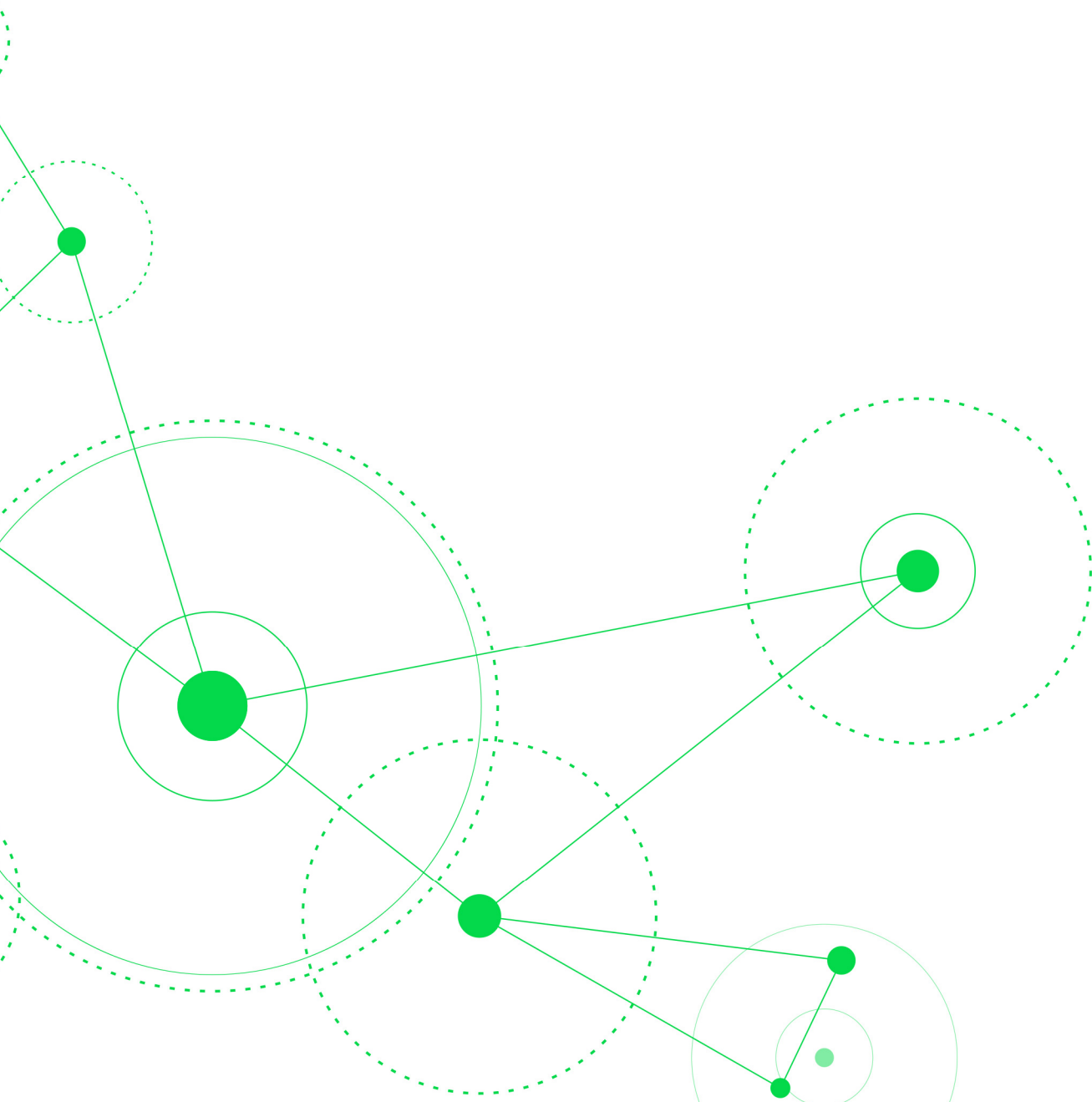
AI Ethics and Start-Ups _Yong Lim

- **Codifying ethics into law and the challenges it poses for start-ups.**

  - In the field of AI governance, we are now observing a progression from *ethics* to *law*. While this shift was expected, the major legislative proposals currently being discussed in the EU and elsewhere attempt to codify what was previously discussed as ethical demands into normative rules.

  - However, this may create problems.
    - Ethics is inherently idealistic and aspirational, and therefore naturally flexible.
    - Codification could result in unattainable, conflicting, and rigid rules.
    - This can increase the costs of compliance, particularly for start-ups.

- **Normative discussions continue to focus on Big Tech.**

  - Despite being a *universal* requirement, AI ethics discussions in the past centered around big tech. This trend seems to continue in the discourse surrounding AI law.

  - Current legislative proposals do allow for some considerations for start-ups.
    - EU's *Artificial Intelligence Act* states that additional measures (that are different from those applied to gig tech) are necessary to "support innovation, in particular through AI regulatory sandboxes and other measures to reduce the regulatory burden and to support Small and Medium-Sized Enterprises ('SMEs') and start-ups."

- Specifically, protocols for verification should be differentiated between big tech and start-ups to support innovation by SMEs and start-ups.
- The *Act* also provides that the SMEs' situation should be considered when assessing penalties.
  - Korea's draft AI law also stipulates financial support for SMEs.

- **We need a more concerted and focused effort to approach AI governance from the perspective of start-ups.**

  - It is crucial to involve SMEs in the process of establishing technology standards and collaboratively develop those standards. This ensures that the standards take into account the unique circumstances and realities faced by SMEs.

  - There is currently significant uncertainty surrounding AI ethics, and this is more of a problem for start-ups.
    - There are still many aspects to consider, such as the allocation of roles and responsibilities among stakeholders, the balance between innovation and safety, and the prevention of undesirable futures.

Session I.
AI Ethics on the Ground: The Case of the Hyper-Scale AI Industry
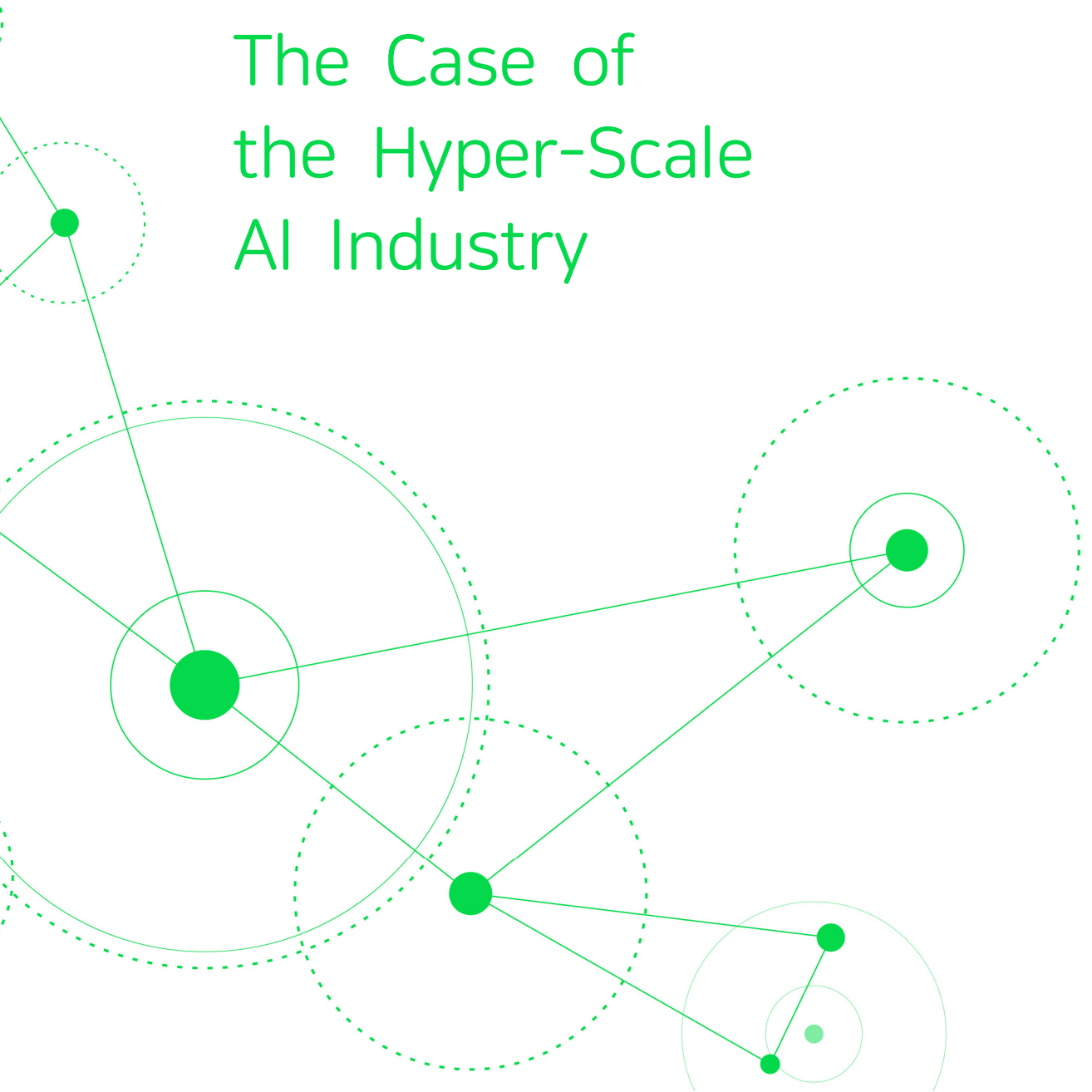
CLOVA Studio AI Ethics Guide:
Background and Directions _Woo-Chul Park

- **NAVER services incorporate various AI technologies.**
  - NAVER is developing technology so that AI can be utilized as a daily tool.

- **AI HyperCLOVA (revealed in May 2021)**
  - Generative AI can turn a good idea into a business opportunity.
    - Unlike traditional businesses where capital or technological requirements pose significant barriers, anyone with a great idea can start a business.
  - Generative AI is capable of performing various tasks such as sentence generation, classification, summarization, and more. It opens up diverse business opportunities.

- **LOVA Studio is built on a large language model, *i.e.*, HyperCLOVA, to allow users to access the functionalities they need, including sentence generation, conversion, summarization, data classification.**

- **NAVER's effort to tackle ethical concerns accompanying the opportunities and possibilities provided by AI - *What does society demand from AI companies?***
  - NAVER AI Ethics Principles
    - Developed to reflect NAVER's corporate philosophy and perspective on AI, in consideration of social demands on AI companies.
  - CLOVA Studio AI Ethics Guide

- Written to help understand the various situations users may encounter while using CLOVA Studio from an AI ethics perspective, and to provide assistance in preventing and addressing related issues.
- NAVER/User Responsibilities
  - NAVER explains the NAVER AI Ethics Principles applicable to CLOVA Studio, as well as technological tools to implement these principles.
  - NAVER engages in communication with users and takes necessary measures to improve user experience within reasonable boundaries when issues are raised.
  - Users should refrain from engaging in malicious or harmful use of the service.
- Procedural & Technical Efforts
  - Procedural Efforts
    - Service App Review Process
      - Sharing AI ethics-related video materials.
      - Providing support for service policies and testing.
  - Technical Efforts
    - AI Filter
      - Prohibiting the output of inappropriate content.

CLOVA Studio Service Direction

- Complete pre-control of AI is difficult. Creating an AI ecosystem offers numerous business opportunities but also comes with challenges.
- If start-ups encounter issues while using CLOVA Studio, they are encouraged to report such issues to NAVER to foster interactive collaboration.

eCream's Reflections on AI Ethics _Seon-Yeop Kim

- **Ethics is a heavy subject**.
  - If AI ethics is not regulated, start-ups directly facing consumers will be the most affected.

- **"ANATE Philosophy" and Ethical Issues**
  - ANATE is an AI service that specializes in creativity.
    - The business model focuses on "AI coexisting with humans" and using AI as an "assistive" tool.
  - ANATE serves as an AI toolkit that enables web novel creation.
    - Exclusive tool for author communities or authors.
  - Ethical Issues
    - Ethical considerations regarding entertainment.
      - Positive information vs. negative information
        - Since AI development requires training data, if the data provided contains negative information, the output may also have negative implications.
      - Adult content vs. family-friendly content:
        - Currently, the training data includes filtering for adult content, but the demand for such content exists in markets like gaming.
        - What are the limits? Should AI be limited to family friendly content when humans do not have such limits?
        - Maintaining HyperCLOVA's internal regulations regarding prohibited content.
        - Planning self-imposed restrictions and solutions on prohibited content or generated material.

- Artificial limitations on creativity.

- Ethical considerations regarding information:
  - Copyright/intellectual property
    - Lack of clarity necessitates the need for guidelines ⇒ eCream ensures that all AI creations are credited to the prompt writers.
  - Responsibility for the leakage of personal information:
    - Sensitive area for companies – if personal data protection is not guaranteed, users will be reluctant to use the product.
    - How is the privacy of training data ensured? AI retains information learned, similar to humans.
  - Plagiarism prevention:
    - A sensitive issue for authors.
    - Currently using different search engines internally to check for plagiarism.
  - Hallucination:
    - Users are informed during sign-up or service usage.

- Ethical considerations regarding humans:
  - What if AI harms humans?
  - Addressing the tendency towards bias:
    - Preventing the systematization of hate and extremism.
  - Can AI misuse be prevented?
    - Excluding services and solutions that can replace humans entirely.
    - Highlighting the existence of areas where humans are irreplaceable within the service
  - A channel of coexistence
    - Creating opportunities for human creative services and planning using AI.
    - Creation of novels, creative writing contests, *etc.*

Responsible AI Development
– Balancing Innovation and Ethics _Deok-Min Ryu

- **First developer to use GPT3 for the purpose of Korean writing.**

  - Initially, only OpenAI's GPT was utilized, but later CLOVA was introduced. Since CLOVA has been trained on a significant amount of Korean language data, it excels at generating content that captures Korean sentiments effectively.

- **Reverse engineering**

  - Reverse engineering can be employed to exploit vulnerabilities in AI systems ("confuse" the AI) and bypass regulations.

- **Due to the absence of legal advisors on the issue of AI, the responsibility for content generation had to be placed on humans through a process of "AI initial draft, followed by human editing."**

  - However, there were concerns about the ownership and attribution of AI-generated content. AI-generated content also often lacked the desired length or structure required by authors.

- **Ethical regulation poses a challenging issue.**

  - It instills significant fear in individuals intending to start a business ⇒ can be compared to the question of where one should use dynamite, but the decision ultimately rests with the users.

  - Start-ups thrive on pushing boundaries and introducing novel ideas. Therefore, excessive filtering or regulation is deemed unsuitable as

it hampers the spirit of innovation.

- **Poetry publication project.**
  - Able to write poems with only the given theme and keywords ⇒ effective time-saving tool.
  - It is possible to publish a poetry collection in just one day if you set your mind to it (excellent literary tool!).
  - However, no one revealed that AI has written the poetry ⇒ raises questions of whether this denial is ethically appropriate.

- **Various forms of creative expression are possible (such as drawing, composing, *etc.*).**
  - But there are issues regarding copyright and originality.
    - For example, Naver's blog algorithm prevents users from simply copying another user's content and using it as their own.
      - AI solves the imitation issue ⇒ however, as it is also AI that polices the plagiarized content, it becomes a catch-22 situation.
  - It is also possible to draw using prompts.
    - There are ethical issues such as the generation of adult content ⇒ a scary problem for AppPlatform.

- **Use of AI in elementary schools.**
  - Often used as a tool to teach students English.
  - The biggest problem is Google's age limit for using their services, which is 13 years or older.
  - Another concern is ensuring safe use of children.

- **"Bad User"**
  - Issues are frequently raised when AI produces undesirable or

inappropriate outcomes. Who is liable for this, the start-up or the "Bad Users"? (a question of agency)

- AI ethics issues require social consensus and a common denominator.
  - The industry faces numerous challenges, and ethical concerns are extensive in scope.
- Regulatory Sandbox
  - It is believed that matters concerning artificial intelligence should be placed within a sandbox for protection.

Experiencing AI Ethics at a US Start-up

_Unso Jo

- **Hugging Face is a US open-source platform for AI models in modalities such as vision, language, and voice.**
  - It is like YouTube – on YouTube, users can upload and share videos, while on Hugging Face, users can contribute and share natural language processing (NLP) models, datasets, and other NLP-related resources.
  - It does not do its own research or develop NLP models.
    - Researchers, students, or private developers develop the models, and upload them to the platform.
    - There are more than 200,000 models.

- **Hugging Face does not require the models to be reviewed and does not take responsibility for the models' contents.**
  - However, Yannic's GPT-4chan became an issue in 2022.
    - GPT-4chan was trained on 4chan's "Politically Incorrect" platform.
    - If this model is taken down, other "inappropriate" models must also be taken down. Decisions about model/dataset content are made case-by-case, and Hugging Face takes a media centric approach to reviewing models.

- **Emphasis on Transparency**
  - Wishes to shape the platform's culture in a way that emphasizes transparency.

● **Ethical Charter type cultural/media-oriented AI literacy is important for Hugging Face to show they take initiative in leading the AI community towards better and more ethical AI.**

NAVER SAPI