Correspondence

Check for updates

My career path at a medical artificial intelligence company, working as a physician outside of clinical practice

Chang Ho Ahn

Lunit, Seoul, Korea

After serving as a clinical instructor in the endocrinology department at Seoul National University Bundang Hospital in Seongnam, Korea, I transitioned to Lunit, a medical artificial intelligence (AI) company, in 2023. While my experience as a clinical professor was not extensive, and I am still growing into my new role, I am eager to share my journey through this essay. I hope to provide insights to medical students and junior colleagues who are considering a career in the AI industry.

The artificial intelligence company where I work

The company I work for is dedicated to combating diseases that afflict humanity, using artificial intelligence (AI) as its primary tool. It has specifically prioritized the fight against cancer among various human diseases. Our approach to tackling cancer involves two main strategies: the early and accurate detection of cancer, and recommending the optimal treatment method by understanding the biology of each patient's cancer. I am part of the oncology group, which focuses on developing AI biomarkers from tissue pathology image data to recommend the most effective treatment options [1]. The company was founded in 2013 and went public on the KOSDAQ in 2022. It is currently in a phase of expansion, marked by an increase in both the number of employees and revenue. However, further development is still necessary.

My role in the company

Within the company, I am part of the medical affairs department, which consists of teams focusing on clinical research, biomedical research, data management, and medical product management. As indicated by the names of these teams, our department's responsibilities include conducting various clinical studies with developed AI products, collecting and processing the data necessary for the development of new AI products, maintaining and updating existing AI products, and transforming the primary analysis results from AI models into biomarkers that are more intuitively understandable for humans.

Since AI is a new technology, AI companies are relatively new startups. In the traditional pharmaceutical industry, the role of physicians is well-established; however, in AI companies, this role remains undefined. Drawing on my medical expertise, I engage in a variety of tasks required

by the company. My responsibilities span from direct involvement in AI product development to strategizing on revenue generation with AI products within the business domain. Additionally, I focus on navigating product approvals and building clinical evidence in the clinical research and approval domain. Despite my expertise, I collaborate with field-specific experts who possess even greater knowledge in these diverse areas. One of the highlights of my tenure at the company has been the opportunity to learn from these experts across various disciplines and to demonstrate my value by contributing vital ideas through our interactions.

Product management as an essential task

One common question I receive from acquaintances and juniors curious about my role at the company is whether my work primarily involves simple tasks such as reviewing medical images or pathology tests, and how much programming knowledge is necessary for actual AI development. While my duties do include processing medical data directly, I find that a physician's expertise is most valuable in the area of product management. The company's goal is to sell products that generate profit. Product planning requires identifying what consumers in the healthcare industry—patients, physicians, insurance companies, and government agencies—need, and determining how our products can meet those needs. Physicians have a distinct advantage in this area due to their direct experience with clinical unmet needs. However, product management involves more than just presenting ideas. It also includes the realization of these ideas by leveraging the company's human and material resources. Therefore, launching a product on the market requires knowledge of AI development, software development, product approval processes, and even basic business and accounting principles. These skills collectively address the question of how much programming knowledge is necessary. Although I do not engage in coding at the company, I have studied the fundamentals of programming and have actively pursued knowledge in AI technology. Such a technological understanding is essential to fully utilize medical expertise. To take a more proactive role in the field of medical AI, rather than merely acting as a medical advisor, one must be capable of addressing issues such as why the performance of an AI model under development is unsatisfactory. This involves determining whether the issue stems from specific characteristics of the medical data and deciding which of various solutions—such as modifying the data or altering the training method—has the highest likelihood of success.

Incorporation of various artificial intelligence models

Although I work in the AI industry, I cannot confidently predict how AI technology will transform healthcare. This new technology is characterized by a mix of excessive expectations and high possibilities. At my company, our staff members are constantly striving to remain at the forefront of technological advancements. They are working to integrate foundation models, zero-shot learning, and vision language models into our products. A foundation model [2], also known as a base model or pre-trained model, is a large neural network that has been trained on a vast amount of data in a self-supervised manner without being optimized for any specific downstream task. These models learn general representations and patterns from the data, capturing broad knowledge that can then be adapted or fine-tuned for various downstream tasks through transfer learning. Examples of well-known foundation models include GPT-3 for natural language processing, Bidirectional Encoder Representations from Transformers (BERT),

and Contrastive Language-Image Pre-training (CLIP) for vision-language tasks.

Zero-shot learning [3] is a paradigm where a machine learning model is trained to generalize and perform tasks without explicit training on examples from those tasks. The model utilizes knowledge gained during pre-training on a broad dataset to make inferences and predictions about new, unseen tasks or classes. In zero-shot learning, the model receives either a natural language description or a few examples of the new task, using its understanding of underlying concepts and relationships to adapt its knowledge accordingly. This approach enables the model to handle tasks it has never encountered during training, potentially diminishing the need for extensive data collection and annotation for each new task.

A vision-language model [4] is a type of multimodal neural network that is designed to simultaneously process and understand both visual and textual data. These models are trained on extensive datasets that include images along with their corresponding textual descriptions or captions. Through this training, they learn to link visual features with related language representations. Vision-language models are utilized in a variety of tasks that involve both visual and textual inputs, such as image captioning, visual question answering, and multimodal retrieval. Notable examples of vision-language models include CLIP, Vision and Language BERT (ViLBERT), and UNiversal Image-TExt Representation (UNITER).

CLIP, developed by OpenAI, employs a contrastive learning method to train on a vast dataset of image-text pairs. It is capable of performing zero-shot classification and image-text matching tasks without the need for fine-tuning. CLIP is celebrated for its flexibility and its proficiency in adapting to new tasks. ViLBERT modifies the BERT architecture to accommodate both visual and textual inputs. It processes images and text through two distinct streams, which are later merged to create a unified representation. ViLBERT is specifically designed for tasks such as visual question answering and image-text retrieval. UNITER offers a unified architecture that integrates image and text processing into a single stream. It leverages a transformer-based model to develop joint representations of visual and textual information. UNITER is adept at handling a variety of vision-language tasks and has demonstrated robust performance in visual question-answering and image-text matching tasks.

Exciting experience of overcoming technical limitations

Participating in such projects is particularly exciting and can be thoroughly enjoyed within the company. While there are moments when I am amazed by the potential of these experiences, there are also times when I clearly see the limitations of current technology. In developing products, engineers do not always apply the latest AI techniques; instead, they sometimes opt for more traditional AI architectures because of their distinct advantages. Overcoming these limitations is a major concern that I share with the AI research team.

The work environment at the company is less structured than in a hospital, with unexpected tasks frequently arising. This can be viewed either as an advantage, lending to a dynamic atmosphere, or as a disadvantage, contributing to instability. I have progressed past the adaptation stage and am now focused on developing my competitiveness. Although I no longer provide direct patient care, I am committed to engaging in meaningful work that extends beyond merely seeking profit. My limited experience may still offer valuable insights to the readers.

ORCID
Chang Ho Ahn: http://orcid.org/0000-0002-0702-0608

Authors' contributions

All work was done by Chang Ho Ahn.

References

1. Eom HJ, Cha JH, Choi WJ, Cho SM, Jin K, Kim HH. Mammographic density assessment: comparison of radiologists, automated volumetric measurement, and artificial intelligence-based computer-assisted diagnosis. *Acta Radiol* 2024;2:2841851241257794.
   https://doi.org/10.1177/02841851241257794
2. Wikipedia. Foundation model [Internet]. San Francisco (CA): Wikipedia; c2024 [cited 2024 Feb 28]. Available from: https://en.wikipedia.org/wiki/Foundation_model
3. Xian Y, Schile B, Akata Z. Zero-shot learning: the good, the bad and the ugly. arXiv: 1703.04394v2 [Preprint]. 2020.
   https://doi.org/10.48550/arXiv.1703.04394
4. Bordes F, Pang RY, Ajay A, Li AC, Bardes A, Petryk S, et al. An introduction to vision-language modeling. arXiv: 2405.17247v1 [Preprint]. 2024.
   https://doi.org/10.48550/arXiv.2405.17247