김건희 교수님 세미나 보고서:

During this lecture, professor Kim mainly talked about his project called "SODA" and mentioned two research papers that he and his team published last year. SODA is mostly about a language model that works more efficiently and better than language datasets that are in the world.

Professor Kim believes that there is a lack of dialogue data given to us, therefore people used to hire two people to talk to each other. But this was very not efficient as human beings are difficult to have a conversation in their own due to their contrast in age, gender, etc. Therefore it is better to use existing dialogue datasets. For example, an existing dataset would be Wizard of Wikipedia, aka WoW, which allows the AI chatbot model to elaborate more on what the human is giving. But the datasets are not enough. There is a lot of vacuum space that hasn't been covered yet. And a solution to this is crowdsourcing. Chatgpt is buying all the datasets from the world and is getting stronger by day. Samsung, kakao, and naver are all giving up in this field of chat bot because they cannot compete with chatgpt. If chatgpt is able to freely use Korean, it will hugely affect the Korean companies. Chatgpt has already posed as a threat to Google, who is number 1 in the field of search.

He also talked about ProsocialDialog, which is a dialogue model that give you positive feedbacks no matter what you say. This is bad somehow because it doesn't really think about how agents should react to problematic contexts. Chatgpt has somehow solved this by when being asked about a controversy, chatgpt talks about two sides, also fades away from the topic.

Professor Kim believes we should use these advances in technology to find an opportunity to use it to our own.