Question: Could you talk about Schroders track record of working with startups? E.g., I know that last year Schroders was involved in the OIP as well. Did that exercise a success?

Answer: Yes, the project was a success. The project is near completion and they have recently presented their project in IMAS Digital Summit 2021.

Question: Could you please name the aspects of spaCy that made you look for a different solution? Is that the special types of entities you need to extract, terminology customisation, or something else?

Answer: I mentioned spacy, it was in used in our LIBOR use case because spacy is NLP and we used it to help us leverage the lemmatization. For example, when we are working with our legal colleagues and they gave us the search terms, it included terms like act of God, act of Gods – singular and plurals. So, we leverage libraries like spaCy to handle the variations or lemmatizations of the word rather than, having to type them all out. But spaCy is for NLP – for the 4 documents types that I have shown, the challenge is identifying the relationship between the figures and the roles itself. Having an accurate extraction is actually the challenge, it's not so much on NLP side.

Question: What languages do you need to support?

Answer: Let us stick to English, that is where all the information is coming from, that is where it is being processed. It covers 80-90 plus percent.

Question: Is there an internal repository with source documents to be assumed tagged for each client/account? How standardized (document file types like pdf.) are these?

Answer: The likes of the data providers from the Bloomberg's, FactSets, Morning Star, Refinitiv,, they do have financial information and data but where we see the challenge is actually in terms of emerging market or growth markets as well – that's where the data are sparser, the coverage is lesser but that's where a lot of activities are because our teams do need to analyse and understand the performance or even the modelling of the data from such countries. By saying tagging from it, where I think you are trying to relate to, would be more towards to machine learning – to train the models against a specific set. Where I see this is that it is going to be challenging in terms of finding out all the source to really tag it out – because our team, while we have covered Singapore, Indonesian, London, and Zurich, we are a very small team. One key thing which is why we leverage the OIP and IMAS partnership is that we do not have the bandwidth to research and build up the product on our own. That is why we are doing very selected business use cases to solve for the business challenge based on whichever client that send us more documents – it is a bigger load for the business, that is when we will go in and solve it. There is a benefit analysis that we work with in order to identify which processes and which are the things we should solve for. Where I have heard of, or at least discussions with several vendors as well – not going to name them, I think if you go

down the ML route, it is possible but it's going to be a massive effort as well. I think I have a certain reservation for it because it is a niche, we are looking for a much more narrow or dedicated AI solutions if we really want to say in that way.

Question: Are you open to proprietary taxonomy / term management solutions?

Answer: That is fine. I came from a vendor space previously. Our team, our main focus is delivery the solution for the business. It does not have to be something that we build, we have leverage (inaudible), new algorithm, or research results that are public from Github etc. What I would go is always how well does the solution work, can it fit our business use case – that means if I send you a document, we can give you certain parameters, can it extract the information, is it accurate, and from the information extracted, can we then use it for downstream processes? That is why I would rather draw a blue box, or a black box would work as well, but I think everyone should understand it.

Question: What are the current sustainability standards Schroders are compiling with? (e.g., GRI, SASB)?

Answer: I need some time to check on this but let us not restrict the participants based on this.

Question: In terms of IP that could be developed to answer this challenge, would Schroders be willing to own the algorithms being developed or would the remain the propriety of the Start-up?

Answer: This will be subject to further discussions.

Question: Will you provide samples to train models for text extraction?

Answer: Yes. A selected set of samples and directions can be provided as guidelines

Question: Since you mentioned LIBOR as a use case, wondering if you will also be looking to address the challenge related to extraction of LIBOR fallback information too?

Answer: No. The LIBOR use case has been completed. But if there is an existing solution that can be shared we would be open to hear about it

Question: The documents are electronic PDF or scanned PDF? (since electronic PDF we don't need to do OCR)
Answer: Let us focus on electronic PDF for now.

Question: Are you open to proprietary taxonomy / term management solutions?
Answer: Yes.

Question: Are you looking at an app or web platform for your solution?

Answer: No restrictions as the data and information would be used in further downstream processes.

Question: Will you provide samples to train models for text extraction?

Answer: Sample documents are available but they are not tagged as BAU activity is focused on extraction rather than tagging of the document

Question: May I know who will be the main user of your platform? Sustainability compliance team or operation team?
Answer: The Operations Innovation team would front it to feed it to downstream processes

Question: Is the focus to just extract and store in database? Not in scope: a. Information retrieval: index, search sort, etc; Inference: b. Summarisation, paraphrasing models in scope
Answer: The storage is for ease of reference. If there are additional processing to enrich the information retrieval that would depend on its relevance to the information

Question: How many different formats for PDFs needs to be supported?
Answer: Let us focus on the general PDF standards. But if its externally produced PDFs the format might change (e.g. PDF/A, PDF/E) but that can be kept for later stages.

Question: What are the main document types to be process? Is it solely on financial statement type of data?
Answer: The financial document varies but financial statement is a key example in this case.