Goal : delete all manual setup to build a goal-oriented bot (document creation, price estimation, etc…)

Problem : the manual setup of those kind of chatbots is time-consuming and error-prone.
No scripted solutions seem to be efficient enough to reduce setup duration and errors.
We always turn around a useless question : is it better to use filters or gotoressources / bots.
Can be explained deeper if needed.

Hypothesis : switching the way we see the problem could offer a solution.
The final goal of these chatbots is to perform an action.
Let's say the action is creation a contract for a bank's corporate customers (guarantee contracts).

Let's also say we have access to 10k+ of those contracts.

Here's our hypothesis (so the question for Dino) :

1) Know the list of attributes needed
2) Turn those 10k+ pdf files into text files (OCR, some open source stuff should exist)
3) Manually train the NLU to extract entities (needed attributes) from a sample of this source of contract
4) Apply the created algorithm to the 10k examples
5) Develop a Neural Network able to score the probability of needing an attribute compared to one another
6) Setup the bot and run the logic from the created NN

We can provide:
- the list of attribute needed
- a trained NLU to extract the entities from the PDF

Need to be done:
- OCR: PDF to Text
- Entities extractor that can run on 10k text files
- Create the NN for the logic of the bot