Text comprehension vs. discursive chatbots

Chatbots, so popular these days, are all large language models (LLM). These are probabilistic models which try to learn the probability of words to form sentences from extremely extensive training on human-written language. They are based on semi-supervised deep learning techniques and are aimed at discursive applications, essentially answering questions, or writing essays.

While they have amazing capabilities they also have substantial drawbacks:
- The patterns identified in the unsupervised phase are not subject to the huge number of constraints characterizing human language, those of "making sense". Since they do not incorporate any semantics they are prone to many embarrassing blunders.
- They require extreme training and are thus very costly.
- They cannot tackle applications in which text comprehension is required, for example the summarization of large, unknown documents.

There is a huge number of applications in which text comprehension is required, notably the "Holy Grail" of discovery, the finding of knowledge not previously known to any human. Discovering a new drug goes way beyond passing the bar exam…and is also much more profitable. InfoCodex has been designed for this type of application.

Semantic is typically subdivided in lexical semantics, dealing with the meaning and relations between words, and logical semantics, addressing the more difficult tasks of reference, presupposition and implication. Lexical semantics can be hard-wired in principle, although this is a huge task. This is what InfoCodex has achieved, by constructing possibly the largest linguistic structure in the world, which combines universal thesauri with ontologies. This linguistic structure is then combined with machine learning techniques to produce a software that can comprehend text. Accordingly, InfoCodex was possibly the first to prove machine discovery, by finding previously unknown biomarkers for diabetes only based on large text data. Of course, it can also summarize unknown texts of any size. By construction, InfoCodex does not need to be trained! It can thus be deployed immediately in any new situation. This abates enormously the cost of its applications.

Instead of pursuing the speeding train of large language models, where several tech giants have a sizable advantage, we propose to conquer the terrain of text comprehension applications where InfoCodex has a clear advantage by its unique technology. The market is as large, if not larger and the construction of the linguistic structure took 100 man/years, providing a sizable competitive advantage. InfoCodex is a small company that does not have the power to do this alone. Its product, however, has proven its worth with many customers of first quality.