**CDL Cohort Project 4**

**1. Quantum Natural Language Processing**

Quantum Natural Language Processing (QNLP) uses quantum algorithms to process and analyze natural language. As with conventional NLP, the aim is to extract information and insights, or even create documents from scratch. However QNLP promises to be much more powerful than conventional NLP, because it naturally accommodates things like grammar and context, and its power will scale quickly as quantum hardware develops. This document gives an overview of a project to apply QNLP in finance.

**2. Examples of the types of real-world problems this solution can solve**

Natural Language Processing concerns the interface between human language and machine language, and so is useful whenever humans are trying to control computers, or computers are trying to understand what the humans are talking about. The potential business applications are therefore limited only by the accuracy of the algorithms.

Some popular applications for NLP are chatbots and virtual assistants (they never sleep); human resources (e.g. scanning resumes and job applications); speech recognition and machine translation (voice-to-text translation in real time is the aim); social media monitoring (by companies, or state actors); survey and sentiment analysis (how that new product release is going); text analysis and text extraction (e.g. in aviation, scan airplane maintenance manuals for relevant instructions); autocorrect spelling and grammar (because people can’t write proper); business and stock market analysis (parsing financial reports); credit scoring (companies like Lenddo include things like social media); healthcare (managing patient data and feedback, scanning for signs of pandemics); defence (“I could tell you, but then I would have to kill you”); and auto-generation of reports (was this written by a machine?)

While NLP is improving all the time, when it comes to human-machine interactions, some things still remain lost in translation. With online search, for example, search terms are generally treated as distinct keywords rather than being combined into a grammatical whole. And when it comes to things like nuance and irony, it seems that humans are just not machine-readable. Or is the problem that we are using the wrong kind of machines?

While “meaning aware” NLP tasks are believed to be so difficult that they will remain out of reach of classical computers for the foreseeable future, quantum NLP promises to move beyond these limitations by treating language as a quantum system in its own right. Unlike conventional NLP, which deals with words as distinct entities, the quantum version treats them as quantum states which are entangled through grammar and meaning. As DisCoCat co-inventor Bob Coecke puts it, language is “quantum native” in the sense that “language seems to want to live on quantum. Quantum systems want to be simulated on a quantum computer.”[[1]](#footnote-0)

This explains why companies such as Cambridge Quantum Computing, where Coecke serves as senior scientific advisor, are moving into the area. According to CEO Ilyas Khan, their quantum-based approach “could improve a broad range of applications where reasoning in complex systems and quantifying uncertainty are crucial, including medical diagnoses, fault-detection in mission-critical machines, and financial forecasting for investment management.”

**3. Core business model: QNLP for foresight**

The business model that we focus on is providing market foresight, based on the interpretation of financial reports, market sentiment, and transaction data.

In today’s competitive and data intensive business environment companies need to analyze their competitive environment more profoundly and be able to anticipate future opportunities and risks faster, creating a need for more accurate competitive intelligence. Foresight study, especially weak signal detection, is an effective technique to get this information. (A weak signal is an indicator of a potentially emerging issue that may become significant in the future.) This research area has been widely studied for its utility, and many corporations have their own teams dedicated to spotting weak signals and emerging trends. The biggest problem with weak signal detection but it is limited by the need of human expert judgments on these signals. Moreover, the increase in the volume of information on the Internet through blogs and web news has made the detection process difficult, which has created a need for automation.

Detecting weak signals and future scenarios scanning have been studied by many researchers and increasingly, in the age of big data, text-mining experts, AI developers and data scientists have entered in the game using “new” technologies, like NLP, to automate foresight processes. Despite all the research in the area, most of the automation systems still need involvement of human experts to get any usable results out of them.

Understanding and manipulating language are tasks that are too complicated for most of the current systems. For example, one popular machine learning approach in finance is to look at investor sentiment, as measured by things like hashtags on Twitter. The limitations of such approaches is shown by the fact that the Eurekahedge AI Hedge Fund Index, which tracks the returns of 13 hedge funds using machine learning, has had an average annual return for the past five years of 5.5 percent, as compared to 12.5 percent for the S&P 500. We believe that quantum can do better.

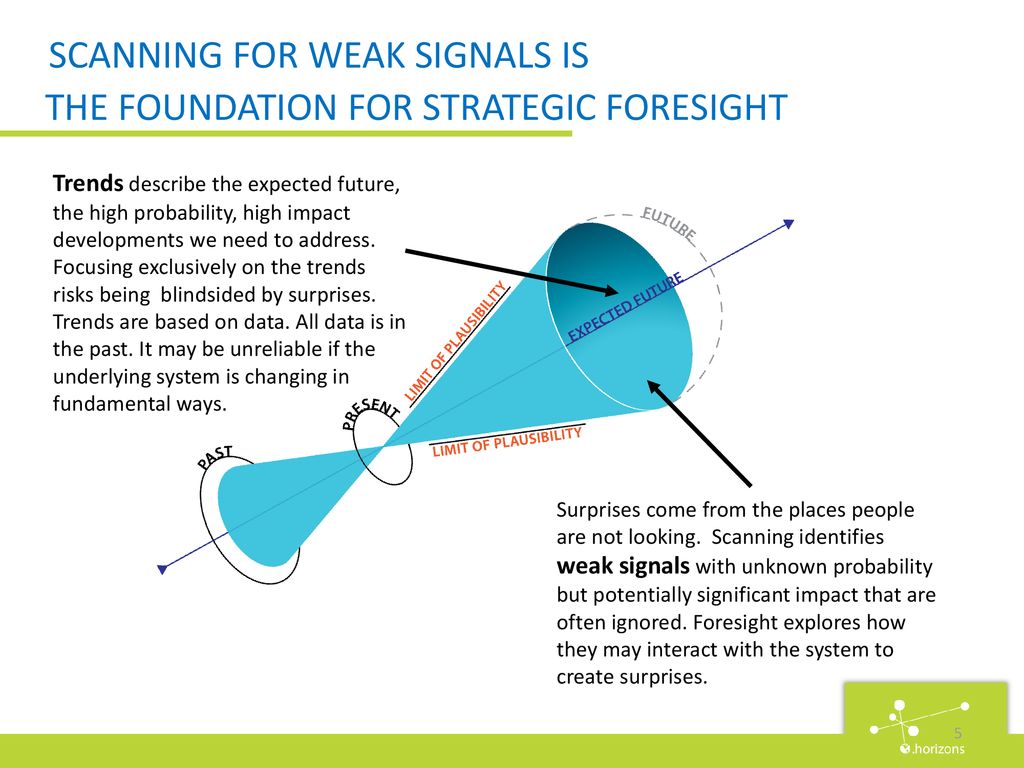
Our approach is to use Quantum NLP (QNLP) on a noisy intermediate-scale quantum (NISQ) hardware. We use a categorical compositional distributional semantics (DisCoCat) model of natural language (a tensor network language model), which combines two different approaches, categorial grammar (rule-based approach to language syntax) and distributional semantics (statistical approach). We use variational quantum circuits in the place of deep neural networks.

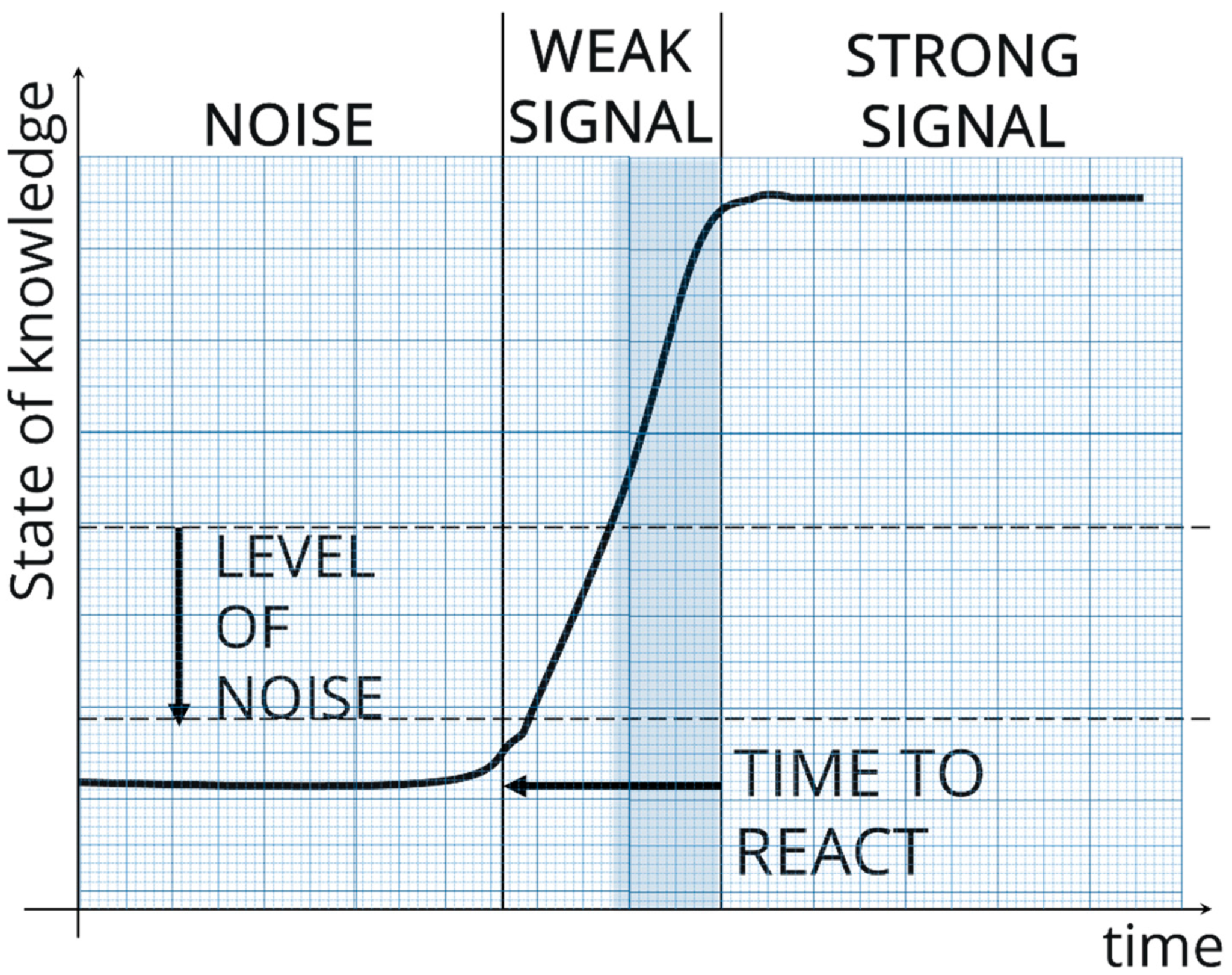
The system is based on a compositional grammar model (pregroup grammar), which is endowed with a distributional embedding of words into a d-dimensional vector space, where vector geometry captures the correlations between words according to some corpus. This allows the perfect match of grammatical structure in terms of pregroups and the compositional quantum structure of bipartite entanglement. In addition to vector spaces and inner-products, DisCoCat also includes several other quantum features, like projector spectra for representing meanings of adjectives, verbs and relative pronouns; density matrices for representing linguistic ambiguity and lexical entailment; and entanglement for representing correlated concepts, all of which exist “natively” on quantum hardware.

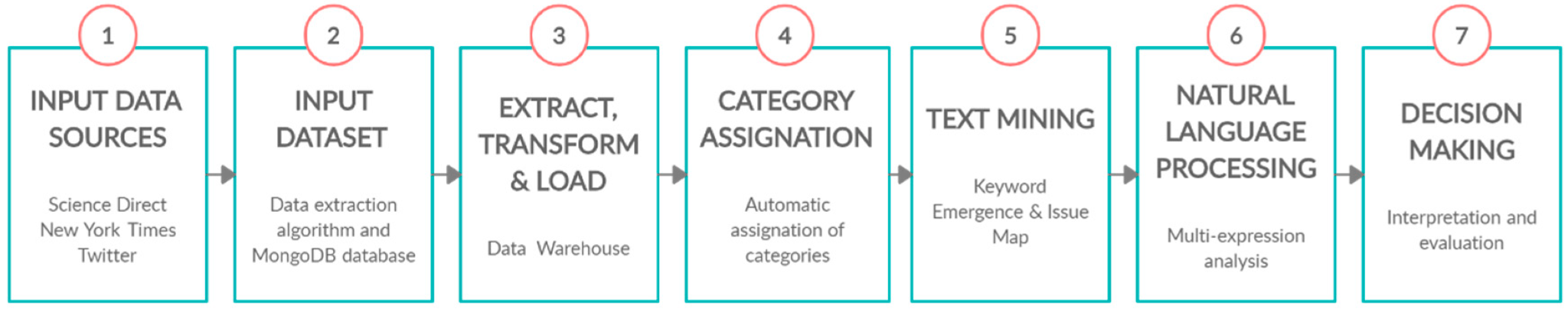
Using DisCoCat allows an exponential reduction of space resources as compared to implementations on classical hardware. Other direct benefits are the nativeness of density matrices, and the availability of quantum algorithms that provide an algorithmic quantum advantage for typical NLP tasks such as classification.

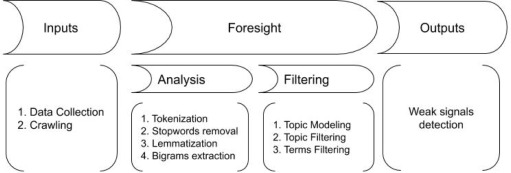
**4. Potential customers for this solution**

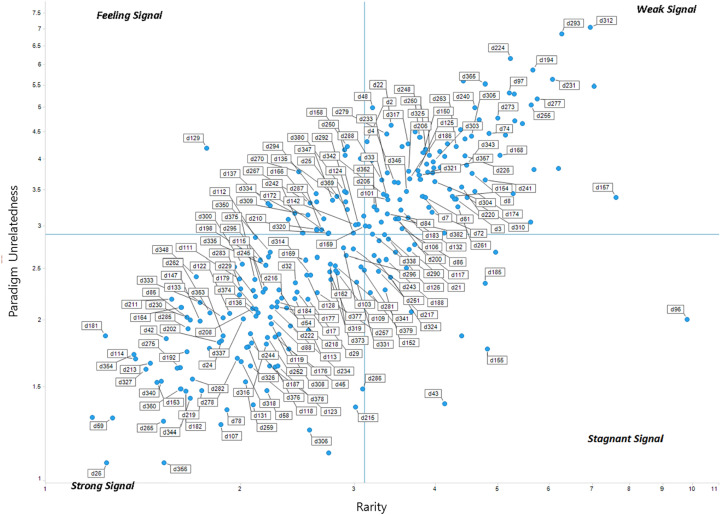
Customers will be investment firms such as Goldman Sachs, hedge funds such as Winton Capital, and business consultancy firms such as McKinsey.

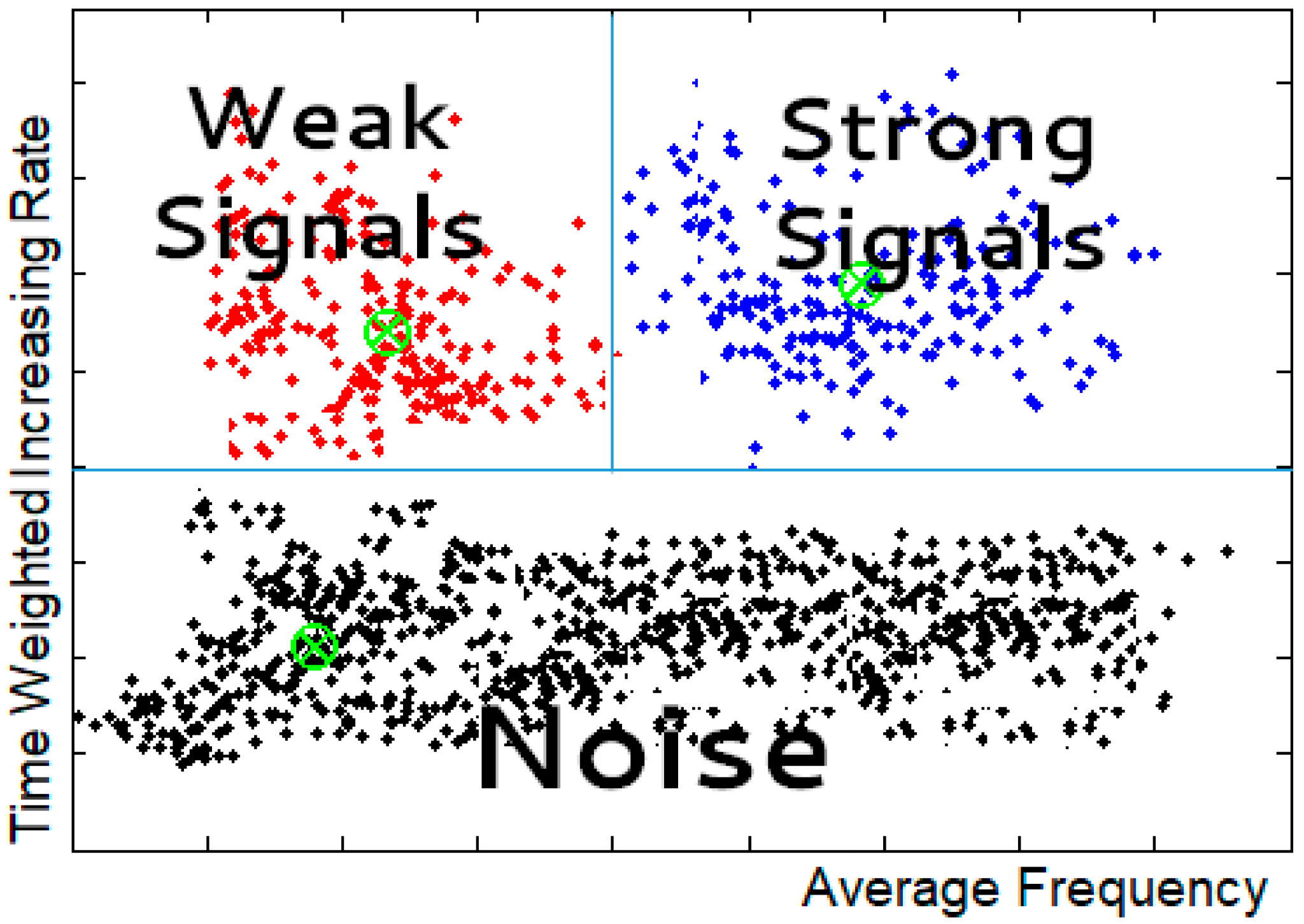












1. https://thequantumdaily.com/2020/12/10/meaning-aware-computers-cqc-researchers-make-major-nlp-advance-in-using-quantum-computers-to-understand-language-and-towards-achieving-meaningful-quantum-advantage/ [↑](#footnote-ref-0)