Project Plan	Merck
	Student Team
	20.05.2021

Basic Question of the Project

The goal of the project is:

1) To utilize QNLP (Quantum Natural Language Processing) to speed, scale, and make more accurate post market investigations of pharmaceutical drug effects.
2) To develop a prototype of this application that will add value to the R&D processes of Merck by the end of August

Merck is very open minded towards the result of this project, and is considering this project as an experiment. However, there are some angles where our relevant backgrounds, the domains Merck is interested in exploring, and opportunities with outside partners & academics coalesce where we think we would have a great opportunity for generating value. From iterative discussions between these groups, we have honed in on the Quantum Natural Language Processing (QNLP) space, specifically in pharmacovigilance and drug safety contexts (AE reaction studies in post-market users).

Perspective and Context

What are relevant industry trends?

- The skyrocketing amount of social media postings increases the effort for screening possible Adverse Events
- Healthcare/Pharmaceuticals are transitioning from reactive medicine to predictive medicine
- Efficiency in big data, AI, and computing allow for new opportunities to leverage software in healthcare
- Quantum computing advancements have great potential synergy in the pharmaceutical industry applications (high dimensionality of inputs, large calculation times for classical options (e.g. protein folding, patient scanning))
- Due to rapid increase in development costs in R&D, medicine is becoming more expensive.
- Future decades will lead to historical patent losses, leading to secure cash flows disappearing; new drugs need to be discovered.
- Many patents are nearing their expiry date which increases the pressure to invent new drugs. High rate of patent expirations means that new innovations are needed to maintain the advancement of the industry for dealing with the rapid development of new illnesses or strains
- Increased focus on prevention rather than treatment to speed up diagnosis by empowering doctors to leverage advanced technologies to become patient educators

What is the company's current situation?

- Merck is heavily invested in various emerging technologies to tackle computational chemistry problems as drug development becomes more resource-consuming and drugs become complicated in structure
- Merck is pushing quantum computing technologies by the formation of the Quantum Task Force and its investment in various quantum startups through its venture capital arm, M Ventures
- Reduced confidence in pharma industry due to high failure rate of drug trials
- Failure rates in preclinical and clinical stages are at 90% according to a study by McKinsey [1]

What challenges could be encountered in the QEL project?

- We are attempting to combine two new technologies (quantum computing and NLP) which are both in their infancy
- Our end product might not be possible to access sufficient resources
- No access to quantum hardware currently (Would have to work with a server based annealer/DWAVE,etc.)
- No success-proven valuable application that outperform the classical counterparts in the industry yet
- Problem of identifying quantum algorithm to our specific use case

What resources are available for developing the QEL project?

- Merck is willing to provide connections with industry experts that could provide key insights to our project such as Bob Coecke from University of Oxford who is an expert of QNLP (Quantum Natural Language Processing) and Harsha Gurulingappa, an expert in NLP
- Merck has lots of expertise in the quantum computing field and NLP
- DisCoPy can be used as an already existing toolbox
- Access to qiskit via IBM

Success Criteria and Targets

The success criteria is a tangible prototype with the following characteristics:

- Be able to be simulated in a demo environment
- Provide genuine benefit over comparable classical implementations (speed, dimensionality, etc).
- Potential to extrapolate from pharma/healthcare industry into other industries
- Realizable and technically feasible within the timeline of the class
- Realize a product that could have a positive NPV/payoff for Merck
- Ability to set up a foundation for future QNLP related projects at Merck
- Contribution to the emerging field of QNLP

Our target is to meet as many of the aforementioned criteria, but we are aware that due to the already mentioned difficulties, it may not be possible to fulfill all success criteria.

For the process, we have the following measurable objectives:

- Constructive collaboration between Merck and our team
- Efficient workflow and great team spirit within our team
- Open communication and insightful feedback from teaching assistants

Scope of the Project and Constraints within Solution Space

This project will examine the feasibility and scope of implementing QNLP solutions in classically dense and noisy healthcare processes. Our goal is to create a QNLP Adverse Effect (AE) product that will 1.) monitor adverse drug effects and 2). determine if any action is needed to remediate the adverse reactions/effects in patients in post market drug studies. If we can develop a successful proof of concept, this would lead to potential future extrapolation of using quantum technology in a variety of other pharma/healthcare subfields. As a proof of concept, we intend to build a "demo" deliverable that could be a litmus test for further commitment and investment from the parties involved (QEL, E-Lab Team, Merck). Our implementation is in regards to general benefits to the pharma/healthcare industry, i.e., the limitations of classical models to predict/analyze adverse effects become less of a technical issue, and more of a domain implementation and strategy decision (that Merck can decide or seek expertise with the E-Lab team on). The client (Merck) will also be deeply involved in the entire pipeline, leading to faster iteration and a more tailored and nuanced end product.

What is not in scope of this project:

- Review of patient data regulations and any other legal limitations
- Examination of QNLP for other pharmaceutical companies aside from Merck

Project Deliverables

The main deliverable is a prototype fulfilling the success criteria described in the section above. We will provide the deliverables based on the following timeline:

- 1st deliverable (Due date 21.05.2021): Ideation Deliverable (This project plan and a document explaining our idea)
- 2nd deliverable (Due date 11.06.2021): Mid-term Pitch
- 3rd deliverable (Due date 30.07.2021): Demo day pitch
- 4th deliverable (Due date 31.08.2021): Final deliverable

Stakeholders

- Single Point of Contact (SPoC):
    - Philipp Zent
- Team members:
    - Clarissa Anjani: Sales & Marketing Specialist (Mail: e.clarissa.anjani@gmail.com)
    - Friederike Butt: Quantum Specialist (Mail: friederike.butt@rwth.aachen.de)
    - Sebastian Soldner: Finance & Business Specialist (Mail: soldners121@gmail.com)
    - Kevin Shen: Quantum Specialist (Mail: kevin.shen@tum.de)
    - Shruti Pistolwala: Software Specialist (Mail: shruti.pistolwala@tum.de)
    - Philipp Zent: Finance & Management Specialist (philipp.zent@icloud.com)

- Key information resources:
    - Thomas Ehmer: Innovation Incubator Lead at Merck
    - Harsha Gurulingappa: Head of Text Analytics at Merck
    - Phillipp Harbach: Head of Silico Research at Merck
    - Insa Mohr: Associate director at Merck
    - Bob Coecke (And Team): Professor of Quantum Foundations, Logics and Structures at Oxford University

- Project sponsor:
    - Merck
    - Contact Persons: Thomas Ehmer and Insa Mohr

Communication and Decision Making

- Weekly team meeting on Friday at 5 pm
- Continuous alignments within the team via Whatsapp and ad-hoc meetings
- Weekly TA session on Thursday at 1:30 pm
- Meetings with Merck (Thomas Ehmer, Insa Mohr, and possibly other key information resources) on a regular basis to make decisions about goals, the direction etc. together with them
- Important decisions are made by consensus of the entire team

Signature Company	Signature Student Team QEL Team

Additional References

A Heid, I Ostovic, 'Recalculating the future of drug development with quantum computing', McKinsey & Company,2020,https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/recalculating-the-future-of-drug-development-with-quantum-computing

QEL Merck Challenge Ideation

Clarissa Anjani, Friederike Butt, Sebastian Soldner,
Kevin Shen, Shruti Pistolwala, Philipp Zent

May 2021

Contents

1 Focus of the project

The focus of the project is to utilize Quantum Natural Language Processing (QNLP) to 1.) monitor adverse drug effects and 2). determine if any action is needed to remediate the adverse reactions/effects in patients. [1].

1.1 What is the context of the project?

This course (QEL) is taught at the Technical University Munich in a joint teaching format with PushQuantum Munich and Tum Venture Labs, as well as industry partners. The goal of QEL is to examine various possible applications of quantum computing in the industry in a practical hands-on approach, by creating realistic and comprehensive business models in interdisciplinary teams. Our partner for this project is Merck, a healthcare/pharmaceuticals team, who we are exploring the sub-domain of QNLP with to see potential practical applications this technology would have in their drug development (DD) opportunities (by enhancing or highlighting AE events in the DD process by higher accuracy, greater speed, or investigative opportunities). As AE studies are often extremely complex and handle a large variety of factors, classic calculations of large datasets is limited in scope or unfeasible. As Quantum Technology is a field which has seen intense interest and investment from industry partners in recent years, we believe that we are for the first time at the intersection of a point where we could realistically enhance AE outcomes with quantum technology.

1.2 What is the scope of the project?

The goal is, to develop a prototype to solve specific problems faced by the health care and pharmaceutical industry. This project will examine the feasibility and scope of implementing QNLP solutions in classically dense and noisy healthcare processes. Our goal is to create a QNLP Adverse Effect (AE) product that will 1.) monitor adverse drug effects and 2). determine if any action is needed to remediate the adverse reactions/effects in patients in post market drug studies. If we can develop a successful proof of concept, this would lead to potential future extrapolation of using quantum technology in a variety of other pharma/healthcare subfields. As a proof of concept, we intend to build a "demo" deliverable that could be a litmus test for further commitment and investment from the parties involved (QEL, E-Lab Team, Merck). Our implementation is in regards to general benefits to the pharma/healthcare industry, i.e., the limitations of classical models to predict/analyze adverse effects become less of a technical issue, and more of a domain implementation and strategy decision (that Merck can decide or seek expertise with the E-Lab team on). The client (Merck) will also be deeply involved in the entire pipeline, leading to faster iteration and a more tailored and nuanced end product.

[1] An Adverse Effect (AE) is defined as a reaction to a pharmaceutical product and which does not necessarily have a causal relationship with a treatment [1]

2 Shortlist of Problems

Quantum Natural Language Processing (QNLP) has the potential to solve several problems faced by companies of healthcare / pharmaceutical industry like Merck. In the following, we focus on four specific cases in the context of analysing unstructured text data.

2.1 Identifying adverse reactions

Studying side effects of drugs is one important part in the healthcare and pharmaceutical industry, especially in the context of pharmacovigilance. In addition, our contact persons at Merck underlined the significance of this problem and suggested to focus on this use case. As healthcare companies like Merck develop drugs to deal with the world's most challenging diseases, one of the challenges is to make sure that patients are safe, while still benefiting from an effective drug. Before a company can address the patient's complaint, QNLP can analyze if the adverse reaction came from the respective drug or an unrelated event. As an example, a patient posts about his/her headache on social media or reports to the patient hotline, QNLP can then be used to support in identifying if this symptom is related to an adverse reaction of the patient's medication.

2.2 Education

Education is one of Merck's keystones [2]. An automated system that can evaluate if a given answer to some question is correct or incorrect can simplify the learning process and make it more accessible and flexible.

2.3 Second opinion

A researcher in the company wants to confirm their opinion or look for some specific method with an unknown terminology. A system that takes in a description and delivers the terminology of the desired method, would speed up research and make information more easily accessible.

2.4 Internal expert finder

Someone in the company needs expertise in a specific field. For example a marketing person needs to know about quantum computing, in order to adapt some marketing strategy. A process that could find the fitting expert for this task, would make the exchange of information more effective and could connect researchers to solve specific problems.

3 Shortlist of potential solutions

3.1 Classification of the task

3.1.1 Classical approach, Natural Language Processing (NLP)

This method can classically be implemented using Open AI GPT-3 (Generative Pre-trained Transformer) [3]. Using the API token in a chatbot or any text-based input system, we can develop a knowledge management system. In this system, a text based input will be given by a human and, using GPT-3 trained model, an answer would be generated. The model can be trained based on the inputs from doctors, organizations and many more. It helps solving issues related to the speedup of a diagnosis. Information can be acquired faster, as well as the proper diagnosis, and time can be utilized properly. The ideal case to get information is to search on Google which is not recommended in healthcare profession so issues related to wrong information can be solved using this technique [4].

3.1.2 Quantum Natural Language Processing (QNLP) approach

QNLP is based on diagrammatic reasoning: The quantum model interprets language as a quantum process by means of the diagrammatic representation in terms of DisCoCat and categorical quantum mechanics (CQM) [5]. Then, the resulting quantum representation can be translated into a quantum circuit form using the ZX-calculus [5]. Thereby, it becomes implementable on quantum hardware. So, a key idea is the translation of linguistic structures into quantum circuits by assigning meanings to words and letting them interact. Within this model, it is possible to formulate question-answering problems by looking for a word that maximizes some closest vector problem, e.g.

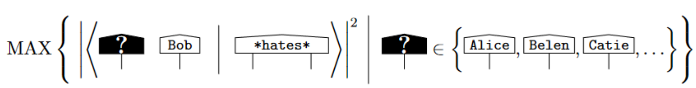

Figure 1: Question-answering by maximizing the inner product and computing the closest vector [5]. Here, one wants to find the answer to the question: 'Who hates Bob?'.

This grammar-aware question-answering has been realized on quantum computers using the toolbox DisCoPy [6]. If we, for example, replace 'Bob' in figure 3 with 'patient' and 'hates' with 'coughs', we could look for the answer to this question, in terms of a diagnosis, e.g. a 'cold'. This concept could be applied to the above explained problems in different variations.

General Message

One natural way to do QNLP is through variational circuits. Similar as in variational quantum eigensolver, we can first prepare a quantum circuit model whose gates are dependent on variable parameters $\{\beta_i\}$. We then train the model with classical optimization algorithms, or namely to prepare initial states, execute the circuit, do measurements, feed the results into cost functions, update parameters $\{\beta_i\}$ and iterate over and over until a certain precision is reached.

The advantage of this approach is that it is NISQ-friendly. It is perhaps more suitable to say that NISQ is NLP-friendly as NLP is claimed to be a "quantum native" problem. "What quantum theory and natural language share at a fundamental level is an interaction structure. This interaction structure, together with the specification of the spaces where the states live, determines the entire structure of processes of a theory. So the fact that quantum theory and natural language also share the use of vector spaces for describing states—albeit for very different reasons—makes those two theories (essentially) coincide."[5] This similarity between quantum theory and natural language makes it particularly convenient to encode words into quantum states.

From there we can benefit from the superposition and entanglement of quantum theory. "Whatever exponential space blow up we have due to the tensor product (on classical computers) would immediately vanish (on quantum computers)". This is saying that, ideally we would be able to train NLP models with an exponentially smaller amount of (qu)bits. It is estimated that with $50 - 100$ qubits, with the quantum method we will be able to solve problems that are beyond the capability of today's best classical computers.

	1 verb, nouns 2K D	10K verbs, nouns 2K D	10K verbs, nouns 1M D
Classical	8×10^9 bits	8×10^{13} bits	8×10^{22} bits
Quantum	33 qubits	47 qubits	73 qubits

Figure 2: Comparison of the number of (qu)bits required for NLP problems of various sizes [5]

3.1.3 State of the Art

Bob Coecke's group has performed several proof-of-concept experiments on their DisCoCat QNLP model on simulators and real quantum devices. This year, they tested their model with a dataset of size > 100 sentences on the quantum computer ibmq-bogota, which possesses 5 qubits and a quantum volume of 32. They obtained results with reasonable error rates and statistical significance [7].
Currently, a python tutorial notebook on DisCoPy is available on the internet, as well as an interactive platform [8] [9].
There are other researches on QNLP for far term applications. These are beyond the paradigm of NISQ and hence beyond the scope of our project. However, just for completeness we briefly mention some results here [10]. In the long run, it is shown that with fault-tolerant quantum computers with access to QRAM, there will be polynomial speedup to many NLP problems including but not limited to:

1. Kernal correspondence analysis (Word2vec), which is a multivariate sta-

tistical tool giving geometric representations of categorical random variables.

2. Latent semantic analysis, which compares words and documents for the purpose of topic identification, document retrieval and document indexing.
3. Slow feature analysis, which reduces the dimension of input sentences by projecting them onto a smaller subspace.

3.1.4 Application to specified problems

Identifying adverse reactions

By combing through user inputed data, we can generate a large dataset a model could train off of for identifying adverse effect reactions. We would assign weights to certain key words or phrases; the maximal overlap of some input sentence or statement with the category of an adverse event could be computed using the technical methods described method. For example, if a person 'has sore feet', one could look for an overlap with 'adverse event'.

Education

Training and testing understanding by typing ones answer to some test-question. QNLP could be used to check the correctness of the given answer.

Second opinion

To get a second opinion, a researcher in the biotech company could type in the attributes he wants to classify, for example a molecule that fulfills certain criteria. Then the researcher could check, if the given output aligns with his opinion.

Internal expert finder

Some areas of knowledge are assigned to internal experts within a company. One could look for the maximal overlap of e.g. a 'researcher who is an expert in quantum computing' and the names of employees of the company.

We plan to do a proof-of-concept experiment for our use cases first with $5 \sim 7$ qubits on either a quantum computer or a simulator, in combination with classical computation for certain subroutines, given the current limitation in quantum hardware resources. Later we would also experiment on a larger scale when quantum devices with more qubits become available.

3.2 Market potential

As part of defining the market potential, the total available market refers to the size of at least the pharmaceutical and healthcare industry, but it can potentially be extended further to other industries. We have also defined possible customers for our use case.

Total Available Market: Pharma/Healthcare Market Sizing

The global pharma/healthcare market is estimated at 8.45 trillion USD as of 2018. The global physician market is estimated at 1.43 trillion USD in 2020, growing at an estimated 5.3% CAGR up to 1.505 trillion USD in 2021 [11]. With regard to the top 27 OECD countries, the number of doctors visits per capita in 2018 was 7.19 a year [12]. With a global healthcare system that is growing at a larger rate than global GDP estimates, (3.3% and 3.4% respectively in 2020 and 2021) [13], there is foreseeable growth in healthcare and patient requirements [14]. In Europe, the healthcare analytics market is estimated at 2.68 billion USD in 2020, scaling to 8.68 billion USD by 2025 [15]. Areas that are seeing immense growth that will lead to benefits in pharma/healthcare includes big data and AI, aiding the transition of the global healthcare industry from reactive medicine to prescriptive medicine.

In addition to the global physician market, there are significant investments going into research and innovation in the pharmaceutical industry. An example is Horizon 2020, Europe's research and innovation program as part of addressing the pandemic. The European Commission has pledged over 1 billion €from Horizon 2020 to address the pandemic with a focus on topics ranging from global preparedness of outbreaks and digital technologies [16].

NLP Market Sizing

The global market size of NLP was at 10.2 billion USD in 2019 and is expected to grow to an estimated 26.4 billion USD by 2024, which is equivalent to a compound annual growth rate (CAGR) of 21.0% for the period of 2019-2024 . [17]. A major part of the growth is assigned to the increased investments from the pharma/healthcare industry.
Examining the pharma/healthcare industry's share in NLP, the CAGR is estimated even slightly higher at 21.02% with a market size of 1.12 billion USD in 2016 and an estimated 6.24 billion USD by 2025 [18].

Adverse reactions in NLP Market Sizing

There are many weak factors influencing the potential size of this market that are hard to quantify. Within the pharma/healthcare industry, this application would rather be an add-on to existing markets. Figure 2 demonstrates the huge potential of AI-enabled solutions in the drug discovery and clinical trials market. The market potential for our application would be a certain percentage of the market sizes depicted in Figure 2. However, we want to emphasize, that our use case is likely to be applicable in other industries as well and therefore promises enormous overall potential.

Figure 3: "The solution segment of the market is expected to grow to 3,385.0 million USD by 2030 with the CAGR of 24.48%." [19]

Target Customer: Identifying Adverse Reactions

All pharmaceutical and healthcare companies which are involved in drug safety can be considered as target customers. In the context of pharmacovigilance, QNLP can be used as part of monitoring adverse effects that were not discovered during clinical trials, research, and other activities before the drug's market launch. Our target consumers would benefit from a continuous tracking to identify the direct correlation of the drug to a patient's complaint on social media platforms or patient hotlines. In addition, physicians and individual consumers could be identified as potential target customers.

3.3 Value proposition

Based on the choice from our shortlist of problems, we have come up with the following value proposition:

"We help Merck (and the pharma/healthcare industry leading experts to) speed up AE discovery, scale dimensionality in AE inputs, and uncover hidden insights by developing a QNLP AE solution such that we detect potential adverse event signals in post market investigation data."

References

[1] *Adverse event*, https://en.wikipedia.org/wiki/Adverse_event, Accessed: 2021-05-20.

[2] *Education at merck*, https://www.merckgroup.com/ro-ro/company/responsibility/our-strategy/providing-young-people-with-access-to-education-and-culture.html, Accessed: 2021-05-17.

[3] *Openai*, https:https://openai.com/, Accessed: 2021-05-18.

[4] B. 2021, *Babylon health*, https://www.babylonhealth.com/, Accessed: 2021-05-18.

[5] B. Coecke, G. de Felice, K. Meichanetzidis, and A. Toumi, *Foundations for near-term quantum natural language processing* (2020).

[6] K. Meichanetzidis, A. Toumi, G. de Felice, and B. Coecke, *Grammar-aware question-answering on quantum computers* (2020).

[7] R. Lorenz, *Qnlp in practice: running compositional models of meaning on a quantum computer*, Accessed: 2021-05-18 (2021).

[8] *Python notebook*, https://discopy.readthedocs.io/en/main/notebooks/qnlp-tutorial.html.

[9] *Platform*, https://qnlp.cambridgequantum.com/generate.html?fbclid=IwAR3TYeOZtXeRQ6UPYuDWAkUBHnlmief5-1dfsr8G1J1uYvIOfEiY88qx_t8.

[10] A. Luongo and A. Bellante, *Quantum algorithms for nlp: lsa, qsfa and ca. talk at qnlp ii, quantum week of fun*, https://www.youtube.com/watch?v=d-Lfdfy-xw", 2020.

[11] *The state of healthcare industry – statistics for 2021*, https://policyadvice.net/insurance/insights/healthcare-statistics/, Accessed: 2021-05-18.

[12] *Physicians and other health practitioners global market report 2021: covid-19 impact and recovery to 2030*, https://www.thebusinessresearchcompany.com/report/physicians-and-other-health-practitioners-global-market-report-2020-30-covid-19-impact-and-recovery:~:text=The\%20global\%20physicians\%20and\%20other,(CAGR)\%20of\%205.3\%25.andtext=The\%20market\%20is\%20expected\%20to,at\%20a\%20CAGR\%20of\%207\%25., Accessed: 2021-05-18.

[13] *Number of doctor visits per capita in selected countries 2018*, https://www.statista.com/statistics/236589/number-of-doctor-visits-per-capita-by-country/, Accessed: 2021-05-18.

[14] *World economic outlook reports*, https://www.imf.org/Errors/500.aspx, Accessed: 2021-05-18.

[15] *Europe healthcare analytics market research report – segmented by type, application, end user, country (uk, france, spain, germany, italy, russia, sweden, denmark, switzerland, netherlands, turkey, czech republic and rest of eu) - size, share, growth, trends and forecast — 2020 to 2025*, https://www.marketdataforecast.com/market-reports/europe-healthcare-analytics-market, Accessed: 2021-05-18.

[16] M. Gabriel, *Eu research and innovation in action against the coronavirus: funding, results and impact*, https://ec.europa.eu/info/sites/default/files/research_and_innovation/research_by_area/documents/ec_rtd_eu-research-innovation-against-covid.pdf/, Accessed: 2021-05-18.

[17] *The global natural language processing (nlp) market size*, :https://www.businesswire.com/news/home/20191230005197/en/Global-Natural-Language-Processing-NLP-Market-Size, Accessed: 2021-05-18.

[18] *Natural language processing (nlp) in healthcare and life sciences market 2021*, https://www.marketwatch.com/press-release/natural-language-processing-nlp-in-healthcare-and-life-sciences-market-2021-worldwide-opportunities-driving-forces-covid-19-impact-analysis-future-potential-2025-2021-03-31, Accessed: 2021-05-18.

[19] *Secondary research, expert views, and bis research analysis*, Provided by Merck: 20.05.2021.