



Manuel Burghardt
Lisa Dieckmann
Timo Steyer
Peer Trilcke
Niels-Oliver Walkowski
Joëlle Weis
Ulrike Wuttke (Hg.)

Fabrikation von Erkenntnis

Experimente in den Digital Humanities



**Quantencomputing in den Digital Humanities:
innovativ oder übertrieben?**
Experimente mit Quanten-basiertem Maschinellen Lernen

Johanna Barzen
Frank Leymann

Fabrikation von Erkenntnis: Experimente in den Digital Humanities
Herausgegeben von Manuel Burghardt, Lisa Dieckmann, Timo Steyer, Peer Trilcke,
Niels-Oliver Walkowski, Joëlle Weis, Ulrike Wuttke



Melusina Press 2021

Publiziert im Jahr 2021 durch **Melusina Press**
11, Porte des Sciences
L-4366 Esch-sur-Alzette
<https://www.melusinapress.lu>
Melusina Press ist eine Initiative der Universität Luxemburg.

Konzept: Niels-Oliver Walkowski, Johannes Pause
Lektorat: Carolyn Knaup, Niels-Oliver Walkowski
Cover: Valentin Henning, Erik Seitz

Die digitale Version dieser Publikation steht frei zur Verfügung unter
<https://www.melusinapress.lu>.

ISBN (Online): 978-2-919815-25-8
DOI (Publikation): 10.26298/melusina.8f8w-y749-qidd
DOI (Band): 10.26298/melusina.8f8w-y749

Das vorliegende Werk steht unter einer CC BY-SA 4.0 Lizenz. Informationen zu dieser Lizenz finden Sie unter <https://creativecommons.org/licenses/by-sa/4.0/deed.de>. Die in diesem Werk enthaltenen Bilder und Ressourcen unterliegen der selben Lizenz, sofern sie keiner anderen Quelle entnommen oder mit einer anderen Lizenz versehen sind.



1. Einleitung

Durch das Aufkommen erster allgemein anwendbarer und über Cloud-Zugänge verfügbarer Quantencomputer¹ stellt sich die Frage, wo und wie diese neue Technologie auch für Fragestellungen aus den Digital Humanities eingesetzt werden kann. Seitdem die ersten kommerziell verfügbaren Quantencomputer auf dem Markt sind, entwickeln sich diese rasch weiter und sowohl für Forschung wie auch für Anwendungsmöglichkeiten in der Industrie werden Quantencomputern eine Bandbreite an Potenzialen zugesprochen und umwälzende Neuerungen durch deren Einsatz erwartet.² Anbieter wie IBM erwarten bis 2023 beispielsweise Quantencomputer mit mehr als 1000 Qubits zur Verfügung stellen zu können,³ Fraunhofer stellt seit Anfang 2021 eine umfassende Forschungsinfrastruktur mit eigenem Quantencomputer bereit,⁴ welcher dem europäischen Datenschutz unterliegt, und das aktuelle Bundeskonjunkturprogramm sieht Quantencomputer als eines von drei Zukunftsthemen für Forschung und Industrie, in das signifikant investiert wird.⁵

Die Frage ist also nicht mehr, ob diese neue Art von Rechner Realität wird, sondern wie man mit den gegebenen Schwierigkeiten, die diese Technologie noch hat, bereits erstes Anwendungswissen aufbauen kann, (i) um zu überprüfen, ob die erwarteten Potenziale (siehe Abschnitt 3) auch für Forschung in den Geisteswissenschaften nutzbar sind, (ii) für welche Fragestellungen sich der Einsatz lohnt und, (iii) wenn es sich als fruchtbar erweist, wie erste wiederverwendbare Komponenten zur Verfügung gestellt werden können, welche weiterführende komplexere Anwendungen erlauben. Für die Nutzung des Quantencomputers – über den Einsatz des klassischen Computers hinaus –, um existierende und neue Fragestellungen aus den Digital Humanities zu adressieren, haben wir den Begriff der *Quantum Humanities* geprägt.⁶ Warum braucht es hierfür einen neuen Begriff? Dafür spricht im Besonderen die so anders geartete Handhabung von Quantencomputern, die sich aus den zugrundeliegenden Annahmen der Quantenmechanik und deren technischen Umsetzungen ergibt. Was das Verständnis zur Nutzung dieser Technologie beispielsweise erschwert, sind die zugrundeliegenden physikalischen Phänomene wie die Verschränkung der Qubits oder die Tatsache, dass das Kopieren von Zuständen nicht möglich ist, so dass klassische Ansätze nicht einfach auf das Quantencomputing überführt werden können (eine Einführung in das Quantencomputing findet sich in Abschnitt 2.1). Aber auch deshalb, weil mittels Quantencomputern Problemklassen adressiert werden können, die bisher praktisch nicht lösbar oder sogar prinzipiell mit klassischen Computern nicht lösbar sind,⁷ gilt es deren Potenzial auszuloten. Daher die provokante Frage: Ist die Nutzung des Quantencomputers für die Digital Humanities innovativ oder übertrieben?

1 Vgl. Karalekas et al. 2020; Leymann et al. 2020.

2 Vgl. National Academies of Sciences, Engineering, and Medicine 2019, S. 51–55.

3 Vgl. Gambetta 2020.

4 Fraunhofer-Kompetenznetzwerk Quantencomputing; Kompetenzzentrum Quantencomputing Baden-Württemberg.

5 Vgl. Bundesfinanzministerium 2020, S. 11f.

6 Vgl. Barzen / Leymann 2019; Barzen / Leymann 2020; *Quantum Humanities*.

7 Vgl. Nielsen / Chuang 2010, S. 40–42; Raz / Tal 2019.

Das Arbeiten mit Quantencomputern ist aktuell noch stark experimenteller Natur und für Nutzer ohne mathematisches und physikalisches Fachwissen oft nicht intuitiv erschließbar. Um diese Einstiegshürden abzubauen und sich der Beantwortung des aufgespannten Spektrums bestehender Fragen zu nähern, ist unsere Werkzeugumgebung *QHAna* (Quantum Humanities Analysis Tool) entstanden.⁸

Im Folgenden werden in Abschnitt 2 die Kernkonzepte des Quantencomputings aufgezeigt. Da Quantencomputer im Besondern geeignet sind, Optimierungsprobleme zu lösen, ist der oft auf solche Probleme zu reduzierende Bereich des Maschinellen Lernens für den Einsatz des Quantencomputers als besonders zielführend einzuschätzen⁹ – ein Bereich mit einem Methoden- und Werkzeugarsenal, dessen Anwendung sich auch in unterschiedlichen Projekten der Digital Humanities wiederfinden lässt. Abschnitt 3 widmet sich daher den Potenzialen des Quantencomputers für die Digital Humanities. Da Quanten-basiertes Maschinellen Lernens als vielversprechend einzuschätzen ist, ist auch unser Anwendungsfall als Grundlage der Experimente in Abschnitt 4 in diesem Bereich angesiedelt. Das verfolgte Ziel der Experimente ist, auf früheren ersten Ergebnissen aufbauend,¹⁰ diese in größerem Maßstab – sowohl in der Anwendung auf größere Datensätze als auch durch die Erweiterung der Bandbreite der zur Verfügung stehenden Algorithmen – zu evaluieren und zu verifizieren, um die Nutzbarkeit für mögliche Quantum Humanities und die Angemessenheit der Lösungen auszuloten.

2. Quantencomputing

In den letzten Jahren hat sich das Quantencomputing von einer stark theoretischen Disziplin hin zu einem lebendigen und aufstrebenden Forschungsfeld mit praktischer Anwendung entwickelt, dessen zugesprochene Bedeutsamkeit sich in Industrie und Wissenschaft mehr und mehr etabliert und dessen Bandbreite potenzieller Anwendungen von den Grundlagen der Physik und Informatik bis hin zu Anwendungen in den Bereichen Technik, Finanzen, Chemie und Optimierungsprobleme reicht.¹¹ Um sich dem Quanten-basierten Maschinellen Lernen als potenziell vielversprechendes Anwendungsfeld innerhalb möglicher Quantum Humanities anzunähern, sollen im Folgenden die Kernkonzepte des Quantencomputings knapp dargestellt werden.¹²

2.1 Grundlagen Quantencomputing

Die Quantenphysik, als Grundlage für das Quantencomputing, unterscheidet sich grundlegend von der klassischen Physik durch ihren nicht-deterministischen Charakter

8 Vgl. Barzen 2021, S. 35–46.

9 Vgl. Schuld / Petruccione 2018; Dunjko et al. 2016.

10 Vgl. Barzen 2021, S. 41–45.

11 Vgl. Meichanetzidis et al. 2020, S. 1.

12 Da dies nur ein erster Einblick in ein äußerst komplexes Thema sein kann, findet sich bei Bedarf eine umfassende Einführung bei: Nielsen / Chuang 2010; Rieffel / Polak 2011.

und die Tatsache, dass Messen eine beeinflussende Auswirkung auf den zu messenden Zustand hat. So kann man etwa prinzipiell nicht gleichzeitig den Ort und die Geschwindigkeit eines (mikroskopischen) Teilchens exakt bestimmen (das ist die wohlbekannte Heisenbergsche Unschärferelation), und die möglichst exakte Messung einer dieser Größen führt zur sehr großen Ungenauigkeit der Messung der anderen Größe.

Übertragen auf den Bereich des Quantencomputings bedeutet dies, dass der Zustand eines Bits, der im klassischen Fall immer genau entweder 0 oder 1 ist, in der Quantenphysik prinzipiell nicht mehr genau bekannt ist. D.h. der Zustand eines solchen Bits, das man nun treffend *Quantenbit* oder kurz *Qubit* nennt, ist gleichzeitig 0 und 1. Genauer formuliert ist der Zustand eines Qubits eine Überlagerung oder *Superposition* der Zustände 0 und 1. Ein Qubit hat dann nicht mehr nur die beiden möglichen Werte 0 oder 1, sondern unendlich viele mögliche Werte, nämlich alle Werte, die sich aus der Kombination der Zustände $|0\rangle$ und $|1\rangle$ ergeben (wobei man nun statt 0 und 1 in der Quantenwelt die sogenannte Dirac-Notation $|0\rangle$ und $|1\rangle$ verwendet). Mathematisch sind $|0\rangle$ und $|1\rangle$ orthogonale Basisvektoren des zweidimensionalen komplexen Vektorraums der Qubits, d.h. ein Qubit lässt sich in dieser Basis als $\alpha|0\rangle + \beta|1\rangle$ darstellen (sogenannte Linearkombination), wobei α und β komplexe Zahlen sind: so entstehen die erwähnten unendlich vielen möglichen Kombinationen aus $|0\rangle$ und $|1\rangle$.

Um am Ende einer Berechnung mit Qubits das Ergebnis der Berechnung zu erfahren, muss eine Messung der Qubits durchgeführt werden. Diese Messung beeinflusst nun aber den aktuellen Zustand $\alpha|0\rangle + \beta|1\rangle$ des Qubits und es ‚kollabiert‘ durch die Messung in einen der Zustände $|0\rangle$ oder $|1\rangle$, so dass die Messung dann die klassischen Werte 0 oder 1 liefert. Die Wahrscheinlichkeit, mit der das Qubit in den Zustand $|0\rangle$ bzw. $|1\rangle$ übergeht, beträgt $|\alpha|^2$ bzw. $|\beta|^2$ (sogenannte Bornsche Regel). Da einer diese Zustände mit Sicherheit gemessen wird, summieren sich diese Wahrscheinlichkeiten zu Eins auf: $|\alpha|^2 + |\beta|^2 = 1$, d.h. ein Qubit ist ein Vektor der Länge 1.

Nun verarbeitet man im Allgemeinen nicht einzelne Qubits, sondern Gruppen davon, die man *Quantenregister* nennt. Jedes einzelne dieser Qubits ist in einer Superposition von $|0\rangle$ und $|1\rangle$, d.h. ein Quantenregister mit n Qubits ist dann in einem Zustand der Superposition der resultierenden 2^n möglichen Kombinationen $|00\dots 0\rangle$, $|10\dots 0\rangle$ bis $|11\dots 1\rangle$. Beispielsweise befindet sich ein Quantenregister mit 50 Qubits in Superposition von $2^{50} > 10^{15}$ dieser Werte. Im Vergleich dazu hat ein klassisches Register lediglich einen einzelnen Wert. Diese exponentiell vielen Werte eines Quantenregisters können mit jeder einzelnen Operation auf dem Quantenregister gleichzeitig manipuliert werden, ein Phänomen, welches *Quantenparallelität* genannt wird. Dieses Phänomen ist eine der Ursachen für die Überlegenheit des Quantencomputings für die Lösung einer Reihe von Problemen (*Quantenvorteil*).

Ein weiteres quantenphysikalisches Phänomen ist zentral für das Quantencomputing: die *Verschränkung* von Zuständen. Durch Verschränkung werden die Zustände der Qubits in einem Quantenregister korreliert, sodass deren gemeinsame Manipulation im Sinne der Quantenparallelität erst ermöglicht wird. Verschränkung ist einzigartig für das Quantencomputing und dessen Effekt kann auf klassischen Rechnern nur eingeschränkt nachvollzogen werden. Somit ist die Verschränkung eine weitere Ursache für den Quantenvorteil.

Eine Operation eines Quantencomputers überführt einen Zustand des Quantenregisters, d.h. einen Vektor der Länge 1, in einen anderen Zustand, also wieder einen Vektor der Länge 1. Solche längenerhaltenden Operationen werden *unitär* genannt. Ein Quantenalgorithmus ist eine Kombination solcher unitären Operationen. Jede einzelne Operation eines Quantenalgorithmus wird durch ein sogenanntes *Gatter* repräsentiert, und der Quantenalgorithmus als Kombination solcher Gatter wird als *Quantenschaltkreis* bezeichnet. Das Resultat der Ausführung eines solchen Quantenschaltkreises ist ein Quantenzustand, der eine Superposition möglicher Lösungen darstellt. Dieser Ergebniszustand muss gemessen werden, um das eigentliche Ergebnis der Berechnung zu erhalten. Durch das Messen kollabiert der Ergebniszustand zu einer in der Superposition enthaltenen möglichen Lösung mit einer bestimmten Wahrscheinlichkeit. Daher muss ein Quantenalgorithmus stets mehrfach ausgeführt und der Ergebniszustand gemessen werden, um eine Wahrscheinlichkeitsverteilung aller möglichen Lösungen zu erhalten. Diejenige Lösung, die dann am häufigsten auftritt, wird als ‚die‘ Lösung betrachtet. In diesem Sinne ist die Berechnung mit Hilfe eines Quantenschaltkreises als Zufallsexperiment aufzufassen.

2.2 Aktuelle Herausforderungen

Ein Quantencomputer ist kein abgeschlossenes System, sondern steht prinzipiell immer in Wechselwirkung mit seiner Umgebung. Dadurch wird der Zustand eines Qubits gestört. Dieses Phänomen wird als *Dekohärenz* bezeichnet, die entsprechende Zeit heißt *Dekohärenzzeit*. So kann beispielsweise der Zustand $|0\rangle$ nach einer bestimmten Zeit spontan in den Zustand $|1\rangle$ übergehen. Die Berechnung eines Quantenalgorithmus muss daher innerhalb der Dekohärenzzeit beendet sein, damit das Ergebnis möglichst geringe Fehler aufweist. Dies beschränkt gegenwärtig die Algorithmen, die auf heutigen Quantencomputern hinreichend korrekt ausgeführt werden können.¹³

Ebenso wie Qubits wird auch die Ausführung eines Gatters durch Wechselwirkung mit der Umgebung gestört. Somit ist das Ergebnis eines Gatters stets mit einem kleinen Fehler behaftet. Diese kleinen Fehler summieren sich im Laufe der Ausführung des Algorithmus auf. Um diese akkumulierten Fehler zu beschränken, sollte ein Algorithmus ‚wenig‘ Gatter enthalten: eine weitere Einschränkung derjenigen Algorithmen, die auf heutigen Quantencomputern ausgeführt werden können.¹⁴

Allerdings gibt es die Möglichkeit der Korrektur solcher Fehler. So können etwa mehrere Qubits derart gruppiert werden, dass Fehler, die in einer solchen Gruppe auftreten, erkannt und behoben werden können. Eine solche Gruppe von Qubits wird als *logisches Qubit* bezeichnet. Da mehrere physikalische Qubits ein einziges logisches Qubit realisieren, wird die Anzahl der Qubits, die für die eigentlichen Berechnungen zu Verfügung stehen, signifikant reduziert.¹⁵ Die Anzahl der für eine Berechnung verfügbaren Qubits ist aber wesentlich, um die Vorteile des Quantencomputings zu erreichen.

13 Vgl. Nielsen / Chuang 2010, S. 278f.

14 Vgl. Leymann / Barzen 2020, S. 13f.

15 Vgl. Rieffel / Polak 2011, S. 245f.

Somit werden heutzutage keine Methoden der Fehlerkorrektur angewendet, sprich: man versucht Vorteile trotz dieser Fehleranfälligkeit zu erreichen. Diese fehleranfälligen Quantenrechner mit nur einer geringen Anzahl an zur Verfügung stehenden Qubits nennt man NISQ (Noisy Intermediate-Scale Quantum)-Maschinen.¹⁶

Das impliziert, dass Algorithmen nur vergleichsweise ‚kurz‘ einen Quantencomputer verwenden können. Durch ‚geschicktes‘ Aufteilen eines Algorithmus in einen Teil, der auf einem klassischen Rechner ausgeführt wird, und in einen Teil, der ‚kurz‘ auf einem Quantencomputer läuft, wird dieses Ziel erreicht. Solche *hybriden* Algorithmen stehen gegenwärtig im Fokus der Anwendungen.¹⁷ Das Identifizieren von Teilen einer Problemlösung, die vom Quantencomputer profitieren, ist als nicht-trivial einzuschätzen.

Für den Bereich des Maschinellen Lernens sind eine Reihe solcher hybriden Algorithmen vorgeschlagen worden – man spricht dann vom Quanten-basierten Maschinellen Lernen oder Quanten Maschinellen Lernen (QML). Die Anwendungsbereiche, in denen diese QML-Algorithmen erfolgreich verwendet werden können (d.h. in denen Ergebnisse beispielsweise schneller oder mit höher Präzision erzielt werden), müssen identifiziert werden. Unser Fokus ist, entsprechendes Anwendungswissen für die Digital Humanities zu Verfügung zu stellen.

2.3 Quanten-basiertes Maschinelles Lernen

Viele Algorithmen des Maschinellen Lernens lassen sich auf Optimierungsprobleme reduzieren. Beispielsweise kann man Clustering als Suche sogenannter maximaler Schnitte auffassen (der Algorithmus wird als *Maximum Cut* bezeichnet) und deren Bestimmung wiederum auf die Lösung eines quadratischen binären Optimierungsproblems zurückführen.¹⁸ Für solche Optimierungsprobleme sind effiziente (hybride) Quantenalgorithmen bekannt.¹⁹ Eine andere Betrachtung führt die Suche maximaler Schnitte auf die Bestimmung von Eigenwerten zurück, für die es (hybride) Quantenalgorithmen gibt, die exponentiell schneller sind als die besten bekannten klassischen Algorithmen.²⁰ Somit lassen sich also bestimmte Probleme des Maschinellen Lernens mit Quantencomputern schneller lösen.

Ebenso können Vorteile in der Genauigkeit der Ergebnisse erzielt werden. Im Bereich der Klassifizierung durch Support Vektor Maschinen (SVM) etwa werden oft sogenannte Kern-Methoden angewendet.²¹ Hierbei werden die zu klassifizierenden Datenpunkte in einen extrem hochdimensionalen Raum abgebildet, in dem sie leicht zu trennen, also zu klassifizieren sind. Hierzu müssen aber Berechnungen im hochdimensionalen Raum durchgeführt werden (Skalarprodukte sind zu bestimmen), was sehr ineffizient ist. Diese ineffizienten Berechnungen lassen sich aber oft auf Berechnungen im

16 Vgl. Preskill 2018.

17 Vgl. Leymann / Barzen 2020, S. 1f.

18 Vgl. Barzen 2021, S. 31–33.

19 Vgl. Farhi et al. 2014; Crooks 2018.

20 Vgl. Poljak / Rendl 1995; Peruzzo et al. 2014.

21 Vgl. Skiena 2017, S. 371

ursprünglichen (niederdimensionalen) Raum durch spezielle Funktionen (Kerne; englisch: Kernel) effizient zurückführen. Solche Funktionen zu finden ist schwierig, aber im Quantencomputer lassen sich Berechnungen im hochdimensionalen Raum sehr effizient ausführen, um solche Kerne zu bestimmen oder anzunähern. Diese im Quantencomputer berechneten Kerne können Klassifizierungen erzielen, die präziser sind als mit klassischen Methoden durchgeführte.²²

Hieran lässt sich ein allgemeiner Vorteil des Quantencomputings für das Quantenmaschinelle Lernen erkennen: Da die Zustände in einem Quantenregister mit n Qubits Vektoren in einem n -dimensionalen Raum sind, lassen sich Daten mit sehr vielen Merkmalen — sprich: Vektoren mit sehr vielen Dimensionen — in einem Quantencomputer in natürlicher Weise darstellen. Solche hochdimensionalen Daten sind für klassische Algorithmen aber ein großes Problem, weshalb hier der Quantencomputer inhärente Vorteile bietet.

Allerdings ist das Laden von Daten in einen Quantencomputer, also die Transformation von Daten in einen Quantenzustand, der von einem Quantencomputer verarbeitet werden kann, ein Problem an sich. Hierzu sind separate Quantenalgorithmen notwendig, die vor der Ausführung des eigentlichen Quantenalgorithmus auszuführen sind. Diese zustandserzeugenden Algorithmen können in einigen Fällen die potenziellen Quantenvorteile des eigentlichen Algorithmus wieder zunichte machen.²³ Außerdem müssen nun weiter Gatter ausgeführt werden, sodass die gesamte Ausführungszeit und somit die akkumulierten Fehler ansteigen. Auch benötigen manche dieser Algorithmen zusätzliche Qubits, sodass für den eigentlichen Algorithmus weniger Qubits zu Verfügung stehen. Somit ist nicht einfach ersichtlich, ob eine QML-Lösung sinnvoll ist oder nicht. Es gilt also in vielen Fällen die theoretisch erreichte Verbesserung auch praktisch zu überprüfen.

3. Quantum Humanities als Chance

Anders als der Ansatz der *Quantum Social Science*,²⁴ bei welchem auf die Formalisierung von Informationsverarbeitung in sozialen Systemen, speziell Ökonomie und Finanzwesen, mittels der Quantenmechanik fokussiert wird, ist unser Ziel die praktische Anwendung von Quantencomputern. Denn trotz der gegebenen Schwierigkeiten (siehe Abschnitt 2.2) verspricht der Einsatz des Quantencomputers unter bestimmten Voraussetzungen und für bestimmte Anwendungsfälle einiges an Potenzial, was den schon genannten ‚Quantenvorteil‘ komponiert: nämlich (i) eine (exponentielle) Beschleunigung von Algorithmen²⁵ wie bei der unstrukturierten Suche²⁶ oder der Eigenwertbestimmung für die Hauptkomponentenanalyse²⁷ bei der Dimensionsreduktion von Da-

22 Vgl. Sierra-Sosa et al. 2020; Havlicek et al. 2018.

23 Vgl. Leymann / Barzen 2020, S. 3–8.

24 Haven / Khrennikov 2013.

25 Vgl. Rønnow et al. 2014; Shor 1997.

26 Vgl. Grover 1996.

27 Vgl. Abrams / Lloyd 1997.

tenräumen, (ii) die Verarbeitung großer Datenmengen, (iii) Präzision der Ergebnisse wie etwa bei der Klassifizierung,²⁸ (iv) die Lösung praktisch bisher unlösbarer Probleme,²⁹ (v) die Lösung von klassisch vorher prinzipiell nicht lösbarer Problemkassen (etwa der Erzeugung echter Zufallszahlen),³⁰ (vi) einen kostengünstigen Zugang im Verhältnis zu der ermöglichten Rechenleistung³¹ und (vii) Energieeffizienz.³²

Einiges davon ist noch stark theoretischer Natur und gilt für den ‚idealen‘ Quantencomputer. Wie in Abschnitt 2.2 verdeutlicht, ist dies technisch nicht der Stand, den wir heute haben, und noch sind die theoretischen Potenziale teils nicht ausschöpfbar, so beispielsweise bei Potenzial (vi) ‚kostengünstiger Zugang‘, wo, sobald es über die kostenlos zur Verfügung gestellten Testumgebungen hinausgeht,³³ aktuell noch auf teure prototypische Infrastruktur zurückgegriffen werden muss,³⁴ oder auch Potenzial (vii) ‚Energieeffizienz‘, bei dem die aktuell nötige extreme Kühlung von supraleitenden Qubits³⁵ der theoretischen Energieersparnis bei der eigentlichen Berechnung entgegen wirkt. Allerdings werden die verfügbaren Quantencomputer kontinuierlich besser, mit zunehmender Anzahl von Qubits und ausgereifteren Programmiermöglichkeiten.³⁶ Es gilt also zu klären, welche der Potenziale aktuell schon eingesetzt werden können.

Im Umfeld der Digital Humanities – insbesondere im Bereich des Natural Language Processing (NLP), speziell des Quantum Natural Language Processing (QNLP)³⁷ – finden sich erste Ansätze, um sich die Potenziale (i) ‚Beschleunigung‘, (ii) ‚Verarbeitung großer Datenmengen‘ und (iv) ‚Lösung praktisch bisher unlösbarer Probleme‘ zunutze zu machen. Hier zielen experimentelle Verfahren auf die Entwicklung von NLP-Modellen, die explizit für die Ausführung auf Quanten-Hardware ausgelegt sind. Dabei werden als potenzieller Quantenvorteil für NLP eine Reihe von Möglichkeiten aufgezeigt, so beispielsweise: eine algorithmische Beschleunigung von Such- oder Klassifizierungsaufgaben als eine der dominierenden Aufgaben im NLP, das Abbilden komplexer linguistischer Strukturen, welches durch exponentiell große Quantenzustandsräume ermöglicht wird, und auf Dichtematrizen basierende neuartige Bedeutungsmodelle, welche auf natürliche Weise linguistische Phänomene wie Hyponymie und linguistische Mehrdeutigkeit modellieren. Hierbei werden Sätze als parametrisierte Quantenschaltkreise instanziiert, Wortbedeutungen in Quantenzuständen kodiert und explizit die grammatikalische Struktur berücksichtigt, indem sie in Verschränkungsoperationen überführt werden.³⁸ Das Vorgehen wurde nach ersten Simulationen auch bereits auf

28 Vgl. Havenstein et al. 2018.

29 Vgl. Nielsen / Chuang 2010, S. 40–42.

30 Vgl. Raz / Tal 2019.

31 Vgl. Dickel 2018.

32 Vgl. Barzen 2021 S. 5–7 für eine vertiefte Darstellung diese sieben Vorteile in Bezug zu den Digital Humanities.

33 Wie beispielsweise von IBM die [IBM Quantum Experience](#).

34 Wie beispielsweise der Zugang zum Quantencomputer, der über das [Kompetenzzentrum Quantencomputing von Fraunhofer](#) „für externe Kundinnen und Kunden zur Zeit für 11.621 Euro pro Monat“ zur Verfügung gestellt wird.

35 Was die vorherrschende Implementierung bei den kommerziell verfügbaren Quantencomputern wie beispielsweise die von IBM oder Google ist.

36 Vgl. Gambetta 2020.

37 Vgl. Palmer 2020; Meichanetzidis et al. 2020; Lorenz et al. 2021.

NISQ-Hardware erprobt und stellt die Basis für mögliche Skalierbarkeit dar, wenn sich die Leistungsfähigkeit der Quantenhardware in naher Zukunft verbessert.³⁹

Darüber hinaus ist aber im Besonderen das Potenzial (iii) ‚Präzision der Ergebnisse‘ für unterschiedliche Anwendungen in den Digital Humanities als äußerst vielversprechend zu betrachten (siehe das konkrete Beispiel in Abschnitt 4.4). Mit zunehmenden Datenmengen in Digital Humanities-Projekten ist deren Analyse mit Techniken des Maschinellen Lernens in vielen der Vorhaben zu beobachten. Zudem ist, wie Abschnitt 2.3 ausführt, der Einsatz von Quantencomputern für diesen Bereich bereits besonders zielführend und somit das Nutzen der Vorteile für die Digital Humanities-Community naheliegend. Um allerdings Quantencomputer zu nutzen, gilt es in der Ausführung eine weitere Hürde zu überwinden: Denn die in den Digital Humanities bereits bestehende Hürde des Findens einer gemeinsamen Sprache zwischen den beteiligten Disziplinen, wie der Informatik und der jeweiligen geisteswissenschaftlichen Disziplin, wird für die Quantum Humanities noch höher, da sich hier die Anzahl der beteiligten Disziplinen um die Physik und Mathematik vermehrt. Es stellt sich also die Frage, wie erstes Anwendungswissen und wiederverwendbare Implementierungsbausteine in diesem Bereich aufgebaut werden können, um die gegebene Einstiegshürden abzubauen. Um diesen Prozess für einen ersten Anwendungsfall zu vereinfachen und die technischen Details weitestgehend vor dem Endnutzer zu verstecken, haben wir unser Analyserwerkzeug QHAna (Quantum Humanities Analysis Tool)⁴⁰ entwickelt, welches den Vergleich klassischer und Quanten-basierter Verfahren erlaubt, um mittels Experimenten für den Bereich Quanten-basiertes Maschinelles Lernen in den Digital Humanities Anwendungswissen für weiterführende Anwendungsfälle in anderen Domänen, aber auch für größere Datenmengen (sobald die Quantencomputer dies erlauben) aufzubauen.

4. QHAna als Experimentierumgebung für Quanten-basiertes Maschinelles Lernen

Als Grundlage, um das Experimentieren mit dem Quanten-basierten Maschinellen Lernen mittels QHAna darzustellen, soll im Folgenden erst das Projekt MUSE⁴¹ als aktuelle Untersuchungsgrundlage vorgestellt und eine knappe Einführung in die Basisfunktionalitäten von QHAna gegeben werden. Die zwei folgenden Experimente konzentrieren sich darauf – aufbauend auf Cluster- und Klassifizierungsansätzen –, die Daten aus MUSE zu analysieren.

38 Vgl. Meichanetzidis et al. 2020, S. 1.

39 Vgl. Lorenz et al. 2021, S. 12f.

40 QHAna.

41 MUSE.

4.1 Das Projekt MUSE als Untersuchungsgrundlage

Zur Erprobung eines ersten Anwendungsfall dienen die Daten aus unserem Digital Humanities-Projekt MUSE.⁴² MUSE (als Akronym für *MUster Suche und Erkennen*) hat zum Ziel, Konventionen zu identifizieren, die sich im Film entwickelt haben, um mittels Filmkostümen beispielsweise Stereotypen, Charaktereigenschaften etc. zu kommunizieren, und diese als eine Mustersprache⁴³ für Kostüme darzustellen, um das Verständnis filmisch vestimentärer Kommunikation zu schärfen. Dazu ist ein umfassender Datensatz zu Filmkostümen entstanden.⁴⁴ Das Korpus besteht aus 60 repräsentativen Filmen, jeweils 20 aus den Genres Western, High-School-Comedy und Märchen, von denen zum Zeitpunkt der Verfassung dieses Artikels bereits 55 komplett erfasst sind. Dies beinhaltet Informationen zu den Filmen, den Rollen und sehr detaillierte Angaben zu den Kostümen. Von den Kostümen sind 5.245 in dem Datensatz zu finden, was 28.822 Basiselemente (wie beispielsweise ‚Wintermantel‘, ‚Badehose‘, ‚Faltenrock‘ etc.) und 63.773 Teilelemente (wie beispielsweise ‚Fledermausärmel‘, ‚Reißverschluss‘, ‚Kellerfalte‘ etc.) mit beispielsweise 159.526-mal bestimmten Farben oder 180.274-mal bestimmten Materialien als Attributzuweisung inkludiert. Die Basis der Kostümerkennung stellt das MUSE-Repository dar, welches wiederum auf einer Ontologie der potentiell relevanten Parameter zur Beschreibung von Kostümen aufbaut.⁴⁵ Hierbei wurden Teilbereiche der Ontologie – wie die der Basiselemente, der Farben, der Materialien, der Designs oder Funktionen, um nur ein paar zu nennen – als Taxonomien abgebildet, um mittels dieser hierarchischen Strukturen die Erfassung zu unterstützen, aber auch Analysen auf unterschiedlichen Abstraktionsebenen zu ermöglichen.⁴⁶ Hier gilt es nun, die erfassten Kostümdaten zu untersuchen, wobei Techniken der Data Analytics⁴⁷ und des Maschinellen Lernens eingesetzt werden.⁴⁸ So sollen beispielsweise einerseits häufig wiederkehrende Strukturen sichtbar gemacht werden, um Hinweise auf mögliche Kostümmuster zu identifizieren, und andererseits sollen die neu erfassten Kostüme den schon gefundenen Gruppierungen automatisch zugeordnet werden. Ersteres kann mittels Clusterverfahren unterstützt werden (siehe Abschnitt 4.3), während letzteres mittels Klassifizierungsalgorithmen erzielt werden kann (siehe Abschnitt 4.4). Um diese Schritte weiter zu optimieren, kann die Analyse der Filmkostümdaten nicht nur mit klassischen Ansätzen aus dem Maschinellen Lernen, sondern wie in Abschnitt 2.3 eingeführt, auch mit Quanten-basierten Ansätzen des Maschinellen Lernen durchgeführt werden.

42 Vgl. Barzen et al. 2018; Barzen 2018.

43 Das Verständnis von Mustern folgen hier dem von Alexander et al. geprägten Musterbegriff. (Vgl. Alexander et al. 1977, S. ix–xvii.)

44 Vgl. Barzen et al. 2021a.

45 Vgl. Barzen 2013. Das Verständnis von Ontologie folgt hier dem der Informatik, wobei der Begriff (anders als in der Philosophie) als Werkzeug der Wissensrepräsentation mit klare Begriffsdefinitionen durch deren Merkmale sowie deren Einbettung in ein Ordnungsschema verstanden wird. (Vgl. Hesse 2014, S. 306; Vgl. Furrer 2014, S. 308–309).

46 Vgl. Barzen 2018, S. 94–128.

47 Vgl. Falkenthal et al. 2017.

48 Vgl. Barzen 2021, S. 11–16.

Die MUSE-Daten eignen sich aus mehreren Gründen als Anwendungsfall für eine Machbarkeitsstudie und um erste wiederverwendbare Bausteine für Quantum Humanities zu identifizieren: (i) die Daten stehen, den FAIR-Prinzipien folgend, über DaRUS⁴⁹ jedem Interessierten frei zur Verfügung, was die Reproduzierbarkeit der Experimente und damit die Nachvollziehbarkeit der Ergebnisse unterstützt; (ii) es ist ein sehr umfassender und gut strukturierter Datensatz, der beispielsweise durch die strukturelle Besonderheit der in Baumstrukturen vorliegenden Daten eine zielgenaue Übersetzung der kategorialen Daten in numerische erlaubt,⁵⁰ was wiederum die Voraussetzung für die Anwendbarkeit vieler Analysealgorithmen ist; (iii) das MUSE-Forschungsprojekt hat ein intrinsisches Interesse an der optimalen Auswertung des aufwendig erstellten Datensatzes, um möglichst viele Kostümmuster zu identifizieren; und (iv) – dieser Punkt ist nicht zu unterschätzen – kennen wir die Daten sehr gut, da uns MUSE schon seit 2013 begleitet, was für die Interpretation der Analyseergebnisse fundamental ist.

4.2 QHAna

QHAna ist unsere Analyse-Werkzeugumgebung, um Auswertungen von Daten mittels klassischen und Quanten-basierten Maschinellen Lernens zu unterstützen. Hierbei verfolgt QHAna mehrere Ziele: (i), wie schon ausgeführt, aus dem Anwendungsfall heraus das intrinsische Interesse, Hinweise auf Kostümmuster zu identifizieren; (ii) hierzu eine Möglichkeit bereit zu stellen, diese Analyse mittels klassischen und Quanten-basierten Maschinellen Lernens zu unterstützen; (iii) diese mittels einer modularen Struktur des Werkzeuges so zur Verfügung zu stellen, dass eine einfache Erweiterung um neue Komponenten erlaubt wird; (iv) hierbei die Komplexität so zu ‚verstecken‘, das Einstiegshürden abgebaut werden und die Anwendung der Algorithmen auch für mathematisch und physikalisch nicht speziell geschulte Personengruppen durchführbar ist; (v) durch die gegebene Vergleichbarkeit von klassischen und Quanten-basierten Lösungen den Aufbau von Anwendungswissen zu stärken, so dass dieses auch in anderen, später auch größeren Anwendungsfällen genutzt werden kann; (vi) was wiederum zur Schärfung des Verständnisses der Potentiale des Quantencomputers für Anwendungen in den Digital Humanities und darüber hinaus beiträgt.⁵¹

Wie Abbildung 1 zeigt, ist der in Sektion A abgebildete Ausschnitt der Datenanalyse-Pipeline – bestehend aus den meist iterativ auszuführenden Schritten der ‚Daten-Vorbereitung‘, ‚Feature Engineering‘, ‚Clustering‘ und ‚Klassifikation‘ – die Basis des Hauptmenüs (Sektion B) von QHAna. Die im Hauptmenü darüber hinaus zu findenden Reiter ‚Documentation‘ und ‚Overview‘ geben übergeordnete Informationen zu den vorhandenen Daten und Komponenten sowie zur verwendeten Literatur und Software, wie sie auch die Ergebnisse der Berechnungen zusammenfassen. Wie in Sektion C von Abbildung 1 ersichtlich wird, sind zur Durchführung der einzelnen Schritte der

49 Da R U S .

50 Vgl. Barzen et al. 2021b, S. 40–47.

51 Vgl. Barzen 2021, S. 35.

Analyse bereits verschiedene Komponenten und Algorithmen in QHana integriert (oder befinden sich gerade in der Integration, wenn mit doppeltem Sternchen gekennzeichnet, wie beispielsweise die Principal Component Analysis (PCA) oder die klassischen und Quanten-basierten Autoencoder). Die mit einem Sternchen versehenen Komponenten stehen bereits in mehreren klassischen und Quanten-basierten Ausführungen zur Verfügung. Dies gilt beispielsweise für die Ausführungen der Clusteralgorithmen, bei denen für den Maximum Cut-Algorithmus⁵² bereits drei verschiedene klassische und eine Quanten-basierte und für den k-Means-Algorithmus⁵³ eine klassische und vier Quanten-basierte Implementierungen zur Verfügung stehen. Für die Klassifikationen stehen für die Support Vector Machines (SVM)⁵⁴ eine klassische und zwei Quanten-basierte Versionen sowie die Möglichkeit, den Klassifizierer mittels neuronaler Netze zu errechnen, sowohl mittels eines klassischen und einem Quanten-basierten neuronalen Netz, zur Verfügung.

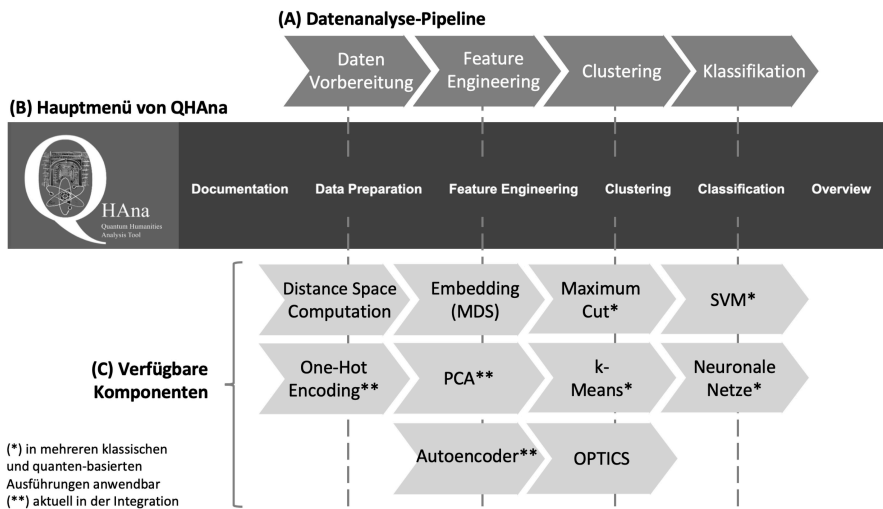


Abb. 1: Übersicht QHana

Um die unterschiedlichen Implementierungen für den Nutzer so einfach wie möglich zur Verfügung zu stellen, können diese über Dropdown-Menüs ausgewählt werden, wobei unabhängig von der Quanten-basierten oder klassischen Implementierung eines Algorithmus die Komplexität der Algorithmen verborgen bleibt und auf der Oberfläche von QHana gezielt nur die Einstellungen der Hyperparameter, die Auswahl der jeweiligen Optimierer oder des Backends (klassischer Computer oder Quantencompu-

52 Vgl. Crooks 2018.

53 Vgl. Arthur / Vassilvitskii 2007; Khan et al. 2019.

54 Vgl. Burges 1998; Bennett / Campbell 2000.

ter) zur Ausführung angepasst werden können. Abbildung 2 zeigt exemplarisch einen Ausschnitt der Auswahlmöglichkeiten, wenn die Quanten-basierte SVM in der Qiskit⁵⁵-Implementierung für die Klassifizierung ausgewählt wird, mit den jeweiligen Default-Werten. Alle Algorithmen sind mit Default-Werten versehen, so dass sie direkt anwendbar sind. Nicht immer sind diese allerdings für den vorliegenden Anwendungsfall die optimalen. Wie in Abbildung 2 ersichtlich, kann hier beispielsweise mittels der Anzahl der sogenannten *Shots* ausgewählt werden, wie häufig der Algorithmus ausgeführt werden soll, was wiederum einen Einfluss auf die Güte des Ergebnisses haben kann. Denn, wie in Abschnitt 2.1 ausgeführt, jedes Messen ist ein Zufallsexperiment (welches aber sehr schnell durchlaufen wird), so dass mehrere Durchläufe nötig sind, um über eine Wahrscheinlichkeitsverteilung das am häufigsten auftretende Ergebnis als ‚das‘ Ergebnis auszuzeichnen. In QHANA sind neben echten Quantencomputern – aktuell die kostenlos von IBM in der IBM Quantum Experience⁵⁶ zur Verfügung gestellten⁵⁷ – für die Ausführung auch mehrere Simulatoren von Quantencomputern anwählbar.

The screenshot shows the 'Initialize Classification' interface. A dropdown menu is open, showing options: 'classicSklearnSVM', 'qkeQiskitSVM', 'variationalQiskitSVM' (selected), 'classicSklearnNN', and 'HybridQNN'. Below the menu is an 'Initialize' button. Underneath, there is a 'set parameters' button and a table of parameters for the selected 'variational SVM (using qiskit)' classification type.

Parameter	Input	Description
ClassificationType	variational SVM (using qiskit)	Name of chosen classification type
Feature Map	ZFeatureMap	Feature Map : {'ZFeatureMap', 'ZZFeatureMap', 'PauliFeatureMap'}, (default='ZFeatureMap') Feature map module used to transform data.
Entanglement	linear	Entanglement : {'full', 'linear'}, (default='full') Specifies the entanglement structure.
Repetitions	2	reps : int, (default=2) The number of repeated circuits.
Shots	1024	Shots : int, (default=1024) Number of repetitions of each circuit, for sampling.
QuantumBackend	aer_statevector_simulator	Backend : Enum default(aer_statevector_simulator) A list of possible backends. aer is a local simulator and ibmq are backends provided by IBM.
IBMQ-Custom-Backend		str default("") The name of a custom backend of ibmq.
IBMQ-Token		IBMQ-Token : str, (default='') IBMQ-Token for access to IBMQ online service
Variational Form	RyRz	Variational Form : {'RealAmplitudes', 'ExcitationPreserving', 'EfficientSU2', 'RyRz'}, (default='RyRz') The variational form instance.
Repetitions (variational Form)	3	reps : int (default=3) For variational form; Specifies how often a block consisting of a rotation layer and entanglement layer is repeated.
Optimizer	SPSA	Optimizer : {'ADAM', 'AQGD', 'BOBYQA', 'COBYLA', 'NELDER_MEAD', 'SPSA', 'POWELL', 'NFT', 'TNC'}, (default='SPSA') The classical optimizer to use.
Max iterations	100	Max iterations : int (default=100) For optimizer; Maximum number of iterations to perform.

Abb. 2: Auswahlmöglichkeiten der Hyperparameter für eine Quanten-basierte Klassifizierung

55 Qiskit.

56 IBM Quantum Experience.

57 Da die Anzahl an kostenlosen Berechnungen limitiert ist, muss nach Anmeldung bei der IBM Experience das Nutzer-Token in QHANA eingegeben werden, um die Berechnungen auf den zur Verfügung stehenden Quantencomputern durchzuführen.

Diese haben zwar den Vorteil, weniger fehleranfällig in den Qubits und den Gates zu sein, können allerdings nur eine geringe Anzahl an Qubits simulieren. Da es sich bei dem in Abbildung 2 ausgewählten Algorithmus um einen – wie in Abschnitt 2.2 beschrieben – ‚hybriden‘, spezieller einen ‚variationellen‘ Quantenalgorithmus handelt, spricht: dieser Quanten-klassische Iterationen beinhaltet, bei denen in jedem Durchlauf die Parameter für den Quantenanteil klassisch optimiert werden,⁵⁸ ist hier die Auswahl des angewandten Optimierers essenziell, um gute Ergebnisse zu erzielen. Um erstes Anwendungswissen über die Güte der Optimierer zu schaffen, sind hier bereits verschiedene Optimierer zur Verfügung gestellt, so dass über die Auswahlmöglichkeiten diese angewendet, ausprobiert und Ergebnisse verglichen werden können. Gleiches gilt natürlich auch für die anderen Hyperparameter. Die folgenden Experimente sollen zwei exemplarische Anwendungen, für welche QHAna genutzt werden kann, näher beleuchten.

4.3 Experimente mit Quanten-basiertem Clustering

Um diejenigen Kostüme zu gruppieren, die gleiche Elemente aufweisen und daher potenziell auf Kostümmuster hinweisen, gilt es, die Kostümdaten mittels Clusteralgorithmen zu untersuchen. In Vorarbeiten haben wir bereits erste Ergebnisse mittels Quanten-basiertem Clustering auf einem kleinen Datensatz bestehend aus 10 Kostümen zeigen können.⁵⁹ Hierbei wurden die Kostüme mit dem Maximum Cut-Algorithmus sowohl mittels klassischer als auch mittels Quanten-basierter Implementierung geclustert. Wie Abbildung 3 verdeutlicht, konnten wir die Experimente auf einen deutlich größeren Datensatz ausweiten sowie QHAna um vielversprechende Clusteralgorithmen durch die Integration mehrerer k-Means-Algorithmen bereichern. Für die in Abbildung 3 zu sehenden Ergebnisse sind 40 Kostüme aus dem Genre High-School-Comedy ausgewählt worden, wobei jedes Kostüm über seine ausgewählten Attribute (hier Farbe, Zustand, Charaktereigenschaft und Alterseindruck) und deren Lage in der jeweiligen Taxonomie zu jedem anderen Kostüm mit einem Ähnlichkeitswert versehen wurde.⁶⁰ Diese Ähnlichkeitswerte zwischen den Kostümen sind in ein Distanzmaß überführt worden und sodann mittels Multidimensionaler Skalierung in einen 2-dimensionalen Raum eingebettet worden, der die Abstände des Ähnlichkeitswerte möglichst genau widerspiegelt. Sektion A von Abbildung 3 zeigt das Clusterergebnis der Anwendung des klassischen k-Means-Algorithmus auf diese eingebetteten Kostümdaten, während Sektion B das Ergebnis einer von vier aktuell in QHAna integrierten Quanten-basierten Implementierungen des k-Means-Algorithmus zeigt. Die für Sektion B von Abbildung 3 angewandte Implementierung wird in QHAna mit *Positive Correlation Quantum k-Means* betitelt, basiert auf den Arbeiten von Sarma et al.⁶¹ und wurde auf

58 Vgl. Sierra-Sosa et al. 2020.

59 Vgl. Barzen 2021, S. 41f.

60 Vgl. Wu / Palmer 1994.

61 Vgl. Sarma et al. 2020.

dem von der IBM Quantum Experience zur Verfügung gestellten *ibm_qasm_simulator* ausgeführt.

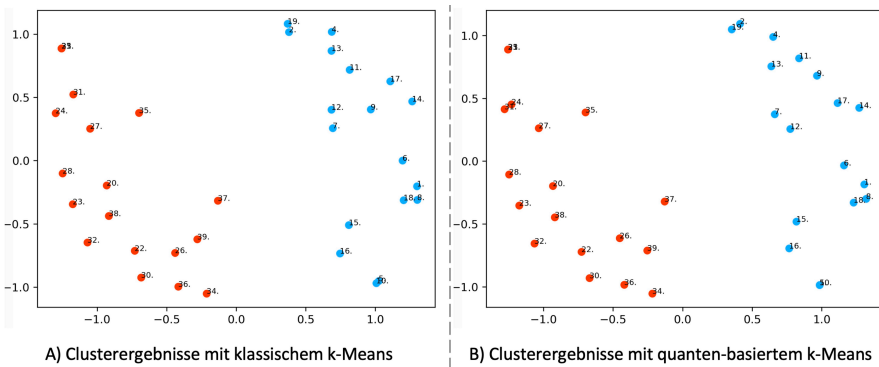


Abb. 3: Vergleich des klassischen und Quanten-basierten Clustering mittels k-Means

Auch wenn durch die Neuberechnung der Einbettung der Kostüme in den 2-dimensionalen Raum Abweichungen in der Lage der Datenpunkte entstanden sind, stimmen die in Abbildung 3 zu sehenden Cluster in Sektion A und B überein: Beide Male sind in Cluster 1 (blaue Punkte) die gleichen Kostüme zusammen gruppiert worden wie in Cluster 2 (rote Punkte). Betrachtet man die Attribute der Kostüme näher, beinhalten die Kostüme bei Cluster 1 in den Charaktereigenschaften die Werte ‚aktiv‘, ‚gut‘ und ‚stark‘, in der Farbe ‚Hellblau‘, ‚Hellgrün‘, ‚Hellbraun‘ und ‚Hellgrau‘, im Zustand ‚ordentlich‘, ‚neu‘, ‚sauber‘ und ‚gebügelt‘ und im Alterseindruck ‚30ern‘, ‚40ern‘ und ‚50ern‘. Bei Cluster 2 wiederum sind es Kostüme mit den Werten ‚böse‘, ‚passiv‘ und ‚schwach‘ in den Charaktereigenschaften, ‚Dunkelblau‘, ‚Dunkelgrün‘, ‚Dunkelbraun‘, und ‚Dunkelviolett‘ in der Farbe, ‚alt‘, ‚verwaschen‘, ‚ausgewaschen‘, ‚verschwitzt‘, ‚fleckig‘, ‚nass‘, ‚abgetragen‘, ‚dreckig‘, ‚beschmutzt‘, ‚kaputt‘ und ‚unordentlich‘ im Zustand und ‚20ern‘ und ‚Jugendlicher‘ im Alterseindruck. Hier lässt sich als erster Hinweis auf eine mögliche Hypothese festhalten, dass die Kostüme aus Cluster 1 eher positiv behaftet sind, während die in Cluster 2 eher negativ konnotiert sind. Die so gefunden Cluster müssen also durch einen interpretatorischen Schritt auf ihre Aussagekraft hin untersucht werden und können nach der Verifikation durch weitere Analysen der Daten im Idealfall in ein Kostümmuster überführt werden.

Was unsere bisherigen Experimente mit Quanten-basierten Clusteralgorithmen zeigen, ist, dass Quanten-basiertes Clustering möglich ist – allerdings im Rahmen der bisher angewandten Implementierungen noch keinen Quantenvorteil bietet. Hier gilt es in einem nächsten Schritt noch weitere Verfahren und Implementierungen zu integrieren und deren Ergebnisse zu vergleichen. Trotzdem können diese ersten Ergebnisse, auch ohne einen Quantenvorteil, als Aufbau von Anwendungswissen sowohl über die

angebotenen Lösungen und deren Ergebnisse als auch über den Einfluss der jeweilig ausgewählten Hyperparameter auf die Ergebnisse gewertet werden.

4.4 Experimente mit Quanten-basiertem Klassifizieren

Dem zugrundeliegenden Konzept von Mustern folgend, stehen die Kostümmuster – als abstrahierte Lösungen für Designprobleme – mit den konkreten Lösungen, also den konkreten Kostümen aus den Filmen, auf denen die Muster basieren, in einer engen Beziehung. In unserem Muster-Repository wird dieser Beziehung in der Kategorie ‚Anwendungsbeispiele‘ des Musters Rechnung getragen, und Verlinkungen von dem Muster zu den Kostümbeschreibungen im MUSE-Repository lassen Nutzer in die konkreten Umsetzungen von Kostümbildner*innen navigieren.⁶² Dies ist sowohl für die theoretische Auseinandersetzung als auch für die Verwendung der Muster in der Praxis von Interesse und stärkt die Nachvollziehbarkeit der Musterinhalte. Da im Projekt immer noch Kostüme erfasst werden, ist es für MUSE wichtig, die neu hinzukommenden Kostüme denjenigen Mustern zuzuweisen, zu denen sie beitragen. Sind beispielsweise bestimmte Farben, Materialien oder Basiselemente als signifikant für die Kommunikation eines speziellen Stereotypen identifiziert worden, sollen diejenigen Kostüme identifiziert werden, welche ähnliche Attributsausprägungen haben, um möglicherweise dem Muster zugeordnet werden zu können.

Basierend auf gelabelten Trainingsdaten, deren Label durch die identifizierten Cluster im Schritt davor bestimmt sind (die später aber auch manuell vergeben werden können, um den interpretatorischen Schritt der Musterzuordnung einfließen lassen zu können), werden hier mittels Klassifizierungsalgorithmen möglichst optimale Trennlinien zwischen den Clustern gefunden, welche erlauben, neue Datenpunkte direkt in die richtige Klasse einzusortieren. Zur Klassifizierung sind in QHAna unter anderem sowohl klassische Implementierungen der Support Vector Machine (SVM) als auch eine Implementierung einer Quanten-basierte Methode für SVM mittels Quantum Kernel Estimation (QKE)⁶³ umgesetzt. Abbildung 4 zeigt exemplarisch vier Ergebnisse der Anwendung von drei klassischen (Sektion A–C) und einer Quanten-basierte (Sektion D) Klassifizierung mittels SVM. Als Daten sind die gleichen 40 Kostüme ausgewählt worden, welche auch für das Clustering ausgewählt waren. Das Label ‚positiv‘ entspricht Cluster 1 (blaue Punkte in Abb. 3) und das Label ‚negativ‘ entspricht Cluster 2 (rote Punkte in Abb. 3). Um die Mächtigkeit der SVM zu zeigen, sind hier zu Anschauungszwecken mit Rollenrelevanz und Stereotype bewusst andere Attribute für die Kostüme, über deren Ähnlichkeitswerte die Kostüme in dem 2-dimensionalen Raum dargestellt werden, ausgewählt worden. Wie Abbildung 3 veranschaulicht, sind die Daten mit den hierfür ausgewählten Attributen linear separabel, also im Zweidimensionalen durch eine gerade Linie oder im Mehrdimensionalen durch eine Hyperebene trennbar, was für die Datenpunkte in Abbildung 4 nicht gilt. Zum Berechnen des Klassifizierers

62 Vgl. Barzen 2018, S. 208f.

63 Havlicek et al. 2018.

wird nun die in Abschnitt 2.2 beschriebene Kern-Funktion wichtig. Für Abbildung 4 sind dafür mit dem *Radial Basis Function (RBF) Kernel* (Sektion A), der *Polynomial Kernel Function* (Sektion B–C) und der *Quantum Kernel Estimation* (Sektion D) drei verschiedene Varianten zum Einsatz gekommen. Hierbei sind bei allen Berechnungen, mit Ausnahme der zweiten Anwendung der Polynomial Kernel Function (Sektion C), die Default-Werte von QHANA angewendet worden.

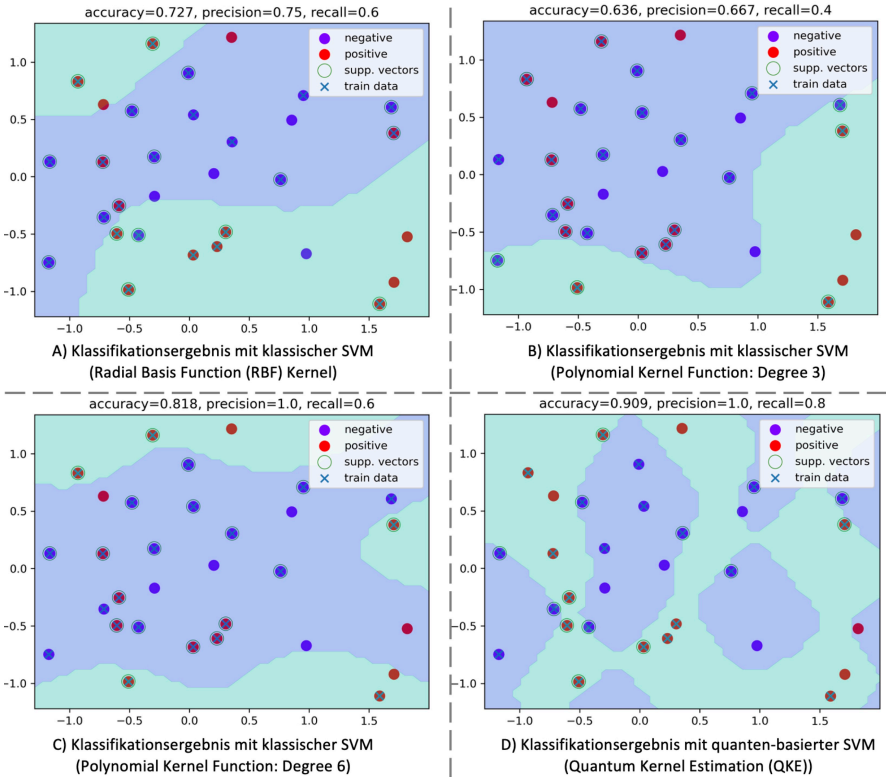


Abb. 4: Vergleich von drei klassischen (A–C) und einer Quanten-basierten (D) Klassifizierung mittels SVM

Als Trainingsdaten wurden willkürlich 25 der gelabelten Datenpunkte (teils übereinanderliegend in Abb. 4) ausgewählt, um die Trennlinien zwischen den Clustern zu berechnen, welche dann die Zuordnung der übrigen Kostümdatenpunkte erlauben (hierbei ist zu beachten, dass – abweichend zu Abbildung 3 – die roten Punkte für die ‚positiv‘ und die blauen Punkte für die ‚negativ‘ gelabelten Datenpunkte stehen). Die Präzision, mit der die Datenpunkte richtig zugeordnet wurden, wird über die ‚accuracy‘ angegeben. Was Abbildung 4 deutlich macht, ist, dass die Präzision, mit welcher Datenpunkte dem richtigen Cluster zugeordnet werden können, bei dem Einsatz der QKE

(91% richtig klassifiziert) höher ist als bei dem Einsatz der klassischen SVMs (Sektion A: 73 %, Sektion B: 64 %, Sektion C: 82 % richtig klassifiziert). Sektion B und Sektion C von Abbildung 4 verdeutlichen zudem, dass mit dem Anpassen der Hyperparameter Einfluss auf die Ergebnisse erzielt werden kann. Hierzu bietet QHAna die Möglichkeit, sowohl für klassische wie auch für Quanten-basierte Implementierungen die Einstiegshürden zur Partizipation stark abzubauen, um Erfahrungswissen aufzubauen. Die Nutzung von QHAna kann uns hier beispielsweise einen klaren Hinweis darauf geben, dass die Zuordnung neu erfasster Kostüme zu bereits identifizierten Kostümmustern mittels QKE präziser vorgenommen werden kann, als wenn man diese rein mit klassischen Verfahren vornimmt.⁶⁴

5. Fazit und Ausblick

Da Quantencomputer eine ganze Bandbreite an Potenzialen bieten, ist die grundlegende Fragestellung, der es nachzugehen gilt, inwieweit diese Potenziale auch für Anwendungen in den Digital Humanities eingesetzt werden können. Das bereits Vorteile, wie beispielsweise in der Präzision der Klassifizierungsergebnisse durch den Einsatz des Quantencomputers, erzielt werden können, hat der vorliegende Anwendungsfall in kleinerem Rahmen zeigen können. Für diesen Anwendungsfall lässt sich die im Titel aufgespannte Frage, ob Quantencomputing in den Digital Humanities als innovativ oder übertrieben eingestuft werden kann, mit ‚innovativ‘ beantworten, da nur durch den Einsatz des Quantencomputers diese Präzision der Ergebnisse möglich war. Um der Frage aber in einem größeren Rahmen nachzugehen, will der vorliegende Artikel Interessierte dazu anregen, sich selber ein Bild davon zu machen, was bereits möglich ist. Um in dem Bereich des Quanten-basierten Maschinellen Lernens Anwendungsfälle aus den Digital Humanities zu erproben, ist QHAna entwickelt worden. Indem QHAna der Öffentlichkeit frei zur Verfügung steht, soll hiermit, ganz im Sinne von wissenschaftlicher Nachvollziehbarkeit und Reproduzierbarkeit der Ergebnisse, der Aufbau von Anwendungswissen für mögliche Quanten Humanities unterstützt werden.

Die entwickelte Werkzeugumgebung QHAna führt einerseits durch den Analyseprozess von Daten und ermöglicht andererseits, indem sowohl klassische wie auch Quanten-basierte Verfahren angewendet werden können, durch den Vergleich beider Ansätze (aktuell in den Bereichen: Daten-Vorbereitung, Feature Engineering, Clustering, Klassifikation) die Angemessenheit, teils sogar Überlegenheit, Quanten-basierter Lösungen zu beurteilen. In einem nächsten Schritt können bewährte Realisierungen von Vorgehensmodellen, Algorithmen und deren Parameter der Digital Humanities-Community zur Verfügung gestellt werden und so die Umsetzung weiterführender Anwendungen potenzieller Quantum Humanities unterstützen. Den Ansatz (wie auch das Werkzeug) beabsichtigen wir für die Analyse musikwissenschaftlicher Daten auszuweiten und zu erproben.⁶⁵

64 Dies verifiziert und stärkt erste Aussagen zur erhöhten Präzision mittels QKE (vgl. Barzen 2021, S. 44f) auch für die Anwendung unterschiedlicher Implementierung klassischer SVMs.

Vor dem Hintergrund der aktuellen Beschränkungen der existierenden Quantencomputer können diese Vorarbeiten aktuell allerdings nur als experimentelle Pionierarbeiten für Quantum Humanities gesehen werden: Als eine Machbarkeitsstudie also, die, solange die Maschinen für groß angelegte Experimente noch zu verrauscht und fehleranfällig sind, bereits auslotet, wie weiterführende Anwendungen aussehen könnten. Um die praktischen Vorteile, wie beispielsweise die gezeigte Präzision in der Klassifizierung, nutzen zu können, gilt es genau hier und heute bereits das Anwendungswissen aufzubauen für eine Community, die stark davon profitieren kann, deren Hürden zur Partizipation jedoch gleichzeitig als hoch einzuschätzen sind. Um daher einen einfachen Zugang noch weiter zu verbessern, arbeiten wir kontinuierlich an der Erweiterung von QHAna, was sowohl die Integration von weiteren Analyseverfahren und -Implementierungen, die Einbindung weiterer Quantencomputer-Backends wie auch die Bestimmung optimaler Hyperparameter beinhaltet. Weiter wird QHAna aktuell so modifiziert, dass es auch für andere Datensätze nutzbar ist und eigene Implementierungen von Algorithmen integriert werden können. Die darüberhinausgehende Vision hybride und Quanten-basierte Teilbereiche in einen benutzerdefinierten und automatisch durchlaufenden Analyse-Workflow zu integrieren, haben wir in einer ersten prototypischen Implementierung bereits umgesetzt.⁶⁶

Danksagungen

Wir möchten uns bei Fabian Bühler, Daniel Fink, Marcel Messer, Felix Truger und Philipp Wundrack und für ihren wertvollen Input und die Implementierung verschiedener Aspekte des Anwendungsfalls und von QHAna bedanken. Teile dieser Arbeit wurde durch das BMWi-Projekt PlanQK (01MK20005N) und das Terra Incognita Projekt Quantum Humanities der Universität Stuttgart gefördert.

Bibliografie

- Daniel S. Abrams / Seth Lloyd 1997: Simulation of Many-Body Fermi Systems on a Universal Quantum Computer. In: Phys. Rev. Lett. 79 (1997), 2586. DOI: 10.1103/PhysRevLett.79.2586. [online]
- Christopher Alexander / Sara Ishikawa / Murry Silverstein / Max Jacobson / Ingrid Fiksdahl-King / Shlomo Angel: A Pattern Language: Towns, Buildings, Constructions. New York 1977.
- David Arthur / Sergei Vassilvitskii: k-means++: The Advantages of Careful Seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (2007), S. 1027–1035.

65 Vgl. Barzen et al. 2017.

66 Leymann / Barzen 2021, S. 2f.

- Johanna Barzen / Fabian Bühler / Frank Leymann: MUSE Datenset. V1. DaRUS, 2021. DOI: 10.18419/darus-1805. (Barzen et al. 2021a) [[online](#)]
- Johanna Barzen / Frank Leymann / Michael Falkenthal / Daniel Vietz / Benjamin Weder / Karoline Wild: Relevance of Near-Term Quantum Computing in the Cloud: A Humanities Perspective. In: Cloud Computing and Services Science. CLOSER 2020. Hg. von Donald Ferguson / Claus Pahl / Markus Helfert. Cham 2021, S. 25–58. DOI: 10.1007/978-3-030-72369-9_2. (Barzen et al. 2021b) [[online](#)]
- Johanna Barzen / Frank Leymann: Quantum Humanities: A First Use Case for Quantum-ML in Media Science. In: Digitale Welt 4 (2020), H. 1, S. 102–103. DOI: 10.1007/s42354-019-0243-2. [[online](#)]
- Johanna Barzen / Frank Leymann: Quantum humanities: a vision for quantum computing in digital humanities. In: SICS Software-Intensive Cyber-Physical Systems 35 (2019), H. 1, S. 153–158. DOI: 10.1007/s00450-019-00419-4. [[online](#)]
- Johanna Barzen / Michael Falkenthal / Frank Leymann: Wenn Kostüme sprechen könnten: MUSE - Ein musterbasierter Ansatz an die vestimentäre Kommunikation im Film. In: Digital Humanities. Perspektiven der Praxis. Hg. von Peggy Bockwinkel / Beatrice Nickel / Gabriel Viehhauser. Berlin 2018, S. 223–241.
- Johanna Barzen / Uwe Breitenbücher / Linus Eusterbrock / Michael Falkenthal / Frank Hentschel / Frank Leymann: The vision for MUSE4Music. Applying the MUSE method in musicology. In: Computer Science – Research and Development 32 (2017), H. 3–4, S. 323–328. DOI: 10.1007/s00450-016-0336-1. [[online](#)]
- Johanna Barzen: From Digital Humanities to Quantum Humanities: Potentials and Applications. arXiv:2103.11825 [cs.LG], 17.3.2021. [[online](#)]
- Johanna Barzen: Taxonomien kostümrelevanter Parameter: Annäherung an eine Ontologisierung der Domäne des Filmkostüms. Technischer Bericht Nr. 2013/04, Stuttgart 2013.
- Johanna Barzen: Wenn Kostüme sprechen – Musterforschung in den Digital Humanities am Beispiel vestimentärer Kommunikation im Film. Köln 2018. [[online](#)]
- Kristin P. Bennett / Colin Campbell: Support Vector Machines: Hype or Hallelujah? In: ACM SIGKDD Explorations Newsletter 2 (2000), H. 2. DOI: 10.1145/380995.380999. [[online](#)]
- Bundesfinanzministerium: Corona-Folgen bekämpfen, Wohlstand sichern, Zukunftsfähigkeit stärken. Ergebnis Koalitionsausschuss 3.06.2020. [[online](#)]
- Christopher J.C. Burges: A Tutorial on Support Vector Machines for Pattern Recognition. In: Data Mining and Knowledge Discovery 2 (1998), S. 121–167.
- Gavin Crooks: Performance of the Quantum Approximate Optimization Algorithm on the Maximum Cut Problem. arXiv:1811.08419 [quant-ph], 20.11.2018. [[online](#)].
- Christian Dickel: A Cloud Quantum Computer Business Plan. In: QuTech, 18.07.2018. [[online](#)]
- Vedran Dunjko / Jacob M. Taylor / Hans J. Briegel: Quantum-enhanced machine learning. arXiv:1610.08251 [quant-ph], 26.10.2016. [[online](#)]

- Michael Falkenthal / Johanna Barzen / Uwe Breitenbücher / Sascha Brüggemann / Daniel Joos / Frank Leymann / Michael Wurster: Pattern Research in the Digital Humanities: How Data Mining Techniques Support the Identification of Costume Patterns. In: *Computer Science – Research and Development* 32 (2017), H. 3–4, S. 311–321. DOI: 10.1007/s00450-016-0331-6. [\[online\]](#)
- Edward Farhi / Jeffrey Goldstone / Sam Gutmann: A Quantum Approximate Optimization Algorithm. arXiv:1411.4028 [quant-ph], 14.11.2014. [\[online\]](#)
- Frank J. Furrer: Eine kurze Geschichte der Ontologie. Von der Philosophie zur modernen Informatik. In: *Informatik Spektrum. Organ der Gesellschaft für Informatik e.V. und mit ihr assoziierter Organisationen* 37 (2014), H. 4, S. 308–317.
- Jay Gambetta: IBM's Roadmap For Scaling Quantum Technology. In: *IBM Research Blog*, 15.09.2020. [\[online\]](#)
- Lov Grover: A fast quantum mechanical algorithm for database search. In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing* (1996), S. 212–219. DOI: 10.1145/237814.237866. [\[online\]](#)
- Emmanuel Haven / Andrei Khrennikov: *A Quantum Social Science*. Cambridge, New York u. a. 2013.
- Christopher Havenstein / Damarcus Thomas / Swami Chandrasekaran: Comparisons of Performance between Quantum and Classical Machine Learning. In: *SMU Data Science Review* 1 (2018), H. 4, Artikel 11. [\[online\]](#)
- Vojtech Havlicek / Antonio D. Córcoles / Kristan Temme / Aram W. Harrow / Abhinav Kandala / Jerry M. Chow / Jay M. Gambetta: Supervised learning with quantum enhanced feature spaces. arXiv:1804.11326 [quant-ph], 5.6.2018. [\[online\]](#)
- Wolfgang Hesse: Ontologie und Weltbezug – vom philosophischen Weltverständnis zum Konstrukt der Informatik. In: *Informatik Spektrum. Organ der Gesellschaft für Informatik e.V. und mit ihr assoziierter Organisationen* 37 (2014), H. 4, S. 298–307.
- Peter J. Karalekas / Nikolas A. Tezak / Eric C. Peterson / Colm A. Ryan / Marcus P. da Silva / Robert S. Smith: A quantum-classical cloud platform optimized for variational hybrid algorithms. In: *Quantum Sci. Technol.* 5 (2020), 024003. DOI: 10.1088/2058-9565/ab7559. [\[online\]](#)
- Sumsam Ullah Khan / Ahsan Javed Awan / Gemma Vall-Llosera: K-Means Clustering on Noisy Intermediate Scale Quantum Computers. arXiv:1909.12183 [cs.ET], 26.9.2019. [\[online\]](#)
- Frank Leymann / Johanna Barzen: The bitter truth about gate-based quantum algorithms in the NISQ era. In: *Quantum Sci. Technol.* 5 (2020), 044007. DOI: 10.1088/2058-9565/abae7d. [\[online\]](#)
- Frank Leymann / Barzen, Johanna: Hybrid Quantum Applications Need Two Orchestrations in Superposition: A Software Architecture Perspective. arXiv:2103.04320 [quant-ph], 7.3.2021 [\[online\]](#)

- Frank Leymann / Johanna Barzen / Michael Falkenthal / Daniel Vietz / Benjamin Weder / Kaoline Wild: Quantum in the Cloud: Application Potentials and Research Opportunities. In: Proceedings of the 10th International Conference on Cloud Computing and Services Science (CLOSER) (2020), S. 9–24. DOI: 10.5220/0009819800090024. [[online](#)]
- Robin Lorenz / Anna Pearson / Konstantinos Meichanetzidis / Dimitri Kartsaklis / Bob Coecke: QNLP in Practice: Running Compositional Models of Meaning on a Quantum Computer. arXiv:2102.12846 [cs.CL], 25.2.2021. [[online](#)]
- Konstantinos Meichanetzidis / Alexis Toumi / Giovanni de Felice / Bob Coecke: Grammar-Aware Question-Answering on Quantum Computers. arXiv:2012.03756 [quant-ph], 7.12.2020. [[online](#)]
- National Academies of Sciences, Engineering, and Medicine: Quantum Computing: Progress and Prospects. Washington, DC 2019. DOI: 10.17226/25196. [[online](#)]
- Michael A. Nielsen / Isaac L. Chuang: Quantum Computation and Quantum Information. Cambridge, New York u. a. 2010.
- Maija Palmer: Quantum computers will create better versions of Alexa and Siri. In: Sifted, 10.12.2020. [[online](#)]
- Alberto Peruzzo / Jarrod McClean / Peter Shadbolt / Man-Hong Yung / Xiao-Qi Zhou / Peter J. Love / Alán Aspuru-Guzik / Jeremy L. O’Brien: A variational eigenvalue solver on a quantum processor. In: Nature Communications 5 (2014), 4213. DOI: 10.1038/ncomms5213. [[online](#)]
- Svatopluk Poljaka / Franz Rendl: Solving the max-cut problem using eigenvalues. In: Discrete Applied Mathematics 62 (1995), H. 1–3, S. 249–278. DOI: 10.1016/0166-218X(94)00155-7. [[online](#)]
- John Preskill: Quantum Computing in the NISQ era and beyond. In: Quantum 2 (2018), S. 79–99. DOI: 10.22331/q-2018-08-06-79. [[online](#)]
- Ran Raz / Avishay Tal: Oracle Separation of BQP and PH. In: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (2019), S. 13–23. DOI: 10.1145/3313276.3316315. [[online](#)]
- Eleanor Rieffel / Wolfgang Polak: Quantum Computing: A Gentle Introduction. Cambridge u. a. 2011.
- Troels F. Rønnow / Zhihui Wang / Joshua Job / Sergio Boixo / Sergei V. Isakov / David Wecker / John M. Martinis / Daniel A. Lidar / Matthias Troyer: Defining and detecting quantum speedup. In: Science 345 (2014), 6195, S. 420–424. DOI: 10.1126/science.1252319. [[online](#)]
- Abhijat Sarma / Rupak Chatterjee / Kaitlin Gili / Ting Yu: Quantum Unsupervised and Supervised Learning on Superconducting Processors. In: Quantum Information and Computation 20 (2020), H. 7–8, S. 541–552. DOI: 10.26421/QIC20.7-8-1. [[online](#)]
- Maria Schuld / Francesco Petruccione: Supervised Learning with Quantum Computers. Cham 2018. DOI: 10.1007/978-3-319-96424-9. [[online](#)]
- Peter W. Shor: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. In: SIAM Journal on Computing (1997). DOI: 10.1137/S0097539795293172. [[online](#)]

Daniel Sierra-Sosa / Juan Arcila-Moreno / Begonya Garcia-Zapirain / Cristian Castillo-Olea / Adel Elmaghraby: Dementia Prediction Applying Variational Quantum Classifier. arXiv:2007.08653 [quant-ph], 14.7.2020. [[online](#)]

Steven S. Skiena: The Data Science Design Manual. Cham 2017. DOI: 10.1007/978-3-319-55444-0. [[online](#)]

Zhibiao Wu / Martha Palmer: Verb Semantics and Lexical Selection. In: Proceedings of the 32nd annual meeting on Association for Computational Linguistics (1994), S. 133–138. DOI: 10.3115/981732.981751. [[online](#)]