Assessing a combined human coding and natural language processing method for qualitative analysis in physics education research

Alexis Buzzell and Ramón Barthelemy
Department of Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT, 84112

Timothy J. Atherton
Department of Physics and Astronomy, Tufts University, 574 Boston Avenue, Medford, MA, 02155

As artificial intelligence becomes increasingly integrated into education research, it is critical to assess the capabilities and limitations of such tools in complex disciplinary contexts. This study explores the effectiveness of natural language processing (NLP) for topic identification using the `BERTopic` package to analyze undergraduate quantum mechanics syllabi. By comparing AI-generated topics with human-coded themes from a corpus of syllabi from 50 US institutions, we find that while NLP reliably identifies broad and structural elements of syllabi (e.g., policies, textbook references), it is limited in its ability to capture the nuanced and context-dependent content typical of upper-level physics instruction. The AI-derived topics were often underrepresented and diverged from expert human interpretation, particularly in areas requiring deep disciplinary knowledge. Our findings support the use of NLP as a complementary tool for theme generation and cross-validation, but emphasize the continued necessity of expert human analysis for rigorous and meaningful educational research.

2025 PERC Proceedings edited by Pawl, Zwolak, and Leak; Peer-reviewed, doi.org/10.1119/perc.2025.pr.Buzzell

I. INTRODUCTION

As artificial intelligence (AI) tools become increasingly integrated in education research [1], including the domain of physics education research (PER) [2–5], there is a growing need to carefully assess their capabilities and limitations. Among such tools, natural language processing (NLP) has shown promise for identifying patterns and latent themes within large textual datasets, such as student work [2] and qualitative interview transcripts [4].

Course syllabi offer a particularly valuable dataset for exploring the potential of NLP in PER. Syllabi are publicly available or easy to obtain and contain no confidential or private information from participants, making them a low-risk source of textual data. Moreover, the syllabi reflect the epistemological perspectives and pedagogical methods of the instructors [6], offering insight into curricular decisions and institutional norms. Analyzing these documents enables researchers to trace the inclusion or omission of specific topics and how these vary across institutions.

A subsequent study used only human coding to analyze the four-year undergraduate quantum curriculum across 50 US physics programs [7]. That study grouped course topics using content structures derived from textbooks by Griffiths [8] and McIntyre [9], highlighting both the diversity and commonality of quantum instruction across institutions. The themes and topic codes from that analysis form the foundation for comparison in the present work (see Tables I and II).

In a previous study on modern physics courses, it was found that NLP was valuable for identifying both topics and policy-level decisions in the syllabi [10]. However, NLP often under counted the frequency of topics in comparison with human analysis. Building on these prior studies, we investigate whether NLP can successfully extract meaningful curricular content in more advanced contexts.

In this study, we combine NLP using `BERTopic` with prior human-coded analysis to examine quantum mechanics syllabi from the same 50 institutions. Our research question is: *How effective is natural language processing, specifically using BERTopic, at identifying coherent and meaningful curricular themes in upper-level undergraduate quantum mechanics syllabi, when compared to human-coded analysis?* By leveraging a corpus large enough for computational analysis yet small enough for manual validation, we aim to assess both the affordances and limitations of NLP methods in a rigorous physics education context.

II. METHODS

This study used topic modeling, a class of NLP techniques that identify latent semantic structures within a collection of texts [11–13]. Topic modeling involves a series of computational steps: converting documents into smaller linguistic units (e.g., tokens), transforming these tokens into numerical representations (embeddings), and then using clustering algorithms to group similar embeddings into coherent topics. Each stage can be configured with different algorithms or models, and the outcome typically depends on both the selected parameters and the structure of the underlying text data. For example, consider a sentence in a quantum mechanics syllabus such as "The time-independent Schrödinger equation will be applied to the infinite square well and harmonic oscillator." `BERTopic` tokenizes this sentence and transforms it into a dense numerical embedding using a pretrained language model (e.g., `BERT`). The model captures the contextual relationship between phrases like "Schrödinger equation," "infinite square well," and "harmonic oscillator." After dimensionality reduction, a clustering algorithm groups this embedding with other similar sentence fragments from the corpus. If enough similar sentences are found, `BERTopic` will form a topic characterized by representative keywords such as "Schrödinger," "well," "oscillator," and "potential," indicating that this concept is frequently discussed across the syllabi in the corpus.

Syllabi were collected as part of a larger, previously conducted study [7]. The subset of syllabi utilized for this analysis was collected from 50 US institutions. In the 2021-2022 academic year, these institutions collectively granted 1,558 bachelor's degrees, representing 18.5% of the total bachelor's degrees in physics granted that year [14]. Of these 50 institutions, 84% are classified as having very high research activity and 14% are classified as having high research activity, according to Carnegie's classification of higher education institutions [15], with an institution whose research activity is not defined by these metrics. Furthermore, 70% of the institutions are public. Additionally, 14% are minority-serving institutions (MSIs, $n = 7$), including Hispanic-serving institutions (HSIs, $n = 6$), and Asian American and Native American Pacific Islander-serving institutions (AANAPISIs, $n = 2$) [16].

These 50 institutions required between one and four quantum-related courses in order to graduate with a bachelor's degree in physics. Eleven institutions required only one course, nineteen required two courses, another nineteen institutions required three courses, and an additional one institution required four courses. Therefore, the corpus analyzed comprised $n = 121$ syllabi documents. To compare the quantum curriculum from institution to institution, as in the previous study [7], documents for different courses taught at the same institution were combined into one single document so that each institution, $n = 50$, had one document associated with it. The 50 documents were then converted to plain text for further analysis using the `pdftotext` utility for PDF files and `pandoc` for the word documents.

This analysis utilized `BERTopic` [17], a relatively recent algorithm that builds on transformer-based language models. `BERTopic` begins by embedding the text using a pre-trained model such as `BERT` (Bidirectional Encoder Representations from Transformers), which captures contextual information by analyzing the position and meaning of words within a broader textual window. These embeddings are then reduced

TABLE I. Subset of topics combined into overall themes for coding syllabi, as used in [7]

Schrödinger equation	Early quantum experiments	3D quantum mechanics
Time independent Schrödinger equation	de Broglie wavelengths	Spherical coordinates
1D quantum mechanics	Blackbody radiation	Hydrogen atom (in context of
Solutions to different potentials	Planck's hypothesis	Schrödinger equation,
Infinite potential well	Photoelectric effect	not Bohr model)
Harmonic oscillator	Double slit experiment	Angular momentum
Bound states	Frank Hertz experiment	

TABLE II. Subset of themes found in prior undergraduate quantum curriculum study and frequency counts across syllabi, full table is available in [7]

Themes	Freq. (%)
Schrödinger Equation	98
3D QM	94
Formalism	78
Quantum Theory of Light	76
Uncertainty principle	76
Spin	76
Wavefunctions	74
Particles	70
Tunneling	70
Early QM Experiments	54
Perturbation Theory	46
Variational Principle	24
Scattering Theory	22
Adiabatic Approximation	18

in dimensionality to make the clustering more computationally tractable. The reduced representations are then grouped using `HDBSCAN`, a density-based clustering algorithm that determines the number of clusters automatically based on the density of the data points, rather than requiring it to be set in advance. Finally, representative keywords and example sentences are selected for each topic.

Following best practices for `BERTopic`, documents were segmented into sentence-like units using the `sent_tokenize` function from the `nltk` package. This approach ensures that each unit is short enough to be meaningfully embedded by the language model. Unlike traditional topic modeling methods that recommend removing stop words, `BERTopic` leverages these function words as part of its contextual understanding and therefore does not require their removal.

III. RESULTS

Running `BERTopic` on the dataset of 50 institutions' complete set of undergraduate quantum syllabi produced 70 topics. The algorithm provided the ten most associated words with each topic. Each document, which included the entirety of the respective institutions' quantum syllabi, was divided into sentence fragments, the classified groups then recognize and assign one topic per sentence fragment. The algorithm also provides the document from which the sentence fragment(s) came from, so we were therefore able to map the sentence fragment back to their parent syllabi document to compute the frequency with which each topic occurred in the original dataset. There were several topics identified by the algorithm that upon human supervision could be generalized and grouped together. Some of these topics were not directly related to the curricular content, but instead reflected administrative or structural aspects of the syllabi, such as textbook information, grading policies, or accommodations.

Table III illustrates an example of one such non-curricular topic groupings: references to textbooks. The table includes the original topics generated by `BERTopic`, the ten most associated words for each topic, and a prototypical example sentence. These example sentences were automatically generated by the algorithm and are representative of the center of each topic cluster. The ten combined most frequently found topics were *common words* (90%), *textbook* (90%), *exam/quiz policies* (88%), *assignment submission policies* (88%), *grades/grading policies* (86%), *academic misconduct* (82%), *course preliminaries* (82%), *accommodations* (78%), *attendance* (68%), and *learning goals* (66%).

These ten most commonly identified topics were concerned with the structural and administrative content of the syllabi. However, in the human-coded results, these elements were not coded, since the previous study [7] focused exclusively on curricular content related to quantum mechanics. This contrast highlights a key strength of `BERTopic`: its ability to reliably identify consistent surface-level features in documents, even without domain-specific training. At the same time, it unscores a limitation: `BERTopic` assigns equal weight to policy-related content and disciplinary content, even when only the latter is relevant to the research question. This result reinforces the importance of human supervision in filtering or redefining NLP output in studies that target complex disciplinary knowledge rather than document structure.

In contrast, Table IV presents the subset of AI-identified topics that are aligned more directly with curricular content in quantum mechanics. These were compared with previously established human-coded topics from [7] to assess thematic overlap and coverage of the two methods.

`BERTopics` identified 14 topics that corresponded to course content. By comparing the ten most relevant words for each topic, with the codes used for the human coding process, each NLP generated content topic was assigned to one or more human-coded topic. Three of the topics were

TABLE III. Topics identified using BERTopic and combined into the general theme of *textbooks* with human supervision through analysis of their associated keywords and example phrases.

Combined Freq.	**Interpretation**	**Individual Freq.**	**Topics Identified by BERTopic**	**Representative Example Phrases**
45 (90%)	Textbook	37	textbook, textbooks, lectures, lecture, reading, classroom, books, readings, lecturing, sections	"Readings since lecture activities will consist of the students working individually or in groups to solve problems it is vital that you prepare for class in ..."
		36	quantum, textbook, books, textbooks, introductory, book, introduction, griffiths, library, isbn	"Textbook introduction to quantum mechanics by David J Griffiths and Darrell F Schroeter rd ed Cambridge University Press"
		28	textbooks, textbook, books, fundamentals, texts, tiplermodernphysicsec, library, isbns, book, wiley	"[redacted] version page required texts materials Modern Physics for Scientists and Engineers nd edition by John R Taylor University Science Books ISBN ..."
		11	feynman, lectures, books, book, topics, lecture, [redacted], quantum, physics, library	"Other books that may be of interest R P Feynman R B Leighton and M Sands The Feynman Lectures on Physics Vol"
		6	wileyplus, wiley, textbook, bookstore, registration, assignments, pdf, [redacted], edition, purchasing	"If the new textbook does not come with a WileyPlus registration we should be able to request a WileyPlus registration if we show proof that a ..."

not assigned to human-coded topics, as they included content related to relativity or nuclear physics, and the human-coding process was only concerned with quantum related topics. Additionally, a new theme was generated using human interpretation based upon the ten more relevant words that BERTopic identified. These 14 topics can be seen in Table IV.

IV. DISCUSSION

In Table IV, we compare the frequencies between human-coded topics and AI-generated topics by aligning the ten most relevant AI-identified words with previously defined human-coded topics. However, this comparison introduced an inherent asymmetry. Although human coding treated topics like "Schrödinger equation" as a single category, AI-derived topics often contained the word "Schrödinger" in multiple, distinct clusters, each corresponding to a different conceptual context. This is likely a reflection of how BERTopic disambiguate terms based on the surrounding context, effectively treating each usage of "Schrödinger" as its own semantic instance. As a result, the same human-coded topic may appear in several AI-interpreted themes, leading to what may seem to be an inflated combined frequency.

We acknowledge that this difference in granularity complicated direct comparisons. Rather than interpreting the "combined frequency" as a precise match, we present it as an indicator of overlap between the sets of concepts represented by each method. This framework reflects the strength and limitation of each approach: Human coding applied broader, discipline-informed categories, while AI identified finer-grained, context-sensitive clusters. Future work may benefit from developing a more nuanced alignment framework to better bridge these differences.

From the last column of Table IV, it is shown that the ten most relevant words identified by BERTopic could be assigned a coherent human-interpreted theme for 13 of the 14 identified topics. Furthermore, it is noted that BERTopic identified three different topics that could be more broadly defined as topics within modern physics [10]. Human analysis did not identify these topics, as the objective of the previous paper was to code topics taught on quantum concepts throughout the four-year undergraduate curriculum [7]. Therefore, topics related to relativity, atomic and nuclear physics were ignored during human coding.

The remaining ten topics identified by BERTopic were assigned the more general categories of: *spin*, *formalism*, *tunneling*, *Schrödinger equation*, *types of particles*, *uncertainty principle*, *Schrödinger equation*, *simple harmonic oscillator*, *wave-particle duality*, and *perturbations*. It can therefore be drawn that the AI model did produce a list of coherent and relevant topics to the undergraduate quantum curriculum, although the themes of the topics are not in agreement with those chosen by human analysis.

V. CONCLUSIONS

We find BERTopic to be well suited to identify structural or administrative features in syllabi, such as grading policies or textbook references, but less effective in extracting complex, discipline-specific content such as quantum mechanics topics. This limitation matters because automated tools

TABLE IV. Comparison of human-coded topics and AI-identified relevant words. Column 1 lists topics from the previous human-coded analysis [7] that align with the ten most relevant words generated by the AI. Column 2 shows their frequency in the human-coded data. Column 3 shows how often all the associated human coded topics appeared in the syllabi according to the AI-generated word sets. Column 4 reports the frequency of each AI-identified topics, and column 5 lists the ten most relevant words. Column 6 provides a human interpretation of each AI generated topics.

Human Analysis			**AI Analysis**		
Human coded topics from [7]	**Freq. (%) [7]**	**Combined freq. (%)**	**Freq. (%)**	**Ten most relevant words**	**Interpretation of AI Identified Topic**
3D quantum mechanics	94	98	42	spin, spinors, rotations, rotational, orbital, angular, gyromagnetic, rotate, momenta, proton	Spin
Spin	76				
Formalism	78	78	40	eigenfunctions, hermitian, observables, operators, observable, eigenstates, operator, unitary, hermiticity, hermite	Formalism
Schrödinger equation	98	100	40	potentials, scattering, wells, quantummechanical, harmonic, tunneling, potential, particle, tunnelling, amplitude	Tunneling
Tunneling	70				
Particles	70				
Not identified			40	topics, relativity, particles, schrodinger, atomic, particle, [redacted], cosmology, matter, waves	Modern physics topics
Not identified			36	relativity, relativistic, lorentz, einstein, dilation, spacetime, clocks, morley, velocity, frames	Modern physics topics
Schrödinger equation	98	100	36	schrodinger, quantum, solutions, dimensional, potentials, dimension, wave, equation, schrodinger, threedimensional	Schrödinger eq.
Wave functions	74				
3D quantum mechanics	94				
Particles	70	70	34	fermions, bosons, symmetries, particles, fermion, symmetry, wavefunction, symmetrization, antisymmetric, particle	Types of particles
Uncertainty principle	76	98	32	uncertainty, schrodinger, heisenberg, observables, quantum, expectation, wavefunction, probabilistic, variance, momentum	Uncertainty
Schrödinger equation	98				
Formalism	78				
Wave functions	74				
Schrödinger equation	98	98	32	schroedinger, eigenfunctions, wavefunctions, wavefunction, schro, schrdinger, eigenvalue, stationary, dinger, eigenavlues	Schrödinger eq.
Formalism	78				
Wave functions	74				
Schrödinger equation	98	98	32	oscillators, oscillator, harmonic, operators, perturbative, operator, equations, solutions, solvable, coupled	SHO
Formalism	78				
Perturbation theory	46				
Early quantum experiments	54	98	30	planck, broglie, schrodinger, waves, wave, wavelike, particles, particle, atom, wavelengths	Duality
Schrödinger equation	98				
Wave functions	74				
Quantum theory of light	76				
Perturbation theory	46	90	32	perturbations, perturbation, perturbative, perturbed, energies, wavefunctions, hamiltonian, nonhw, iterative, expansion	Perturbation
Wave functions	74				
Formalism	78				
Not identified			22	fission, nuclei, nucleus, nucleosynthesis, nuclear, radioactive, atomic, fusion, radioactivity, reactions	Modern physics topics
Perturbation theory	46	92	12	entanglement, decoherence, approximations, particles, variational, perturbation, adiabatic, scattering, theory, phases	Not identified
Particles	70				
Adiabatic approximation	18				
Variational principle	24				
Scattering theory	22				

like `BERTopic` are increasingly being used in education research, and it is critical to understand both their capabilities and their blind spots. In PER, where curricular nuance and conceptual depth are central to our research questions, relying on general-purpose language models may lead to misleading conclusions if not combined with human interpretation.

This study highlights the need for caution when applying NLP to technical educational data and points toward the importance of developing or adopting domain-specific models, such as `PhysBERT`, that are better equipped to handle the semantic complexity of physics instruction. More broadly, our findings suggest that hybrid approaches that combine human expertise with AI tools can enhance scalability while preserving disciplinary accuracy. As PER continues to grow in scope and methodology, this work contributes to a foundational understanding of how to responsibly integrate NLP in curriculum analysis and other text-based studies.

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, International Journal of Educational Technology in Higher Education **16**, 1 (2019).

[2] J. Wilson, B. Pollard, J. M. Aiken, M. D. Caballero, and H. J. Lewandowski, Phys. Rev. Phys. Educ. Res. **18**, 010141 (2022), URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.18.010141.

[3] T. O. B. Odden, A. Marin, and M. D. Caballero, Phys. Rev. Phys. Educ. Res. **16**, 010142 (2020), URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.16.010142.

[4] P. Tschisgale, P. Wulff, and M. Kubsch, Phys. Rev. Phys. Educ. Res. **19**, 020123 (2023), URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.19.020123.

[5] G. Kortemeyer, Phys. Rev. Phys. Educ. Res. **19**, 010132 (2023), URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.19.010132.

[6] L. Parson, The Qualitative Report **21**, 9 (2016), URL https://nsuworks.nova.edu/tqr/vol21/iss1/9/.

[7] A. Buzzell, T. J. Atherton, and R. Barthelemy, Phys. Rev. Phys. Educ. Res. **21**, 010102 (2025), URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.21.010102.

[8] D. J. Griffiths, *Introduction to Quantum Mechanics* (Cambridge University Press, 2005), 2nd ed., ISBN 978-0521537369.

[9] R. McIntyre, *Quantum Mechanics: A Paradigms Approach* (Wiley, Hoboken, NJ, 2017), ISBN 978-1-119-29451-3.

[10] A. Buzzell, T. J. Atherton, and R. Barthelemy, Phys. Rev. Phys. Educ. Res. **21**, 010139 (2025), URL https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.21.010139.

[11] A. Abdelrazek, Y. Eid, E. Gawish, W. Medhat, and A. Hassan, Information Systems **112**, 102131 (2023).

[12] R. Mitkov, ed., *The Oxford Handbook of Computational Linguistics* (Oxford University Press, Oxford, 2021), 2nd ed., ISBN 9780199573691.

[13] I. Vayansky and S. A. Kumar, Information Systems **94**, 101582 (2020).

[14] S. Nicholson and P. Mulvey, *Roster of physics departments with enrollment and degree data* (2021), accessed: 2024-07-22.

[15] *Carnegie classification of institutions of higher education*, URL https://carnegieclassifications.acenet.edu/.

[16] *University of california, irvine, msi directory*, URL https://inclusion.uci.edu/core-programs/minority-thriving-campus/msi-directory/.

[17] M. Grootendorst, arXiv preprint arXiv:2203.05794 (2022).