The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)

Word2Vec4Kids: Interactive Challenges to Introduce Middle School Students to Word Embeddings

Nathan Wiatrek [1, 2], Yash Verma [1], Fred Martin[1]

[1]University of Texas at San Antonio
[2]Southwest Research Institute
nathan.wiatrek@swri.org, yash.verma@my.utsa.edu, fred.martin@utsa.edu

Abstract

As Artificial Intelligence (AI) continues to integrate into more aspects of society, equipping younger generations with foundational AI knowledge becomes increasingly critical. This paper presents Word2Vec4Kids (W2V4K), an interactive application designed to familiarize middle school students with word embeddings, a key aspect of Natural Language Processing (NLP). W2V4K leverages the Word2Vec model, allowing students to explore word associations, similarity, and vector arithmetic through engaging game modes. The application was tested with 38 middle school students aged 11–14 at a Science Technology Engineering Math (STEM)-focused charter school. Data were collected on students' interactions with the application, including screen recordings, audio, and survey responses. Results demonstrated that W2V4K effectively introduces NLP concepts to students. Qualitative observations revealed high levels of engagement with students expressing excitement and curiosity about word relationships. As they progressed through the game modes, students showed increasing confidence in predicting word associations, brainstorming relevant words, and connecting the concepts to real-world applications. Quantitative data from post-interaction surveys indicated positive learning outcomes with 44.5% of students achieving perfect scores on concept-related items. Additionally, students demonstrated an ability to critically think about language representation. This study suggests that W2V4K provides an effective and engaging method for introducing NLP concepts to middle school students, contributing to the broader goal of enhancing AI literacy among younger generations.

Introduction

Artificial Intelligence (AI) based technologies have integrated into many aspects of society. AI algorithms play a significant role in content curation, personalized recommendations, and targeted advertising on social media platforms like TikTok, Instagram, and Facebook (Akgun and Greenhow 2022). As social media platforms are ramping up AI usage, younger generations are increasing their social media usage, with 70% of teenagers using social media multiple times a day vs. 33% of teenagers in 2012 (Abi-Jaoude, Naylor, and Pignatiello 2020). This indicates that young people are interacting with AI technology at an astonishing rate.

Given this increasing prevalence of AI in daily life, developing AI literacy for younger generations has become increasingly important. AI literacy enables individuals to critically evaluate and effectively utilize AI technologies both at home and in the workplace (Long and Magerko 2020). Being able to understand what the purpose of a specific AI system is, and how it uses data to make a prediction/recommendation can help users understand the result. This understanding can also help a user select an AI system and understand how to utilize it in a work environment to assist in solving complex problems.

Recognizing this need for AI literacy, Touretzky et al. proposed five ideas as guidelines for teaching AI concepts to children (Touretzky et al. 2019). Despite the increasing importance of AI literacy, there remains a need for effective and engaging tools to teach AI concepts in a manner consistent with the big ideas. To address this need, we developed Word2Vec4Kids (W2V4K), a novel education tool that empowers children to learn about Natural Language Processing (NLP) and focuses on the representational aspect of big idea 2, Representation and Reasoning. Specifically, we focus on the second Learning Objective for middle school students which is to "[explain] how word embeddings ... represent words as sequences of numbers" (AI4K12 2020).

W2V4K introduces children to this fundamental concept of Machine Learning (ML), with word embeddings. Word embeddings are high dimensional vector representations of words that capture semantic relationships in a way that simpler representations cannot. These embeddings are trained on large corpus of text, allowing computers to understand the relationship between words. The word embedding model serves as a representation of language that the computer maintains and uses for reasoning in tasks such as finding similar words and understanding semantic relationships. Language is an intuitive domain for AI education for students since they do the act of natural language processing every day in conversations.

W2V4K teaches NLP concepts by explaining the role the model Word2Vec and word embeddings have for ML tasks to middle-school-aged students. This paper demonstrates that W2V4K effectively introduces students to key NLP principles, resulting in positive engagement, despite no prior NLP knowledge for the majority of participants.

Related Work

AI education for K12 students is a significant field with over 300 studies published since 2010 (Heeg and Avraamidou 2023). Many of these studies have focused on awareness and ethical usage of AI. Some studies have focused on using robots to bring AI into the physical world for children to use (Karalekas, Vologiannidis, and Kalomiros 2023; Chiu et al. 2022). While physical robots and machinery provide tangible interactions, allowing students to observe how AI can interact with the world, they can also be untenable due to high costs or limited teaching space. Digital applications, particularly those focusing on NLP, have emerged as another way to help students interact with machine learning in a way that isn't cost or space prohibitive.

There have been many studies focused on NLP and word embeddings as an aspect of AI to teach, likely due to how intuitive it is for children to understand processing of words. NLP has gained increasing prominence in AI education, comprising over 7% of all work in the field of AI education, and continues to grow (Fitria 2021). Many studies focusing on NLP have utilized web applications as the medium for teaching AI concepts. Bandyopadhyay et al. (Bandyopadhyay et al. 2022) provides a comprehensive list of online demos regarding word embeddings, and their focuses.

Bandyopadhyay et al. found that 10 out of 12 demos covered nearest neighbors, 7 covered analogies with vectors, 6 covered plotting a vector embedding in a 2D/3D view, and 4 covered word similarity. These 4 topics give the student a deeper understanding of what word embeddings are and how they can be used. Building upon the themes in existing tools, W2V4K aims to provide a comprehensive learning experience that engages the user and covers the most prevalent topics in the field.

In addition to a comprehensive list of online demos, Bandyopadhyay et al. contributed their own graphical display of embeddings. There are many similarities between (Bandyopadhyay et al. 2022) and W2V4K. Both target school aged children and focus on educating the user about what word embeddings are. Both use the Word2Vec model originally proposed by (Mikolov et al. 2013) and attempt to display this information in a 2D/3D manner. The reduction of dimensions between Bandyopadhyay et al. and W2V4K is one of the primary distinguishing factors. Bandyopadhyay et al. reduces the dimensions of Word2Vec by using two dimensions to represent user selected semantic features, and utilizes a residual dimension to maintain semantic relations between similar words. W2V4K utilizes Principal Components Analysis (PCA) for displaying the embedding in two different situations, biased and unbiased, which will be discussed further in the methodology section.

Another NLP tool used in education is NaturalLanguageProcessing4All (NLP4All) by Arthur Hjorth (Hjorth 2021). The focus of NLP4All was to teach high school children about NLP in an interactive way, creating text classifiers that correlated with a social studies learning unit. This helps to show students how NLP can be used at a higher level, and give them a deeper understanding of the technology.

Beyond these differences W2V4K also distinguishes itself in the medium for learning selected. All of the demos found by Bandyopadhyay et al. focused on web applications, while W2V4K is a native MacOS application. This is relevant because 90% of teenagers own an Apple device (Sattelberg 2023), and therefore, would likely be familiar with the patterns of MacOS applications over that of web applications. Some of these features include offline accessibility, application performance, and intuition of how to interact with an Apple based application, enhancing the learning experience.

Research Study

The W2V4K application was developed as part of a broader study investigating the effectiveness of teaching AI concepts to children. This study was conducted over two separate 1.5-hour sessions, involving distinct groups of students for each session. The study had a total of seven AI centered application available to the over 120 participants.

The study population consisted of middle-school students aged 11–14 from a Science, Technology, Engineering, and Mathematics (STEM) focused charter school in a major Texas city. All computer hardware was provided by the researchers, as such there were three MacOS workstations were available for student use. Of the 41 students who interacted with the W2V4K project, 38 consented to their data being used in the study. Participation required signed consent forms from guardians and students.

Each of the two research sessions consisted of three phases:

1. Pre-interaction survey: Participants completed a questionnaire designed to assess their initial interest in and understanding of AI concepts.
2. Interaction phase: Each student was allotted 5–10 minutes to interact with the W2V4K application.
3. Post-interaction survey: Following the interaction, students completed a post-survey to evaluate changes in their perceptions and understanding.

Methodology

W2V4K consists of six screens for students to interact with and learn from. The name, description, and picture of each screen can be seen in Table 1. The Tutorial, Emotion Explorer, and Similarity Slider screens give the children an overview of what word embeddings are and how they can be used. The Neighbor Navigator, Word Quest, and Word Alchemy screens give students agency to interact with and explore the word embedding.

Implementing Word2Vec

There were multiple engineering challenges to implementing Word2Vec in the Swift programming language for use on Apple hardware. The first challenge was getting the vector representation to load into memory. Apple provides a MLWordEmbedding package where you load a list of vectors associated with words and it will determine the nearest neighbor (Apple 2024b). Taking the accepted word2vec vector list and putting it through the MLWordEmbedding package gave questionable results, such as *later* being a nearest neighbor to *king*.

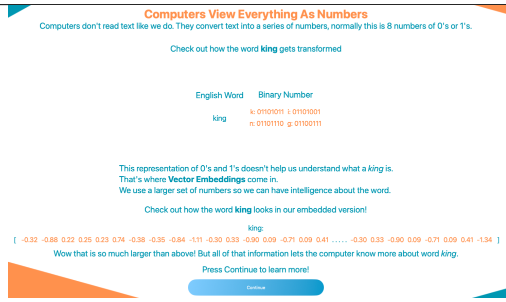
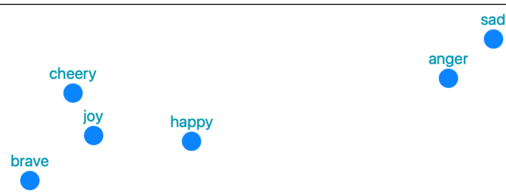
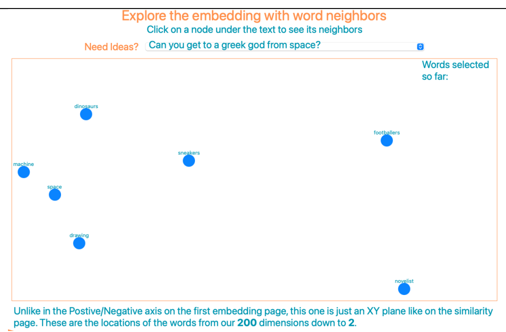
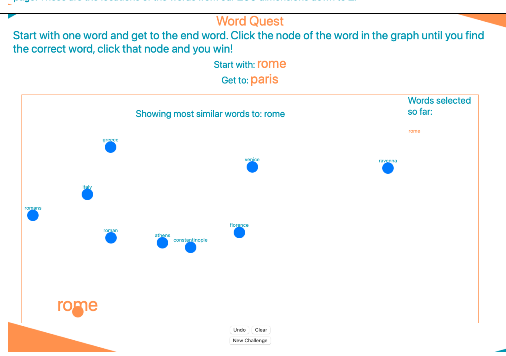
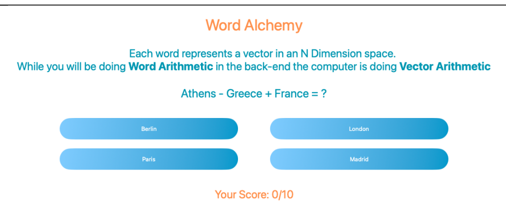
Screen Name	Picture	Description
Tutorial		The first screen a student see's when using W2V4K. This screen introduces the students to how computers see text with number representations. This is first done by taking the word "king" and translating to ASCII binary, then taking "king" and transforming it to the word2vec representation of the word. The students are directed to notice the difference in size between the two representations of the same word.
Emotion Explorer		We reduced this embedding from 200 to two dimensions using PCA. This embedding was a biased embedding to show a more stark contrast between positive and negative emotions. The embedding was biased in a manner similar to (Chanin 2021), in which we created an emotion vector by applying PCA to pairs of positive and negative emotions. The students are instructed to enter various emotions to see where they land on the screen
Similarity Slider		Similarity Slider introduces the concept of word similarity to the children. A single word is entered and 15 nearest neighbors were found, the students then move a slider that would represent words being "similar enough" to each other for their use case.
Neighbor Navigator		Students are given an initial set of words and can click on each of the words to discover that words ten nearest neighbors. Multiple prompts were provided to create goals and help students explore the embedding. These prompts were a mix of open-ended and goal oriented, providing direction for the student if they wished to use it. Some of the prompts included were: "Can you get to a greek god from space?", and objective prompt, and "Can you get to your favorite animal from dinosaur?", which was open-ended.
Word Quest		Word Quest has a similar interface to Neighbor Navigator, however there were concrete goals for, quests, for the students to achieve. For each quest the student clicked on the starting word, and was presented with the 15 nearest neighbors, by cosine similarity in the 200-dimension embedding. The student has to continue this pattern to get to the end word. For example, in one of the challenges, the start *woman* and the end is *queen*, one path is to go from *woman* to *lady* to *queen*, as *queen* is not one of the 15 nearest neighbors to *woman*, being a minimum of two neighbors away.
Word Alchemy		This screen asks students to figure out what a resulting word would be given a prompt. These prompts take the form of equations such as "Athens - Greece + France = ?" and gives the student four options to select from.

Table 1: W2V4K's Screens

These results were especially questionable when compared to the widely accepted word2vec application `gensim` (Řehůřek and Sojka 2010). To get the correct nearest neighbors, we created a new Swift package called WordEmbedding that takes in a list of words and their vectors and returns results that are more closely related to the `gensim` output.

Nearest neighbors are defined as a set of words that are most similar to a given word or vector. The nearest neighbor of a word can be defined using two methods, Euclidean distance and cosine similarity. The Euclidean distance is the distance of a straight line between two points in any dimension, and can be calculated with Equation 1. The smaller the euclidean distance, the closer the two vectors are to each other and therefore are closer neighbors.

$$D_{ij} = \sqrt{\sum_{v=1}^{n} (X_{vi} - X_{vj})^2} \tag{1}$$

Cosine similarity is calculated by taking the dot product of the vectors divided by their magnitude. This equation can be seen in Equation 2. The resulting value can be between -1 and +1, with the closer the two vectors are to having the same orientation, the higher the value is.

$$similarity(A, B) = cos(\theta) = \frac{A \cdot B}{\|A\| \, \|B\|} \tag{2}$$

There are strengths and weaknesses to both equations for finding neighbors of words. Implementing both in an optimized way allows the end user to select the correct equation for their use case. To optimize equations 1 and 2 Apple's Digital Signal Processing library (Apple 2024a) is used. This library contains highly optimized math functions, such as the sum of squares, for Apple's hardware.

To validate our algorithm's results, we compared nearest neighbors of words in our application as well as with `gensim`. To do this the same initial word was given to each system and the nearest neighbors were checked for correctness. We also compared the results of vector addition and the resulting nearest neighbors with those from `gensim`. As an example, the equation "king + man" was given to both systems. `Gensim` returned the words, in order, "king, man, father, brother, son", while our application returned back "king, father, brother, son, man."

Another engineering challenge in creating W2V4K was adapting a high-dimensional vector space with the two-dimensional screen displays. We partially addressed this using Principal Component Analysis (PCA) to reduce the 200-dimensional space to two dimensions. However, applying either Equation 1 or 2 to this reduced set yielded poor results due to insufficient separation in the lower-dimensional space. To overcome this, we implemented a two-step process: first, we ran the nearest neighbors algorithm on the high-dimensional vector set to identify relevant words and their corresponding vectors. Then, we mapped these words to their 2D representations to display on the screen.

The W2V4K application employs the singleton pattern to optimize memory usage while prioritize fast data retrieval by loading the embedding information in Random Access Memory (RAM). The application was developed for Mac computers but could be adapted to fit on iOS devices. The memory size of the application is 1 gigabyte, well below the minimum amount of RAM Apple computers and cell phones, are currently sold with. This memory utilization ensures compatibility across various hardware configurations.

Word2Vec Model Adaptation

The original Word2Vec model (google 2013) contains three million vectors that represent words, phrases, and named entities in a 300-dimension embedding. This model was found to be unsuitable for our application due to two problems stemming from its large size. The first problem was that the original model was too large to run necessary algorithms, such as cosine similarity, in a timely manner on a resource-constrained laptop. The second problem was that the large number of vectors known by the model proved detrimental as outputs since many results would be words not in a middle school student's vocabulary. After performance and accuracy testing, we found that 200 dimensions was an appropriate number of dimensions for accurately representing words while also being more performant. We also used a slimmed down vector set (Daniel, Nikhil, and Nicholson 2019; tensorflow 2016) that had 70,000 vectors, 2.3% of the of the original model. This model was further tuned by removing words that were found in CMU's offensive/profane word list (von Ahn 2009), to keep the application school age appropriate.

Game Modes

The games section of W2V4K presents the students with three game modes: Neighbor Navigator, Word Quest, and Word Alchemy. Neighbor Navigator's initial set of words was created with two goals in mind, the first was having topics that would be interesting and enjoyable for children. The second goal was to have a diverse set of endpoints, meaning there was at least one word in a quadrant of a graph for the 2D word embedding. With these two goals in mind, the final words were chosen to be: machine, space, dinosaurs, drawing, sneakers, footballers, and novelist.

Word Alchemy, had children explore vector addition with words. In this game mode children were asked what word would be the result of additions and subtractions of various words. One such question is what is the resulting word from the following equation "Athens - Greece + France". The answer to this question is the word "Paris", which is found due to capitals of countries having similar distances in the vector-space regardless of the country. There were 11 total questions for the children to interact with, at the end of each question the students were presented with the correct answer.

These three game modes were developed based on their ability to engage the user and display useful properties of a vector embedding.

Data Collection

Data gathered consisted of three parts: the pre-survey, student interaction with W2V4K, and the post-survey.

Question	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree
I believe AI has the potential to positively impact our daily lives	17	17	1	1	0
I am worried about bias in how AI systems make decisions	3	12	16	2	3
I am excited about the possibilities that AI could bring to different careers	16	10	4	4	2
I am scared of AI	1	15	7	5	8
I am planning to major in a STEM field in college	11	8	12	1	4

Table 2: Pre survey multiple choice questions and responses

Pre-Survey The pre-survey asked a series of questions on their background, and their awareness and comfort level with AI. The AI sentiment questions and their responses can be seen in Table **??**. After these questions, the students were asked if they had interacted with any AI or if they knew concepts that would be covered in the activity, namely if they knew anything about vector representation of language.

Application Interaction Data The second set of data, students interaction with the application, included screen and audio recordings as well as all meaningful interactions (e.g., selecting or typing words) stored in a SQLite database for post-experiment analysis.

Post-Survey The third set of data, post-survey, was used to test students' understanding of word embeddings. It asked multiple-choice and open-ended questions tied to the application's screens and game modes that exposed relevant concepts. In the post survey, there are four multiple choice questions, along with two open ended questions for the student's to provide input and reflect on their learning.

Results

The two major outcomes of W2V4K is high levels of student engagement and student understanding of word embeddings. Through a combination of quantitative application usage date, qualitative observations, and post-survey results, we found that students not only enjoyed interacting with W2V4K but developed a substantive understanding of word embedding concepts.

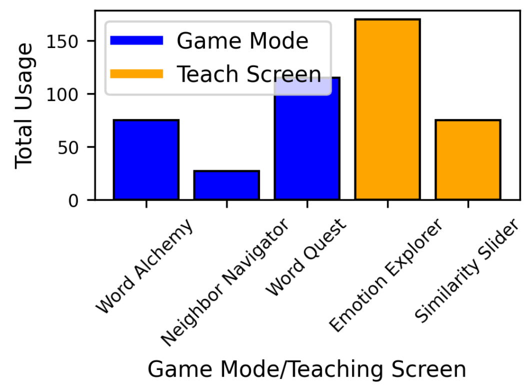


Figure 1: W2V4K Usage by screen

Student Engagement

Students showed high levels of engagement with W2V4K during their 10 minutes with the applications. The total usage of W2V4K can be seen in Figure 1. Out of the three game modes Word Quest was the most played, with 75% of the students having attempted at least one of the challenges. The engagement was evident with multiple students giving enthusiastic reactions, such as "woah!" during the PCA visualization and "Yeah, it was actually fun!" as a summary of their time with the application. Some students preferred different game modes or teaching screens than others, but when a user found one they enjoyed they would spend most of their time on it, usually laughing throughout the process.

As the students used the application more, their understanding of word embeddings grew. Students were quickly able to grasp where words would appear in the biased PCA visualization, with groups of students successfully predicting the area that the next word they entered would appear. Through usage of the biased PCA the students showed comprehension of word clustering, commenting how words that were visually clustered were often synonyms or antonyms.

Analysis of student interactions revealed diverse word exploration patterns, demonstrating growing engagement with the concept of word embeddings. Students initially focused on common words (e.g., "sad", "mad"), but progressively introduced more complex and nuanced terms (e.g., "sorrow", "ecstatic"). Additionally, some students attempted to input a colloquial term, such as "hangry," indicating a desire to connect the technology to their lived experiences.

Multiple students paired antonyms (e.g., "jolly" and "grumpy") and commented on their relative positions in the embedding space. This behavior aligns with the increased accuracy in predicting locations of words in the visualization. Students also had a curiosity towards words that had no emotional context such as their names or a number. Students remarked on how these words were in the middle of the embedding indicating they understood there was little to no emotional bias with these words and numbers. These exploration patterns reflect an increased engagement while also suggesting that students were constructing mental models of how word embeddings capture semantic relationships.

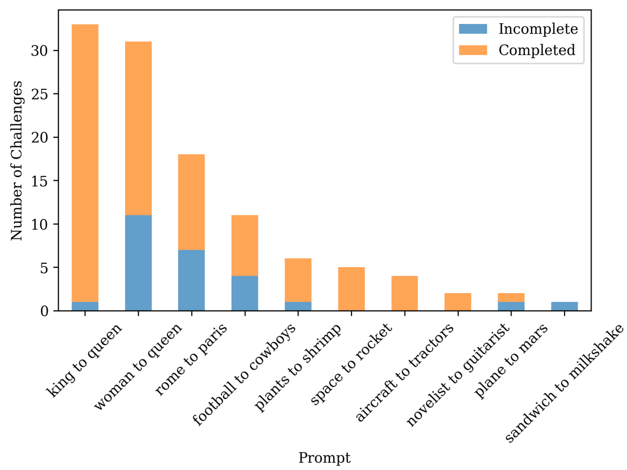


Figure 2: Challenge completion by prompt

In Word Quest, student confidence and understanding grew noticeably. Initially, many students had questions on how feasible the challenge was and expressed uncertainty, such as "How are we supposed to get to queen [from woman]." By later challenges, they were making predictions on paths to get to the final word such as "[I'll] do Italy and [W2V4K will] show France, and that'll have Paris," and "[I'll] do boy and see if [I] can get prince." These moments show an understanding of the semantic relationships that word embeddings can capture. Figure 2 shows student interactivity with challenge mode; students had less time for later challenges and thus completed fewer.

Students in the study were able to connect W2V4K to other AI applications that are widely used, one student asked, "ChatGPT uses this right?" when discussing word similarity. The students also understood the complexity of the problem of language representation and semantic capturing, commenting on how much larger the vector representation of a word was compared to its binary representation. These help to underscore that students were able to not just learn what the app was trying to teach, but conceptualize how it could be used outside of the single application.

The students were able to help teach other students concepts presented by W2V4K. In a notable interaction between a pair of students, the first said "I don't get [this]." The second responded explaining "whenever we write one word it'll appear up here [gestures in a region]." Immediately after they wrote the word "amazing" and remarked how positive that word was. Students helping others learn is vital to teaching students these concepts, being able to contextualize information in a way no application can.

W2V4K also provided learning opportunities from misconceptions. In Similarity Slider one student was confused about the word "smart" being near the word "dumb" in the embedding. This provided an opportunity to spark discussion about how the Word2Vec model was trained to the student. The student was able to recognize that in many sentences replacing "smart" with "dumb" will alter the meaning of the sentence, but still result in an understandable sentence.

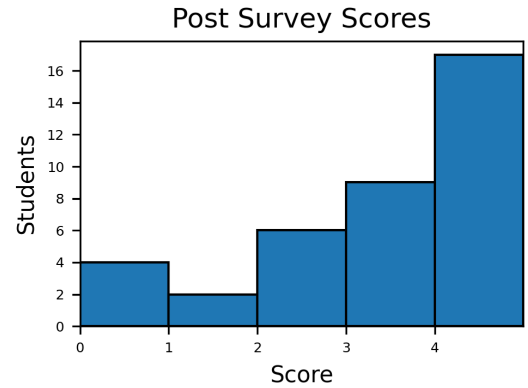


Figure 3: Number of students with each score

NLP Understanding

The post survey consisted of four multiple choice questions, that had objective answers, and two open-ended questions. The responses to the four multiple choice questions resulted in the score for the student. The scores range from a 4, meaning every question was correct, down to 0 where no question would be correct. The distribution of scores can be seen in Figure 3. This distribution shows that the majority of students were able to have high scores. Students did exceedingly well on questions where the answers were shown to them through the learning half of W2V4K (question 1 and 2) or a prominent part of each of the games (question 3).

One post-survey question asked the students to apply their understanding of word embeddings to determine the placement of a given word. Students were presented with a quadrant graph (Figure 4) and asked to determine where the word "cow" would be placed. This question required students to build upon their experience with the Emotion Explorer screen. The results of this question were encouraging: 87% of students correctly identified that "cow" would be placed in the bottom-right quadrant, demonstrating their ability to apply what they have learned. While 10% of students said the word would go in the top-right quadrant and 3% in the bottom-left, no student chose the top-left quadrant. This suggests most students grasped the underlying concepts, even if some showed uncertainty in their application.

When asked to describe the benefits of vector embeddings over binary representation, the results were mixed. 55% of the students either left this question blank or indicated that they did not know the answer, and one student (3%) provided an incorrect interpretation. However, 42% of the students demonstrated understanding consistent with the benefits of Word2Vec. These correct responses ranged from basic recognition that "it helps the computer gain more information about the word" to more advanced understanding, with some students noting that Word2Vec "gets words closely related to [a given word]" and can make "[resulting word queries] more accurate and filtered." This range of answers from basic to more nuanced understanding suggests that students were able to not only absorb the provided information but also extrapolate its implications.

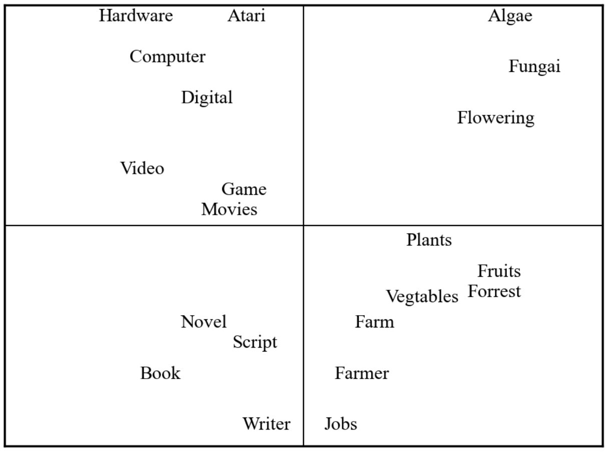


Figure 4: Post Survey Question - Where would “cow” land?

Grade	Size	Mean	Median	IQR	Std. Dev.
6	13	2.31	3	4.00	1.70
7	15	3.40	4	1.00	0.99
8	10	2.80	3	1.75	1.03
All	38	2.87	3	2.00	1.33

Table 3: Score Statistics by Grade

The high rate of non-responses to the open-ended question about vector embeddings may be partially attributed to time constraints at the end of the session. Many students were deeply engaged with the various applications and continued exploring the game even when prompted to complete the post-survey. This enthusiasm for the interactive elements of W2V4K, while positive, resulted in some students rushing through or skipping parts of the survey.

Despite this, the responses to multiple-choice questions suggest that students who did complete the survey took it seriously. To assess the impact of W2V4K on student understanding, we analyzed the post-survey results across different grade levels (see Table 3). We found no significant difference in performance between the three grades studied. This lack of variation suggests that the application was equally accessible and effective for all age ranges of the students.

We also examined the relationship between application usage and post-survey scores but found no strong correlation. This could be due to the self-selecting nature of our sample, students who volunteered for an AI experiment may already have had a higher baseline interest or aptitude for learning ML topics.

To ensure that the high scores were not simply a result of random guessing on the multiple-choice questions, we calculated the probability distribution of scores that would result from chance and compared it to the actual distribution of student scores (see Table 4). The stark difference between these distributions provides strong evidence that students were not guessing randomly but were indeed demonstrating learned knowledge in their responses.

Score	Probability of Score	Students with score
0	15.0625%	10.5%
1	37.5%	5%
2	34.375%	16%
3	12.5%	24%
4	1.5625%	44.5%

Table 4: Comparison of Score Probability and Results

These findings collectively suggest that while time constraints may have impacted responses to open-ended questions, the multiple-choice sections of the post-survey provide reliable data on student understanding outcomes. The application appears to have been effective across grade levels, though further research with a more diverse sample might be needed to confirm this trend.

Limitations and Future Directions

As part of a larger research study, student time for interacting with W2V4K was constrained. While the allocated time was sufficient to assess student engagement with the application, it did not provide students with enough opportunity to fully explore the entire application. Being part of a broader research study also limited the pre-survey. In a future study, we would include pre-study items that are directly related to W2V4K, including the one shown in Figure 4.

This iteration of W2V4K was very promising, with students going from being unfamiliar with word embeddings to half of the group getting a perfect score on the objective questions. Students described their experience with W2V4K as very positive and fun. Many of the students were highly engaged, but reactions were varied on game modes. Some students were absorbed into exploring word relationships and the visualization of it, while others showed less interest specifically in the explore game mode. These different reactions highlight areas for potential improvement, removing areas that were less used and less enthusiastically received by the students.

In the future, we would focus on bringing this application to more diverse schools and age ranges. This is due to the likelihood of bias in the student sample, as students that are more interested in AI topics may be more understanding of these topics. This specific consideration is important as over half of the students declared that the were strongly or somewhat strongly planning on majoring in a STEM field.

Conclusion

A novel application for teaching natural language programming concepts was developed for use by students at a partner middle school. While engaging with the application, students thought about language and how it is processed. Students explored the word embeddings with words that were meaningful to them such as “hangry” and connected their understanding of language to the dimensionality of word embeddings on the screen. The usage of W2V4K along with the post-survey scores indicate that it was an engaging application that effectively introduces students to NLP.

Acknowledgments

This material is based upon work supported in part by the National Science Foundation under Grant IIS-2112633.

References

Abi-Jaoude, E.; Naylor, K. T.; and Pignatiello, A. 2020. Smartphones, social media use and youth mental health. *Cmaj*, 192(6): E136–E141.

AI4K12. 2020. Big Idea 2 – Representation & Reasoning. https://ai4k12.org/big-idea-2-overview/. Accessed: 2024-09-06.

Akgun, S.; and Greenhow, C. 2022. Artificial intelligence in education: Addressing ethical challenges in K-12 settings.

Apple. 2024a. Apple Developer Documentation Accelerate vDSP. https://developer.apple.com/documentation/accelerate/vdsp/. Accessed: 2024-08-01.

Apple. 2024b. Apple Developer Documentation MLWordEmbedding. https://developer.apple.com/documentation/createml/mlwordembedding. Accessed: 2024-08-01.

Bandyopadhyay, S.; Xu, J.; Pawar, N.; and Touretzky, D. 2022. Interactive visualizations of word embeddings for K-12 students. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 12713–12720.

Chanin, D. 2021. Exploring Gender Bias in Word2Vec — chanind.github.io. https://chanind.github.io/nlp/2021/06/10/word2vec-gender-bias.html. [Accessed 30-04-2024].

Chiu, T. K. F.; Meng, H.; Chai, C.-S.; King, I.; Wong, S.; and Yam, Y. 2022. Creation and Evaluation of a Pretertiary Artificial Intelligence (AI) Curriculum. *IEEE Transactions on Education*, 65(1): 30–39.

Daniel, S.; Nikhil, T.; and Nicholson, C. 2019. Embedding projector - visualization of high-dimensional data — projector.tensorflow.org. https://projector.tensorflow.org/. [Accessed 30-04-2024].

Fitria, T. N. 2021. Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. In *Prosiding Seminar Nasional & Call for Paper STIE AAS*, 134–147.

google. 2013. Google Code Archive - Long-term storage for Google Code Project Hosting. — code.google.com. https://code.google.com/archive/p/word2vec/-. [Accessed 30-04-2024].

Heeg, D. M.; and Avraamidou, L. 2023. The use of Artificial intelligence in school science: a systematic literature review. *Educational Media International*, 60(2): 125–150.

Hjorth, A. 2021. NaturalLanguageProcesing4All: -A Constructionist NLP tool for Scaffolding Students' Exploration of Text. In *Proceedings of the 17th ACM Conference on International Computing Education Research*, 347–354.

Karalekas, G.; Vologiannidis, S.; and Kalomiros, J. 2023. Teaching Machine Learning in K–12 Using Robotics. *Education Sciences*, 13(1): 67.

Long, D.; and Magerko, B. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–16.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Řehůřek, R.; and Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA. http://is.muni.cz/publication/884893/en.

Sattelberg, W. 2023. Android isn't cool with teenagers, and that's a big problem.

tensorflow. 2016. GitHub - tensorflow/embedding-projector-standalone — github.com. https://github.com/tensorflow/embedding-projector-standalone/. [Accessed 30-04-2024].

Touretzky, D.; Gardner-McCune, C.; Martin, F.; and Seehorn, D. 2019. Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 9795–9799.

von Ahn, L. 2009. cs.cmu.edu. https://www.cs.cmu.edu/~biglou/resources/bad-words.txt. [Accessed 30-04-2024].