

LingoLoop: Final Report

I. Basic project information

Name of your project: Lingo Loop

Link to project website: <https://nets-2130.vercel.app/login>

Name of your teammates: William Bush, Ryder Sitcawich, Areeb Alam, Edmund Doerksen

One sentence description of your project:

Lingo Loop is a crowdsourcing platform that collects, refines, and quizzes users on generational meanings of slang terms, leveraging cross-generation contributions to produce a richer understanding of slang.

Logo for your project:



What problem does it solve?

Lingo Loop addresses the difficulty of understanding slang terms that vary in meaning across generations. Younger and older generations often have different interpretations of the same slang word. By collecting definitions from multiple generational cohorts, Lingo Loop clarifies slang meanings and helps bridge linguistic and cultural gaps.

What similar projects exist?

- Online slang dictionaries (like Urban Dictionary) where users submit definitions. However, these lack structured generational tagging.
- Lexicographical surveys (e.g., academic studies on language change) but these are not crowd-powered and are limited in scale.

What type of project is it?

- Human computation algorithm (crowd contributions for defining slang)
- A tool for crowdsourcing (collecting and comparing generational definitions)
- Possibly a business idea using crowdsourcing for linguistic insights

What was the main focus of your team's effort?

Something in between: We engineered a functioning system (front-end, back-end with Supabase) and conducted data analysis (comparing generational responses).

How does your project work? Describe each of the steps involved in your project. What parts are done by the crowd, and what parts will be done automatically.

1. **User Login:** Users log in and provide their birth year.
2. **Contribution Step (Crowd):** Users from different generations submit their definition and usage examples of slang terms.
3. **Aggregation (Automatic):** Once three targeted generations (Gen Z, Millennials, Gen X/Baby Boomers) have contributed definitions, the system automatically generates a multiple-choice quiz question using these definitions. There are four answer choices, three for each of the generation groups, and an option to select "I don't know."
4. **Quiz (Crowd):** Other users attempt to guess the correct meaning from the provided options, supplying user responses that help identify which definition resonates the most with them.
5. **Data Processing (Automatic):** The system aggregates quiz responses and presents statistics showing how different generations interpret the slang term.

See section XII for selected screenshots of these features.

Provide a link to your final presentation video.

 Nets 2130 Final Video.mp4

Which two sections below did you pick for your in-depth analysis?

We selected **Quality Control** and **Aggregation** for our in-depth analysis.

II. The Crowd

Who are the members of your crowd?

Our crowd consisted of classmates from the course as well as our friends and family members.

For your final project, did you simulate the crowd or run a real experiment?

Simulated crowd.

If the crowd was real, how did you recruit participants?

We recruited classmates through a course assignment ("Be the Crowd" HW) and asked our friends and family members to participate.

How many unique participants did you have?

We had approximately 27 unique participants.

III. Incentives

What motivation does the crowd have for participating in your project?

- Altruism: Volunteers were interested in contributing to a fun linguistic project.

- Enjoyment: Classmates and family members found it entertaining to share and guess slang meanings.
- Relationships: Our crowd primarily consisted of friends and family, so there was an incentive to contribute as a means of supporting a loved one

How do you incentivize the crowd to participate?

We relied on non-monetary incentives. Classmates participated as part of their course homework, ensuring that they spent time contributing and completing tasks. We also recruited volunteers: friends and family. Our volunteers were motivated by curiosity, the opportunity to see aggregated results, and an eagerness to help their loved ones with a final project. The interface was designed to be engaging as a means of motivating users to provide high-quality contributions.

If we were to use a real crowd, we would link our website on an MTurk task and solicit volunteers to create accounts, after which they could be paid on a per-contribution basis. Calculating fair and accurate compensation for our Turkers would pose a significant challenge given that metrics on the number of slang terms contributed and quiz questions answered for each user are stored in our Supabase database rather than on MTurk itself.

Did you perform any analysis comparing different incentives?

No direct incentive comparison (like monetary vs. altruistic) was done since we only used altruistic and course-based incentives. However, we observed that classmates who participated for course credit tended to provide more consistent and timely input than our family members.

If you compared different incentives, what analysis did you perform? If you have a graph analyzing incentives, include a graph here.

N/A

IV. What the crowd gives you

What does the crowd provide for you?

The crowd provides generational-specific definitions and examples of slang terms used in sentences and phrases. They also provide quiz responses that indicate which generational definition resonates the most with them.

Is this something that could be automated? If it could be automated, say how. If it is difficult or impossible to automate, say why.

Basic slang definitions could be approximated with large language models. However, capturing nuanced, generational perspectives on slang is difficult with LLMs. This is because these models generate responses based on the internet data they have been trained on. If one queries a model for a slang definition, it is likely to respond based on the definitions it has encountered most often. And so modern, non-traditional use cases for words that have long existed in the English lexicon will not be captured sufficiently.

For the other feature of our project (the slang quiz), automation becomes essentially impossible as our goal is to survey human respondents of different generations to determine their perspective on slang. As such, it is obviously impossible for a machine to replace human participation.

Did you train a machine learning component from what the crowd gave you?

No, we did not train a machine learning model on the collected data. Our project focused on human contributions and direct aggregation rather than ML-driven inference.

Next 3 Questions do not apply

Did you create a user interface for the crowd workers?

Yes. Contributors saw a web form to submit slang terms, definitions, and examples, and a multiple-choice quiz interface to guess meanings.

Describe your crowd-facing user interface.

Our UI is a clean, responsive React/Next.js application deployed via Vercel. When contributing, users see a form asking for a slang term, its definition and an example sentence with the term. When quizzing, users see a slang term at the top and multiple-choice definitions below. They click on an option and are given immediate feedback (progression to the next question or a message indicating completion).

V. Ethics

Should my application exist at all?

We believe so. Understanding how language varies by generation can foster cross-generational communication.

Does this task potentially expose workers to harm?

Minimal harmful content is involved. Definitions are broadly benign. Workers may find the occasional vulgar slang definition distressing, which would be a concern were this project to be expanded in the future. We could mitigate this issue by instituting content moderation to delete entries for unapproved terms or even ban users who contributed such terms.

Are you fairly compensating the workers for their time?

We did not pay contributors. Classmates participated as part of coursework, and volunteers participated by choice. While not monetary compensation, all participants were informed of the purpose and potential educational value. We believe the compensation to be fair since our contributors' role was to provide the service of improving our project. They knew in advance that they were contributing to help us out, rather than for any financial gain.

If you are creating a dataset for machine learning:

We did not create a dataset for ML. Data is stored securely and is non-private. Users provided slang terms willingly, and no sensitive data was used.

We did create a json dataset from our database, however, and it is linked [here](#).

Is there the potential for introducing bias?

Generational bias may occur if certain groups dominate participation. Some slang may not be well-understood by older generations, skewing data. Recognizing these limitations, we present raw results on our stats page with transparency.

Is your evaluation sound?

Our evaluation—comparing how different generations interpret slang—is straightforward and transparent. While not a formal scientific study, our conclusions are supported by data collected systematically. We do have some concern, however, that user provided definitions of slang terms are not of sufficient quality to make the dictionary feature on our website useful.

VI. Skills

Do your crowd workers need specialized skills?

No specialized linguistic skills were required. Workers only needed basic English proficiency and familiarity with slang. Different generations bring their familiarity with the slang of their era, which ensures we have diverse contributions.

Do the skills of individual workers vary widely?

Yes. Some participants wrote very detailed, context-rich definitions, while others wrote minimal or vague explanations. Familiarity with certain slang terms varied greatly among generations. Given the low average length of the user definitions (just 2 words on average), it appears that most contributors either lacked the skill to produce quality responses or more likely were not interested in generating such responses.

Did you analyze the skills of the crowd?

We performed a qualitative assessment of definition richness. Volunteers sometimes provided creative examples, while some classmates wrote shorter, more literal definitions, which at times were just single synonyms. We did not produce a quantitative skill metric or a formal graph on worker skill.

VII. Quality Control (QC) [In-Depth Analysis]

Is the quality of what the crowd gives you a concern?

Yes. Definitions could be incomplete, off-topic, or not truly reflective of generational usage. In addition, users may not be trying their best on our quiz, such that our data does not accurately represent differences in generations' perceptions of slang.

How do you ensure the quality of what the crowd provides?

We implemented the following QC measures:

- General Requirements:** We only considered a slang term “complete” once we had definitions from three generational categories. If one generation provided a low-quality definition, the other two often clarified the intended meaning.
- Attention Check During Quiz:** Midway through the quiz, we asked a non-slang-related attention question. Participants had to confirm they were still engaged by selecting the answer choice we instructed them to. If not, their submission would not be recorded. There was an additional attention check that would require the user to press a button. If the button was not pressed within a 20-second timeout period, the results were also invalidated. These two attention-check mechanisms are shown in the screenshots on this page.
- Manual Spot Checks:** We reviewed random submissions to ensure participants followed instructions (e.g., no nonsense input, actual slang terms). And for example sentences, we required that they be at least 5 words long.

What Does This Term Mean?

“SELECT “sweet””

Food	→	Poor	→
I don't know	→	Sweet	→

Question 14 of 14

The screenshot shows a mobile application interface. At the top, it asks "What Does This Term Mean?" Below that, it says "“SELECT “sweet””". There are two rows of buttons. The first row has "Food" on the left and "Poor" on the right, with arrows between them. The second row has "I don't know" on the left and "Sweet" on the right, also with arrows. At the bottom, it says "Question 14 of 14".

Are you still paying attention?

Yes, I am still here!

Did you analyze the quality of what you got back?

Yes.

What analysis did you perform on quality?

We performed an analysis based on the word length of provided responses. Our goal was to use a simple proxy to assess the quality of our respondents’ submissions. Specifically, we compared the word length of user definitions and example sentences with those found on Urban Dictionary. Data from Urban Dictionary was sourced via [this Kaggle dataset](#). The dataset included information on more than 2 million Urban Dictionary entries prior to 2018. Our results are displayed in the graph at the end of this section.

A second analysis we performed was on the effectiveness of our attention check question: “SELECT ‘sweet’” at filtering out low quality responses. We queried our database of quiz

responses to determine the impact of the attention check, and saw that all 27 contributors passed this check.

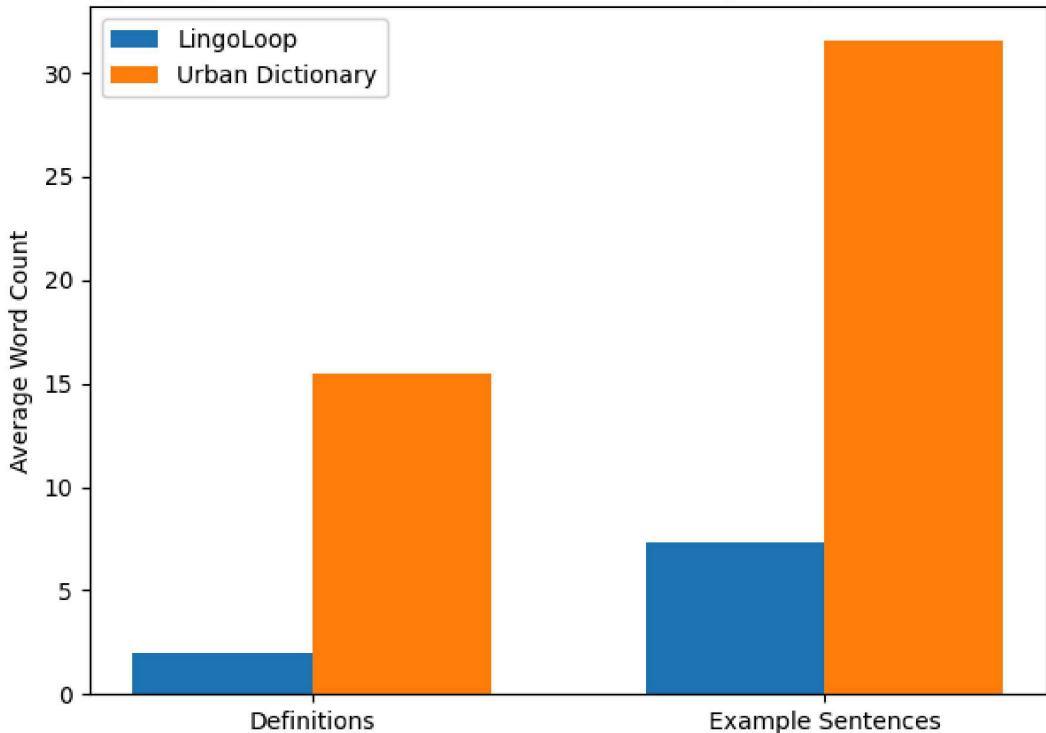
What questions did you investigate and what conclusions did you reach?

As mentioned in our response to the previous question, we sought to determine the quality of user contributed dictionary entries as well as the effectiveness of one of our quality control measures.

With regards to quality control, since all 27 contributors passed our check, it suggests that they were paying attention while taking the quiz and put in their best effort. However, it means we learn little about the effectiveness of this QC method. It makes sense that our contributors would pass the QC because we knew them all personally, either as classmates or family, and collected their emails. As such, there is a social stigma attached to putting minimal effort towards our quiz. To truly test our QC method, we would have to open up our website to anonymous crowdworkers, who are more likely to attempt to complete tasks quickly without paying attention to the work they are doing, since there is more of a profit motive for them.

With regards to user contributions towards our slang dictionary, we found the average length of definitions to be just 2.0 words. Meanwhile, the average definition length on Urban Dictionary was computed to be 15.8 words. For average example sentences, our users averaged 7.3 words while Urban Dictionary users averaged 31.6. A closer inspection at our users' contributions reveals definitions such as "super fun", "cool", and "very, extremely". Evidently, our respondents did not put much thought into their slang definitions, instead resorting to vague descriptors or even synonyms to define their terms. Such "definitions" would be appropriate in the context of casual conversation but are not of any use to an online reader unfamiliar with a particular slang term. In conclusion, going by response length alone, our crowdsourced dictionary entries are of significantly inferior quality to those provided by users on Urban Dictionary, the world's largest crowdsourced slang dictionary.

Average Length of Definitions and Sentences: LingoLoop vs Urban Dictionary



VIII. Aggregation [In-Depth Analysis]

How do you aggregate the results from the crowd?

We aggregated three generational definitions into a single multiple-choice question. The quiz collected user guesses, and we tallied how each generation responded to each option.

Did you analyze the aggregated results?

Yes. We looked at how often each generation selected each definition. This helped identify patterns—for instance, maybe Millennials and Gen Z frequently picked the same definition, suggesting shared linguistic understanding.

What analysis did you perform on the aggregated results?

We compared generational response rates to each option. For each slang term, we produced a histogram showing how each generation's (Gen Z, Millennials, Gen X, and Boomers) participants distributed their votes across the four options (three definitions + "I don't know").

What questions did you investigate and what conclusions did you reach?

We investigated whether younger generations tend to agree more with each other compared to older generations. We found some clustering across the 13 terms for which we had sufficient responses to make a useful analysis: Gen Z and Millennials aligned on modern internet slang

more often, while Gen X/Baby Boomers occasionally diverged, reflecting different cultural backgrounds.

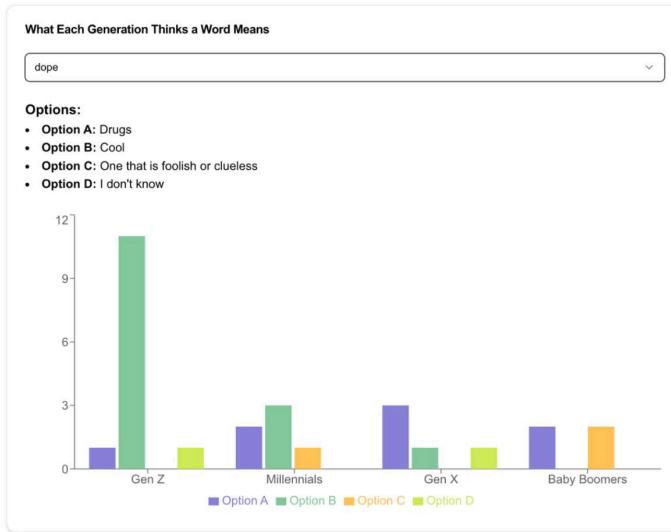
Did you create a user interface for end users to see the aggregated results?

Yes, we built a Stats page that shows a bar chart for each slang term, illustrating generational response patterns. This also doubles as the analysis graph for our aggregation. An example Stats page for the term 'dope' is shown in the second image below.

What Each Generation Thinks a Word Means

Select a slang term

- bet
- cap
- delulu
- dope
- fake
- fam
- fire
- flex
- ghost
- lit
- low-key



Describe what your end user sees in the interface.

The Stats page presents a dropdown to select a slang term (first image above). Once selected,

users see a chart with four bars (Option A, B, C, D) grouped by generation. Labels and color-coding make it clear how each generation responded. This visual summary helps users quickly grasp generational differences in slang interpretations.

IX. Scaling Up

What is the scale of the problem you are trying to solve?

Currently, we tested with a few dozen slang terms. Scaling to thousands of terms would create a rich linguistic corpus capturing generational slang differences.

Would your project benefit from contributions from thousands of people?

Yes. More contributors mean more robust data, covering more slang terms and more reliable generational patterns.

If it would benefit from a huge crowd, how would it benefit?

A larger crowd would produce statistically significant patterns, reducing the influence of outliers. It would enable us to cover a wider range of slang and refine definitions even further. We could even introduce an option for users to vote on definitions for their generation, ensuring that only the best quality contributions are added to our quiz.

What challenges would scaling introduce?

- **Cost & Incentives:** Without monetary compensation, relying on volunteers at large scale might be difficult.
- **Quality Control:** Ensuring QC with a huge crowd may require more automated checks. This is especially true given that one of our QC methods failed to weed out respondents.
- **Data Management:** Handling thousands of terms and tens of thousands of responses requires more sophisticated database and analysis pipelines.

Scaling-Up Analysis

From the first lecture, we know that the average worker on MTurk makes \$5 per hour. We gauge that users should be compensated at a rate slightly higher than this average, \$6, because our task interface is hosted on an external site, meaning more hoops for the Turkers to jump through in order to contribute. In addition, we estimate it should take users 1 minute on average to provide a slang entry (word, definition, sentence), while completing the quiz should take about 2 minutes (an average of 8.6 seconds per multiple-choice question). This means we would pay our Turkers at a rate of 10 cents per definition and 20 cents per quiz attempt. Scaling up so that our platform rivals Urban Dictionary in size (with its 2.3 million entries as of 2018) would thus require a \$230,000 budget, and we have not even accounted for the collection of quiz responses to capture differences in generations' understanding of slang. Clearly, creating a quality product with crowd worker contributions would require a significant financial outlay. As such, we think that adding a gamification layer (leaderboards, achievement badges) could keep large crowds engaged for less or even no money. We also think caching partial results and employing simpler UIs would allow for faster contributions.

X. Project Analysis

Did your project work? How do you know?

Yes, our project successfully collected generational definitions, created quizzes, and analyzed results. Users found the interface usable, and the quizzes ran smoothly. We can show that certain terms are consistently interpreted similarly by some generations and differently by others.

Do you have a graph analyzing your project?

We included bar charts (see Aggregation and QC section) analyzing generational responses and a search page to look at word definitions across generations

What were the biggest challenges?

- Ensuring consistent participation without monetary incentives.
- Implementing attention checks to maintain data quality.
- Aggregating results and presenting them meaningfully.

Were there major changes between what you originally proposed and your final product?

Yes. Initially, we planned to just build a slang dictionary displaying generational differences in term definitions. After consultation with the professor, we added the gamified component to allow for a more quantitative display of generational perceptions. We also had planned to analyze a much larger set of terms but scaled down due to time constraints.

Limitations of your product:

- Non-monetary incentives may limit participant diversity at scale.
- Definitions are subjective; some slang lacks a single “correct” meaning.
- Limited number of participants may reduce statistical robustness.
- Access to participants of older generations is limited due to lack of a real crowd. We did get our family members to help, but as can be seen in our results, we have comparatively few of them to draw from.

Did your results deviate from what you would expect based on previous work or class learnings?

We expected younger generations to dominate modern slang understanding, which largely held true. However, some older participants provided surprisingly accurate definitions of modern slang, challenging initial assumptions about generational divides. We also noted that Gen Z and Millennial perception of modern slang was more similar than either generation's was to Gen X/Boomers which makes sense given the nature of language evolution.

If results deviated, why might that be?

Volunteer interest and individual linguistic competence vary widely. Some older users might be very internet-savvy, skewing results. They may also have deliberately tried to adopt the mindset

of a young respondent when completing our quiz, given that much of the Gen Z slang that appeared in the quiz lacked an informal definition in generations past.

XI. Technical Challenges

Did your project require substantial technical components?

Yes. We implemented:

- Next.js front-end for user interaction.
- Supabase for database and authentication.
- Serverless functions to manage quiz generation and data retrieval.

Largest technical challenge:

Synchronizing data insertion and quiz generation was tricky. After a slang term reached full definition coverage, we had to trigger a quiz question generation automatically.

How did you overcome this challenge?

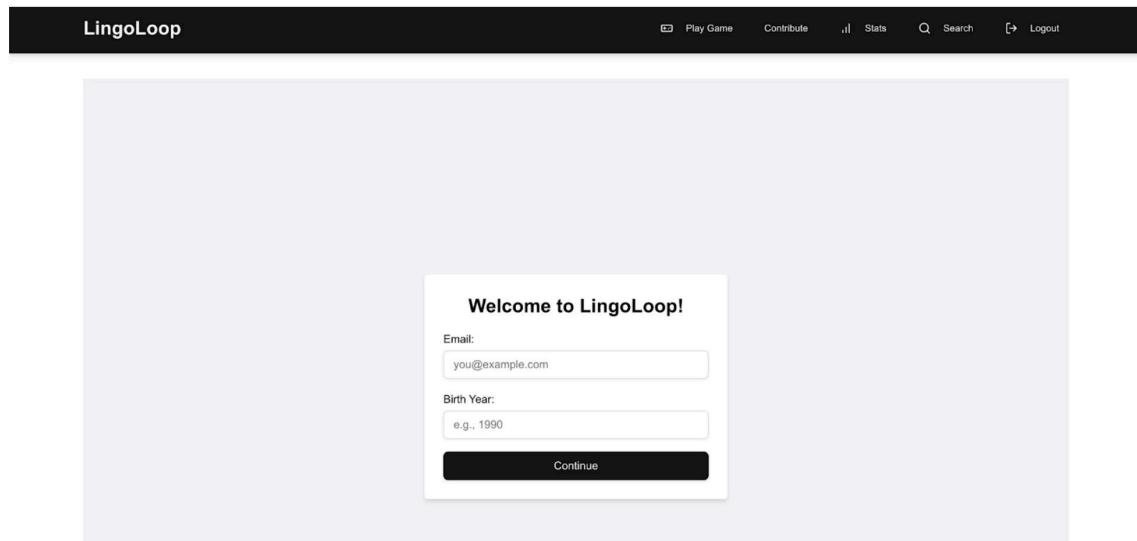
We used Supabase triggers and carefully structured database tables. We tested extensively to ensure that the quiz question is generated only when all three definitions are present.

Do you have screenshots or diagrams?

A flow chart of our application can be found [here](#).

XII. Project Screenshots

1- Homepage/Login page



2- Quiz interface

What Does This Term Mean?

“dope”

Drugs	→	Cool	→
I don't know	→	One that is foolish or clueless	→

Question 6 of 14

```
graph LR; A[Drugs] --> B[Cool]; C[I don't know] --> D[One that is foolish or clueless]
```

3- Slang Search

Search Slang Terms