

# Investigating data to text approaches to achieve diversity of generated marketing text in the music industry

Alexandre Mercier

11.03.2024, Master Thesis Final Presentation

Chair of Software Engineering for Business Information Systems (sebis)  
Department of Computer Science  
School of Computation, Information and Technology (CIT)  
Technical University of Munich (TUM)  
[www.matthes.in.tum.de](http://www.matthes.in.tum.de)

1. Motivation & Goal
2. Problem Statement
3. Approach
4. Results & Findings
5. Conclusion & Future Work

# Motivation



DJ, Aschaffenburg (DE)

Rafe Pressler



Noch keine Bewertung

DJ für Partys, Hochzeiten und Firmenevents

Dance

Schlager

Latin

Rap



DJ, dortmund (DE)

patty



Noch keine Bewertung

Party

Tanzmusik

Rock

Pop



DJ, Duisburg (DE)

DJ Walli



Noch keine Bewertung

Sie haben einen Grund zu feiern? Klar! Denn es gibt immer einen guten Grund für gute Partys! Egal ob Hochzeiten, Geburtstage, Firmenevents, Abendgalas oder ande...

Party

Unterhaltungsmusik

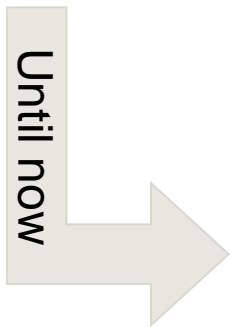
Lounge

Companies and platforms are scrambling to leverage the newest Natural Language Generation (NLG) technologies:

- Assist time-intensive tasks (specifically: requiring effort from users)
- Automate writing similar texts

Companies and platforms are scrambling to leverage the newest Natural Language Generation (NLG) technologies:

- Assist time-intensive tasks (specifically: requiring effort from users)
- Automate writing similar texts



## Templates

Welcome to the energetic world of \_\_\_\_\_! This \_\_\_\_\_ hails from the town of \_\_\_\_\_ and is ready to conquer the clubs in their region. With an impressive radius of \_\_\_\_\_ km, they are prepared to spread their unique blend of \_\_\_\_\_ and \_\_\_\_\_ far and wide. \_\_\_\_\_ consists of \_\_\_\_\_ talented musicians who captivate audiences with their creative arrangements and dynamic live performances. Immerse yourself in the world of \_\_\_\_\_ and let their compelling sound sweep you away!

# Motivation



DJ, Aschaffenburg (DE)

Rafe Pressler



Lasst euch von Rafe Pressler, dem Schlager-DJ aus Aschaffenburg, in eine unvergessliche musikalische Welt entführen! In einem Umkreis von 300 km steht Rafe Pressler bereit, die Clubs der Region mit mitreißenden Schlagerklän...

Dance

Schlager

Latin

Rap



DJ, dortmund (DE)

patty



Lasst euch von patty, dem Rock-DJ aus Dortmund, in eine unvergessliche musikalische Welt entführen! In einem Umkreis von 300 km steht patty bereit, die Clubs der Region mit mitreißenden Partyklängen zu erobern.

Party

Tanzmusik

Rock

Pop



DJ, Duisburg (DE)

DJ Walli



Noch keine Bewertung

Sie haben einen Grund zu feiern? Klar! Denn es gibt immer einen guten Grund für gute Partys! Egal ob Hochzeiten, Geburtstage, Firmenevents, Abendgalas oder ande...

Party

Unterhaltungsmusik

Lounge

# Motivation



DJ, Aschaffenburg (DE)

Rafe Pressler



Zu jedem Fest gehört Musik. Ob Hochzeit, Firmenevent oder Party, meine Künste bringen selbst den faulsten Tiger zum tanzen. Ich komme zu euren Events in Aschaffenburg und Umgebung ...

Dance

Schlager

Latin

Rap



DJ, dortmund (DE)

patty



Hier DJ Patty, fresh aus Dortmund mit den besten underground tunes für euch. Tanzt ihr lieber zu Rock und Pop? Dann steht auch da mein Repertoire für euch zur Verfügung ...

Party

Tanzmusik

Rock

Pop



DJ, Duisburg (DE)

DJ Walli



Noch keine Bewertung

Sie haben einen Grund zu feiern? Klar! Denn es gibt immer einen guten Grund für gute Partys! Egal ob Hochzeiten, Geburtstage, Firmenevents, Abendgalas oder ande...

Party

Unterhaltungsmusik

Lounge

LLMs still suffer from repetitions:

**“Over 90% of 1008 generated jokes were the same 25 Jokes.”**

S. Jentzsch and K. Kersting, “ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models,” Jun. 2023. [Online]. Available: <http://arxiv.org/pdf/2306.04563v1>





**Startup aiming to simplify how artists/bands connect with event organizers and locations.**

**Three use-cases:**

1) Descriptive Texts of bands	2) Press kit for bands	3) Regional summary
We are The Letters, bringing the trendiest songs to your festival.	The Letters are an inspiring duo from Regensburg. They last played at the Puch Open Air ...	In Bavaria you will encounter a wide a variety of traditional wind orchestra as well as ...

## Related Work

Templates: learned through neural models, most rigid generation. [9]

Plan-then-Generate: multi-step process of ordering the data then generating a text. ([PlanGen](#)) [2][4][8][10]

Large language models: impressive results but hard to evaluate. ([T5](#), [GPT](#), ...) [3][5][7]

## Challenges

Controllability: How can one orient a model's training or generate the output they require.

Creativity: Variety of content is limited compared to human content.

Lack of input data: The model relies on structured data but a text will often have additional context, analysis or details.

Evaluation: of generated content is complicated and often manual.

1. Motivation & Goal

2. Problem Statement

3. Approach

4. Results & Findings

5. Conclusion & Future Work

RQ1

Which generative data-to-text technique yields the best overall results?

RQ2

How can we compare similar generated text? How can we measure diversity in generated content?

RQ3

How can creativity/variety of generative models be controlled and promoted?

RQ4

Are generated texts as fluent and coherent as human written text?

1. Motivation & Goal
2. Problem Statement
3. Approach
4. Results & Findings
5. Conclusion & Future Work



- Database snapshot from April 2023
- ~900 bands, but only 440 with descriptions. 360 samples left after cleaning.

- Done with the Deepl API

- Manual annotations required for finetuning and dataset split

---

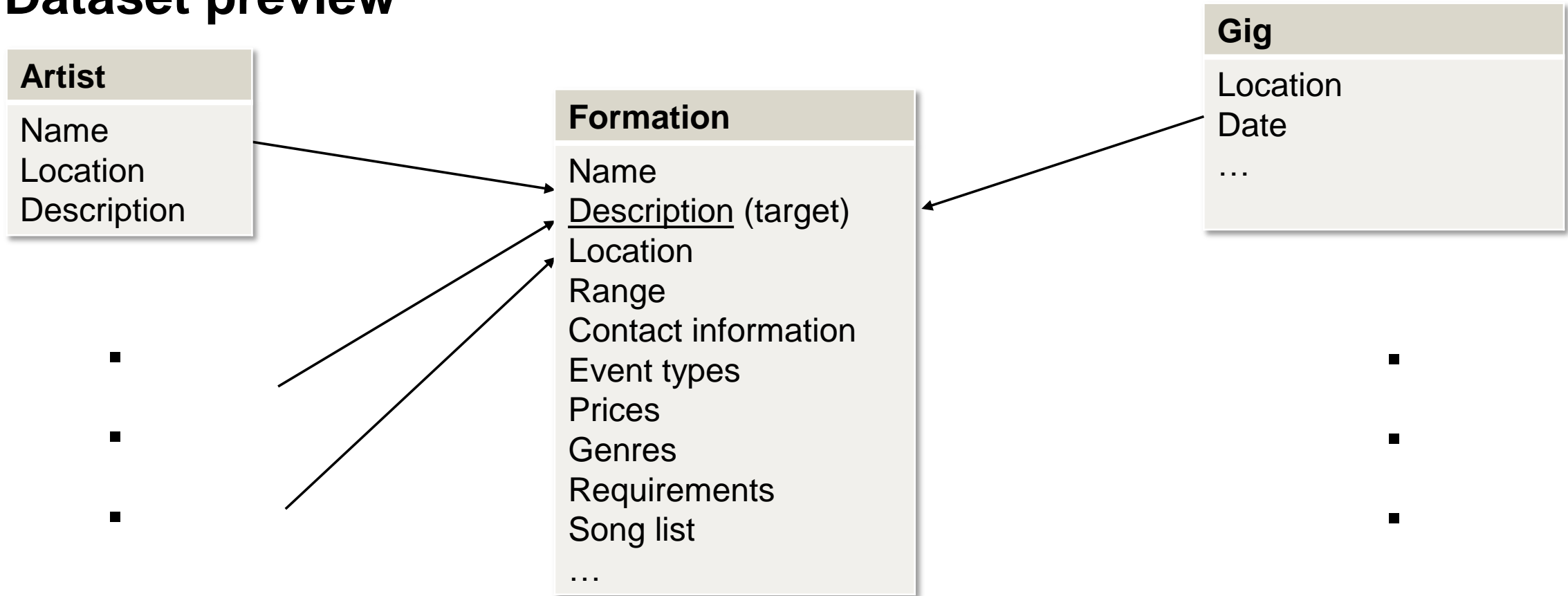
## “Parameter“ Split

- 10 bands and DJs
- For experimenting with parameters while staying statistically relevant

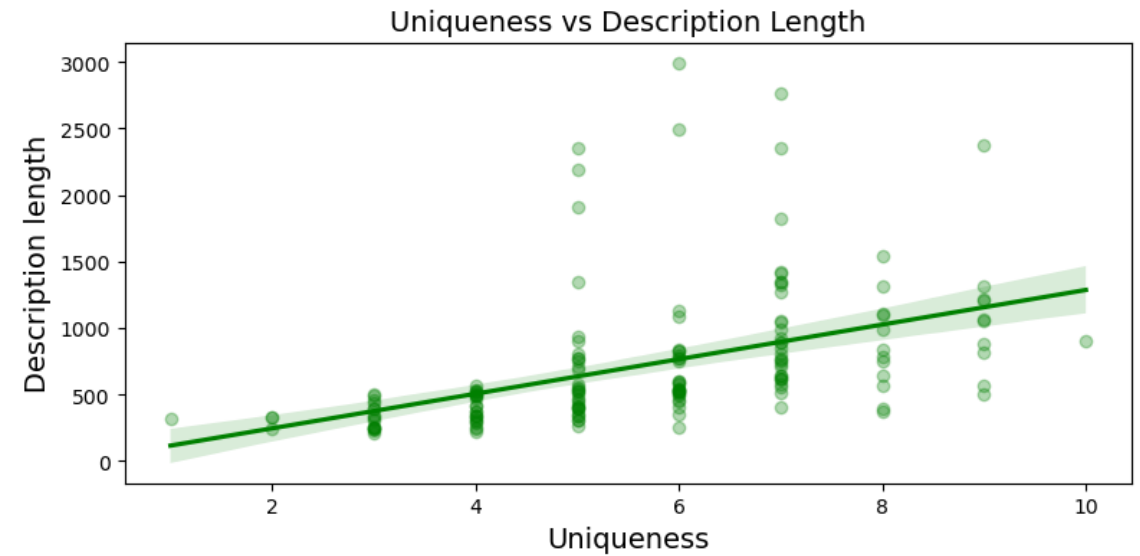
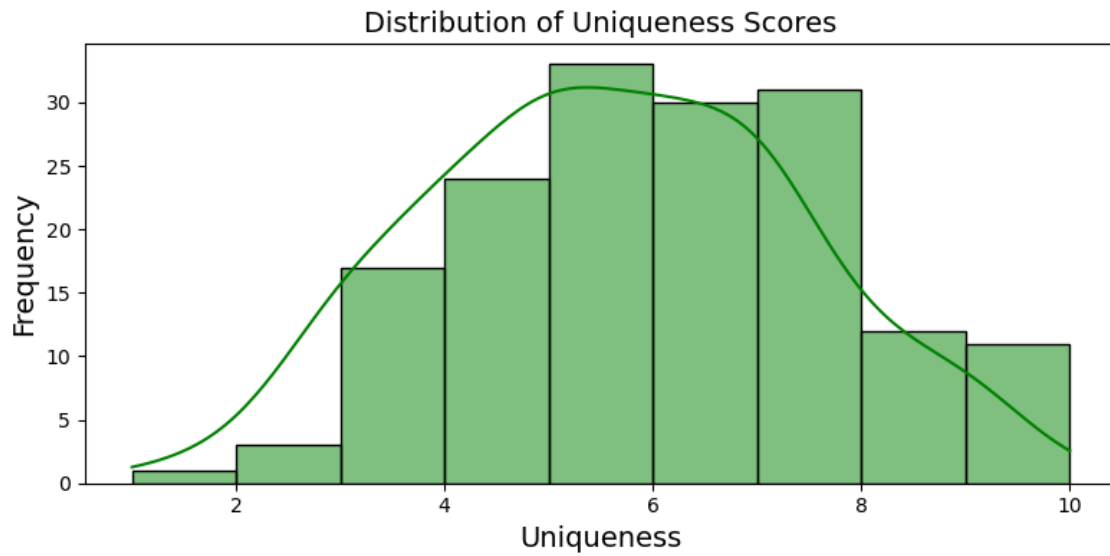
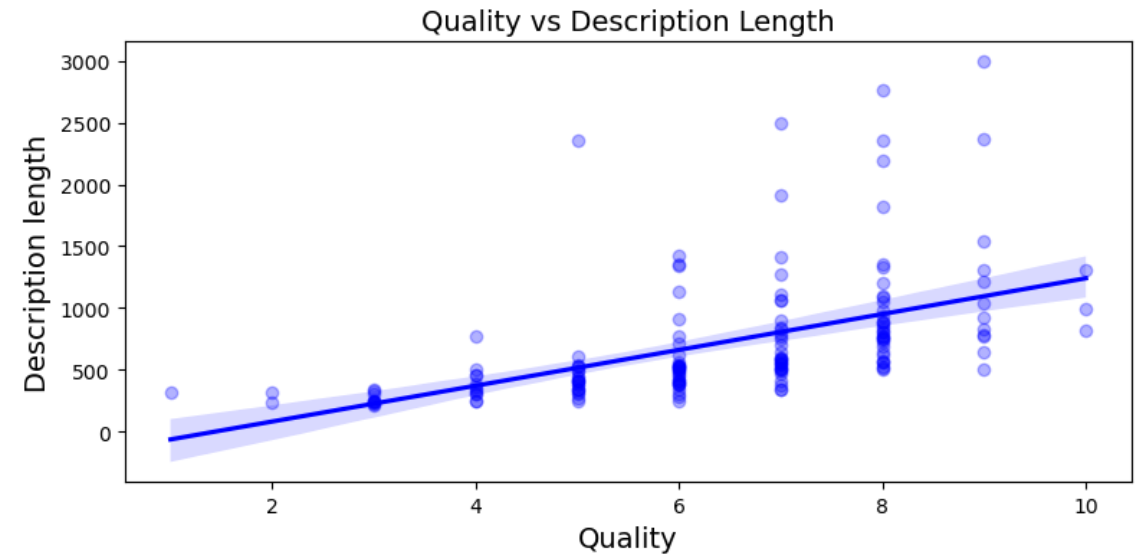
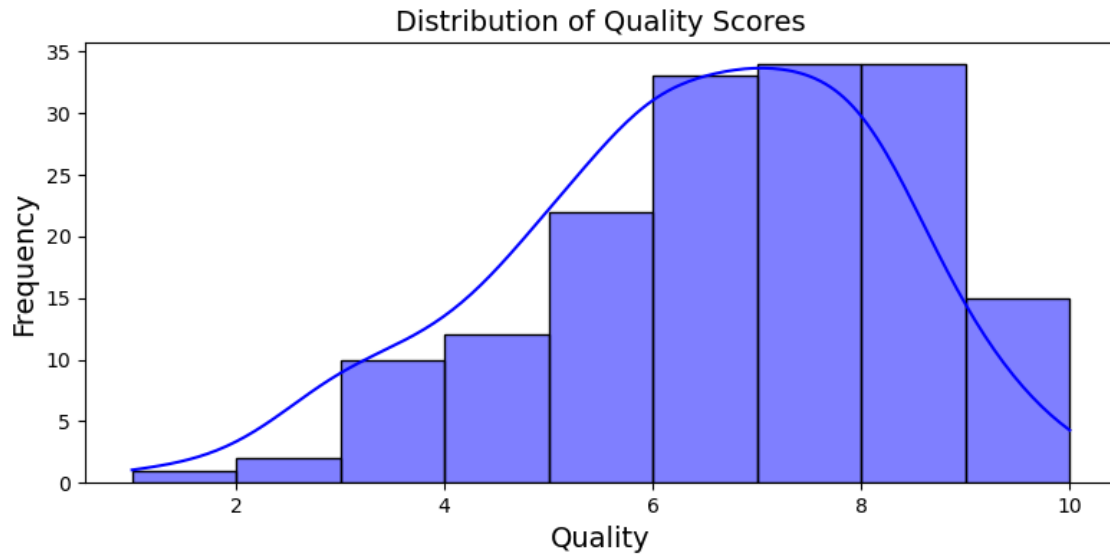
## “Diversity“ Split

- 50 bands from Bavaria
- Dataset for measuring and comparing techniques and metrics

## Dataset preview



# Is the manual annotation biased?





## How do we evaluate the results?

### Quality

Automated evaluation using Geval:

- Engagingness
- Fluency
- Naturalness
- Informativeness

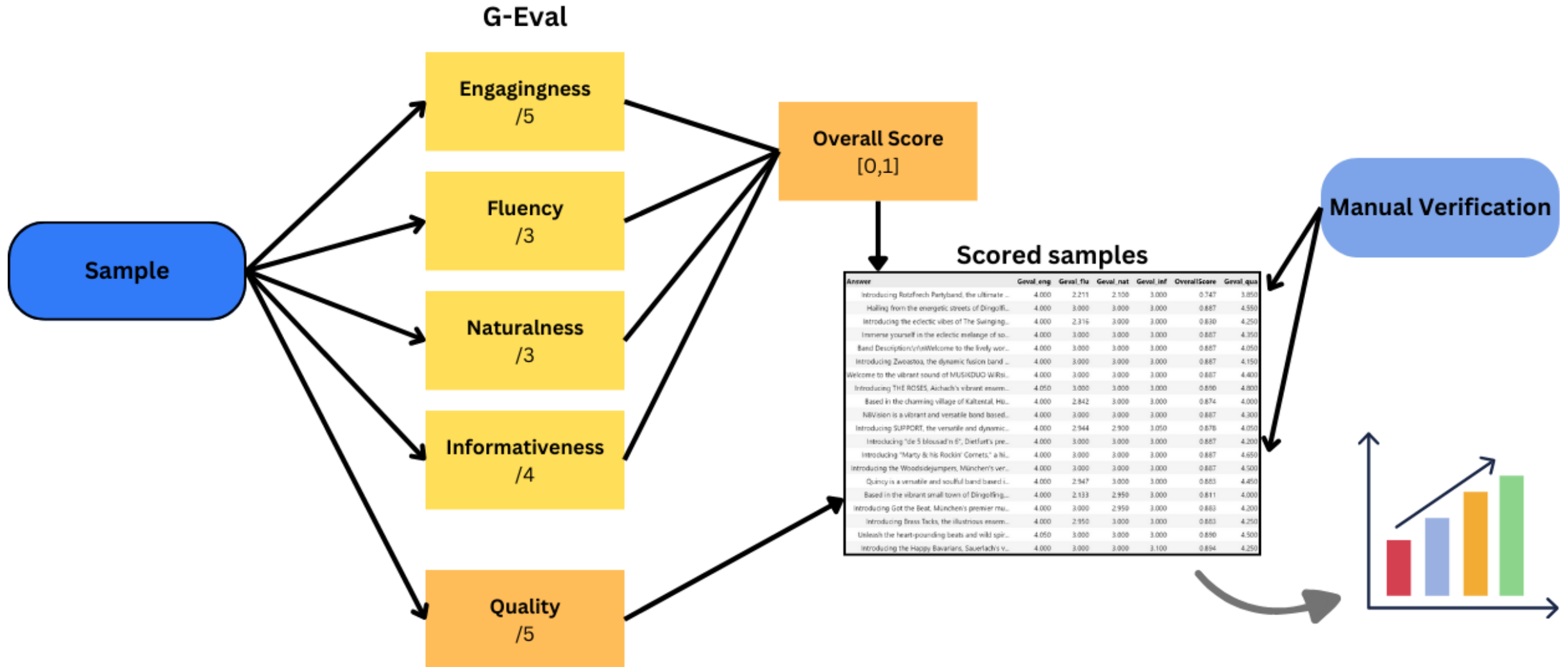
- Overall Score
- Quality

### Similarity

Jaccard similarity score:

- Pairwise comparison
- ngram based
  
- easy to visualize





## Advantages

- Reference-free metric
- Task agnostic
- Leverages Chain-Of-Thought (COT)
- Highly adaptable

## Disadvantages

- In practice: has a bias towards machine-generated content
  - Uses the same LLM to generate and to evaluate the results
  - LLMs favor integer scores
  - Output is sensitive
  - Relies on opaque model
- Still comparable?
  - Conflict of interest?
  - Average over multiple runs
  - Requires debugging logic
  - API/Version can change

Yang Liu et al. "G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment" May 2023 <https://arxiv.org/abs/2303.16634>

# G-eval

You will be given one description written in english for a band or DJ.  
Your task is to rate the descriptions on one metric.

Please make sure you read and understand these instructions carefully.  
Please keep this document open while reviewing and refer to it as needed.

The final score is  
the average over 20  
scores

## Evaluation Criteria:

Engagingness (1-5) Is the text dull/interesting?

- **-1:** This is not a description of a band or DJ.
- **1:** (Low Engagingness) Dull, lacks creativity and excitement, no captivating details. Uninspiring and unlikely to hire.
- **2:** (Below Average Engagingness) Somewhat interesting, but lacks uniqueness. Basic information, bland language. Might consider if no better options.
- **3:** (Average Engagingness) Decent level of engagement, informative but not exciting. Might be considered if criteria align.
- **4:** (High Engagingness) Exciting and enthusiastic. Unique selling points, captivating language. Highly likely to hire.
- **5:** (Exceptional Engagingness) Extremely engaging, persuasive, and creative. Creates strong desire to hire without hesitation.

## Evaluation Steps:

1. Read the description carefully to get a sense of its overall content and style. If it is not a description of a band or DJ, rate it -1 and skip the other steps.
2. Consider the language used, the information within and the structure of the sentences. Is it engaging and attention-grabbing, or does it lack creativity?
3. Rate the description on a scale of 1 to 5, with 1 being dull and 5 being highly interesting.

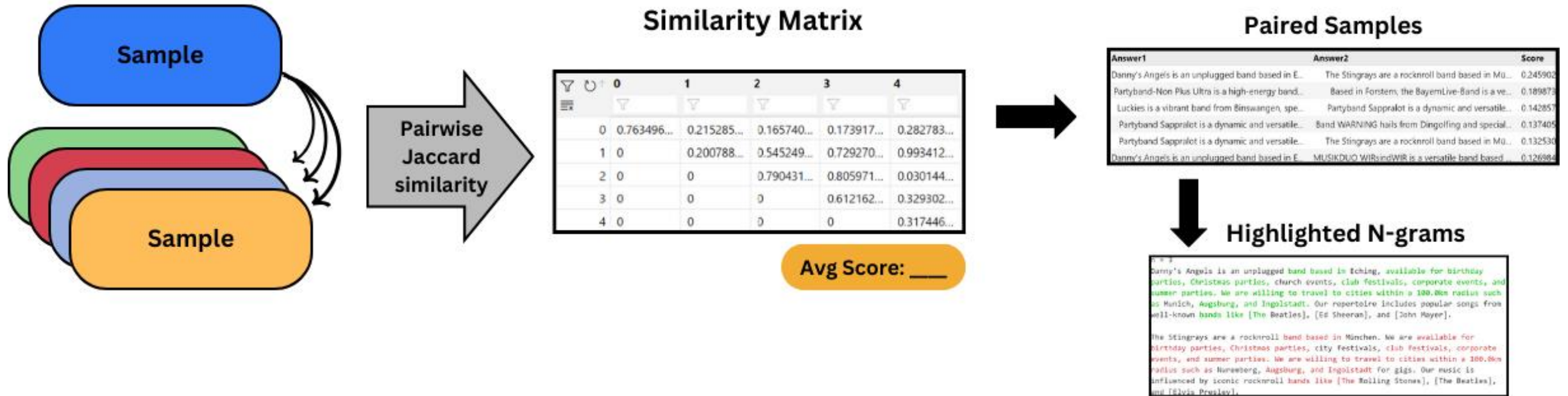
Remember to focus on the engagement factor of the text, rather than the accuracy of the description or personal preferences for music genres.

Description:

{{Description}}

Evaluation Form (scores ONLY):

Form diversity vs. Content diversity  
 → Since the content is limited by the data we check for form diversity



$$Jaccard(U, V) = \frac{|U \cap V|}{|U \cup V|}$$

With U and V the sets of ngrams for sample 1 and 2 respectively

## Parameter Split

Temperature variation	10 values between 0 and 2
Top-p variation	5 values between 0.1 and 0.9
Logit Bias	5 values between -5 and -100

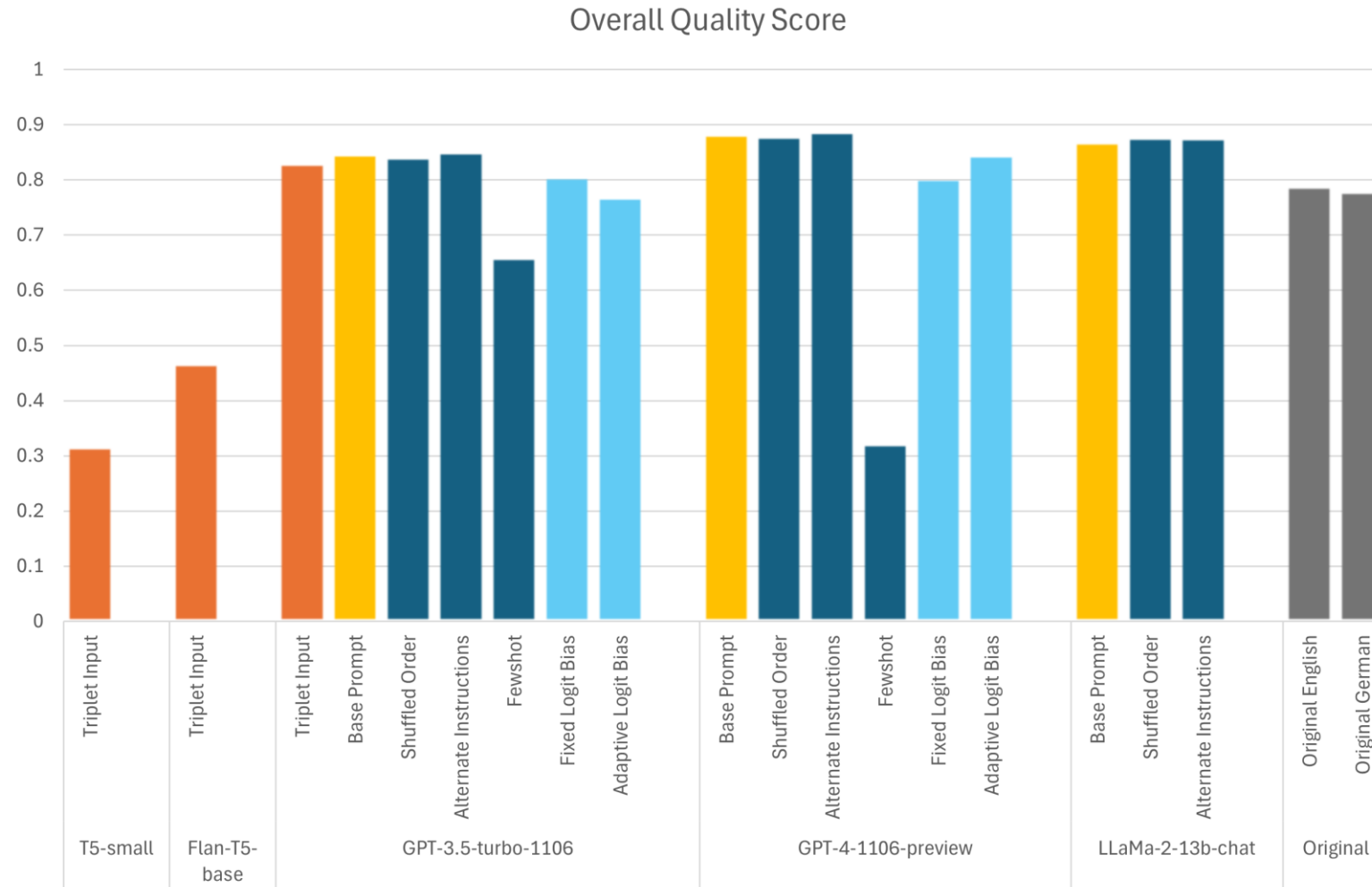
## Diversity Split

Triplet-style input	(subject, predicate, object) without additional instructions
Base prompting	The same prompt is used for every sample
Shuffled Order	The order of the data is shuffled inside the prompt
Alternate Instructions	The base prompt's instructions are modified for each sample
Fewshot	2 random samples are passed along with a prompt to generate similar text
Logit Bias (Fixed)	Most used tokens are suppressed in next generation
Logit Bias (Adaptive)	Most used tokens are adaptively suppressed in next generation

## Outline

1. Motivation & Goal
2. Problem Statement
3. Approach
4. Results & Findings
5. Conclusion & Future Work

# Which generative data-to-text technique yields the best overall results?





## Which generative data-to-text technique yields the best overall results?

- Finetuning worked badly → T5-small and FlanT5 didn't produce good samples
- Switching from triplets to text prompts removed a lot of ambiguity but also restricted creativity
- Zero-shot approaches all achieved great texts
  - GPT-3.5 was very concise, only using provided data
  - GPT-4 and LLaMa2 extrapolated more details, sometimes too long.
- Fewshot experiments failed (probably due to mistranslations or mixup of data)
- The biggest influence was the choice of model itself
  - GPT-4 and LLaMa2 the best choice

## How can we compare similar generated text? How can we measure diversity in generated content?

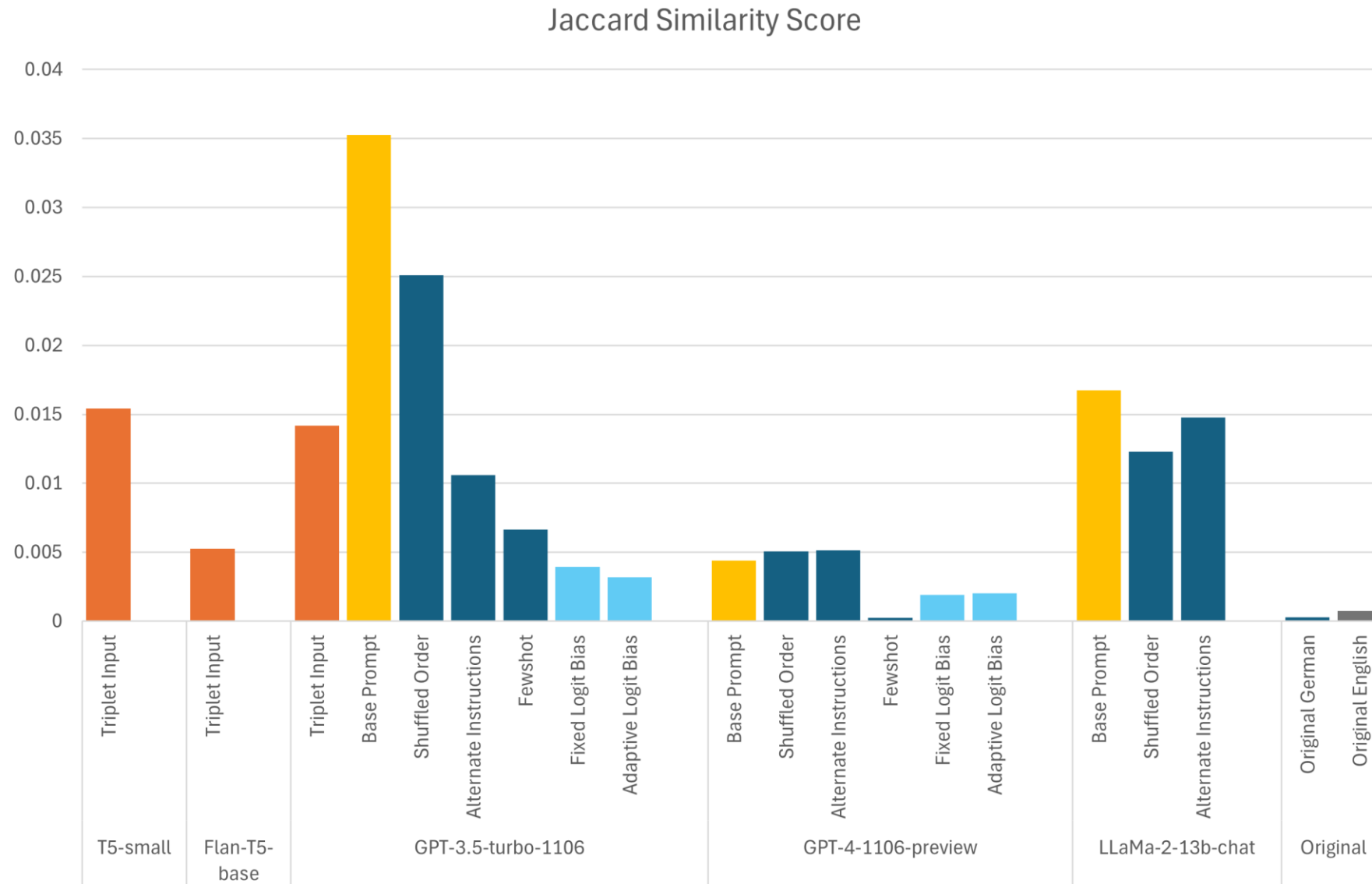
- N-gram based method
    - Efficient, easy, understandable
    - Might not generalize to other tasks
    - Required a lot of manual inspection
- Only works for form diversity

n = 3

Jaccard score: 0.24

Danny's Angels is an unplugged band based in Eching, available for birthday parties, Christmas parties, church events, club festivals, corporate events, and summer parties. We are willing to travel to cities within a 100.0km radius such as Munich, Augsburg, and Ingolstadt. Our repertoire includes popular songs from well-known bands like [The Beatles], [Ed Sheeran], and [John Mayer].

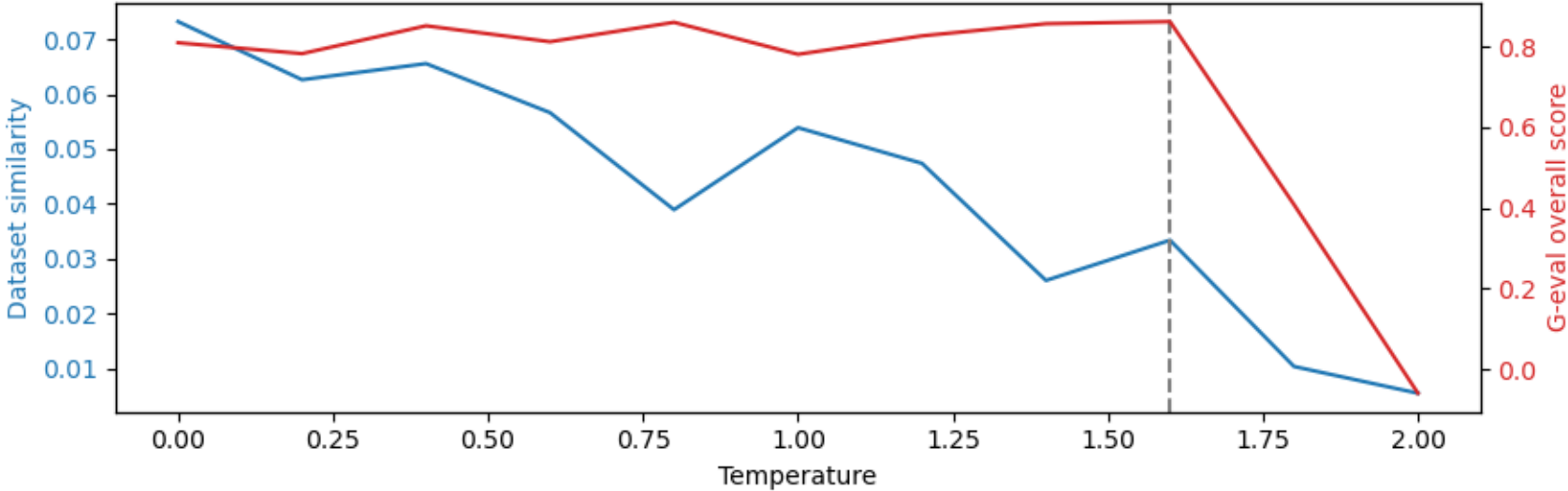
The Stingrays are a rocknroll band based in München. We are available for birthday parties, Christmas parties, city festivals, club festivals, corporate events, and summer parties. We are willing to travel to cities within a 100.0km radius such as Nuremberg, Augsburg, and Ingolstadt for gigs. Our music is influenced by iconic rocknroll bands like [The Rolling Stones], [The Beatles], and [Elvis Presley].



# Temperature experiment



Temperature has a weak effect on similarity without affecting quality up to a threshold (here 1.6)

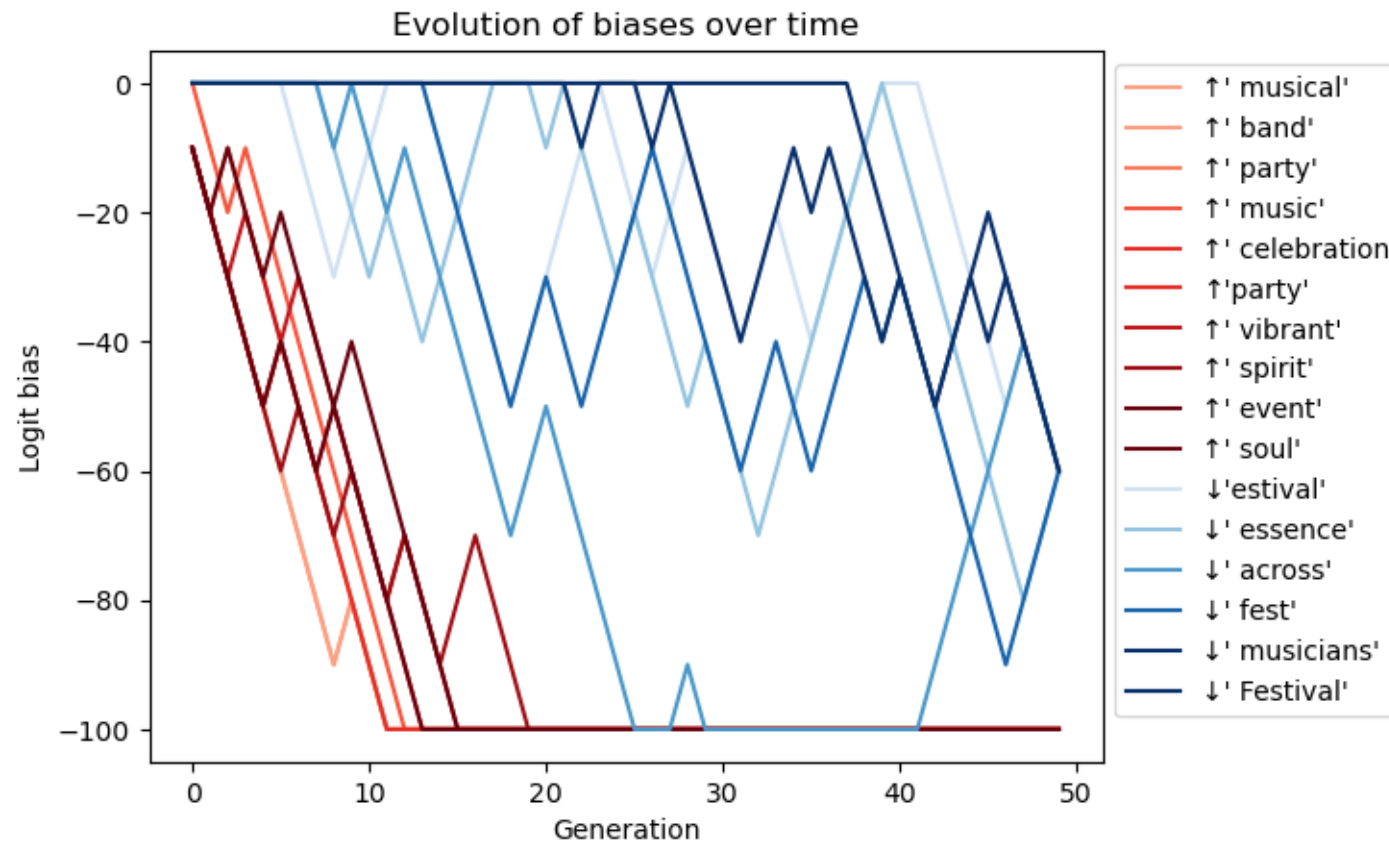


# Logit Bias (adaptive)

The 100 most used logits are suppressed according to the following formula:

$$-(\#in\_top\_100 - \#not\_in\_top\_100) * 10$$

Range: 0 to -100



Are generated texts as fluent and coherent as human written text?

Ideally: generated samples indistinguishable from human written text

Model	Fluency		Naturalness		#samples
	Avg.	#Scores $\geq 2.9/3$	Avg.	#Scores $\geq 2.9/3$	
Original DE	2.48	4	2.75	18	50
Original EN	2.42	22	2.74	28	50
GPT-3.5	2.74	41	2.98	49	50
GPT-4	2.94	46	2.98	49	50
LLaMa2	2.99	20	2.97	19	20

Hard to judge:

- Geval biased towards LLMs
- Difference between input data and content of references

Reality:

- Similar grammatical quality, but different content due to input data
- Fluent and enjoyable to read but longer texts felt hollow

## Task 2 and 3: Pilot experiments

### Task 2: Writing an article containing recent and upcoming gigs

- GPT-4 was able to incorporate specific data about each event but failed to aggregate and sort them by importance.
- Trouble at separating recent and upcoming events despite passing the current date to the model.

### Task 3: Regional Summaries

- Only enough data for Bavaria
- Overall, good summarization of common/popular events. However, the dataset doesn't contain information about each events size/importance

**No references to  
compare them to!**

# Implementation

- Not live yet, but fully implemented
- A usage study should then be done by comparing proposed samples to the descriptions used by the artists
- Drawbacks of using LLMs-as-a-service:
  - Slow responses (20-30sec for GPT-4)
  - Expensive in the long run
  - API and model constantly changing

**Kategorien & Sucheinstellung**

**Besetzung**

Band x Big Band x

**Genres**

Blasmusik x Hinzufügen

**Buchbar für**

Geburtstagsfeier x Karneval und Fasching x Weihnachtsfeier x Kirche x Stadtfest x Club / Pub / Bar x

Hinzufügen

**Heimatort**

Heimatort\*  
Eggenfelden

50 km 100 km 150 km 200 km

Suche Gigs im Umkreis von: 150 km

**Beschreibung** ⓘ

Beschreibung Generieren ⓘ

Meet the Broccoli Catz—an exuberant Blasmusik Big Band hailing from the charming town of Eggenfelden, DE. This lively ensemble turns any occasion—a Geburtstagsfeier, Karneval und Fasching, Weihnachtsfeier, you name it—into a musical extravaganza that guests will never forget. Each festivity, whether within the hallowed halls of a Kirche or under the exuberant skies of a Stadtfest, is graced with their thrilling performance.

With the versatile Magic Jerry at the helm with a cornucopia of talents (from alphorn to vocals), and his dynamic members such as Wolfgang, Bertl, Uwe, and the enigmatic Magda, this band embodies an extraordinary fusion of musicality and technical prowess. Keeping up with the anonymous Ulknudels, they create an enthralling Blasmusik experience, conjuring up a rambunctious, brass-heavy ambiance in Clubs and Pubs alike.

Abbrechen Speichern



## Outline

1. Motivation & Goal
2. Problem Statement
3. Approach
4. Results & Findings
5. Conclusion & Future Work

## Conclusion

- Implemented a robust pipeline for quality and diversity evaluation
- Showed that the diversity promoting techniques had a great influence on some models, yet the choice of model still had the largest impact
- CoT techniques had a positive effect on relevancy and quality of samples by augmenting the knowledge but restricted the output diversity

## Future work

- Rerun similar experiments with a simpler/more controlled dataset that has reliable references
- Investigate embedding- or entropy-based metrics for content similarity evaluation
- Find data augmentation strategies that don't restrict the model's creativity
  
- Study impact of automatic generation on digital platforms

- [1] an Chenxin, J. Feng, K. Lv, L. Kong, X. Qiu, and X. Huang, “CoNT: Contrastive Neural Text Generation,” May. 2022. [Online]. Available: <http://arxiv.org/pdf/2205.14690v4>
- [2] R. Eisenstadt and M. Elhadad, *Neural Micro-Planning for Data to Text Generation Produces more Cohesive Text*, Proceedings of the Workshop on Discourse Theories for Text Planning. Dublin, Ireland: Association for Computational Linguistics. [Online]. Available: <https://aclanthology.org/2020.dt4tp-1.2>
- [3] M. Freitag and S. Roy, “Unsupervised Natural Language Generation with Denoising Autoencoders,” Apr. 2018. [Online]. Available: <http://arxiv.org/pdf/1804.07899v2>
- [4] R. Puduppully, Y. Fu, and M. Lapata, “Data-to-text Generation with Variational Sequential Planning,” Feb. 2022. [Online]. Available: <http://arxiv.org/pdf/2202.13756v1>
- [5] L. F. R. Ribeiro, M. Schmitt, H. Schütze, and I. Gurevych, “Investigating Pretrained Language Models for Graph-to-Text Generation,” Jul. 2020. [Online]. Available: <http://arxiv.org/pdf/2007.08426v3>
- [6] E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds., *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018.
- [7] Y. Su, Z. Meng, S. Baker, and N. Collier, “Few-Shot Table-to-Text Generation with Prototype Memory,” 2021.
- [8] Y. Su, D. Vandyke, S. Wang, Y. Fang, and N. Collier, “Plan-then-Generate: Controlled Data-to-Text Generation via Planning,” Aug. 2021. [Online]. Available: <http://arxiv.org/pdf/2108.13740v1>
- [9] S. Wiseman, S. Shieber, and A. Rush, “Learning Neural Templates for Text Generation,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018, pp. 3174–3187.
- [10] S. Yang and Y. Liu, “Data-to-text Generation via Planning,” *J. Phys.: Conf. Ser.*, vol. 1827, no. 1, p. 12190, 2021, doi: 10.1088/1742-6596/1827/1/012190.



Prof. Dr.

**Florian Matthes**

Technical University of Munich (TUM)  
TUM School of CIT  
Department of Computer Science (CS)  
Chair of Software Engineering for Business  
Information Systems (sebis)

Boltzmannstraße 3  
85748 Garching bei München

+49.89.289.17132  
matthes@in.tum.de  
[www.matthes.in.tum.de](http://www.matthes.in.tum.de)



## Example prompt

This is an example prompt as used with GPT-3.5, GPT-4 and LLaMa2.  
Other models relied on data in a triplet-format.

Create a captivating and comprehensive band description with the following details:

Name: **Tromposaund**

Location: **Holzhausen**

Music Genres: **blues,brassmusic,folkmusic,jazz,swing**

Event Type(s): **birthday,fairground,weddingparty**

We are a **band,brassband**

Mention in the description some cities within a **500** km radius (without specifying the distance) to which we can travel for gigs, and also mention suitable well-known bands or songs if you believe we play them. Bands and songs should be enclosed in [].

The text should be concise but include all the important information.

### Output with a triplet-style prompt

Hey there, we're Tromposaund! Hailing from the small town of Holzhausen, we're a brass band with a love for blues, folk, jazz, and swing music. Whether you need some lively tunes for a birthday celebration, fairground event, or wedding party, we're ready to bring our unique sound and energetic vibe to your special occasion. With a distance of 500.0, we're willing to travel to deliver our uplifting music wherever it's needed. Book us for an unforgettable musical experience!

### Output with a text-style prompt

Tromposaund is a dynamic brass band based in Holzhausen, specializing in blues, brass music, folk music, jazz, and swing. Whether it's a birthday, fairground, or wedding party, we bring the energy to get everyone on their feet. We are available for gigs in cities within a 500.0km radius, such as Berlin, Munich, and Prague. Our repertoire includes classics from well-known bands like [Louis Armstrong and His Hot Five] and [Duke Ellington], ensuring a lively and entertaining performance for all.

## Example prompt – GPT-4

### Output with a text-style prompt

Introducing Tromposaund, the eclectic ensemble hailing from the quaint village of Holzhausen, renowned for their vibrant fusion of musical styles that span blues, brass music, folk music, jazz, and swing. As a brass band with a unique flair, Tromposaund delivers an unforgettable auditory experience that resonates with the rustic charm of their origins.

Whether you're planning a birthday bash, a lively fairground event, or dreaming up your ideal wedding party, Tromposaund is poised to bring the musical heartbeat to your celebration. The band's versatility ensures that each performance is tailored to the festive atmosphere of your event, promising to leave your guests enchanted and your special occasions imbued with a toetapping rhythm.

Tromposaund is no stranger to the road, their music a passport to cities within a 500.0km radius of Holzhausen. Whether it's the vibrant cultural hubs of Frankfurt or Munich, or the historical richness of Prague and Vienna, Tromposaund is ready to journey forth and bring their soulful sounds to your doorstep.

Fans of time-honored tunes and legendary acts will delight in Tromposaund's repertoire, which often includes nods to iconic bands and songs. Imagine the smooth, sultry tones of [Miles Davis]'s trumpet, the foot-stomping energy of [Glenn Miller]'s swing classics, or the heartfelt narratives woven through [Bob Dylan]'s folk ballads. And, for those who revel in the power of brass, expect homages to the likes of [The Dirty Dozen Brass Band], echoing the festive streets of New Orleans right to your event.

Prepare for a melodic journey with Tromposaund - where passion meets tradition, and every performance is a celebration of musical heritage.

# Example prompt – LLaMa2

## Output with a text-style prompt

Introducing Tromposaund - the unparalleled brass band that's sweeping the nation with their intoxicating blend of blues, jazz, swing, and folk music! Hailing from the charming town of Holzhausen, these sonic visionaries are ready to transport audiences to a world of euphoria and enchantment.

With a dynamic repertoire that boasts an impressive array of genres, Tromposaund seamlessly weaves together the rich textures of blues, the soulful melodies of jazz, the lively energy of swing, and the heartfelt melancholy of folk music. Whether it's a birthday celebration, a fairground spectacle, or a luxurious wedding party, this band is guaranteed to leave a lasting impression on any audience.

As a quintessential brass band, Tromposaund is comprised of talented musicians who have honed their craft in some of the most prestigious musical institutions in the land. Their mastery of brass instruments, combined with their unparalleled showmanship, ensures that every performance is an unforgettable experience.

If you're searching for a band that can bring the house down with a medley of classic tunes, look no further than Tromposaund! Their repertoire includes timeless anthems like Louis Armstrong's "When the Saints Go Marching In," Duke Ellington's "Take the A Train," and Count Basie's "One O'Clock Jump," all of which are sure to get your party started! And if you're looking for something a little more contemporary, Tromposaund can also deliver with their own soulful takes on modern hits.

As for their travel radius, Tromposaund is more than happy to bring their infectious beats and melodies to any location within a 500km radius. They've already captivated audiences in nearby cities like Frankfurt, Cologne, and Heidelberg, and they're ready to do the same in your hometown!

So if you're in search of a brass band that's guaranteed to leave your guests dancing, singing, and begging for more, look no further than Tromposaund!