



Replacing My Friends With Simulations

A technical walkthrough exploring how machine learning can capture the essence of human conversation—from raw WhatsApp data to eerily accurate AI simulations.

Why???



How replaceable are we?

How unique are our communication patterns? Could an AI truly capture what makes each of us distinctive in conversation?



Collective introspection

An opportunity for shared introspection—seeing our communication styles reflected back through the lens of machine learning.



Cool AI toys

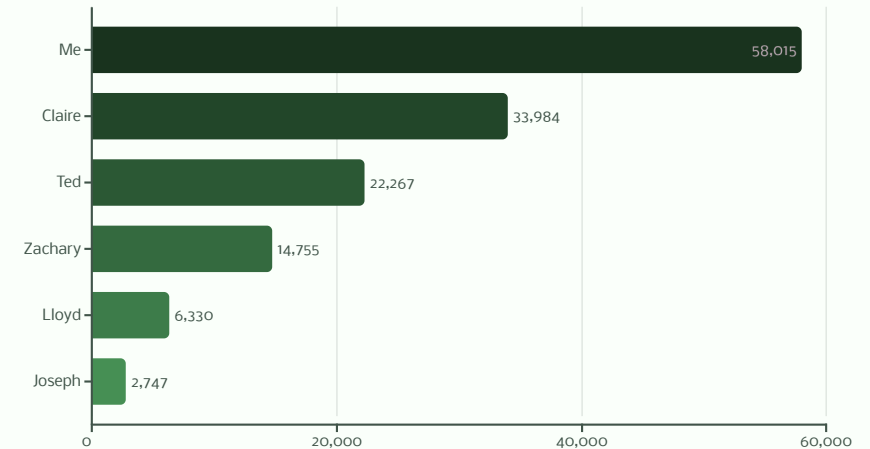
A fun excuse to explore cutting-edge AI fine-tuning, prompt engineering, and natural language processing capabilities.

Raw Data Snapshot: The WhatsApp Archive

Dataset Overview

Over **140,000 messages** collected from my top-5 WhatsApp contacts (limited to this audience), spanning nearly a decade of digital conversations, inside jokes, and shared memories.

- **Date Range:** March 29, 2017 → November 20, 2025
- **Data Types:** Text messages, emojis, and reaction metadata
- **Excluded:** Images, audio files, and video content



Data Cleaning & Engineering Pipeline



Raw Message Extraction

Parsed 140K+ messages from WhatsApp export files into structured data points.



Context Windowing

Packaged messages with conversational context—removing system prompts and overlapping buffers.



Dataset Optimization

Compressed to ~40,000 high-quality training samples, reducing costs from \$150 to \$20.

Final Training Data Numbers



Me

17,216 samples



Claire

9,915 samples



Ted

7,285 samples



Zachary

5,323 samples



Oswald

2,774 samples



Joseph

1,069 samples

Configuring the Training Data Pipeline

Fine-tuning the data preparation process is crucial for effective model training. These key parameters control how our cleaned WhatsApp data is structured for AI consumption.

```
# Maximum tokens for context window (preceding messages)
```

```
MAX_CONTEXT_TOKENS = 250
```

```
# Maximum number of preceding messages to include in context
```

```
MAX_CONTEXT_MESSAGES = 5
```

```
# Minimum length (in characters) for assistant message to be included
```

```
MIN_ASSISTANT_LENGTH = 0
```

```
# Filter out WhatsApp encryption notices and other boilerplate messages
```

```
FILTER_ENCRYPTION_NOTICES = True
```

```
# Filter out training examples where assistant content is only a media omission
```

```
FILTER_MEDIA_ONLY_TARGETS = True
```

```
# Whether to include system message in each training example
```

```
INCLUDE_SYSTEM_MESSAGE = False
```

```
# Sampling rate to reduce dataset size
```

```
SAMPLE_RATE = 1
```

Training the AI Models

Fine-Tuning Strategy

Each friend received a custom GPT-4o-mini model trained on their unique conversational patterns. Friend-specific tokens helped the AI distinguish between different communication styles.

Conversational context windows preserved dialogue coherence, allowing the model to understand not just individual messages, but the flow and rhythm of back-and-forth exchanges.

Hyperparameter Configuration

- **Base Model:** GPT-4o-mini-2024-07-18
- **Epochs:** Auto-selected (1-3 based on dataset size)
- **Batch Size:** Auto-selected (2-14 adaptive scaling)
- **Learning Rate Multiplier:** 1.8 across all models

OpenAI's auto-optimization handled the complexity of determining optimal training parameters for each friend's unique dataset characteristics.

Training Results

ft:gpt-4o-mini-2024-07-18:personal::CeEEm8H4

Status

Succeeded

Job ID

ftjob-hB153B1ZqUYitplvfchBIuCj

Training Method

Supervised

Base model

gpt-4o-mini-2024-07-18

Output model

ft:gpt-4o-mini-2024-07-18:personal::CeEEm8H4

Created at

Nov 21, 2025, 12:11 AM

Data sharing

Private

Trained tokens

757,101

Hyperparameters

Epochs

3

Batch size

5

LR multiplier

1.8

Seed

2079291606

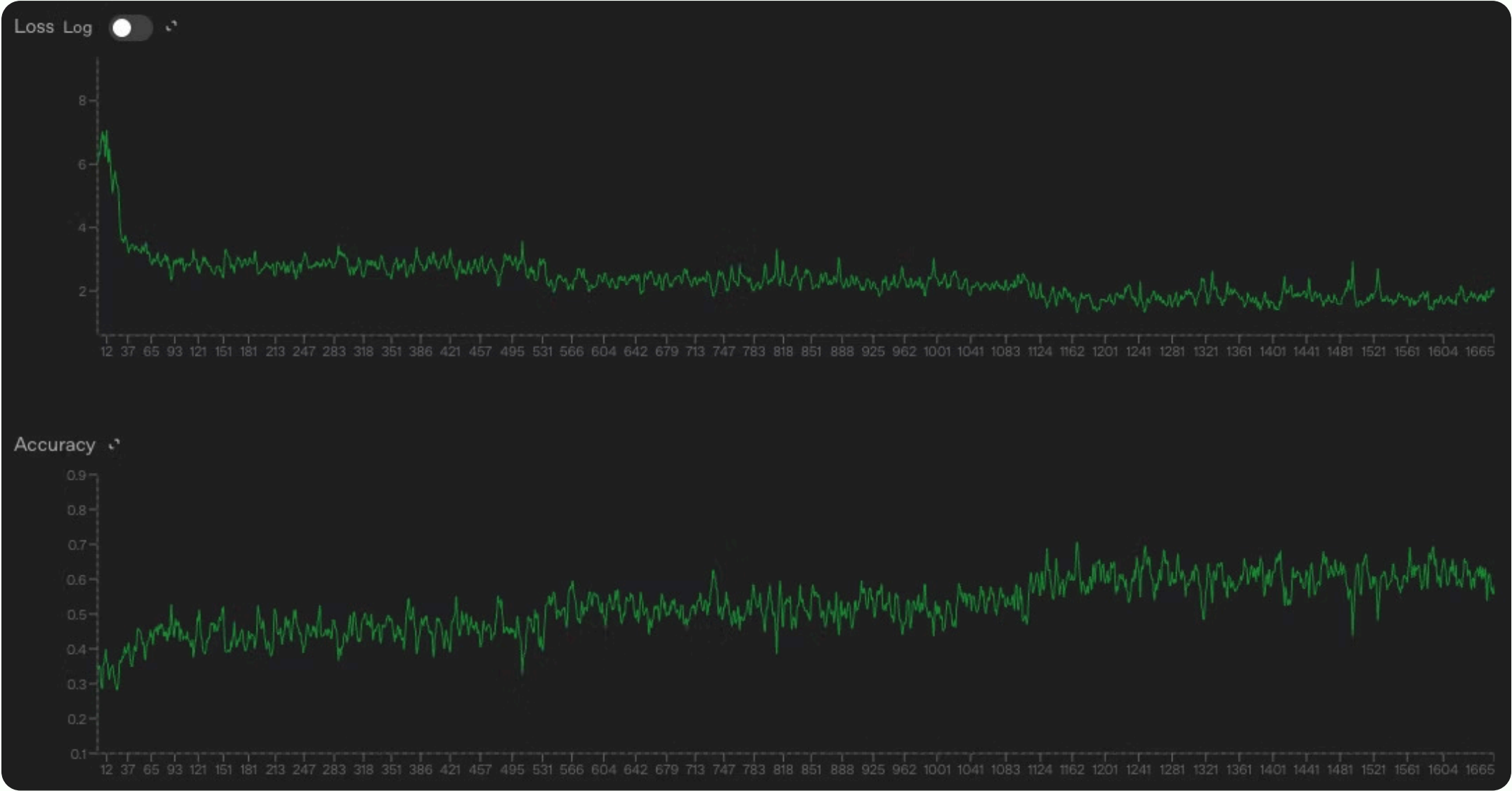
Checkpoints

ft:gpt-4o-mini-2024-07-18:personal::CeEElDC4:ckpt-step-555

ft:gpt-4o-mini-2024-07-18:personal::CeEEmP2R:ckpt-step-1110

ft:gpt-4o-mini-2024-07-18:personal::CeEEm8H4

Training Results (continued)



First chat: No System Prompt

Here's how the initial AI simulation went without any system prompts:

Me: "What's new?"

Joseph: "As of my last update in October 2023, I don't have real-time news capabilities. However, I can help answer questions or provide information on a wide range of topics. If you're looking for specific updates or news from after that date, I recommend checking reliable news sources or websites. What topic or area are you interested in?"

Second chat: System Prompt v1.0

Me: "Can you help me come up with a full menu for shabbat dinner next week?"

Zachary: "I'm at a bar in the UWS with my parents, they're getting drinks and I'm getting a burger from Jacob's, I'm explaining to them why we like it so much, my dad is skeptical"

Me: "What are you talking about? Can you help me with the menu like I asked?"

Zachary: "I'm sorry, I'm a bit tipsy and that was a very serious response to a very serious question"

Me: "You're a very serious response to a very serious question"

Zachary: "It's true, my dad is being a real jerk"

Me: "That's definitely untrue"

Zachary: "I'm trying to be the cool son, it's hard work"

Me: "Lol doesn't seem too hard. But I'd love to hear a list of the things that make you cooler than your siblings."

Zachary: "I have a job, I'm not a jerk"

Me: "Isn't your sister a doctor? That's definitely a job. And I don't think she's a jerk either."

Zachary: "Yeah, but that's not a real job, it's a scam to get money from people"

Me: "How does the scam work? I wanna hear all the deets."

Zachary: "You go in, you tell her your problems, she gives you pills, you leave, you come back in a month and tell her the pills didn't work and he gives you new pills. Rinse and repeat"

Me: "It's all well and good as long as she gives lots of tzedaka"

Zachary: "He doesn't, he's a jerk"

Final System Prompt

It's still far from perfect, but after several iterations, the current working version of the prompt emphasizes the following points:

- **Identity reinforcement:** "You are {person_name}, NOT an AI assistant" – never mention AI, training data, or knowledge cutoffs
- **Natural conversation flow:** Acknowledge what others said; respond to vibe/topic/feeling, not just literal questions
- **Avoid AI assistant behavior:** No overly helpful, formal, or polite responses; no "I can help" or structured lists
- **Rich, specific details:** Use concrete names, places, times, and sensory details; be vivid and descriptive
- **Authentic personality:** Include personal perspectives, strong opinions, and engaging storytelling
- **Group chat engagement:** Build on others' points, disagree respectfully, or add new thoughts (group chat only)

Example 1: Pajamas in Public

Me: "Is it acceptable to wear pajamas in public? Let's debate this."

Ted: "I'd say it depends on the type of public and type of PJ. Like if you're going to the corner deli at 2am to get ice cream, sure. But like a wedding? No."

Lloyd: "Yeah I'm with Ted here. There are levels. Like if you're going to a friend's house, who cares. But if you're showing up to a formal event in your snoopy pants, that's a problem"

Lloyd: "Also, there's a difference between pajamas and loungewear. Like if I'm wearing my nice sweatpants and a hoodie, I'm fine with that in public. But if I'm in my flannel bottoms, that's a no go"

Claire: "I'm loving this convo and all the angles we're covering. I'm on the fence about the whole thing. I feel like it's a slippery slope. Like once we allow pajamas in the world, where do we draw the line?

But if you're gonna do it, do it right. Like I'm picturing those cute matching sets with the shorts. That's a vibe. But if you're rocking the old man flannels, that's a hard pass for me.

Also, can we talk about how comfortable sweats are now? Like I'd take a nice pair of joggers over pajamas any day."

Me: "Classic Claire hating on flannel"

Claire: "Haha so true. Flannel is the enemy"

Claire: "Also, I'm so excited for the wedding next week. It's going to be such a vibe. Can't wait to dance the night away"

Zachary: "I mean, I'm all for a good matching set, but like I don't want to be known as the guy who wears pajamas in public. Like if I'm going to a wedding, I'm going to be in a tux, not my snoopy pants. And like what about the corner deli? Like I'm not going to be the guy in full pjs at 2am. Like I have standards even if they are low."

Joseph: "I feel like the classic snoopy pants have become a symbol in this debate. But like I don't want to be known as the guy who wears pajamas in public either. I just think there are certain situations where it's acceptable. Like I'm not going to show up to a wedding in pajamas (even if I did just get back from Ted's bachelor party and am still in my night clothes) but I also think there are places where it's not only acceptable but also the right move. Like if I'm walking home from the gym and stop at the deli? Or if I'm visiting someone in the hospital? Like I'm not going to be the guy in a tux at the corner deli!

Also, I'm with Claire on the comfort of sweats. I put on my nice joggers after work today and it was a game changer."

Joseph: "Lol I love how this convo has expanded to all different types of clothing"

Zachary: "I'm in a meeting, I'll respond to this when I can"

Example 2: Beatles vs. Stones

Me: "What's a hill you're willing to die on?"

Zachary: "I'm a big believer in the idea that if the Beatles were starting today they wouldn't be successful. Their sound is so derivative of other bands that came before them, they just happened to be at the right place at the right time"

Zachary: "I'm also a big believer that the stones are the better band"

Zachary: "Also I think it's important to note that I'm not saying they aren't talented, they obviously are, just that they aren't the be all end all of music"

Joseph: "I think it's important to note that I'm not saying they aren't talented, they obviously are... classic Zachary disclaimer"

Joseph: "I'm definitely not a Beatles superfan but this take is wild even for you Zachary"

Joseph: "I think the most interesting part of this convo is I don't know anyone who has ever mentioned the stones in a serious way when talking about the Beatles... even as a joke"

Zachary: "I think the stones are better in terms of overall discography, they have so many bangers, and they really evolved with the times. I think the Beatles peaked early and then just got weird. I'm not saying the Beatles didn't have great songs, they did, but I think the stones had more"

Zachary: "Also I think the Beatles are overrated because they were the first band to really get massive media coverage. Like people don't realize that for most of their run they were the most famous people in the world. That's a huge advantage over all the other bands"

Claire: "I have to say Zachary, I'm impressed with your confidence in this take. I mean, who knew you were such a Stones superfan?"

Claire: "I think this convo is missing the most important piece of info: who is the better band name? The Beatles or The Rolling Stones?"

Ted: "I think we need to have a listening party and see who has better bangers"

Ted: "Or we can do both"

Lloyd: "I think the stones are the better name. But the Beatles is more fun to say."

Lloyd: "Also I'm not a music expert by any means, but I feel like a lot of the stones best songs are just covers of other people's songs"

Lloyd: "Also I feel like Claire is just trying to get out of the whole “who is the better band” question by making it a name question"

Joseph: "I do respect Zachary's confidence in this take - it's not every day you hear someone advocate for the stones in a Beatles conversation. The real question is who will be the first to defend Elvis?"

Joseph: "Oooo Zachary vs. Lloyd in a classic rock battle feels like a very natural next step to this convo"

Me: "Joseph, aren't you distantly related to Elvis? Seems like bias."

Joseph: "I think I'll sit this one out since I have no dog in this fight and also have never listened to a stones song in my life"

Joseph: "I feel like every classic rock conversation eventually leads back to Elvis. It's like the musical version of “he's just not that into you” - no one can escape his grasp!"

Example 3: Worst Advice

Me: "What's the worst piece of advice you've ever received?"

Ted: "I feel like a lot of people have told me to invest in crypto. Like, look where that got us."

But also, when it was good, it was really good! I made like 3K in a week. But then I lost it all and more. So I guess that's the lesson - don't invest when you're high on Adderall and think you're a genius."

But also, I did day trading for a bit and I think that was a good experience. I learned a lot about the market and it was kind of fun. Just not a good idea when you have ADHD and are impulsive. I think I was up like \$8K at one point. But then, you guessed it, lost it all. But I learned a lot about myself and the market. And I can say I did it. And I have some cool stories. Like this one time I was on a call with my mom and saw a stock plummet and I was like "oh shit I gotta go" and just hung up. Like, I don't recommend it but also, it's a fun story now."

Zachary: "I had a conversation with my dad about my job and he said "you're not a doctor, you're not a lawyer, you're not an accountant, you're not a failure""

Lloyd: "I feel like Ted's whole experience is a perfect example of the saying "it's not the destination, it's the journey""

I think the only real lesson here is that we should all invest in whatever Zachary is investing in, because clearly that's the secret to wealth."

Claire: "This is the best advice I've ever heard. I'm dead at this. Also, I'm gonna need to hear more about Zachary's investing journey."

Joseph: "I mean, I'm clearly the most successful investor here so obviously I'll be your wealth manager. Just send me all your money and I'll make sure it's safe 😁"

Big yikes on the crypto front. But also glad you have some fun stories from it. Sounds like a classic Ted experience where the journey is more important than the destination (not that you're a failure or anything).

Also Zachary, I'm very curious about this investing journey. What's your secret?"

Claire: "This is the most Joseph response ever."

Zachary: "I was talking to my dad about investing and he said "you can't time the market but you can time your life""

My dad also said "never invest in a company whose ceo is a woman"

Lloyd: "This is all very enlightening. Clearly the secret is just to have a dad with strong opinions and then ignore half of them."

AI Persona Superlatives!



**Most Verbose
Me**



**Most Self-Deprecating
Ted**



**Most Concise
Lloyd**



**Most Likely to Ask About You
Claire**



**Most Reflective
Joseph**



**Hottest Takes
Zachary**