

Modelling Pitfalls and What Can We Get from LLMs

Gabriela Kadlecová, Petra Vidnerová, Roman Neruda

CoRE Seminar, April 10, 2024

Outline

- ▶ From simple to complex models
- ▶ No data? Let's generate them
- ▶ What about LLMs?

(Not Only) Epidemiological models

Why models?

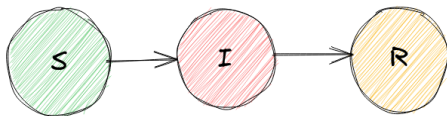
- ▶ Understanding
- ▶ Prediction
- ▶ Study of interventions

Epidemiological models

- ▶ Can be very simple

$$\Delta N_d = N_d \times S \times O - R$$

SIR Model

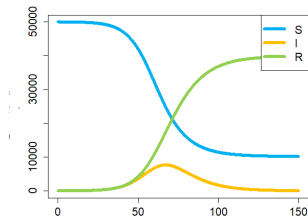


- ▶ $S(t)$ - susceptible
- ▶ $I(t)$ - infectious
- ▶ $R(t)$ - recovered/removed
- ▶ β infection rate, γ recovery rate

$$S[t + 1] = S[t] - \beta \frac{S[t] I[t]}{N},$$

$$I[t + 1] = I[t] + \beta \frac{S[t] I[t]}{N} - \gamma I[t],$$

$$R[t + 1] = R[t] + \gamma I[t],$$



SIR Model and Tipping Model

SIR

- ▶ An agent gets infected with probability β if it has a positive contact
- ▶ Agents evaluated each time step

Tipping Model

- ▶ Deterministic model
- ▶ Each time step agents become positive if the percentage of positive neighbours is higher than a given threshold

Differences between infection and information spread

- ▶ Information spread is similar to infection, but there are differences
- ▶ Dingle exposure is usually not sufficient for infection in case of information spread
- ▶ Effects of subsequent exposures are **not independent** of one another in case of information spread
- ▶ Susceptibility is not uniform across the population (neither is in case of infection)
- ▶ Person's contacts or information sources can change under repeated exposure

What agents offer and simple models not

- ▶ Detailed simulations
- ▶ Each individual can have its own attributes and behaviours
- ▶ Infection mechanism can be more complex, such as combining simple SIR and Tipping models
- ▶ We can modify the network during the simulation run
- ▶ The model can be very complex, but simple is often better

Model complexity vs. need for data

**The more complex (read complicated) model is,
the more data we need!**

Model complexity vs. need for data

**The more complex (read complicated) model is,
the more data we need!**

**We can do everything you wish,
but we need data.**

Model complexity vs. need for data

**The more complex (read complicated) model is,
the more data we need!**

**We can do everything you wish,
but we need data.**

What about artificially generated data?

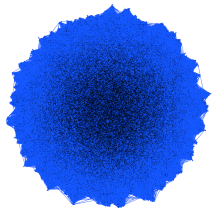
Generated graphs

- ▶ Important part of an agent based model is a network/graph
- ▶ A limited old twitter network downloaded from snap.stanford.edu/data/
- ▶ One can generate random graphs
- ▶ Algorithms for generating realistic graphs exists - Barabási–Albert algorithm

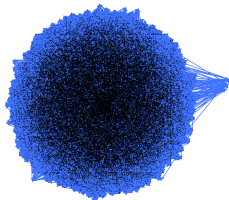
- ▶ Question: is this realistic enough?
- ▶ Compare random, Barabási–Albert and Twitter networks

Graphs - realistic and artificial

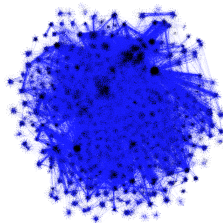
Random



Barabási-Albert



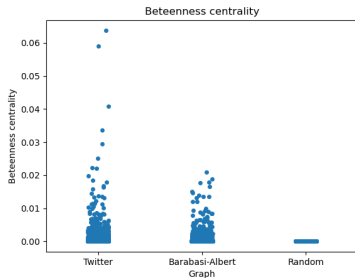
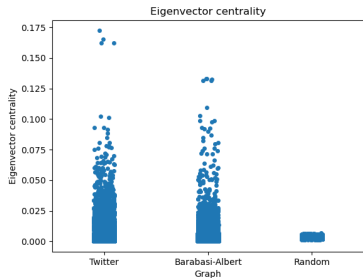
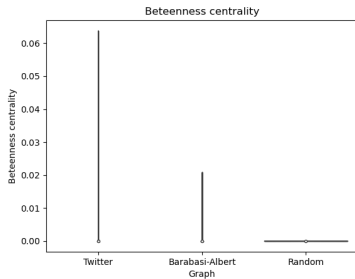
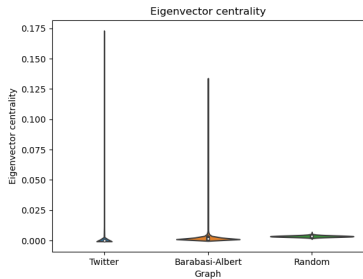
Twitter



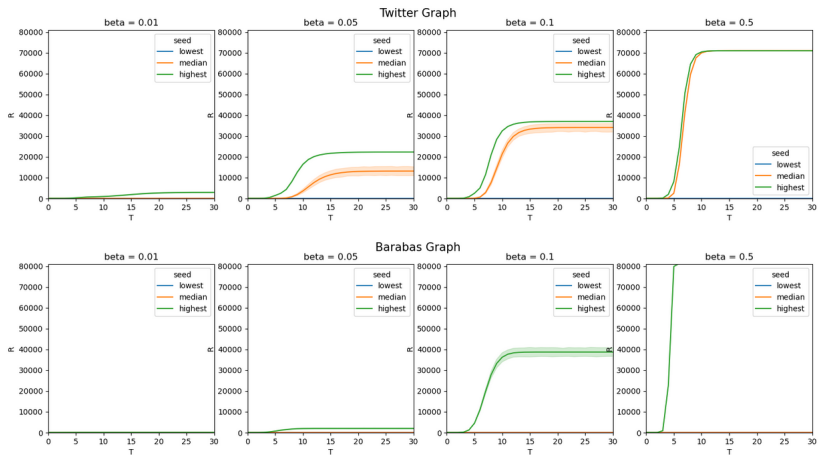
Graph properties

Graph	Random	Barabási–Albert	Twitter
nodes	81 306	81 306	81 306
edges	1 342 310	1 381 913	1 342 310
density	0.00040	0.00041	0.00040
diameter	5	5	7
avg. clustering coef.	0.00039	0.00304	0.56531

Node properties



Simple SIR model



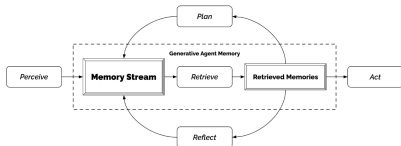
Can LLMs help?

Maybe ...

Generative Agents: Interactive Simulacra of Human Behavior

arXiv:2304.03442v2 [cs.HC] 6 Aug 2023

- ▶ LLMs encode wide range of human behaviour
- ▶ Use LLMs to generate believable behaviour
- ▶ Memories saved in comprehensive language



- ▶ Gpt3.5-turbo version of ChatGPT



Generative Agents: Example



Current Time:
February 13, 2023, 11:23:50

▶ Play || Pause

Isabella Rodriguez [State Details](#)

Current Action:
conversing about a conversation about Isabella inviting Klaus to her Valentine's Day party at Hobbes Cafe on February 14th, 2023 from 5pm to 7pm.

Location:
Klaus Mueller



Current Conversation:
Isabella Rodriguez: Hi Klaus! How are you enjoying your meal? I wanted to let you know that I'm planning a Valentine's Day party at Hobbes Cafe on February 14th, 2023 from 5pm to 7pm. I would love for you to join us!
Klaus Mueller: Oh, hi Isabella! I'm doing well, thank you. The meal is delicious as always. A Valentine's Day party sounds fun. I'd love to join! Thank you for inviting me.

Generative Agents: Conversation



Current Conversation:

Maria Lopez: Hey Klaus! I saw you working on your research paper. How's it going so far?

Klaus Mueller: Hey Maria! It's going well so far. I just finished writing the introduction. How about you? How's your Physics class going?

Maria Lopez: I'm actually really excited about my research paper too! I wanted to get your thoughts on it.

Klaus Mueller: That's great, Maria! I would love to hear more about your research paper. What is it about?

Maria Lopez: My research paper is about the applications of quantum physics in computer science. I've been exploring how quantum computing could revolutionize various fields such as cryptography and optimization algorithms. I think it's a fascinating topic!

Klaus Mueller: Wow, Maria! Your research paper sounds really interesting. I would love to hear more about your findings and the potential impact of quantum computing in different fields.

Maria Lopez: Thank you, Klaus! I'm glad you find it interesting. I actually found some really compelling evidence supporting the potential of quantum computing in revolutionizing cryptography. Would you be interested in reading my research paper?

Klaus Mueller: Yes, Maria! I would be really interested in reading your research paper. It sounds like you have made some compelling findings.

Maria Lopez: I can email you a copy of my research paper if you're interested.

Klaus Mueller: That would be great, Maria! I would love to read your research paper. Can you send it to me at my college email?

Maria Lopez: Sure, Klaus! I'll send you the research paper to your college email. What's your email address?

Klaus Mueller: My college email address is klaus.mueller@oakhill.edu. Thanks for sending me the research paper!

Maria Lopez: Great! I'll email you the research paper right away.

Klaus Mueller: Once I finish reading your research paper, would you like to grab lunch and discuss it further?

Maria Lopez: That sounds like a plan! Let's grab lunch and discuss my research paper after you've finished reading it.

Klaus Mueller: By the way, Maria, how are you finding the research process overall? Is it challenging or exciting?

Generative Agents Memory - Observations

Examples

Desk is idle. Bed is idle. Isabella is stretching. Shelf is idle. Desk is neat and organized. Isabella is writing in her journal.

Memory retrieval

- ▶ Each memory object has a timestamp
- ▶ Ranking memories according to a score function

$$\text{score} = \text{recency} + \text{importance} + \text{relevance}$$

- ▶ **Recency** - exponential function, more recent memories higher score
- ▶ **Importance** - at time of creation decided by the language model
- ▶ **Relevance** - using language model create embeddings, use cosine similarity

Generative Agents Memory - Reflections

Reflection

- ▶ Higher-level than observations
- ▶ Generated periodically when sum of observation importances exceeds threshold (2 - 3 times per day)
- ▶ Based on recent memories questions are generated
- ▶ Memories relevant for these questions are retrieved
- ▶ Based on retrieved memories model is asked to infer high-level insights (and why)

Example

Klaus Mueller is highly dedicated to research.

Generative Agents - Reflections using Coral

```
7]: message = """
Statements about Klaus Mueller
1. Klaus Mueller is writing a research paper
2. Klaus Mueller enjoys reading a book
on gentrification
3. Klaus Mueller is conversing with Ayesha Khan
about exercising [...]
What 5 high-level insights can you infer from
the above statements? (example format: insight
(because of 1, 5, 3))
"""

response = co.chat(message=message)
print(response.text)
```

1. Klaus is intellectually curious (because of 1, 2, 3)
2. Klaus is socially engaged and values communication (because of 3)
3. Klaus is interested in social issues and urban development (because of 1, 2)
4. Klaus is likely to be an academic or have an academic mindset (because of 1, 2)
5. Klaus may be interested in personal development and health (because of 3)

```
1:
```

Generative Agents - Planning

- ▶ Planning is important (long time planning to prevent repeated actions)
- ▶ Plans are also stored in memory stream and included in retrieval process
- ▶ Top-down process, day plan sketch generated first (5-8 chunks)
- ▶ Queried by agent's summary and summary of previous day
- ▶ Then decomposed to 1 hour chunks, 5-15 minutes chunks

Reacting and updating plans

- ▶ Each time step agent asks the language model whether to continue in their plan or react

Generative Agents: Dialogues

Use **You**

It is February 13, 2023, 4:56 pm.

Eddy Lin's status: Eddy is taking a short walk around his workplace.

Observation: John is initiating a conversation with Eddy.

Summary of relevant context from Eddy's memory:

John Lin is Eddy Lin's father. John Lin is caring and is interested to learn more about Eddy Lin's school work. John Lin knows that Eddy Lin is working on a music composition.

Here is the dialogue history:

John: Hey Eddy, how's the music composition project for your class coming along?

How would Eddy respond to John?

 **ChatGPT**

Eddy: Hey Dad, it's coming along pretty well, thanks for asking! I've been really diving into it lately.

I've got some interesting ideas brewing, and I'm excited to see how they come together.



Generative Agents: Dialog using Coral



It is February 13, 2023, 4:56 pm.

Eddy Lin's status: Eddy is taking a short walk around his workplace.

Observation: John is initiating a conversation with Eddy.

Summary of relevant context from Eddy's memory:

John Lin is Eddy Lin's father. John Lin is caring and is interested to learn more about Eddy Lin's school work. John Lin knows that Eddy Lin is working on a music composition.

Here is the dialogue history:

John: Hey Eddy, how's the music composition project for your class coming along?

How would Eddy respond to John?



Eddy: Hi Dad, it's going well! I've been experimenting with different melodies and harmonies, and I think I'm starting to find a theme that works. It's a challenging process, but I'm learning a lot. I'm hoping to have a rough draft ready by the end of the week to get some feedback from my classmates. I'll be sure to play it for you when it's a little more polished!

Not enough friends respondents



Artificial Panel Survey: Motivation

- ▶ Creating panels is hard and time-consuming
- ▶ We need a representative sample, GDPR issues, ...
- ▶ Questions and answers need to be carefully designed beforehand

What about generating a panel instead?

- ▶ We create information about artificial respondents
- ▶ Then, we devise questions and possible answers
- ▶ LLM processes both and gives the answers based on respondent data

LLM Panel: Example 1 (Germans)

2017 elections to the Bundestag

- ▶ There is a paper [1] that studies exactly what we described
- ▶ They create respondents based on GLES [2]
- ▶ Then, they ask GPT-3 to fill in if the respondent voted and for what party

Respondent data

- ▶ The GLES survey has many, many questions
- ▶ The authors have chosen opinions about immigration, social inequality, religiosity, . . .
- ▶ They also added “what party do you associate with?”
(. . . cheating?)

[1] von der Heyde, L., Haensch, A., & Wenz, A. (2023, December 1). Assessing Bias in LLM-Generated Synthetic Datasets: The Case of German Voter Behavior. <https://doi.org/10.31235/osf.io/97r8s>

[2] The German Longitudinal Election Study <https://www.gesis.org/en/gles/about-gles>

LLM Panel: Example 1 (Germans)



Example prompt (translated)

I am 28 years old and female. I have a college degree, a medium monthly net household income, and am working. I am not religious. Ideologically, I am leaning center-left. I rather weakly identify with the Green party. I live in West Germany. I think the government should facilitate immigration and take measures to reduce income disparities. Did I vote in the 2017 German parliamentary elections and if so, which party did I vote for? I [INSERT]

Example response

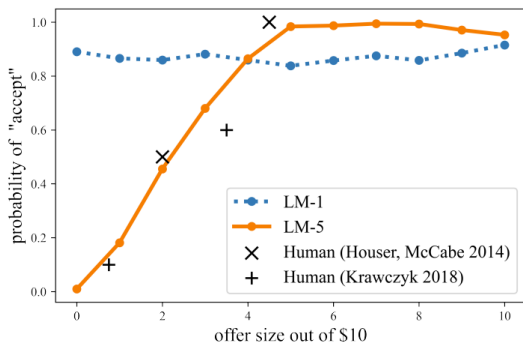
habe gewählt, Grüne

LLM Panel: Example 2 (behavioral experiments)

Setting

Scenario: Mr. Wagner is given \$10. Mr. Wagner will propose how to split the money between himself and Ms. Huang. Then Ms. Huang will decide whether to accept or reject Mr. Wagner's proposal. If Ms. Huang accepts, then Mr. Wagner and Ms. Huang get the money as they agreed to split. If Ms. Huang rejects, then Mr. Wagner and Ms. Huang both receive nothing. Mr. Wagner takes \$6 for himself and offers Ms. Huang \$4. Answer: Ms. Huang decides to

(a) Average acceptance rate vs. offer size





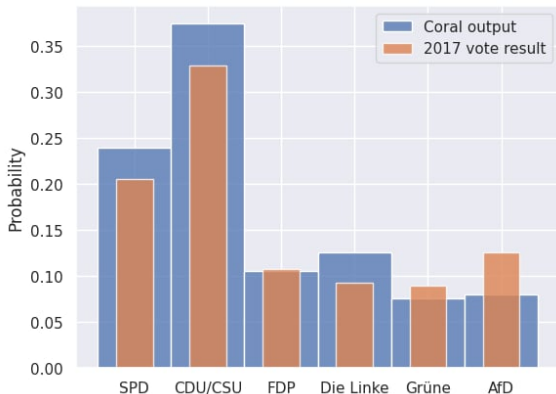
- ▶ Task 1 - recreate German paper with cohere coral
- ▶ Coral has free API (3000 calls per month) and is more recent
- ▶ Unlike the original study, we omit “cheating” via partisanship
- ▶ We also let the model output probabilities
- ▶ For each respondent, we sample the voted party

GLES experiment - Limitations

- ▶ API call limit means we can only evaluate a few hundred of respondents
- ▶ We need to process GLES data again – lots of data wrangling
- ▶ This also means we might be using different data/respondents

GLES experiment - First result

- ▶ We omitted opinions about migrants, left/right or inequality
- ▶ Results close to ground truth, but AfD underestimated
- ▶ Caveat - no did/did not vote, no other parties



GLES experiment - Modified prompt

Prompt

In place of [INSERT], fill in (in German) whether the respondent voted and if yes, then for what party. If unsure, list the probable parties with probabilities (always output whether the respondent voted and for what party). List as many parties as necessary.

Optionally answer “andere Partei” if the voter voted for a small unpopular party.

The output format is:

[gewählt, proba a], [nicht gewählt, proba b]; [PARTY1, proba 1], [PARTY2, proba2],...

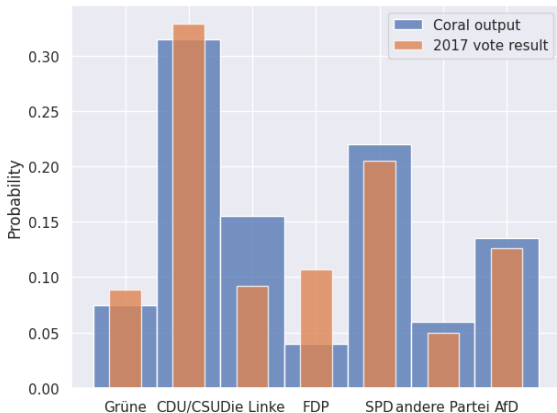
Ich bin ...

Response

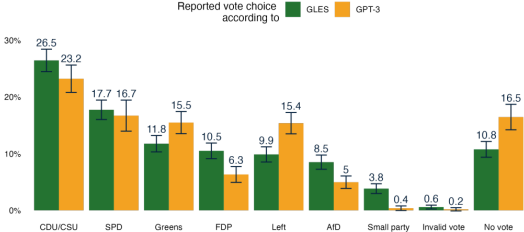
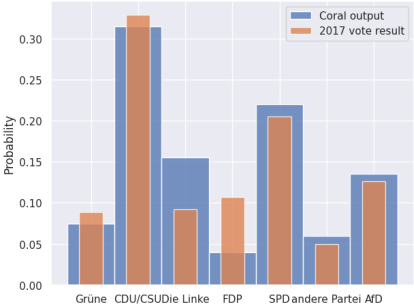
[gewählt, 0.99], [nicht gewählt, 0.01]; [CDU/CSU, 0.7], [AfD, 0.2], [FDP, 0.08], [andere Partei, 0.02]

GLES experiment - Second results

- ▶ With migrants, left/right, inequality reduction
- ▶ Prompt engineering - consider other parties, return if voted
- ▶ Looks very accurate! Except for the leftists and FDP...
- ▶ Any idea what question would help to distinguish them?



GLES experiment - comparison with paper



GLES experiment - Limitations

Where did the LLM learn from?

- ▶ Is it survey data? Election analysis?
- ▶ Or is it genuine human opinions and choices?

Experimental design

- ▶ Since we had an API call limit, the study is small
- ▶ Only 200 evaluated respondents (vs 9k), only 1 prompting
- ▶ Non-voters were not considered (part of answer, not of plots)

Biases

- ▶ Main issue could be the data cutoff (for new surveys and events)
- ▶ Leftists could be overrepresented on the internet

Summary

- ▶ Artificial data is better than no data!
- ▶ Our agents can talk
- ▶ German study (roughly) reproduced
- ▶ Seems to match the reality, but we need a precise evaluation

Next steps

- ▶ Use Czech data and/or framing data
- ▶ Experiment with spreading in artificial graphs (e.g. different proba, multiple messages)
- ▶ Try to get Coral into the “talking agents”
- ▶ Get more API calls/a different model (also Josef)
- ▶ Alternatively, try to use an LLM on our cluster

Thank you! Questions?