

# How to use AI to Analyse your Text Data

Michiel van der Ree  
Center for Information Technology  
team Data Science

House of Connections, April 25



university of  
 groningen

center for  
 information technology

# Doing applied machine learning

Suppose a researcher wants to use machine learning to analyze texts..

Until recently, we had two options

- A trained model is available and we apply it to researcher's data
  - very fast
  - need to be careful about domain shift
  - bit boring?
- Train a model ourselves
  - Requires annotation
    - Requires annotation guidelines, recruiting annotators
    - "How many examples do I need?"
  - Expensive to change your mind

Large Language Models (LLMs) provide a nice middle ground!



university of  
 groningen

center for  
 information technology

# Let's go to the notebook!



**You**

Generate 40 lowercase usernames with a theme of AI in education



**ChatGPT**

Here are 40 lowercase usernames based around the theme of AI in education:

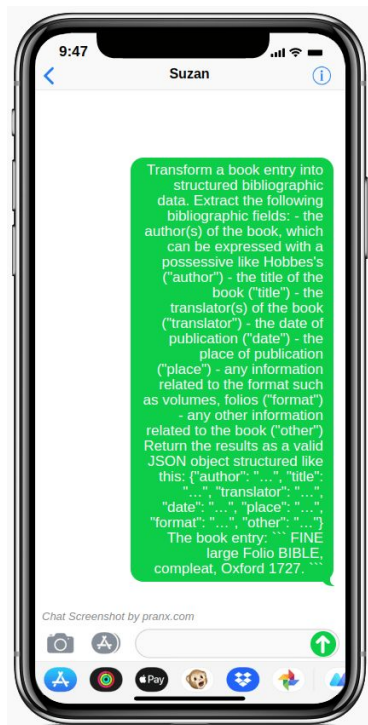
- URL, username on distributed leaflets
- Password is set by using it at first login



university of  
 groningen

center for  
 information technology

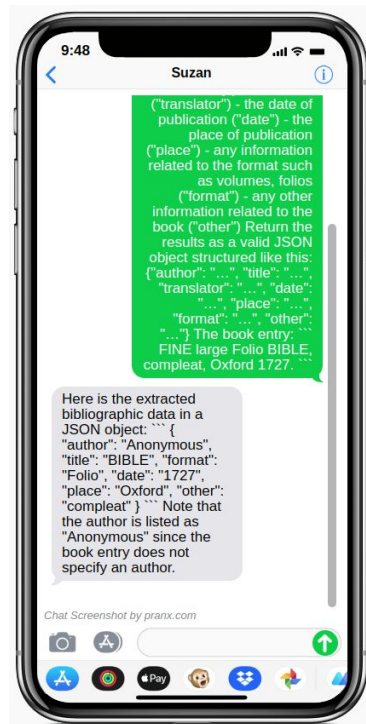
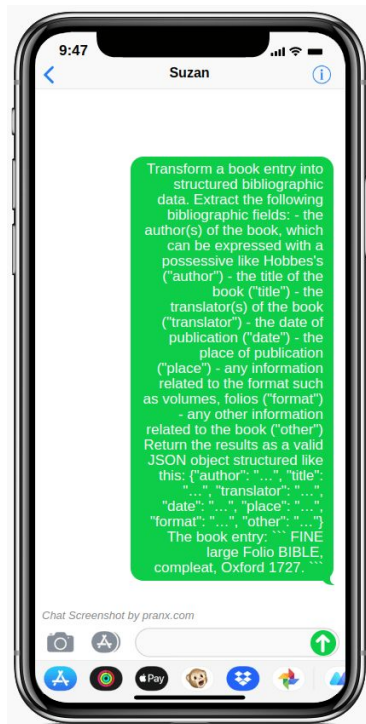
# Changing chat history?



university of  
 groningen

center for  
 information technology

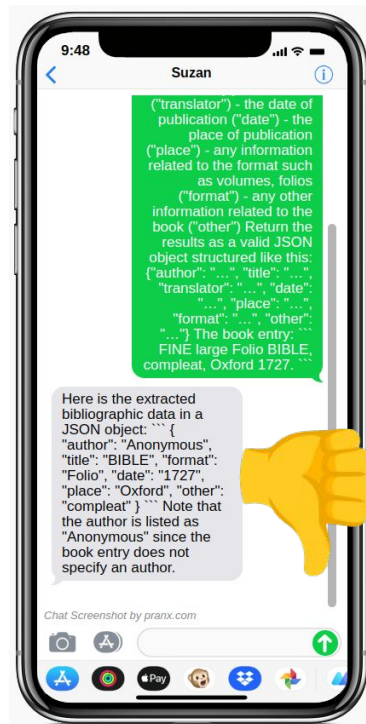
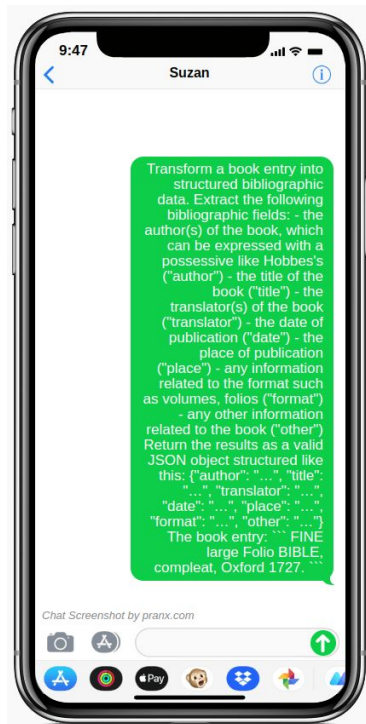
# Changing chat history?



university of  
 groningen

center for  
 information technology

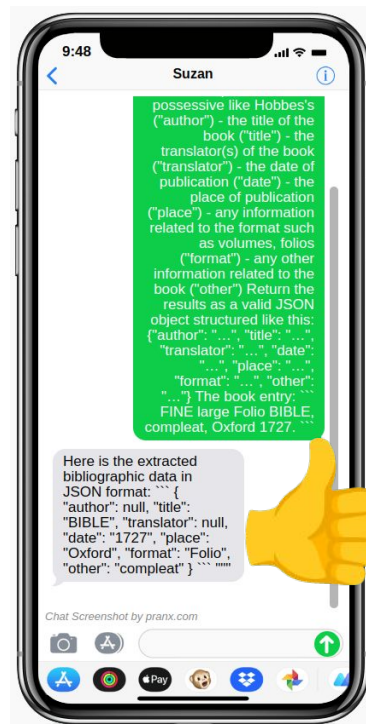
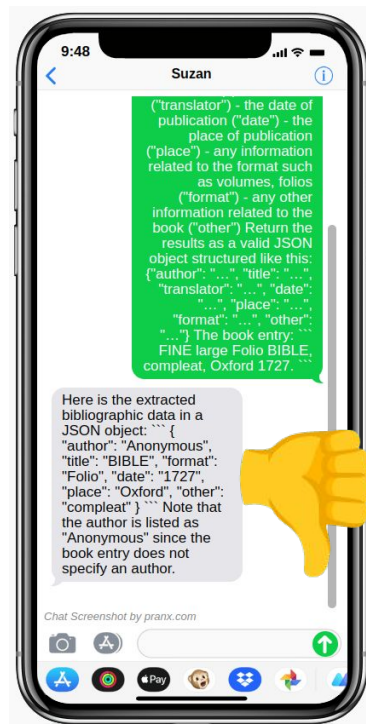
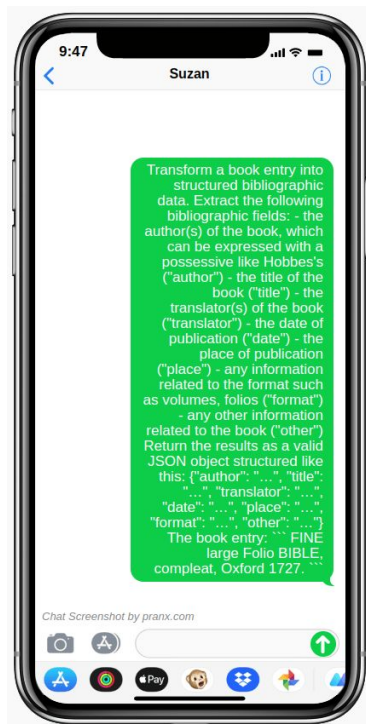
# Changing chat history?



university of  
 groningen

center for  
 information technology

# Changing chat history?

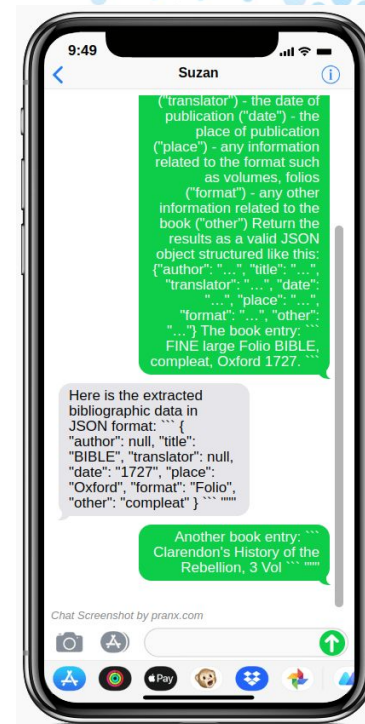
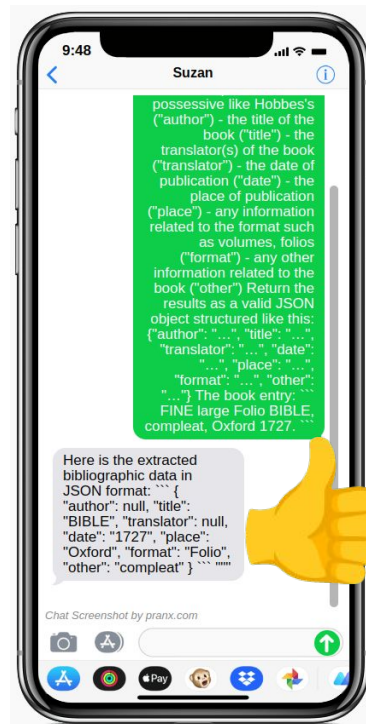
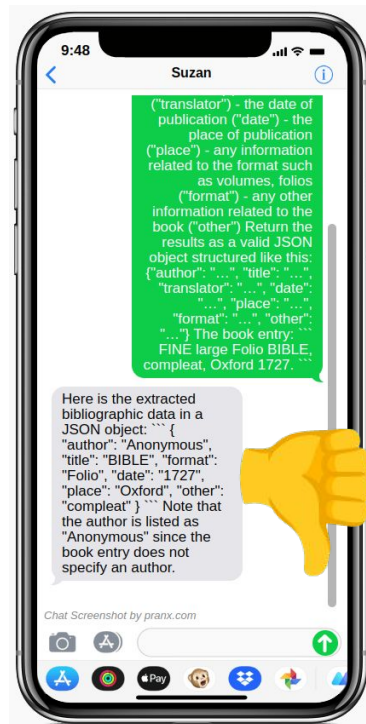
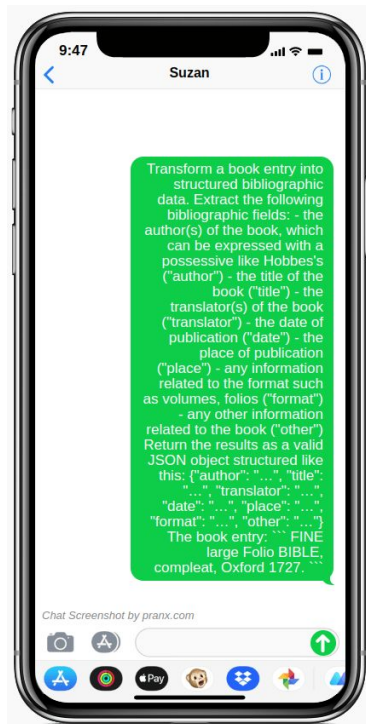


university of  
groningen

center for  
information technology



# Changing chat history?



university of  
 groningen

center for  
 information technology



# Reflection

What potential applications do you see?

What shortcomings did you encounter?

Other points?



university of  
groningen

center for  
information technology

# Many thanks for your participation!

- Materials available at [github.com/UG-Team-Data-Science/ai-in-edu](https://github.com/UG-Team-Data-Science/ai-in-edu)
- Interested in running models on your own hardware?  
Check out [GPT4AI](#)
- In need of support? Reach out at [datascience@rug.nl](mailto:datascience@rug.nl)
- Really want scale up your workflow? UG staff and students can run jobs on the Hábrók compute cluster!



university of  
groningen

center for  
information technology

## Bonus slide: Text generation

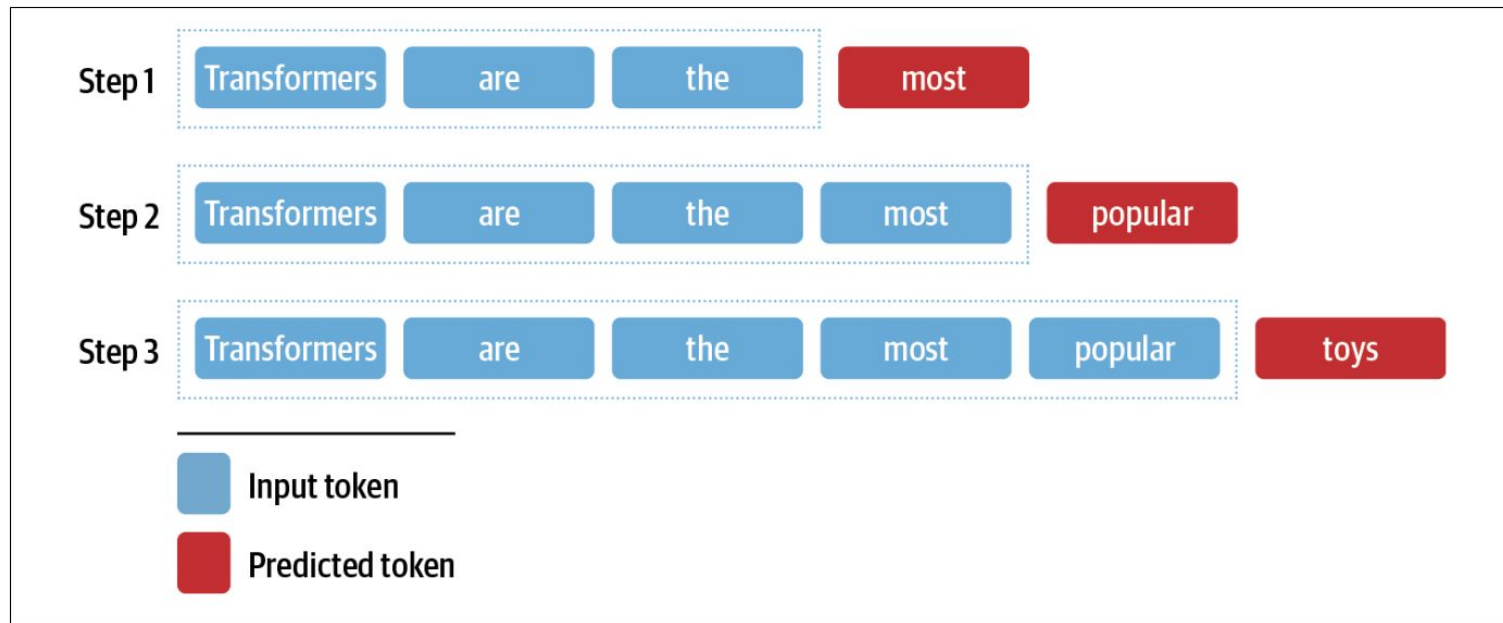


Image from Tunstall et al (2022)



university of  
 groningen

center for  
 information technology