

LLM AS A RESEARCH ASSISTANT: EXPERIMENTING WITH HISTORICAL CLASSIFICATIONS

Liudmila Lyagushkina

PhD researcher,

University of Nottingham (UK)

Goal:

Statistical analysis of the collective experiences of repeatedly arrested victims of Stalinism, 1929-1953

Sources:

Database *Victims of Political Terror in the USSR* (International Memorial, 2017)



Soviet Poster 'Long Live the USSR! Blueprint for the Brotherhood of all Working Classes of all the World's Nationalities!'. 1935.
Source: Wikimedia. Animation: Gemini Veo 2

1. CAN WE RECONSTRUCT PERCEIVED ETHNICITIES FROM NAMES AND BIRTHPLACES?

METHODOLOGY

'TRADITIONAL' MACHINE LEARNING (ML):

- Simple algorithm uses existing examples of classifications from the database to predict classes for new entries (supervised ML)
- Idea: SVM model attempts to determine the optimal boundaries for separating data into classes

LARGE LANGUAGE MODELS (LLMs):

- Top LLM models: OpenAI ChatGPT, Google Gemini, Claude Sonnet
- Samples: 1,000 examples
- Two tests: 1) Names only; 2) Names + birthplaces

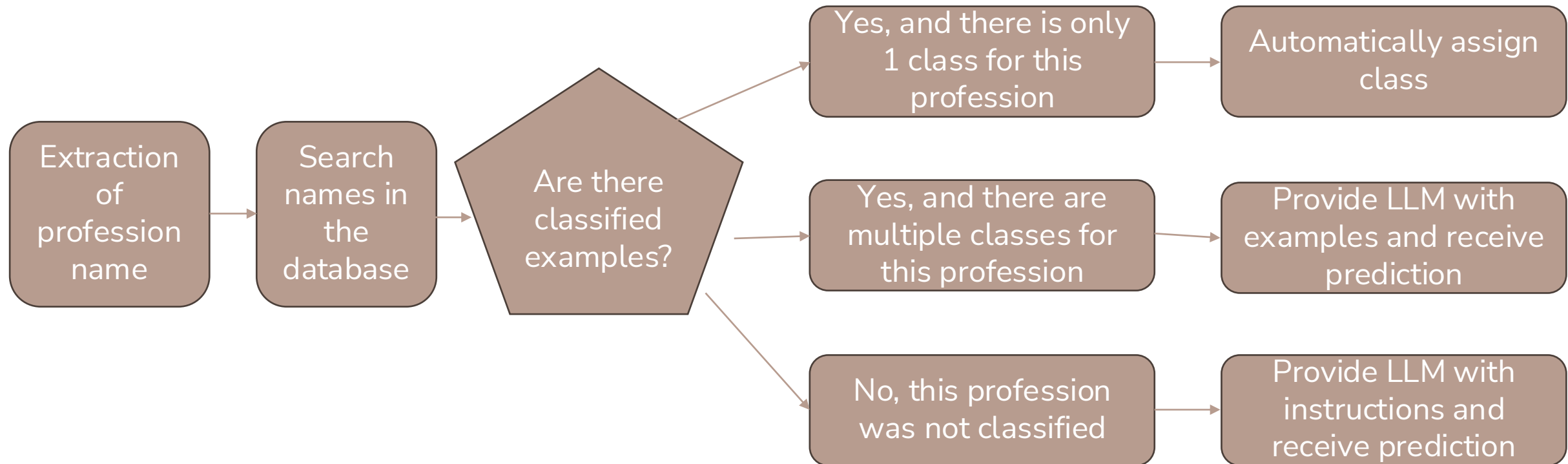
INCONSISTENCIES IN THE DATA

- ‘Traditional’ ML correctly predicted 82% of ethnicities, while LLMs – 78-79%
- ‘Traditional’ ML achieved higher accuracy by under-predicting minorities (defaults to Russian)
- LLMs over-predict minorities (Ukrainians, Poles, Belarusians, and Jews)
- ‘LLM biases’: Reflect contemporary texts, not historical biases
- ‘ML biases’: Do they reflect biases in the database itself?



2. CAN WE GROUP VICTIMS BY THEIR SOCIAL POSITIONS?

‘TEACHING’ TO ‘THINK AS HISTORIANS’: RETRIEVAL AUGMENTED GENERATION (RAG) APPROACH



RESULTS OF CLASSIFICATION

- Both models achieved similar accuracy – 90% and 91% respectively
- LLM outperformed in complex cases, when there were no examples in the training dataset
- Manual verification of the classification results revealed several cases in which LLM was correct, but manually marked data was not
- The RAG approach is easier to improve and adapt to the realities of the 1920s and 1940s

Percentage of correct predictions of social positions

Social position	N of cases in sample	Traditional ML (SVM)	LLM (Gemini)
white-collar workers	509	96%	94%
blue-collar workers	330	90%	87%
collective farm workers	43	77%	79%
non-cooperative artisans	33	82%	88%
clergy	27	96%	100%
military personnel	23	83%	96%
status undefined	11	64%	73%
individual farmers	9	89%	78%
cooperative artisans	8	75%	63%
without a definite occup..	5	80%	20%
housewives, dependents	2	0%	100%
Total	1,000	91%	90%

HISTORICAL REASONING IN THE AGE OF AI

- Traditional ML algorithms are not 'thinking as historians', they 'think' how to fit into the 'training' dataset
- LLMs provide more flexibility in following the instructions, but they are just a tool that needs to be guided on how to solve tasks 'step by step'
- Historians in the age of AI need to do what we do best: critically approach the biases in the sources

Soviet Poster
'Study the
victorious path
of the great
Lenin–Stalin
Party' (1952).
Source:
anticvarium.ru
Animation:
Gemini Veo 2

