Title: Using AI agents to collect biographical data

Project lead(s):
Jérôme Gensel (jerome.gensel@univ-grenoble-alpes.fr), Stories Research Group, LIG
Marlène Villanova (marlene.villanova@univ-grenoble-alpes.fr), Stories Research Group, LIG

Keywords: biographical survey, biographical data, life course, AI agents, schema alignment, ontology population, knowledge graph
Project duration: 6 months
Maximum number of students: 1
Places available: 1

Description

Longitudinal biographical data are data collected during surveys (known as biographical surveys), in which people trained for this task (known as interviewers) ask individuals (known as respondents) about their life course from different perspectives or on different themes: family, work, housing, etc. Each theme is associated with a set of indicators or attributes (for example, address, rent, type of accommodation, etc. for the residential theme). A life course thus traces an individual's life through several thematic (sub-)life courses, and the data are collected using a questionnaire that covers the selected themes and indicators.
Biographical surveys (such as, for example, the Histoires de Vie survey [1]) are generally conducted at regional or national level by national bodies (such as INED or INSEE in France, for example) and collect data on the life courses of thousands of individuals. These biographical surveys and the associated data are essential research tools used primarily in sociology and demography; they enable us to grasp the interplay of thematic life courses, to reveal social mechanisms and inequalities, to study society, and to better understand the factors that determine the life choices of the individuals who make up that society.
From a conceptual point of view [2], each thematic life course can be viewed as a set of chronological sequences (one sequence per attribute or indicator). Within each of these chronological sequences, events and episodes alternate. Each episode describes a stage within a given theme (for example, a residential episode brings together information on a place where the respondent lived), whilst an event most often marks a change of episode (for example, an event such as a house move).
In this context, it is now legitimate to ask whether AI agents could be used to conduct surveys and thereby collect life course data.
Indeed, recent research [3, 4] shows that chatbots can conduct qualitative interviews online, ask questions following a pre-established protocol or questionnaire, and manage follow-up questions. However, other studies [5, 6] also highlight the limitations and inherent biases of these approaches.
As part of this internship, whilst being aware of both the potential of this approach and its limitations or possible biases, we wish to explore the current capabilities of AI agents to conduct surveys autonomously and in a controlled manner, with a view to collecting biographical data within the framework set by a pre-established questionnaire. We will consider two study scenarios here.
In the first scenario, we assume that a written transcript exists of the interview already conducted between the interviewer and the interviewee. The task for the AI agent(s) is therefore to extract life course data from this text in order to instantiate the ontological model corresponding to this biographical survey. The difficulty here lies in aligning, on the one hand, the entities recognised in the text and, on the other hand, the concepts and relationships present in the ontological life course model [7, 8]. An additional constraint to be taken into account is that of identifying the temporal milestones which correspond to events and thus to thematic episodes in a person's life…
In the second scenario, the aim of the AI agent(s) is to collect the respondent's biographical data through a dialogue with them, based on a pre-established biographical survey questionnaire. Here, the objective is to investigate the extent to which one or more AI agents can replace a human interviewer,

regardless of the specific biographical survey in question. In addition to the challenge of aligning and instantiating the ontological model – which was already present in the first scenario – the difficulty here lies in establishing an interactive dialogue between the AI agent(s) and the respondent, enabling the collection, in written form, of a set of data constituting the most comprehensive life histories possible.

Aims of the Internship

- to carry out a state-of-the-art review identifying research that utilises AI agents in the collection of survey data;
- to design an initial architecture based on AI agents capable of extracting data relating to a life course from a textual transcript of a bibliographic survey interview, and of populating the corresponding ontological model;
- design a second architecture based on AI agents capable of collecting biographical data through a questionnaire-guided dialogue, and then populate the corresponding ontological model;
- to implement these architectures based on AI agents;
- to test and validate the operational viability of these architectures;
- where appropriate, to compare the principles and performance of these architectures with similar approaches.

References

[1] F. Guérin-Pace et al. (dir.) : En quête d'appartenances : L'enquête Histoire de vie sur la construction des identités. Ined Éditions, coll. « Grandes Enquêtes », 2009, 222 p.
[2] J. Gensel et al.: Un modèle multi points de vue pour représenter les trajectoires de vie, 5ème colloque international du Collège International des Sciences du Territoire, Paris, 18-20 novembre 2020.
[3] F. Chopra and I. Haaland, Ingar: Conducting qualitative interviews with AI,
CEBI Working Paper Series, No. 06/23, University of Copenhagen, Department of Economics, Center for Economic Behavior and Inequality (CEBI), Copenhagen, (2023).
[4] A. Wuttke et al.: AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers. LaTeCH-CLfL 2025: 179-204, (2025).
[5] S. Barari, et al.:AI-Assisted Conversational Interviewing: Effects on Data Quality and User Experience. CoRR abs/2504.13908, (2025).
[6] R. T. Williams and E. Ingleby:The online survey in qualitative research: can AI act as a probing tool? Frontiers Res. Metrics Anal. 10, (2025).
[7] A. Sahbi et al.: Automatic Ontology Population from Textual Advertisements: LLM vs. Semantic Approach. KES 2024: 3083-3092, (2024).
[8] S. Saki Norouzi, et al.: Ontology Population using LLMs. CoRR abs/2411.01612 (2024)

Required Skills

- Master's student in Computer Science, Artificial Intelligence or Information Systems;
- An interest in AI agent-based architectures and social and human sciences (SHS) approaches;
- Strong Python programming skills (machine learning, deep learning, neural networks, etc.);
- Ability to work independently, attention to detail, curiosity, analytical skills…

Supervision and working environment

The placement will take place at the Grenoble Computer Science Laboratory (LIG) (https://www.liglab.fr/fr), within the stories research team (https://lig-stories.imag.fr/), a group dedicated to the study of semantic trajectories.
The stories team's research areas include symbolic AI, neural AI, knowledge engineering and spatio-temporal information processing.

Prospects for pursuing a PhD

Pursuing a PhD on this topic is envisaged, subject to securing funding (a PhD scholarship), which has yet to be obtained.
