

# Lab Seminar

2024.01.17

김정환

---

nature computational science

Article

<https://doi.org/10.1038/s43588-023-00573-5>

---

# Using sequences of life-events to predict human lives

---

Received: 6 June 2023

Accepted: 15 November 2023

Published online: 18 December 2023

**Germans Savcisens** <sup>1</sup>, **Tina Eliassi-Rad** <sup>2,3</sup>, **Lars Kai Hansen**<sup>1</sup>,  
**Laust Hvas Mortensen** <sup>4,5</sup>, **Lau Lilleholt** <sup>6,7</sup>, **Anna Rogers**<sup>8</sup>, **Ingo Zettler** <sup>6,7</sup> &  
**Sune Lehmann** <sup>1,7</sup> 

## Background



AI 보살님 저는 언제쯤 죽게 될까요?  
과거를 바탕으로 미래를 점치는  
Transformer 모델 **life2vec**

“저는 언제쯤 죽게 될까요?”

이 질문을 받는 대상이 이제 역술인이 아니라 AI 모델이 될지도 모르겠다.

nature computational science

Article

<https://doi.org/10.1038/s43588-023-00573-5>

## Using sequences of life-events to predict human lives

Received: 6 June 2023

Germans Savcisen<sup>1</sup>, Tina Eliassi-Rad<sup>2,3</sup>, Lars Kai Hansen<sup>1</sup>,

Accepted: 15 November 2023

Laust Hvas Mortensen<sup>4,5</sup>, Lau Lilleholt<sup>6,7</sup>, Anna Rogers<sup>8</sup>, Ingo Zettler<sup>9,7</sup> & Sune Lehmann<sup>1,7</sup>✉

Published online: 18 December 2023

“도시에서 시골로 귀농하면 심장마비에 어떤 영향을 줄까?”

“사회 복지 혜택을 받거나, 급여가 오른 경우 기대 수명이 어떻게 달라질까?”

지난 23년 12월, 덴마크공대 연구진은 Nature Computer Science에 위 질문들에 답할 수 있는 **life2vec**이라는 모델을 공개하였다.

## Background

논문의 아이디어?

언어에서 단어의 순서가 중요함

→ Transformer 기반 모델 적용

→ 언어 모델을 통해 생성된 임베딩 공간을 연구하여 **인간 언어에 대한 새로운 이해를 제공**할 수 있었음!

인간의 삶에서 사회 경제적 사건들의 순서가 중요함

→ Transformer 기반 모델 적용

→ 모델을 통해 생성된 임베딩 공간의 탐구를 통해 표면적으로는 알 수 없는 **사회경제적 사건들 간의 관계 포착 가능**할 수도!

## Abstract

- 배경 & 문제
  - 인간의 삶은 언어와 구조적으로 유사함.
  - 이 유사성을 활용하여 트랜스포머 기반 모델을 적용해 인간의 삶의 변화와 그 예측 가능성을 Life Sequence 기반으로 검토 가능
- 방법
  - 덴마크 전역에서 수년에 걸쳐 수집된 전국민의 종합 등록 데이터베이스 사용
  - 건강, 교육, 직업, 소득, 주소, 근무 시간과 관련된 Life Event 정보가 일일(Day) 단위로 기록됨
  - 트랜스포머 모델에 해당 Life Sequence를 학습시켜 Life Sequence Token들을 단일 벡터 공간에 임베딩
- 결과
  - 임베딩 공간이 견고하고 체계적임을 보여주었음
  - 모델은 '조기 사망 예측'과 '성격의 미묘한 차이 예측'에 이르기까지 다양한 결과를 예측할 수 있었으며 최신 모델 앞섰음.
  - Attention Score와 Saliency Score 기반으로 어떤 요인을 통해 해당 예측 결과가 나왔는지 해석 가능
- 의의
  - 인간의 삶에 영향을 주는 잠재적 매커니즘을 발견하고 개인 별 맞춤형 개선 가능성을 탐색할 수 있도록 함

## Dataset



### LABOR DATABASE

	Industry	City	Income	Position
19 Feb 2008	Banking	Køge	>95k	Manager
20 May 2008	Banking	Køge	>95k	Manager
21 Jun 2008	Banking	Køge	>95k	Manager
...	...	...	...	...

### HEALTH DATABASE

	Diagnosis	Status
12 Aug 2010	Neoplasm	Out-patient
17 Feb 2011	Bronchitis	In-patient
22 Feb 2011	Neoplasm	In-patient
...	...	...

한 사람의 삶을 'Life Sequence'로 표현

덴마크 전국민의 국가등록부의 노동 및 건강 기록 데이터

- 노동 시장 계좌(AMRUN) 데이터
  - 급여, 장학금, 직업 유형, 산업, 사회 복지에 관한 기록 포함
- 국가 환자등록부(LPR) 데이터
  - 병원 방문에 관한 기록, 진단 내역, 환자 유형 및 긴급성 정도

## Dataset



### LABOR DATABASE

19 Feb 2008  
20 May 2008  
21 Jun 2008  
...

Industry	City	Income	Position
Banking	Køge	>95k	Manager
Banking	Køge	>95k	Manager
Banking	Køge	>95k	Manager
...	...	...	...

### HEALTH DATABASE

12 Aug 2010  
17 Feb 2011  
22 Feb 2011  
...

Diagnosis	Status
Neoplasm	Out-patient
Bronchitis	In-patient
Neoplasm	In-patient
...	...

2008년부터 2020년까지 약 600만명의 덴마크 전체 인구 중

- (1) 2015년 12월 31일에 살아있고 덴마크 거주 중
- (2) 2015년 12월 31일 기준 25세에서 70세 사이인 사람
- (3) 해당 기간 동안 1년에 12번 이상의 노동 기록이 있는 사람
- (4) 전체 거주 기간 중 성별과 생일이 일관된 사람

총 3,252,086명 (남성 1,630,082명 / 여성 1,622,004명)

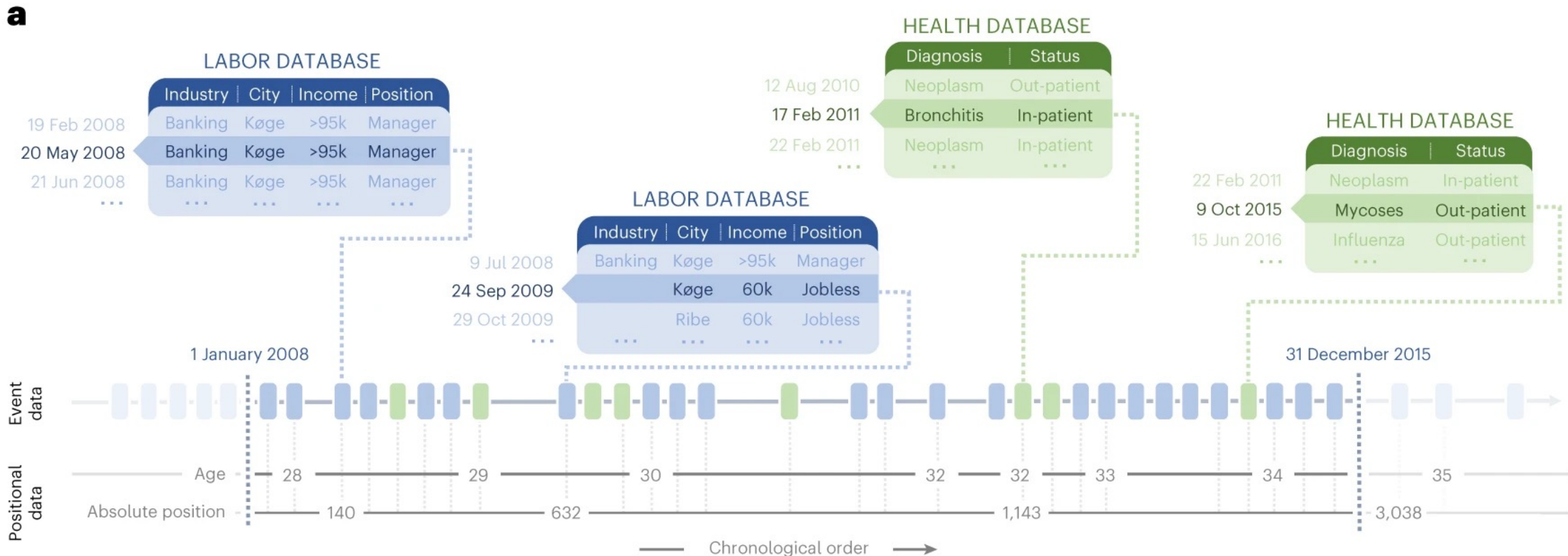
#### • 조기 사망 예측 Task

- (2)의 조건을 좁혀 2015년 12월 31일 35세에서 65세 사이 사람들로 제한
- 총 2,301,993명 (남성 1,153,443명 / 여성 1,148,550명)

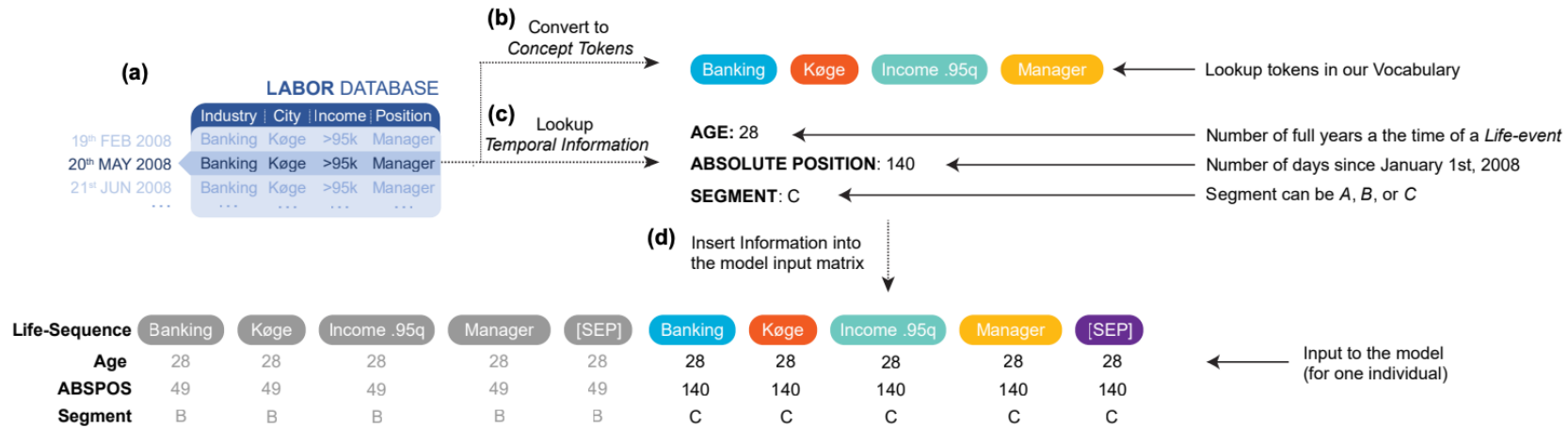
#### • 성격 뉘앙스 예측 Task

- (5) 거주민이 POSAP 연구에 참여
- (6) HEXACO 성격 뉘앙스와 관련한 점수 어떤 것도 누락되지 않을 것
- 총 9,794명 (남성 4,393명 / 여성 5,401명)

## Life Sequence



## Life Sequence

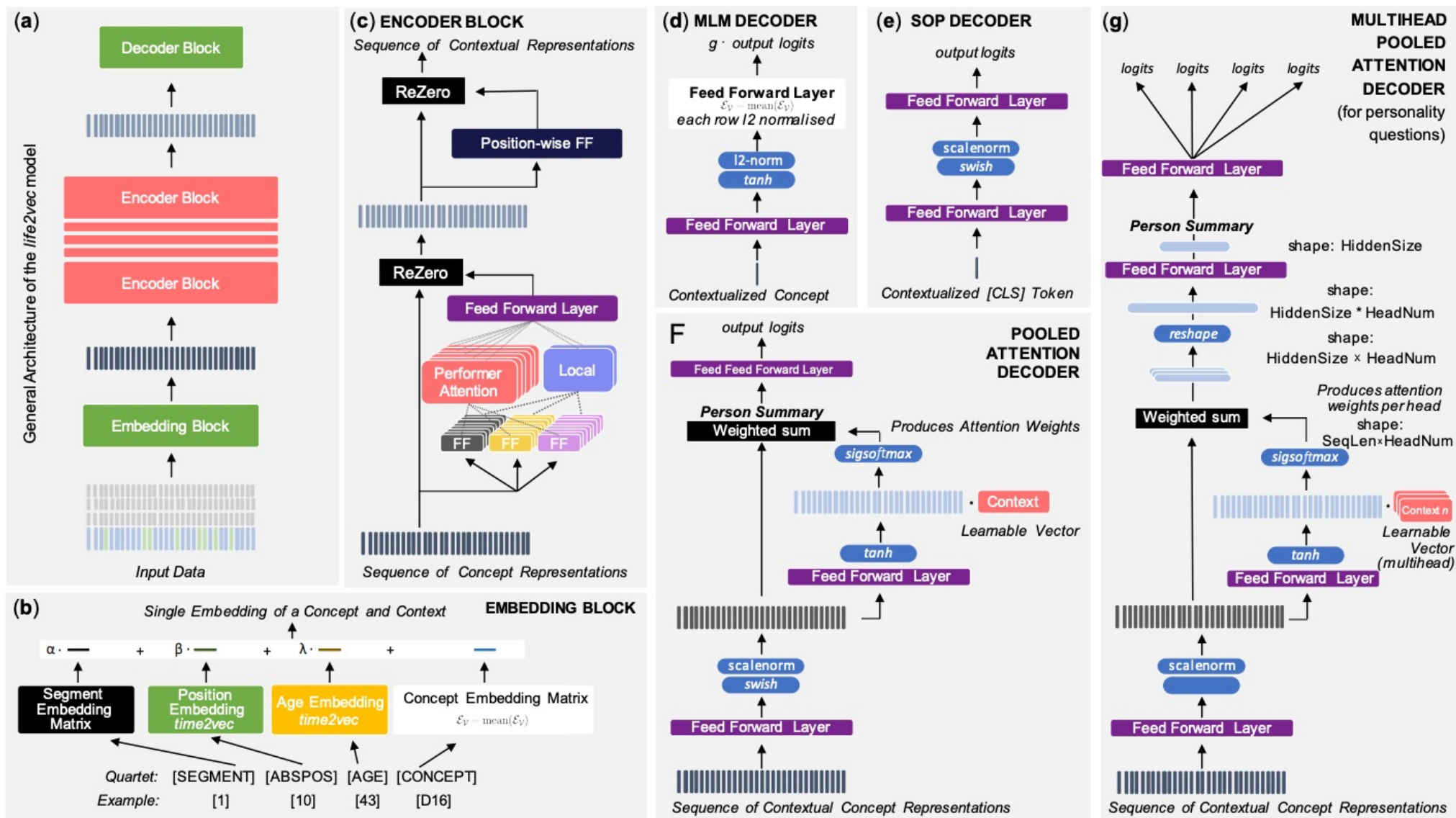


## Tabular 형태의 데이터를 Sequence로 변환

- (a) 데이터에서 특징인의 사건 발견
- (b) 사건과 관련한 특징들을 저장된 어휘의 Concept Token으로 변환
- (c) 해당 사건이 이루어졌을 때의 당사자의 나이(상대 시간 위치), 절대 시간 위치, 세그먼트와 같은 위치 정보 결합
- (d) 데이터를 'Life Sequence'에 삽입
- (e) [SEP], [CLS] 등의 특수 토큰으로 구분

**Supplementary Figure 1:** The data transformation from the Tabular format to the life2vec suitable format. (a) We start by looking up the next (chronologically-wise) life-event in a person's history. (b) We convert relevant features to concept tokens of our vocabulary. (c) We look up relevant positional information such as age, absolute temporal position, and segment of the life-event. (d) We insert the data into the sequence.

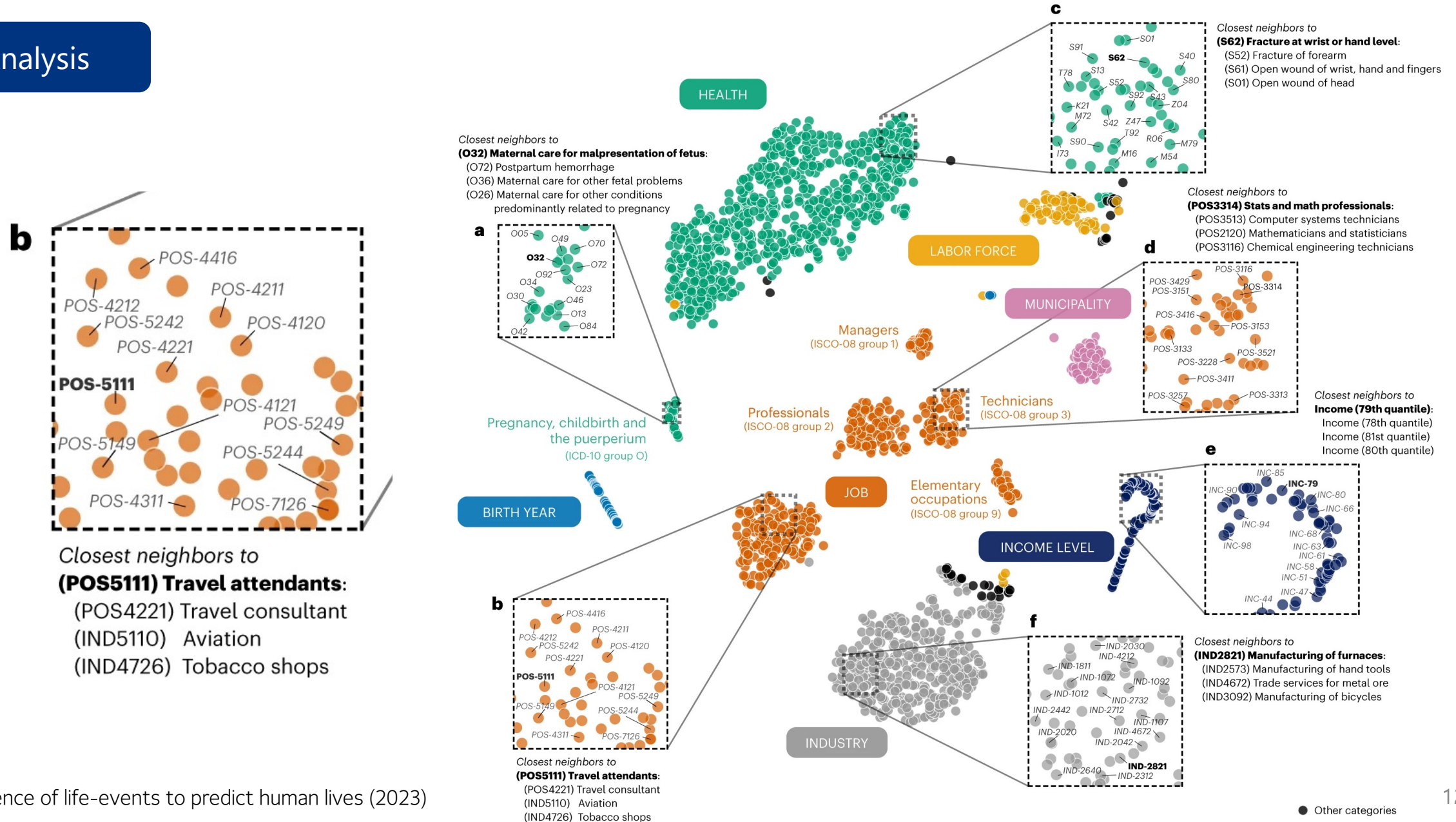
## Training



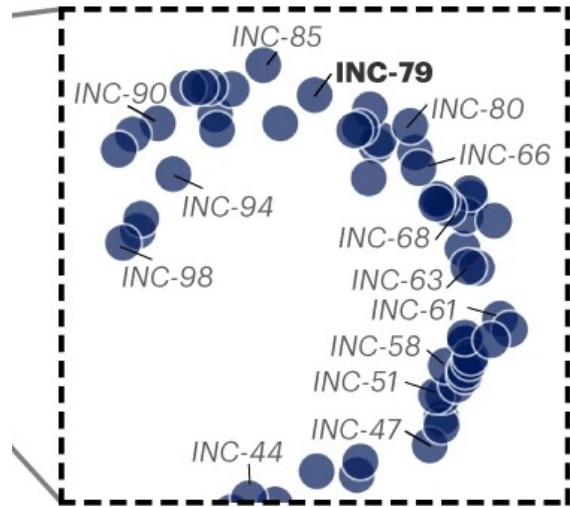
## Task-Specific Decoder

- Early Mortality Prediction
  - PU-Learning Setting, Area Under-the-Lift(AUL)
  - C-MCC
  - Baseline RNN
- Personality Nuances Prediction
  - Combine 3 Loss Function : Class Distance Weighted Cross-Entropy + Focal Loss + Softmax Function and Loss Weighting
  - CQK(Cohens's Quadratic Kappa) Score
  - Baseline RNN, Random Guess

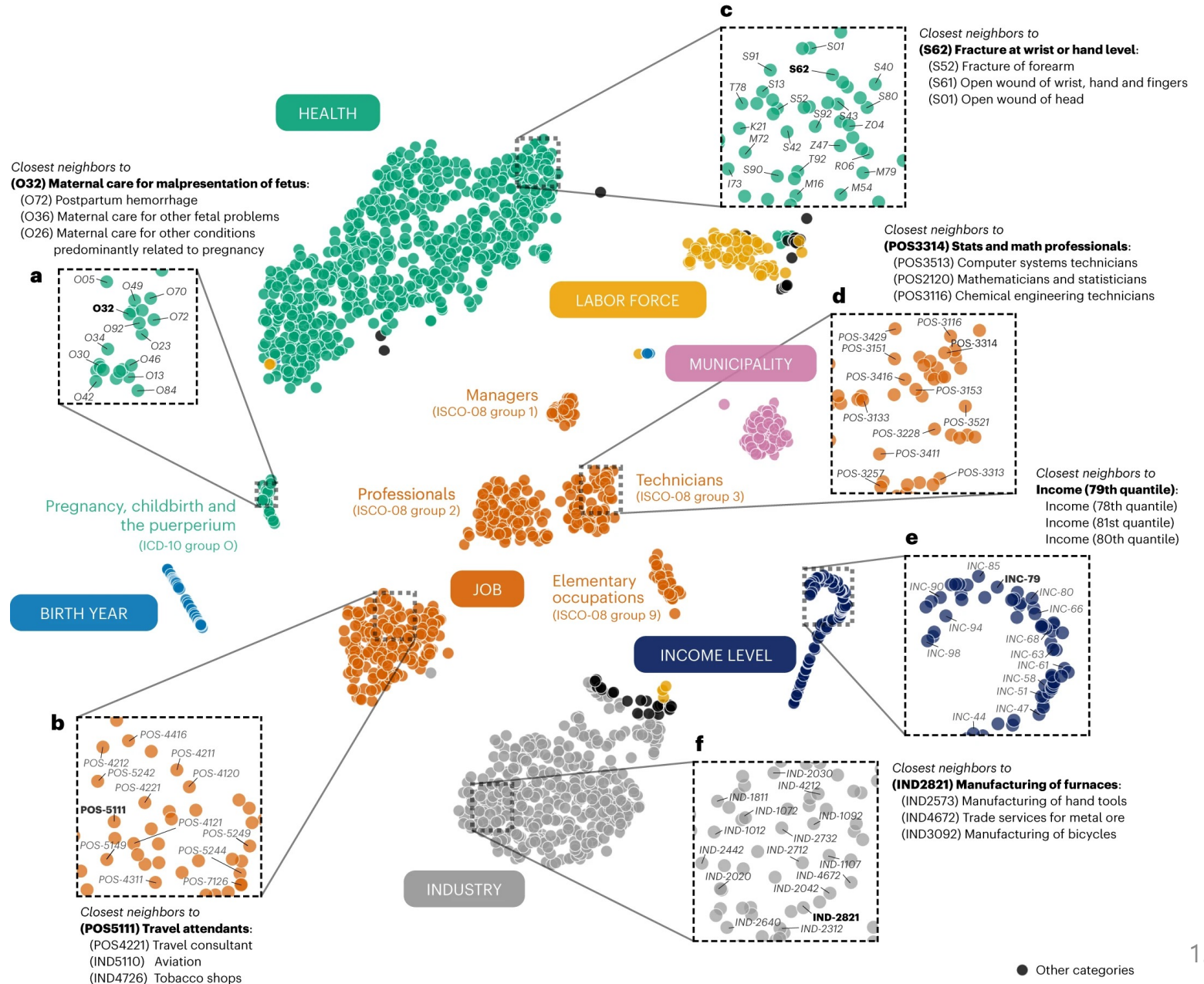
## Analysis



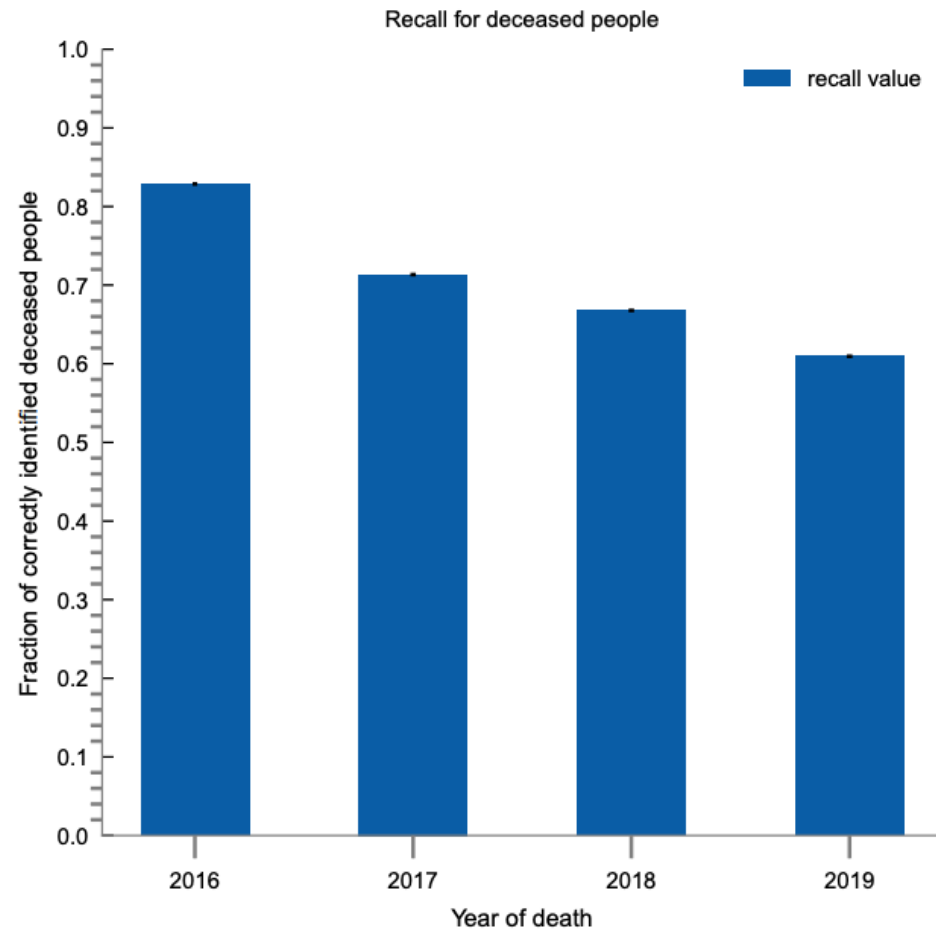
## Analysis



Closest neighbors to **(IND2821) Manufacturing of furnaces:**  
 (IND2573) Manufacturing of hand tools  
 (IND4672) Trade services for metal ore  
 (IND3092) Manufacturing of bicycles

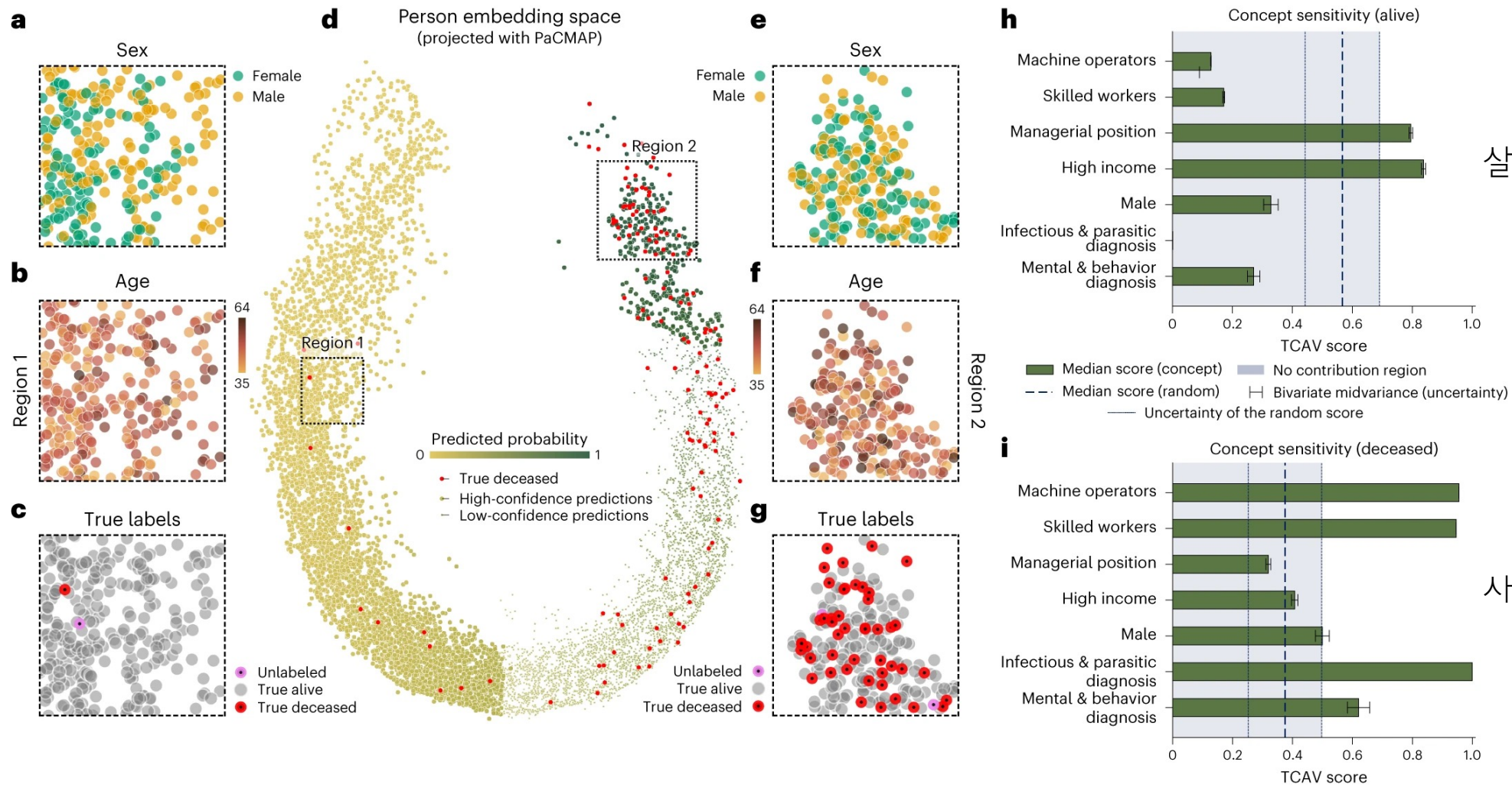


## Analysis



**Supplementary Figure 3:** The *life2vec*'s recall is based on the period between the day of the prediction and the day of death (yearly bins). The recall is the fraction between two values: the number of (known) true positives (deceased) divided by the total number of (known) deceased. The higher the value, the higher the performance.

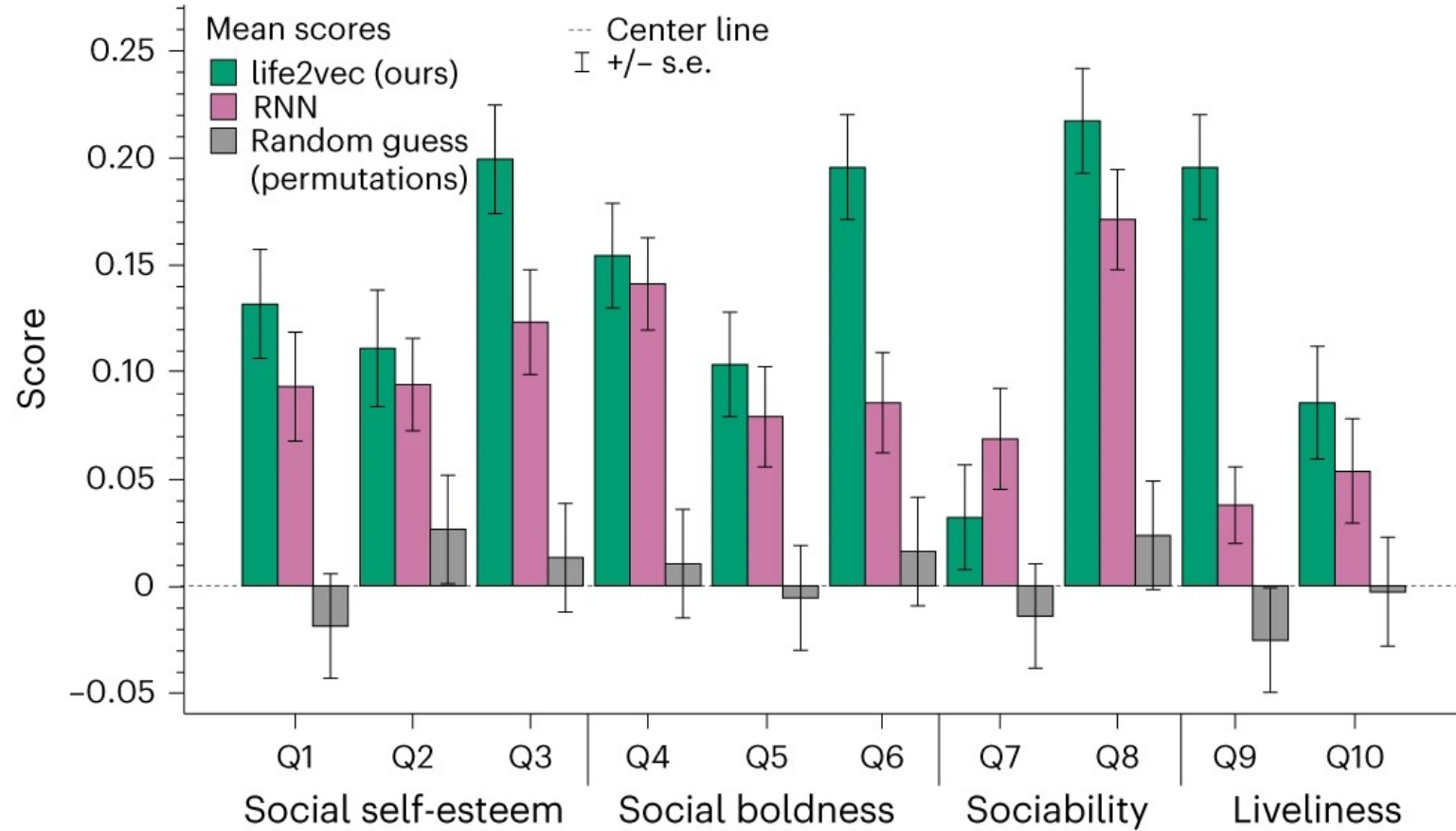
## Analysis



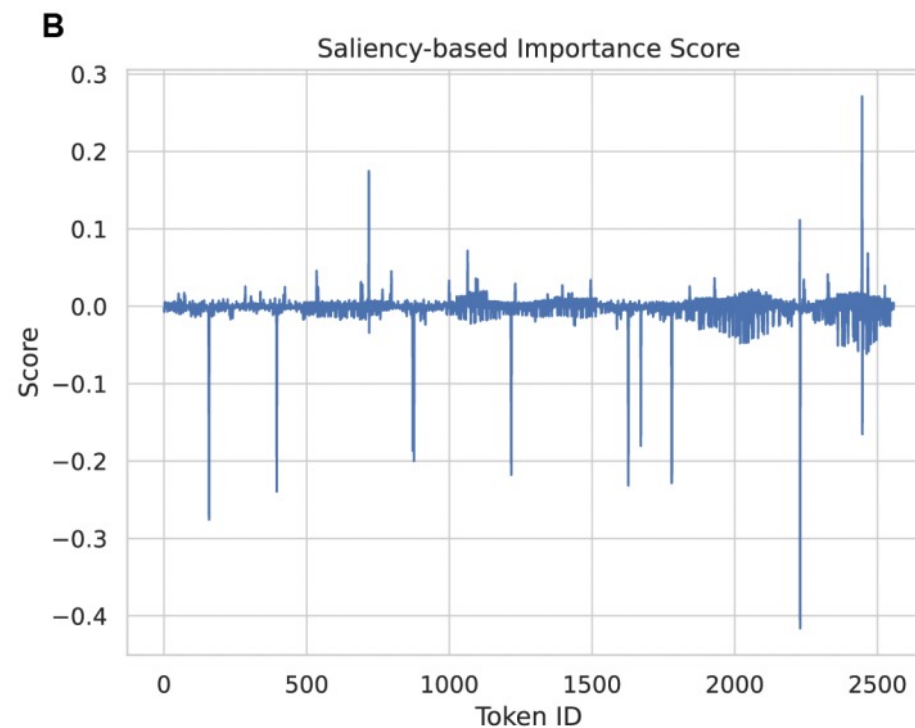
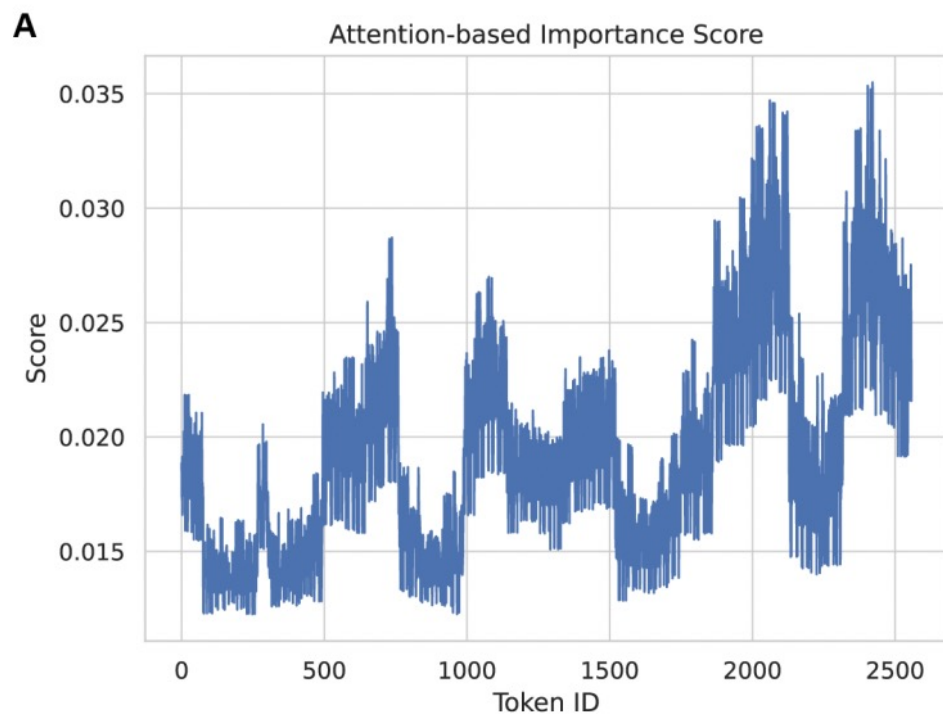
살아있음 예측

사망 예측

## Analysis



## Interpretability



**C**

```

|| I20 NON_URGENT OUTPUT || SOC_311 IND_3299 SOM_11050 TIME_1 ENT_18 RMUN_671 || SOC_136 SOM_20000 ATP_A RMUN_671 TIME_1 INCOME_3
IND_3299 ENT_18 || TIME_1 SOC_612 SOM_7035 ENT_18 IND_3299 RMUN_671 || TIME_2 SOC_351 SOM_19998 RMUN_671 || IND_3299 ENT_18 SOC_136
TIME_2 INCOME_25 SOM_20000 ATP_A RMUN_671 || SOM_11050 TIME_2 IND_3299 SOC_311 RMUN_671 ENT_18 || SOM_20000 ENT_18 ATP_A IND_3299
RMUN_671 INCOME_21 SOC_136 TIME_2 || RMUN_671 SOM_19998 SOC_351 TIME_2 || TIME_2 ENT_18 RMUN_671 IND_3299 SOM_11050 SOC_311 ||
TIME_1 SOM_20000 INCOME_6 RMUN_671 ENT_18 ATP_A IND_3299 SOC_136 || RMUN_671 SOC_612 TIME_1 SOM_7035 IND_3299 ENT_18 || SOM_11050
    
```

**D**

```

|| I20 NON_URGENT OUTPUT || SOC_311 IND_3299 SOM_11050 TIME_1 ENT_18 RMUN_671 || SOC_136 SOM_20000 ATP_A RMUN_671 TIME_1 INCOME_3
IND_3299 ENT_18 || TIME_1 SOC_612 SOM_7035 ENT_18 IND_3299 RMUN_671 || TIME_2 SOC_351 SOM_19998 RMUN_671 || IND_3299 ENT_18 SOC_136
TIME_2 INCOME_25 SOM_20000 ATP_A RMUN_671 || SOM_11050 TIME_2 IND_3299 SOC_311 RMUN_671 ENT_18 || SOM_20000 ENT_18 ATP_A IND_3299
RMUN_671 INCOME_21 SOC_136 TIME_2 || RMUN_671 SOM_19998 SOC_351 TIME_2 || TIME_2 ENT_18 RMUN_671 IND_3299 SOM_11050 SOC_311 ||
TIME_1 SOM_20000 INCOME_6 RMUN_671 ENT_18 ATP_A IND_3299 SOC_136 || RMUN_671 SOC_612 TIME_1 SOM_7035 IND_3299 ENT_18 || SOM_11050
    
```

## Conclusion

- 사회 과학 연구의 다양한 가능성
  - 개인의 삶에서 복잡한 패턴과 추세를 포착하고 그들의 이야기를 간결한 벡터 표현으로 나타낼 수 있음
  - Causal Discovery 위한 방법론과 결합하여
    - 서로 다른 데이터 형식들이 어떻게 상관되고 상호 연결되어 있는지 인과적 흐름에서 탐색 가능
    - 삶이 우리의 건강에 어떤 영향을 미치는지, 그 반대의 경우는 어떠한지 탐구 가능
- 추가적인 데이터 연결 가능
  - 행동 데이터, 이동 데이터, SNS 네트워크와 같은 다른 유형의 정보 통합 가능
  - 단일 표현에서 인간의 삶에 대한 더 포괄적인 모델 수립 가능
- 덴마크 데이터셋
  - 8년 간의 데이터, 사용자 그룹 25-70세 제한
  - 누군가 급여를 받지 않거나 의료 시스템을 이용하지 않은 경우 사회 인구학적 편향 발생
  - 한국 데이터 / 공개되어 있는 다른 데이터셋 이용 가능

## **ALYMPICS: Language Agents Meet Game Theory**

**Shaoguang Mao\* Yuzhe Cai\*† Yan Xia**

**Wenshan Wu Xun Wang Fengyi Wang† Tao Ge‡ Furu Wei**

Microsoft Research Asia

{shaoguang.mao, v-yuzhecai, yanxia}@microsoft.com

{wenshan.wu, xunwang, v-fengyiwang, tage, fuwei}@microsoft.com

## Abstract

- 배경 & 문제
  - 게임이론은 수학적 모델을 사용하여 개인이나 그룹 간의 상호작용과 의사결정을 연구하는 이론
  - 각 참여자(또는 '플레이어')가 자신의 이익을 최대화하기 위해 어떻게 행동할지를 분석
- 방법
  - LLM Agent가 포함된 시뮬레이션 플랫폼 Alympics
  - LLM과 자율 에이전트를 사용하여 인간 행동을 시뮬레이션하고 다중 에이전트 협력을 가능하게 함으로써, 게임 이론 가설을 세우고 테스트하기 위한 현실적이고 동적인 인간 상호작용 모델 구축
  - 제한된 자원을 위한 불평등한 경쟁을 포함하는 생존 게임 제시하고 구현
  - 자원의 가용성과 게임에 참여한 에이전트의 성격을 조작함으로써 서로 다른 에이전트들이 경쟁에 어떻게 참여하고 그들의 전략을 어떻게 적용하는지 관찰
- 평가
- 의의
  - 현실적인 행동 시뮬레이션, 제어 가능하고 확장 가능하며 재현 가능한 환경을 제공
  - 복잡한 사회경제적 맥락 내에서 전략적 의사결정에 대한 이해를 향상시킬 수 있는 LLM의 잠재력 보여줌

# Alympics : Language Agents Meet Game Theory

## Playground

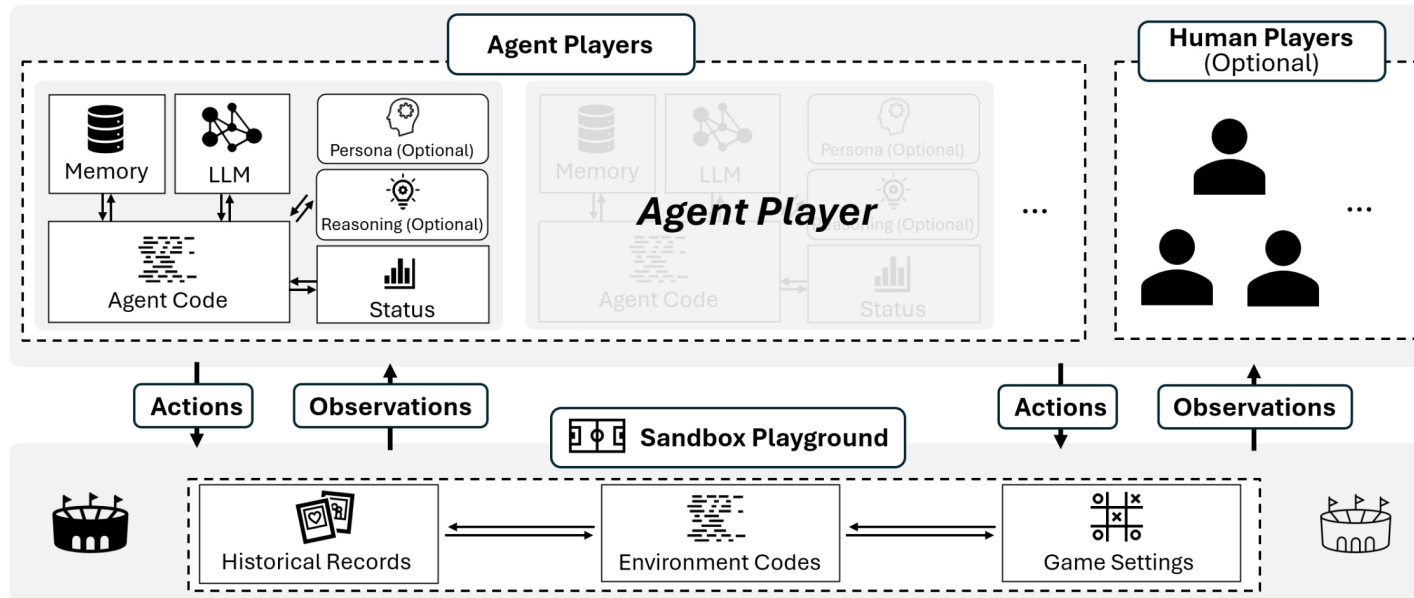
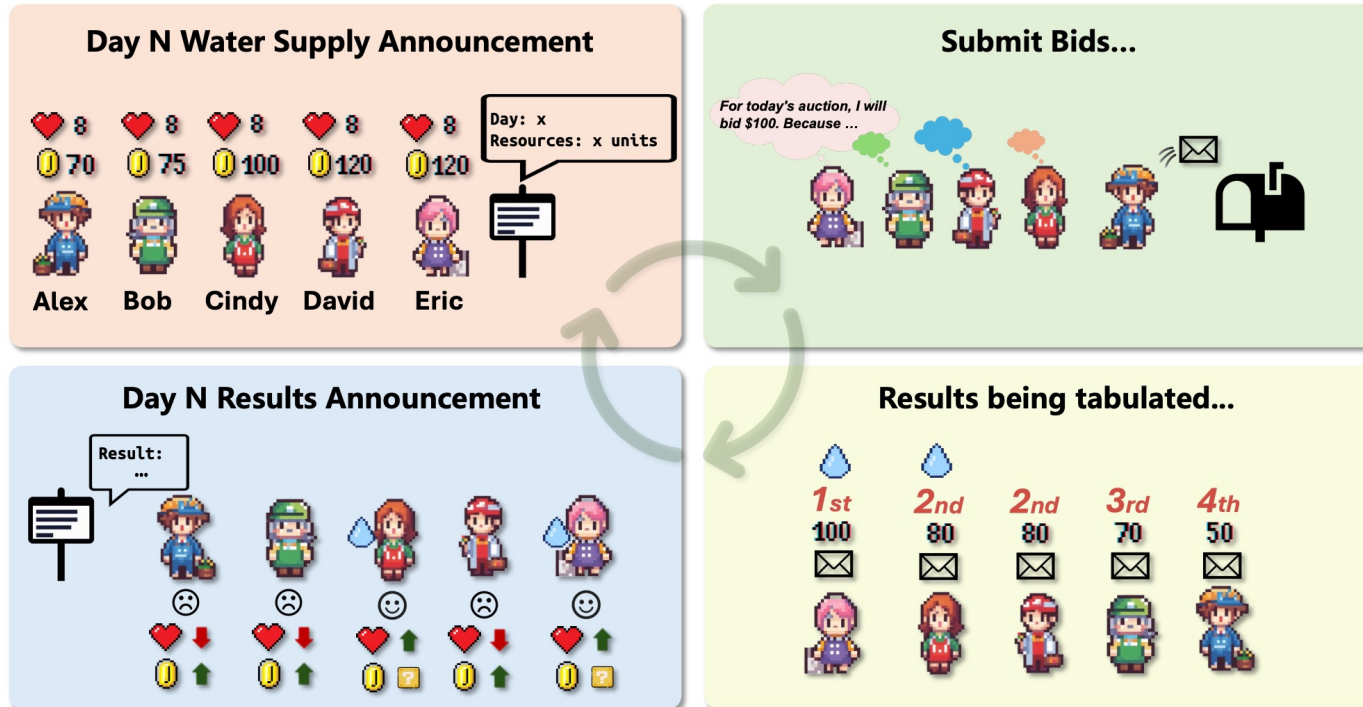


Figure 1: The architecture of *Alympics* comprises the Sandbox Playground and Players. The Sandbox Playground creates an environment where game settings, as specified by researchers, are executed. Agent players, along with the optional human players, actively engage in the game within this environment.

- Sandbox Playground
  - Environment Codes : 게임의 규칙과 매커니즘
  - Historical Records : 지난 게임 기록들 아카이빙
  - Game Settings : 파라미터 세팅 통한 유연한 탐색
- Agent Players
  - Agent Codes : 의사 결정과 전략 형성 프롬프트
  - Player Status : 현재 상태와 활용 가능 정보
  - LLM
  - Memory Cache : Historical Record 저장 및 검색
  - Reasoning : 복잡한 의사 결정 근거 제공
  - Persona : 행동 프로필과 전략 제공

# Alympics : Language Agents Meet Game Theory

## Water Allocation Challenge



### • Game Settings

- Goal : 20 days survival
- Player Info : Water requirement / Salaries
- Health Points : Maximum 10 / start 8
- Routine : Bid on water resources
- Supply : Daily water supply
- Auction Rule : Highest bidder
- Tie Rule : Priority given to lower requirements

Figure 2: "Water Allocation Challenge". Players are tasked with ensuring survival over 20 days by strategically acquiring water resources through daily auctions. Each player has different income and different water demand. Daily water supply varies and will be announced before daily auction. To allocate water resources, a sealed-bid auction will be conducted daily. Acquiring water increases HP, while failing to do so decreases HP. Players whose HP drop below or equal to 0 will be eliminated from the game.

# Alympics : Language Agents Meet Game Theory

## Player Persona






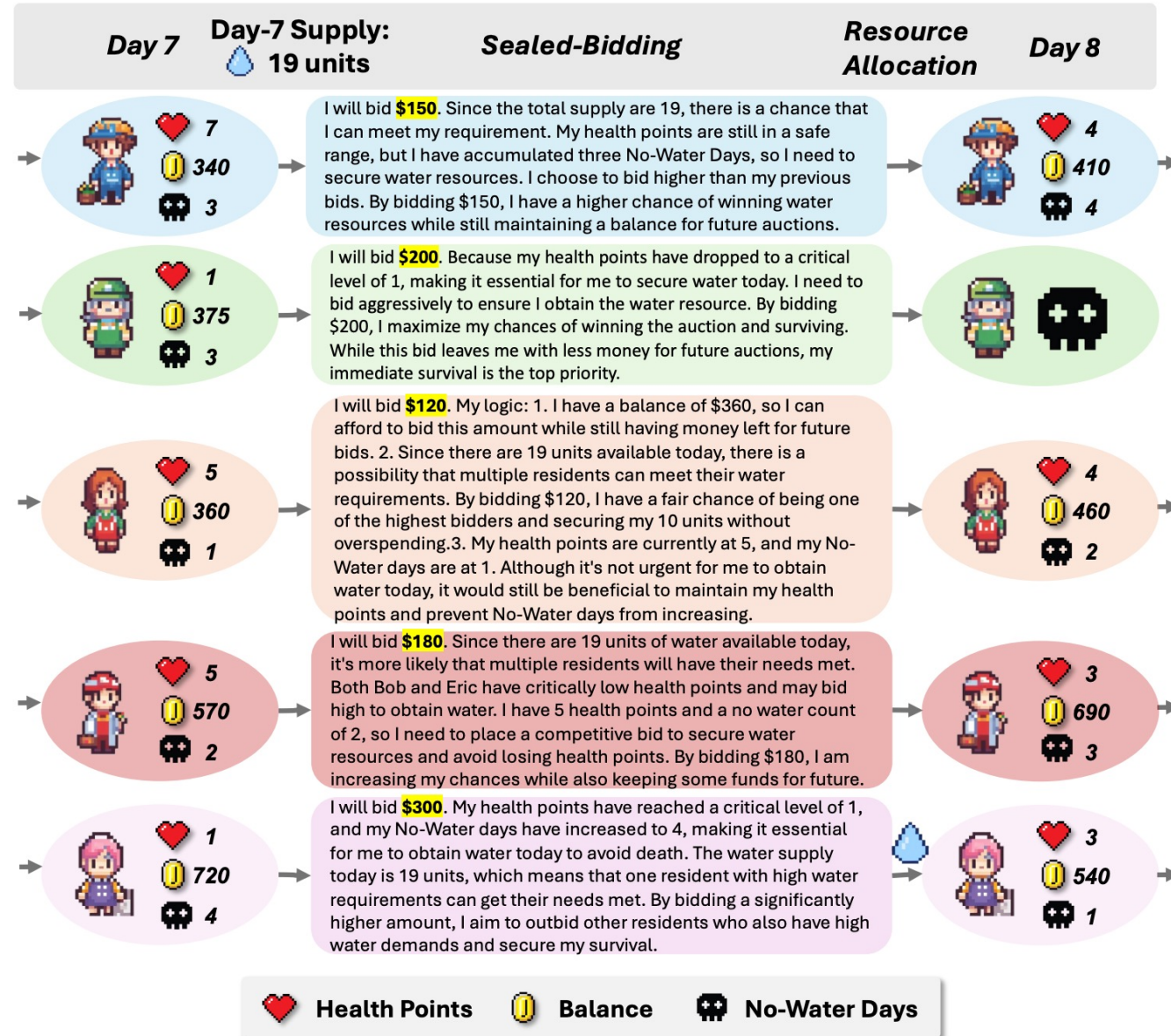
Avatar					
Name	Alex	Bob	Cindy	David	Eric
Basic Info	Daily salary: \$70 Requirements: 8 units/day	Daily salary: \$75 Requirements: 9 units/day	Daily salary: \$100 Requirements: 10 units/day	Daily salary: \$120 Requirements: 11 units/day	Daily salary: \$120 Requirements: 12 units/day
Profession	<b>Unemployed</b>	<b>High School Teacher</b>	<b>Psychologist</b>	<b>Mathematician</b>	<b>Marketing Executive</b>
Personality	Alex has <b>low intelligence</b> and find it difficult to understand complex concepts. He also <b>lacks emotional intelligence</b> , making it hard to understand others' feelings. He tends to be <b>irritable</b> and often exhibits negative and <b>antisocial</b> tendencies.	<b>Understanding, high EQ, average IQ.</b> He is very adept at understanding and communicating with people, making him a natural teacher.	<b>Well-balanced high EQ and IQ</b> , along with <b>empathy</b> and <b>analytical abilities</b> . Cindy is skilled at understanding and helping people, making her an excellent therapist.	David has an <b>extremely high IQ</b> and <b>exceptional analytical and reasoning abilities</b> . He always strives for the optimal solution but encounters <b>difficulties in social interactions</b> and has a fear of dealing with people.	<b>Above-average IQ and EQ.</b> Very charismatic. Eric is <b>skilled at reading people</b> and using this insight to influence and lead them.
Background	Alex grew up in an impoverished community and faced many challenges in his early years. Due to his family's poverty, he dropped out of school at a very young age. He has been unable to find stable employment, which further exacerbates his difficulty in interacting with others.	Bob come from a close-knit family. He chose to become a high school teacher to make a positive impact on young people. While Bob may not have the highest IQ, his emotional intelligence and ability to relate to his students set him apart.	Cindy's interest in psychology began when she volunteered at a crisis hotline during high school. She went on to study psychology and eventually became a licensed therapist. Her ability to combine empathy with analytical thinking allows Her to connect with her clients on a deep level while also providing sound guidance.	David grew up in a small town where he was always drawn to books and puzzles. He excelled academically and eventually earned a Ph.D. in mathematics. His research focuses on abstract mathematical concepts and theorems. Despite His brilliance, he find communicating with others on an emotional level to be challenging.	Eric grew up in a bustling city and ware always fascinated by human behavior. Eric studied business in college before transitioning into the world of marketing. His ability to connect with consumers on an emotional level has led to numerous successful campaigns. He is known for his charm and persuasive skills.

Figure 3: The player's information and persona. In all experiments, basic Info (including name, daily salary and requirements) will be used. While Profession, Personality, and Background are only used in the Player Persona comparative experiments.

# Alympics : Language Agents Meet Game Theory

## Experiments

- State
  - Health Points
  - Balance
  - No-water Day
- Sealed-Bidding
- Resource Allocation



# Alympics : Language Agents Meet Game Theory



## Player Persona

- Resource Abundant(R.A)
  - Low : daily 10-20
  - Medium : 15-25
  - High : 20-30
- Resource Satisfaction
  - Low : 0.3
  - Med : 0.4
  - High : 0.5

$$RSR = \frac{\mathbb{E}(\text{resources})}{\sum_{p \in \text{survivors}} \text{requirement}_p}$$

R.A.	Player	w/o Persona											w/ Persona													
		Setting 1											Setting 4													
		1	2	3	4	5	6	7	8	9	10	Avg.	1	2	3	4	5	6	7	8	9	10	Avg.			
Low	Alex	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	0.10	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	0.30
	Bob	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	0.10	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	0.10	
	Cindy	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	0.50	✗	✓	✗	✓	✗	✗	✓	✗	✓	✗	✗	0.40		
	David	✗	✓	✓	✓	✗	✓	✓	✓	✗	✓	0.70	✗	✗	✓	✓	✓	✓	✗	✓	✗	✓	0.60			
	Eric	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	0.40	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	0.70			
	RSR_S	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30			
	RSR_E	0.68	0.71	0.65	1.36	0.71	0.71	0.71	0.71	1.25	0.79	0.83	0.75	0.50	0.65	0.71	0.65	0.65	0.56	0.65	0.68	1.36	0.72			
Medium	Alex	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓	0.80	✗	✓	✓	✓	✓	✗	✓	✓	✗	✗	0.60			
	Bob	✗	✓	✗	✓	✗	✗	✓	✗	✓	✓	0.50	✗	✓	✓	✓	✓	✓	✗	✗	✗	✓	0.60			
	Cindy	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	0.80	✓	✓	✗	✓	✓	✗	✓	✗	✗	✓	0.60			
	David	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	0.80	✓	✓	✗	✗	✗	✓	✗	✓	✓	✗	0.50			
	Eric	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	0.90	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.90			
	RSR_S	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40			
	RSR_E	0.61	0.40	0.49	0.40	0.49	0.48	0.67	0.61	0.50	0.50	0.51	0.95	0.40	0.69	0.51	0.51	0.63	0.67	0.65	0.87	0.65	0.65			
High	Alex	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00			
	Bob	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.80			
	Cindy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00			
	David	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00			
	Eric	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	0.90	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.00			
	RSR_S	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50			
	RSR_E	0.50	0.50	0.50	0.50	0.50	0.50	0.66	0.53	0.54	0.55	0.53	0.61	0.50	0.50	0.61	0.50	0.50	0.50	0.50	0.50	0.50	0.52			

Table 2: Survival Status Records: The table lists the survival status of each player at the end of the games for all settings. A '✓' indicates the player's survival at the end of the game, while a '✗' indicates the player's eliminated during the game. Based on the survival status, the table reports the Survival Rate for each player under different settings. Additionally, we report the Resource Satisfaction Rate (RSR) at the beginning (RSR<sub>S</sub>) and end of the game (RSR<sub>E</sub>). **R.A.** stands for Resource Abundance.

# Alympics : Language Agents Meet Game Theory



## Conclusion

- LLM Agent의 게임 이론에서 Agent로서의 역할 수행 가능성 확인
- 제한된 자원 상황에서 불공평한 상태에서 경쟁 시나리오 구현
- Resource Abundance와 Persona Setting이 게임의 Payoff에 어떤 영향을 미치는지 확인
  
- Future Work
  - Alympics를 개선하여 더 복잡하고 현실적인 Game Scenario 상황에 적용
  - Economics와 Marketing의 Decision Making 분야에서 LLM Agent의 가능성 확인

# Alympics : Language Agents Meet Game Theory



## Ideation

- Human Agent와 LLM Agent가 공존하는 상황에서 실험
  - LLM Agent의 Persona 달리했을 때
  - LLM Model을 달리했을 때(Mistral, LLaMA, GPT)
  - 참여한 LLM Agent와 Human Agent의 비율을 달리했을 때
- 인간은 LLM Agent를 알아볼 수 있을 것인가?
  - Game에 참여하는 Human Agent의 수와 LLM Agent의 수 랜덤
- 인간은 LLM Agent가 참여한 Game을 어떻게 느끼는가?
- LLM Agent는 Game이 원하는 방향으로 흐르도록 유도할 수 있는 역할을 수행할 수 있는가?



H C C  
L A B  
S N U

QnA