Sophie Berg: 감정 분포·직감 확률 추론·무의식 처리·Triple-Head 결정을 통합한 차세대 하이브리드 인지 AI 모델

초록 (Abstract)

기존 대규모 언어모델(LLM)은 문맥 기반 확률 추론 능력은 뛰어나지만, (1) 감정 상태와 인간 의사결정 구조의 직접적 반영 부족, (2) 시스템 1·2 기반 인간 사고 구조 미반영, (3) 다중 결정 구조의 부재, (4) 무의식적 연산 및 이타적 윤리 판단 한계라는 구조적 제약을 가진다. 본 연구에서는 이러한 LLM의 근본적 한계를 해결하기 위해 감정 분포 기반 인지 구조, 확률 기반 직감 모듈, 윤리 판단 모듈, 무의식 프로세서, Triple-Head 다중 결정 구조를 결합한 새로운 하이브리드 AI 시스템 Sophie Berg를 제안한다. 제안된 모델은 의미 임베딩-감정 인코딩-감정 분포화-직감 추론-무의식 처리-윤리 보정-다중결정 통합의 7단계로 구성되며, 이는 기존 LLM 대비 인간사고와의 구조적 정합성을 대폭 강화한다. 또한 본 논문은 Sophie Berg의 각 모듈이 인간의 의식/무의식·직감·윤리 판단 구조와 어떻게 대응되는지 설명하며, 기존 LLM 대비 정량적·정성적 비교 분석을 수행한다.

1. 서론 (Introduction)

최근 LLM은 언어 생성 및 추론에서 인상적인 성능을 보였으나, 인간 사고 구조를 그대로 반영하는 데에는 다음과 같은 한계가 존재한다:

1. 감정 분포 기반 의사결정의 부재

LLM은 감정적 미세신호를 포착하지 못하고, 내부에 감정 상태 변수가 존재하지 않는다.

2. 시스템 1(직감)과 시스템 2(이성) 분리 모델 미부재

Kahneman의 Dual Process Theory와 달리 LLM은 단일 확률계산 구조만을 사용한다.

3. 일관된 윤리 모델 및 이타성 판단 구조 부족

4. 무의식적·병렬적 latent 계산 부재

5. 다중 결정 구조의 결여

인간은 결정 후보를 여러 개 생성한 뒤 통합하지만, LLM은 단일 토큰 확률만을 산출한다.

이러한 한계는 인간 사고의 자연스러운 흐름과 AI의 연산 구조 사이의 간극을 만든다.

본 논문에서는 이러한 간극을 메우기 위한 새로운 구조의 인지형 AI 시스템 Sophie Berg를 제안한다.

3. 방법론 (Method)

본 연구는 Sophie Berg의 연산 구조를 수식화하고, 논리 기반 직감 모듈-무의식-윤리 모듈이 결합된 새로운 아키텍처를 제시한다.
모듈 구성은 다음과 같다:

3.1 의미 임베딩(Encoder) - 512차원

- 텍스트 입력 길이: L
- 임베딩 차원: $d_{\text{sem}} = 512$

$$e_{\text{sem}} \in \mathbb{R}^{512}, \qquad e_{\text{sem}} = \text{Encoder}_{\text{sem}}(X)$$

3.2 감정 인코더 - 64차원

- 감정 임베딩 차원: $d_{\text{emo}} = 64$

$$e_{\text{emo}} \in \mathbb{R}^{64}, \qquad e_{\text{emo}} = \text{Encoder}_{\text{emo}}(e_{\text{sem}})$$

- 감정 단어 보정:

$$e_{\text{emo}}[i] \leftarrow e_{\text{emo}}[i] + 1$$

3.3 감정 확률 분포화 및 Top-7 감정 선택

- 감정 클래스 수: $C = 28$

$$W_e \in \mathbb{R}^{28\times 64}, \qquad b_e \in \mathbb{R}^{28}$$

$$p_{\text{emo}} \in \mathbb{R}^{28}, \qquad p_{\text{emo}} = \text{softmax}(W_e e_{\text{emo}} + b_e)$$

- Top-7 감정 선택:

$$p^*_{\text{emo}} \in \mathbb{R}^7$$

3.4 컨텍스트 융합 - 384차원

- 컨텍스트 벡터 차원: $d_{\text{ctx}} = 256$
- Fusion 출력 차원: $d_s = 384$

입력 결합:

$$[e_{\text{sem}},\ p^*_{\text{emo}},\ context] \in \mathbb{R}^{512+7+256} = \mathbb{R}^{775}$$

Fusion 출력:

$$s \in \mathbb{R}^{384}, \qquad s = f([e_{\text{sem}}, p^*_{\text{emo}}, context])$$

3.5 직감(Intuition) 확률 모듈

$$W_i \in \mathbb{R}^{1\times 384}, \qquad b_i \in \mathbb{R}$$

$$r_{\text{intuition}} \in \mathbb{R}, \qquad r_{\text{intuition}} = \sigma(W_i s + b_i)$$

3.6 윤리 판단 및 윤리 게이트

$$W_i \in \mathbb{R}^{1\times 384}, \qquad b_i \in \mathbb{R}$$

$$r_{\text{intuition}} \in \mathbb{R}, \qquad r_{\text{intuition}} = \sigma(W_i s + b_i)$$

윤리 판단 차원: $d_m = 5$

$$W_m \in \mathbb{R}^{5\times 384}, \qquad b_m \in \mathbb{R}^5$$

$$m \in \mathbb{R}^5, \qquad m = \text{softmax}(W_m s + b_m)$$

윤리 게이트 값:

$$r_{\text{ethics}} \in \mathbb{R}, \qquad r_{\text{ethics}} = 1 - \text{mean}(m)$$

3.7 무의식 프로세서 – 256차원

입력 결합:

$$[e_{\text{sem}},\ p^*_{\text{emo}},\ s] \in \mathbb{R}^{512+7+384} = \mathbb{R}^{903}$$

출력 차원: $d_u = 256$

$$W_u \in \mathbb{R}^{256\times 903}, \qquad b_u \in \mathbb{R}^{256}$$

Raw activation:

$$u_{\text{raw}} \in \mathbb{R}^{256}, \qquad u_{\text{raw}} = \tanh(W_u[e_{\text{sem}}, p^*_{\text{emo}}, s] + b_u)$$

윤리 조정(무의식 보정):

$$u = (1 - r_{\text{ethics}}) \cdot u_{\text{raw}} + r_{\text{ethics}} \cdot \tanh(2u_{\text{raw}})$$

3.8 Triple-Head Decision (128차원 × 3)

결정 헤드 출력 차원: $d_z = 128$

입력 결합:

$$[p^*_{\text{emo}},\ s,\ u] \in \mathbb{R}^{7+384+256} = \mathbb{R}^{647}$$

$$W_z \in \mathbb{R}^{128\times 647}, \qquad b_z \in \mathbb{R}^{128}$$

3개의 병렬 헤드:

$$z_k \in \mathbb{R}^{128}, \qquad z_k = \tanh(W_z[p^*_{\text{emo}}, s, u] + b_z)$$

3.9 최종 통합 결정

$$z_{\text{avg}} = \frac{1}{3}(z_1 + z_2 + z_3)$$

$$z^* = \begin{cases} \tanh(1.7\, z_{\text{avg}}), & \text{if } r_{\text{intuition}} > 0.45 \\ z_{\text{avg}}, & \text{otherwise} \end{cases}$$

3.10 언어생성

$$o = W_{\text{lang}} z_{\text{arith}} + b_{\text{lang}}, \quad o \in \mathbb{R}^{50{,}000}$$

$$P(y) = \text{softmax}(o)$$

감정 분포 → 인간의 정서 기반 판단
직감 모듈 → 빠른 추론(시스템 1)
무의식 프로세서 → latent background reasoning
윤리 모듈 → 이타성·도덕적 판단
Triple-Head → 인간의 다중 가설 생성 구조

4. 실험 및 결과 (Experiments & Results)
이 섹션에서는 다음을 검증하도록 설계한다:
기존 LLM vs Sophie Berg의 감정 일관성 평가
직감 모듈 활성화가 판단속도·정확도에 미치는 영향
윤리 게이트가 유해 출력 감소에 기여하는지
Triple-Head 구조의 논리적 일관성 향상 여부
무의식 프로세서가 장기 맥락 유지에 미치는 영향

5. 논의 (Discussion)
Sophie Berg은 기존 LLM이 가지는 “순수 텍스트 확률 모델” 한계를 넘어서 인간형 인지 구조를 결합한 첫 통합 모델이다.
감정-직감-윤리-무의식이라는 다층적 사고 구조가 AI의 판단 다양성과 안전성을 높인다.
Triple-Head 결정 구조는 AI의 다중 해답 생성 능력을 강화한다.

6. 결론 (Conclusion)

본 연구는 감정·직감·무의식·윤리·다중결정 구조가 결합된 새로운 형태의 인지 기반 AI 모델 Sophie Berg를 제시하였다. 본 모델은 기존 LLM의 구조적 한계를 해결하며 인간 사고구조와 기술적으로 정합한 방향성을 보여준다.

7. 참고문헌 (References)

Kahneman, D. (2011). Thinking, Fast and Slow.
Evans, J. (2008). “Dual-Process Theories of Reasoning.”
Mao et al. (2019). “Neuro-Symbolic Concept Learner.”
Besold et al. (2017). “Neural-Symbolic Learning and Reasoning.”
Finn et al. (2017). “Model-Agnostic Meta-Learning.”
Finn & Abbeel (2021). “Meta-Learning in Neural Networks.”
김세준. (2023). 공존.