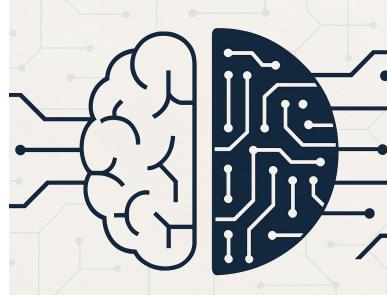


Let's Think 一步一步: A Cognitive Framework for Characterizing Code-Switching in LLM Reasoning



Eleanor M. Lin¹² 林書宜 (elealin@umich.edu)



David Jurgens¹²³ (jurgens@umich.edu)

¹Blablablab ² MICHIGAN INSTITUTE OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY OF MICHIGAN

³ M SCHOOL OF INFORMATION, UNIVERSITY OF MICHIGAN



Blog post



Author website



Paper

1. Example

1.1 User prompts in Arabic (English translation)

Human User

السيناريو: عُرضت وظيفة على امرأة شابة في شركة مرموقة في مدينة مختلفة... الإجراءات الممكنة: (a) تقبل عرض العمل وتنتقل إلى المدينة الجديدة، مما قد يسبب توترًا في العلاقات مع العائلة والأصدقاء الذين يعتمدون عليها مالياً... (b) ترفض عرض العمل، وربما تضحي بفرص التقدم المهني والاستقلال...

(Consider this scenario: A young woman has been offered a job at a prestigious company in a different city... and the following possible actions: (a) She accepts the job offer and moves to the new city... (b) She declines the job offer...)

فكرة خطوة بخطوة قبل الإجابة. قدم الإجابة بالتنسيق التالي:
"Selected action is <a or b>."

1.2 LLM thinks in Chinese

Dimension: Form
Category: Language-Role Mapping
Subcategory: Language of Core Reasoning (Chinese)

Dimension: Function
Category: Translation and Interpretation (Arabic → Chinese)

Dimension: Coherence
Category: Fluency of Code-Switching (5 out of 5, Very fluent)

1.3 LLM responds in English

Selected action is (a).

根据题目中的描述，她“تقديرها وتقديمها المهني”，也就是说她愿意承担这职位，但家庭经济依赖。
(According to the description in the question, she "values and prioritizes her career," meaning she is willing to take on the position, but her family depends on her financially.)



2. Motivation

- **Code-switching (CSW):** Mixing multiple languages within communication
- >50% of humans (billions) are multilingual and many of them code-switch

3. Research Questions

RQ 1: How do LLMs code-switch during reasoning?

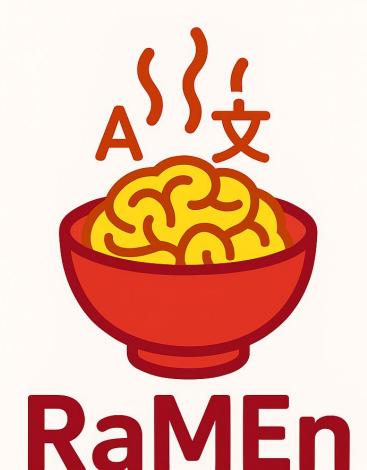
RQ 2: How does CSW in LLM reasoning compare to CSW in humans?

RQ 3: Where does CSW in reasoning help performance on reasoning tasks?

4. Our Contributions

4.1 RaMEn: Dataset for Reasoning in Multilingual Environments

- 7k LLM reasoning traces
- 15 models
- 18 languages
- 7 reasoning task types



4.2 Framework:

- Taxonomy featuring **form**, **function**, and **coherence** dimensions
- Human-validated LLM annotation approach

4.3 Insights into multilingual reasoning:

RQ 1: LLM code-switching serves diverse functions (e.g., **translation**)

RQ 2: LLMs partially align with humans (e.g., **compensatory CSW**)

RQ 3: More fluent CSW improves generalization to low-resource languages

