

An Exploration of Knowledge Editing for Arabic

Basel Mousi Nadir Durrani Fahim Dalvi
 {bmousi, ndurrani, faimaduddin}@hbku.edu.qa

Motivation

LLMs are **static** and quickly become **outdated**
 Knowledge Editing (KE) **updates facts** efficiently **without retraining**
 KE is well studied in English but **underexplored in other languages**, including *Arabic*
 An edit in one language should **generalize across languages**

Research Questions

How do existing methods perform at **knowledge editing in Arabic** compared to English?
 Can an edit in Arabic **propagate crosslingually**?
 How can models be **trained to edit themselves** effectively in **multiple languages**?

Example

Outdated fact: سيعقد EMNLP في ميامي هذا العام
Updated fact: سيعقد EMNLP في سوتشو هذا العام
Multilingual Consistency:
 أين سيعقد EMNLP هذا العام؟ ← سوتشو
 Where is EMNLP this year? → Suzhou
 Wo findet EMNLP dieses Jahr statt? → Suzhou

Experimental Setup

Datasets: Translated ZsRE Translated Counterfact

Knowledge Editing Methods:

ROME MEMIT In-Context Editing Learning To Edit

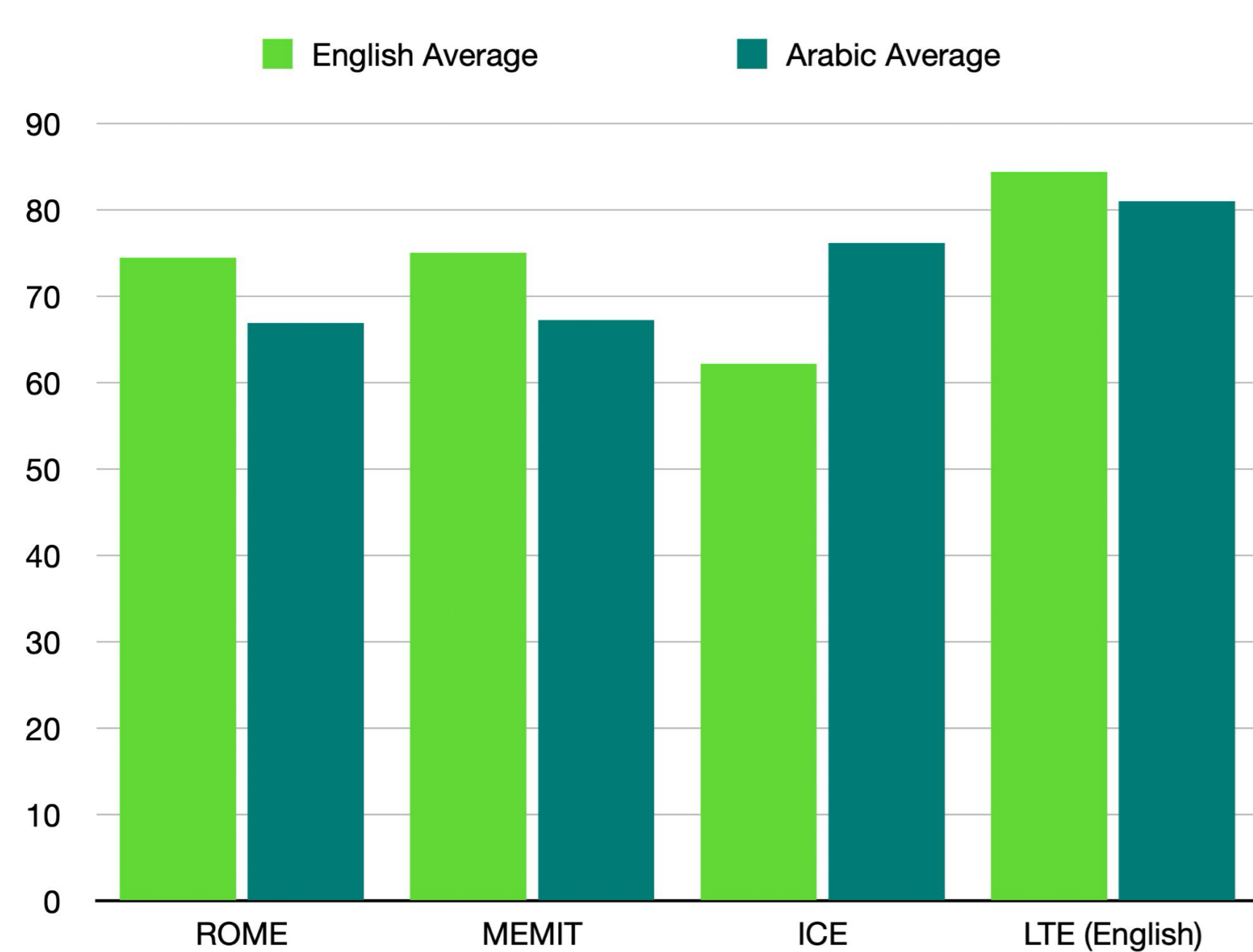
Model: Llama2-7B-Chat

Languages: English, Arabic, and German



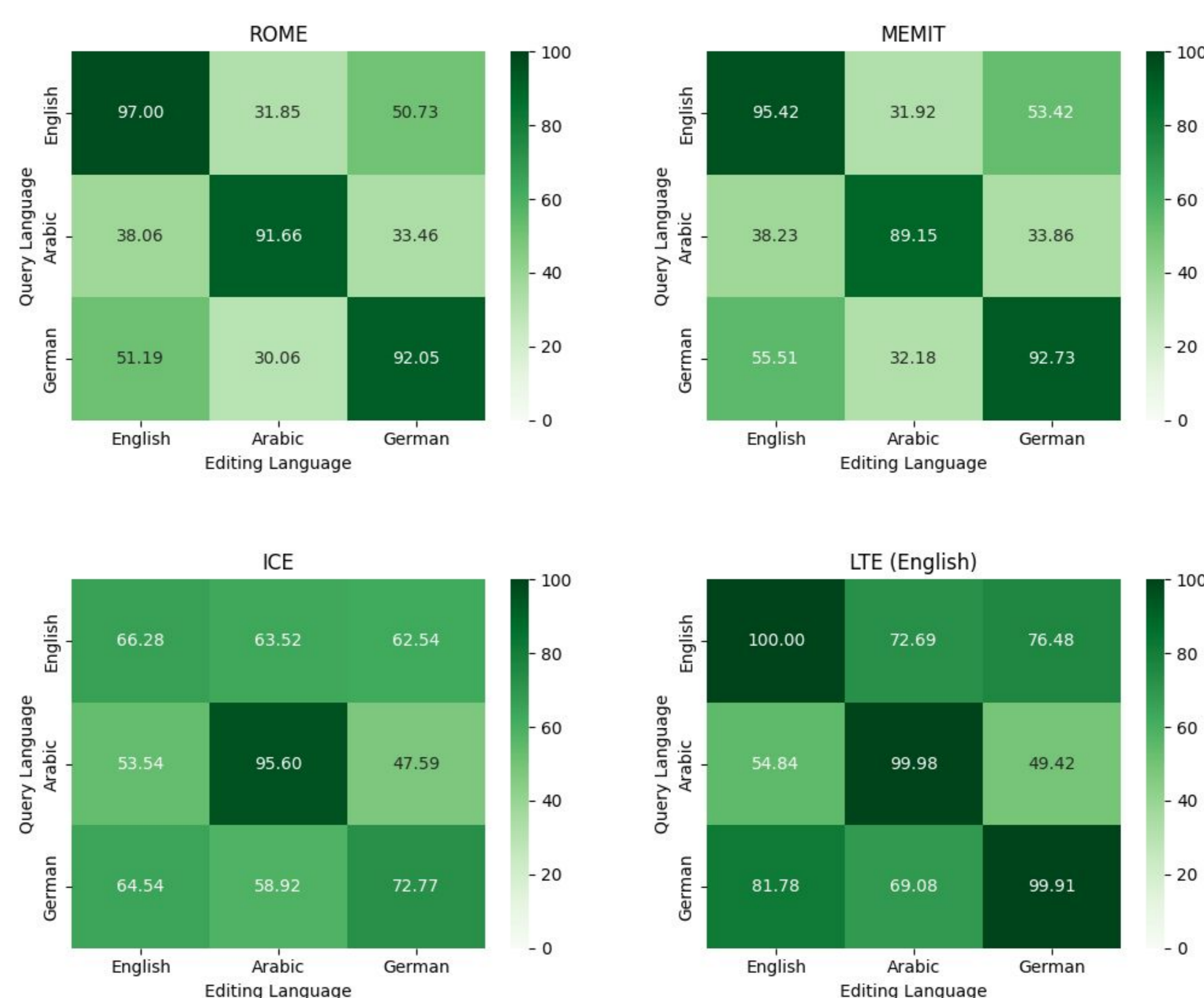
Findings

KE Effectiveness in Arabic



- **LTE (English):** Best overall
- **ICE:** Good reliability & generality, but very weak portability
- **MEMIT:** Strong locality, weaker generality & portability
- **ROME:** Lowest performance — poor transfer to Arabic

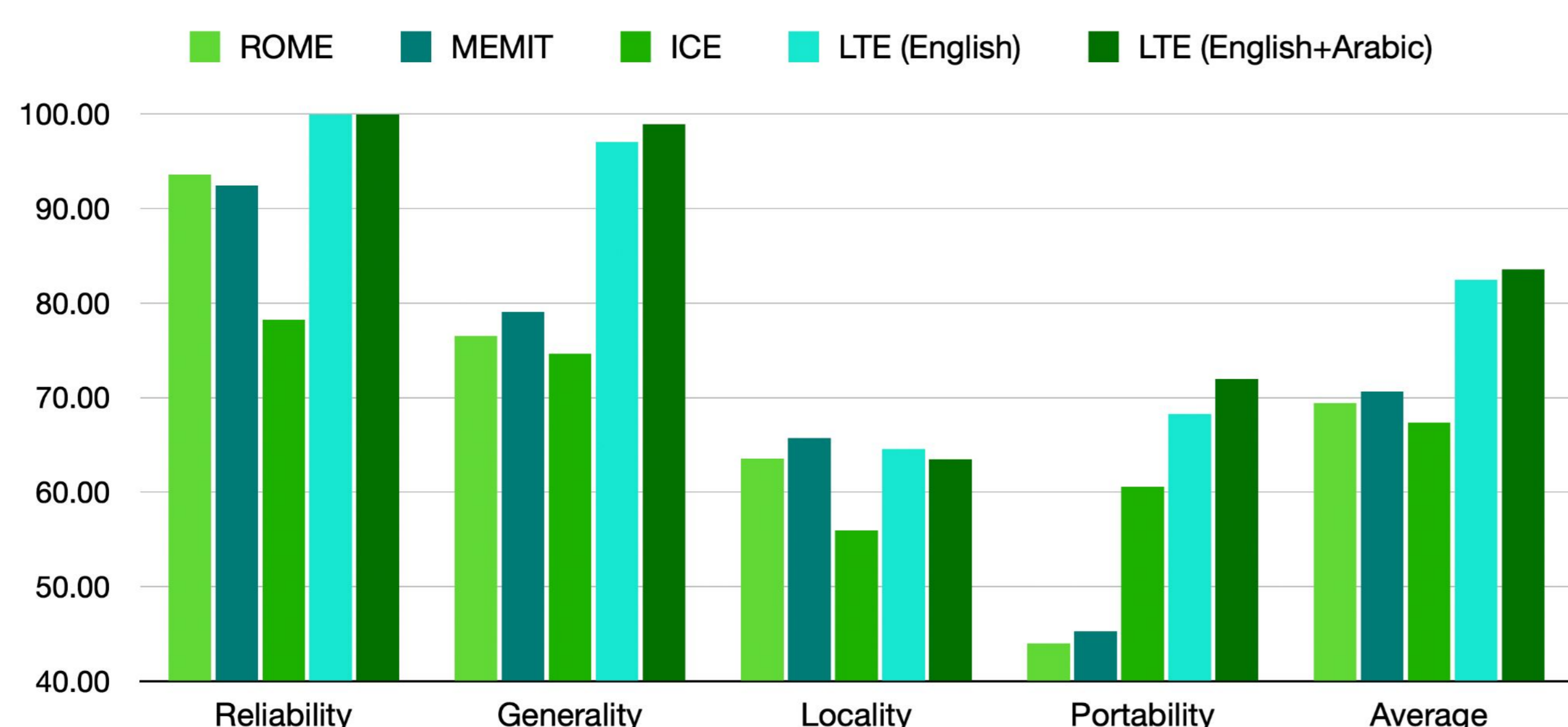
Crosslingual Propagation across languages



Crosslingual reliability scores

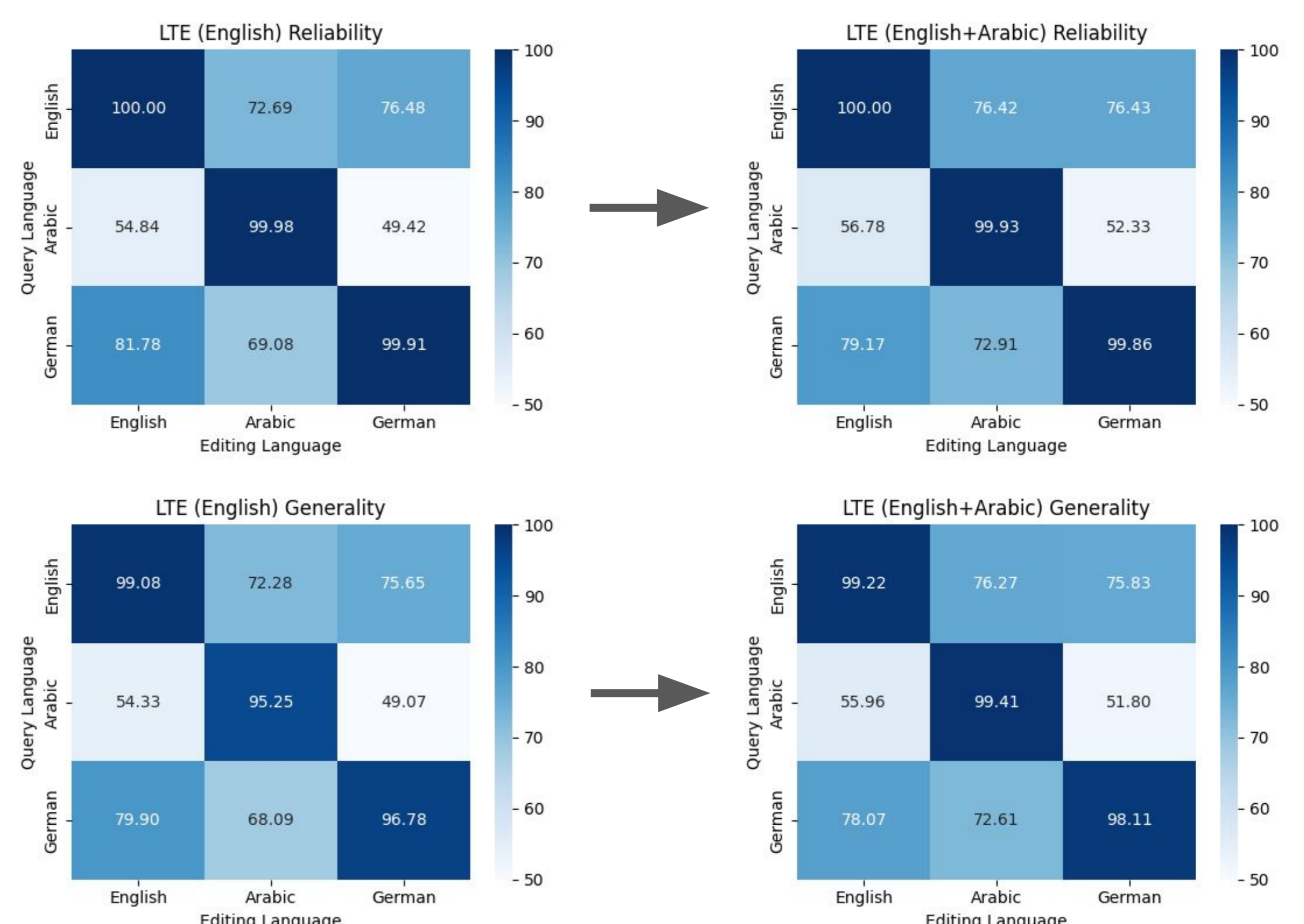
- Edits made in Arabic **fail to propagate reliably** to English or German, and vice versa.
- **LTE (English)** again has the best cross-lingual propagation
- Parametric edit methods like **ROME** and **MEMIT** are especially weak in transferring edits across languages

Enabling models to perform multilingual editing



Average scores across English, Arabic and German

- **LTE (English+Arabic)** improves Generality and Portability within each language
- **LTE (English+Arabic)** also slightly improves crosslingual Generality and Reliability, showcasing the usefulness of multilingual learning to edit training



Crosslingual **Reliability** and **Generality** after tuning