How AI Reads Coptic Ostraca: Formula, Fragment, and the Reconstruction of Meaning in O.Frange 1–27

Gunhyuk Lee
Central European University (Vienna campus)
PhD student, Historical Studies Department

Ostraca are broken pieces of pottery or stone used as writing surfaces. In late antique Egypt, they were widely employed for everyday communication, including letters, receipts, notes, and memoranda. Many were written in Coptic, the latest stage of the Egyptian language, written mainly in the Greek alphabet with additional characters. As a low-resourced ancient language, Coptic presents particular difficulties for translation, especially in documentary texts where meaning depends on convention, formula, context, and social practice.

This paper examines how generative AI responds to ambiguity, fragmentation, and interpretative plurality in Coptic documentary texts. Its case study is the O.Frange ostraca, a corpus of Coptic monastic letters from Djeme in Western Thebes, published by Anne Boud'hors and Chantal Heurtel (2010). Although the edition is larger, papyri.info provides Coptic transcriptions and French translations only for documents 1–27, which form the dataset of this study. Rather than asking only whether AI can produce accurate translations, this paper asks how AI reconstructs meaning when the surviving text is short, formulaic, damaged, or open to more than one interpretation.

The dataset preserves a broad documentary spectrum. Some texts contain practical requests concerning textiles, clothing, travel, illness, and the movement of goods. Others are highly formulaic monastic letters structured by repeated expressions of humility, greeting, blessing, and prayer. Some survive in severely lacunose form, preserving isolated phrases without adequate context. Another group combines epistolary form with scriptural citation, gnomic religious language, or spiritually charged expressions. This range makes it possible to test AI on lexical difficulty, documentary convention, formulaic language, and material incompleteness.

The comparison between AI outputs and the existing French translations reveals four recurring tendencies. First, generative AI shows a strong completion bias. Fragmentary or discontinuous texts are frequently turned into smooth and coherent statements. Second, AI tends toward semantic inflation. Formulaic expressions, such as humility formulas, greetings, and requests for prayer, are often rendered as emotionally intensified and individualized statements. Third, AI produces genre distortion. Practical documentary texts embedded in monastic and social networks are reshaped into personal letters or spiritual reflections. Fourth, AI tends to erase formula by flattening repetitive expressions into modern prose.

Examples from O.Frange 1–27 illustrate these patterns across the corpus. Practical texts (1, 5, 6, 15, 20, 22) concern objects, clothing, travel, and delivery. Formulaic letters (4, 9, 10, 11, 12, 16, 17, 18) are dominated by repeated epistolary conventions. Fragmentary ostraca (7, 19, 23, 26) preserve only partial sense units. Texts 8, 13, 14, and 24 combine letter writing with scriptural or religious discourse. In each case, AI tends to reorganize the surviving Coptic into forms of coherence and sincerity closer to modern expectations than to the discontinuous and formulaic logic of the originals.

This paper argues that AI use in papyrology should be evaluated in terms of interpretative behavior as well as translation accuracy. The central question is not only whether AI translates

correctly, but how it manages uncertainty, fills gaps, and reduces or preserves multiple possible readings. By showing how AI reconstructs meaning in short and fragmented Coptic ostraca, this study contributes to broader discussions of methodological rigor and responsible AI use in the study of ancient low-resourced corpora.