

EXAMENSARBETE Using Small LLMs to Assess and Enhance Skill Management Documents**STUDENT** Maxime Pakula**HANDLEDARE** Pierre Nugues (LTH)**EXAMINATOR** Jacek Malec (LTH)

Mastering Tool Creation: The Power of Formulating Instructions with LLMs

POPULÄRVETENSKAPLIG SAMMANFATTNING **Maxime Pakula**

Med framväxten av allt mer avancerade och tillgängliga språkmodeller verkar skapandet av verktyg för språkbehandling baserade på dessa teknologier nu vara inom räckhåll för alla. Denna studie bedömer genomförbarheten av att designa ett effektivt verktyg genom att enbart fokusera på utformningen av instruktioner.

Natural Language Processing expanderar snabbt och möjliggör för företag att automatisera textrelaterade uppgifter. Traditionellt krävde detta stora datamängder, vilket kan vara utmanande för små eller nischade företag. Nyligen har kraftfulla, öppen källkod stora språkmodeller introducerats som kan bearbeta text effektivt, men som kräver betydande datorkraft och ofta är värdextern, vilket väcker oro för datastyrning. Små språkmodeller erbjuder ett lovande alternativ. De kräver färre resurser och kan hostas lokalt, vilket ger bättre kontroll över data för företag.

Detta examensarbete, som genomfördes på ett franskt konsultföretag, syftade till att skapa ett verktyg för konsulter att skriva dokument som visar deras färdigheter och erfarenheter. Företaget står inför utmaningar, inklusive den extra svårigheten med att bearbeta data på franska, eftersom de flesta modeller är tränade på engelska. Verket som är under utveckling måste kontrollera att texterna överensstämmer med företagets riktlinjer och kunna förbättra dem vid behov. Under min forskning skapade jag först ett testdataset och valde jämförelsemått. Därefter testade jag olika små språkmodeller för att bedöma deras effektivitet baserat på språket som instruktionerna är skrivna på och mängden

exempel som ges. Slutligen jämförde jag de erhållna resultaten med en mänsklig bedömning.

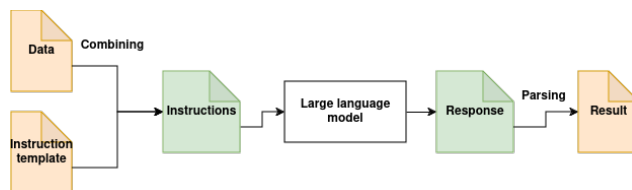


Figure 1: Schematisk vy av verktyget

Även om metoden är ganska enkel, är resultaten mycket uppmuntrande. De visar att vissa små språkmodeller redan är tillräckligt avancerade för att automatisera denna typ av uppgift. Jag fann särskilt att när texterna är på franska, är det bättre att ge instruktionerna på franska, även om modellerna främst har tränats på engelska. Jag observerade också att det finns ett optimalt antal exempel att ge för att uppnå de bästa resultaten, och för texter av några stycken som de i denna studie, är det optimala antalet mellan 2 och 3 exempel. Dessa resultat antyder att vi bör fortsätta utforska mer sofistikerade instruktionstekniker med dessa små modeller för att ytterligare förbättra prestandan hos de utvecklade verktygen.