

EXAMENSARBETE Application Specific Instruction-set Processor Using a Parametrizable multi-SIMD Synthesizeable Model Supporting Design Space Exploration**STUDENT** Tristan Peetre**HANDLEDARE** Marcus Clang (LTH), samt en handledare som representerade ärendeföretaget**EXAMINATOR** Jacek Malec (LTH)

När AI Talar Företagets Språk: Lokala och Molnbaserade LLM-lösningar för Grafdatabaser

POPULÄRVETENSKAPLIG SAMMANFATTNING Tristan Peetre

Denna studie utforskar möjligheten att använda LLM:er för att omvandla en användares frågor för att prata med ett företags grafdatabas. Genom utvärdering av både lokala och molnbaserade modeller visar resultaten att molntjänster ger bäst prestanda, till priset av minskad datasekretess.

Jag har undersökt hur AIs kan användas för att automatiskt översätta frågor från naturligt språk till *Cypher*-frågor, som används för att hämta information från grafdatabaser hos företag. Bakgrunden till arbetet är ett tydligt behov: det är idag ofta krångligt för icke-tekniska användare att förstå hur man får tillgång till ens företags data. Målet var därför att sänka tröskeln för att söka information, utan att förlora dataskyddet, vilket innebär att modellerna borde kunna köras lokalt, utan att skicka känslig data till molnet.

Resultaten visar att det är tekniskt möjligt att använda lokala AI modeller för detta ändamål, särskilt efter *fine-tuning* (träna dem med testfrågor). Dock visade det sig att de klart bästa resultaten uppnåddes av betalda molnmodeller, vilket illustrerar den tydliga fördelningen mellan sekretess och prestanda. En intressant överraskning var att en mindre, *fine-tuned* modell ibland kunde konkurrera med betydligt större modeller på vissa frågor, vilket tyder på att en domänanpassning kan vara viktigare än modellstorlek i vissa fall.

Det här arbetet kan få direkt tillämpning i utvecklingen av smarta assistenter som kan tolka frågor från användare och automatiskt hämta

rätt information, helt utan att användaren behöver förstå en komplicerad uppbyggnad. På sikt kan detta förbättra tillgången till information, effektivisera arbetsflöden och minska behovet av teknisk support inom företag.

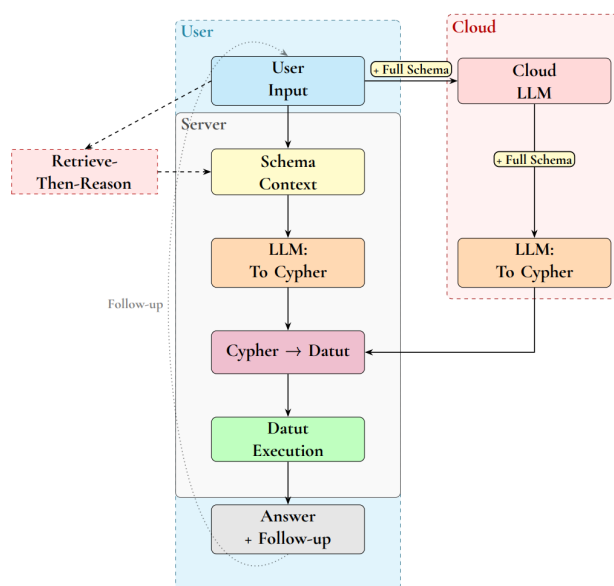


Figure 1: En förenklad version av pipelinen som testades