# THE NATIONAL UNIVERSITY OF LESOTHO

Department of Mathematics and Computer Science

CS3520 – Computer Organisation and Architecture I

Domain Analysis Report

23rd September 2025

## 1. Application Context

The chosen application domain is low-cost AI-enabled mobile phones designed for the African context, with a focus on Lesotho. Mobile phones are an essential in most people’s lives in Lesotho for many reasons like communication, banking, health care and many others. With the current growth in the demand for AI-enabled mobile phones , there is an increase to the cost of those phones which results in majority of people not having access to them.

The AI-enabled mobile phone is going to help most of Basotho citizens with different uses like using voice assistant to call relatives easily, and to give commands like health check for elderly users who live far from hospitals. This also has features like battery and reserve optimization which is very crucial since most Basotho citizens have limited access to electricity. The opportunity lies in designing AI-powered mobile devices that are affordable and easy to access by all locals.

This domain is relevant because Africa is in need of digital transformation since it is behind in terms of technology and most countries are already up to date with technology, hence this AI-powered mobile phone will enhance the use of technology in Lesotho.

## 2. Representative Applications

|  |  |  |  |
| --- | --- | --- | --- |
| **Application** | **Short Description** | **Key Computational Tasks** | **Notes (Relevance)** |
| Voice Assistants | Enable voice-based commands such as calling or sending notes | Speech-to-text, NLP parsing, intent recognition | Improves accessibility and hands-free use for everyday tasks |
| Mobile Security | Unlock phone/transactions via PIN, password, or voice | Cryptography, hashing, biometric recognition | Provides secure access for mobile banking and e-services |
| Predictive Text & Autocorrect | Suggests next word, fixes spelling errors | N-gram models, deep learning, sequence prediction | Enhances communication efficiency in messaging apps |
| Battery & Resource Optimization | Manages background processes to save energy | Process scheduling, low-power optimization | Ensures longevity of affordable mobile devices |
| Offline Language Translation | Translate Sesotho English for daily communication | Sequence models, embedding search, NLP inference | Bridges communication barriers in education and trade |

## 3. Workload Characteristics

- Arithmetic intensity: Heavy use of linear algebra (dot products, matrix multiplications) and nonlinear activation functions.  
- Data movement: Requires efficient access to model parameters (weights) stored in memory, with frequent reuse.  
- Memory usage: Moderate to high; models like CNNs and translation require large parameter storage, but optimizations such as quantization can reduce footprint for low-cost devices.  
- Latency sensitivity: Real-time interaction (e.g., voice authentication, translation) requires low-latency inference.  
- Energy efficiency: Since devices are low-cost and battery-powered, workloads must be optimized for low energy consumption.

## 4. Initial Insights for Processor Design

- It must be easy to use, and reliable meaning it must not crash while in use.  
- It must assist user to access AI without the use of internet.   
- Should include fast memory access with support for frequent small data fetches.  
- ISA should provide extensions for ML primitives (e.g., SIMD, MAC operations).  
- Optimized for low-latency inference while conserving energy.

## 5. References

## - Sze, V., Chen, Y., Yang, T., & Emer, J. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proceedings of the IEEE. - World Bank (2022). Digital Economy for Africa Report.