[This capstone idea is subject to change please..]

Using Hidden Markov Model in Women in STEM Retention

Women enter STEM programs in African universities but drop out at high rates. Current support programs are ineffective because they don't target the right intervention moments.

Describe the Observations

- **Academic metrics:** Course grades, attendance rates, assignment submissions, GPA trends
- **Social engagement:** Study group participation, STEM club membership, mentorship involvement
- **Digital behaviours:** LMS login frequency, discussion forum activity, resource usage
- **Support utilisation:** Advisor meetings, counselling services, help-seeking patterns
- **Temporal patterns:** Registration behaviours, course load changes, research participation

Type of HMM Problem: This is an **unsupervised HMM/ Learning problem**. Hidden engagement states are unknown; thus, I must:

1. Discover optimal number of engagement states ("Highly Engaged," "Struggling," "At Risk")
2. Learn transition probabilities between states
3. Identify emission probabilities linking behaviours to engagement levels

Training Algorithm = Baum-Welch algorithm for parameter estimation

a. Known Values: Complete observation sequences from historical student data, Predetermined number of hidden states (5 engagement levels), Student demographic information and program details, Academic calendar structure and course requirements, Initial parameter estimates (random or literature-informed)

b. Unknown Values:

- **Hidden state sequences:** Actual engagement progression for each student
- **Transition probabilities (A)**: Likelihood of moving between engagement states
- **Emission probabilities (B):** Probability of observing behaviours given engagement state
- **Initial state distribution (π):** Starting engagement state probabilities

Parameter Updates

- **Transition Matrix (A):** Learning how engagement evolves over time, capturing critical periods like sophomore year challenges or pre-graduation phases.
- **Emission Matrix (B):** Determining which behavioural combinations indicate each engagement state, accounting for diverse student backgrounds and learning styles.
- **Initial State Distribution (π):** Establishing baseline engagement probabilities for incoming female STEM students based on preparation and demographics.