# Data Dictionary

* **Id**: Unique identifier for each keystroke event.
* **StartTime & EndTime**: Timestamps of when a key was pressed and released.
* **KeyPressed**: The individual key that was pressed.
* **FullWord**: The accumulated word being typed.
* **FlightTime**: Time between consecutive key presses.
* **KeyHoldTime**: Duration a key was held down.
* **DeviceId**: Identifier for the device.
* **CreatedAt & ModifiedAt**: Timestamps of when the data entry was created or modified.

# Using HMM to Infer Hidden Variables

A Hidden Markov Model (HMM) can be used to uncover hidden cognitive states based on observed keystroke patterns. Here’s what we can do:

Observed Variables (What we see in the data):

***FlightTime*** (Time between key presses)

***KeyHoldTime*** (Duration a key is held)

***KeyPressed*** (Can be encoded as numerical values)

***Hidden Variables (What we want to infer):***

* Typing states such as "Fluent Typing", "Paused", "Error Correction", "Hesitation".
* Typing behaviour.

# **Approach**

* Encode KeyPressed, FlightTime, and KeyHoldTime as numerical features.
* Use an **HMM with Gaussian emissions** (hmm.GaussianHMM) since FlightTime and KeyHoldTime are continuous variables.
* Train the HMM to learn the hidden typing states.
* Predict hidden states for new keystrokes.

# **Interpreting the Results**

* Each row will now have a **"HiddenState"** column, representing different inferred typing behaviors, such as:
* **State 0:** Fluent Typing
* **State 1:** Hesitation
* **State 2:** Error Correction

# Hidden States

|  |  |  |  |
| --- | --- | --- | --- |
| **Hidden Variable** | **Possible Hidden States** | **Interpretation** | **How to Identify Each State?** |
| **Typing Behavior** | Fluent, Hesitant, Error Correction, Random | Detects if the user is typing smoothly or struggling | - **Fluent:** Low inter-key delay, low variance in flight/hold times |
| - **Hesitant:** High inter-key delay, increased pauses |
| - **Error Correction:** High backspace/delete usage |
| - **Random:** Irregular patterns in flight/hold time |
| **Cognitive Load** | Low, Medium, High | Measures mental effort based on typing speed & pauses | - **Low:** Fast typing, consistent inter-key delay |
| - **Medium:** Slight variation in inter-key delay and hold time |
| - **High:** High variance in keystroke timing, frequent pauses |
| **Typing Fatigue** | Fresh, Slightly Fatigued, Highly Fatigued | Detects user fatigue based on keypress delays | - **Fresh:** Low variation in typing speed, steady rhythm |
| - **Slightly Fatigued:** Increasing key hold time, slower transitions |
| - **Highly Fatigued:** Very high hold time, frequent pauses, increased backspace usage |
| **Typing Skill Level** | Novice, Intermediate, Expert | Distinguishes between different levels of typing expertise | - **Novice:** High inter-key delay, slow flight times, inconsistent rhythm |
| - **Intermediate:** Moderate typing speed, occasional hesitation |
| - **Expert:** Fast, consistent typing with minimal pauses |