Algorithm 1 Dual-LLM Quant Factor Mining Pipeline w/ Self-Correction

Input: Task List $\mathcal{T}$, Data Bundle $\mathcal{D} = \{\mathcal{D}_{stock}, \mathcal{D}_{index}\}$
Output: Factor Library (Python Codes $\mathcal{F}_{code}$, Parquet Data $\mathcal{F}_{data}$, CSV Records $\mathcal{R}$)

Initialize $\mathcal{M}_{idea} \leftarrow$ Ideation Model
Initialize $\mathcal{M}_{code} \leftarrow$ Coding Model
Initialize Run Directory P_{run} with Timestamp
for each seed task s in $\mathcal{T}$ **do**
  // **Phase 1: Ideation (The Brain)**
  $\mathcal{V} \leftarrow \mathcal{M}_{idea}$.Ideate($s$.idea, s.num, DataSchema)
  for each variation v in $\mathcal{V}$ **do**
    $name_{orig} \leftarrow v$.name, $formula \leftarrow v$.formula
    $C_{curr} \leftarrow \mathcal{M}_{code}$.GenCode($v$.desc, $formula$)
    $status \leftarrow$ "Fail", $name_{unique} \leftarrow$ null
  // **Phase 2: Execution & Self-Correction Loop**
    for $attempt \leftarrow 0$ **to** $MAX\_RETRIES$ **do**
      $target \leftarrow (attempt > 0)?name_{unique} :$ null
      $(Func, name_{unique}, path) \leftarrow$ CodeManager.SaveAndLoad($C_{curr}, name_{orig}, target$)
      if $Func ==$ null **then**
        $C_{curr} \leftarrow \mathcal{M}_{code}$.Refine($C_{curr}$, "Syntax Error", $formula$)
        continue
      end if
      $(is\_success, msg) \leftarrow$ Executor.Run($Func, \mathcal{D}$.copy())
      if $is\_success$ **then**
        $status \leftarrow$ "Success"
        Save Factor Data to P_{run}/factors/$name_{unique}.parquet$
        break
      else
        if $attempt < MAX\_RETRIES$ **then**
          $C_{curr} \leftarrow \mathcal{M}_{code}$.Refine($C_{curr}, msg, formula$)
        end if
      end if
    end for
  // **Phase 3: Cleanup & Recording**
    if $status \neq$ "Success" **then**
      Delete file at $path$
    end if
    Record Metadata to $\mathcal{R}$ ($name_{unique}, status, path, timestamp$)
  end for
end for