Based on the input text and image content, consider which submodules of the Linguistic Specialists and Visual Specialists should be invoked when performing sarcasm detection.

Integrate the outputs of each sub-task agent and ultimately make a sarcasm judgment.

Text

oh i so love working late from home…

Image

Input

Task Analysis

Task Allocation

Commander GPT

Module invocation

Linguistic Specialists

Context Modeling

Sentiment Analysis

Rhetorical Device

Visual Specialists

Facial Expression

Scene Text

Image Summarization

Callback

Integration

Decision making

Commander GPT

context_modeling: Analyze broader conversational context
sentiment_analysis: Extract emotional polarity
rhetorical_device: Identify irony, hyperbole, etc.
facial_expression: Analyze facial emotions in image
image_summarization: Describe visual content
scene_text: Extract text from image

Context Modeling :On a Friday night, she…;Sentiment Analysis:love;Rhetorical Device Recognition:irony;Facial Expression Recognition:happy;Image Summarization:arafed woman with a tattoo on her neck sitting in a living room;Scene Text Recognition:None
Taking all of the above into account, it's ultimately judged to be sarcastic.