

Adversarial Text Generation using Large Language Models for Dementia Detection

UMass
Boston



Youxiang Zhu, Nana Lin, Kiran Sandilya Balivada, Daniel Haehn, Xiaohui Liang

1. Introduction

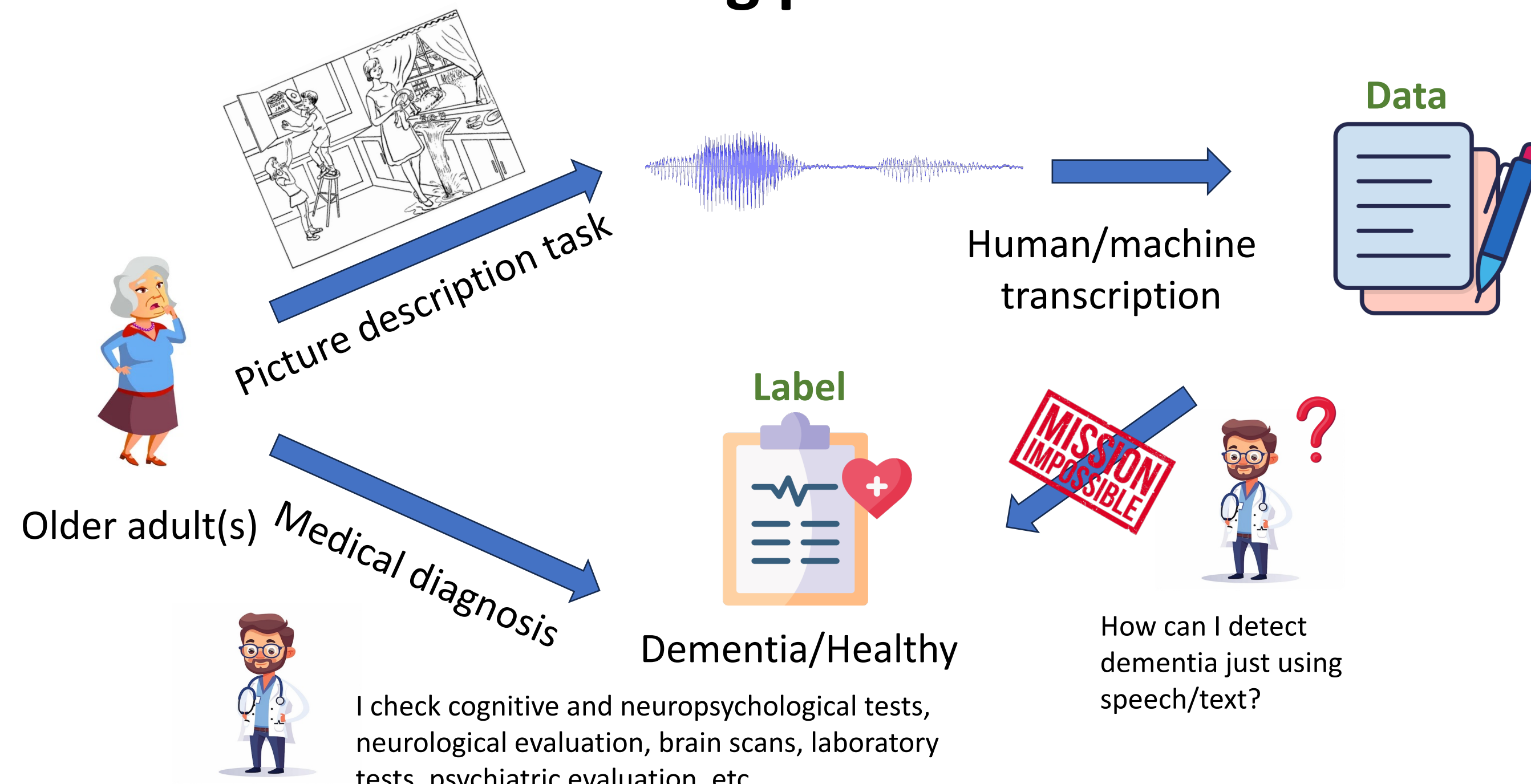
LLMs can Generalize to unseen task with a detailed task instruction Think step by step follows human demonstrations from training or in-context

But what if The task instruction is unclear Humans cannot demonstrate accurate intermediate steps

Dementia detection is such a super-human-level task! LLMs struggle to relate their internal knowledge to dementia detection.

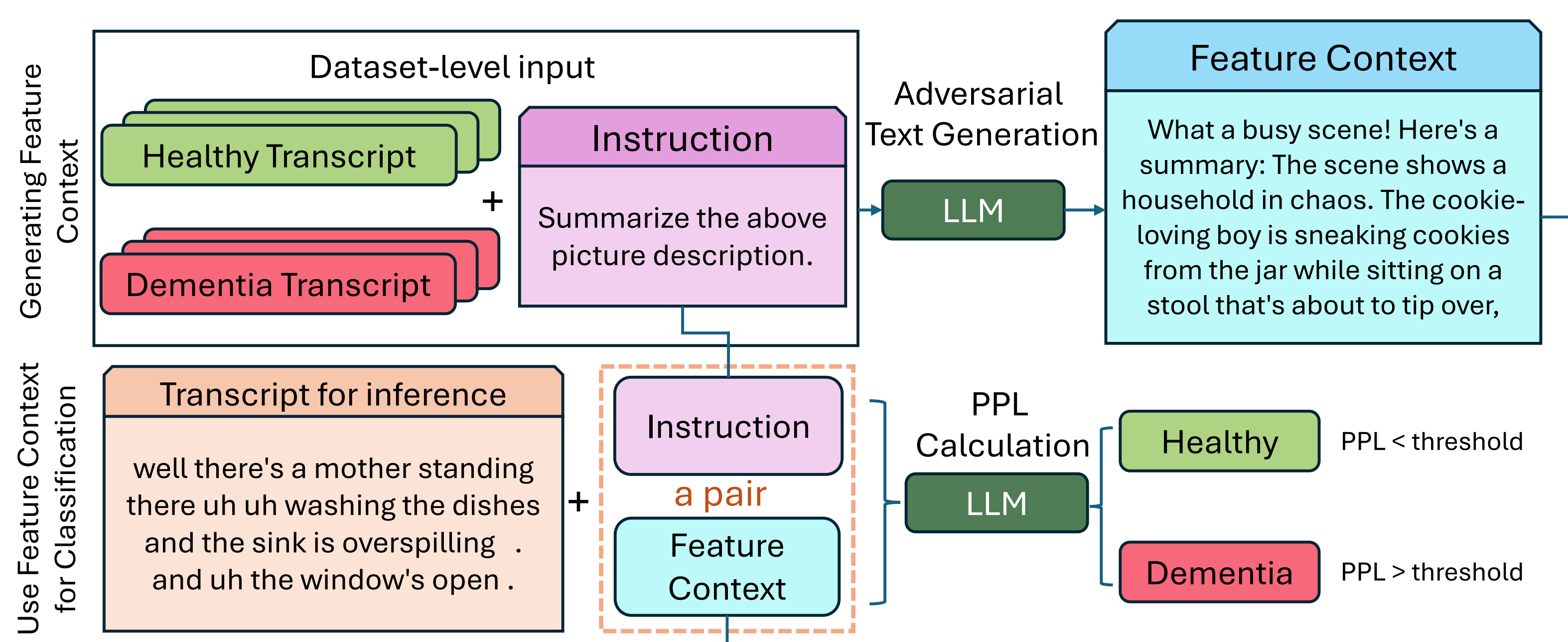
We introduce adversarial text generation (ATG) to relate dementia detection with other tasks!

Data collection and labeling process of dementia detection



LLM Few shot or CoT prompting -> 55-75% ACC, even worse than fine-tuning BERT ~80% ACC

2. How ATG relate dementia detection with other tasks?



Three steps to use ATG for dementia detection:

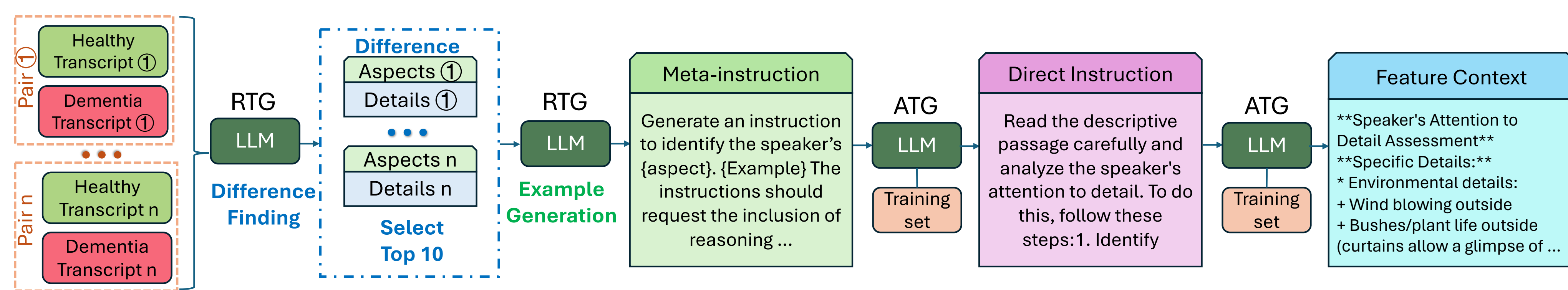
- Step 1: Pick a task instruction
- Step 2: Generate feature context using ATG
- Step 3: Use feature context for classification

Perplexity: a measurement of how a LLM familiar to a text

Regular text generation (RTG) vs. ATG

	RTG	ATG
Input	a text	a dataset and an instruction
Output	a text	a text (feature context)
Objective	pick next token to <u>minimize the perplexity</u> of the text sequence	pick next token to <u>maximize the perplexity difference</u> between two classes

3. How to find best task instructions with ATG?



Four steps: Difference finding -> Meta instruction construction -> Direct instruction generation -> Feature context generation

4. Key takeaways

Performance:

- ATG achieves >10% ACC improvement compared to regular prompting baselines
- ATG less prone to overfitting compared to fine-tuning

Feature contexts:

- Good performed feature contexts highlighted the difference between dementia and healthy samples

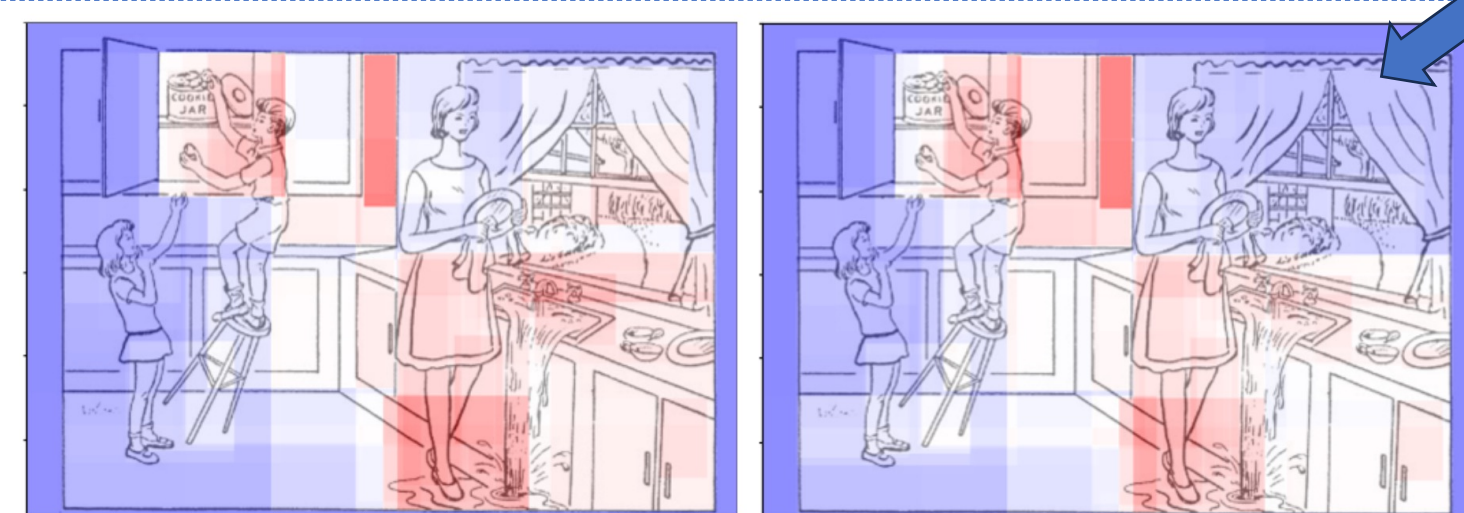


Figure 2: The focused area of HC (left) and AD (right). Red means highly focused and blue means lowly focused. Lowly focused area for dementia participants. Consistent with the finding of image text alignment in previous work.

Tasks and features related to dementia:

- Tasks: assessing attention to detail, language, clarity, etc.
- Features: environment, character, and other picture content, language-related features, etc.

MoCA Full assesses:
• Short-term memory
• Visuospatial abilities
• Executive functions
• Attention, concentration, and working memory
• Language
• Orientation to time and place

Consistent with the aspects to examine of existing cognitive screening tools (e.g., MoCA)

Attention to detail
Assessment Task: Speaker's Attention to Detail Read the descriptive passage carefully and analyze the speaker's attention to detail. To do this, follow these steps: 1. Identify the specific details mentioned in the passage, focusing on the minor elements such as object positions, character actions, and surroundings. 2. Categorize the types of details mentioned, such as: * Environmental details (e.g., wind, curtains, water overflowing) * Character actions (e.g., cookie-stealing, stool overturning, mother drying dishes) * Object positions and arrangements (e.g., stool falling, cookies in the jar) * Character requests (e.g., little girl asking boy to be quiet) 3. Consider the level of precision used to describe each detail. For example, the passage notes specific aspects such as the sink overflowing, but does not describe the temperature or humidity of the environment. 4. Evaluate how the speaker's attention to detail impacts the overall picture painted by the passage. For example, the detailed descriptions of character actions and surroundings create a vivid scene, but the mother's inattention to the overflow suggests that her priorities are elsewhere. 5. Based on these findings, conclude how the speaker's attention to detail affects the passage's overall effect. Is the level of detail meticulous and thorough, or does it prioritize some elements over others? What is the likely effect of the speaker's focus on detail on the reader?< eot_id >< start_header_id >assistant< end_header_id > **Speaker's Attention to Detail Assessment** **Specific Details:** * Environmental details: + Wind blowing outside + Bushes/plant life outside (curtains allow a glimpse of the outdoor view) + Open window/curtains + Overflowing sink + Weather appears to be light enough to cause the stool to overturn (suggesting not too intense or turbulent) * Character actions: + Girl reaching for cookies from the boy + Boy slipping/stalling with cookie (cookie out of the jar, stool tipping over) + Mother drying dishes while oblivious to the overflowing sink + Daughter with finger to her mouth (talking quiet, be quiet request to brother) * Object positions and arrangements: + Cookies in the jar, being transferred between boy and sister + Stool close enough for it to fall + Plates/cups (dry/ not being washed/dried on the counter) * Other details: + Summer (water drying her feet not suggesting soaking-wet surface) **Categorization of details:** The speaker pays attention to both macro-level elements (water running out of the sink, wind, cookie-snatching kids) and micro-level details (apron-wearing mom, short-handed kids, slips on socks). There's an evident effort to note all manner of movement within the space: the stools fall, girl reaching, brother giving/losing cookies, the curtains in motion (tied and back slightly away from the breeze) **Precision of description:** Most of the descriptions in this passage show remarkable detail and specific location: The overflowing sink overflowing into the top or splashing onto the mother's wet shoes wouldn't that um let alone uh plate um wiping dish sink um except outside nothing action there fine okay that's fine. Contrastingly, the outside situation lacks similar specificity (I did say...I guess action alright.) **Impact on the passage:** The passage effectively conveys a chaotic and potentially disordered scene within the kitchen, while implying that the mother might not be attending to all aspects of the space (stool slippers into chaos while wiping the dish, breeze seemingly ignored). Some actions stand out as problematic (stealing, stool overturning), while others might seem relatively innocuous (curtains billowing). **Speaker's attention to detail:** Overall, the speaker demonstrates meticulous attention to the movements and postures within the kitchen, whereas surrounding areas (particularly outside) receive more vague or observational comments. While there are few weak links (ums, laughs, trailing thoughts), the writer skillfully captures multiple simultaneous actions, object arrangements, and sensations (action-packed vs daydreaming daughter). The likely effect of the speaker's focus on detail is a visceral, snapshot impression of kitchen mayhem, tempered by some interpretive uncertainty (summer outside unclear, outside calm hard to describe). The careful reader will anticipate trouble erupting soon (crack on the head?), highlighting the daughter's reach into cookies amidst potential

Example feature context