

GazeSummary: Exploring Gaze as an Implicit Prompt for Personalization in Text-based LLM Tasks

Jiexin Ding¹, Yizhuo Zhang¹, Xinyun Liu², Ke Chen¹, Yuntao Wang³, Shwetak Patel¹, Akshay Gadre¹

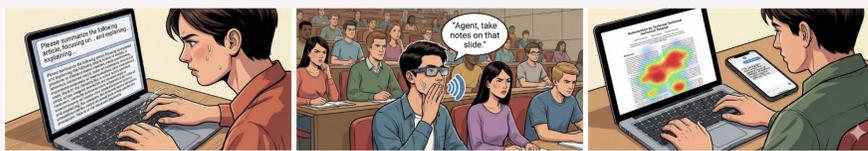
¹University of Washington, ²Google, ³Tsinghua University



Introduction

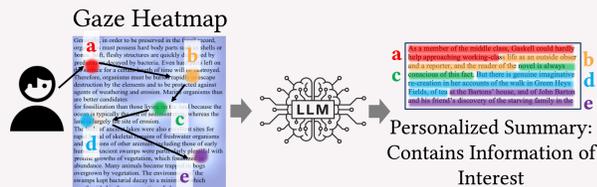
The Problem: Otrusive Personalization

- Current LLM personalization relies on **explicit, high-effort** user prompts.
- Voice and text prompts are **obtrusive** in mobile and wearable contexts.
- Explicit commands often fail to **capture the subtle nuances** of user attention.



Proposed Solution: Gaze as Implicit Prompt

- Gaze: reflect user's **focus** without interrupting user.
- LLM: strength in generating high-quality summaries.

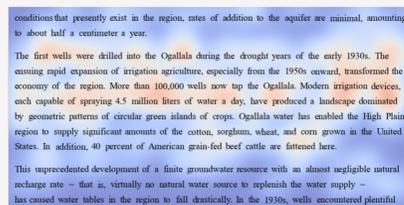


Gaze as Implicit Prompt

We explored several **gaze representation** methods as input to an LLM for generating personalized summaries.

Gaze Representations

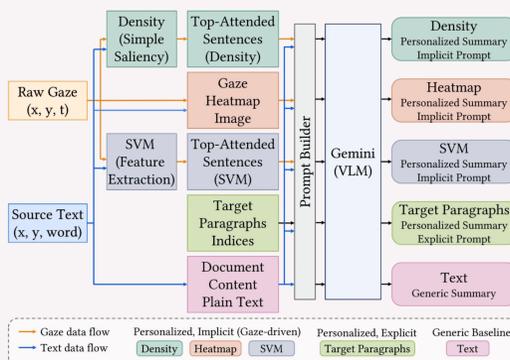
- Gaze **Density**: Rank sentences based on their total **fixation duration** and selecting the top 20%.
- Gaze **Heatmap**: Aggregate gaze points into a heatmap and overlay it onto the text as an **image-text pair**.
- Gaze **SVM**: Extract 26 **gaze features** and classify attention level of each sentence using SVM.



Heatmap

Baseline

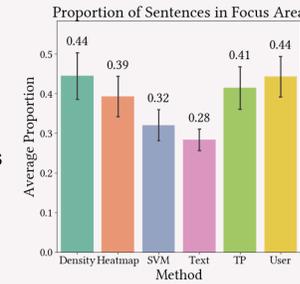
- **Target Paragraphs**: An **explicit-prompt** method that simulates users **manually** inputting paragraphs as interest. The inputs are two target paragraphs indices and source text.
- **Text**: A text-only method that generates a **non-personalized** summary.



Evaluation

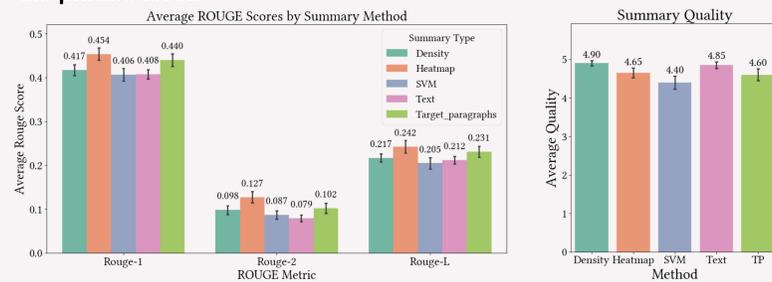
Content Focus Analysis

- Proportion of sentences in the summary that corresponded to these target paragraphs in the source text.
- **Higher ratio** => summary content is **more aligned** with the paragraphs user should focus on => **more personalized**.
- Gaze **effectively guides LLMs** to generate summaries that focus on the content users attend to.



Similarity with the User-Provided Summary

- At the lexical level, **Heatmap** outperforms other gaze representations.
- Heatmap even **surpasses explicit-input** summaries (Target Paragraphs) at the **phrase level**.



Summary Quality

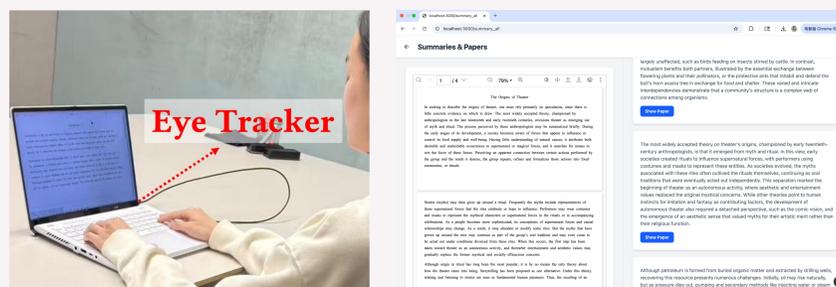
- Incorporating gaze information can make summaries more personalized **without compromising their overall quality**.

Takeaway 1

- **Heatmap** performs **best** as a gaze representation.
- Heatmap **outperforms explicit** user input at the phrase level.
- Gaze-based personalization does **not degrade** overall quality.

Gaze-Guided Assistant

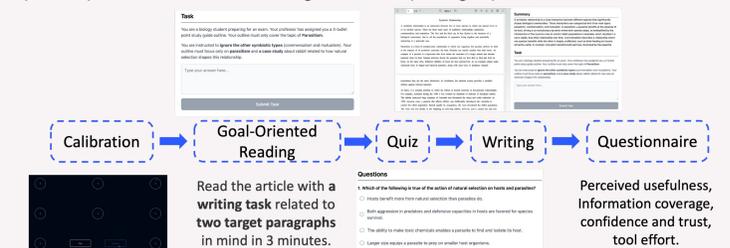
The system **automatically** generates summaries after reading based on **gaze heatmap** and displays them as references during writing.



User Study

Task: Goal-Oriented Reading

- 12 participants. TOEFL reading with 5-6 paragraphs, ~700 words.

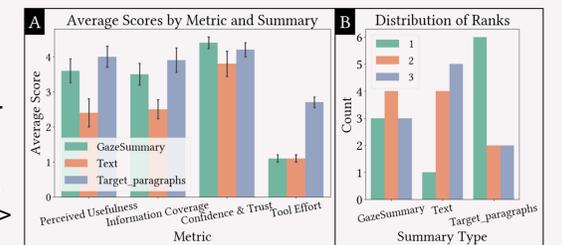


Baseline

- Text: non-personalized generic summary without gaze information.
- Target Paragraphs: user manually input the prompt.

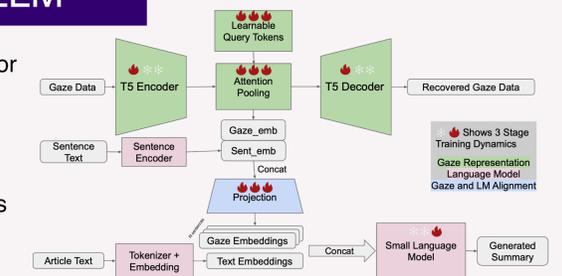
Takeaway 2

- Personalized summary is **more useful and better aligned** with users' focus than the generic summary.
- GazeSummary requires **less effort to use** than inputting prompt manually.
- Explicitly entering prompt > GazeSummary > text-only baseline.

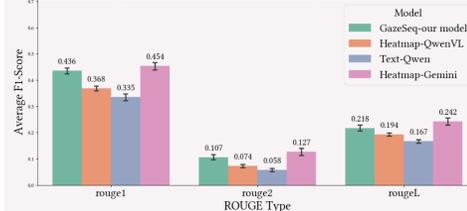


Gaze-Aware LLM

Existing VLM are inefficient for the gaze-oriented task. We propose a **compact, locally deployable** model that achieves performance **comparable** to larger models in gaze-based text summarization.



Average ROUGE Score with User Summary for Different Models



- Our model performs significantly **better than Qwen-3B** and achieves performance **comparable** to Gemini.
- Our model outperforms Qwen-VL, showing that it **handles gaze information more effectively** than Qwen-VL's vision encoder.

Contact Info

- Email: jxding@uw.edu



Full paper Personal website LinkedIn