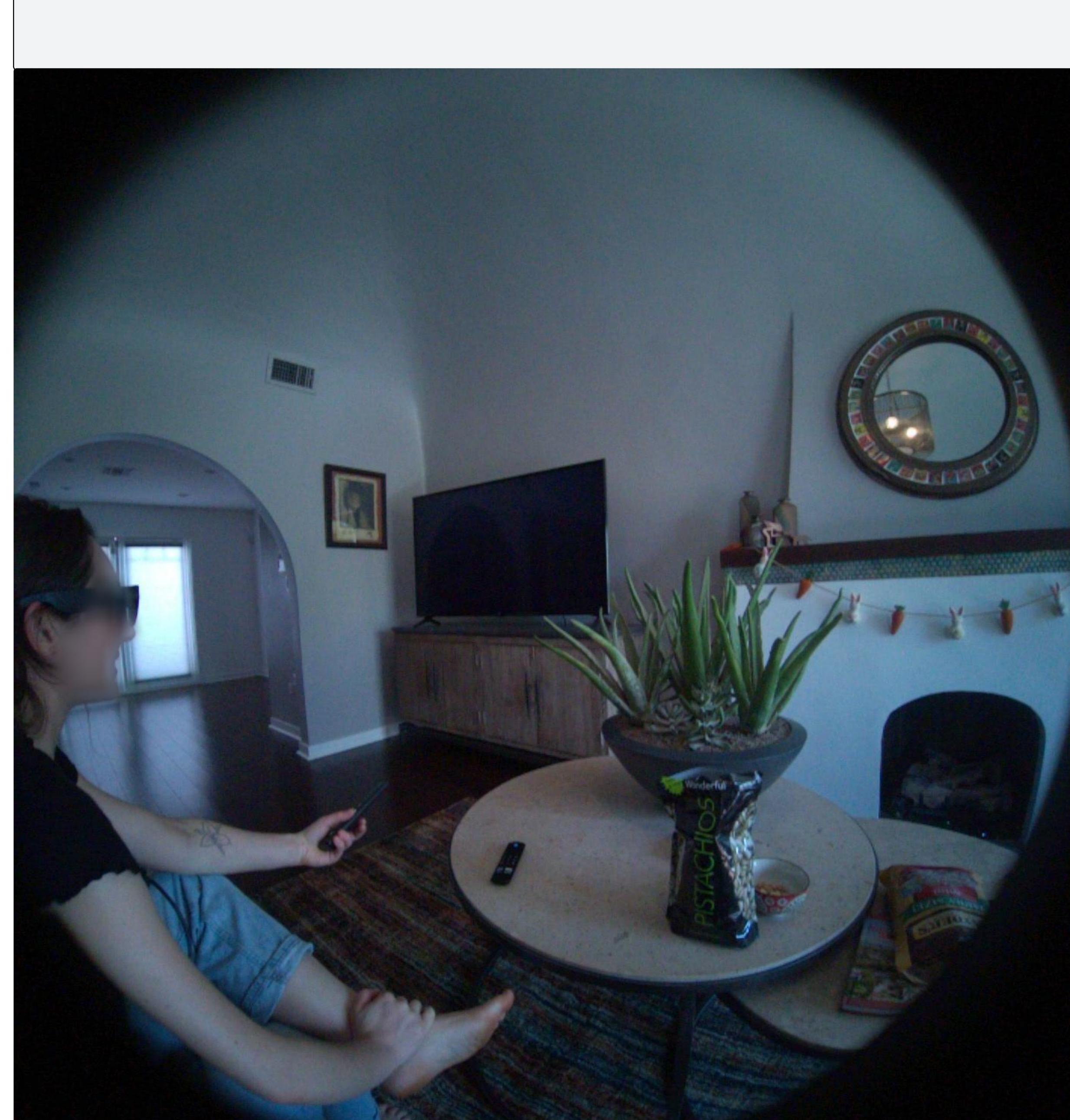
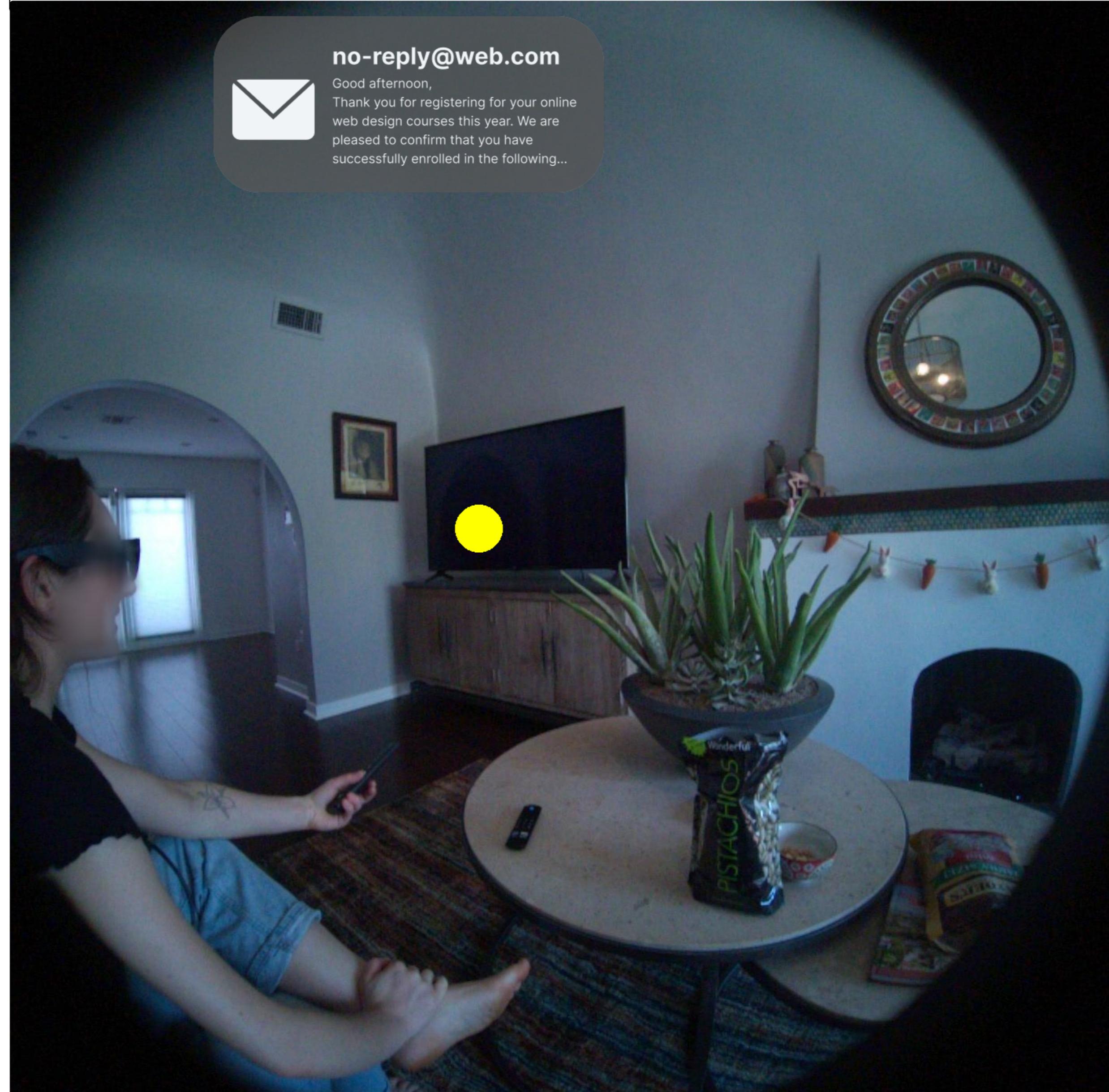


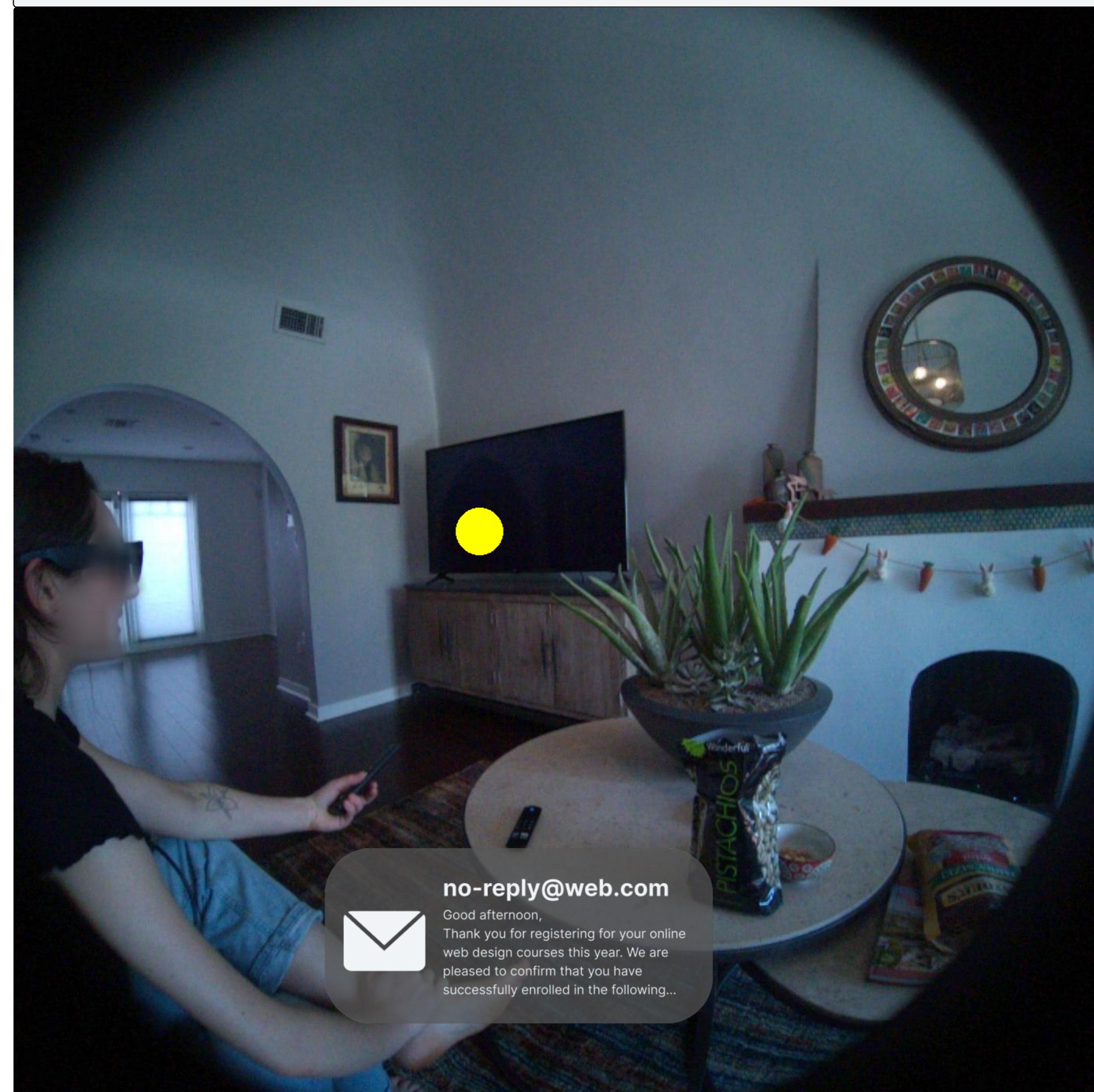
## 1. Current user view



## 2. UI layout 1 with eye gaze and UI element overlaid



## 3. UI layout 2 with eye gaze and UI element overlaid



## 4. Text instruction

Compare these three identical views from a head-mounted camera. The first is the current user view. The second is one potential UI layout with a UI element overlaid, and the third is another potential UI layout with the element appearing in a different location. Which is the better layout, the first (1) or second (2)?

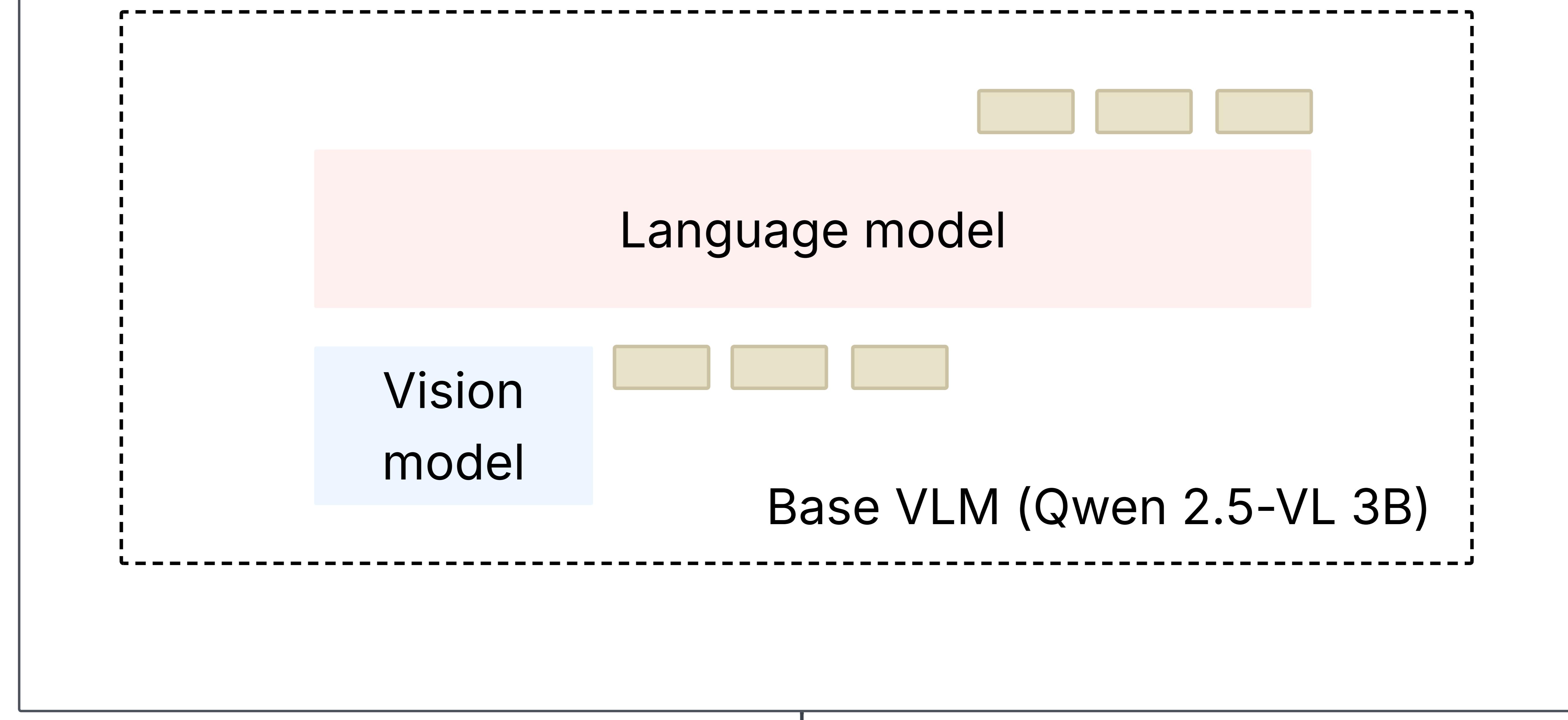
## 5. Target label

1

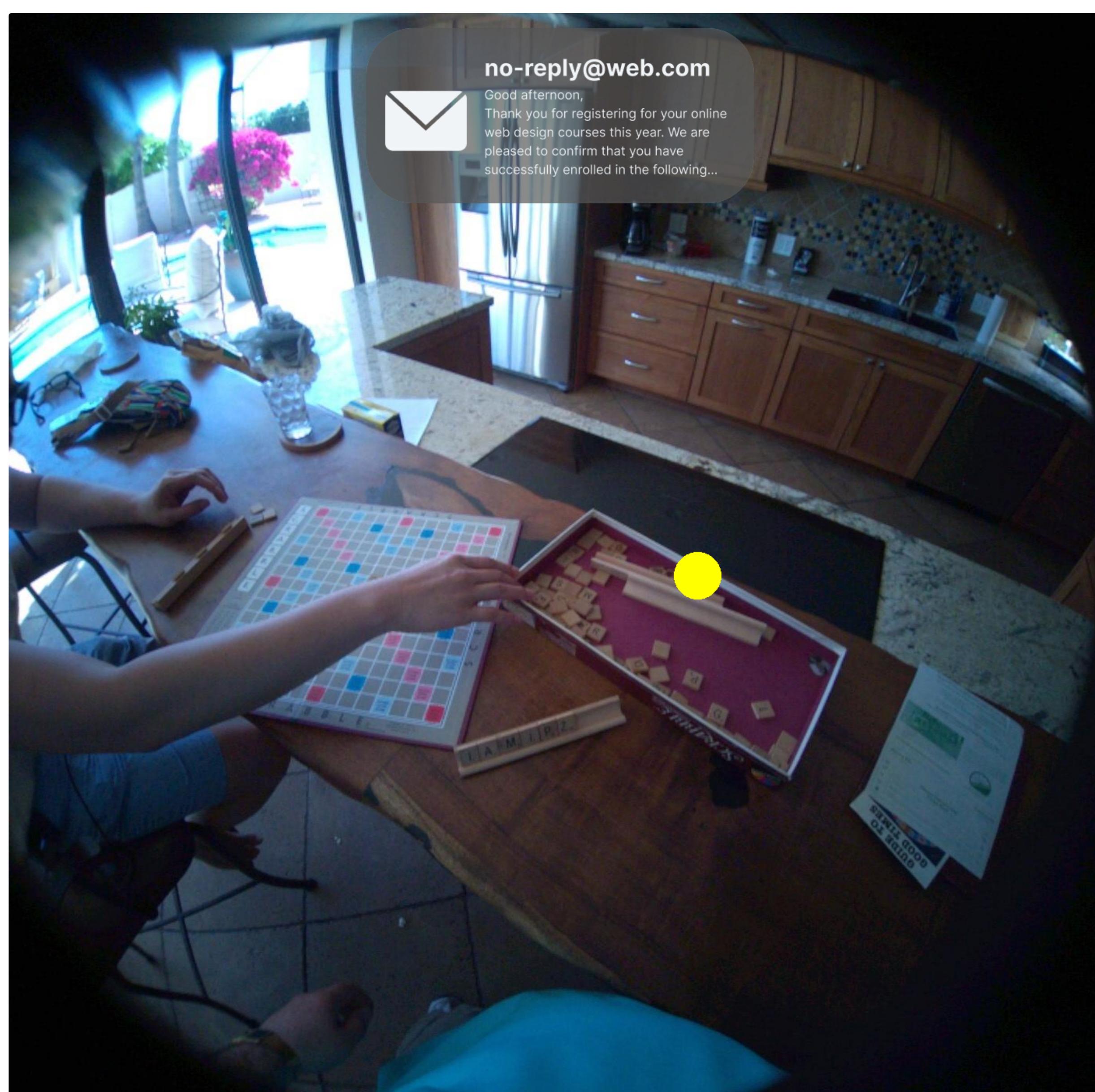
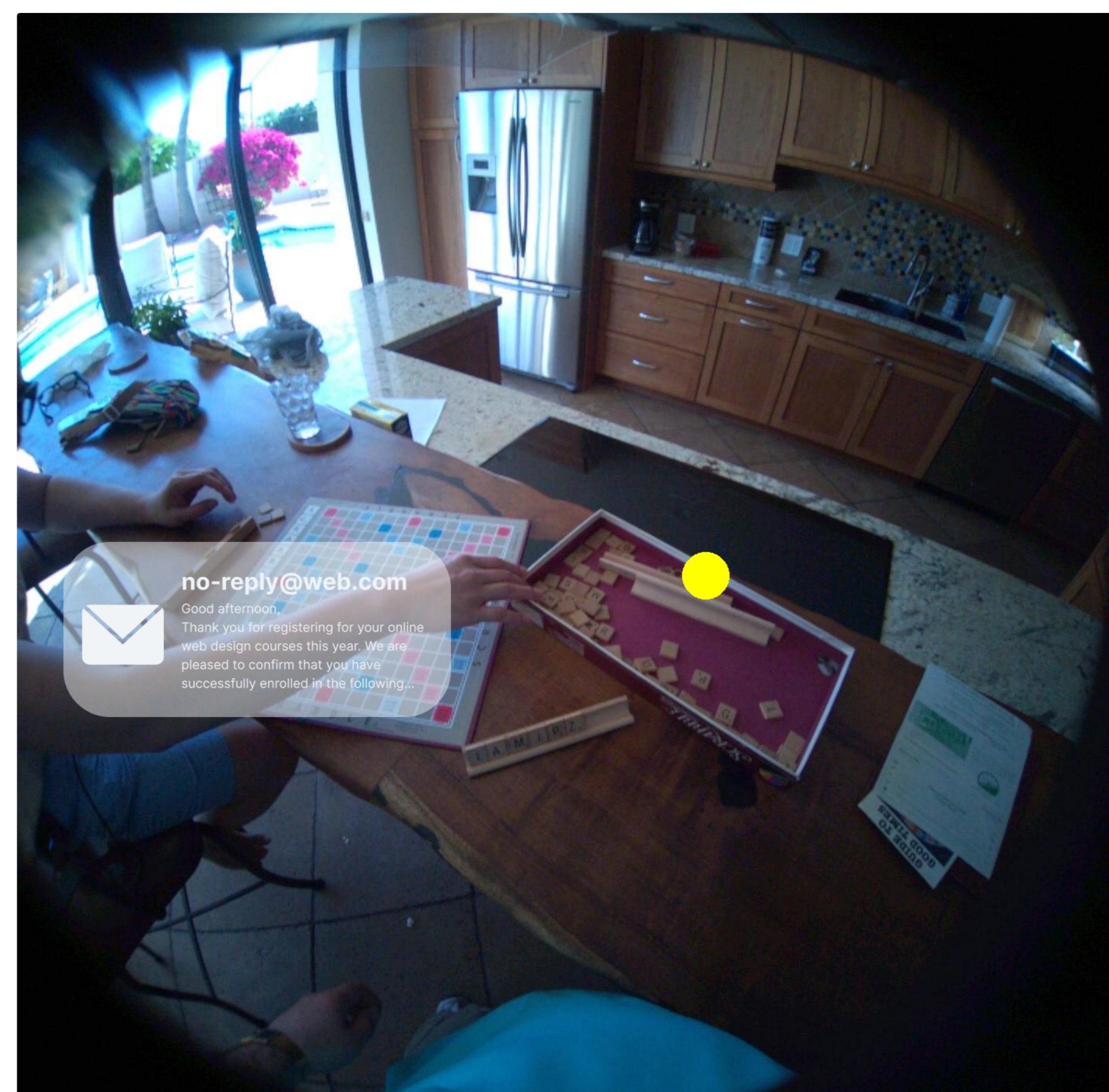
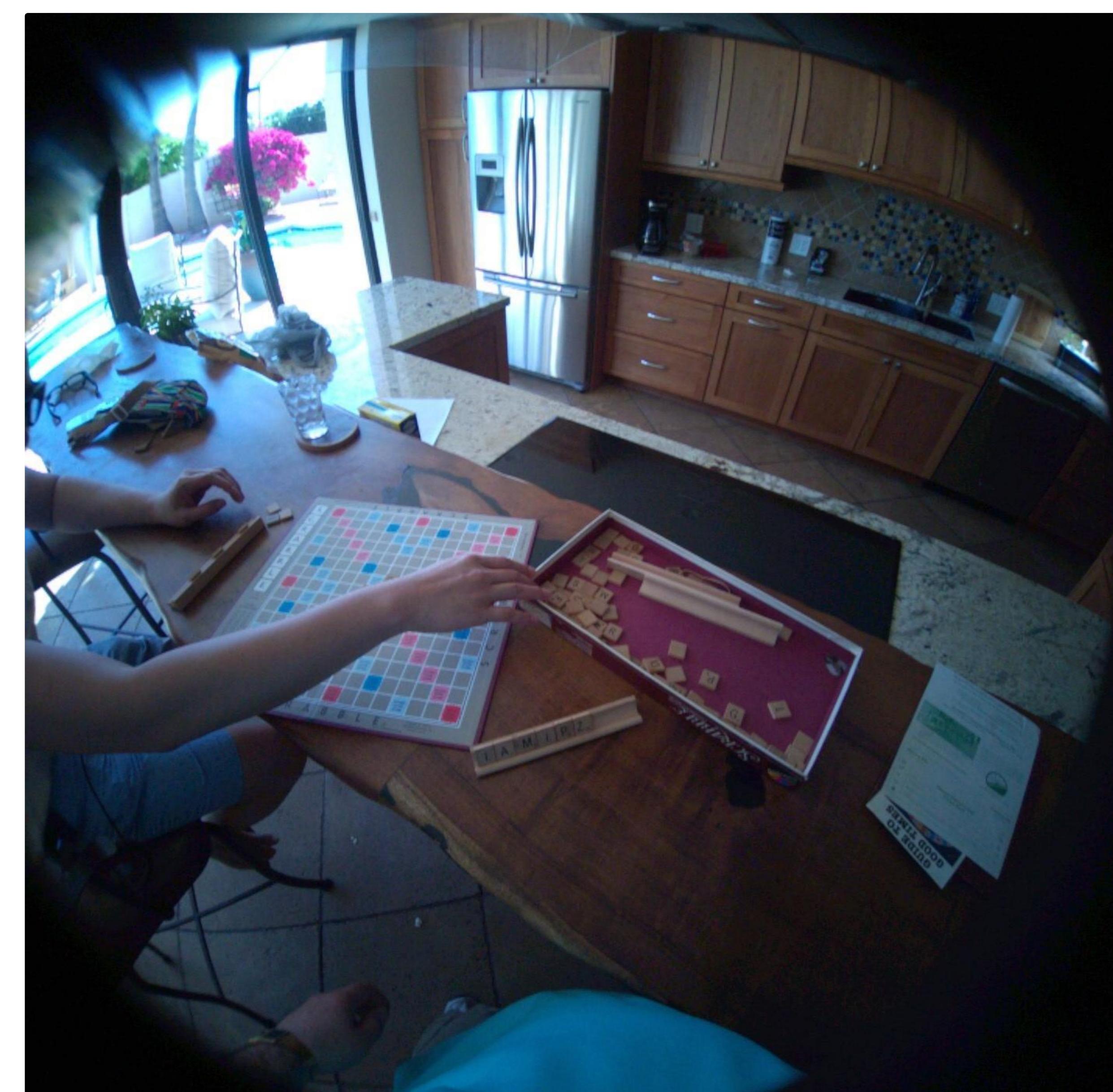
**UiVLA**

Vision-Language Action Model

Fine-tune VLM with UI adaptations



**User:** Compare these three identical views from a head-mounted camera. The first is the current user view. The second is one potential UI layout with a UI element overlaid, and the third is another potential UI layout with the element appearing in a different location. Which is the better layout, the first (1) or second (2)?



**UiVLA:** 2. The second layout is better because it does not block the table or the other person.

U