MEASURING PRODUCTIVITY

ARYAMAN KHANDELWAL, RISHI VIJAYWARGIYA, SANIDHYA SINGH

PROBLEM STATEMENT

Lack of Comprehensive Productivity Metrics: Traditional methods of measuring productivity use subjectivity and dependence on measure by external party which is manual and biased.

Need for Real-Time Interventions: Current systems fail to incorporate dynamic interventions to improve focus and retention during tasks, missing opportunities to enhance user performance and cognitive engagement.

Need for a Holistic Productivity Score: A multi-dimensional score that integrates attention, retention, task efficiency, and intervention response is required for accurate productivity measurement.

POTENTIAL APPLICATIONS/IMPACTS

Potential Applications

- **Workplace Productivity Tools:** Companies could implement a customised version of this solution to measure and improve employee productivity in real-time, allowing for more accurate assessments and personalized interventions.

- **Educational Platforms:** E-learning platforms could integrate this system to monitor student engagement and provide dynamic interventions to improve learning outcomes.

Potential Impacts

- **Enhanced User Engagement:** The ability to implement real-time interventions will help users maintain focus, improve retention, and optimize task execution, leading to higher engagement and performance.

- **Improved Accuracy in Productivity Measurement:** By offering a multi-dimensional, unbiased and objective measure of productivity, the solution will provide a more accurate reflection of user performance, addressing the current limitations of subjective assessments.

TIMECOURSE OF RECOVERY FROM TASK INTERRUPTION

ACT-R is used to model how interruptions affect cognitive performance. It helps simulate the time it takes to resume a complex task after being interrupted by defining memory activation and retrieval processes.

ACT-R predicts how quickly people recover by modeling the reactivation of "mental contexts" step-by-step, influenced by associative activation, which primes future steps and facilitates smoother task resumption.

The paper also delves in the calculation of the Response Time Metric which is crucial for developing the juxtaposition between a persistent task and a task with intermediate interruptions. The **Response Time Metric is calculated by subtracting the Immediate Response Time(IRT for 15-30 seconds) after the interruption and the average response time before the interruption.**

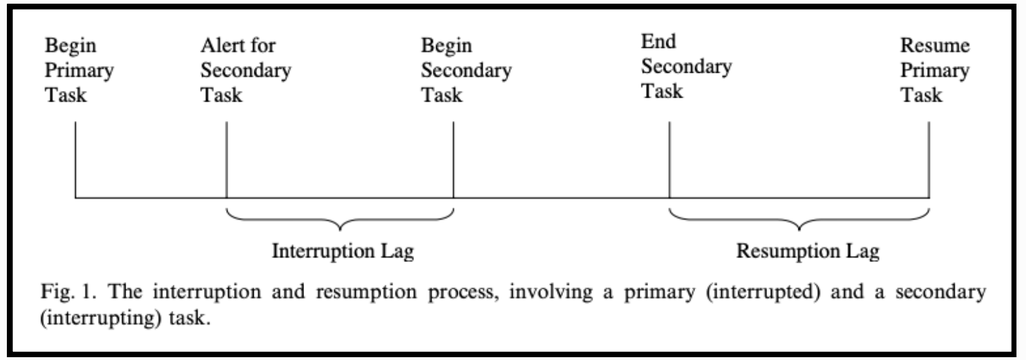


Fig. 1. The interruption and resumption process, involving a primary (interrupted) and a secondary (interrupting) task.

Base-Level Activation Formula

The formula to calculate the activation A_i of a chunk i is:

$$A_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right)$$

Where:

- A_i: The activation level of the memory chunk i.
- n: The number of times the chunk has been accessed (retrieved or used) in the past.
- t_j: The time since the j-th retrieval of the chunk. Essentially, it is the time elapsed between the current moment and when the chunk was last accessed.
- d: A **decay parameter**, typically between 0 and 1, which controls the rate at which memory decays over time.

Takeways

- *ACT-R Calculation for developing the retention metric*
- *Imperative to develop a archetypal environment contemporary to a real scenario of a workplace*

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c154540ad2074d95860a48e76afbdc296693b137

MEASURING ENGAGEMENT LEVEL OF TV VIEWERES

In the paper, the Y targets used for engagement were the categorical labels of viewer engagement: None, Low, Medium, and High. Facial features extracted included distances and angles between **facial points (e.g., lips, eyebrows), head roll, head size, and head position**.

The graphical aggregation methods used to analyze the temporal dynamics of these features included mean, median, standard deviation, minimum, maximum, range, and zero crossings to summarize changes over short time windows.

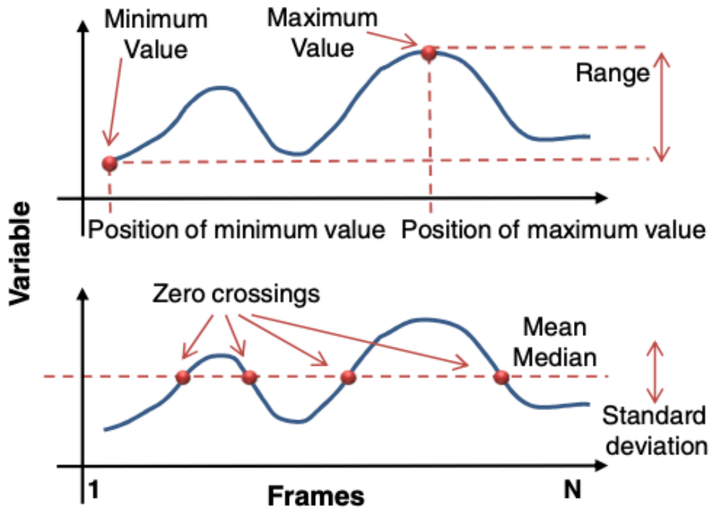


Figure 4. Graphical representation of temporal aggregation methods.

Takeways

- *Temporal Aggregation for Frame Reduction and Faster Annotation*
- *Visual variables for measuring key distances across the face*

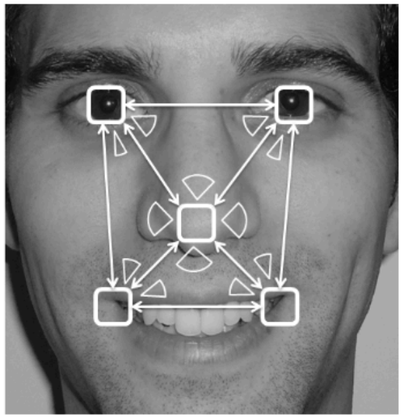
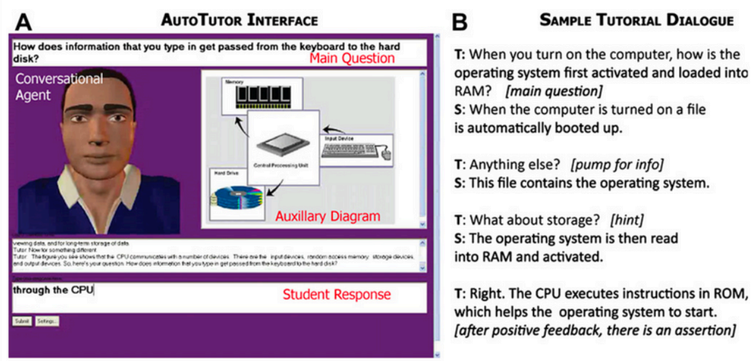


Figure 3. Facial points (squares), distances (lines), and angles (segments).

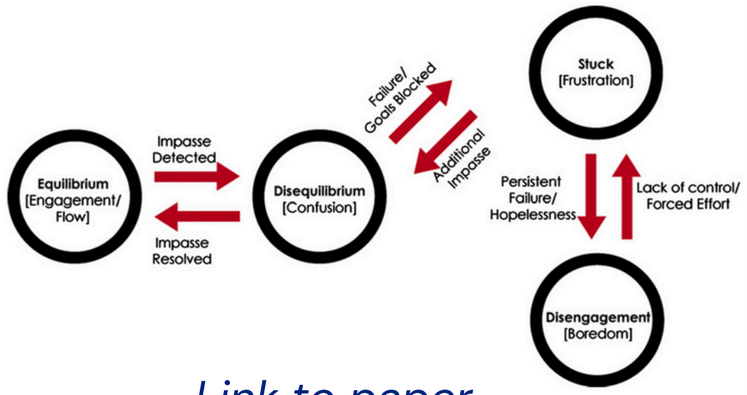
Link to paper

DYNAMICS OF AFFECTIVE STATES DURING COMPLEX LEARNING

- **Motivation:** How **affective states (confusion, frustration, boredom, engagement) influence learning** during complex tasks.
- **Dataset:**
  - Participants interacted with AutoTutor, and emotional states were recorded **110 times** during **32-35 minute sessions.**
  - **30 undergraduate students** (13 male and 17 female) from a mid-south university in the U.S
- **Methodology:**
  - **Retrospective affect judgment protocol.**
  - **Time series analysis** tracked transitions between emotions.
- **Results:**
  - Frequent transitions between confusion and engagement/flow.
  - Frustration arises when impasses are unresolved, leading to boredom.
- **Conclusion:**
  - **Cognitive disequilibrium (confusion) is essential for deep learning.**
  - Systems should challenge learners and support resolution of confusion to prevent frustration and disengagement.



Current State	*N*	Next State						
		Flow	Con	Fru	Bor	Del	Sur	Neu
Study 1								
Flow	373	–	**120**	61	72	22	21	72
Confusion	354	**116**	–	**70**	44	13	14	95
Frustration	245	62	62	–	51	7	9	49
Boredom	288	71	50	**49**	–	5	3	103
Delight	85	20	18	6	5	–	5	30
Surprise	78	22	18	9	7	6	–	15
Neutral	373	75	83	49	107	31	21	–



Link to paper

PREDICTING OFFICE WORKERS' PRODUCTIVITY

- **Motivation:** Predict **office workers' productivity** using **physiological, behavioral, and psychological indicators.**
- **Dataset:**
  - **48** participants, data from **wearable devices**, Kinect camera, and computer interactions.
  - **83 features** including heart rate, skin temperature, electrodermal activity, and facial expressions.
- **Methodology:**
  - Two-phase experiment (**low-stress, high-stress**).
  - Machine learning models (**XGBoost, random forest**) to predict productivity and psychological states.
- **Results:**
  - Extended model outperformed the baseline model.
  - **Key predictors: mood, eustress, skin temperature, electrodermal activity.**
  - **Wearable devices outperformed** workstation addons.
- **Conclusion:**
  - **Psychological states like mood and eustress are essential** for accurate productivity prediction.
  - **Wearable devices offer valuable insights** and are practical tools for assessing productivity in smart workstations.

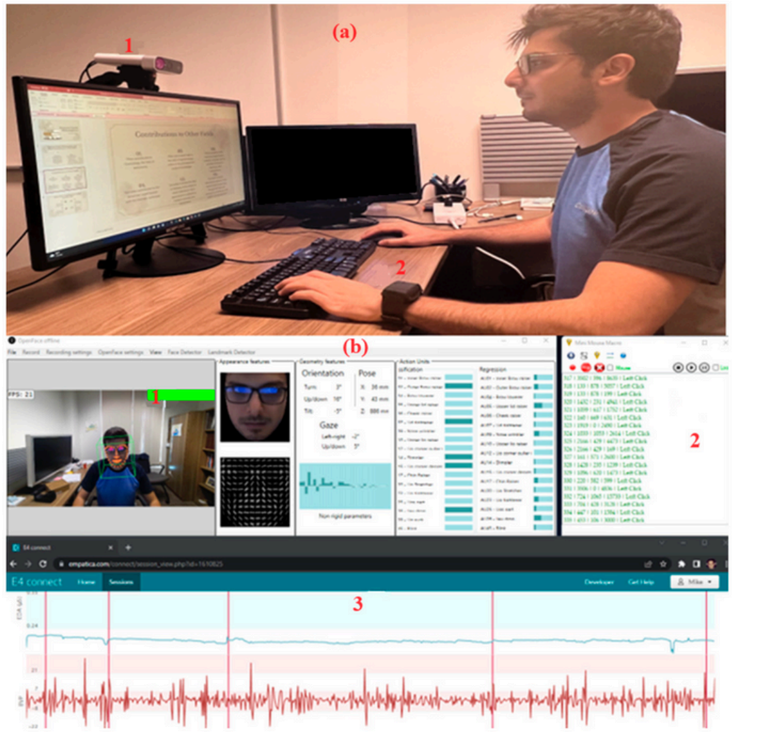


Figure 1. (**a**). Participant taking the experiment: (1) Kinect Camera and (2) Empatica E4 hand-wrist physiological sensor. (**b**). Data collection platform: (1) OpenFace application for facial feature extraction, (2) human–computer interaction application monitoring, and (3) physiological data monitoring.

Table 3. Performance comparison between the baseline and extended productivity models.

Algorithms	Productivity Regression Analysis			
	Baseline Model		Extended Model	
	R^2	MAE	R^2	MAE
Linear regression	0.25	21.59	0.27	16.42
Ridge regression	0.12	23.43	0.13	18.60
Lasso regression	0.10	24.16	0.15	22.60
Random forest	0.44	17.19	0.57	10.91
Gradient boosting	0.40	18.29	0.46	13.67
XGBoost	0.48	16.62	0.60	10.52

Link to paper

ERO-METRIC

ENGAGEMENT

Engagement in this context can be defined as the level of active involvement and interaction between a human and a computer, encompassing both

- ***Physical actions*** *(e.g., mouse clicks, key presses),*
- ***Facial activity for corresponding task****.*

Higher engagement indicates continuous input, capturing both the frequency and consistency of actions performed over time, providing insights into overall productivity.

Let $E(t)$ represent the **engagement score** over a time period t, calculated using physical actions and facial data:

$$E(t) = \quad M(t) + \quad K(t) + \quad F(t)$$

This metric reflects the user's ***mental and physical participation*** *in a productive process combined with temporal aggregation of facial features and activity.*

Engagement will be calculated within a set time window period(30 seconds) a reflection of physical activity(input received by the device) + facial aggregation data during that time.

RETENTION

*Retention is defined as the **user's ability to successfully resume** and continue a primary task after an intermediate interruption, with minimal loss of performance or cognitive capacity.*

***Resumption Lag (RL)**: The time taken to resume the primary task after an interruption.*

***Performance Level Change (ΔP\Delta PΔP)**: The difference in performance (e.g., speed, accuracy) before and after the interruption.*

In the scenario where a user is performing a task and gets interrupted by an intermediate task, retention involves:

- *How effectively the user can get back into the flow of the primary task.*
- *How quickly they can return to their previous performance level.*
- *How much cognitive effort is required to restore context and continue the task.*

$$R(t) = \cdot \frac{1}{RL(t)} + \cdot (1 - \Delta P(t)) -$$

Adaptive Control of Thought-Rational

OUTPUT EFFICIENCY

Task Output Efficiency (TOE) *is defined as the quality of output a user generates while performing tasks within a specific timeframe, particularly when the user experiences interruptions and resumes work.*

$$TOE(t) = A(t)$$

Measures the correctness or quality of the outputs generated by the user, represented as the proportion of correct answers or tasks completed correctly.

FINAL PRODUCTIVITY ESTIMATE

The Consolidated Productivity Score (CPS) represents an overall measure of the user's engagement, retention, and task output efficiency during a productivity scenario, accounting for each of these factors' relative importance.

$$CPS(t) = \quad E_{\text{norm}}(t) + \quad R_{\text{norm}}(t) + \quad TOE_{\text{norm}}(t)$$

For all the Three Tasks, we define a aggreation for all three estimates collected from these tasks independently to develop

- *Engagement - Unbiased Estimates*
- *Retention*
- *Output for that particular Main Task*

X

(Dependent Variables)

PRODUCITIVITY MEASURED(ESTIMATE) ACROSS

(Annotated Labels)

PRODUCTIVITY LEVEL(0-5 Range)

EXPERIMENT PROTOCOL - METRICS

In our experiment, we measured the following key metrics **effectively:**

- *Several key **facial distances** of the user during the experiment.*
- ***Mouse clicks** & **keyboard presses** throughout the experiment, with detailed logging of timestamps.*
- *Accuracy, time taken, error rate and other metrics varying across different tasks.*
- *Heartbeat data with second-level granularity (**omitted in the further calculations - major inconsitencies)***

This would help us compute the **engagement, retention & output** of the user.

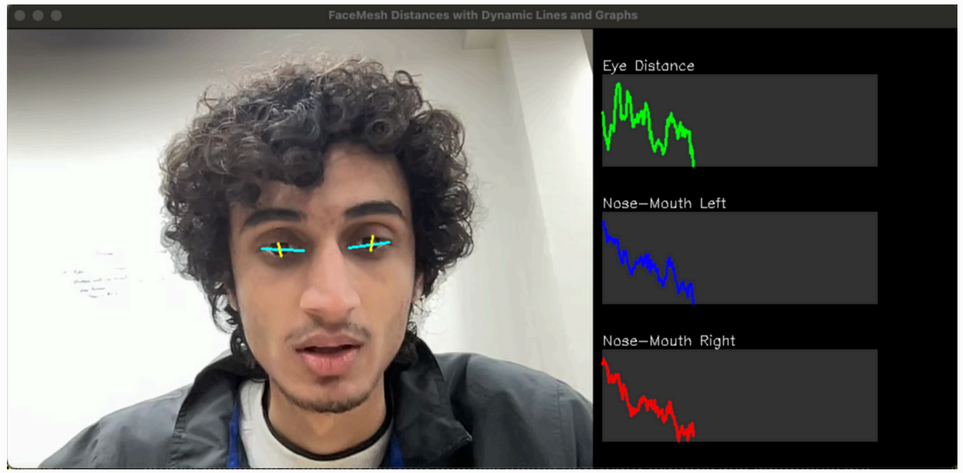
We will be using the following instruments:

- Primary Camera or laptop camera
- Apple Watch

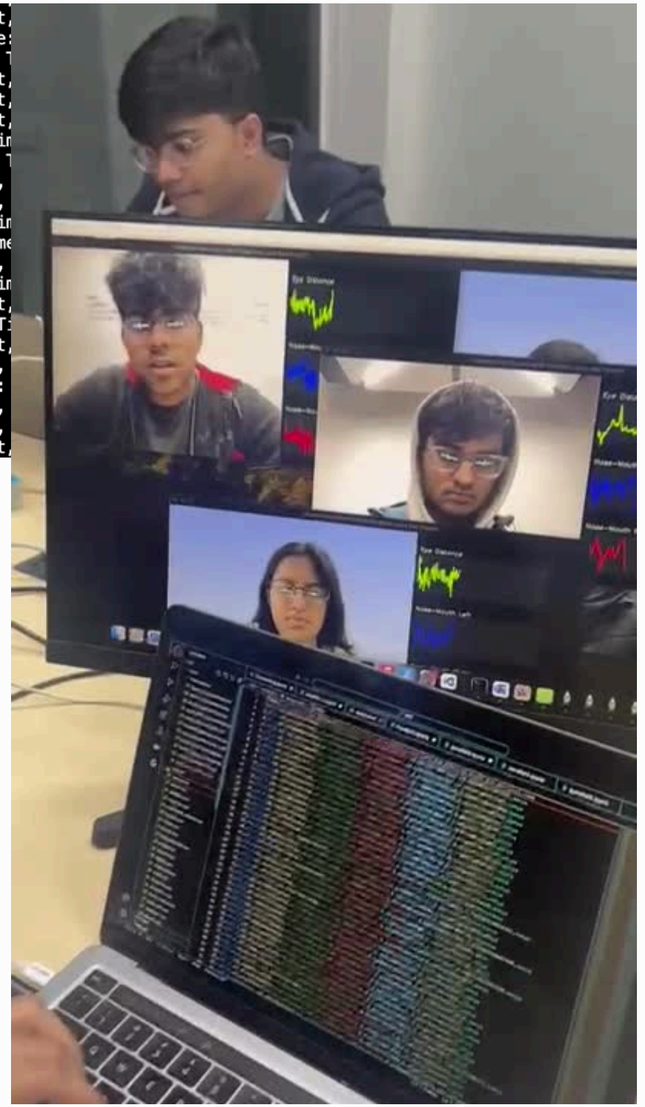
We collected data from **26 participants (24 usable)**, including male and female students from our batch and junior batches, in two rounds: **pre-intervention (treatment)** and **post-intervention (control)**.

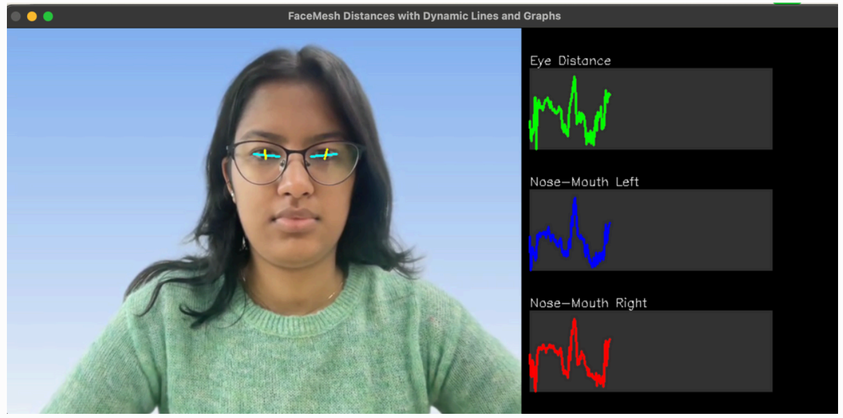
Note: All the trials for the experiment will be performed in the ***same controlled environment***.

EXPERIMENT PROTOCOL - METRICS

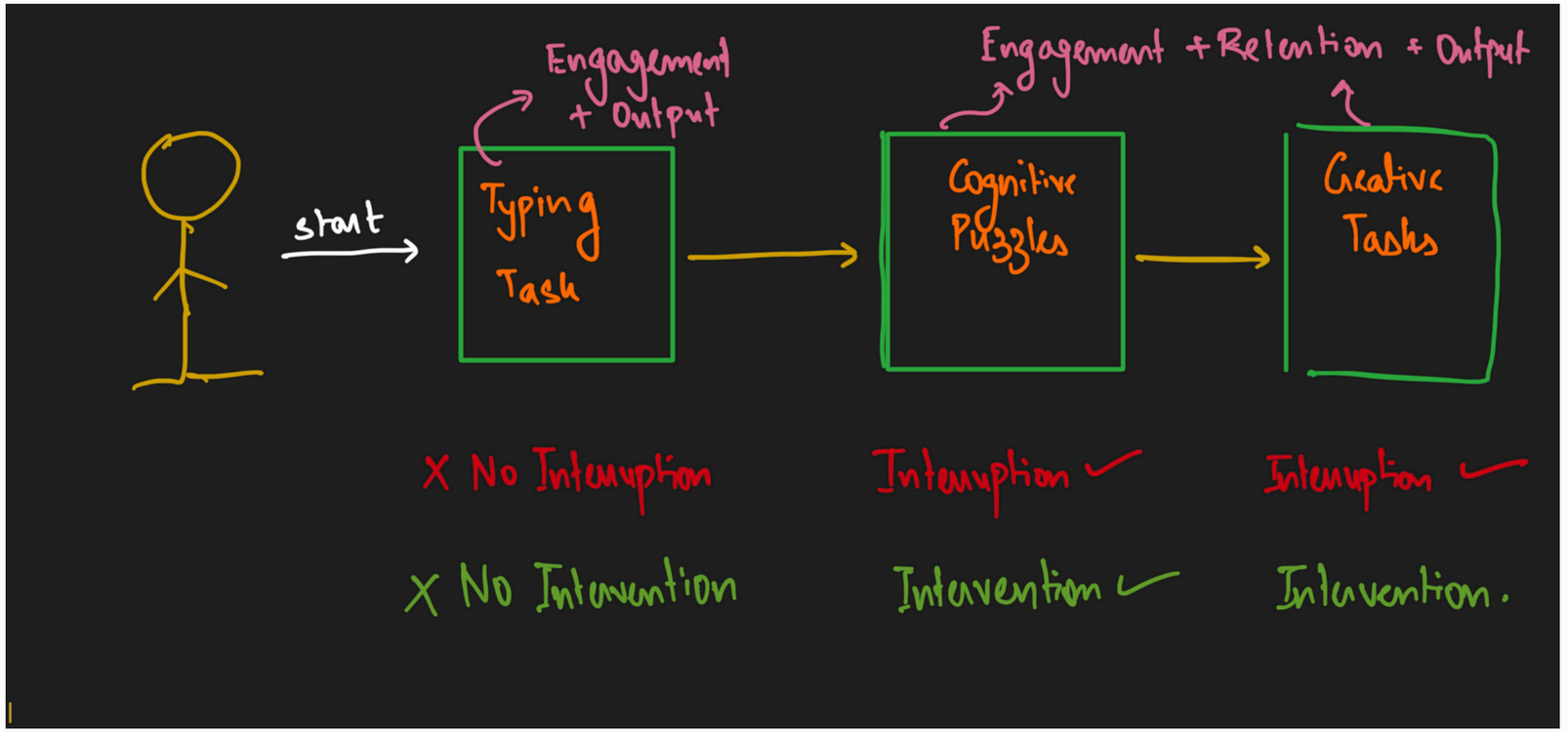



```
Recorded Mouse Click at (400.48828125, 508.02734375), Button: Button.left,
Recorded Mouse Click at (950.2265625, 463.375), Button: Button.left, Time:
Recorded Mouse Click at (953.83984375, 381.171875), Button: Button.left, T
Recorded Mouse Click at (777.86328125, 558.87109375), Button: Button.left,
Recorded Mouse Click at (704.73046875, 445.99609375), Button: Button.left,
Recorded Mouse Click at (506.55078125, 566.79296875), Button: Button.left,
Recorded Mouse Click at (446.84375, 80.26953125), Button: Button.left, Tim
Recorded Mouse Click at (784.953125, 657.59765625), Button: Button.left, T
Recorded Mouse Click at (442.42578125, 85.01171875), Button: Button.left,
Recorded Mouse Click at (701.22265625, 431.4765625), Button: Button.left,
Recorded Mouse Click at (718.1875, 541.42578125), Button: Button.left, Tim
Recorded Mouse Click at (78.296875, 71.4921875), Button: Button.left, Time
Recorded Mouse Click at (725.5703125, 145.01953125), Button: Button.left,
Recorded Mouse Click at (382.5859375, 67.078125), Button: Button.left, Tim
Recorded Mouse Click at (1396.8046875, 763.00390625), Button: Button.left,
Recorded Mouse Click at (856.203125, 517.5234375), Button: Button.left, Ti
Recorded Mouse Click at (1046.41015625, 82.54296875), Button: Button.left,
Recorded Mouse Click at (1060.921875, 869.58984375), Button: Button.left,
Recorded Mouse Click at (74.8671875, 365.875), Button: Button.left, Time:
Recorded Mouse Click at (542.80859375, 387.8828125), Button: Button.left,
Recorded Mouse Click at (991.41015625, 437.6484375), Button: Button.left,
Recorded Mouse Click at (819.54296875, 458.13671875), Button: Button.left,
```





EXPERIMENT PROTOCOL- FLOW



Experiment Total Time: 30 mins (upper bound)

INTERVENTION IN FORM OF INTERRUPTION

Objective:

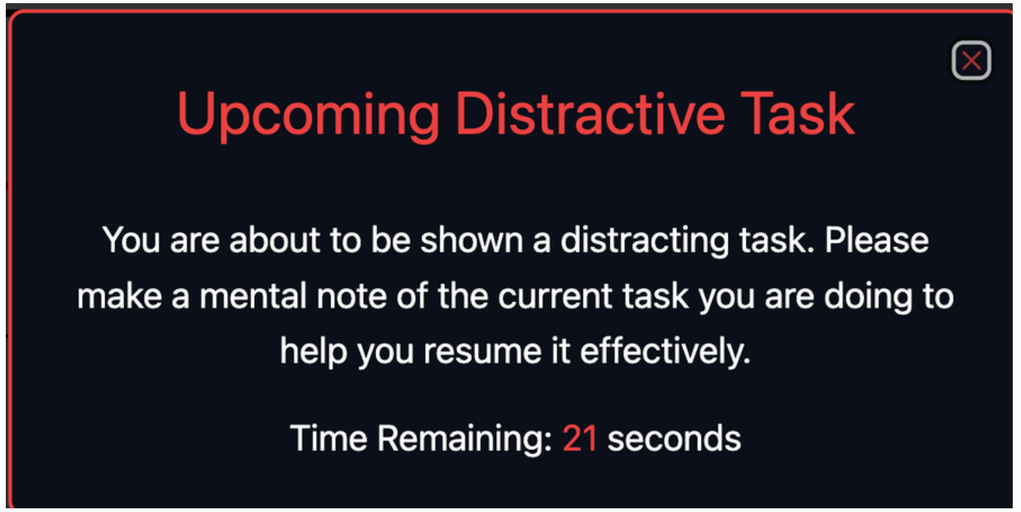
We want to construct a **differentiating factor** between groups of individual based on their control and treatment performance.

These groups allow for measuring the key difference in performance, which in this case are :

- Engagement Metric
- Retention Metric
- Output Metric

Informed vs Immediate Interruption:

- Keeping the existence of the interruption common across tasks, the difference arises in form of **consolidated thought retrieval** against **immediate retrieval**.
- Foucsing on **active recall**



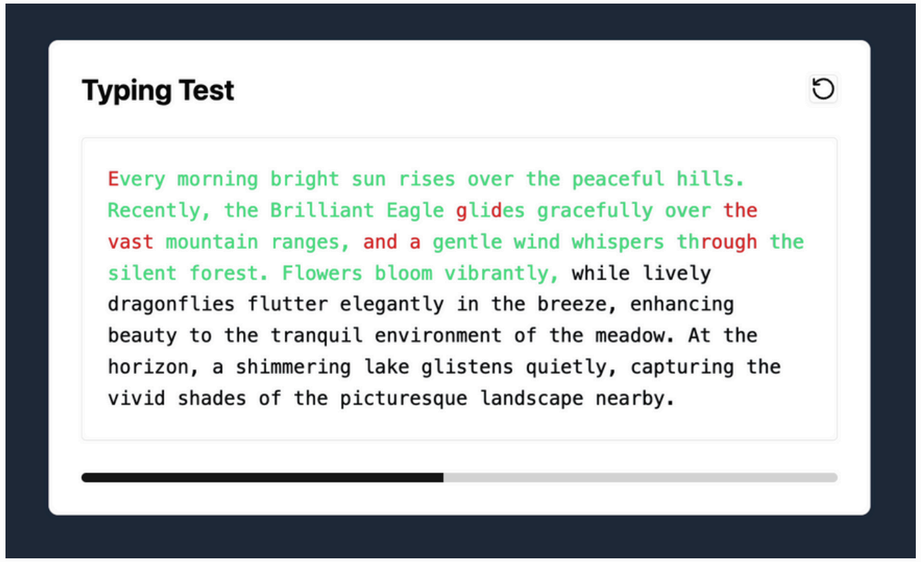
Intervention

What Metrics to Distinguish Between?

- We observe **analytical difference** between performance of users across the different tasks, irrespective of kind of task.

- Certain tasks are intentionally crafted to introduce interventional pauses, allowing users time for reflection and anticipation. These controlled interruptions provide valuable insights into user engagement and cognitive states.

TYPING - TASK 1



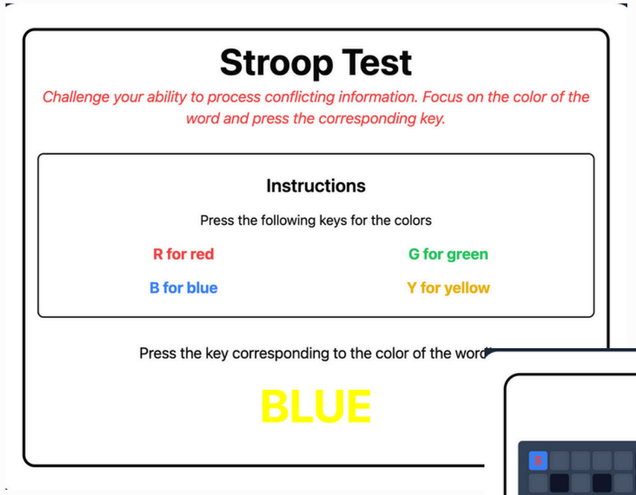
- Engagement & output metrics are measured in this sub-task
- We measured accuracy, time-taken average per word, error rate.
- The frequency of typing the next word will be representative of **intrinsic potential** of the individual.
- There was no **intervention** in this task.

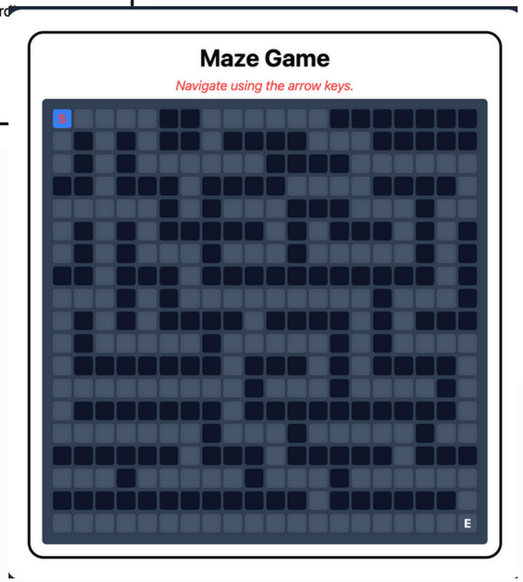
Implications

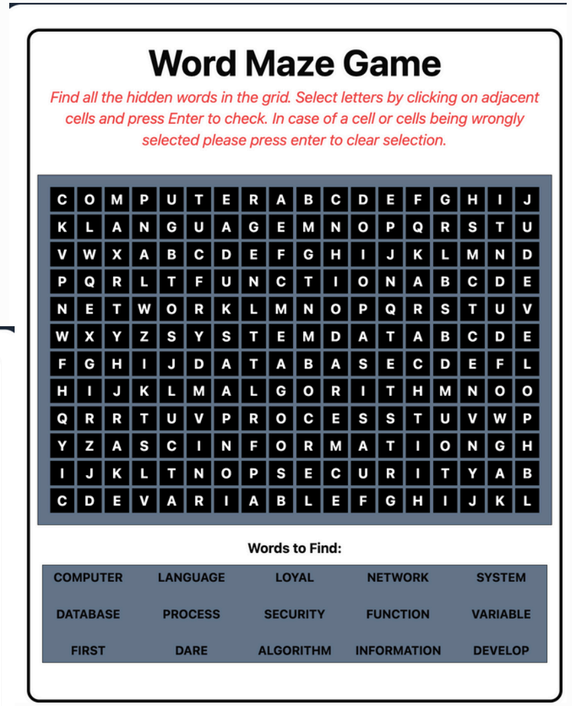
- This sets the users activity to a **baseline**, pointing at the general interest of the individual all-together
- Slow typers interest factor is accommodated by the accuracy achieved.
- A combination of accuracy and speed develops the metric for engagement out of the typing test.

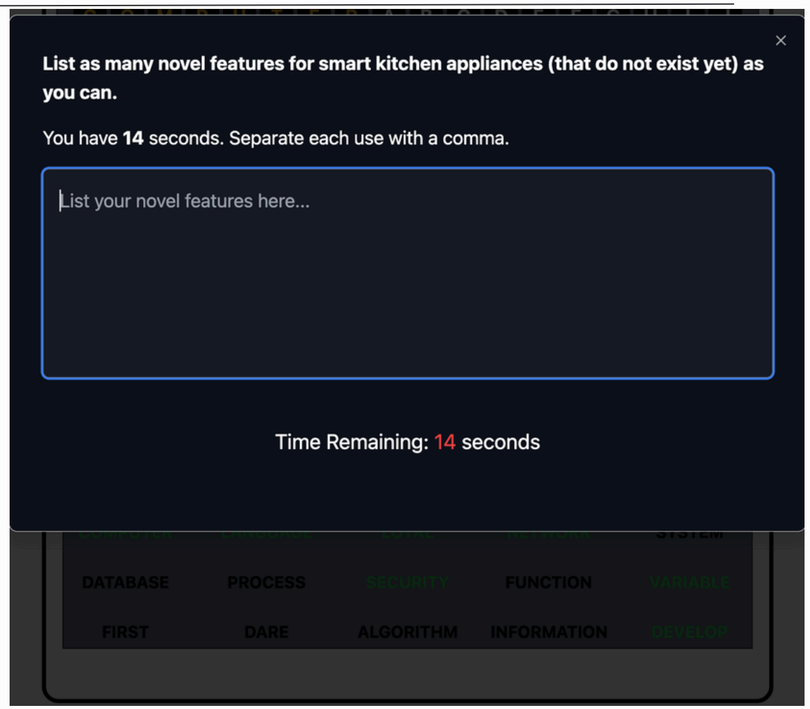
Total Task Time: 5 mins

COGNITIVE PUZZLES - TASK 2









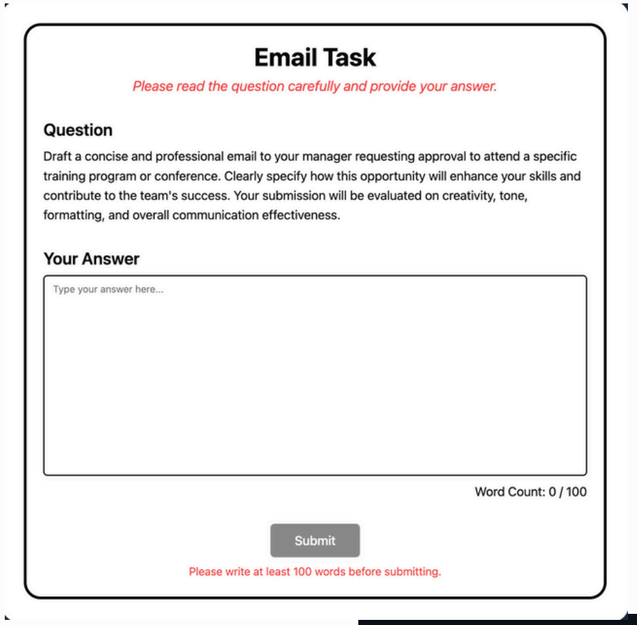
- In this task, the user is expected to solve **cognitive puzzles** like the **maze problem**, **stroop test** and the **word maze.**
- Key Metrics such as **retention**, **output**, and **engagement** will be measured.

Interruption

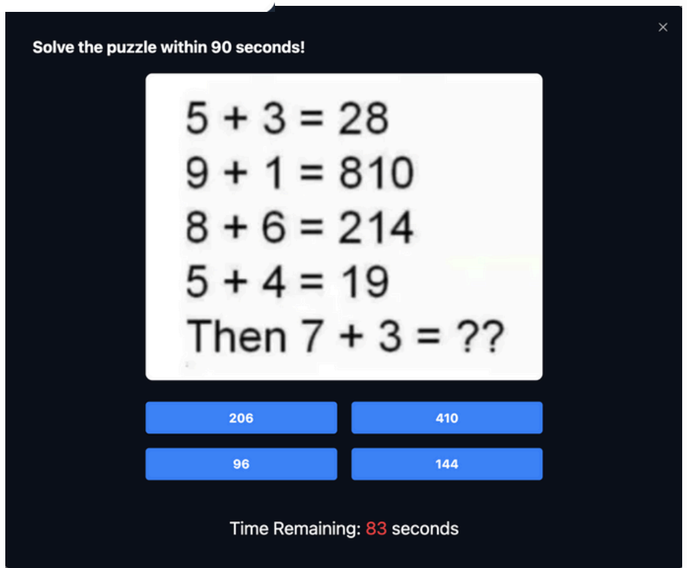
- While solving a puzzle, user will be **interrupted** with another task that will not asses the **user's cognitive abilities.**
- The interrupted task was a **creative** task, asking participants to list as many unusual uses for an inanimate object, taking their field of thought away from the previous task in hand.

Total Task Time: 10 mins

CREATIVE THINKING - TASK 3



- In this task, the user is expected to solve a **creative task.** The task can be, like writing a short mail to your boss for a leave application.
- This task will judge the user's creativity (can be judged via a LLM agent), retention power, and engagement.
- There is **intervention** in this task.

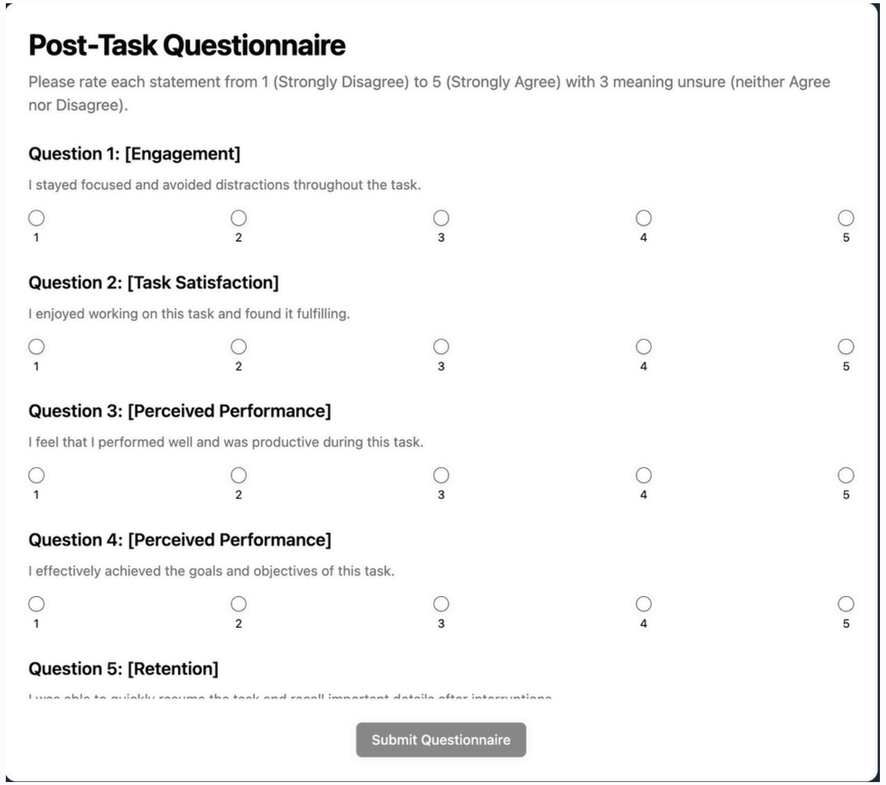


Interruption

- While solving the creative task, user will be **interrupted** with another task that will not asses the user's creative abilities.
- The interrupted task will be a **logical** task, that is solving easy arithmetic problems, pattern matching problems for a **minute.**
- The timer for the main task will be paused once the user receives a pop up for the interrupted task, and the time will resume once the interrupted task has been completed.

Total Task Time: 10 mins

POST -TASK QUESTIONNAIRE (SUBJECTIVE)



Post-Task Questionnaire

Please rate each statement from 1 (Strongly Disagree) to 5 (Strongly Agree) with 3 meaning unsure (neither Agree nor Disagree).

Question 1: [Engagement]

I stayed focused and avoided distractions throughout the task.

1 2 3 4 5

Question 2: [Task Satisfaction]

I enjoyed working on this task and found it fulfilling.

1 2 3 4 5

Question 3: [Perceived Performance]

I feel that I performed well and was productive during this task.

1 2 3 4 5

Question 4: [Perceived Performance]

I effectively achieved the goals and objectives of this task.

1 2 3 4 5

Question 5: [Retention]

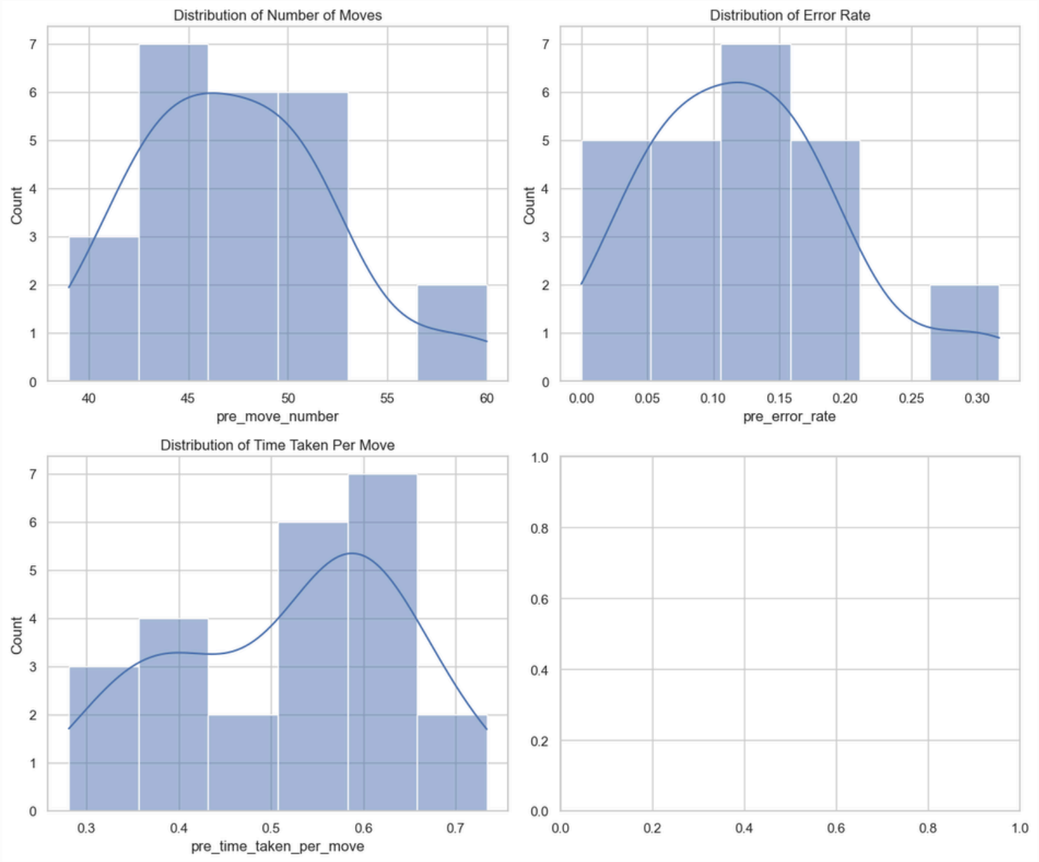
Submit Questionnaire

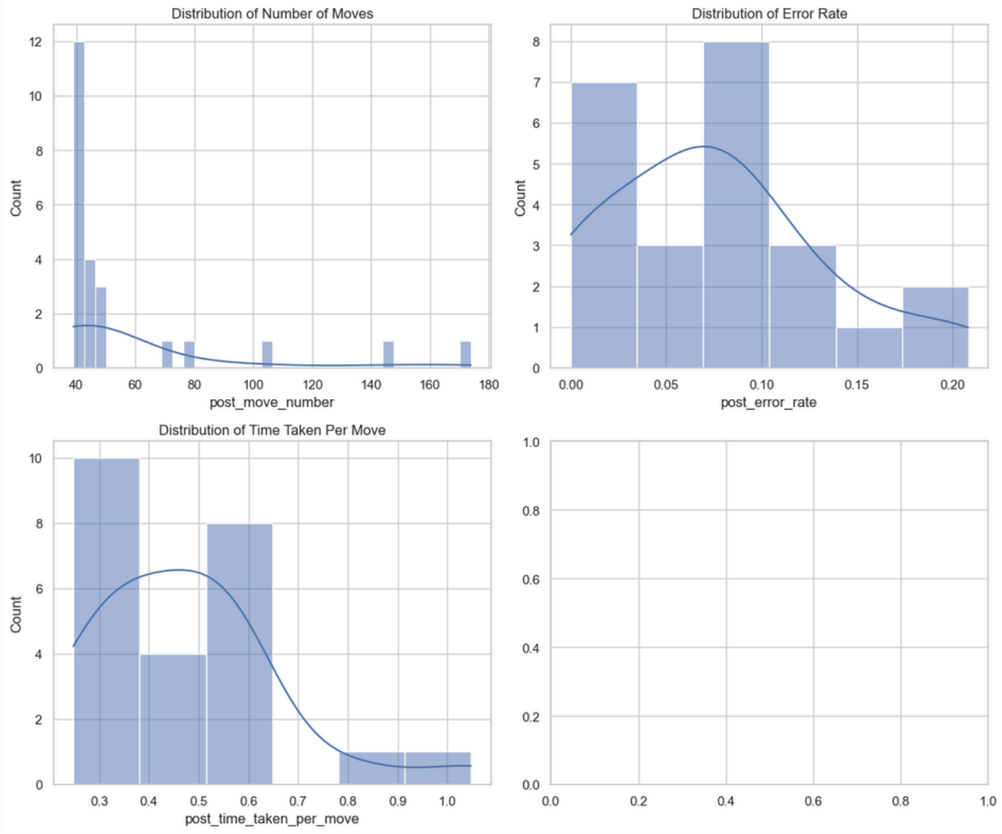
- After the **end of experiment**, the user will be prompted to complete a questionnaire.
- Responses will be collected using a Likert's scale.
- The questionnaire will involve questions focusing on the following areas:
  - **Engagement** during the task.
  - **Retention** (if applicable) during the task.
  - **Task enjoyment** and **satisfaction**.
  - **Perceived performance** and effectiveness.
  - **Task difficulty** and challenges encountered.

- These responses will provide valuable insights into user experience and areas for potential improvement.

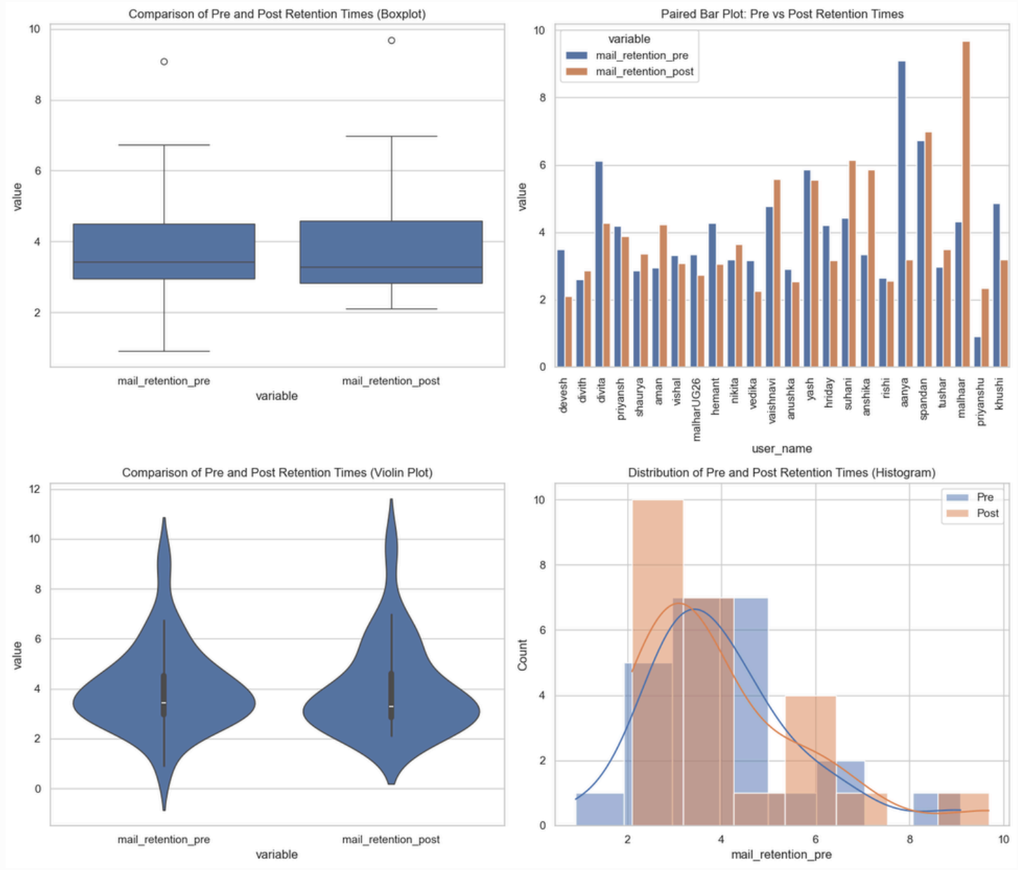
Total Task Time: 1–2 mins

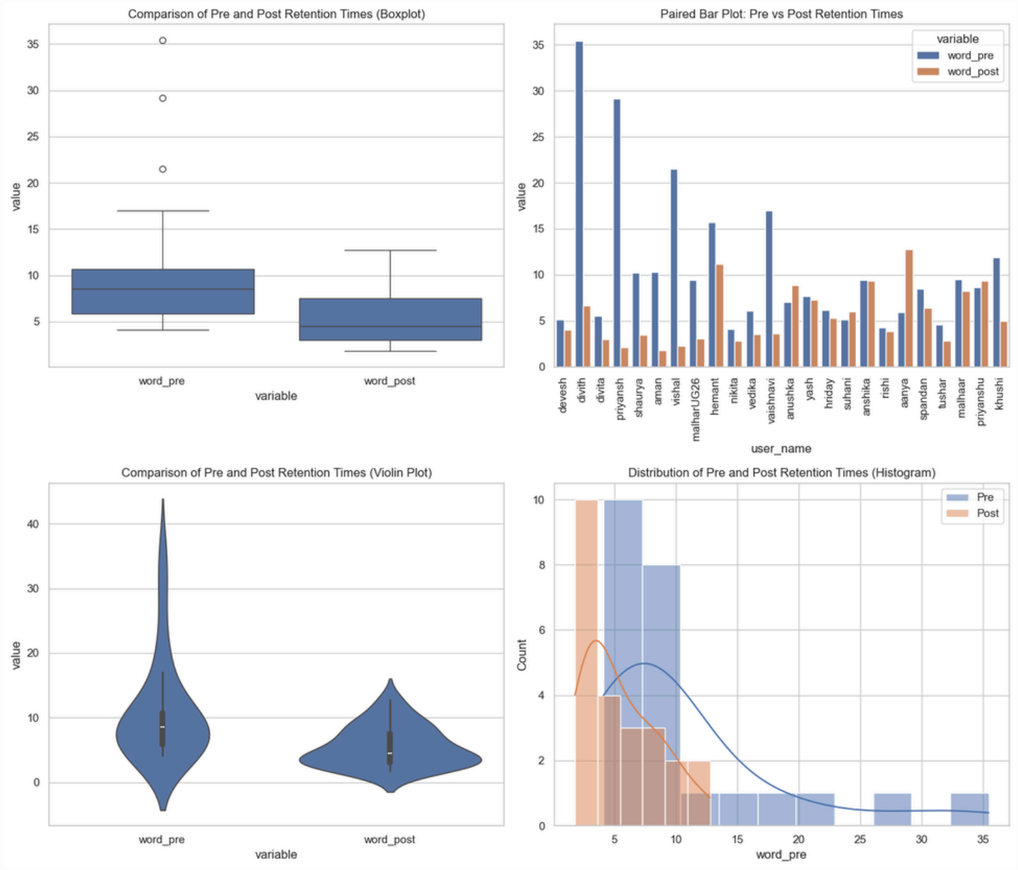
USER ENGAGEMENTS





RETENTION - FLUCTUATIONS





Mail Task

Word Search Task

DATA PRE-PROCESSING

- **Filtering** and **aggregating** productivity scores from the questionnaire.
- **Building LLM pipelines** to evaluate responses in the email task.
- Extracting FACS and calculating **key facial distances**.
- Uniquely **renaming task features** to avoid duplicates and ensure uniqueness.
- Conversion of timestamps to datetime for sorting and difference calculations.
- **Grouping** data by users or sessions to aggregate metrics.
- Calculating differences in time to derive time intervals between events.
- **Filling missing values** with zeros to maintain consistency in feature set.

FEATURE EXTRACTION

- **Path Features:**
  - Number of unique paths taken, Number of wrong moves, Average time per move
- **Mouse Events Features:**
  - Total number of clicks, Average interval between clicks , Variance of mouse x-coordinates, Variance of mouse y-coordinates
- **Productivity Score:**
  - Average productivity score
- **Stroop Test Features:**
  - Average time taken, Accuracy, Average resumption time
- **Typing Test Features:**
  - Average words per minute, Average typing accuracy, Average total errors , Average time taken for typing test
- **Word Search Features:**
  - Total number of words found, Average time taken to find words, Average retention time
- **Facial Features:**
  - Eye Distance Mean, Eye Distance STD, Nouse Mouth Distances, Eye Vertical Positions

ML MODELS

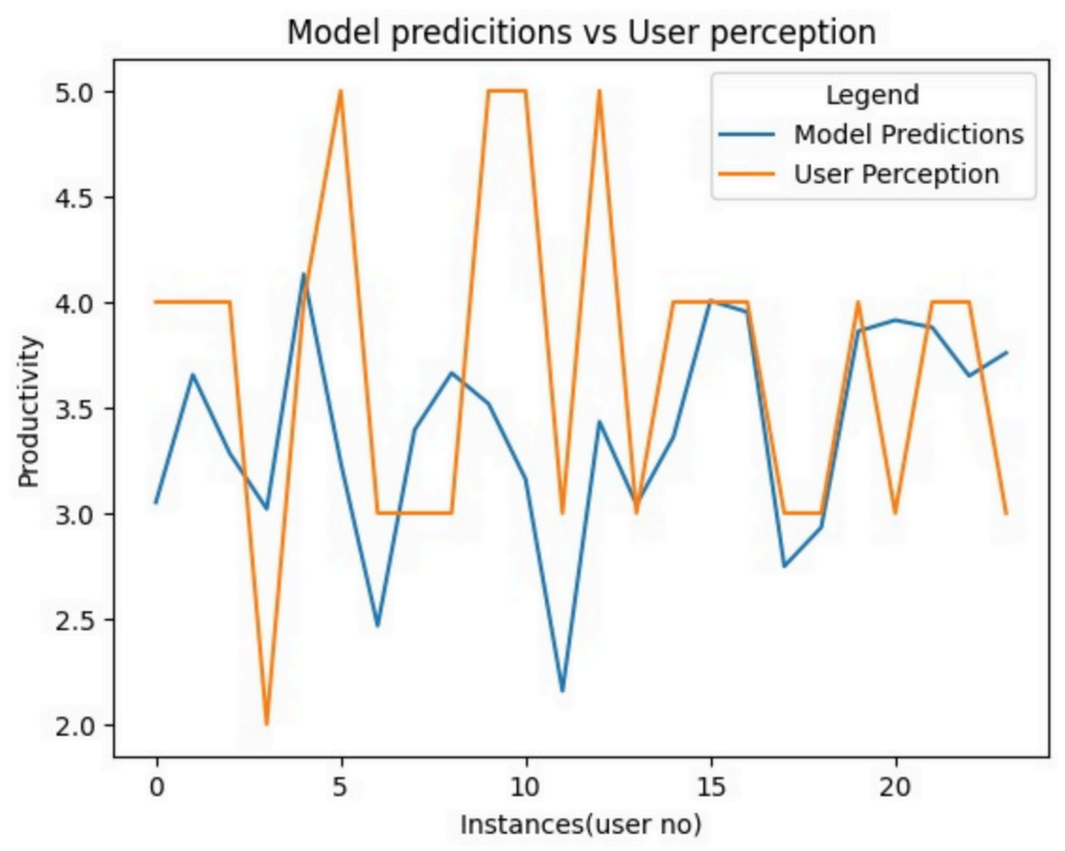
- **Input: Aggregation** of the features in the previous slide.

- **ML Model:**
  - **StandardScaler:** Used for scaling the features, ensuring that the model does not bias towards variables on a larger scale.
  - **SelectKBest with f_regression:** Selects features based on the f-regression statistical test, which is particularly useful for regression problems.
  - **XGBRegressor:** Final estimator in the pipeline, responsible for learning the relationship between the scaled, selected features and the target productivity score.

- **Hyper-parameter Tuning and Model Selection**
  - **GridSearchCV:** Used for hyper-parameter tuning, optimising the **XGBRegressor parameters** such as n_estimators, max_depth, and learning_rate based on the best negative mean absolute error obtained from cross-validation.

- **Metrics Used to Evaluate the Model:**
  - **Mean Absolute Error (MAE) & Root Mean Square Error (RMSE)**

OUTPUTS:




```
Validation MAE: 0.49183039665222167
Validation RMSE: 0.675279007438166
```

```
Test MAE: 0.649202436208725
Test RMSE: 0.8505624519782454
```

SOURCING HARDWARE AND SOFTWARE

- Use laptop webcam as primary camera feed
- Apple Watch series 7 available (if needed)
- Online Drive for storing data

CHALLENGES

- Unable to get HRV data in desired form and granularity.
- Very small dataset
- Only single intervention is done
- Accurate syncing of video, task, HRV
- Webm format for troublesome
- Collecting data was a hassle

DEPLOYABILITY

Whenever a user receives a notification from WhatsApp, the platform will alert them that they've received a distracting notification and remind them to mentally note the task they are currently engaged in.

Challenge while scaling: Incorporating **personalisation** for every user will be quite critical and challenging.

OVERALL IMPACT

- **Holistic Measurement:** Utilising the ERO metric provides a more comprehensive and objective measure of productivity by evaluating Engagement, Retention, and Output, rather than relying on a single aspect.
- **Targeted Interventions**: The multi-dimensional data allows for more precise identification of areas needing improvement, enabling organisations to implement specific strategies that enhance productivity effectively.
- **Improved Employee Satisfaction:** By focusing on engagement and retention, the approach supports a healthier, more motivating work environment, which can increase employee satisfaction and productivity.
- **Broader Industry Application:** This comprehensive approach can set a precedent for other organisations, prompting a shift towards more effective productivity measurement methods across industries.
- **Global Workforce Impact:** By promoting a more engaged and satisfied workforce, this methodology can contribute to a happier, more productive global workforce.

THANK YOU