Appendix

A. Generation of Ads using Henry-SEED

Henry-SEED takes as input a prompt consisting of the title of the ad to be generated and the brand name and generates the scene-by-scene verbalization of cognitive and visual concepts and voiceover. Listings 1-10 gives the input and output of a few randomly chosen examples of advertisements generated by Henry-SEED are shown below. Using the output script generated by Henry-SEED, we next prompt Adobe Firefly [1] with Henry-SEED's output to generate the visualization. Next, OCR from the Henry-SEED generated verbalization is put on top of the generated images manually. We provide visualizations for easier understanding (Figs. 3-6), along with the raw generations (Listings 1-10).

Listing 1. Advertisement generated by Henry-SEED for the Costco Ad. Original Ad: https://www.youtube.com/watch?v=uT72lJhUUS0

```
Input Title: "Maytag Overnight Wash and Dry" for the brand Costco


Generated Ad: "The ad is shot in landscape orientation, at a medium pace.

Scene 1: The scene shows a person standing in front of a washer and dryer
Emotions: comforting, comfortable
Tags: appliance, laundry, laundry room, man
Voiceover: "Maytag has taken the laundry room by storm with their overnight
      wash and dry."
Colors: Gray, Off_White, Silver, White, Black, Dark_Gray

Scene 2: The scene shows a person lying on a bed pressing a button on a cell
      phone
Emotions: gift, birthday, holiday
Tags: button, catch, smartphone, hand
Voiceover: "Imagine going to bed with dirty clothes and waking up to a fresh,
      clean load. All without leaving the house. It's like having your own
      personal laundry assistant."
Colors: Gray, Silver, Black, Dark_Gray, Dark_Brown

Scene 3: The scene shows a man and woman embracing while a washer and dryer are
       in the background
Emotions: love, happiness, togetherness
Tags: couple, hug, laundry, laundry room
Voiceover: "Never fight over the laundry room again."
Colors: Gray, Off_White, Silver, White,Black, Dark_Gray

Scene 4: The scene shows a logo of Maytag
Emotions: trust, confidence
Tags: logo, Maytag
Voiceover: "Get a Maytag and enjoy the luxury of time."
Colors: Off_White, Silver, White, Black, Dark_Gray, Gray
"
```

Listing 2. Advertisement generated by Henry-SEED for the Pruduential Financial Ad. Original Ad: https://www.youtube.com/watch?v=JPfi0P-0nwU

```
Input Title:  "Financial Professional Tips: Planning for Your Own Care" for
      the brand Prudential Financial

Generated Ad: "The ad is shot in landscape orientation, at a low pace.

Scene 1: The scene shows a woman looking thoughtfully at a laptop on her desk
Emotions: concern and contemplation
Tags: computer, laptop, office, woman
Colors: Gray, Dark_Blue

Scene 2: The scene shows the same woman now with a slight frown as she examines
       papers that appear to be medical bills
Emotions: worry and realization
Tags: bills, healthcare, paperwork, desk
Voiceover: "Long-Term Care is one of the most important issues in financial
      planning, but it's not always easy to talk about."
Colors: Gray, Dark_Blue

Scene 3: The scene shows the woman engaging in a conversation with a financial
      professional
Emotions: engagement and interest
Tags: consultation, financial advisor, discussion, office
Voiceover: "As a Financial Professional, I can help you evaluate your options
      and determine how much you may need to set aside each month."
Colors: Olive, Gray

Scene 4: The scene shows the woman in deep thought, looking out of a window
      while holding a pen and notepad, suggesting planning and reflection
Emotions: contemplation and seriousness
Tags: planning, reflection, window, notepad
Colors: Sky_Blue, Light_Gray

Scene 5: The scene transitions back to show the woman nodding in understanding
      as she listens to the financial professional, visibly gaining confidence
Emotions: comprehension and confidence
Tags: financial professional, office, learning, agreement
Voiceover: "Most people believe that they can rely on Medicare to pay for their
       long-term care expenses. But the truth is, Medicare typically only
      covers about 65% of the cost of a nursing home stay and only about 32%
      of the cost of an assisted living stay."
Colors: Gray, Dark_Blue

Scene 6: The scene shows the woman now jotting down notes actively, with a look
       of determination on her face
Emotions: determination and proactivity
Tags: planning, notetaking, strategy, office
Voiceover: "So if you're a typical American, you'll need about $100,000 to
      cover the remaining costs of long-term care."
Colors: Gray, Dark_Blue

Scene 7: The scene shifts to the woman shaking hands with the financial
      professional, symbolizing a plan set in motion
Emotions: satisfaction and trust
Tags: handshake, agreement, financial planning, commitment
Colors: Gray, Dark_Blue

Scene 8: The scene cuts to the woman now at home, relaxed and smiling, speaking
       on the phone, perhaps sharing her relief and positive steps taken with
      a loved one
Emotions: relief and happiness
Tags: home, communication, personal life, reassurance
Voiceover: "Let's talk about a plan for your own care."
Colors: Cream, Soft_Green

Scene 9: The final scene zooms out to a serene image of the woman walking in a
      park, symbolizing peace of mind about her future
Emotions: contentment and peace
Tags: park, walking, future, peace of mind"
Colors: Sky_Blue, Green
```

Listing 3. Advertisement generated by Henry-SEED for an Adidas Ad. Original Ad: https://www.youtube.com/watch?v=fujboh-W7Sk

```
Input Title:  "I'M POSSIBLE | IMPOSSIBLE IS NOTHING | Adidas" for the brand
      Adidas


Generated Ad: "The ad is shot in landscape orientation, at a medium pace.

Scene 1: The scene shows a man facing a rock climbing wall, poised to start his
       ascent
Emotions: anticipation and focus
Tags: athletic, climb, climber, rock
Voiceover: "What is possible is up to you."
Colors: Gray, Dark_Brown

Scene 2: The scene shows a man as he begins to climb, gripping the rock tightly
Emotions: determination and strain
Tags: athletic, climb, climber, hand
Voiceover: "Don't let anyone tell you otherwise."
Colors: Dark_Gray, Dark_Brown

Scene 3: The scene shows the man midway up the rock wall, reaching for a
      distant handhold
Emotions: perseverance and challenge
Tags: athletic, climb, climber, hand
Colors: Black, Dark_Gray

Scene 4: The scene captures a moment where the man slips but regains his grip
      on the rock face, muscles tensing
Emotions: tension and danger
Tags: athletic, climb, climber, hand
Colors: Dark_Gray, Dark_Brown

Scene 5: The scene shows the man pausing to chalk his hands, his breath visible
       in the cool air
Emotions: calm and preparation
Tags: athletic, chalk, climb, climber
Colors: Gray, Dark_Brown

Scene 6: The scene shows the man navigating a particularly difficult part of
      the climb, his movements precise and deliberate
Emotions: concentration and determination
Tags: athletic, climb, climber, hand
Colors: Black, Dark_Gray

Scene 7: The scene shows the climber's face, sweat beading on his forehead,
      reflecting the intensity of his effort
Emotions: grit and resolve
Tags: athletic, climb, climber, determination
Colors: Dark_Gray, Dark_Brown

Scene 8: The scene shows the man reaching for the final ledge, the clear sky
      behind him signaling the nearing end of his ascent
Emotions: anticipation and effort
Tags: athletic, climb, climber, hand
```

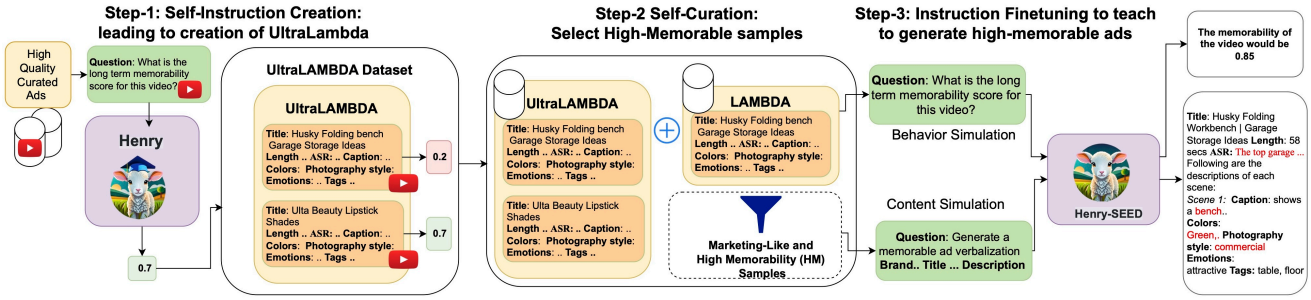


Figure 4. Overview of our SEED method for memorable ad generation. Our self-alignment consists of three steps: (i) **Self-instruction creation**: We first collect 5 million high-quality ads from YouTube, Facebook, and other mediums. Henry (trained on the complete train+test sets of LAMBDA) is then used to rate this curated set in an LLM-as-a-Judge fashion. (ii) **Self-curation**: We select marketing-like and high-memorability samples from the UltraLAMBDA and LAMBDA datasets. (iii) Instruction fine-tuning: Henry-SEED is trained on the self-curated set using two tasks: Behavior Simulation and Content Simulation.

```
Voiceover: "Impossible is not an obstacle."
Colors: Dark_Gray, Gray, Dark_Brown, Blue

Scene 9: The scene shows the man pulling himself onto the summit, his body
      silhouetted against the sky
Emotions: triumph and exhilaration
Tags: athletic, climb, climber, hand
Voiceover: "It's a chance to prove them wrong."
Colors: Dark_Gray, Dark_Brown, Blue

Scene 10: The scene shows the man standing on top of the rock with his arms
      raised in victory, overlooking the landscape below
Emotions: achievement and elation
Tags: athletic, arms up, climb, climber"
Colors: Dark_Gray, Dark_Brown
```

Listing 4. Advertisement generated by Henry-SEED for an Amazon ad. Original Ad: https://www.youtube.com/watch?v=rcFTqs-zSH4

```
Input Title: "Alexa Auto, Listening to Audiobooks" for the brand Amazon

Generated Ad: "The ad is shot in landscape orientation, at a low pace.

Scene 1: The scene shows a car driving down a country road near a mountain
      house
Emotions: anticipation, freedom
Tags: blue, car, drive, house
Colors: White, Lavender, Dark_Blue, Turquoise, Black

Scene 2: The scene shows an image of a car approaching a large sign above the
      road, signaling the start of the highway
Emotions: curiosity, readiness
Tags: car, drive, road sign, highway
Voiceover: "You just pulled onto the highway"
Colors: White, Lavender, Turquoise, Gray

Scene 3: The scene shows an in-car view with a driver's hands on the steering
      wheel, a smartphone mounted on the dashboard displaying the Amazon Music
       app
Emotions: concentration, engagement
Tags: car interior, hands, smartphone, technology
Voiceover: " but you can't remember the name of that band that's been stuck in
      your head all morning."
Colors: White, Dark_Blue, Black, Gray

Scene 4: The scene shows the driver speaking to Alexa through the car's
      infotainment system, the display lighting up in response
Emotions: interaction, ease
Tags: car interior, voice command, infotainment system
Voiceover: "With Alexa and Amazon Music, you don't have to remember. Just ask, "
      Alexa, play the Beatles."
Colors: White, Lavender, Dark_Blue, Black

Scene 5: The scene shows the car passing through a landscape with rolling hills
       and the sun beginning to set in the background
Emotions: satisfaction, serenity
Tags: car, drive, hills, sunset
Colors: White, Lavender, Turquoise, Dark_Blue

Scene 6: The final scene shows the vehicle driving into the horizon, as the
      iconic music of the Beatles fills the car's interior
Emotions: fulfillment, joy
Tags: car, drive, horizon, music"
Colors: White, Lavender, Orange, Dark_Blue
```

Listing 5. Advertisement generated by Henry-SEED for an Gatorade ad. Original Ad: https://www.youtube.com/watch?v=BKPQkjRF4yY

```
Input Title: "Gatorade | Make Your Own Footsteps with Suni Lee" for the brand
      Gatorade

Generated Ad: "The ad is shot in landscape orientation, at a medium pace.

Scene 1: The scene shows Sunisa Manado standing in front of her house
Emotions: determination, brave, confident, persistence, courage
Tags: athletic, face, girl, muscle
Voiceover: "I'm Sunisa Manado, but you can call me Sunisa. And I'm going to
      show you how I stay motivated. Let's go! This is my neighborhood. I'm
      going to take you through my favorite spots. There's my house right
      there."
Colors: Dark_Brown, Black, Brown, Tan

Scene 2: The scene shows Sunisa Manado performing a handstand in the park
Emotions: achievement, determination, persistence, commitment, success
Tags: balance, gymnast, handstand, girl
Voiceover: "And this is the park where I get so much done. This is the park
      where I train."
Colors: Dark_Brown, Dark_Blue, Purple, Gray

Scene 3: The scene shows Sunisa Manado doing a flip on the balance beam
Emotions: brave, courage, determination, persistence, inspiration
Tags: gymnast, flip, beam, girl
Voiceover: "Being an athlete takes a lot of hard work and determination."
Colors: Dark_Brown, Dark_Blue, Purple, Gray

Scene 4: The scene shows Sunisa Manado in a powerful pose in her pink sports
      bra and leotard
Emotions: determination, brave, courage, persistence, inspiration
Tags: athletic, face, girl, gymnast
Colors: Dark_Brown, Dark_Blue, Purple, Gray

Scene 5: The scene shows Sunisa Manado lifting herself on the parallel bars
Emotions: achievement, persistence, determination, courage, commitment
Tags: gymnast, lift, bars, girl
Voiceover: "And being an athlete also means that you have to have good
      nutrition."
Colors: Dark_Brown, Dark_Blue, Purple, Gray

Scene 6: The scene shows Sunisa Manado practicing her routine on the floor of
      the park
Emotions: determination, persistence, inspiration, commitment, discipline
Tags: floor, gymnast, routine, girl
Colors: Dark_Brown, Dark_Blue, Purple, Gray

Scene 7: The scene shows Sunisa Manado spinning on the balance beam
Emotions: success, brave, determination, precision, focus
Tags: beam, spin, gymnast, girl
Colors: Dark_Brown, Dark_Blue, Purple, Gray

Scene 8: The scene shows Sunisa Manado holding a bottle of Gatorade,
      rehydrating after her practice
Emotions: refreshed, healthy, satisfaction, wellbeing, recovery
Tags: athlete, drink, Gatorade, health
Voiceover: "That's why I choose to fuel up with Gatorade. You need to have that
       energy, especially during my toughest training."
Colors: Black, Dark_Green, Dark_Gray, Gray

Scene 9: The scene shows Sunisa Manado smiling towards the camera with a sense
      of pride and accomplishment
Emotions: joy, confidence, pride, optimism, accomplishment
Tags: face, smile, girl, gymnast
Voiceover: "Gatorade helps replenish what I've lost and helps keep me going."
Colors: Black, Dark_Green, Dark_Gray, Gray

Scene 10: The scene shows Sunisa Manado walking away from the park, as the sun
      sets, symbolizing the end of a day's hard work
```

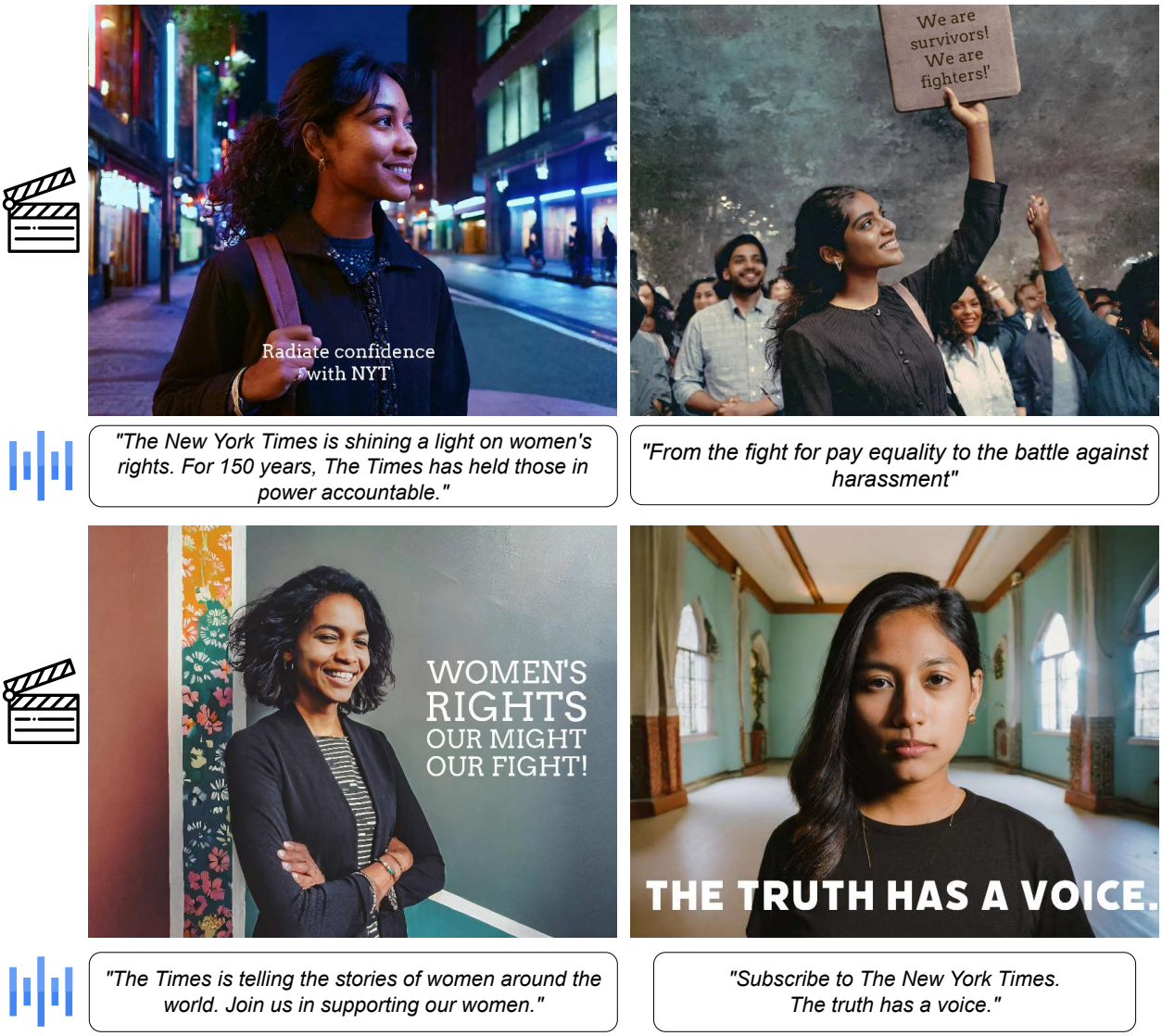


Figure 5. Henry-SEED Prompt: *Generate the detailed description of a 50 second memorable advertisement titled "Shining a Light on Women's Rights — The Truth Has a Voice — The New York Times" for the brand The New York Times* Link to the original ad: https://www.youtube.com/watch?v=bPblzhUzTeg Original memorability score: 65. Memorability score of Generated Ad: 91.

```
Emotions: contemplation, fulfillment, anticipation, resilience, hope
Tags: sunset, walk, reflection, girl
Voiceover: "I'm going to make my own footsteps. Follow your dreams, but don't
      forget to chase your greatness."
Colors: Black, Dark_Green, Dark_Gray, Gray"
```

Listing 6. Advertisement generated by Henry-SEED for a Rolex ad. Original Ad: https://www.youtube.com/watch?v=vA-Lxmh1MdQ

```
Input Title: "Rolesor A Meeting of Metals" for the brand Rolex

Generated Ad: "The ad is shot in landscape orientation, at a low pace."

Scene 1: The scene shows a close up of a watch that is on a wrist
Emotions: anticipation, luxury, elegance
Tags: dial, gold, watch
Colors: Dark_Brown, Brown, Black, Khaki

Scene 2: The scene shows a gold and silver watch sitting on a table, reflecting
      soft light
Emotions: sophistication, status, success
Tags: dial, gold, watch
Colors: Brown, Dark_Brown, Khaki, Black

Scene 3: The scene shows a person's hands elegantly wrapping the gold and
      silver watch in delicate gift paper
Emotions: care, generosity, surprise
Tags: wrapping, gift, luxury
Colors: Brown, Khaki, Gold, White

Scene 4: The scene shows a recipient's joyous reaction as they receive the
      watch as a gift at a celebratory gathering
Emotions: happiness, gratitude, celebration
Tags: emotion, celebration, gift
Colors: Gold, Silver, Cream, Warm hues, Soft lighting

Scene 5: The scene shows the watch being worn at an elegant event, the gold and
      silver glinting under ambient lighting
Emotions: pride, accomplishment, refinement
Tags: elegance, luxury, watch
Colors: Gold, Silver, Black, Soft lighting, Dark hues
```

Figure 6. Henry-SEED Prompt: *Generate the detailed description of a 18 second memorable advertisement titled "Maytag Overnight Wash and Dry" for the brand Costco*. Link to the original ad: https://www.youtube.com/watch?v=uT72lJhUUS0 Original memorability score: 76. Memorability score of the generated Ad: 83

Listing 7. Advertisement generated by Henry-SEED for an NVIDIA ad. Original Ad: https://www.youtube.com/watch?v=HnaOhOJ6AcM

```
Input Title: "Pac Man 256 on NVIDIA SHIELD" for the brand Nvidia

Generated Ad: "The ad is shot in landscape orientation, at a medium pace."

Scene 1: The scene shows a black and white logo.
Emotions: danger, tough
Tags: brand, logo
The text shown in the scene is 'SHIELD'
Colors: Black, White, Gray

Scene 2: The scene shows a black screen with a red light on it
Emotions: danger, tough
Tags: brand, logo, text
The text shown in the scene is 'SHIELD'
Colors: Black, White, Red

Scene 3: The scene shows a computer screen showing a game of Pac-Man
Emotions: nostalgia, fun
Tags: arcade, arcade machine, game, Pac-Man
The text shown in the scene is 'HIGH SCORE'
Colors: Cyan, Yellow, Gray, Dark Green, Red

Scene 4: The scene shows a video game character navigating a maze filled with
    obstacles
Emotions: excitement, challenge
Tags: arcade, character, game, maze
The text shown in the scene is 'LEVEL UP'
Colors: Cyan, Red, Pink, Orange, Gray

Scene 5: The scene shows a video game with a racing theme, cars zooming on the
    track
Emotions: thrill, speed
Tags: racing, cars, game, screenshot
The text shown in the scene is 'NEW RECORD'
```

```
Colors: Red, Yellow, Blue, Gray, Black

Scene 6: The scene shows a strategy video game with a detailed world map
Emotions: focus, strategy
Tags: strategy game, world map, empire, video game
The text shown in the scene is 'EXPAND YOUR EMPIRE'
Colors: Green, Brown, Blue, White, Tan

Scene 7: The scene shows a futuristic virtual reality game with immersive
       graphics
Emotions: amazement, innovation
Tags: VR, virtual reality, future, game
The text shown in the scene is 'ENTER NEW WORLDS'
Colors: Neon Blue, Neon Green, Black, Dark Blue

Scene 8: The scene shows a family gathered around playing a multiplayer video
       game, laughing and enjoying
Emotions: joy, togetherness
Tags: family, multiplayer, game, living room
The text shown in the scene is 'PLAY TOGETHER'
Colors: Multi-colored, Warm Hues

Scene 9: The scene transitions to a player wearing headphones, deeply
       concentrated in an esports competition
Emotions: determination, competition
Tags: esports, competition, gamer, video game
The text shown in the scene is 'RISE TO THE CHALLENGE'
Colors: Black, Neon, Dark, Spotlights

Scene 10: The scene closes with the NVIDIA SHIELD logo emerging brightly on the
        screen
Emotions: empowerment, assurance
Tags: brand, logo, closing
The text shown in the scene is 'NVIDIA. SHIELD - Game Your Way'
Colors: Red, White, Black
```

Listing 8. Advertisement generated by Henry-SEED for a Zillow ad. Original Ad: https://www.youtube.com/watch?v=P6gydrvTkaA

```
Input Title: "Find It, Tour It and Own It With Zillow (:06)" for the brand
       Zillow

Generated Ad: "The ad is shot in landscape orientation, at a medium pace.

Scene 1: The scene shows a person holding a cell phone with their hand.
Emotions: achievement, ambition
Tags: catch, smartphone, hand, person, phone, screen, text message, woman
Colors: Gray, Dark_Blue, Dark_Brown, Dark_Pink, Off_White

Scene 2: The scene shows a woman in a sweater looking at a cell phone.
Emotions: happy, smile
Tags: smartphone, earphone, girl, person, smile, woman
Colors: Gray, Dark_Brown, Dark_Pink, Dark_Blue

Scene 3: The scene shows a woman looking at a phone with a map on it.
Emotions: confident, smiling
Tags: catch, smartphone, earphone, girl, person, phone, stand, talk, woman
Colors: Gray, Dark_Brown, Dark_Pink, Dark_Blue

Scene 4: The scene shows a woman looking at a cell phone with a real estate app
       on it.
Emotions: togetherness, happiness
Tags: smartphone, earphone, girl, person, woman
Colors: Gray, Dark_Brown, Dark_Pink, Dark_Blue

Scene 5: The scene shows a woman looking at a cell phone with a real estate app
       displayed.
Emotions: happy, smiling
Tags: smartphone, earphone, girl, person, woman
Colors: Gray, Dark_Brown, Dark_Pink, Dark_Blue

Scene 6: The scene shows a woman using a cell phone to speak with an agent.
Emotions: confident, happy
Tags: smartphone, earphone, girl, person, talk, woman
Colors: Gray, Dark_Brown, Dark_Pink, Dark_Blue
Voiceover: "Getting the perfect home is a journey, so we help you find it, tour
       it, and own it."

Scene 7: The scene shows a woman completing a transaction on a cell phone.
Emotions: achievement, satisfied
Tags: smartphone, earphone, girl, person, woman
Colors: Gray, Dark_Brown, Dark_Pink, Dark_Blue"
Voiceover:  "For moving made simple, there's no place like Zillow"
```

Listing 9. Advertisement generated by Henry-SEED for a Kroger ad. Original Ad: https://www.youtube.com/watch?v=SqwqI01q3fA

```
Input Title: "How to Make Taco Seasoning | Kroger Recipes | Kroger" for the
       brand Kroger

Generated Ad: "The ad is shot in landscape orientation, at a low pace.

Scene 1: The scene shows a person pouring chipotle adobo sauce into a glass jar
       .
Emotions: care, comfort
Tags: bottle, can, container, hand, food, person, jar, liquid, pepper, pour,
       red, sauce, tomato sauce, tin, tray, woman
Colors: Black, Gray, Dark_Brown, Maroon, Dark_Red
Voiceover: "The audio in the ad says Taco seasoning is one of those spices that
       everyone loves on their food. It's so delicious and it's so easy to
       make."

Scene 2: The scene shows a woman in a red sweater adding ground cumin to the
       mix.
Emotions: anticipation, focus
Tags: blender, container, food, hand, person, ingredient, measuring cup, spice,
       woman
Voiceover: "All you have to do is get your chopstick and add some of the garlic
       and some of the onion and some of the cumin"
Colors: Black, Gray, Dark_Brown, Maroon, Dark_Red

Scene 3: The scene shows the addition of chili powder to the seasoning mix.
Emotions: precision, satisfaction
Tags: blender, container, food, food processor, hand, person, ingredient,
       mixture, spice, woman
Voiceover: " and some of the dried chilies and some of the oregano and some of
       the salt"
Colors: Black, Gray, Dark_Brown, Dark_Pink, Maroon

Scene 4: The scene shows the woman sprinkling crushed red pepper into the mix.
Emotions: excitement, zeal
Tags: blender, container, food, food processor, hand, person, ingredient,
       measuring cup, mixture, spice, woman
Voiceover: "and some of the paprika and some of the pepper and just give it a
       little bit of a squeeze"
Colors: Black, Gray, Dark_Brown, Maroon, Dark_Red

Scene 5: The scene shows the woman measuring and adding ground oregano to the
       seasoning mix.
Emotions: happiness, creativity
Tags: blender, container, food, food processor, hand, person, ingredient,
       measuring cup, mixture, spice, woman
Colors: Black, Gray, Dark_Brown, Maroon, Dark_Red

Scene 6: The scene shows the final steps of making the taco seasoning mix, with
       the woman adding garlic powder and onion powder.
Emotions: accomplishment, pride
Tags: blender, container, food, food processor, hand, person, ingredient,
       mixture, spice, woman
Voiceover: "You're going to taste it, so don't be afraid to add more or less of
       any of these ingredients, depending on how spicy or how flavorful you
       want your taco seasoning to be."
Colors: Black, Gray, Dark_Brown, Maroon, Dark_Red"
```

Listing 10. Advertisement generated by Henry-SEED for a Publix ad. Original Ad: https://www.youtube.com/watch?v=HH0VWvs-cFg

```
Input Title: "GreenWise Ice Cream Scoop from Publix" for the brand Publix
       Supermarkets

Generated Ad: "The ad is shot in landscape orientation, at a low pace.

Scene 1: The scene shows a woman standing in front of a refrigerator.
Emotions: anticipation, satisfaction
Tags: beverage, bottle, catch, doorway, drink, fill, girl, home appliance,
       juice, laugh, pour, refrigerator, shirt, stand, woman
Voiceover: "Greenwise, the better energy choice from Publix."
Colors: Black, Gray, Dark_Gray, Dark_Brown, Silver

Scene 2: The scene shows a woman holding a roll of Greenwise ice cream in her
       hand.
Emotions: curiosity, interest
Tags: bottle, catch, hand, woman
Voiceover: " All the deliciousness of ice cream, a fraction of the calories.
       And now it comes in a roll."
Colors: Black, Gray, Dark_Gray, Dark_Brown

Scene 3: The scene shows the woman as she easily scoops the ice cream onto a
       plate, displaying the convenience of the new roll format.
Emotions: ease, delight
Tags: plate, roll, scoop, serve, woman
Voiceover: " So you can easily scoop and serve exactly what you need."
Colors: Black, Gray, Dark_Gray, Dark_Brown

Scene 4: The scene shows the woman enjoying a bite of the Greenwise ice cream,
       her expression reflecting the product's deliciousness.
Emotions: enjoyment, satisfaction
Tags: bite, enjoyment, ice cream, satisfaction, taste, woman
Voiceover: "It's a win-win. Learn more at publix.com."
Colors: Black, Gray, Dark_Gray, Dark_Brown"
```

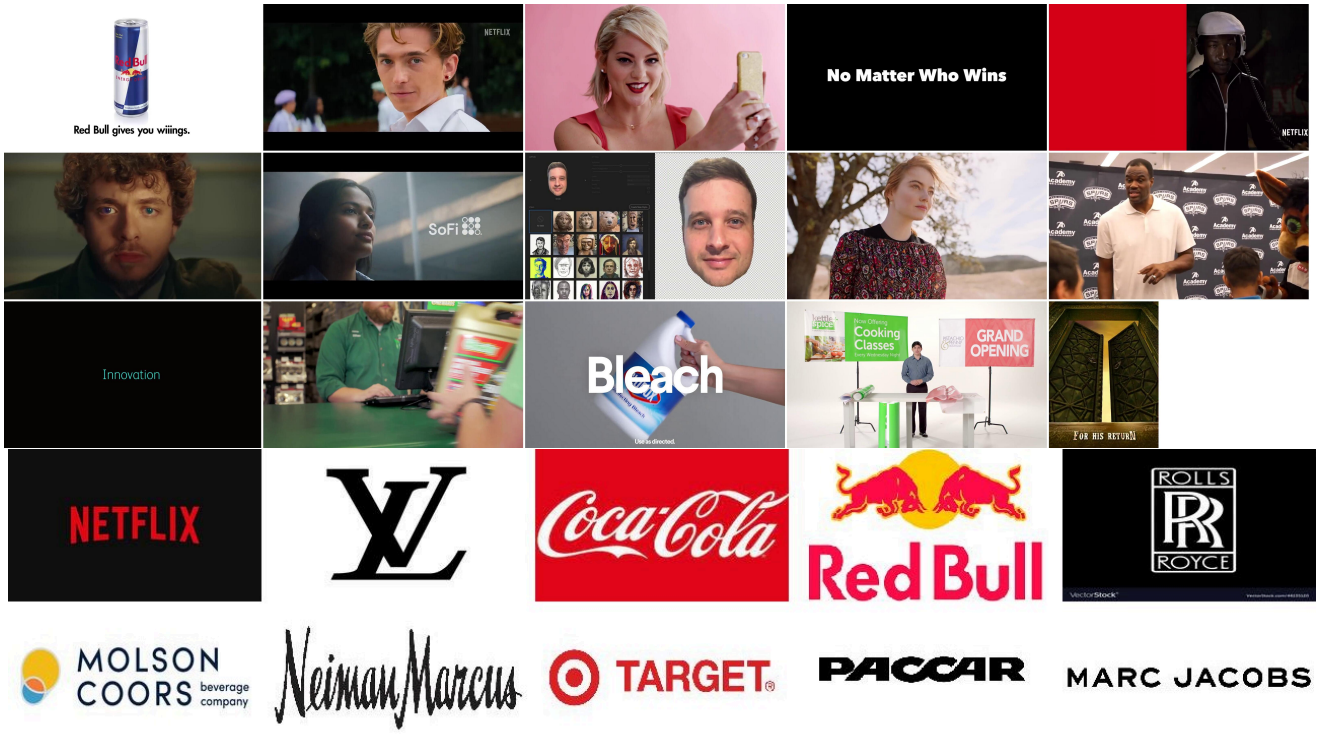


Figure 7. The top three rows show the keyframes from videos in our dataset, LAMBDA, arranged from most to least memorable. The bottom two rows show brands arranged from the most memorable brands to the least.

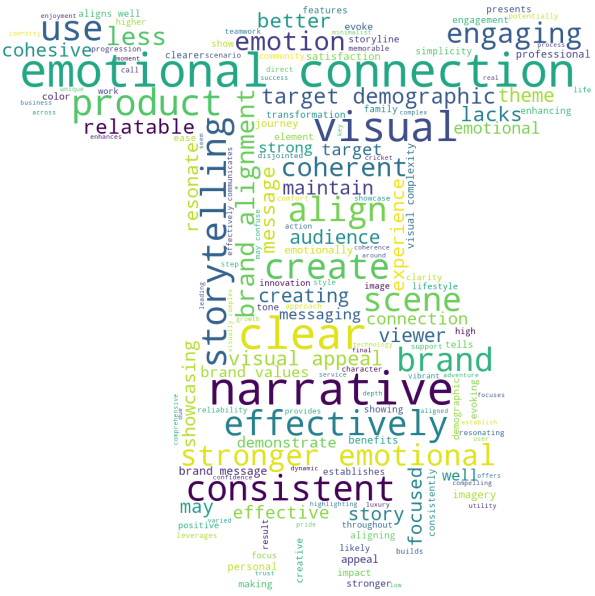


Figure 8. Word Cloud for the GPT-4 reasoning on the 75/88 generations where it rates Henry-SEED Generated Ads to be better than the Original.

B. Extraction And Use Of Cognitive And Perceptual Signals In Advertisements

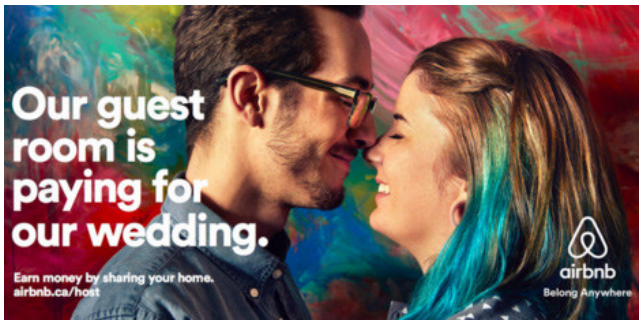


Figure 9. Airbnb advertisement showing the visual concepts of two adults, and the text “Our guest room is paying for our wedding”. “World knowledge” captured by LLMs helps identify the two adults as partners, and helps relate the text with the two adults and the Airbnb logo to infer what the ad is talking about.

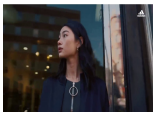
Image	Semantic Category	Verbalization	Semantic Category	Verbalization
	**OCR**	The text shown in the scene is "Adidas".	**Clutter**	The clutter in the scene is low.
	ASR	The audio in the scene is "To take hold of the world's spotlight overnight".	**Photo Style**	The photography style of the scene is commercial photography.
	Human Presence	The scene has 1 person with prominent face.	**Emotion**	The emotion of the scene is ambitious, determined.
	Caption	The scene shows a young woman sitting in a glass door looking out.	**Aesthetics**	The image has medium aesthetic value.
	Colors	The foreground colors of the scene are Black, Dark Brown, Dark Blue, Dark Gray, Mud Green and the background colors are Dark Blue, Black, Dark Brown. The dominant tone of the scene is neutral.	**Object Tags**	This scene is categorized by the tags: person, woman, blazer, facing, template, fashion, street fashion, cold, client, cardigan, sweat.
	Audio Type	The scene has music and speech.	**Logo Presence**	There is a logo in the scene.

Table 5. To augment the scene understanding of LLM, we verbalize video scenes and images using a diverse set of cognitive and perception tools and pass it to the LLM in the format shown in the table. For image memorability datasets, we use the following semantic categories: caption, color, photo style, emotion, clutter, human presence, object tags, OCR, and aesthetics. For video scene memorability datasets, we use the following semantic categories: caption, color, emotion, human presence, object tags, ASR, OCR, Audio-type, Logo-presence. We use the following models to extract the features: OCR [19], clutter [37], ASR [57], Photo style [44], human presence [46], emotion [65], caption [44], aesthetics [33], colors [56], object tags [74], audio-type [25], and logo presence [74]. Black colored text is the verbalization template, and red text indicates the model outputs.

C. Ablation Experiments

C.1. Ablation of Data and Architectural Choices on Memorability Prediction

We also conduct extensive ablations to understand the effect of different kinds of data and architectural choices. Tables 2 and Table 6 show the data ablations. We see that combining datasets actually worsens the performance across all the datasets except the SUN dataset. Further, we find that in zero-shot settings, STM helps in predicting LTM relatively much better than vice versa. This corroborates with the studies in psychology which show that for a content to get committed to LTM, it has to pass through STM [53]. Therefore, content memorable, according to STM, has an effect on LTM but, interestingly, not vice versa. Further, we observe that Henry loses performance for unseen brands. This underscores the importance of scaling the study across more brands. Next, we evaluate the impact of various architectural choices (Table 7). We find that Henry's vision branch is not strong enough by itself to produce good results. Cognitive features that were found important in our study also improve prediction performance. Low-level features like objects and colors have the maximum impact on STM, but higher-level features like emotion, ASR, and aesthetics have a higher impact on LTM.

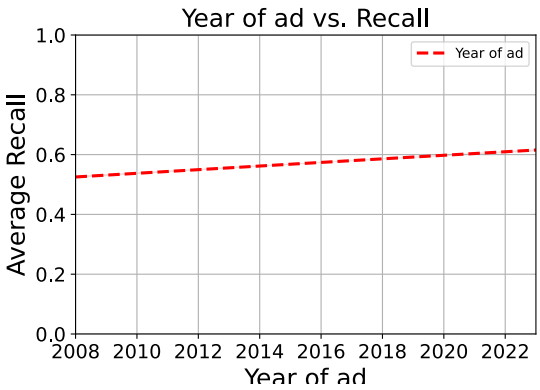


Figure 10. The graph shows the relationship of the year the ad is uploaded on youtube vs the recall.

C.2. Ablation of Data and Architectural Choices on Memorable Ad Generation

We also run some ablation studies to find the impact of the amount of data (Fig. 11) and the impact of behavior simulation and content simulation tasks (Table 8) on ad quality and memorability. A few trends are noticeable. Performance increases as the amount of data increases. Interestingly, the performance converges very slowly with the amount of increase in data. We test the performance in three conditions: brand-split, time-split, and random-split. In the brand-split testing, we leave some randomly chosen brands out of training and only test on them. For the time-split testing, we

Generalization Type	Train on	Zero-shot Testing	Lamem	Memcat	SUN	VideoMem	Memento10k	LAMBDA
Memory-type	Short-term	Long-term	-	-	-	0.31	-	0.18
Memory-type	Long-term	Short-term	0.06	0.08	0.07	0.15	0.1	-
Modality	Videos	Images	0.55	0.65	0.55	-	-	-
Modality	Images	Videos	-	-	-	0.44	0.54	0.09
Brands	All except 20 brands	Left-out 20 brands	-	-	-	-	-	0.42
Dataset	All except Memento	Memento	-	-	-	-	0.59	-
Dataset	All except Memcat	Memcat	-	0.68	-	-	-	-

Table 6. Ablation across data to understand how memorability prediction generalizes across the type of memory, datasets, modality (image/video), and brands. The reported values are correlations between model and human memorability scores. A few trends can be observed from the table: (i) STM generalizes better on LTM in zero-shot than vice versa (rows 1 and 2), (ii) Henry trained on either videos or images generalizes to both (rows 3 and 4), (iii) There is a significant performance loss in modeling memorability for brands not seen during training (row 5), (iv) Zero-shot generalization to Memento (video) and Memcat (image) is near to the current trained state of the art literature models on Memento [20] and Memcat [28] (rows 6 and 7).

	Lamem	Memcat	VideoMem(ST)	Memento10k	VideoMem(LT)	LAMBDA
Henry on individual datasets	0.74	0.82	0.64	0.75	0.48	0.55
Henry vision only	0.20	0.17	0.17	0.21	0.15	0.11
Henry language only	0.51	0.53	0.42	0.54	0.37	0.44
Henry -object tags	0.67	0.71	0.57	0.69	0.46	0.52
Henry -colors	0.65	0.74	0.55	0.67	0.45	0.51
Henry -emotion	0.71	0.78	0.61	0.73	0.42	0.46
Henry -aesthetics	0.72	0.79	0.61	0.71	0.46	0.53
Henry -clutter	0.73	0.81	0.60	0.74	0.45	0.53
Henry -asr	-	-	-	-	-	0.46
Henry -asr-emotion	-	-	-	-	-	0.42
Henry on Silent Ads	-	-	-	-	-	0.56
Henry on Ads with audio	-	-	-	-	-	0.52

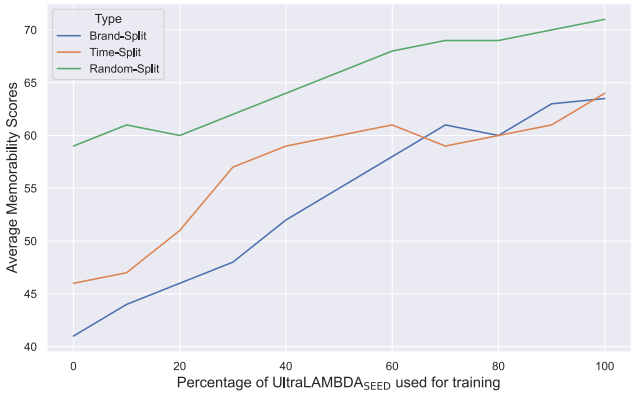
Table 7. Ablation across architectural choices. "-" denotes non-speech dataset. A few trends are visible from the table: (i) Despite having a vision branch, object tags and colors have a net positive impact on the overall performance (rows 2,3,4), (ii) For LTM (LAMBDA, VideoMem (LT)), dropping cognitive features such as emotion, aesthetics, and clutter cause a larger performance drop than dropping visual features such as objects and colors. The trend is the opposite for STM (Lamem, Memcat, VideoMem (ST), Memento10k).

Task	LAMBDA (ρ)	Δ Memorability
BS-only	0.541	-
CS-only	-	+28.41
BS+CS	0.547	+30.66

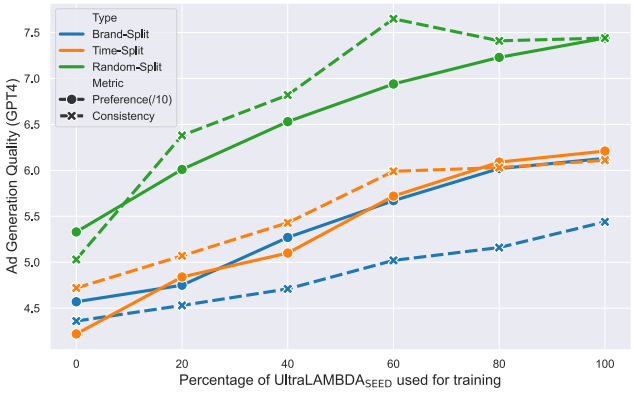
Table 8. Ablation on modeling behavior simulation (BS) or memorability prediction and Content Simulation (CS) on memorable ad generation together. For memorability prediction, we again show the Spearman rank correlation on the test set similar to Table 2; for generation, we measure the change in memorability according to Henry Oracle similar to Table 3. We observe that mixing the two tasks together increases the performance across both tasks.

put a cutoff time; we train our model before that cutoff time and test on ads after that time. For the random-split testing, we test on randomly selected advertisements. Brand-split performs worse than time-split testing, indicating that brands have a higher contribution to determining memorability. This trend is observed only in ad memorability but not in ad quality.

We also test the impact of various subsets of UltraLAMBDA on the memorability of the ads generated by Henry-SEED. Table 3 shows the results. It can be seen that adding the high memorable samples from LAMBDA train set, increases the memorability of generated ads substantially. We also train LLaVA on the complete set of 2 million UltraLAMBDA ads without filtering it via Henry assigned memory labels. Interestingly, this model, while trained on 2.5 times more data than UltraLAMBDA filtered via Henry, has lesser average memorability than it.



(a)



(b)

Figure 11. Graphs showing the importance of the amount of synthetic data on (a) Ad memorability score and (b) Ad quality for the generated ads. As we can see from the graphs, both the ad memorability and quality increase with the increase in the amount of synthetic data.

D. Questionnaire to Gather Human Preferences over Generated Ads

Below is the web-based form used to annotate the human preferences between the generated and original ad stories. Participants for this task were working professionals in the software, marketing, advertising, and creative industries. Participation was voluntary, and participants were invited to judge the efficacy of generated advertisements. Participants had a general interest in the creative and advertising industries and generative technologies; therefore, they were not interested in getting paid but rather in seeing and trying out the generative technology stack. We have a roughly 65-35 distribution of males to females with the age range between 22-50.

```
Instructions:

Shown next are 10 pairs of advertisements. Determine which ad within each pair
       is more effective based on the title, brand, and scene-by-scene
       descriptions provided. You will also be expected to provide reasons for
       your choice wherever asked.


Question 1
Choose the advertisement you find more effective. Also provide reasons for your
       choice.

Title: Bike to Work Day at NVIDIA
Brand: Nvidia
Nvidia is a technology company focusing on graphics processing units (GPUs) for
       gaming, professional visualization, data centers, and automotive
       markets, driving innovation in visual computing.

Advertisement A:

The ad is shot in landscape orientation, at a medium pace. The audio in the ad
       is silent.
Scene 1: The scene shows the camera takes a photo from the inside of the person
       on the bicycle
Colors: White, Dark_Pink, Olive, Gray, Pink, Dark_Brown
Emotions: danger, dangerous, warning
Tags: attach, bicycle, catch, smartphone

Scene 2: The scene shows the person riding a bicycle down the road
Colors: White, Dark_Gray, Mud_Green, Olive, Gray
Emotions: danger, quiet
Tags: bicycle, path, grass, motorbike
The text shown in the scene is 'NVIDIA'

Scene 3: The scene shows a man on a bike taking a ride
Colors: Off_White, Dark_Gray, Silver, Black, Gray
Emotions: danger, exciting, fun
Tags: bicycle, biker, bridge, hand
The text shown in the scene is 'DVIDIA'

Scene 4: The scene shows a bike rider going under a bridge under a road
Colors: Dark_Gray, Silver, Light_Green, Green, Olive, Gray, Bright_Green
Emotions: danger, dangerous, funny
Tags: bridge, car, curve, highway
The text shown in the scene is 'NVIDIA'

Scene 5: The scene shows a man riding a bicycle down a tree lined street
Colors: White, Dark_Gray, Mud_Green, Dark_Pink, Olive, Black, Gray
Emotions: thrill, adventure, romantic
Tags: bicycle, biker, hand, person
The text shown in the scene is 'NVIDIA'

Scene 6: The scene shows a man riding on a bicycle down the street
Colors: Emerald, Dark_Gray, Silver, Light_Green, Olive, Gray
Emotions: funky, enjoyable
Tags: bicycle, hand, person, man
The text shown in the scene is 'NVIDIA'

Scene 7: The scene shows a closeup of someone riding a bicycle down a road
Colors: White, Dark_Gray, Silver, Dark_Pink, Olive, Gray
Emotions: danger, majestic
Tags: bicycle, bicycle helmet, biker, hand
The text shown in the scene is 'NVIDIA'

Scene 8: The scene shows a person is riding a bike on the side of the road
Colors: White, Dark_Gray, Mud_Green, Olive, Gray, Lavender
Emotions: enjoy, enjoyment
Tags: car, person, man, motorcycle
The text shown in the scene is 'NVIDIA'

Scene 9: The scene shows someone riding a bike in front of a small city
Colors: White, Dark_Gray, Olive, Black, Gray
Emotions: funky
Tags: bicycle, biker, bin, car
The text shown in the scene is 'NVIDIA'

Scene 10: The scene shows a cyclist riding his bike on a gravel road
Colors: White, Brown, Mud_Green, Olive, Gray, Dark_Brown, Cyan
Emotions: recreational, adventure
Tags: bicycle, biker, hand, person

Advertisement B:

The ad is shot in landscape orientation, at a low pace. The audio in the ad is
       silent.

Scene 1: The scene shows a man wearing a hard hat holding a bike helmet
Colors: Dark_Gray, Brown, Mud_Green, Cream, Olive, Black, Dark_Brown
Emotions: protective, protective
Tags: building, construction worker, hat, jumpsuit

Scene 2: The scene shows a man riding a bike on a path near a creek
Colors: Emerald, Dark_Gray, Mud_Green, Olive, Black, Dark_Brown
Emotions: recreational, relaxation
Tags: bicycle, bicycle helmet, biker, path

Scene 3: The scene shows a man holding a bike up while standing in front of a
       building
Colors: Dark_Gray, Brown, Mud_Green, Cream, Olive, Black, Dark_Brown
Emotions: pride, achievement
Tags: building, professional, hat, bicyclist

Scene 4: The scene shows a man riding a bike down a street with trees lining
       the road
Colors: Brown, Cream, Green, Olive, Dark_Brown
Emotions: cheery, freedom
Tags: bicycle, bicycle helmet, biker, man
```

```
Scene 5: The scene shows a man riding a bike down a street in front of a house
Colors: Dark_Gray, Mud_Green, Olive, Black, Dark_Brown
Emotions: cheery
Tags: bicycle, bicycle helmet, biker, car

Scene 6: The scene shows a closeup of the man's face as he adjusts his bike
      helmet, showcasing determination
Colors: Cream, Olive, Black, Gray, Dark_Brown
Emotions: determined, prepared
Tags: man, helmet, focus, detail

Scene 7: The scene shows the man holding his bike next to other cyclists at a
      traffic light, promoting community and camaraderie
Colors: Mud_Green, Cream, Olive, Dark_Brown
Emotions: community, anticipation
Tags: cyclists, traffic light, group, waiting

Scene 8: The scene shows the man arriving at work, parking his bike in a bike
      rack
Colors: Mud_Green, Cream, Olive
Emotions: satisfaction, accomplishment
Tags: office building, bike rack, arrival, work

Scene 9: The scene shows the man walking into the building, greeting colleagues
       who are also carrying bike helmets
Colors: White, Cream, Olive, Black, Gray
Emotions: friendly, inclusive
Tags: workplace, colleagues, greeting, professional attire

Scene 10: The scene shows the man at his workstation with a helmet on his desk,
       looking out the window at the sunny day, hinting at the ride home
Colors: White, Cream, Olive, Gray
Emotions: thoughtful, accomplished
Tags: office, workstation, helmet, window


Select preferred advertisement:
Option 1: A
Option 2: B
Option 3: Both are equally effective


Give reasons for your choice:
-------------------------------------
```

D.1. Expert Feedback Collected For Generated Ads

1. Feedback for ad generation for the Maytag Ad shown in Fig 6

   (a) **Expert 1**: "I appreciate the prominent use of the logo in the advertisement. Its placement towards the end, accompanied by a compelling slogan, is in alignment with the brand's advertising strategy."

   (b) **Expert 2**: "In my opinion, the color scheme of the advertisement is stunning. It complements the tone of the advertisement exceptionally well."

   (c) **Expert 3**: "The emotional portrayal in scene 2 could be enhanced. I anticipated a sense of 'recreation' and 'relaxation' to be more effectively conveyed."

2. Feedback for ad generation for the New York Times Ad shown in Fig 5

   (a) **Expert 1**: "One noteworthy aspect in the generated ad description is the concept of 'blocking.' In the ad, the main actor is depicted moving and protesting against various backdrops, including a static background and a subtly shifting frame. This technique is reminiscent of the famous concept utilized in cinematography. While this is not in reflected in the image, I will attribute it to the image generation and not the description generation."

   (b) **Expert 2**: "I like the generated voiceover a lot in terms of story, but I find it hard to fit over the scenes, perhaps this is because the generations dont incorporate transitions/animations."

   (c) **Expert 3**: "I find the overall generated story exceptional in terms of its storytelling in a few ways. 1. The flow of the generated ad, A woman exploring nightlife, protesting, achieving, and nonetheless standing defiant. 2. The slogans are great. 3. The changing head tilt of the woman from sideways to center is a very precise details cinematographer use to paint an overall story or emotion."

3. Feedback for ad generation for the Brainly Ad shown in Fig 3

   (a) **Expert 1**: "I find the overall story formulation to be decent. It portrays kids encountering challenges in solo learning, showcasing easy accessibility and a gradual improvement in confidence and engagement throughout the story. I would still prefer a scene where the UI of the app is somehow shown to the user.[8]"

   (b) **Expert 2**: "I like the use of animated scenes, but I find the incorporation of different main characters slightly jarring. Either they should have been in a common scene, or the main character should not change with every scene. The standout feature of the ad is the utilization of color themes and their harmonization with the emotional tone of each scene."

   (c) **Expert 3**: "Having created Ed-Tech advertisements, I find the storytelling to be excellent. This ad is very persuasive, although it lacks novelty, I still find it to be effective."

E. Perplexity evaluation

A common approach to measuring language modeling performance on some data distribution D is to measure *perplexity*, which is defined as the exponential of the average negative loglikelihood per token [9, 10, 31], that is:

$$PPL = \exp\left(\frac{-1}{\sum_{j=1}^{|D|} N_j} \sum_{j=1}^{|D|} \sum_{i=1}^{N_j} \log P(y_{j_i} | y_{j_1}, \ldots, y_{j_{i-1}})\right), \tag{1}$$

where $|D|$ is the number of documents in the dataset, y_j is the j-th document in D, N_j is the total number of tokens in y_j, and y_{j_i} represents the i-th token of y_j.

[8] The generated description of the ad actually shows the student interacting with a visible UI that the image generation model could not respect properly

To calculate perplexity on a selected dataset D, each dataset document y is tokenized and fed into a language model (following the procedure described below) via computing $\log P(y|x)$, where x is set to either the empty string or a beginning-of-text token. Thus, given $\log P(y)$, for each document $y \in D$ we can sum up the per-document loglikelihoods and divide by the number of total dataset tokens.

Given our language model, we aim to compute the conditional (log) probability (or "loglikelihood") of a target string y conditioned on input x, denoted as $\log P(y|x)$. This can be performed in a single LM call.

Let $x = x_0, x_1, ..., x_{n-1}$ be an input sequence of n tokens and $y = y_0, y_1, ..., y_{m-1}$ be the target sequence of m tokens, where x_i and y_i represent individual tokens. To compute $\log P(y|x)$, we follow these steps:

1. Concatenate x and y to form a new sequence, but discard the final token y_{m-1}. The resulting sequence is $x_0, x_1, ..., x_{n-1}, y_0, y_1, ..., y_{m-2}$.

2. Pass this concatenated sequence through the language model to obtain logits l of shape $(n+m-1, |V|)$, where $|V|$ is the size of the vocabulary. The last m positions in these logits correspond to the predicted probability distributions for the target tokens y_0 to y_{m-1}, conditioned on the input x and the preceding target tokens.

3. Apply a log-softmax function to the last m logits to obtain log probabilities for the completion tokens only.

4. Calculate the conditional loglikelihood of the target string y given the input x by summing the log probabilities of each target token:

$$\log P(y|x) = \sum_{i=0}^{m-1} \log p(y_i|x, y_0, ..., y_{i-1}) = \sum_{i=0}^{m-1} l(n+i, y_i), \tag{2}$$

where $\log p(y_i|x, y_0, ..., y_{i-1})$ is the log probability of the i-th target token conditioned on the full input x and the preceding target tokens. (and where $x, y_0, ...y_{-1}$ denotes conditioning on only x.)

We follow the above procedure to calculate perplexity over three equally divided parts of the dataset, *i.e.*, 33-percentile cuts where samples are ranked as per their memorability values. The lower the perplexity of an LLM over a category of samples, the better it is at generating those samples. Therefore, for example, if an LLM has a lower perplexity over high memorable samples, it is easier for it to generate highly memorable samples than lower memorable ones.

F. Annotation Protocol and Participant Details for the LTM Study

Figure 12 shows a visualization of the annotation protocol we followed.

The participants in the study were students who were offered optional course credit and freebies like eatables and a chance to see research and know their memorability scores. The participation was voluntary. The students were shown a protocol of the study and were required to sign the IRB approval, which was prominently displayed. The approval contained details about what kind of data was being collected and how the data would be used. The data collection protocol was approved by the IRB of the participating institution. The aggregate statistics were reported to each candidate after completing the study. Three emails were sent to take-home participants; if they didn't reply within the given time frame, their data was discarded from the experiment.

The participants were primarily graduate and undergraduate students. The participants are from two universities spread across two locations in India. The participants are bilingual and speak a variety of languages, including English. The age range is from 16 to 35 years, and all genders/sexes are encouraged. We saw a roughly 30-70 distribution of females to males.

F.1. Memorability Questionnaire

This section contains the questions we asked before the study, the attention check questions that were asked during the study, and finally, the recognition questions to check which brands were remembered.

F.1.1 Introductory Questionnaire (to be filled before the study starts)

1. I remember seeing ads for the following brands this year:
   - List 15 randomly selected from the list of brands that we have
2. I remember using products of the following brands this year:
   - List 15 randomly selected from the list of brands that we have (non-intersecting list from above)
3. Have you installed any Ad Blocking software in your browser(s)?
   a. Yes
   b. No
4. Do you use a Youtube subscription?
   a. Yes

Figure 12. The study protocol we followed for our long term memorability human study. All the previous works follow a game-like annotation protocol, where the study participants compete with each other to get best memorability scores and a participant is excluded from the study if their annotations fall below a certain threshold. We follow a more natural way in which participants fill an initial questionnaire, then watch 10 ads with attention checks on day 1 and in subsequent days, receive a form asking them to fill in what do they remember seeing. Further, using Stable Diffusion, we also ask them to recreate the scenes they remember.

b. No

5. Approximately how much percentage of time do you spend on Youtube mobile vs Youtube web?
   - <10% on mobile
   - >10% but <30% on mobile
   - >30% but <70% on mobile
   - >70% on mobile
6. How do you apprise yourself of the latest products and brands? (Multi correct)
   - Primarily friends and family
   - Amazon, Flipkart or any other e-commerce stores
   - Television and OTT Platform Ads (like Youtube, Netflix, Hotstar, etc)
   - Email Ads
   - Store Visits
   - Website Ads
   - I primarily search for products

F.1.2 Checks (to be answered during the experiment)

1. **Attention check**: A factual question like, What is the capital of India? (Asked randomly between videos, needs to be answered in <10s)
   a. Kanpur
   b. Delhi
   c. Goa
   d. Mumbai
2. **Consistency Check**: Do you remember watching this video in this experiment (Asked after showing the 11th video)
   a. Yes
   b. No

F.1.3 Recognition Questions (asked after a few days after watching the videos)

1. In the study, I remember seeing Ads of the following brands:
   - (Randomly selected list of 20 brands which contains the brands shown to the participant)
   - {For each brand in the list which the participant has selected}
2. Brand: X (already filled in)
   - For the {brand} ad, I remember seeing the following (Write Scene Descriptions, feel free to write any scenes, music, characters, emotions, objects you remember seeing):

G. Collection of all the Prompts used in the Paper

G.1. GPT-4 Prompts

Listing 11. GPT-4 Prompt to calculate preference between Real Ad (A) and Generated Ad (B)

```
As a seasoned marketer, evaluate the effectiveness of the following two ads
      using a comprehensive set of metrics:

Creativity and Innovation: Originality and uniqueness in conveying the message.
      Use of unexpected ideas or elements that grab viewers' attention.

Emotional Connection: Ability to evoke strong, relevant emotions in the target
      audience. Establishing a connection between the brand and the viewers'
      emotions.

Storytelling: Crafting a compelling narrative that engages and retains the
      audience. Creating a memorable experience through a coherent and
      impactful story.

Visual Appeal: Use of strong visual elements, such as striking visuals, colors,
      and graphics. Ensuring that the visual elements align with the overall
      message and brand image.

Brand Alignment: How well the ad aligns with the values, mission, and
      personality of the brand. Consistency with the brand's visual identity,
      tone, and messaging. The ad's ability to leave a lasting impression on
      viewers regarding the brand. Incorporating brand elements that make it
      easy for the audience to remember and recognize.

Target Demographics: Relevance of the ad content and message to the target
      audience. Appropriateness of visuals, language, and themes for the
      specific demographic group.

Based on these criteria, analyze and determine which of the two ads is more
      effective. I will provide you with the Voiceover, followed by their
      scene-by-scene descriptions, including the emotions shown in the scene,
      the text, objects, colors, and style of the image.
```

```
Ad (A): {Verbalization for Ad (A)}

Ad (B): {Verbalization for Ad (B)}

Give me your answer in a json format, with the following keys:
- ad_a_score: Score between 0 and 10 for Ad A
- ad_b_score Score between 0 and 10 for Ad B
- winner The winner of the two ads
- reason line separated Reasons for the winner in not more than 3 lines
```

Listing 12. GPT-4 Prompt to measure consistency of an Ad

```
You are now a seasoned marketer that judges the consistency of an advertisement
      well. The consistency of an Ad can be determied by a few metrics (in no
      particular order) such as:
1. Does the voiceover match with the Scenes in the Ad?
2. Do the scene description make a good story?
3. Are the emotions depicted in the scenes consistent with the overall ad?
4. Does the ad represent the product and the brand well?

Rate the consistency of the following ad out of 10. Give me the rating only and
      nothing else, or you will be penalized.
{Advertisement Description}
```

Listing 13. GPT-4 Prompt to generate ad verbalization with In-Context-Learning (ICL)

```
You are now a seasoned marketer that creates memorable ads given its duration,
      brand and title.
Your output should follow the writing style of the input exactly. For example,
      each scene should look like:
The scene shows {}. The foreground colors of the scene are {}, and the
      background colors are {}. The dominant tone of the scene is {}. The
      photography style of the scene is {}. The scene has {} visual complexity
      . The emotions shown in the scene are {}. This scene is categorized by
      the tags {}.
You are only supposed to fill in the {}

Generate the detailed description of a {DURATION_AD1} second memorable
      advertisement titled "{TITLE_AD1}" for the brand {BRAND_AD1}
Generate the detailed description of a {DURATION_AD2} second memorable
      advertisement titled "{TITLE_AD2}" for the brand {BRAND_AD2}
...
Generate the detailed description of a {DURATION_AD5} second memorable
      advertisement titled "{TITLE_AD5}" for the brand {BRAND_AD5}
Generate the detailed description of a {DURATION_TARGET} second memorable
      advertisement titled "{TITLE_TARGET}" for the brand {BRAND_TARGET}
```

G.2. Henry Prompts

Given below are the verbalization templates we use to teach Henry and Henry-SEED behavior simulation and content simulation tasks:

Listing 14. Verbalization pattern to predict memorability given advertisement. The same template is used to prompt GPT-3.5, GPT-4, Henry, Henry-Oracle, and Henry-SEED. Note that video tokens are optional.

```
Students are shown ads and their
    memorability is tested after 1 to 3 days
    . For the given ad:
<video> .. </video>
They watch a 15 second advertisement for
    Chanel.
The title of the advertisement is " Comes
    in Red for a Limited Edition CHANEL
    Fragrance".
The ad is shot in landscape orientation, at
     a medium pace.
The audio in the ad says: Number 5. Limited
     Edition. Chanel.
Following are the descriptions of each
    scene:
     Scene 1:
         The scene shows a red bottle of
             perfume that is on a dark
             surface.
         The foreground colors of the scene
             are Black, and the background
             colors are Dark_Brown,Maroon,
             Black,Gray.
         The dominant tone of the scene is
             neutral.
         The photography style of the scene
             is product.
         The scene has Low visual complexity
             .
         The emotions shown in the scene are
              gift, romantic, celebration.
         This scene is categorized by the
             tags bottle, man, perfume, red,
             woman.
         The text shown in the scene is 'N5
             ', 'CHANEL', 'PARIS', 'PARFUM'
         ....
What would be the memorability score of
    this video?

Output: 71
```

Listing 15. Henry Prompt to generate ad verbalization used to train and evaluate Henry-SEED

```
Generate the detailed description of a {DURATION_TARGET} second memorable
      advertisement titled "{TITLE_TARGET}" for the brand {BRAND_TARGET}
```

G.3. Mistral prompt for filtering marketing ads

Listing 16. Mistral Prompt for Ad Filtering

```
"Based on the topic_tags_vocab = {'politics': 'The art and science of governing
      societies and making decisions that affect collective interests.', '
      marketing': 'The process of promoting, selling, and distributing
      products or services to consumers, often involving market research,
      advertising, and branding strategies.'} provided, please identify the
      top most relevant topic tag from the topic_tags_vocab keys that
      represent the following advertisement based on content and page_name.
      Please use only the most relevant tag and make sure to choose from
      provided topic tags only. Do not include any other tags not mentioned in
      the prompt.Answer with the most relevant topic tag only. The
      advertisement is posted by the page Donald J. Trump and has the
      following content : ['President Trump is coming to town! Get your free
      tickets now >>>']. Answer in only politics or marketing."

cleaned_text = "The advertisement is posted by the page {page_name} and has the
      following content : {page_content}"
```

H. Computing Infrastructure and Hyperparameters

H.1. Modeling Memorability

All the experiments were conducted on 8x40 A100 instances. All experiments were performed leveraging DeepSpeed ZeRO stage-3 with cpu offload [58–60] and Flash-attention [18] with gradient-checkpointing [14] at bf16 precision. We use AdamW as the optimizer (with fused gelu), the learning rate was kept 2e-5 for all experiments. The

maximum context length for image-only datasets is 500, including public video datasets is 800 and including our dataset is 2048. The corresponding batch sizes are 32,16,8. The gradient accumulation is set to 1 and weight decay is disabled. The warmup steps are set to 20 and residual dropout was kept at 0.25. We train all models for two epochs, but use the checkpoint with best validation spearman correlation.

For all experiments, where we combine datasets, we use a custom sampler to account for dataset imbalance, that ensures a maximum proportion of the dataset in an epoch, here are the maximum proportions. For validation we take 5% of each dataset. We use the provided test splits for public datasets and we use a 15% test split for our dataset

H.1.1 Images

1. **Lamem** 50%
2. **Memcat** 100%
3. **SUN** 100%

H.1.2 Videos

1. **VideoMem** 75%
2. **Memento** 75%
3. **AdsData** 100%
4. **MediaEval** 100%

H.2. Generating Memorable Ads

All the experiments were conducted on 8x80 A100 instances. All experiments were performed leveraging DeepSpeed ZeRO stage-2, Flash Attention and Gradient-Checkpointing. $\alpha = 0.001$, awac_scale$= 1$, $\gamma = 0.99$, $\beta = 0$ cql_scale$= 0.1$

H.2.1 Inference hyperparameters

$\beta = 4$, temperature$= 0.8$, steps_for_target_sync 10, $\tau = 0.7$, two_qs: True, lr=1e-5

I. License and Terms of Release

LAMBDA and UltraLAMBDA are sourced from brand videos from YouTube, Facebook Ads, and CommonCrawl. The dataset annotations and video links contained in LAMBDA and UltraLAMBDA will be released under CC BY-NC 4.0 license. The videos themselves are released as per their creators' licenses. The videos or the released data do not contain or disclose any identities of their annotators or any specific persons. Since it is handcrafted, LAMBDA makes sure that none of the videos are offensive; UltraLAMBDA being sourced from the internet is noisier. While the videos themselves originate from brands, the content of some brands may seem offensive to certain people.

We used Llama, GMHRA, ViT, EVA-CLIP, and Qformer models in accordance with their licenses to train Henry.

J. Limitations and Potential Risks

In this paper, we try to fill a gap in the existing literature about long-term memorability modeling and datasets. Therefore, we conduct the first study for that purpose. While doing that, we have made initial efforts starting with the English language advertisements. Future work would be needed to address other languages. Further, given the limitations of the study, we conducted it in an academic environment with a student population consisting of undergraduate and graduate student volunteers. We will expand the scope to a wider audience in the future work. We trained a model, Henry, on the collected dataset, showing good performance on all literature datasets. However, since the literature datasets are all English-based and deal with a majorly uniform population, the training will be scaled to more languages and population types in future work. We also observed a decrease in performance for brands not seen during the training and for videos with longer verbalizations exceeding 1500 tokens. Additionally, the model exhibits a slight inaccuracy when advertisements have significant musical content. In our opinion, the model does not pose any potential risk or harm besides the limitations mentioned here. We also conduct a review of the generated ads through experts and non-expert annotators. Both experts and non-expert annotators preferred Henry-SEED generated ads 3/5 times.